INFERENCIA ESTADÍSTICA PARA ECONOMÍA Y ADMINISTRACIÓN DE EMPRESAS

JOSÉ M. CASAS SÁNCHEZ

EDITORIAL CENTRO DE ESTUDIOS RAMÓN ARECES, S. A.

ÍNDICE

Capítulo 1
MUESTREO Y DISTRIBUCIONES EN EL MUESTREO

1.1. INTRODUCCIÓN

Anteriormente hemos estudiado conceptos fundamentales, como eran el concepto de variable aleatoria y su distribución de probabilidades, estudiamos diferentes modelos de distribuciones tanto de tipo discreto como de tipo continuo y analizábamos sus características básicas (media, varianza, etc.). A partir de ahora estaremos interesados en saber qué modelo sigue la población, y para ello nos basaremos en la información que se obtenga de un subconjunto o parte de esa población que llamaremos **muestra**.

Cuando realizamos una introducción general de la estadística decimos que uno de los objetivos fundamentales es el obtener conclusiones basándonos en los datos que se han observado, proceso que se conoce con el nombre de **inferencia estadística,** es decir utilizando la información que nos proporciona una muestra de la población se obtienen conclusiones o se infieren valores sobre características poblacionales.

En este capítulo daremos una serie de conceptos básicos que serán fundamentales para el desarrollo posterior de la inferencia estadística.

1.2. MUESTRA ALEATORIA

Sabemos que hay diferentes métodos para investigar u observar una población (observación exhaustiva o censo, subpoblación, muestra y observación mixta), aquí nos vamos a referir a la observación parcial mediante una **muestra** y diremos que se ha investigado la población a partir de una muestra cuando los elementos que componen la muestra no reúnen ninguna característica esencial que los diferencie de los restantes, representando, por tanto, a toda la población. Las conclusiones sacadas de la muestra se pueden inferir o extender a la población total. Así por ejemplo, supongamos que deseamos conocer el precio medio o valor medio de las viviendas en una zona de Madrid en el año 1994. Para conocer la característica precio de la vivienda en esa zona, necesitaríamos saber el precio de venta de cada una de las viviendas vendidas durante ese período de tiempo y el precio por el cual cada propietario vendería la suya. Esta lista completa de viviendas con sus precios, constituye la población en la que estamos interesados, cuya característica, precio medio de la vivienda o media poblacional, deseamos conocer. Pero, en esta y en otras muchas situaciones prácticas no será posible o no será fácil, por diversas razones el obtener la población entera en la cual estamos interesados. Sin embargo, si podemos obtener la información necesaria, precio de la vivienda, para una muestra representativa de la población y a partir de la cual inferir y obtener conclusiones para toda la población total.

La muestra debe de ser representativa de toda la población y, por tanto, tendrá características similares a las que se observarían en la población entera, de tal manera que si observando los precios de las viviendas que han sido incluidas en la muestra resulta que el precio medio de las viviendas de la muestra, media muestral $\bar{x}$, ha resultado ser 8.970.540 ptas. podremos inferir que la media poblacional precio medio de la vivienda en toda la población o zona que estamos considerando está en torno a 8.970.540 ptas.

La razón principal para investigar una muestra en lugar de la población completa es que la recogida de la información para toda la población daría lugar a un coste muy elevado tanto en recursos económicos como en tiempo. Incluso en ciertos casos en que los recursos fueran suficientes para investigar la población completa, puede ser preferible el investigar sólo una muestra muy representativa, concentrando sobre ella un mayor esfuerzo para obtener medidas más precisas de las características que nos interesen. De esta forma se puede evitar lo que algunas veces ocurre en las grandes operaciones censales, por ejemplo, en el censo decenal de población de los Estados Unidos, en donde se investigó toda la población, se observó que ciertas características y grupos poblacionales estaban muy poco representados, lo cual era debido a la problemática que lleva consigo una gran operación censal, tanto por el volumen de cuestionarios como por la cantidad de información.

Cuando se selecciona una muestra de una población, un objetivo fundamental es el poder hacer inferencias sobre características poblacionales u obtener conclusiones que sean válidas para toda la población. Por tanto, es muy importante que la muestra sea representativa de la población; así pues la calidad de la inferencia o conclusión obtenida a partir de la muestra, sobre las diferentes características poblacionales estará directamente relacionada con la representatividad de la muestra. Por ejemplo, supongamos que un director comercial desea conocer la opinión sobre un nuevo producto de limpieza. No sería correcto que limitara la correspondiente encuesta a sus amigos y a las personas que viven en su barrio, pues tales personas no reflejarían la opinión de toda la población ya que la muestra no sería representativa de toda la población, ni aleatoria. Para evitar estos problemas y poder realizar una inferencia correctamente sobre toda la población a partir de una muestra es necesario que se verifique la **representatividad** y la **aleatoriedad** de la muestra.

Un objetivo básico en muestreo es seleccionar una muestra que garantice con un costo razonable una buena representatividad de la población.

El procedimiento de selección de la muestra puede conducir a diferentes tipos de muestreo, como veremos al estudiar el muestreo en poblaciones finitas. Aquí nos vamos a referir a un solo tipo de muestreo, aunque inicialmente consideremos dos:

— muestreo con reemplazamiento, y

— muestreo sin reemplazamiento.

El **muestreo con reemplazamiento** consiste en seleccionar, por mercanismos aleatorios, los elementos de la población que entran a formar parte de la muestra, pero de tal manera que cuando se observa la característica, que estamos investigando, del primer elemento seleccionado, se devuelve el elemento a la población, se selecciona el segundo elemento entre todos los elementos de la población, se anota la característica que se está investigando y se devuelve a la población, y así sucesivamente. Este procedimiento permite que un elemento de la población pueda ser seleccionado en más de una ocasión para formar parte de una muestra, pues la selección se realiza con reemplazamiento, es decir, con devolución del elemento seleccionado a la población.

En el **muestreo sin reemplazamiento,** los elementos de la población que entran a formar parte de la muestra también se seleccionan aleatoriamente, pero después de observar la característica que estamos investigando no se devuelve el elemento de nuevo a la población, con lo cual no pueden volver a ser seleccionados como ocurría en el muestreo con reemplazamiento.

Así pues, si tenemos una población de N elementos y queremos seleccionar una muestra de tamaño n resulta que la probabilidad de que un elemento de la población sea seleccionado en la primera extracción para formar parte de la muestra será $\frac{1}{N}$, en ambos tipos de muestreo. Sin embargo, en la selección del segundo elemento las probabilidades son diferentes, pues en el muestreo con reemplazamiento continúa siendo $\frac{1}{N}$, ya que el número de elementos de la población sigue siendo N, pero en el muestreo sin reemplazamiento el tamaño de la población es $N - 1$, pues el primer elemento seleccionado no se devuelve a la población y entonces la probabilidad de seleccionar un elemento concreto será: $\frac{1}{N-1}$. Vemos pues que en el muestreo con reemplazamiento la probabilidad de seleccionar uno a uno los n elementos de la muestra permanece constante y en el muestreo sin reemplazamiento no sucede lo mismo ya qe en cada extracción no se devuelve el elemento a la población y ésta va disminuyendo a medida que se selecciona la muestra, siendo los tamaños poblacionales N, $N - 1$, $N - 2$, ..., $N - (n - 1)$.

Luego, la probabilidad de seleccionar una muestra concreta de n elementos será:

	1.ª extracción		2.ª extracción		n.ª extracción
Muestreo con reemplazamiento	$\frac{1}{N}$	$\cdot$	$\frac{1}{N}$	$\cdot \cdots \cdot$	$\frac{1}{N}$
Muestreo sin reemplazamiento	$\frac{1}{N}$	$\cdot$	$\frac{1}{N-1}$	$\cdot \cdots \cdot$	$\frac{1}{N-n+1}$

Si el tamaño de la población es infinito o muy grande, entonces el tamaño de la muestra n en comparación con ese tamaño N infinito o muy grande de la población es prácticamente despreciable, y entonces no existe diferencia significativa entre ambos tipos de muestreo. No sucede lo mismo cuando el tamaño N de la población es finito, dando lugar a tratamientos diferentes, como veremos en el capítulo dedicado al muestreo en poblaciones finitas.

Para llegar al concepto de **muestra aleatoria simple** y poder dar una definición rigurosa de la misma consideramos una población, cuya función de distribución es $F(x)$, constituida por un número infinito de posibles valores de una característica medible X, esta característica puede ser, por ejemplo, el tiempo de espera para recibir un servicio, o el valor de las ventas de un determinado producto. Entonces para seleccionar una muestra aleatoria de tamaño n de esta población se diseña un experimento, de tal manera que la primera realización de ese experimento nos proporciona la observación X_1 de la característica medible X, repitiendo sucesivamente el experimento bajo las mismas condiciones, para todos los factores controlables, tendremos las n observaciones:

$$X_1, X_2, ..., X_n$$

que constituyen la muestra aleatoria simple.

Cada observación X_i correspondiente a la repetición i-ésima del experimento es una variable aleatoria cuya distribución de probabilidad es idéntica a la de la población de la característica X, para todo $i = 1, 2, ..., n$.

Si la población consta de un número finito de elementos, por ejemplo, personas, viviendas, establecimientos comerciales, etc., y realizamos un **muestreo aleatorio con reemplazamiento,** es decir, se selecciona aleatoriamente un elemento de la población, se observa la característica medible que estamos investigando y esta observación sería la X_1. Se devuelve el elemento a la población, después se selecciona un segundo elemento y observando la característica medible tendríamos la observación X_2. Reiterando el proceso n veces tendríamos las n observaciones:

$$X_1, X_2, ..., X_n$$

de la característica medible X de la población, que constituyen la muestra aleatoria simple.

Cada una de estas observaciones, $X_1, X_2, ..., X_n$, también es una variable aleatoria cuya función de probabilidad es idéntica a la de la población, ya que cada selección de un elemento que da lugar a una observación procede de la población original.

Luego las observaciones $X_1, X_2, ..., X_n$ constituyen un **conjunto de variables aleatorias independientes e idénticamente distribuidas** ya que como la selección se ha realizado con reemplazamiento, ninguna observación se ve afectada por otra, es decir, el hecho de que una observación sea seleccionada no depende en absoluto de las que se han seleccionado anteriormente, pues los elementos se devuelven a la población después de anotar la característica a investigar, y la probabilidad de selección permanece constante.

Si en la población con un número finito de elementos, se seleccionan análogamente n elementos **sin reemplazamiento** tendríamos una muestra aleatoria sin reemplazamiento de observaciones:

$$X_1, X_2, ..., X_n$$

de la característica X que estamos investigando.

Estas observaciones, $X_1, X_2, ..., X_n$ también son variables aleatorias cuyas funciones de probabilidad son iguales a las de la población inicial[1], pero las

[1] Se puede demostrar que aunque la selección de las observaciones muestrales se hace sin reemplazamiento, la función de probabilidad no condicionada de las observaciones X_i, es idéntica a la función de probabilidad de la población, para $i = 1, 2, ..., n$.

observaciones no son independientes como ocurría en el caso del muestreo aleatorio con reemplazamiento, y por tanto, no constituyen una muestra aleatoria simple.

En consecuencia, a partir de ahora nos vamos a referir a poblaciones de tamaño infinito o muy grandes, de tal manera que no haremos distinción ni referencia alguna a que el muestreo sea con reemplazamiento o sin reemplazamiento pues la diferencia existente entre ambos será irrelevante para nuestro estudio. No obstante hemos de tener en cuenta que si el tamaño N de la población es finito y realizamos un muestreo con reemplazamiento entonces le daremos el mismo tratamiento que si la población fuese de tamaño infinito, pues como hemos visto también dan lugar a un conjunto de variables aleatorias independientes e idénticamente distribuidas, es decir, a muestras aleatorias simples. Una **muestra aleatoria simple** de tamaño n de una población X está constituida por un conjunto de n-variables aleatorias $X_1, ..., X_n$ independientes e idénticamente distribuidas a la población X, es decir está constituida por un conjunto de observaciones muestrales independientes e idénticamente distribuidas.

Cuando el experimento se realiza, a cada una de las variables aleatorias se le asignará un valor numérico. Es decir, tendremos la **realización de la muestra**

$$X_1 = x_1, ..., X_n = x_n$$

y diremos que ha sido seleccionada una muestra.

En la práctica la muestra se suele obtener utilizando una tabla de números aleatorios, Tabla 1.1, la cual constituye un ejemplo de cómo son estas tablas, pues están constituidas por muchas páginas de números obtenidas aleatoriamente y colocadas en ese formato.

TABLA 1.1. *Número aleatorios.*

Fila	Columna									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	2034	5600	2400	7583	1104	8422	9868	7768	2512	9575
2	8849	5451	8504	3811	0132	8635	1732	4345	9047	0199
3	8915	2894	5638	4436	9692	8061	4665	9252	6729	9605
4	6989	0682	0085	5906	8542	6884	5719	5081	8779	9071
5	5093	8880	3466	0212	9475	4957	8474	8580	9572	6770
6	7940	3305	1183	8918	4397	3167	7342	7780	6745	4688
7	9808	7499	9925	0695	4721	7597	0922	4715	6821	2259
8	5667	7590	8599	5032	3042	3666	1160	3413	2050	1796
9	0644	2848	7347	7161	6813	8276	8175	6534	6107	8350
10	4153	0293	0882	9755	5109	1484	4798	8039	3593	6369

TABLA 1.1. *(Continuación)*

Fila	Columna									
	1	2	3	4	5	6	7	8	9	10
11	4621	0121	0251	9783	7697	4079	8952	4884	8838	1587
12	8490	4941	5203	2932	1008	6544	1137	1018	5123	0347
13	3160	4107	2194	1314	1310	7060	3075	5273	6592	8875
14	0140	1600	8468	6585	5257	4874	9097	8684	7877	8881
15	0483	7097	5973	4235	7466	0821	3261	1359	3706	4676
16	2657	1386	6896	3132	2648	8947	9518	7472	9285	3067
17	4286	7434	3848	9128	5350	0407	6215	4059	4546	5170
18	8445	5087	0964	2800	9369	1680	8490	7760	7548	1060
19	4943	4327	0966	7861	8381	5865	4447	9063	2085	3635
20	9786	8853	0667	9100	2303	4455	0389	6145	2618	5401
21	8784	7084	9040	2542	8494	5685	2677	0233	5952	5003
22	3359	8450	6832	2107	4014	7988	6136	8709	0673	9533
23	0880	9436	5314	5364	6338	4492	9529	4216	9564	1936
24	6835	7154	9725	6945	5068	6102	4967	5868	3672	0240
25	3257	1159	1117	1878	7482	4219	4079	1215	7278	5363
26	5919	6661	9142	8506	2159	7663	4231	9216	0181	4264
27	6174	5167	6056	2688	4973	8345	7830	0483	8133	9319
28	4341	0850	5955	2510	7506	8247	2714	2732	3351	0755
29	8108	3203	3206	5417	4958	8194	1472	7316	5006	4701
30	3917	6944	3774	5552	7344	0332	8616	5099	6335	5902
31	6155	7292	8273	6565	5251	6444	6572	4532	9120	3635
32	3695	8581	6196	7053	0808	1319	7161	2077	7051	1663
33	7094	3242	8861	2713	0739	6480	8261	3716	1115	2739
34	2770	1314	9206	6758	8199	5785	5115	7492	8320	1888
35	1541	9360	9483	1700	2519	4393	5989	2669	4385	5264
36	6714	6622	4753	7661	4173	0740	2112	1204	5611	6928
37	1538	8256	5467	1631	6954	2451	0388	1597	0692	2061
38	6185	3229	9022	2250	7568	4105	7855	4253	3408	9607
39	5861	8254	4408	6331	5158	3646	4213	1484	4749	1772
40	0188	2813	3713	6040	8213	3180	4207	1758	5029	3479
41	5984	5650	5334	0834	0568	9593	7276	2242	2692	0774
42	7575	1609	0006	4395	6678	0966	7304	4884	6223	5828
43	2378	3307	7449	7135	9320	5201	2429	8007	3009	8697
44	3984	0440	9423	3348	1581	1283	4426	1156	0875	7948
45	1588	9349	9377	3424	2191	9673	0955	9909	5854	0890
46	1672	2549	5555	2342	5362	7581	4545	1364	8223	3789
47	4303	1048	9157	0308	9748	4780	8206	0071	9129	6210
48	2718	7400	2207	3906	7271	5283	2421	6430	4418	1688
49	3331	4479	5609	6361	6905	3507	7438	6188	1971	0925
50	7593	6059	1417	2254	1813	8232	0652	9408	9972	5511

Una muestra aleatoria simple de tamaño n de una población de tamaño N puede ser obtenida de la siguiente forma: se enumeran los miembros de la población de 1 a N. Se elige arbitrariamente un lugar en la tabla de números aleatorios, por ejemplo, la fila 2, columna 4, y como cada columna consta de cuatro dígitos, resulta que nos situaríamos en el 3811 y avanzando por filas o por columnas, lo haremos por filas, seleccionaremos los n primeros números distintos que nos encontremos entre 1 y N, que en este caso serían:

3811, 0132, 8635, 1732, 4345, 9047, 0199, 8915 ...

Estos números entran a formar parte de la muestra aleatoria simple. Observemos que es un muestreo sin reemplazamiento.

En este caso estamos suponiendo que N es como máximo 9999 pues los números aleatorios aparecen agrupados en bloques de cuatro dígitos, pero se podían haber agrupado en bloques de cinco dígitos como ocurre en la Tabla A.12 de números aleatorios, que aparece en el anexo A de tablas.

Ejemplo 1.1

Consideremos la población formada por los 100 alumnos de una clase de segundo curso de Económicas cuyas edades aparecen en la Tabla 1.2.

TABLA 1.2. *Edades de los cien alumnos de una clase de segundo de Económicas.*

Alumno	Edad	Alumno	Edad	Alumno	Edad	Alumno	Edad
1	20	26	20	51	20	76	21
2	19	27	19	52	20	77	19
3	20	28	19	53	20	78	19
4	19	29	20	54	21	79	20
5	20	30	19	55	20	80	19
6	19	31	19	56	19	81	21
7	19	32	19	57	21	82	19
8	20	33	19	58	19	83	20
9	19	34	22	59	22	84	19
10	20	35	20	60	19	85	21
11	19	36	20	61	20	86	19
12	19	37	20	62	20	87	19
13	19	38	20	63	20	88	19
14	19	39	20	64	21	89	19
15	20	40	19	65	20	90	20
16	19	41	19	66	19	91	19
17	19	42	19	67	20	92	20
18	19	43	20	68	21	93	21
19	20	44	20	69	20	94	21

TABLA 1.2. *(Continuación)*

Alumno	Edad	Alumno	Edad	Alumno	Edad	Alumno	Edad
20	19	45	19	70	19	95	19
21	20	46	20	71	20	96	20
22	20	47	21	72	19	97	20
23	20	48	21	73	19	98	19
24	20	49	20	74	19	99	19
25	19	50	20	75	20	100	19

Utilizando la Tabla 1.1 de números aleatorios, para la selección de una muestra aleatoria de seis estudiantes, tendremos que empezar seleccionando seis número aleatorios, así pues si entramos en la tabla por el ángulo superior izquierdo y considerando números aleatorios de dos dígitos[2], pues la población es de tamaño 100, tenemos los siguientes números aleatorios:

20, 34, 56, 24, 75, 83

que se corresponden con los seis estudiantes de la muestra seleccionada, cuyas edades son:

19, 22, 19, 20, 20, 20

Esta situación aparece en el Gráfico 1.1.

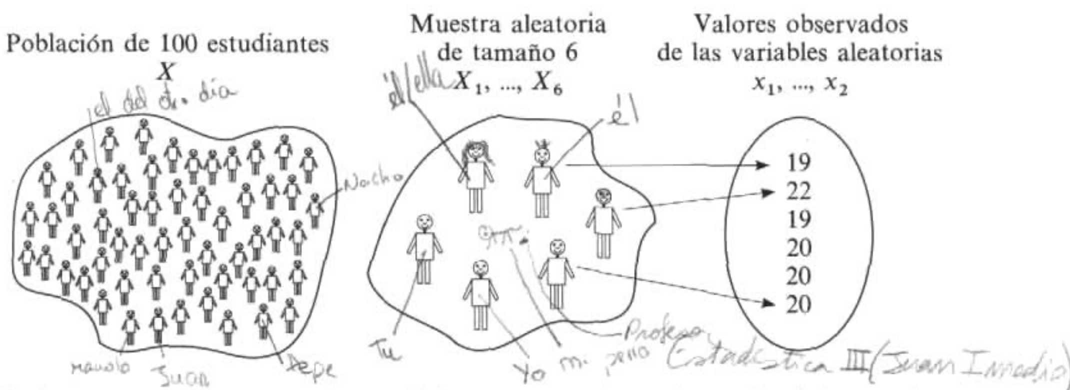


GRÁFICO 1.1. *Esquema de selección de la muestra y valores observados de las variables aleatorias.*

[2] Si nos aparece el 00 entenderemos que corresponde al alumno 100.

Como en este ejemplo estamos interesados en la edad del estudiante, consideramos la variable aleatoria

$$X\text{: edad del estudiante seleccionado}$$

Análogamente se podría hacer para las variables aleatorias estatura, peso, etcétera.

La distribución de probabilidades de la variable aleatoria X, edad del estudiante, viene dada en la Tabla 1.3, en donde se dan los diferentes valores de la variable X y sus probabilidades.

TABLA 1.3. *Distribución de probabilidades de la variable aleatoria X, edad del estudiante, correspondiente a la población de 100 estudiantes.*

Valores de la variable aleatoria X	Probabilidades $P(X=x)$
19	0,46
20	0,41
21	0,11
22	0,02

Si seleccionamos una muestra con reemplazamiento, de seis estudiantes de la población, para observar la edad de cada uno, entonces definiremos seis variables aleatorias:

X_1: edad del primer estudiante seleccionado.

X_2: edad del segundo estudiante seleccionado.

$\vdots$

X_6: edad del sexto estudiante seleccionado.

Cada variable aleatoria tendrá una distribución de probabilidad asociada. Así pues, la distribución de la variable aleatoria X_1 será exactamente la misma que la distribución de la variable aleatoria X dada en la Tabla 1.4, ya que ambas variables aleatorias se refieren a la edad de un estudiante seleccionado aleatoriamente, es decir, la distribución de probabilidades de la variable aleatoria X_1 ésta dada en la Tabla 1.4, en donde aparecen los diferentes valores de la variable aleatoria X_1 y sus probabilidades.

TABLA 1.4. *Distribución de probabilidades de la variable aleatoria X_1, edad del primer estudiante seleccionado.*

Valores de la variable aleatoria X_1	**Probabilidades $P(X_1 = x_1)$**
19	0,46
20	0,41
21	0,11
22	0,02

Pero como el muestreo se ha realizado con reemplazamiento, se puede ver que la variable aleatoria X_2 tiene la misma distribución de probabilidades que X o que X_1 y que X_1 y X_2 son independientes. Análogamente, las variables aleatorias X_3, X_4, X_5 y X_6 tienen la misma distribución que X, y en consecuencia la sucesión de variables aleatorias $X_1, X_2, ..., X_6$ son independientes e idénticamente distribuidas[3].

Definición 1.1. Muestra aleatoria simple.

Sea X la variable aleatoria correspondiente a una población con función de distribución $F(x)$. Si las variables aleatorias $X_1, X_2, ..., X_n$ son independientes y tienen la misma función de distribución, $F(x)$, que la de la distribución de la población, entonces las variables aleatorias $X_1, X_2, ..., X_n$ forman un conjunto de variables aleatorias independientes e idénticamente distribuidas que constituyen una **muestra aleatoria simple** de tamaño n de la población $F(x)$[4].

Al ser las variables aleatorias $X_1, X_2, ..., X_n$ independientes, resulta que la función de distribución conjunta será igual al producto de las funciones de distribución marginales, es decir:

$$F(x_1, ..., x_n) = \prod_{i=1}^{n} F(x_i) \qquad [1.1]$$

Si la población de partida es tipo discreto y la función de probabilidad de la población es:

$$p_i = P(X = x_i) \quad i = 1, 2, ..., r$$

[3] Observemos que si la muestra se selecciona sin reemplazamiento, la correspondiente sucesión de variables aleatorias no son independientes, aunque tengan la misma distribución de probabilidades.

[4] En lo sucesivo y si no indicamos lo contrario las muestras que utilizaremos serán aleatorias simples, aunque a veces por abreviar digamos simplemente muestra aleatoria.

entonces la función de probabilidad de la muestra será:

$$P(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} P(X = x_i) = \prod_{i=1}^{n} p_i \qquad [1.2]$$

Si la muestra aleatoria simple procede de una población de tipo continuo con función de densidad $f(x)$, entonces la función de densidad de la muestra será:

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i) \qquad [1.3]$$

1.3. PARÁMETROS POBLACIONALES Y ESTADÍSTICOS MUESTRALES

En general diremos que los **parámetros poblacionales** son las características numéricas de la población. En concreto, un **parámetro** es una caracterización numérica de la distribución de la población. El conocimiento del parámetro permite describir parcial o totalmente la función de probabilidad de la característica que estamos investigando. Así por ejemplo, si la característica a investigar sabemos que sigue una distribución exponencial de parámetro a su función de densidad será:

$$f(x) = \begin{cases} ae^{-ax} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

pero esta función de densidad no estará totalmente descrita hasta que no se dé el valor del parámetro a, y entonces será cuando podremos formular preguntas concretas sobre esa distribución, es decir, podremos calcular las diferentes probabilidades.

Si la característica a investigar sigue una distribución normal, $N(\mu, \sigma)$, cuya función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\cdot\frac{(x-\mu)^2}{\sigma^2}}$$

observamos que aparecen dos parámetros μ y σ, que no se han especificado, y para describir totalmente la función de densidad tendremos que dar valores a los dos parámetros μ y σ, pues si damos valor a un solo parámetro entonces diremos que está descrita parcialmente.

En la mayoría de los modelos probabilísticos nos encontraremos parámetros cuyos valores tendremos que fijar para especificar completamente el modelo y poder calcular las probabilidades deseadas[5]. De manera más concreta podemos decir que uno de los problemas centrales en estadística se nos presenta cuando deseamos estudiar una población con función de distribución $F(x, \theta)$, donde la forma de la función de distribución es conocida pero depende de un parámetro θ desconocido, ya que si θ fuese conocido tendríamos totalmente especificada la función de distribución. Si el parámetro θ no se conoce entonces se selecciona una muestra aleatoria simple $(X_1, ..., X_n)$ de tamaño n de la población, y se calcula para las observaciones de la muestra el valor de alguna función $g(x_1, ..., x_n)$, que representa o estima el parámetro desconocido θ. El problema es determinar qué función será la mejor para estimar el parámetro θ, lo cual será resuelto en el capítulo dedicado a la estimación.

A continuación exponemos el concepto de estadístico que es fundamental para estimar los parámetros poblacionales, pues los estimaremos mediante estadísticos definidos a partir de las observaciones de una muestra aleatoria.

Definición 1.2. Estadístico.

Un **estadístico** es cualquier función real de las variables aleatorias que integran la muestra, es decir, es una función de las observaciones muestrales, la cual no contiene ningún valor o parámetro desconocido.

Continuando con la población de función de distribución $F(x, \theta)$, en donde θ es un parámetro desconocido, y considerando una muestra aleatoria simple, $(X_1, ..., X_n)$, constituida por n variables aleatorias independientes e idénticamente distribuidas, podemos definir algunos estadísticos o funciones de esas variables aleatorias, como por ejemplo:

$$g_1(X_1, ..., X_n) = \frac{X_1 + \cdots + X_n}{n}$$

$$g_2(X_1, ..., X_n) = \frac{X_1^2 + \cdots + X_n^2}{n}$$

$$g_3(X_1, ..., X_n) = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

[5] En la Estadística clásica un *parámetro* se puede considerar como una constante fija cuyo valor se desconoce.

los cuales se determinan totalmente a partir de las observaciones muestrales.

En general un estadístico T lo representaremos como[6]:

$$T = g(X_1, ..., X_n)$$

es decir, como una función g de las observaciones muestrales, que a su vez será también una variable aleatoria, pues para cada muestra el estadístico T tomará un valor diferente, así pues para una muestra concreta $(x_1, ..., x_n)$ el estadístico tomará el valor:

$$T = g(x_1, ..., x_n)$$

y a medida que vamos tomando muestras diferentes se obtienen distintos valores del estadístico, resultando que efectivamente el estadístico T es también una variable aleatoria y por consiguiente tendrá su correspondiente distribución, a la que llamaremos **distribución muestral del estadístico,** como veremos posteriormente.

Vemos pues que un parámetro y un estadístico son conceptos muy diferentes, pues el parámetro es una constante y cuando se conoce determina completamente el modelo probabilístico, sin embargo el estadístico es una variable aleatoria cuyo valor dependerá de las observaciones muestrales.

En diferentes ocasiones se han estudiado medidas numéricas correspondientes a conjuntos de datos, así pues estudiamos, entre otras, la media y la desviación típica. Ahora vamos a distinguir entre medidas numéricas calculadas con conjuntos de datos poblacionales y las calculadas con datos muestrales. Así pues, si la medida numérica se calcula para el conjunto de datos poblacionales le llamaremos **valor del parámetro poblacional** y si se calcula para el conjunto de datos muestrales, le llamaremos **valor del estadístico muestral.**

---

[6] Seguiremos como norma general el utilizar letras mayúsculas para indicar las variables aleatorias, para los estadísticos, estimadores y para representar una muestra aleatoria general, y utilizaremos letras minúsculas para los valores concretos que puedan tomar las variables aleatorias, las estimaciones y la realización de una muestra o muestra concreta.

Definición 1.3. Parámetros media, varianza y proporción poblacional.

En una población finita de tamaño N los **parámetros poblacionales media, varianza** y **proporción poblacional** vienen dados por[7]:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad [1.4]$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (X_i - \mu)^2 \qquad [1.5]$$

$$p = \frac{X}{N} = \frac{\text{número de éxitos en } N \text{ pruebas}}{\text{número de pruebas}} \qquad [1.6]$$

Definición 1.4. Estadístico media, varianza y proporción muestral.

Para una muestra aleatoria simple de tamaño n, $(X_1, ..., X_n)$ los **estadísticos media, varianza** y **proporción muestral** se definen como:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad [1.7]$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad [1.8]$$

$$P_X = \frac{X}{n} = \frac{\text{número de éxitos en } n \text{ pruebas}}{\text{número de pruebas}} \qquad [1.9]$$

El estadístico **varianza muestral,** S^2, se puede formular también mediante las siguientes expresiones algebraicas:

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right) \qquad [1.10]$$

[7] Si la población es infinita utilizaremos la misma notación para designar estos parámetros poblacionales, pero estos no pueden ser calculados a partir de estas sumas finitas, sino que tendremos que recurrir al cálculo de valores esperados de variables aleatorias de tipo continuo.

En efecto para ver la equivalencia de la expresión [1.8] con la [1.10], consideramos el numerador de la [1.8] y tendremos:

$$\begin{aligned}\sum_{i=1}^{n} (X_i - \bar{X})^2 &= \sum_{i=1}^{n} (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \sum_{i=1}^{n} X_i^2 - 2\bar{X} \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \bar{X}^2 \\ &= \sum_{i=1}^{n} X_i^2 - 2\bar{X}(n\bar{X}) + n\bar{X}^2 \\ &= \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \qquad [1.11]\end{aligned}$$

Si en lugar de considerar las n variables aleatorias, independientes e idénticamente distribuidas $(X_1, ..., X_n)$, que constituyen la muestra aleatoria simple, consideramos una muestra concreta $(x_1, ..., x_n)$ entonces los valores de estos estadísticos muestrales son:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad [1.12]$$

$$s^2 = \frac{1}{n-1} (x_i - \bar{x})^2 \qquad [1.13]$$

$$\hat{p} = \frac{x}{n} \qquad [1.14]$$

Luego vemos que efectivamente el estadístico es una función de las observaciones muestrales, y en estos casos asigna a cada muestra observada la media de los valores, la varianza o la proporción, respectivamente[8].

1.4. FUNCIÓN DE DISTRIBUCIÓN EMPÍRICA

Sabemos que la función de distribución de una variable aleatoria X estaba definida como:

$$F(x) = P(X \leqslant x)$$

[8] Se observa que al definir el estadístico varianza muestral se divide por $n - 1$ en lugar de por n, la razón las veremos con más detalle después, pero aquí ya adelantamos que se ha definido así la varianza muestral s^2, para que esta s^2 sea un estimador insesgado de la varianza poblacional σ^2, pues si hubiéramos dividido por n entonces el estadístico no sería un estimador insesgado.

y puede representar la proporción de valores que son menores o iguales que x.

De manera similar podemos definir la función de distribución empírica para una muestra.

Definición 1.5. Función de distribución empírica de la muestra.

Consideremos una población con función de distribución $F(x)$ y sean $(x_1, ..., x_n)$ los valores observados correspondientes a una muestra aleatoria simple procedente de esa población, y designamos por $N(x)$ el número de valores observados que son menores o iguales que x. Entonces definimos la **función de distribución empírica de la muestra,** que la notaremos por $F_n(x)$, como:

$$F_n(x) = \frac{N(x)}{n} \qquad [1.15]$$

Ejemplo 1.2

Dada una muestra aleatoria formada por las observaciones muestrales (3, 8, 5, 4, 5). Obtener la función de distribución empírica y su correspondiente representación gráfica.

Solución:

Utilizando la expresión [1.15] podemos obtener la función de distribución empírica que aparece en la Tabla 1.5.

TABLA 1.5. *Función de distribución empírica.*

Observaciones muestrales x	**$N(x)$**	**$F_5(x)$**
—	<3, 0	0,0
3	$\leqslant 3$, 1	0,2
4	$\leqslant 4$, 2	0,4
5	$\leqslant 5$, 4	0,8
8	$\leqslant 8$, 5	1,0

La representación gráfica de esta función de distribución la tenemos en el Gráfico 1.2.

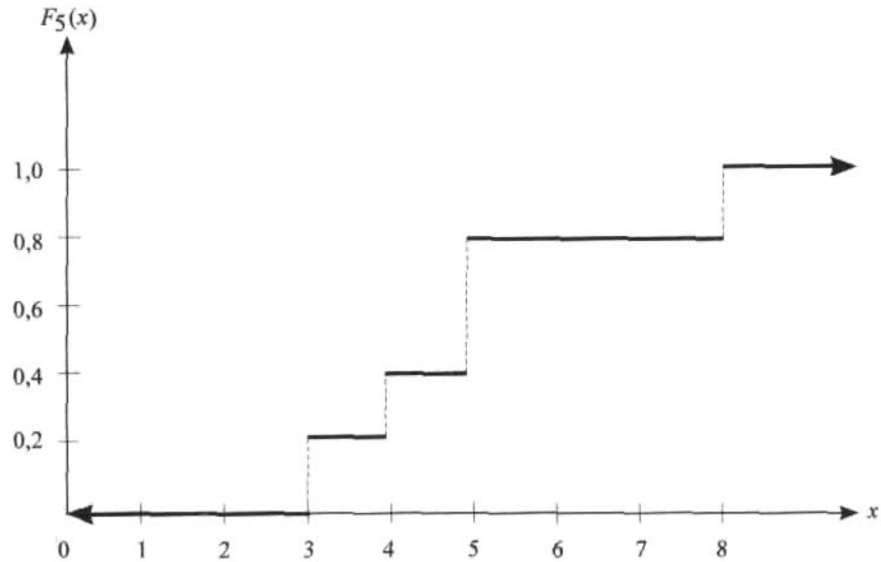


GRÁFICO 1.2. *Función de distribución empírica.*

La función de distribución empírica tiene las mismas propiedades que la función de distribución de la variable aleatoria, y, se puede demostrar, utilizando el **teorema de Glivenko-Cantelli**[9], que $F_n(x)$ converge en probabilidad a $F(x)$. Lo cual, a efectos prácticos, implica que cuando el tamaño de la muestra crece la gráfica de la función de distribución empírica se aproxima bastante a la de la función de distribución de la población, y se puede utilizar como estimador de la misma.

De todo esto se deduce que la función de distribución empírica o su gráfica se puede utilizar para determinar la forma general de la distribución poblacional. También es fácil y muy frecuente el reconocer la forma de la distribución observando el histograma correspondiente que nos daría idea de la función de densidades.

[9] El *teorema de Glivenko-Cantelli*, llamado también *Teorema fundamental de la Estadística*, por su papel fundamental en la inferencia estadística, indica que la función de distribución empírica de la muestra $F_n(x)$ converge en probabilidad a la función de distribución de la población $F(x)$. Es decir, para $\varepsilon > 0$, se verifica:

$$\lim_{n \to \infty} P[\sup_{-\infty < x < +\infty} | F_n(x) - F(x) | \geq \varepsilon] = 0$$

Lo cual significa que si la muestra es suficientemente grande y se verifica el teorema, entonces la muestra puede proporcionar información casi exacta sobre la distribución de la población.

1.5. DISTRIBUCIÓN MUESTRAL DE ESTADÍSTICOS

Como veremos posteriormente los estadísticos muestrales (proporción, media y varianza muestral) se pueden utilizar para estimar los correspondientes parámetros poblacionales. Así pues, para estudiar propiedades de estos estadísticos, como estimadores de los parámetros poblacionales, será necesario estudiar las características de la distribución de probabilidad de estos estadísticos.

Sabemos que los estadísticos muestrales se calculan a partir de los valores $(X_1, ..., X_n)$ de una muestra aleatoria, y estos estadísticos son también variables aleatorias. Como tales variables aleatorias tienen su distribución de probabilidad, así pues los estadísticos muestrales: proporción, media, varianza, etc., tendrán su correspondiente distribución de probabilidad. Si tales distribuciones de probabilidad se pueden obtener, entonces será posible establecer afirmaciones probabilísticas sobre esos estadísticos.

La distribución exacta de los estadísticos dependerá del tamaño muestral n. Así, en muchas situaciones, encontrar la distribución de probabilidad exacta del estadístico media muestral $\bar{X}$, incluso para n pequeño y variables aleatorias discretas, será bastante pesado, pero sin grandes dificultades teóricas. En muchos casos esto será relativamente sencillo, mientras que en otros lo mejor que se puede hacer es tomar una muestra grande y utilizar la distribución límite apropiada.

El término distribución muestral se utiliza para poner de manifiesto que hay diferencia entre la distribución de la población de la cual se ha extraído la muestra y la distribución de alguna función de esa muestra.

Conceptualmente, la distribución muestral de un estadístico puede ser obtenida tomando todas las posibles muestras de un tamaño fijado n, calculando el valor del estadístico para cada muestra y construyendo la distribución de estos valores.

En esta sección estamos interesados en determinar las distribuciones de probabilidad de algunos estadísticos muestrales, en concreto, para la media $\bar{X}$ y varianza S^2 muestral, que serán de bastante utilidad en diferentes aplicaciones estadísticas.

Así, por ejemplo, si el estadístico es la media muestral $\bar{X}$, la distribución muestral de $\bar{X}$ puede construirse tomando todas las muestras posibles de tamaño n, calculando el valor del estadístico $\bar{X}$ para cada muestra, que lo notaremos por $\bar{x}$, y formando la distribución de los valores $\bar{x}$.

Ejemplo 1.3

Supongamos una población formada por las cinco tiendas existentes en un municipio. La característica a investigar será el número de horas que diariamente permanecen abiertas esas tiendas y que representaremos por la variable aleatoria X, estando los valores poblacionales expresadas en la Tabla 1.6.

TABLA 1.6. *Valores poblacionales de la variable aleatoria X.*

Tiendas	Valores de X
T_1	$x_1 = 12$
T_2	$x_2 = 10$
T_3	$x_3 = 14$
T_4	$x_4 = 9$
T_5	$x_5 = 10$

Los valores de los parámetros media y varianza poblacional serán:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$= \frac{1}{5}(12 + 10 + 14 + 9 + 10) = 11$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2$$

$$= \frac{1}{5}[(12 - 11)^2 + (10 - 11)^2 + (14 - 11)^2 + (9 - 11)^2 + (10 - 11)^2] = \frac{16}{5}$$

Las diez posibles muestras aleatorias simples de tamaño 3 que se pueden tomar y el valor del estadístico media muestral aparecen en la Tabla 1.7.

La distribución de probabilidad del estadístico media muestral $\bar{X}$ viene dada por la Tabla 1.8.

TABLA 1.7. *Posibles muestras de tamaño 3 y valores del estadístico media muestral.*

Muestras	Observaciones muestrales x_i	Estadístico media muestral $\bar{X}$ $\bar{x}$
$(T_1\ T_2\ T_3)$	(12, 10, 14)	12,0
$(T_1\ T_2\ T_4)$	(12, 10, 9)	10,3
$(T_1\ T_2\ T_5)$	(12, 10, 10)	10,6
$(T_1\ T_3\ T_4)$	(12, 14, 9)	11,6
$(T_1\ T_3\ T_5)$	(12, 14, 10)	12,0
$(T_1\ T_4\ T_5)$	(12, 9, 10)	10,3
$(T_2\ T_3\ T_4)$	(10, 14, 9)	11,0
$(T_2\ T_3\ T_5)$	(10, 14, 10)	11,3
$(T_2\ T_3\ T_5)$	(10, 9, 10)	9,6
$(T_3\ T_4\ T_5)$	(14, 9, 10)	11,0

TABLA 1.8. *Distribuciones muestral del estadístico media muestral $\bar{X}$.*

Valores del estadístico media muestral $\bar{X}$ $\bar{x}$	Función de probabilidad $P(\bar{x}) = P(\bar{X} = \bar{x})$
9,6	0,1
10,3	0,2
10,6	0,1
11,0	0,2
11,3	0,1
11,6	0,1
12,0	0,2

La representación gráfica de la distribución muestral del estadístico media muestral $\bar{X}$, se tiene en el Gráfico 1.3.

Veamos ahora otro ejemplo más completo para muestras de tamaño dos en el cual obtendremos las distribuciones de probabilidad de los estadísticos media, $\bar{X}$, y varianza, S^2, muestral. También obtendremos las medias y varianzas de ambos estadísticos.

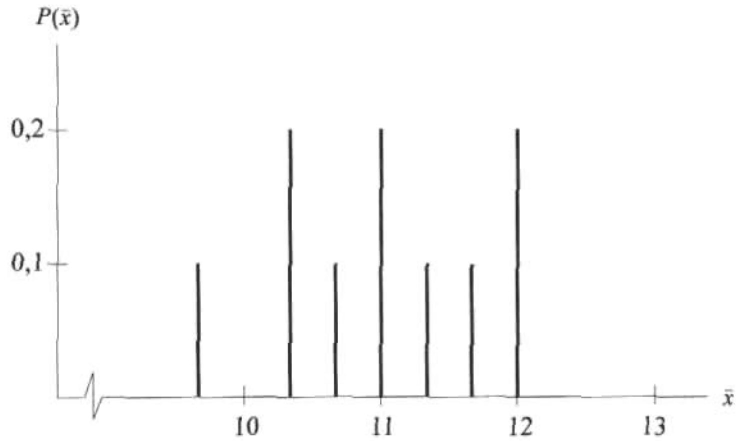


GRÁFICO 1.3. *Distribución muestral del estadístico media muestral $\bar{X}$.*

Ejemplo 1.4

Sea una empresa dedicada al transporte y distribución de mercancías, la cual tiene una plantilla de 50 trabajadores. Durante el último año se ha observado que 25 trabajadores han faltado un solo día al trabajo, 20 trabajadores han faltado dos días y 5 trabajadores han faltado tres días. Si se toma una muestra aleatoria, con reemplazamiento, de tamaño dos (X_1, X_2) del total de la plantilla, obtener:

1. La distribución de probabilidad del número de días que ha faltado al trabajo un empleado, su media y su varianza.
2. Distribución de probabilidad del estadístico media muestral $\bar{X}$.
3. La distribución de probabilidad del estadístico varianza muestral, S^2.
4. La media y varianza del estadístico media muestral.
5. La probabilidad de que el estadístico media muestral, $\bar{X}$, sea menor que 2.
6. La media y varianza del estadístico varianza muestral.
7. La probabilidad de que el estadístico varianza muestral, S^2, sea menor o igual que 0,5.

Solución:

1. Empezaremos obteniendo la distribución de probabilidad de la variable aleatoria:

X: número de días que ha faltado al trabajo un empleado elegido aleatoriamente de la plantilla total.

La variable aleatoria X, puede tomar los valores 1, 2 ó 3, y como la selección se hace de manera aleatoria, todos los trabajadores tendrán la misma probabilidad de ser seleccionados, luego la distribución de probabilidad de la variable aleatoria X viene dada en la Tabla 1.9, y será la distribución de probabilidad de la población.

TABLA 1.9. *Distribución de probabilidad de la variable aleatoria X.*

Valores de la variable aleatoria X x	**Probabilidades $P(X=x)=P(x)$**
1	$P(X = 1) = P(1) = \frac{25}{50} = 0{,}5$
2	$P(X = 2) = P(2) = \frac{20}{50} = 0{,}4$
3	$P(X = 3) = P(3) = \frac{5}{50} = 0{,}1$

A partir de esta distribución de probabilidad tenemos que la media será:

$$\begin{aligned}\mu = E[X] &= \sum_i x_i \cdot P(X = x_i)\\ &= 1(0{,}5) + 2(0{,}4) + 3(0{,}1)\\ &= 1{,}6\end{aligned}$$

y la varianza

$$\begin{aligned}\sigma^2 = \text{Var}\,(X) = E[(X - \mu)^2] &= \sum_i (x_i - \mu)^2 \cdot P(X = x_i)\\ &= (1 - 1{,}6)^2(0{,}5) + (2 - 1{,}6)^2(0{,}4) + (3 - 1{,}6)^2(0{,}1)\\ &= 0{,}44\end{aligned}$$

Observamos que si sumamos el número total de faltas al trabajo que se han producido en la población de los 50 empleados y dividimos por los 50 empleados tenemos la media.

$$\frac{25 \cdot 1 + 20 \cdot 2 + 5 \cdot 3}{50} = \frac{80}{50} = 1{,}6$$

Análogamente sucede con la varianza.

Por esto, en lo sucesivo μ y σ^2 serán consideradas como la media y la varianza poblacional, respectivamente.

2. Seleccionamos una muestra aleatoria, con reemplazamiento, de tamaño dos (X_1, X_2), siendo:

X_1: variable aleatoria correspondiente al número de días que falta el primer trabajador seleccionado.

X_2: variable aleatoria correspondiente al número de días que falta el segundo trabajador seleccionado.

Ambas variables aleatorias X_1 y X_2 tienen la misma distribución de probabilidad que la de la variable aleatoria X, correspondiente a la población.

Pero como nos interesa obtener la distribución de probabilidad de estadístico media muestral:

$$\bar{X} = \frac{1}{2}(X_1 + X_2)$$

ésta estará relacionada con la distribución de probabilidad de las variables aleatorias X_1 y X_2.

Para tener las distribuciones de probabilidad de los estadísticos media $\bar{X}$ y varianza S^2 muestral necesitaremos tener los diferentes valores que puede tomar y sus probabilidades. Para ello empezaremos obteniendo las posibles muestras, con reemplazamiento, de tamaño dos, sus probabilidades y los valores correspondientes de los estadísticos media y varianza muestral, que vienen dados en la Tabla 1.10.

TABLA 1.10. *Muestras de tamaño dos y valores obtenidos para las distribuciones de probabilidad de $\bar{X}$ y S^2.*

Muestras de tamaño dos (x_1, x_2)	$\bar{X}$	S^2	$P(X_1=x_1, X_2=x_2)$
(1, 1)	1,0	0,0	0,25
(1, 2)	1,5	0,5	0,20
(1, 3)	2,0	2,0	0,05
(2, 1)	1,5	0,5	0,20
(2, 2)	2,0	0,0	0,16
(2, 3)	2,5	0,5	0,04
(3, 1)	2,0	2,0	0,05
(3, 2)	2,5	0,5	0,04
(3, 3)	3,0	0,0	0,01

Para obtener las probabilidades correspondientes a los diferentes valores muestrales, tendremos en cuenta que las variables X_1 y X_2 son independientes, pues el muestreo se ha realizado con reemplazamiento. Luego

$$\begin{aligned}
P(\bar{X} = 1) &= P(X_1 = 1, X_2 = 1) \\
&= P(X_1 = 1) \cdot P(X_2 = 1) \\
&= (0{,}5)(0{,}5) = 0{,}25 \\
P(\bar{X} = 1{,}5) &= P[(X_1 = 1, X_2 = 2) \text{ ó } (X_1 = 2, X_2 = 1)] \\
&= P(X_1 = 1, X_2 = 2) + P(X_1 = 2, X_2 = 1) \\
&= P(X_1 = 1) \cdot P(X_2 = 2) + P(X_1 = 2) \cdot P(X_2 = 1) \\
&= (0{,}5)(0{,}4) + (0{,}4)(0{,}5) \\
&= 0{,}20 + 0{,}20 = 0{,}40
\end{aligned}$$

Análogamente obtendremos las restantes probabilidades.

La información que nos proporciona la Tabla 1.10 la utilizaremos para obtener la distribución de probabilidad del estadístico media muestral $\bar{X}$, así pues:

$$\begin{aligned}
&P(\bar{X} = 1) = 0{,}25 \\
&P(\bar{X} = 1{,}5) = 0{,}20 + 0{,}20 = 0{,}40 \\
&P(\bar{X} = 2) = 0{,}05 + 0{,}16 + 0{,}05 = 0{,}26 \\
&P(\bar{X} = 2{,}5) = 0{,}04 + 0{,}04 = 0{,}08 \\
&P(\bar{X} = 3) = 0{,}01
\end{aligned}$$

Luego la distribución de probabilidad del estadístico media muestral $\bar{X}$ la tenemos en la Tabla 1.11.

TABLA 1.11. *Distribución de probabilidad del estadístico media muestral $\bar{X}$.*

Valores del estadístico $\bar{X}$ $\bar{x}$	**Probabilidades** $P(\bar{X}=\bar{x})=P(\bar{x})$
1	0,25
1,5	0,40
2	0,26
2,5	0,08
3	0,01

3. Análogamente podemos obtener la distribución de probabilidad del estadístico varianza muestral S^2. Los diferentes valores del estadístico S^2 aparecen en la tercera columna de la Tabla 1.10, así pues, para la primera muestra tenemos:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{2-1} [(1-1)^2 + (1-1)^2] = 0$$

Para la segunda muestra será:

$$s^2 = \frac{1}{2-1} [(1-1,5)^2 + (2-1,5)^2] = 0,5$$

y de manera análoga tendríamos los restantes valores.

Las probabilidades correspondientes a los diferentes valores del estadístico S^2, las obtenemos a partir de la Tabla 1.10, así pues:

$$P(S^2 = 0,0) = 0,25 + 0,16 + 0,01 = 0,42$$

$$P(S^2 = 0,5) = 0,20 + 0,20 + 0,04 + 0,04 = 0,48$$

$$P(S^2 = 2,0) = 0,05 + 0,05 = 0,10$$

Y la distribución de probabilidad del estadístico varianza muestra S^2 viene dada en la Tabla 1.12.

TABLA 1.12. *Distribución de probabilidad del estadístico varianza muestral S^2.*

Valores del estadístico S^2 s^2	**Probabilidades** $P(S^2 = s^2) = P(s^2)$
0,0	0,42
0,5	0,48
2,0	0,10

4. Para el cálculo de la media y varianza del estadístico media muestral tendremos en cuenta su distribución de probabilidad dada en la Tabla 1.11.

Utilizando la definición de valor esperado de una variable aleatoria de tipo discreto tenemos:

$$\begin{aligned}
\mu_{\bar{X}} = E[\bar{X}] &= \sum_i \bar{x}_i \cdot P(\bar{X} = \bar{x}_i) \\
&= 1(0{,}25) + 1{,}5(0{,}40) + 2(0{,}26) + 2{,}5(0{,}08) + 3(0{,}01) \\
&= 1{,}60 \\
\sigma^2_{\bar{X}} = \text{Var}\,(\bar{X}) &= E[(\bar{X} - E(\bar{X}])^2] \\
&= \sum_i (\bar{x}_i - 1{,}60)^2 \cdot P(\bar{X} = \bar{x}_i) \\
&= (1 - 1{,}60)^2(0{,}25) + \cdots + (3 - 1{,}60)^2(0{,}01) \\
&= 0{,}09 + \cdots + 0{,}019 \\
&= 0{,}22
\end{aligned}$$

5. Teniendo en cuenta la distribución de probabilidad del estadístico media muestral $\bar{X}$, Tabla 1.11, se tiene:

$$\begin{aligned}
P(\bar{X} < 2) &= P(\bar{X} = 1) + P(\bar{X} = 1{,}5) \\
&= 0{,}25 + 0{,}40 \\
&= 0{,}65
\end{aligned}$$

6. Teniendo en cuenta la distribución de probabilidad del estadístico varianza muestral, S^2, dada en la Tabla 1.12, y procediendo de manera análoga a como lo hemos hecho para el estadístico media muestral, tendremos

$$\begin{aligned}
\mu_{S^2} = E[S^2] &= \sum_i s_i^2 \cdot P(S^2 = s_i^2) \\
&= 0{,}0(0{,}42) + 0{,}5(0{,}48) + 2{,}0(0{,}10) \\
&= 0{,}44 \\
\sigma^2_{S^2} = \text{Var}\,(S^2) &= E[(S^2 - E[S^2])^2] \\
&= \sum_i (s_i^2 - 0{,}44) P(S^2 = s_i^2) \\
&= (0{,}0 - 0{,}44)^2(0{,}42) + (0{,}5 - 0{,}44)^2(0{,}48) + (2{,}0 - 0{,}44)^2(0{,}10) \\
&= 0{,}0813 + 0{,}0017 + 0{,}2434 \\
&= 0{,}32
\end{aligned}$$

7. Basándonos en la distribución de probabilidad del estadístico varianza muestral S^2, Tabla 1.12, se tiene:

$$\begin{aligned} P(S^2 \leqslant 0{,}5) &= P(S^2 = 0{,}0) + P(S^2 = 0{,}5) \\ &= 0{,}42 + 0{,}48 \\ &= 0{,}90 \end{aligned}$$

Con este ejemplo, se pone de manifiesto que incluso para muestras de tamaño pequeño y estadísticos con pocos valores posibles se hace pesado el obtener la distribución de probabilidad de los estadísticos muestrales. Para evitar esto en los siguientes apartados daremos algunos resultados que simplifican estos problemas.

1.6. MEDIA Y VARIANZA DE ALGUNOS ESTADÍSTICOS

En el Ejemplo 1.4 hemos obtenido:

— La media, μ, y varianza, σ^2, poblacional.

— Los estadísticos media $\bar{X}$ y varianza S^2 muestral.

— La media y varianza de los estadísticos media muestral, $\bar{X}$, y varianza muestral, S^2, para una muestra de tamaño $n = 2$.

Estos resultados se recogen en la Tabla 1.13, en donde se observa:

1.° Que $\boldsymbol{E[\bar{X}] = E[X]}$,

es decir, que la media del estadístico media muestral es igual a la media de la población.

2.° Que $\boldsymbol{E[S^2] = \mathrm{Var}\ (X)}$,

es decir, que la media del estadístico varianza muestral es igual a la varianza de la población.

3.° Que $\boldsymbol{\mathrm{Var}\ (\bar{X}) = \dfrac{\mathrm{Var}\ (X)}{2}}$,

es decir, que la varianza del estadístico media muestral es igual a la varianza de la población dividida por el tamaño de la muestra, n.

TABLA 1.13. *Media y varianza poblacional y de los estadísticos media y varianza muestral del ejemplo 1.4, para n = 2.*

	Poblacional X	**Estadístico media muestral** $\bar{X}$	**Estadístico varianza muestral** S^2
Media	$\mu = E[X] = 1{,}6$	$\mu_{\bar{X}} = E[\bar{X}] = 1{,}6$	$\mu_{S^2} = E[S^2] = 0{,}44$
Varianza	$\sigma^2 = \text{Var}(X) = 0{,}44$	$\sigma^2_{\bar{X}} = \text{Var}(\bar{X}) = 0{,}22$	$\sigma^2_{S^2} = \text{Var}(S^2) = 0{,}32$

Estos resultados no sólo se verifican para este ejemplo sino que se verifican en general, como veremos en los siguientes teoremas.

Teorema 1.1

Si $(X_1, ..., X_n)$ es una muestra aleatoria simple de tamaño n procedente de una población, descrita por la variable aleatoria X, con media $E[X] = \mu$ y varianza $\text{Var}(X) = \sigma^2$, entonces la **esperanza de la media muestral** es igual a la media de la población, μ, y la **varianza de la media muestral** es igual a la varianza poblacional, σ^2, dividida por n, es decir,

$$E[\bar{X}] = \mu \quad \text{y} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \qquad [1.16]$$

Demostración:

Teniendo en cuenta la definición de muestra aleatoria simple, resulta que las variables aleatorias $X_1, ..., X_n$ son independientes, todas tienen la misma distribución de probabilidad que la población X y en consecuencias todas tienen la misma media y la misma varianza que la población X, es decir:

$$E[X_1] = \cdots = E[X_n] = E[X] = \mu$$

$$\text{Var}(X_1) = \cdots = \text{Var}(X_n) = \text{Var}(X) = \sigma^2$$

Luego si tenemos en cuenta las propiedades de los valores esperados, resulta que la media o esperanza matemática del estadístico media muestral será:

$$
\begin{aligned}
E[\bar{X}] &= E\left[\frac{X_1 + \cdots + X_n}{n}\right] \\
&= \frac{1}{n} E[X_1 + \cdots + X_n] \\
&= \frac{1}{n} (E[X_1] + \cdots + E[X_n]) \\
&= \frac{1}{n} (\mu + \cdots + \mu) = \frac{n\mu}{n} = \mu
\end{aligned}
$$

Análogamente para la varianza, y dado que las variables aleatorias $X_1, \ldots, X_n$ son independientes, resulta:

$$
\begin{aligned}
\text{Var}\,(\bar{X}) &= \text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2} \text{Var}\,(X_1 + \cdots + X_n) \\
&= \frac{1}{n^2} (\text{Var}\,(X_1) + \cdots + \text{Var}\,(X_n)) \\
&= \frac{1}{n^2} (\sigma^2 + \cdots + \sigma^2) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}
$$

Luego vemos que se puede obtener la media y la varianza del estadístico media muestral $\bar{X}$ sin necesidad de conocer la distribución probabilidad del estadístico $\bar{X}$, y sin importar la distribución de probabilidad de la población siempre y cuando la varianza tenga un valor finito.

A la correspondiente **desviación típica** del estadístico $\bar{X}$ se le llama **error estándar de la media** y viene dado por:

$$\text{error estándar de la media muestral } \bar{X} = \frac{\sigma}{\sqrt{n}} \qquad [1.17]$$

Observando los resultados de la expresión [1.16] se pone de manifiesto que el valor central del estadístico media muestral es la media poblacional μ, y como la dispersión del estadístico media muestral $\bar{X}$ en torno a su media μ es:

$$\text{Var}\,(\bar{X}) = E[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n}$$

resulta que cuanto mayor sea el tamaño muestral n menor será la $\operatorname{Var}(\bar{X})$, es decir, menor será la dispersión de $\bar{X}$ en torno a la media poblacional μ, y el valor observado del estadístico $\bar{X}$ estará más próximo a μ, lo cual nos permite decir que el estadístico media muestral $\bar{X}$ puede ser considerado como un buen estimador de la media poblacional μ.

En el Gráfico 1.4 se indica la distribución muestral del estadístico media muestral, $\bar{X}$, para muestras de tamaño $n = 25$ y $n = 110$ procedentes de una población normal $N(100, 6)$, en donde se observa que cada distribución muestral está centrada sobre la media poblacional, pero cuando el tamaño muestral aumenta la distribución muestral del estadístico media muestral está más concentrada en torno a la media de la población. En consecuencia el error estándar de la media muestral es una función decreciente del tamaño n de la muestra, y la probabilidad de que la media muestral difiera de la media poblacional en una cantidad fija, disminuye cuando el tamaño de la muestra crece.

Luego, si el tamaño de la muestra aumenta, la precisión de la media muestral para estimar la media de la población también aumenta. Por ejemplo, si se toma una muestra aleatoria de tamaño $n = 16$, entonces:

$$\text{error estándar de la media muestral} = \frac{\sigma}{\sqrt{16}} = \frac{\sigma}{4}$$

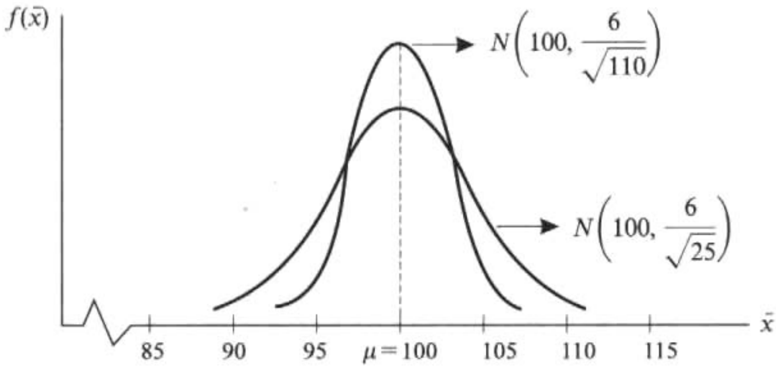


GRÁFICO 1.4. *Representación gráfica de las funciones de densidad del estadístico media muestral para muestras de tamaño $n = 25$ y $n = 110$, de una población $N(100, 6)$.*

y la media muestral $\bar{X}$ tiene una precisión $\sqrt{16} = 4$ veces mayor para estimar la media poblacional que la que tendría si se hubiera tomado una muestra con una sola observación, pero el aumento de la muestra tiene un límite, pues llega un momento que aunque el tamaño de la muestra siga aumentando la precisión prácticamente no aumenta. En efecto, supongamos una población con $\sigma = 12$ y calculamos la desviación estándar del estadístico $\bar{X}$ para diferentes valores de n, obteniéndose la Tabla 1.14.

TABLA 1.14. *Diferentes valores de la desviación estándar de $\bar{X}$ cuando $\sigma = 12$ para $n = 5, 10, 20, 30, \ldots$*

Valores de n	5	10	20	30	40	50	60	70	80	90	100
Desviación estándar $\frac{\sigma}{\sqrt{n}}$	5,38	3,79	2,68	2,19	1,89	1,69	1,55	1,43	1,34	1,26	1,20

Observando los valores de la Tabla 1.14 y su correspondiente representación gráfica, Gráfico 1.5, se observa que la desviación estándar de $\bar{X}$ disminuye sustancialmente a medida que n aumenta, pero cuando n pasa de 40 esta disminución se reduce hasta tal extremo que cuando n sigue creciendo y toma valores superiores a 80 ó 90 la desviación estándar de $\bar{X}$ prácticmente no disminuye. En consecuencia, podemos decir que si utilizamos el estadístico media muestral $\bar{X}$ para tener conocimiento o hacer inferencias sobre el parámetro media poblacional μ no es conveniente tomar muestras de tamaño demasiado grande pues el aumento del coste no compensa con la escasa disminución de la precisión.

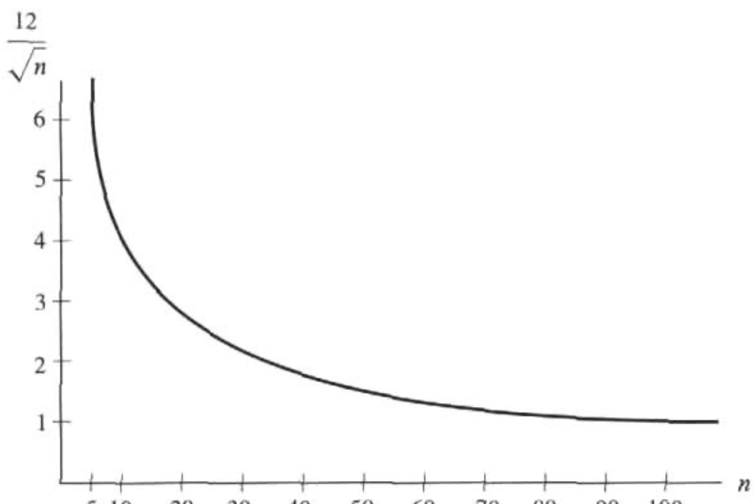


GRÁFICO 1.5. *Representación de la evolución de la desviación estándar del estadístico $\bar{X}$ en función de n.*

El resultado obtenido en el Teorema 1.1 es válido cuando el muestreo se hace de una población infinita, o bien de una población finita, pero con reemplazamiento, pues las variables aleatorias $X_1, ..., X_n$, tienen que ser independientes. Si el muestreo se hace sin reemplazamiento en una población finita de tamaño N, las variables aleatorias $X_1, ..., X_n$ no son independientes y entonces tendríamos que:

$$E[\bar{X}] = \mu \quad , \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Al término $\dfrac{N-n}{N-1}$ se le suele llamar **factor de corrección de población finita.**

Teorema 1.2

Si $(X_1, ..., X_n)$ es una muestra aleatoria simple de tamaño n, procedente de una población, descrita por la variable aleatoria X, con varianza, $\text{Var}(X) = \sigma^2$, entonces la esperanza de la varianza muestral S^2 es igual a la varianza poblacional σ^2 y la varianza de la varianza muestral es función del momento central de orden cuatro, es decir[10]:

$$E[S^2] = \sigma^2 \quad \text{y} \quad \text{Var}(S^2) = \frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\sigma^4 \qquad [1.18]$$

la media del estadístico varianza muestral = varianza de la pobla

Demostración:

Sabemos que el estadístico varianza muestral viene dado por:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

pero otra forma de expresarla es la siguiente:

[10] Si la población de partida es $N(\mu, \sigma)$ entonces como $\mu_4 = 3\sigma^4$, tenemos:

$$\text{Var}(s^2) = \frac{3\sigma^4}{n} + \frac{3-n}{n(n-1)}\sigma^4 = \frac{1}{n(n-1)}(3n\sigma^4 - 3\sigma^4 + 3\sigma^4 - n\sigma^4) = \frac{2\sigma^4}{n-1}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}[(X_i - \mu) - (\bar{X} - \mu)]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}[(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)(n\bar{X} - n\mu)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2n(\bar{X} - \mu)^2\right]$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{n}{n-1}(\bar{X} - \mu)^2 \qquad [1.19]$$

Tomando valores esperados resulta:

$$E[S^2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{n}{n-1}(\bar{X} - \mu)^2\right]$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}E[(X_i - \mu)^2] - \frac{n}{n-1}[E(\bar{X} - \mu)^2]$$

$$= \frac{1}{n-1}n\sigma^2 - \frac{n}{n-1}\frac{\sigma^2}{n}$$

$$= \frac{n\sigma^2}{n-1} - \frac{\sigma^2}{n-1} = \frac{(n-1)\sigma^2}{n-1}$$

$$= \sigma^2$$

Luego vemos que la esperanza del estadístico varianza muestral es igual a la varianza poblacional. Resultado que también será de bastante utilidad cuando estudiemos la estimación.

La segunda parte no la demostraremos, pues aunque no presenta dificultad los desarrollos son algo pesados[11].

1.7. DISTRIBUCIONES DE ESTADÍSTICOS MUESTRALES DE POBLACIONES NORMALES

En este apartado estudiaremos las distribuciones de algunos estadísticos para muestras procedentes de poblaciones normales, cuyos parámetros pueden, o no, ser conocidos.

Sabemos que muchos fenómenos que se observan en la realidad tienen distribuciones de frecuencias relativas que al representarlas tienen una forma parecida a la distribución normal, por ello podemos suponer que la mayoría de las poblaciones con las que nos encontraremos serán normales, y las variables aleatorias observadas en una muestra aleatoria $(X_1, ..., X_n)$ serán independientes y tienen la misma distribución.

1.7.1. DISTRIBUCIÓN DE LA MEDIA MUESTRAL CUANDO SE CONOCE LA VARIANZA POBLACIONAL

Al estudiar las propiedades que se deducían de la distribución normal, la primera que considerábamos era la referente a la distribución de una combinación lineal de variables aleatorias normales. Así pues, sabemos que si $X_1, ..., X_n$ son variables aleatorias independientes distribuidas según una $N(\mu_i, \sigma_i)$, para $i = 1, ..., n$ y si $a_1, ..., a_n$, son números reales, entonces la variable aleatoria

$$Y = a_1X_1 + \cdots + a_nX_n$$

sigue una distribución

$$N(a_1\mu_1 + \cdots + a_n\mu_n, \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2})$$

Este resultado nos será de bastante utilidad para obtener la distribución de la media muestral, como veremos en el Teorema 1.3.

---

[11] Pueden verse en ROHATGI (1976), pág. 305.

Teorema 1.3

Sea $(X_1, ..., X_n)$ una muestra aleatoria simple de tamaño n, procedente de una población $N(\mu, \sigma)$. Entonces la distribución del estadístico media muestral tendrá una distribución normal, es decir:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \qquad [1.20]$$

y como consecuencia el estadístico

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to N(0, 1)$$

Demostración[12]:

Sabemos que la función generatriz de momentos de una variable aleatoria X, $N(\mu, \sigma)$ es:

$$g_X(t) = E[e^{tX}] = e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

y como las variables X_i son independientes y todas tienen la misma distribución $N(\mu, \sigma)$, entonces la función generatriz de momentos del estadístico media muestral será:

$$\begin{aligned}
g_{\bar{X}}(t) &= E[e^{t\bar{X}}] = E\left[e^{t\left(\sum_{i=1}^{n} \frac{X_i}{n}\right)}\right] \\
&= E\left[e^{t\left(\frac{X_1}{n} + \cdots + \frac{X_n}{n}\right)}\right] \\
&= E\left[e^{\frac{t}{n}X_1}\right] \cdot \ldots \cdot E\left[e^{\frac{t}{n}X_n}\right] \\
&= e^{\frac{t}{n}\mu + \frac{1}{2}\frac{t^2}{n^2}\sigma^2} \cdot \ldots \cdot e^{\frac{t}{n}\mu + \frac{1}{2}\frac{t^2}{n^2}\sigma^2} \\
&= \left(e^{\frac{t}{n}\mu + \frac{1}{2}t^2\frac{\sigma^2}{n^2}}\right)^n \\
&= e^{t\mu + \frac{1}{2}t^2\frac{\sigma^2}{n}}
\end{aligned}$$

[12] Ver CASAS y SANTOS (1995). *Introducción a la Estadística para Economía y Administración de Empresa*, cap. 12. La demostración es una consecuencia inmediata de la propiedad I de la distribución normal, bastará hacer $a_i = \frac{1}{n}$ y $b = 0$.

que es la función generatriz de momentos de una variable aleatoria distribuida según una $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Luego, teniendo en cuenta la unicidad de la función generatriz de momento, resulta que:

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

En muchas situaciones la población de partida de la cual se extrae la muestra no es normal. En tales casos la distribución muestral del estadístico media muestral $\bar{X}$, seguirá siendo normal con media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$, es decir,

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \qquad [1.21]$$

siempre que el tamaño muestral sea grande, $n \geqslant 30$. Este resultado es una consecuencia inmediata del Teorema Central del Límite[13].

En el Gráfico 1.6, de la página siguiente, podemos observar la evolución de la forma de la distribución muestral del estadístico media muestral $\bar{X}$ cuando el tamaño de la muestra aumenta.

Observamos que cuando la población de partida no es normal, la forma de la distribución muestral no es exactamente normal, pero cuando el tamaño muestral $n \geqslant 30$, entonces es aproximadamente normal, siendo la aproximación tanto mejor cuanto mayor sea n. También observamos que la dispersión de la distribución muestral de $\bar{X}$ disminuye cuando el tamaño muestral, n, aumenta.

De la aproximación dada por la expresión [1.21], tipificando se tiene:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to N(0, 1) \qquad [1.22]$$

[13] En el Teorema Central del Límite no importa la distribución que siguen las variables aleatorias, pero si era necesario que las variables $X_1, \ldots, X_n$, fuesen idénticamente distribuidas, con media y varianza finitas.

1. Distribución poblacional (no es normal).

2. Distribución muestral de $\bar{X}$ para $n = 5$

3. Distribución muestral de $\bar{X}$ para $n = 15$

4. Distribución muestral de $\bar{X}$ para $n = 30$
   Aproximadamente normal

5. Distribución muestral de $\bar{X}$ para $n = 70$
   Aproximadamente normal

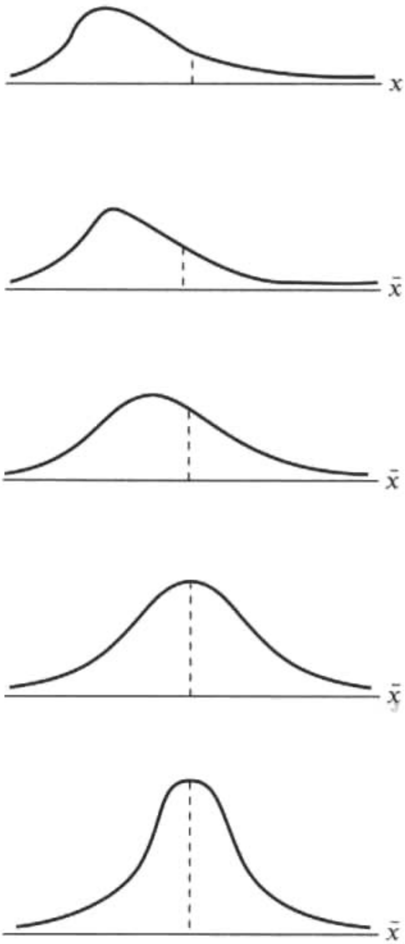


GRÁFICO 1.6. *Distribución poblacional y evolución de la distribución muestral de $\bar{X}$.*

Cuando la población de la que se ha extraído la muestra es normal la distribución dada por la expresión [1.22] es buena aunque el tamaño de la muestra sea pequeño. Pero si la población de partida no es normal entonces la aproximación [1.22] será buena para valores de n grandes, es decir, $n \geqslant 30$.

También es interesante el conocer la distribución muestral de la suma de los cuadrados de variables aleatorias $N(0, 1)$ e independientes, como se indica en el siguiente teorema.

Teorema 1.4

Sea $(X_1, ..., X_n)$ una muestra aleatoria simple de tamaño n, procedente de una poblacioín $N(\mu, \sigma)$. Entonces las variables aleatorias

$$Z_i = \frac{X_i - \mu}{\sigma}$$

son $N(0, 1)$ e independientes y tales que

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

sigue una distribución χ^2 con n grados de libertad.

La demostración no presenta dificultad, pues bastará tener en cuenta la definición y propiedades dadas de la distribución χ^2.

Ejemplo 1.5

El número de libros encuadernados diariamente por una máquina automática sigue una variable aleatoria cuya distribución no se conoce, con una desviación típica de 16 libros por día. Si se selecciona una muestra aleatoria de 49 días, determinar la probabilidad de que el número medio de libros encuadernados durante esos días (la media muestral) se encuentre a lo sumo a 3 libros de la verdadera media poblacional.

$\sigma = 16$
$n = 49.$
$\bar{X} =$

Solución:

Aunque la distribución de la población no es conocida pero como $n = 49$, mayor que 30, entonces la distribución de la media muestral se aproximará a una $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, es decir,

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \equiv N\left(\mu, \frac{16}{\sqrt{49}}\right)$$

O bien, la distribución de la variable aleatoria

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{16/\sqrt{49}} \to N(0, 1)$$

La probabilidad que nos piden, utilizando la Tabla A.7 del anexo A de tablas, será:

$$\begin{aligned}
P(|\bar{X} - \mu| \leqslant 3) &= P(-3 \leqslant (\bar{X} - \mu) \leqslant 3) \\
&= P\left(-\frac{3}{16/\sqrt{49}} \leqslant \frac{\bar{X} - \mu}{16/\sqrt{49}} \leqslant \frac{3}{16/\sqrt{49}}\right) \\
&= P(-1{,}31 \leqslant Z \leqslant 1{,}31) \\
&= F(1{,}31) - F(-1{,}31) \\
&= 0{,}9049 - 0{,}0951 \\
&= 0{,}8098
\end{aligned}$$

Ejemplo 1.6

Refiriéndonos al ejemplo 1.5. Determinar el tamaño de la muestra para que la media muestral se encuentre a lo sumo a 3 libros de la media poblacional con una probabilidad del 0,95.

Solución:

Ahora se tiene que verificar:

$$P(|\bar{X} - \mu| \leqslant 3) = P(-3 \leqslant (\bar{X} - \mu) \leqslant 3) = 0{,}95$$

Dividiendo cada término de la desigualdad por $\dfrac{\sigma}{\sqrt{n}}$, pero $\sigma = 16$, resultará:

$$P\left(-\frac{3}{16/\sqrt{n}} \leqslant \frac{\bar{X} - \mu}{16/\sqrt{n}} \leqslant \frac{3}{16/\sqrt{n}}\right) = P(-0{,}187\sqrt{n} \leqslant Z \leqslant 0{,}187\sqrt{n}) = 0{,}95$$

Luego utilizando la Tabla A.7, del anexo de tablas, se tiene que:

$$P(-1{,}96 \leqslant Z \leqslant 1{,}96) = 0{,}95$$

pues

$$F(-1{,}96) = P(Z \leqslant -1{,}96) = 0{,}025$$
$$F(1{,}96) = P(Z \leqslant 1{,}96) = 0{,}975$$

de donde resulta que

$$0{,}187\sqrt{n} = 1{,}96$$

$$n = \left(\frac{1{,}96}{0{,}187}\right)^2 \simeq 110$$

1.7.2. DISTRIBUCIÓN DE LA MEDIA MUESTRAL CUANDO NO SE CONOCE LA VARIANZA POBLACIONAL

Hasta ahora estábamos admitiendo que se conoce la varianza de la población de la que se extrae la muestra, pero ésta no será la situación general, sino que la mayoría de las veces no conocemos la varianza de la población, entonces como se dispone de una muestra aleatoria simple de tamaño n, podemos calcular la varianza muestral S^2 y utilizarla en lugar de la varianza poblacional σ^2 desconocida, pues S^2 es, como veremos después, un buen estimador de σ^2.

Al hacer esta sustitución si el tamaño de la muestra, n es grande, es decir, $n \geqslant 30$ la distribución del estadístico:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

sigue siendo aproximadamente una $N(0, 1)$.

Si el tamaño de la muestra es pequeño, $n < 30$, los valores de la varianza muestral S^2 varían considerablemente de muestra en muestra, pues S^2 disminuye a medida que n aumenta, y la distribución del estadístico

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

ya no será una distribución normal.

Este problema fue resuelto en 1908 por el estadístico Gosset[14] a partir del siguiente teorema.

---

[14] El estadístico W. S. Gosset trabajaba en una empresa cervecera Irlandesa, la cual prohibía que sus empleados difundieran los resultados de sus investigaciones, y para eludir esta prohibición él publicaba sus trabajos bajo el seudónimo de Student, y de aquí el nombre de la distribución t-Student.

Teorema 1.5

Si $(X_1, ..., X_n)$ es una muestra aleatoria simple, de tamaño n, procedente de una población $N(\mu, \sigma)$ con σ desconocida, entonces el estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \to t\text{-Student con } n - 1 \text{ grados de libertad}$$

Demostración:

Sabemos que

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

y posteriormente veremos que el estadístico:

$$\frac{(n-1)S^2}{\sigma^2} \to \chi^2_{n-1}$$

y que los estadísticos $\bar{X}$ y S^2 son independientes.

Tipificando la variable $\bar{X}$ se tiene:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to N(0, 1)$$

pero incluye el parámetro σ desconocido que es conveniente eliminar.

Recordemos que la variable aleatoria t-Student estaba definida[15] como un cociente entre una variable aleatoria $N(0, 1)$ y la raíz cuadrada de una variable aleatoria χ^2 dividida por sus grados de libertad, ambas independientes. Luego podemos escribir:

---

[15] Sean U y V dos variables aleatorias independientes distribuidas según una $N(0, 1)$ y una χ^2_n respectivamente, entonces la variable aleatoria

$$T = \frac{U}{\sqrt{\frac{V}{n}}} \to t_n \quad (t\text{-Student con } n \text{ grados de libertad})$$

$$T = \frac{U}{\sqrt{\dfrac{V}{n-1}}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \to t_{n-1}$$

pues los estadísticos $\bar{X}$ y S^2 son independientes como veremos en el Teorema 1.6, y en consecuencia también lo son las variables:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad \text{y} \quad \frac{(n-1)S^2}{\sigma^2}$$

1.7.3. DISTRIBUCIÓN DE LA VARIANZA MUESTRAL

Así como al estudiar la distribución del estadístico media muestral decíamos que era de gran utilidad para realizar inferencias, aquí no podemos decir lo mismo del estadístico varianza muestral, pues, la distribución muestral del estadístico S^2 tiene pocas aplicaciones prácticas en estadística, sin embargo, si las tiene el estadístico $\dfrac{(n-1)S^2}{\sigma^2}$ y por ello será el estadístico del que nos ocuparemos en este apartado. Daremos la distribución del estadístico $\dfrac{(n-1)S^2}{\sigma^2}$ mediante el siguiente **teorema de Fisher.**

Teorema 1.6. Teorema de Fisher

Sea $(X_1, ..., X_n)$ una muestra aleatoria simple de tamaño n, procedente de una población $N(\mu, \sigma)$. Entonces se verifica que:

1. Los estadísticos $\bar{X}$ y S^2 son independientes.
2. El estadístico

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \to \chi^2_{n-1} \qquad [1.23]$$

   sigue una distribución χ^2 con $n-1$ grados de libertad.
3. El estadístico

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \to t_{n-1}$$

   Sigue una distribución t-Student con $n-1$ grados de libertad.

Demostración:

1. Para demostrar que los estadísticos media $\bar{X}$ y varianza muestral S^2, son independientes, demostraremos que $\bar{X}$ es independiente de $X_i - \bar{X}$ para cada i, y procederemos directamente calculando la función generatriz de momentos conjunta de $\bar{X}$ y $X_i - \bar{X}$, y tendremos[16]:

$$
\begin{aligned}
g(t_1, t_2) &= E[e^{t_1\bar{X} + t_2(X_i - \bar{X})}] = E[e^{t_2X_i + (t_1 - t_2)\bar{X}}] \\
&= E\left[e^{t_2X_i + (t_1 - t_2)\left(\frac{X_1 + \cdots + X_i + \cdots + X_n}{n}\right)}\right] \\
&= E\left[e^{\left(t_2 + \frac{t_1 - t_2}{n}\right)X_i + \frac{(t_1 - t_2)}{n}\sum\limits_{\substack{j=1 \\ j\neq i}}^{n} X_j}\right] \\
&= E\left[e^{\left(t_2 + \frac{t_1 - t_2}{n}\right)X_i}\right]\cdot E\left[e^{\frac{t_1 - t_2}{n}\sum\limits_{\substack{j=1 \\ j\neq i}}^{n} X_j}\right] \\
&= e^{\left(t_2 + \frac{t_1 - t_2}{n}\right)\mu + \frac{1}{2}\left(t_2 + \frac{t_1 - t_2}{n}\right)^2\sigma^2}\cdot e^{\frac{n-1}{n}(t_1 - t_2)\mu + \left(\frac{t_1 - t_2}{n}\right)^2(n-1)\frac{\sigma^2}{2}} \\
&= e^{t_1\mu + \frac{1}{2}\frac{t_1^2}{n}\sigma^2}\cdot e^{\frac{1}{2}t_2^2\frac{(n-1)}{n}\sigma^2}
\end{aligned}
$$

que son las funciones generatrices de momentos correspondientes a una

$$
N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \text{y} \quad N\left(0, \sigma\sqrt{\frac{n-1}{n}}\right)
$$

respectivamente, con lo cual hemos demostrado que:

I. $\bar{X}$ y $X_i - \bar{X}$ son independientes, y en consecuencia también son independientes $\bar{X}$ y $\sum\limits_{i=1}^{n} (X_i - \bar{X})^2$ y por tanto $\bar{X}$ y S^2 son independientes[17].

---

[16] Como la muestra es aleatoria simple las observaciones son independientes, y tanto X_i como $\sum\limits_{j\neq i} X_j$ son normales, luego bastará tener presente la función generatriz de momentos de la distribución normal.

[17] Recordemos que $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n} (X_i - \bar{X})^2$.

II. $\bar{X}$ es $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$, y

$X_i - \bar{X}$ es $N\left(0, \sigma\sqrt{\dfrac{n-1}{n}}\right)$.

2. Para demostrar que el estadístico $\dfrac{(n-1)S^2}{\sigma^2}$ sigue una χ^2_{n-1}, partimos del estadístico varianza muestral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

de donde podemos escribir:

$$\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2 \\
&= \sum_{i=1}^{n}[(X_i - \mu) - (\bar{X} - \mu)]^2 \\
&= \sum_{i=1}^{n}[(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) + n(\bar{X} - \mu) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2
\end{aligned}$$

y de aquí se tiene:

$$\sum_{i=1}^{n}(X_i - \mu)^2 = (n-1)S^2 + n(\bar{X} - \mu)^2$$

dividiendo ambos miembros por la varianza problacional, resulta:

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

o bien:

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\frac{(n-1)S^2}{\sigma^2}+\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \qquad [1.24]$$

Teniendo en cuenta la definición de la distribución χ_n^2 y su propiedad reproductiva resulta que

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \to \chi_n^2$$

pues tenemos una suma de variables aleatorias $N(0, 1)$ independientes y elevadas al cuadrado.

Análogamente:

$$\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \to \chi_1^2$$

pues se trata de una variable aleatoria $N(0, 1)$ y elevada al cuadrado.

Como admitimos que las variables aleatorias $\frac{(n-1)S^2}{\sigma^2}$ y $\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ son independientes, teniendo en cuenta la propiedad reproductiva de la distribución χ^2, resulta que como:

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \to \chi_n^2$$

entonces tendrá que verificarse que[18]:

$$\frac{(n-1)S^2}{\sigma^2} \to \chi_{n-1}^2$$

La función de densidad de este estadístico $\frac{(n-1)S^2}{\sigma^2}$ será la correspondiente a una χ_{n-1}^2 y por tanto a una $\Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$, y su media será la de una $\Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$ es decir:

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right]=(n-1) \qquad \text{c.q.d.}$$

---

[18] Para mayor rigor podemos calcular la función generatriz de momentos conjunta de $\frac{(n-1)S^2}{\sigma^2}$ y $\frac{n(\bar{X}-\mu)^2}{\sigma}$ y llegaríamos a ver que la función generatriz de momentos de $\frac{(n-1)S^2}{\sigma^2}$ es la correspondiente a la variable aleatoria χ_{n-1}^2.

De aquí, tenemos:

$$\frac{(n-1)}{\sigma^2} E[S^2] = (n-1)$$

$$E[S^2] = \sigma^2$$

Análogamente, la varianza de una $\Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$ es:

$$\text{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

de donde deducimos:

$$\frac{(n-1)^2}{\sigma^4} \text{Var}(S^2) = 2(n-1)$$

$$\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

Luego vemos que las propiedades de la distribución χ^2 se pueden utilizar para encontrar la varianza de la distribución de la varianza muestral, siempre y cuando el modelo de la población de partida sea normal.

Veamos qué significado tiene el término **grados de libertad.** Para ello consideramos el estadístico varianza muestra S^2:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

el cual incluye la suma de cuadrados de las cantidades

$$(X_1 - \bar{X}), \ldots, (X_n - \bar{X})$$

las cuales no son independientes de la información, pues la suma de todas ellas debe ser igual a cero

$$\sum_{i=1}^{n} (X_i - \bar{X}) = \sum_{i=1}^{n} X_i - n\bar{X} = \sigma$$

pues según la definición de $\bar{X}$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Luego si conocemos $(n - 1)$ cualesquiera de estas cantidades $(X_i - \bar{X})$, podemos calcular la restante; así pues, ya que

$$\sum_{i=1}^{n} (X_i - \bar{X}) = 0$$

se deduce que

$$X_n - \bar{X} = -(X_1 - \bar{X}) - \cdots - (X_{n-1} - \bar{X}) = -\sum_{i=1}^{n-1} (X_i - \bar{X})$$

Luego sólo tendremos $n - 1$ cantidades $(X_i - \bar{X})$ independientes.

La situación se puede clarificar algo más, en efecto, supongamos que queremos hacer una inferencia sobre la varianza poblacional σ^2 desconocida. Si la media poblacional μ fuera conocida, esta inferencia se podría basar en la suma de cuadrados de las cantidades.

$$(X_1 - \mu), ..., (X_n - \mu)$$

Estas cantidades son independientes unas de otras, y podríamos decir que tenemos n grados de libertad para estimar la varianza poblacional σ^2. Sin embargo, como la media de la población, en la práctica no suele ser conocida, tiene que ser sustituida por su estimación, es decir, por $\bar{X}$, utilizando por tanto uno de estos grados de libertad, quedando $(n - 1)$ observaciones independientes para utilizarlas en la inferencia sobre la varianza poblacional y entonces decimos que tenemos $(n - 1)$ grados de libertad.

Supongamos que tenemos una población normal y tomamos una muestra aleatoria de esta población con el fin de hacer alguna inferencia sobre la varianza poblacional, entonces utilizando la distribución χ^2 veremos que efectivamente esto es posible, como lo prueba el ejemplo siguiente.

Ejemplo 1.7

En una fábrica conservera se admite que la distribución de pesos de las latas de conservas es normal. El director comercial está muy interesado en que el peso neto del producto incluido en el interior de la lata tenga poca variabilidad, pues en ciertas ocasiones ha observado diferencias entre el peso real y el peso anunciado en la etiqueta. Si se selecciona una muestra aleatoria de 25 latas, obtener los valores k_1 y k_2 tales que

$$P\left(\frac{S^2}{\sigma^2} \leqslant k_1\right) = 0{,}05 \quad \text{y} \quad P\left(\frac{S^2}{\sigma^2} \geqslant k_2\right) = 0{,}05$$

Solución:

Multiplicando ambos miembros de la desigualdad por $(n-1)$ tenemos:

$$0{,}05 = P\left(\frac{S^2}{\sigma^2} \leqslant k_1\right) = P\left(\frac{(n-1)S^2}{\sigma^2} \leqslant (n-1)k_1\right)$$
$$= P(\chi^2_{24} \leqslant 24k_1)$$

Utilizando la Tabla A.9 del anexo de tablas resulta:

$$24k_1 = 13{,}848$$
$$k_1 = 0{,}577$$

Luego

$$P(S^2 \leqslant 0{,}577\sigma^2) = 0{,}05$$

Es decir, existe una probabilidad del 0,05 de que la varianza muestral sea inferior o igual al 57,7 % de la varianza poblacional.

Análogamente calculamos el valor k_2 de manera que:

$$0{,}05 = P\left(\frac{S^2}{\sigma^2} \geqslant k_2\right) = P\left(\frac{(n-1)S^2}{\sigma^2} \geqslant (n-1)k_2\right)$$
$$= P(\chi^2_{24} \geqslant 24k_2)$$

o bien

$$0{,}95 = P(\chi^2_{24} \leqslant 24k_2)$$

Luego de la Tabla A.9 se tiene:

$$24k_2 = 36{,}42$$
$$k_2 = 1{,}517$$

y sustituyendo en la expresión inicial resulta

$$P(S^2 \geqslant 1{,}517\sigma^2) = 0{,}05$$

Es decir, la probabilidad de que la varianza muestral sea mayor o igual que el 151,7 % de la varianza poblacional, es del 0,05.

Gráficamente tendríamos representadas ambas probabilidades en el Gráfico 1.7.

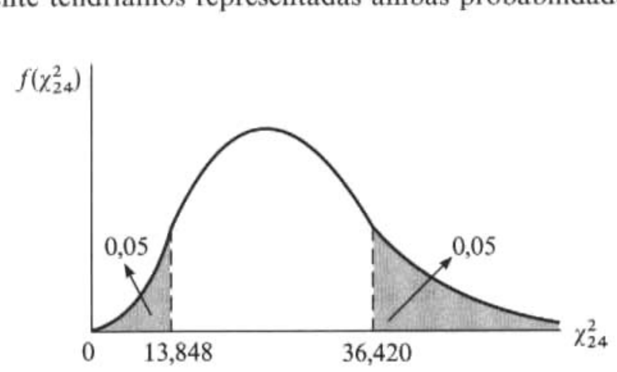


GRÁFICO 1.7. *Representación gráfica de la probabilidad de que la variable aleatoria χ^2_{24} es menor o igual que 13,848 y también de que sea mayor o igual que 36,420.*

1.7.4. DISTRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES CUANDO SE CONOCE LA VARIANZA POBLACIONAL

En muchas situaciones surge la necesidad de comparar las medias muestrales de dos poblaciones distintas. Por ejemplo, supongamos que estamos interesados en comparar los tiempos medios de duración de dos tipos de tubos fluorescentes. La fabricación de ambos tipos de tubos fluorescentes se realiza por compañías distintas y con diferentes procesos de fabricación. Por tanto, los tubos producidos por cada compañía tendrán una distribución diferente, una de la otra, de los tiempos de duración de los tubos.

Designamos por X la variable aleatoria que representa el tiempo de duración del primer tipo de tubos y admitimos que sigue una distribución $N(\mu_x, \sigma_x)$. Análogamente la variable aleatoria Y representa el tiempo de duración del segundo tipo de tubos que sigue una distribución $N(\mu_y, \sigma_y)$. Se selecciona una muestra aleatoria de tamaño n_x del primer tipo de tubos y una muestra aleatoria de tamaño n_y, del segundo tipo de tubos, ambas muestras independientes. Entonces si designamos por $\bar{X}$ e $\bar{Y}$ los estadísticos medias muestrales de ambas muestras, estamos interesados en conocer la distribución muestral de la diferencia $\bar{X} - \bar{Y}$ para las muestras respectivas de tamaño n_x y n_y, procedentes de dos poblaciones normales e independientes.

De manera análoga el Teorema 1.3 que anunciábamos para la distribución muestral de la media, podemos enunciar el siguiente teorema para la diferencia de medias muestrales.

Teorema 1.7

Sean $(X_1, ..., X_{n_x})$ e $(Y_1, ..., Y_{n_y})$ dos muestras aleatorias simples e independientes de tamaños n_x y n_y, procedentes de las poblaciones $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$ respectivamente. Entonces la distribución muestral de la diferencia de medias $\bar{X} - \bar{Y}$, tendrá[19] una distribución normal con media y desviación típica:

$$\mu_{\bar{X}-\bar{Y}} = \mu_x - \mu_y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

es decir

$$\bar{X} - \bar{Y} \to N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)$$

De donde el estadístico

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \to N(0, 1)$$

Demostración:

Por el Teorema 1.3 sabemos que:

$$\bar{X} \to N\left(\mu_x, \frac{\sigma_x}{\sqrt{n_x}}\right)$$

$$\bar{Y} \to N\left(\mu_y, \frac{\sigma_y}{\sqrt{n_y}}\right)$$

[19] Si las distribuciones no son normales y los tamaños muestrales n_x y n_y son grandes, mayores o iguales que 30, entonces por el Teorema Central del Límite la aproximación normal para la distribución de $\bar{X} - \bar{Y}$ es muy buena. Sin embargo si n_x y n_y son pequeños entonces la forma de la distribución muestral de $\bar{X} - \bar{Y}$ dependerá de la naturaleza de la población muestreada.

y sus respectivas funciones generatrices de momentos son:

$$g_{\bar{X}}(t) = E[e^{t\bar{X}}] = e^{t\mu_x + \frac{1}{2}t^2 \frac{\sigma_x^2}{n_x}}$$

$$g_{\bar{Y}}(t) = E[e^{t\bar{Y}}] = e^{t\mu_y + \frac{1}{2}t^2 \frac{\sigma_y^2}{n_y}}$$

Luego la función generatriz de momentos de $\bar{X} - \bar{Y}$ será:

$$\begin{aligned} g_{\bar{X}-\bar{Y}}(t) &= E[e^{t(\bar{X}-\bar{Y})}] = E[e^{t\bar{X}}]\cdot E[e^{-t\bar{Y}}] \\ &= e^{t\mu_x + \frac{1}{2}t^2 \frac{\sigma_x^2}{n_x}} \cdot e^{-t\mu_y + \frac{1}{2}t^2 \frac{\sigma_y^2}{n_y}} \\ &= e^{t(\mu_x - \mu_y) + \frac{1}{2}t^2 \left(\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)} \end{aligned}$$

Y teniendo en cuenta la unicidad de la función generatriz de momentos resulta que:

$$\bar{X} - \bar{Y} \to N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right) \qquad [1.25]$$

Si las dos muestras provienen de poblaciones tales que $\mu_1 = \mu_2$, entonces:

$$\bar{X} - \bar{Y} \to N\left(0, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right) \qquad [1.26]$$

o bien, si $\sigma_x^2 = \sigma_y^2 = \sigma^2$, es decir, tienen la misma varianza, entonces:

$$\bar{X} - \bar{Y} \to N\left(\mu_x - \mu_y, \sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\right) \qquad [1.27]$$

De la expresión [1.25], tipificando se tiene:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}} \to N(0, 1) \qquad [1.28]$$

Ejemplo 1.8

Analizando los salarios de los trabajadores de dos Comunidades Autónomas se deduce que en la Comunidad A el salario medio es de 129.000 ptas. con una varianza de 2.500 ptas.^{2}, y en la Comunidad B el salario medio es de 128.621 ptas. con una varianza de 3.000 ptas.^{2}. Si tomamos una muestra aleatoria de 36 personas en la Comunidad A y de 49 personas en la Comunidad B, determinar la probabilidad de que la muestra procedente de la Comunidad A tenga un salario medio que sea al menos 400 ptas. superior al salario medio de la Comunidad B.

Solución:

Observamos que no hemos dicho que las poblaciones, de partida son normales, pues no es necesario ya que como los tamaños muestrales $n_x = 36$ y $n_y = 49$, son mayores o iguales que 30, la aproximación a la distribución normal dada por la expresión [1.26] es muy buena, sin necesidad de que las poblaciones de partida sean normales.

La información que tenemos es:

Población A: $\mu_x = 129.000$, $\sigma_x^2 = 2.500$ $n_x = 36$
Población B: $\mu_y = 128.621$, $\sigma_y^2 = 3.000$ $n_y = 49$

Aplicando el Teorema 1.7, la distribución muestral de la diferencia de los salarios medias muestrales $\bar{X} - \bar{Y}$ será:

$$\bar{X} - \bar{Y} \to N\left(129.000 - 128.621, \sqrt{\frac{2.500}{36} + \frac{3.000}{49}}\right)$$

$$\bar{X} - \bar{Y} \to N(379, 11{,}43)$$

La representación gráfica de esta distribución $N(379, 11{,}43)$, está dada en el Gráfico 1.8.

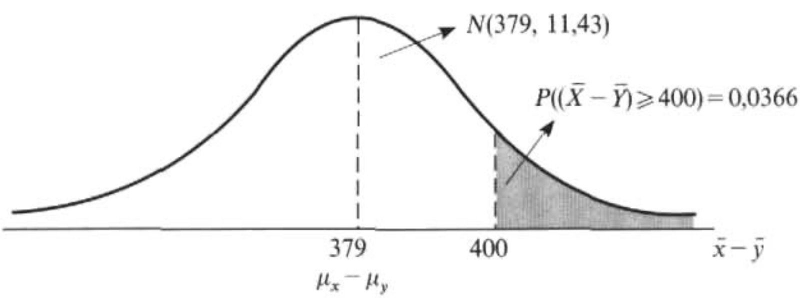


GRÁFICO 1.8. *Representación gráfica de la distribución muestral de la diferencias de medias correspondiente al ejemplo 1.8.*

La probabilidad de que el salario medio muestral de la Comunidad A sea al menos 400 ptas. superior al salario medio muestral de la Comunidad B corresponde a la zona sombreada y viene dado por:

$$
\begin{aligned}
P((\bar{X} - \bar{Y}) \geqslant 400) &= P\left(\frac{\bar{X} - \bar{Y} - 379}{11{,}43} \geqslant \frac{400 - 379}{11{,}43}\right) \\
&= P(Z \geqslant 1{,}83) \\
&= 1 - P(Z \leqslant 1{,}83) \\
&= 1 - F(1{,}83) \\
&= 1 - 0{,}9664 \\
&= 0{,}0336
\end{aligned}
$$

Este resultado nos dice que la probabilidad, de que la media de una muestra aleatoria de 36 salarios de la Comunidad A exceda en 400 o más pesetas a la media de una muestra aleatoria de 49 salarios de la Comunidad B, es 0,0336.

1.7.5. DISTRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES CUANDO NO SE CONOCE LA VARIANZA POBLACIONAL

En general, en situaciones reales las varianzas poblacionales no suelen ser conocidas. Así pues, ahora queremos obtener la distribución de la diferencia de medias muestrales $\bar{X} - \bar{Y}$ cuando el muestro se realiza sobre dos poblaciones normales, independientes y con varianzas desconocidas.

Es decir, consideramos dos poblaciones normales e independientes, $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$ y seleccionamos una muestra aleatoria simple de tamaño n_x de la primera población y otra muestra aleatoria simple de tamaño n_y, independiente de la anterior, y procedente de la segunda población, entonces pueden presentarse dos situaciones:

a) $\sigma_x = \sigma_y = \sigma$ (las varianzas poblacionales son iguales).

b) $\sigma_x \neq \sigma_y$ (las varianzas poblacionales son distintas).

a) Las varianzas poblacionales son iguales

$$\sigma_x = \sigma_y = \sigma$$

Por los Teoremas 1.3 y 1.6 sabemos que:

$$\bar{X} \to N\left(\mu_x, \frac{\sigma}{\sqrt{n_x}}\right) \quad , \quad \frac{(n_x - 1)S_x^2}{\sigma^2} \to \chi^2_{n_x - 1}$$

$$\bar{Y} \to N\left(\mu_y, \frac{\sigma}{\sqrt{n_y}}\right) \quad , \quad \frac{(n_y - 1)S_y^2}{\sigma^2} \to \chi^2_{n_y - 1}$$

Como las muestras son independientes, también serán independientes las varianzas muestrales S_x^2 y S_y^2 y por tanto los estadísticos

$$\frac{(n_x - 1)S_x^2}{\sigma^2} \quad \text{y} \quad \frac{(n_y - 1)S_y^2}{\sigma^2}$$

son variables aleatorias independientes distribuidas según una χ^2 con $n_x - 1$ y una χ^2 con $n_y - 1$ grados de libertad, respectivamente.

Teniendo en cuenta la propiedad reproductiva de la distribución χ^2 resulta que la variable aleatoria W

$$W = \frac{(n_x - 1)S_x^2}{\sigma^2} + \frac{(n_y - 1)S_y^2}{\sigma^2} \to \chi^2_{n_x + n_y - 2}$$

también sigue una distribución χ^2 con $n_x + n_y - 2$ grados de libertad.

También sabemos, por el Teorema 1.7, que

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma^2}{n_x} + \dfrac{\sigma^2}{n_y}}} \to N(0, 1)$$

y como las variables aleatorias Z y W son independientes, teniendo en cuenta la definición de la variable t-Student, resulta que:

$$T = \frac{Z}{\sqrt{\dfrac{W}{n_x + n_y - 2}}} \to t_{n_x + n_y - 2} \qquad [1.29]$$

Luego, sustituyendo en la expresión [1.29] tenemos:

$$T = \frac{\dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}}}{\sqrt{\dfrac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{\sigma^2} \Big/ (n_x + n_y - 2)}} =$$

$$= \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}\ \sqrt{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}} \sqrt{n_x + n_y - 2}$$

$$= \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}} \frac{\sqrt{n_x + n_y - 2} \cdot \sqrt{n_x \cdot n_y}}{\sqrt{n_x + n_y}} \to t_{n_x + n_y - 2} \qquad [1.30]$$

es decir, sigue una distribución t-Student con $n_x + n_y - 2$ grados de libertad.

b) *Las varianzas poblacionales son distintas.*

$$\sigma_x \neq \sigma_y$$

En este caso encontrar una distribución de la diferencia de medias poblacionales que nos pueda ser útil después para la obtención de un intervalo de confianza, no es fácil, y se le conoce con el nombre de **problema de Behrens-Fisher.** Bajo condiciones especiales se puede encontrar alguna distribución, pero el obtener una solución general no es sencillo, nosotros proporcionaremos algunas aproximaciones.

Si las varianzas poblacionales son distintas y desconocidas utilizamos las varianzas muestrales S_x^2 y S_y^2 como estimadores de σ_x^2 y σ_y^2.

Cuando los tamaños muestrales son grandes, es decir, $n_x \geqslant 30$ y $n_y \geqslant 30$, entonces el estadístico

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}}} \to N(0,\ 1) \qquad [1.31]$$

pues para n_x y n_y grandes S_x^2 y S_y^2 son muy buenos estimadores de σ_x^2 y σ_y^2, ya que, como veremos después, la varianza muestral es un estimador insesgado de la varianza poblacional.

Si las muestras son pequeñas, el estadístico

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}}} \to t_v \qquad [1.32]$$

es decir, sigue una t-Student con v-grados de libertad, siendo:

$$v = \frac{\left(\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}\right)^2}{\dfrac{(S_x^2/n_x)^2}{n_x - 1} + \dfrac{(S_y^2/n_y)^2}{n_y - 1}} \qquad [1.33]$$

Tomaremos por valor de v el valor entero más próximo.

1.7.6. DISTRIBUCIÓN DEL COCIENTE DE VARIANZAS

Sean dos poblaciones X e Y normales $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$ e independientes, de las cuales seleccionamos dos muestras aleatorias simples e independientes, de tamaños n_x y n_y, $(X_1, ..., X_{n_x})$ e $(X_1, ..., X_{n_y})$, entonces pueden presentarse fundamentalmente dos situaciones:

a) μ_x y μ_y conocidas.

b) μ_x y μ_y desconocidas.

a) Las medias poblacionales son conocidas

Al ser conocidas las medias poblacionales μ_x y μ_y las podemos utilizar para el cálculo de las varianzas muestrales S_x^2 y S_y^2 y como las muestras son independientes y además proceden de distintas poblaciones, entonces los estadísticos:

$$S_x^{*2} = \frac{1}{n_x} \sum_{i=1}^{n_x} (X_i - \mu_x)^2$$

$$S_y^{*2} = \frac{1}{n_y} \sum_{i=1}^{n_y} (Y_i - \mu_y)^2$$

son independientes y podemos expresarlos como:

$$n_x S_x^{*2} = \sum_{i=1}^{n_x} (X_i - \mu_x)^2 \quad \Rightarrow \quad \frac{n_x S_x^{*2}}{\sigma_x^2} = \sum_{i=1}^{n_x} \left(\frac{X_i - \mu_x}{\sigma_x}\right)^2 \rightarrow \chi^2_{n_x}$$

$$n_y S_y^{*2} = \sum_{i=1}^{n_y} (Y_i - \mu_y)^2 \quad \Rightarrow \quad \frac{n_y S_y^{*2}}{\sigma_y^2} = \sum_{i=1}^{n_y} \left(\frac{Y_i - \mu_y}{\sigma_y}\right)^2 \rightarrow \chi^2_{n_y}$$

pues la suma de n variables aleatorias $N(0, 1)$, independientes y elevadas al cuadrado siguen una χ^2_n.

Y recordando que la variable aleatoria F de Snedecor con n_x y n_y grados de libertad, F_{n_x, n_y}, se define como un cociente entre dos variables aleatorias χ^2 independientes y divididas cada una de ellas por sus grados de libertad, tendríamos:

$$F = \frac{\dfrac{n_x S_x^{*2}}{\sigma_x^2} \Big/ n_x}{\dfrac{n_y S_y^{*2}}{\sigma_y^2} \Big/ n_y} = \frac{S_x^{*2}}{S_y^{*2}} \cdot \frac{\sigma_y^2}{\sigma_x^2} \rightarrow F_{n_x, n_y}$$

b) Las medias poblacionales son desconocidas

Al ser desconocidas las medias poblacionales, que será lo que casi siempre ocurra, y ser las muestras independientes y además procedentes de distintas poblaciones, entonces los estadísticos:

$$S_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (X_i - \bar{X})^2$$

$$S_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (Y_i - \bar{Y})^2$$

son independientes y teniendo en cuenta el Teorema 1.6 resulta:

$$(n_x - 1)S_x^2 = \sum_{i=1}^{n_x} (X_i - \bar{X})^2 \quad \Rightarrow \quad \frac{(n_x - 1)S_x^2}{\sigma_x^2} = \sum_{i=1}^{n_x} \left(\frac{X_i - \bar{X}}{\sigma_x}\right)^2 \to \chi^2_{n_x - 1}$$

$$(n_y - 1)S_y^2 = \sum_{i=1}^{n_y} (Y_i - \bar{Y})^2 \quad \Rightarrow \quad \frac{(n_y - 1)S_y^2}{\sigma_y^2} = \sum_{i=1}^{n_y} \left(\frac{Y_i - \bar{Y}}{\sigma_y}\right)^2 \to \chi^2_{n_y - 1}$$

Análogamente a como ocurría en la situación anterior, llegaremos a una F-Snedecor con $n_1 - 1$ y $n_2 - 1$ grados de libertad, en efecto:

$$F = \frac{\dfrac{(n_x - 1)S_x^2}{\sigma_x^2} \Big/ (n_x - 1)}{\dfrac{(n_y - 1)S_y^2}{\sigma_y^2} \Big/ (n_y - 1)} = \frac{S_x^2}{S_y^2} \cdot \frac{\sigma_y^2}{\sigma_x^2} \to F_{n_x - 1, n_y - 1}$$

A partir de aquí podremos obtener la distribución del cociente de varianzas $\dfrac{\sigma_y^2}{\sigma_x^2}$, así pues la función de distribución será:

$$P\left(\frac{S_x^2}{S_y^2} \leqslant v\right) = P\left(\frac{S_x^2}{S_y^2} \cdot \frac{\sigma_y^2}{\sigma_x^2} \leqslant \frac{\sigma_y^2}{\sigma_x^2}\, v\right)$$

$$= P\left(F_{n_x - 1, n_y - 1} \leqslant \frac{\sigma_y^2}{\sigma_x^2}\, v\right)$$

que será el valor que toma la función de distribución de una F-Snedecor con $n_x - 1$ y $n_y - 1$ grados de libertad en el punto $\dfrac{\sigma_y^2}{\sigma_x^2}\, v$.

También podríamos estudiar otras situaciones:

μ_x conocida y μ_y desconocida

μ_x desconocida y μ_y conocida

pero son similares a los casos anteriores; así pues llegaríamos a tener: $F_{n_x, n_y - 1}$ y $F_{n_x - 1, n_y}$, respectivamente.

1.8. DISTRIBUCIÓN DE LA PROPORCIÓN MUESTRAL

Sea $(X_1, ..., X_n)$ una muestra aleatoria simple de tamaño n, procedente de una $B(1, p)$, y sabemos que el estadístico **proporción muestral** será también una variable aleatoria,

$$P_X = \frac{X}{n}$$

que tomará diferentes valores para cada una de las posibles muestras, así pues para una muestra concreta $(x_1, ..., x_n)$ el valor del estadístico proporción muestral será:

$$\hat{p} = \frac{x}{n}$$

en donde x representa el número de elementos de la muestra que poseen la característica que estamos investigando y la variable aleatoria X sigue una distribución binomial $B(n, p)$.

Luego, la distribución muestral del estadístico proporción muestral tendrá la misma forma que la distribución binomial de X y como la distribución binomial se puede aproximar a la normal cuando n es grande, $n \geqslant 30$ entonces teniendo en cuenta el Teorema Central del Límite resulta que el estadístico proporción muestral sigue una distribución normal, es decir:

$$\hat{p} = P_X = \frac{X}{n} \to N\left(p, \sqrt{\frac{pq}{n}}\right) \qquad [1.34]$$

pues,

$$E[\hat{p}] = E[P_X] = E\left[\frac{X}{n}\right] = \frac{1}{n} E[X] = \frac{1}{n} \cdot np = p \qquad [20]$$

$$\operatorname{Var}(\hat{p}) = \operatorname{Var}(P_X) = \operatorname{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \operatorname{Var}(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} = \frac{pq}{n}$$

[20] Lo cual nos permite decir, cómo veremos en el capítulo siguiente que el estadístico proporción muestral P es un estimado insesgado de la proporción poblacional.

También se verifica, para muestras grandes, que

$$Z = \frac{P_X - p}{\sqrt{\frac{pq}{n}}} \rightarrow N(0, 1) \qquad [1.35]$$

A la **desviación estándar** de la proporción muestral, que es la raíz cuadrada de la varianza, le llamaremos **error estándar de la proporción** y viene dado por:

$$\text{error estándar del estadístico proporción muestral } \hat{p} = \sqrt{\frac{pq}{n}} \qquad [1.36]$$

De manera análoga a como ocurría con el estadístico media muestral, aquí resulta que para un parámetro p fijo, el error estándar de la proporción muestral disminuye cuando el tamaño de la muestra aumenta. Lo cual implica que cuando el tamaño de la muestra aumenta la distribución del estadístico proporción muestral $\hat{p}$ está más concentrada en torno a su media, es decir, en torno a la proporción poblacional como se indica en el Gráfico 1.9.

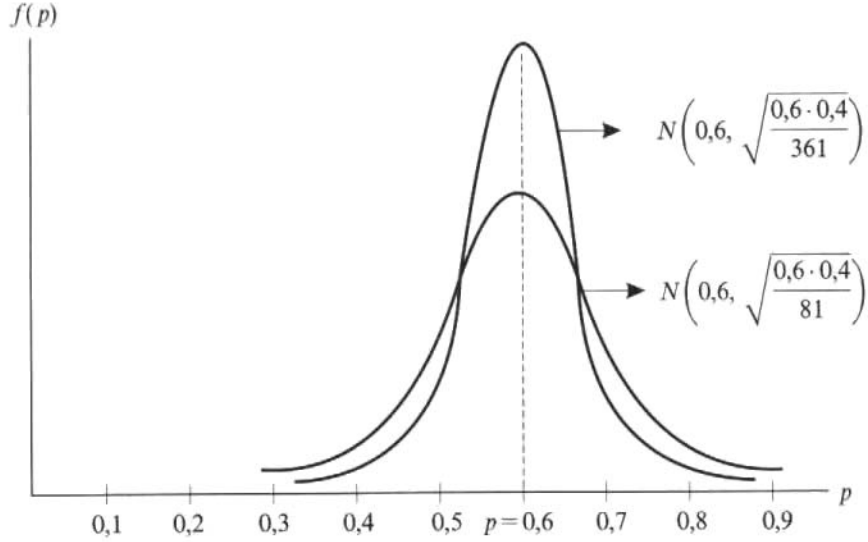


GRÁFICO 1.9. *Representación gráfica de las funciones de densidad del estadístico proporción muestral para muestras de tamaño $n = 81$ y $n = 361$, de una población cuya proporción poblacional es $p = 0{,}6$.*

Ejemplo 1.9

Supongamos que el 30 % de la población de viviendas de un país tienen más de un cuarto de aseo. Con el fin de obtener una información más precisa se toma una muestra aleatoria de tamaño 400 viviendas. Obtener:

1.º La probabilidad de que la proporción de viviendas de la muestra con más de un aseo esté comprendida entre 0,25 y 0,32.

2.º La probabilidad de que el porcentaje de viviendas de la muestra con más de un aseo sea superior al 33 %.

Solución:

Sabemos que el parámetro proporción poblacional es $p = 0{,}3$ y de la expresión [1.34] resulta que el estadístico proporción muestral $\hat{p} = \frac{X}{n}$ sigue una distribución $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

1.º Si notamos por $\hat{p} = \frac{X}{n}$ el estadístico proporción muestral, desearemos encontrar:

$$P(0{,}25 \leqslant \hat{p} \leqslant 0{,}32) = P\left(\frac{0{,}25 - p}{\sqrt{\frac{pq}{n}}} \leqslant \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \leqslant \frac{0{,}32 - p}{\sqrt{\frac{pq}{n}}}\right)$$

$$= P\left(\frac{0{,}25 - 0{,}3}{\sqrt{\frac{0{,}3 \cdot 0{,}7}{400}}} \leqslant Z \leqslant \frac{0{,}32 - 0{,}3}{\sqrt{\frac{0{,}3 \cdot 0{,}7}{400}}}\right)$$

$$= P(-2{,}18 \leqslant Z \leqslant 0{,}873)$$

Utilizando la Tabla A.7 del anexo, resulta:

$$\begin{aligned} P(0{,}25 \leqslant \hat{p} \leqslant 0{,}32) &= P(-2{,}18 \leqslant Z \leqslant 0{,}873) \\ &= F(0{,}873) - F(-2{,}18) \\ &= 0{,}8078 - 0{,}0146 \\ &= 0{,}7932 \end{aligned}$$

Luego la proporción muestral de viviendas que tienen más de un aseo, caerá en el interior del intervalo (0,25, 0,32) para aproximadamente el 79,32 % de las muestras de tamaño 400 procedentes de esta población.

2.° Análogamente, tenemos:

$$P(\hat{p} > 0{,}33) = P\left(\frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} > \frac{0{,}33 - p}{\sqrt{\frac{pq}{n}}}\right)$$

$$= P\left(Z > \frac{0{,}33 - 0{,}3}{\sqrt{\frac{0{,}3 \cdot 0{,}7}{400}}}\right)$$

$$= P(Z > 1{,}31)$$

$$= 1 - P(Z \leqslant 1{,}31)$$

$$= 1 - F(1{,}31)$$

$$= 1 - 0{,}9049$$

$$= 0{,}0951$$

Ejemplo 1.10

Examinados los incrementos salariales de los altos ejecutivos de un amplio grupo de empresas se observa que se distribuyen según una distribución normal de media 12,1 % y de desviación típica 3,5 %. Se toma una muestra aleatoria de 16 observaciones de la población de incrementos salariales. Determinar la probabilidad de que la media muestral sea igual o inferior al 10 %.

Solución:

Sabemos que:

la media poblacional $\mu = 12.1$

la desviación típica poblacional $\sigma = 3{,}5$

tamaño $n = 16$

$$\bar{X} \to N\left(12{,}1, \frac{3{,}5}{\sqrt{16}}\right) \equiv N(12{,}1, 0{,}875)$$

La media muestral es $\bar{X}$ y deseamos obtener:

$$P(\bar{X} \leqslant 10) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqslant \frac{10 - \mu}{\sigma/\sqrt{n}}\right)$$
$$= P\left(Z \leqslant \frac{10 - 12{,}1}{0{,}875}\right)$$
$$= P(Z \leqslant -2{,}4)$$

Utilizando la Tabla A.7 del anexo, resulta:

$$P(\bar{X} \leqslant 10) = P(Z \leqslant -2{,}4)$$
$$= F(-2{,}4)$$
$$= 0{,}0082$$

Luego la probabilidad de que la media de la muestra sea menor o igual que el 10 % es de solamente 0,0082.

1.9. DISTRIBUCIÓN DE LA DIFERENCIA DE PROPORCIONES

Otro problema que se suele presentar es el de comparar las proporciones p_x y p_y de dos poblaciones binomiales, $B(1, p_x)$ y $B(1, p_y)$, basándose en muestras aleatorias simples de tamaño n_x y n_y, respectivamente, extraídas de ambas poblaciones.

Así pues, sean dos muestras aleatorias simples e independientes de tamaño n_x y n_y y procedentes de poblaciones binomiales con parámetros p_x y p_y respectivamente, entonces la distribución muestral de la diferencia de proporciones muestrales

$$\hat{p}_x - \hat{p}_y = \frac{X}{n_x} - \frac{Y}{n_y}$$

tendrá aproximadamente (para n_x y n_y grandes) una distribución normal con media y desviación típica

$$\mu_{\hat{p}_x - \hat{p}_y} = p_x - p_y$$

$$\sigma_{\hat{p}_x - \hat{p}_y} = \sqrt{\frac{p_x q_x}{n_x} + \frac{p_y q_y}{n_y}}$$

es decir,

$$\hat{p}_x - \hat{p}_y \to N\left(p_x - p_y, \sqrt{\frac{p_x q_x}{n_x} + \frac{p_y q_y}{n_y}}\right) \qquad [1.37]$$

Capítulo 2
ESTIMACIÓN PUNTUAL

2.1. INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

Sabemos que una población puede ser caracterizada por los valores de algunos parámetros poblacionales, por ello es lógico que en muchos problemas estadísticos se centre la atención sobre esos parámetros poblacionales. Por ejemplo, supongamos la población de tubos fluorescentes, en donde la característica que estamos investigando es el tiempo de duración del tubo y nos interesa conocer la duración media, es decir el parámetro poblacional μ. El valor de este parámetro poblacional μ podía ser calculado utilizando cada tubo fluorescente de la población, anotando su tiempo de duración y calculando la media de todos los tiempos de duración de todos los tubos de la población. Pero, evidentemente, no sería posible calcular el valor de μ de esta forma, pues el proceso de observar el tiempo de duración de cada tubo de la población es destructivo, y no quedarían tubos fluorescentes para la venta. Un método alternativo sería, seleccionar una muestra de tubos fluorescentes, observar el tiempo de duración de cada uno y calcular su media, la cual sería la estimación o valor aproximado de μ. En este caso el estadístico media muestral $\bar{X}$, función de las observaciones muestrales, o variables aleatorias de la muestra $X_1, X_2, ..., X_n$, es el utilizado para la estimación del parámetro poblacional μ. Como después veremos, el estadístico media muestral es el mejor estadístico para estimar la media poblacional μ.

Vemos pues que en muchos casos no será posible determinar el valor de un parámetro poblacional analizando todos los valores poblacionales, pues el proceso a seguir para determinar el valor del parámetro puede ser destructivo, como en el ejemplo anterior, o nos puede costar mucho tiempo o mucho dinero el analizar cada unidad poblacional. En estas situaciones la única salida que tenemos es utilizar, la **inferencia estadística** para obtener información sobre los valores de los parámetros poblacionales, basándonos en la información contenida en una muestra aleatoria.

Parámetros poblacionales importantes son: la media, la desviación típica y la proporción poblacional[1]. Así, por ejemplo, nos puede interesar tener información sobre:

— La renta media de todas las familias de una ciudad.

— El tiempo medio de espera en la caja de un supermercado.

— La desviación estándar del error medida de un instrumento electrónico.

— La proporción de familias que poseen televisor en color.

— La proporción de automóviles que se averían durante el primer año de garantía, etc.

El objetivo básico de la inferencia estadística es hacer inferencias o sacar conclusiones sobre la población a partir de la información contenida en una muestra aleatoria de la población. Más específicamente, podemos decir que la **inferencia estadística** consiste en el proceso de selección y utilización de un estadístico muestral, mediante el cual, utilizando la información que nos proporciona una muestra aleatoria, nos permite sacar conclusiones sobre características poblacionales.

Un esquema de la inferencia estadística aparece en el gráfico 2.1, en donde la población se representa por su función de distribución y el parámetro poblacional se nota por θ, y toma valores dentro del **espaio paramétrico Ω**; el parámetro puede ser cualquiera, por ejemplo, la media μ, la desviación típica σ, o la proporción poblacional p. Seleccionamos una función de las variables aleatorias muestrales $X_1, X_2, ..., X_n$, que la notaremos por $\hat{\theta} = g(X_1, X_2, ..., X_n)$ y la utilizaremos para obtener la inferencia sobre el valor del parámetro θ.

La función $\hat{\theta}$ es un **estadístico** cuyo valor depende de los valores de las variables aleatorias muestrales $X_1, X_2, ..., X_n$, es decir, el estadístico $\hat{\theta}$ es función de las observaciones muestrales, luego para cada muestra determinada $(x_1, x_2, ..., x_n)$ tomará un valor diferente, y por tanto $\hat{\theta}$, será una variable aleatoria.

[1] En lo sucesivo continuaremos con la norma de utilizar las letras mayúsculas para designar las variables aleatorias, los estadísticos, los estimadores, y la muestra aleatoria en general, y usuaremos letras minúsculas para designar los valores concretos que pueden tomar las variables aleatorias, los estadísticos, y la muestra aleatoria particular o concreta.

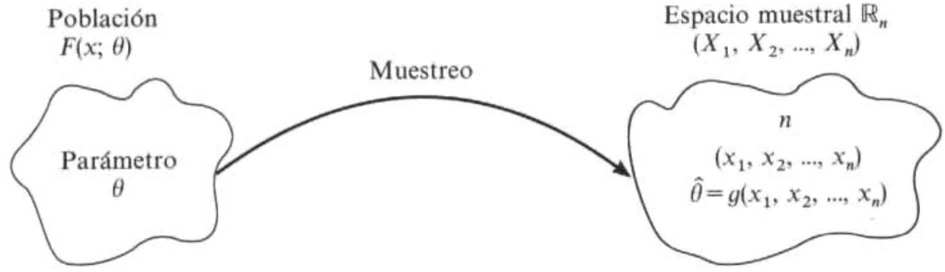


GRÁFICO 2.1. *Esquema de inferencia estadística sobre el parámetro θ.*

Por ejemplo, supongamos que estamos interesados en el parámetro varianza poblacional σ^2. El estadístico muestral que utilizaremos para obtener la inferencia sobre σ^2 es la varianza muestral S^2, es decir

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

en donde las observaciones $(x_1, x_2, ..., x_n)$ corresponden a los valores de una muestra aleatoria determinada por las variables muestrales $X_1, X_2, ..., X_n$.

Un esquema gráfico aparece en el gráfico 2.2, en donde el parámetro poblacional se nota por σ^2.

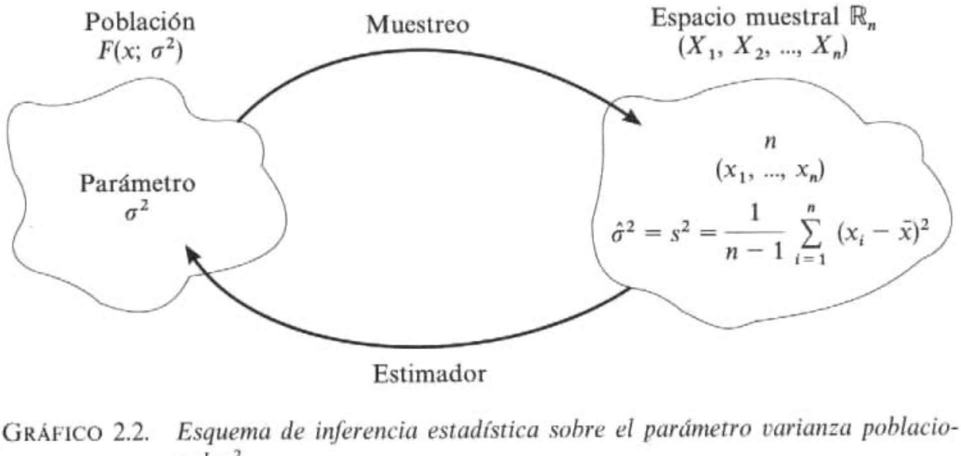


GRÁFICO 2.2. *Esquema de inferencia estadística sobre el parámetro varianza poblacional σ^2.*

Cualquier inferencia o conclusión obtenida de la población, necesariamente, estará basada en un **estadístico muestral,** es decir, en la información proporcionada por la muestra[2]. La elección del estadístico apropiado dependerá de cuál sea el parámetro poblacional que nos interese. El valor verdadero del parámetro será desconocido y un objetivo sería estimar su valor, por lo que tal estadístico se denomina **estimador.**

Las inferencias sobre el valor de un parámetro poblacional θ se pueden obtener básicamente de dos maneras: a partir de **estimación** o bien a partir de la **contrastación de hipótesis.**

En la **estimación,** basta seleccionar un estadístico muestral cuyo valor se utilizará como estimador del valor del parámetro poblacional.

En la **contrastación de hipótesis,** se hace una hipótesis sobre el valor del parámetro θ y se utiliza la información proporcionada por la muestra para decidir si la hipótesis se acepta o no. Por ejemplo, supongamos que estamos interesados en el parámetro proporción poblacional, es decir la proporción de personas que no piensan votar en las próximas Elecciones Generales. Hacemos una hipótesis previa que podría ser: que el valor de la proporción poblacional p será 0,40 o mayor, $p \geqslant 0{,}40$. Se toma una muestra aleatoria de votantes de la población total, y la proporción muestral $\hat{p}$ de aquellos electores que no piensan votar se utilizan para decidir si la hipótesis formulada era razonable o no.

Ambos métodos de inferencia estadística utilizan las mismas relaciones teóricas entre resultados muestrales y valores poblacionales. Así pues, una muestra es sacada de la población y un estadístico muestral es utilizado para hacer inferencias sobre el parámetro poblacional. En estimación, la información muestral es utilizada para estimar el valor del parámetro θ. En el contraste de hipótesis, primero se formula la hipótesis sobre el valor de θ y la información muestral se utiliza para decidir si la hipótesis formulada debería ser o no rechazada.

Pero cuando se utiliza la inferencia para estimar un parámetro poblacional debemos decir cómo de buena es esa inferencia, osea debemos de dar una medida de su **bondad.** Para ello será necesario conocer la diferencia existente entre la estimación del parámetro poblacional, calculada a partir de una muestra específica de tamaño n, y el valor verdadero del parámetro poblacional. En el contraste de hipótesis la **bondad** de la inferencia se mide por la probabilidad de que la decisión de rechazar o no rechazar el valor dado en la hipótesis sobre parámetro poblacional sea correcta.

En este capítulo nos ocuparemos de la estimación estadística y más concretamente de la **estimación puntual** y dejaremos para capítulos posteriores la **estimación por intervalos** y la **contrastación de hipótesis.**

---

[2] Formalmente definimos un **estadístico** como una función de las observaciones muestrales.

2.2. EL PROBLEMA DE LA ESTIMACIÓN: ESTIMACIÓN PUNTUAL

La estimación estadística se divide en dos grandes grupos: la **estimación puntual** y la **estimación por intervalos.** La **estimación puntual** consiste en obtener un único número, calculado a partir de las observaciones muestrales, y que es utilizado como estimación del valor del parámetro θ. Se le llama estimación puntual porque a ese número, que se utiliza como estimación del parámetro θ, se le puede asignar un punto sobre la recta real. En la **estimación por intervalos** se obtienen dos puntos (un extremo inferior y un extremo superior) que definen un intervalo sobre la recta real, el cual contendrá con cierta seguridad el valor del parámetro θ. Por ejemplo, si el parámetro poblacional es la duración de la población de tubos fluorescentes, basándonos en la información proporcionada por una muestra podríamos obtener una estimación puntual del parámetro μ, que lo notaremos por $\hat{\mu}$, $\hat{\mu} = 525$ horas, sin embargo, el intervalo de estimación para μ sería de la forma (475, 575), es decir, de 475 a 575 horas, con un cierto margen de seguridad.

Un esquema de la estimación puntual aparece en el gráfico 2.3 en donde la población viene representada por su función de distribución $F(x; \theta)$, siendo θ el parámetro poblacional desconocido que tomará valores en el **espacio paramétrico Ω** y la muestra aleatoria de tamaño n, está compuesta por las n variables aleatorias $X_1, X_2, ..., X_n$.

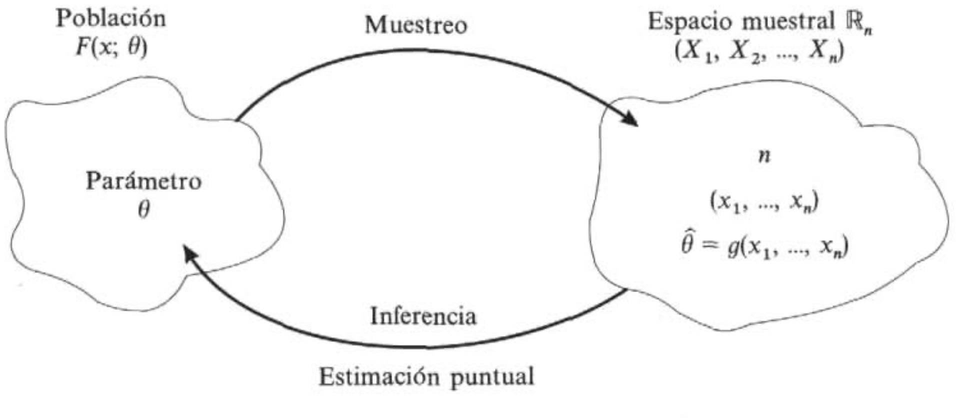


GRÁFICO 2.3. *Esquema de estimación puntual del parámetro θ compuesta por las n variables aleatorias $X_1, X_2, ..., X_n$.*

El **estimador** del parámetro poblacional θ es una función de las variables aleatorias u observaciones muestrales y se representa por

$$\hat{\theta} = g(X_1, X_2, ..., X_n)$$

Para una realización particular de la muestra $(x_1, x_2, ..., x_n)$ se obtiene un valor específico del estimador que recibe el nombre de **estimación** del parámetro poblacional θ y lo notaremos por

$$\hat{\theta} = g(x_1, x_2, ..., x_n)$$

Vemos pues que existe diferencia entre **estimador** y **estimación.** Utilizaremos el término estimador cuando nos referimos a la función de las variables aleatorias muestrales $X_1, X_2, ..., X_n$, y los valores que toma la función estimador para las diferentes realizaciones o muestras concretas serán las **estimaciones.** El estimador es un estadístico y, por tanto, una variable aleatoria y el valor de esta variable aleatoria para una muestra concreta $(x_1, x_2, ..., x_n)$ será la **estimación puntual.**

El estimador $\hat{\theta}$ tendrá su distribución muestral, así pues para diferentes realizaciones de una muestra de tamaño n se tendrá el gráfico 2.4.

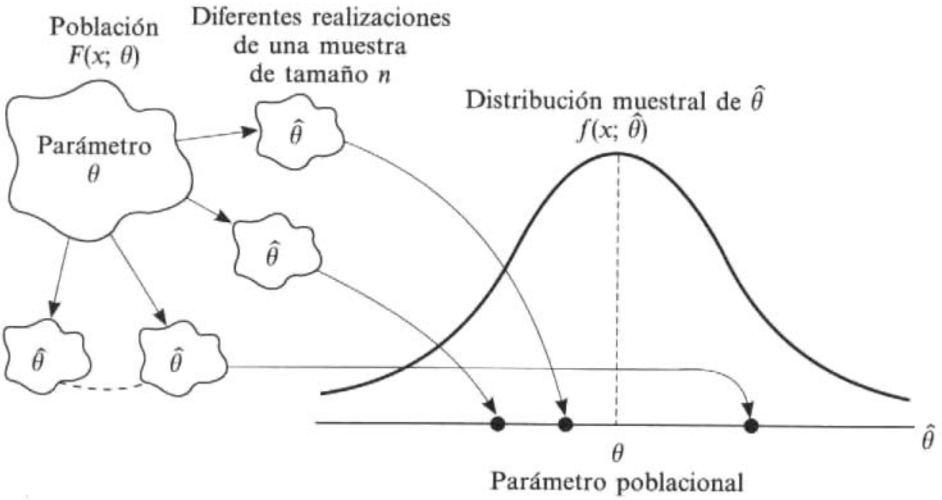


GRÁFICO 2.4. *Representación gráfica de la distribución muestral del estimador $\hat{\theta}$.*

Para seleccionar el estadístico que utilizaremos como estimador del parámetro poblacional tendremos en cuenta las propiedades de la distribución muestral del estadístico. Generalmente nosotros trataremos de obtener un estimador cuyos valores para diferentes realizaciones de una muestra, estén concentrados alrededor del verdadero valor del parámetro θ. Así, por ejemplo, supongamos que consideramos dos estadísticos muestrales, $\hat{\theta}_1$ y $\hat{\theta}_2$, cuyas distribuciones muestrales aparecen en el gráfico 2.5, como estimadores del parámetro θ.

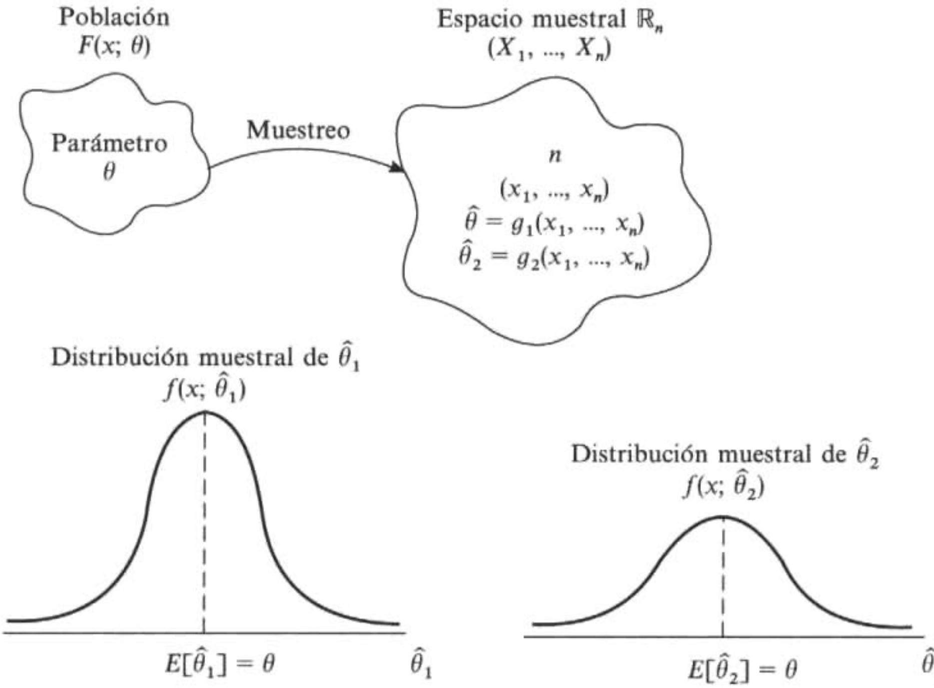


GRÁFICO 2.5. *Distribución muestral de los estadísticos $\hat{\theta}_1$ y $\hat{\theta}_2$.*

Evidentemente seleccionaremos el estadístico $\hat{\theta}_1$ como estimador del parámetro θ, pues los valores del estadístico $\hat{\theta}_1$ para las diferentes realizaciones están más próximas al parámetro θ, que los del estadístico $\hat{\theta}_2$, pues el estadístico $\hat{\theta}_1$, presenta menor varianza que el estadístico $\hat{\theta}_2$ como se observa en el gráfico 2.5.

Para clarificar la diferencia entre **estimador** y **estimación** consideremos el siguiente **ejemplo:** supongamos que pretendemos estimar la renta media μ de todas las familias de una ciudad, para ello parece lógico utilizar como estimador de la media poblacional μ la media muestral $\bar{X}$ siendo necesario seleccionar una muestra aleatoria que supondremos de tamaño $n = 80$, a partir de la cual obtendríamos la renta media de la muestra, por ejemplo, $\bar{x} = 114.380$ ptas. Entonces el **estimador** de la media poblacional μ será, $\hat{\mu} = \bar{X}$, es decir, el estadístico media muestral $\bar{X}$ y la estimación puntual será $\hat{\mu} = \bar{x} = 114.380$ ptas. Observemos que designamos por $\bar{X}$ la variable aleatoria media muestral de las variables aleatorias muestrales $X_1, X_2, ..., X_n$ y por $\bar{x}$ designamos una realización para una muestra específica $(x_1, x_2, ..., x_n)$, que nos da la correspondiente estimación puntual del parámetro μ, es decir, $\hat{\mu} = \bar{x}$.

En la Tabla 2.1 expresamos diferentes parámetros poblacionales, sus estimadores y sus estimaciones.

TABLA 2.1. *Parámetros poblacionales, estimadores y estimaciones.*

Parámetro poblacional	Estimador	Estimación
Media μ	$\hat{\mu} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
Varianza σ^2	$\hat{\sigma}^2 = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Proporción p	$\hat{p} = p_X = \dfrac{X}{n} = \dfrac{\text{Número de éxitos}}{\text{Número de pruebas}}$	$\hat{p} = p_x = \dfrac{x}{n}$

Ejemplo 2.1

Las ventas de una muestra aleatoria de diez grandes establecimientos comerciales de España, el día 5 de enero de 1996, fueron respectivamente: 16, 10, 8, 12, 4, 6, 5, 4, 10, 5 millones de pesetas, respectivamente. Obtener estimaciones puntuales de la venta media, de la varianza de las ventas de todos los establecimientos comerciales y de la proporción de éstos cuyas ventas fueron superiores a 5 millones de pesetas.

Solución:

Las expresiones de las tres estimaciones puntuales que nos piden, aparecen en la última columna de la Tabla 2.1. Así pues la estimación puntual de la media poblacional es la media muestral $\bar{x}$, dada por:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{80}{10} = 8$$

La estimación puntual de la varianza poblacional es la varianza muestral s^2, la cual se obtiene utilizando un desarrollo análogo al de la expresión [1.10]:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

$$= \frac{1}{9}(782 - 10 \cdot 8^2) = 15{,}8$$

Por último, el estimador de la proporción poblacional es la proporción muestral. Para calcular esta proporción muestral necesitamos saber el número de establecimientos comerciales con ventas superiores a 5 millones, que en este caso son 6. De aquí que la estimación puntual de la proporción poblacional es:

$$\hat{p} = p_x = \frac{x}{n} = \frac{6}{10} = 0{,}6$$

Para la elección de estos estimadores puntuales nos hemos basado, principalmente en la intuición y en la posible analogía de los parámetros poblacionales con sus correspondientes valores muestrales, pero éste *no será el método más* adecuado para la obtención de estimadores puntuales, aunque en este caso se obtienen estimadores satisfactorios para los parámetros poblacionales. En general, el problema de obtener estimadores puntuales no será tan sencillo, por ello tendremos que dar propiedades que serían deseables que se cumplieran por los diferentes estimadores puntuales obtenidos. Pero no existe un mecanismo o método único que nos permita obtener el mejor estimador puntual en todas las circunstancias.

Nuestro objetivo ahora será doble:

En primer lugar, daremos algún criterio y propiedades deseables de los estimadores puntuales, con el fin de poder conocer la bondad de los mismos, pues cuantas más propiedades verifiquen los estimadores puntuales mejores serán.

En segundo lugar, daremos varios métodos de obtención de estimadores puntuales.

2.3. PROPIEDADES DE LOS ESTIMADORES PUNTUALES

Sea una población con función de distribución $F(x;\, \theta)$, en donde θ es un parámetro poblacional desconocido, que pretendemos estimar con la ayuda de la muestra aleatoria simple de tamaño n, $(X_1, X_2, ..., X_n)$, a partir del estimador

$$\hat{\theta} = g(X_1, X_2, ..., X_n)$$

que como sabemos es un estadístico y, por tanto, una variable aleatoria que tendrá su correspondiente distribución muestral, su media y su varianza. Pero nos interesa encontrar un estadístico $g(X_1, ..., X_n)$ que nos proporcione el mejor estimador del parámetro desconocido θ, para lo cual tendremos que utilizar alguna medida que nos permita dar algún criterio para seleccionar el mejor estimador. Esta medida será el **error cuadrático medio del estimador.**

Definición 2.1. Error cuadrático medio del estimador $\hat{\theta}$.

Definimos el **error cuadrático medio** del estimador $\hat{\theta}$, que lo notaremos por $\text{ECM}(\hat{\theta})$, como el valor esperado del cuadrado de la diferencia entre el estimador $\hat{\theta}$ y el parámetro θ, es decir

$$\text{ECM}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E[\hat{\theta} - \theta]^2 \qquad [2.1]$$

Desarrollando la expresión [2.1] tendremos:

$$\begin{aligned} \text{ECM}(\hat{\theta}) &= E[\hat{\theta} - \theta]^2 = E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] \\ &= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 = (\text{sumando y restando } (E[\hat{\theta}])^2) \\ &= E[\hat{\theta}^2] - (E[\hat{\theta}])^2 + (E[\hat{\theta}])^2 - 2\theta E[\hat{\theta}] + \theta^2 \\ &= \text{Var}(\hat{\theta}) + (E[\hat{\theta}] - \theta)^2 = \text{Var}(\hat{\theta}) + (\text{sesgo}(\hat{\theta}))^2 \qquad [2.2] \end{aligned}$$

resultando que el ECM del estimador $\hat{\theta}$ se puede descomponer en suma de dos cantidades no negativas:

- La varianza del estimador:

$$\text{Var}(\hat{\theta}) = E[\hat{\theta}^2] - (E[\hat{\theta}])^2 \qquad [2.3]$$

- El cuadrado del sesgo del estimador:

$$(\text{Sesgo}(\hat{\theta}))^2 = (E[\hat{\theta}] - \theta)^2$$

Evidentemente, ambas cantidades deberán de ser tenidas en cuenta para las propiedades deseables de un estimador. Así pues, ambos sumandos, varianza y sesgo, deben de ser lo más pequeños posibles, lo cual equivale a que la distribución muestral del estimador $\hat{\theta}$ debe de concentrarse en torno al valor del parámetro θ, tanto más cuanto menor sea la varianza.

El problema aparentemente parece muy sencillo, pues bastaría seleccionar como mejor estimador del parámetro θ, aquel estimador $\hat{\theta}$ que tenga el error cuadrático medio, ECM, más pequeño de entre todos los posibles estimadores de θ. Pero no es nada fácil el obtener entre todos los posibles estimadores del parámetro θ el que nos dé un error cuadrático medio mínimo para todos los valores posibles del parámetro θ, es decir, no siempre existirá un estimador $\hat{\theta}$ que haga mínimo su error cuadrático medio para todos los valores posibles de θ, pues un estimador $\hat{\theta}$ puede dar lugar a un ECM mínimo para algunos valores del parámetro θ, mientras que otro estimador $\hat{\theta}'$ también dará lugar a un ECM mínimo pero para otros valores diferentes de θ.

Resulta, por tanto, que la utilización del error cuadrático medio para la elección de un buen estimador es insuficiente, siendo necesario dar otros criterios, de tal manera que la elección de un buen estimador puntual dependerá de otras propiedades que satisfaga ese estimador.

Ejemplo 2.2

Sea X_1, X_2, X_3 una muestra aleatoria simple de tamaño 3, cuyos valores son siempre positivos y procedentes de una población con media μ y varianza $\sigma^2 = 25$. Consideramos como posibles estimadores de μ los estadísticos

$$\hat{\mu}_1 = \frac{1}{4}(X_1 + 2X_2 + X_3)$$

$$\hat{\mu}_2 = \frac{1}{5}(X_1 + 2X_2 + X_3)$$

Obtener los errores cuadráticos medios de $\hat{\mu}_1$ y $\hat{\mu}_2$ y comparar sus valores para diferentes valores del parámetro poblacional μ.

Solución:

Empezamos calculando la media y varianza de $\hat{\mu}_1$:

$$\begin{aligned} E[\hat{\mu}_1] &= E\left[\frac{1}{4}(X_1 + 2X_2 + X_3)\right] = \\ &= \frac{1}{4}(E[X_1] + 2E[X_2] + E[X_3]) = \\ &= \frac{1}{4}(\mu + 2\mu + \mu) = \mu \end{aligned}$$

luego

$$\text{Sesgo}(\hat{\mu}_1) = E[\hat{\mu}_1] - \mu = \mu - \mu = 0$$

$$\begin{aligned} \text{Var}(\hat{\mu}_1) &= \text{Var}\left[\frac{1}{4}(X_1 + 2X_2 + X_3)\right] = \\ &= \frac{1}{16}(\text{Var}(X_1) + 4\,\text{Var}(X_2) + \text{Var}(X_3)) = \\ &= \frac{1}{16}(\sigma^2 + 4\sigma^2 + \sigma^2) = \\ &= \frac{6\sigma^2}{16} = \frac{3\sigma^2}{8} = \frac{3 \cdot 25}{8} = \frac{75}{8} \end{aligned}$$

teniendo en cuenta la expresión [2.2], tendremos:

$$\begin{aligned}\mathrm{ECM}(\hat{\mu}_1) &= \mathrm{Var}(\hat{\mu}_1) + (\text{sesgo}\,(\hat{\mu}_1))^2 \\ &= \frac{75}{8} + 0 = \frac{75}{8}\end{aligned}$$

Análogamente para el estimador $\hat{\mu}_2$:

$$\begin{aligned}E[\hat{\mu}_2] &= E\left[\frac{1}{5}(X_1 + 2X_2 + X_3)\right] \\ &= \frac{1}{5}(E[X_1] + 2E[X_2] + E[X_3]) \\ &= \frac{1}{5}(\mu + 2\mu + \mu) = \frac{4\mu}{5}\end{aligned}$$

luego

$$\begin{aligned}\mathrm{Sesgo}(\hat{\mu}_2) &= E[\hat{\mu}_2] - \mu = \frac{4\mu}{5} - \mu \\ &= \frac{4\mu - 5\mu}{5} = -\frac{1}{5}\mu \\ \mathrm{Var}(\hat{\mu}_2) &= \mathrm{Var}\left[\frac{1}{5}(X_1 + 2X_2 + X_3)\right] \\ &= \frac{1}{25}(\mathrm{Var}(X_1) + 4\,\mathrm{Var}(X_2) + \mathrm{Var}(X_3)) \\ &= \frac{1}{25}(\sigma^2 + 4\sigma^2 + \sigma^2) \\ &= \frac{6\sigma^2}{25} = \frac{6 \cdot 25}{25} = 6\end{aligned}$$

y su error cuadrático medio será:

$$\begin{aligned}\mathrm{ECM}(\hat{\mu}_2) &= \mathrm{Var}(\hat{\mu}_2) + (\text{sesgo}\,(\hat{\mu}_2))^2 \\ &= 6 + \frac{\mu^2}{25}\end{aligned}$$

Igualando $\mathrm{ECM}(\hat{\mu}_1) = \mathrm{ECM}(\hat{\mu}_2)$ tendremos:

$$\begin{aligned}\frac{75}{8} &= 6 + \frac{\mu^2}{25} \\ \frac{\mu^2}{25} &= \frac{75}{8} - 6\end{aligned}$$

$$\frac{\mu^2}{25} = \frac{27}{8} \quad ; \quad 8\mu^2 = 675$$

$$\mu = \sqrt{\frac{675}{8}}$$

luego si

$$\mu < \sqrt{\frac{675}{8}} \quad \Rightarrow \quad \text{ECM}(\hat{\mu}_2) < \text{ECM}(\hat{\mu}_1)$$

y el estimador $\hat{\mu}_2$ será mejor que el estimador $\hat{\mu}_1$, pero si

$$\mu > \sqrt{\frac{675}{8}} \quad \Rightarrow \quad \text{ECM}(\hat{\mu}_1) < \text{ECM}(\hat{\mu}_2)$$

resultando que el estimador $\hat{\mu}_1$ será mejor que el estimador $\hat{\mu}_2$. Este resultado confirma lo indicado anteriormente, siendo por ello necesario dar otros criterios o propiedades adicionales para la selección de un buen estimador puntual. Así pues estudiaremos la insesgadez, eficiencia, consistencia y suficiencia que darán lugar a los estimadores puntuales: **insesgados, eficientes, consistentes** y **suficientes.**

2.3.1. ESTIMADOR INSESGADO

Hemos definido el sesgo del estimador $\hat{\theta}$ como:

$$\text{Sesgo}(\hat{\theta}) = E[\hat{\theta}] - \theta \qquad [2.4]$$

Veíamos anteriormente, en la expresión [2.2] del ECM, que en el segundo sumando nos aparecía el cuadrado del sesgo, también decíamos que el ECM$(\hat{\theta})$ debería ser lo más pequeño posible y para ello era necesario que la varianza del estimador y el cuadrado del sesgo también fueran lo más pequeños posibles. Es decir será conveniente que el sesgo en valor absoluto sea lo menor posible, siendo deseable que sea nulo y en tal caso la media del estimador $\hat{\theta}$ coincidirá con el valor del parámetro θ que se está estimando, es decir

$$E[\hat{\theta}] = \theta$$

siendo entonces el estimador $\hat{\theta}$, un estimador **insesgado** del parámetro θ y la distribución muestral del estimador $\hat{\theta}$ se encontrará centrada alrededor del parámetro θ.

Definición 2.2. Estimador insesgado.

Diremos que el estadístico $\hat{\theta} = g(X_1, ..., X_n)$ es un estimador **insesgado** o **centrado** del parámetro θ si la esperanza matemática del estimador $\hat{\theta}$ es igual al parámetro θ, esto es:

$$E[\hat{\theta}] = \theta \qquad [2.5]$$

para todos los valores de θ, y entonces:

$$\text{Sesgo}(\hat{\theta}) = E[\hat{\theta}] - \theta = 0$$

En caso contrario diremos que el estimador es **sesgado** o **descentrado**, es decir

$$E[\hat{\theta}] = \theta + b(\hat{\theta}) = \theta + \text{sesgo}(\hat{\theta}) \qquad [2.6]$$

en donde $b(\hat{\theta}) = E[\hat{\theta}] - \theta = \text{sesgo}(\hat{\theta})$

El sesgo del estimador, sesgo $(\hat{\theta})$, puede ser positivo, negativo e incluso nulo, así pues si es positivo entonces se dice que el estimador sobreestima el valor del parámetro desconocido y si es negativo lo infraestima, siendo por tanto, deseable que sea nulo para que sea insesgado.

El gráfico 2.6 muestra la representación gráfica de las distribuciones muestrales de dos estimadores del parámetro, uno sesgado $\hat{\theta}_1$ y otro insesgado $\hat{\theta}_2$.

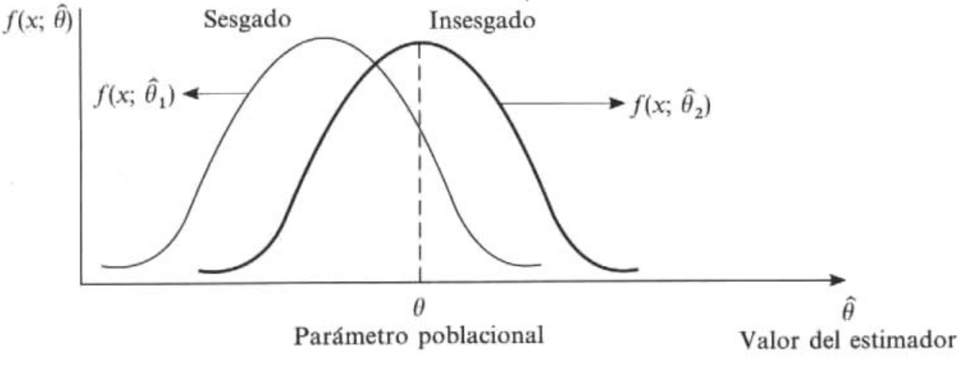


GRÁFICO 2.6. *Representación gráfica de las funciones de densidad $f(x; \hat{\theta}_1)$ y $f(x; \hat{\theta}_2)$ de dos estimadores, uno sesgado $\hat{\theta}_1$ y otro insesgado $\hat{\theta}_2$.*

Algunos estimadores para parámetros poblacionales se obtienen intuitivamente por analogía. Por ejemplo, parece lógico utilizar el estadístico media muestral $\bar{X}$, como estimador del parámetro media poblacional μ, análogamente la proporción muestral $\hat{p} = \hat{p}_X = \frac{X}{n}$ como estimador de la proporción poblacional p y la varianza muestral, S^2, como estimador de la varianza poblacional σ^2. La misma intuición nos vale para seleccionar un estimador puntual de la diferencia de dos parámetros poblacionales. Así pues, el estimador puntual de la diferencia de los parámetros medias poblacionales $\mu_X - \mu_Y$, será la diferencia de medias muestrales $\bar{X}_1 - \bar{X}_2$ y análogamente el estimador de la diferencia entre proporciones poblacionales $(p_X - p_Y)$ será la diferencia entre las proporciones muestrales $(\hat{p}_X - \hat{p}_Y)$.

Estos cinco estadísticos o estimadores $\bar{X}$, $\hat{p}$, S^2, $\bar{X} - \bar{X}$ y $\hat{p}_X - \hat{p}_Y$ son funciones de las observaciones muestrales $X_1, X_2, \ldots, X_n$, cuyos respectivos valores esperados y varianzas aparecen en la Tabla 2.2.

TABLA 2.2. *Algunos parámetros poblacionales, sus estimadores puntuales insesgados, media y varianza*[3].

Parámetro poblacional θ	**Estimador puntual insesgado** $\hat{\theta}$	**Valor esperado de** $\hat{\theta}$	**Varianza de** $\hat{\theta}$
Media μ	$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	μ	$\frac{\sigma^2}{n}$
Varianza[4] σ^2	$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	σ^2	$\frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\sigma^4$
Proporción p	$\hat{p} = p_X = \frac{X}{n}$	p	$\frac{pq}{n}$
Diferencia de medias $\mu_X - \mu_Y$	$\hat{\mu}_X - \hat{\mu}_Y = \bar{X} - \bar{X}$	$\mu_X - \mu_Y$	$\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}$
Diferencia de proporción $p_X - p_Y$	$\hat{p}_X - \hat{p}_Y = \frac{X}{n_X} - \frac{Y}{n_Y}$	$p_1 - p_2$	$\frac{p_x q_x}{n_x} + \frac{p_y q_y}{n_y}$

[3] n_X y n_Y son los tamaños muestrales, σ_X^2 y σ_Y^2 las varianzas poblacionales, X indica el número total de éxitos en n pruebas, y análogamente Y.

[4] En una distribución $N(\mu, \sigma)$, sabemos que $\mu_4 = 3\sigma^4$, luego

$$\mathrm{Var}(S^2) = \frac{3\sigma^4}{n} + \frac{3-n}{n(n-1)}\sigma^4 = \frac{1}{n(n-1)}(3n\sigma^4 - 3\sigma^4 + 3\sigma^4 - n\sigma^4) = \frac{2\sigma^4}{n-1}$$

Los cinco estimadores puntuales que aparecen en la Tabla 2.2 son insesgados, pues teniendo en cuenta lo estudiado en el capítulo anterior, se comprueba fácilmente que:

$$E[\hat{\mu}] = E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$$

$$E[\hat{\sigma}^2] = E[S^2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2\right] = \sigma^2$$

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = p$$

$$E[\hat{\mu}_X - \hat{\mu}_Y] = E[\bar{X} - \bar{X}] = \mu_X - \mu_Y$$

$$E[\hat{p}_X - \hat{p}_Y] = E\left[\frac{X}{n_X} - \frac{Y}{n_Y}\right] = p_X - p_Y$$

Ejemplo 2.3

Dado el estadístico

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

demostrar que es un estimador sesgado de la varianza poblacional[5].

Solución:

Podemos expresar S'^2 de otra forma para tomar valores esperados con más facilidad, en efecto:

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_2) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu + \mu - \bar{X})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} [(X_i - \mu) - (\bar{X} - \mu)]^2$$

[5] Nosotros definimos, como la mayoría de los autores, la **varianza muestral** como:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

aunque algunos autores le llaman a esta expresión **cuasivarianza** muestral y la representan por S_c^2. Estos autores cuando el denominador es n entonces utilizan el término varianza muestral.

$$= \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)]$$
$$= \frac{1}{n} \left[\sum_{i=1}^{n} (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) \right]$$
$$= \frac{1}{n} \left[\sum_{i=1}^{n} (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)(n\bar{X} - n\mu) \right.$$
$$= \frac{1}{n} \left[\sum_{i=1}^{n} (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2n(\bar{X} - \mu)^2 \right]$$
$$= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - (\bar{X} - \mu)^2$$

Tomando valores esperado resulta:

$$E[S'^2] = E\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - (\bar{X} - \mu)^2 \right]$$
$$= \frac{1}{n} \sum_{i=1}^{n} E[(X_i - \mu)^2] - E[(\bar{X} - \mu)^2]$$
$$= \frac{1}{n} \cdot n\sigma^2 - \frac{\sigma^2}{n}$$
$$= \frac{n-1}{n} \sigma^2$$

Luego vemos que efectivamente se trata de un estimador sesgado, pues

$$E[S'^2] \neq \sigma^2$$

La varianza muestral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

sí que es un estimador insesgado de la varianza poblacional, pues fácilmente podemos comprobar que:

$$E[S^2] = E\left[\frac{nS'^2}{n-1} \right] = \frac{n}{n-1} E[S'^2]$$
$$= \frac{n}{n-1} \frac{n-1}{n} \sigma^2$$
$$= \sigma^2$$

Sin embargo no es cierto que la desviación típica muestral sea un estimador insesgado de la desviación típica de la población, es decir

$$E[S] \neq \sigma$$

ya que la raíz cuadrada de una suma de números no es igual a la suma de las raíces cuadradas de los mismos números. Para probar esto utilizaremos el ejemplo 2.4.

Si consideramos el estadístico

$$S^{*2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

fácilmente se comprueba que también es un estimador insesgado de la varianza poblacional, en efecto:

$$E[S^{*2}] = \frac{1}{n} E\left[\sum_{i=1}^{n} (X_i - \mu)^2\right] = \frac{1}{n} \sum_{i=1}^{n} E[(X_i - \mu)^2] = \frac{1}{n} n\sigma^2 = \sigma^2$$

Ejemplo 2.4

Sea una población formada por los elementos (1, 2, 3). Obtener $E[S^2]$ y $E[S]$.

Solución:

Para calcular la media de la varianza muestral, $E[S^2]$, y la media de la desviación estándar muestral, $E[S]$, construimos la Tabla 2.3 con todas las posibles muestras con reemplazamiento de tamaño $n = 2$.

TABLA 2.3. *Muestras de tamaño n = 2, sus medias y varianzas.*

Muestras (x_1, x_2)	$\bar{x}$	$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2$	$S^2 = \frac{1}{2-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$	$S = \sqrt{S^2}$
(1, 1)	1	0,00	0,00	0,00
(1, 2)	1,5	0,50	0,50	0,71
(1, 3)	2	2,00	2,00	1,41
(2, 1)	1,5	0,50	0,50	0,71
(2, 2)	2	0,00	0,00	0,00
(2, 3)	2,5	0,50	0,50	0,71
(3, 1)	2	2,00	2,00	1,41
(3, 2)	2,5	0,50	0,50	0,71
(3, 3)	3	0,00	0,00	0,00
Total	18	6,00	6,00	5,66

La media de la varianza muestral será:

$$E[S^2] = \frac{6}{9} = \frac{2}{3}$$

La varianza de la población $\sigma^2 = \frac{2}{3}$, y la desviación estándar, $\sigma = 0{,}8164$, como se puede comprobar.

Luego

$$E[S^2] = \frac{2}{3} = \sigma^2$$

es decir, como ya sabíamos, la varianza muestral es un estimador insesgado de la varianza poblacional.

La última columna de la Tabla 2.3, nos da las desviaciones típica de cada muestra y su media será:

$$E[S] = \frac{5{,}66}{9} = 0{,}6288$$

resultando que:

$$E[S] = 0{,}6288 \neq 0{,}8164 = \sigma$$

de donde se deduce que, efectivamente, la desviación típica muestral no es un estimador insesgado de la desviación estándar poblacional.

Existen factores de corrección para la desviación estándar muestral S dependientes de la forma de la distribución poblacional, que la convierten en un estimador insesgado de σ.

Proposición 2.1.

> Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores insesgados del parámetro θ, entonces el estimador $\hat{\theta}$ definido como
>
> $$\hat{\theta} = \lambda\hat{\theta}_1 + (1 - \lambda)\hat{\theta}_2 \quad , \quad \lambda \in (0, 1) \qquad [2.7]$$
>
> es también un estimador insesgado del parámetro θ.

Demostración:

Bastará tomar valores esperados en la expresión [2.7].

$$\begin{aligned} E[\hat{\theta}] &= E[\lambda\hat{\theta}_1 + (1 - \lambda)\hat{\theta}_2] = \lambda E[\hat{\theta}_1] + (1 - \lambda)E[\hat{\theta}_2] \\ &= \lambda\theta + (1 - \lambda)\theta = \theta \end{aligned}$$

Esto nos permite decir que si tenemos dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ insesgados del parámetro θ, entonces cualquier combinación lineal convexa de ambos estimadores insesgados será también un estimador insesgado del parámetro θ, y además existirán infinitos de estos estimadores, pues podemos obtener infinitas combinaciones lineales convexas de ambos estimadores.

2.3.2. ESTIMADOR INSESGADO DE VARIANZA MÍNIMA

Ya indicábamos anteriormente que no era posible obtener un estimador $\hat{\theta}$ que haga mínimo su error cuadrático medio para todos los valores posibles del parámetro θ. Sin embargo, sí podemos considerar los estimadores que son insesgados y de éstos determinar el que tenga su error cuadrático medio, ECM$(\hat{\theta})$, mínimo. Es decir, si el estimador $\hat{\theta}$ es insesgado, entonces:

$$E[\hat{\theta}] = \theta \quad \text{y} \quad \mathrm{ECM}[\hat{\theta}] = \mathrm{Var}(\hat{\theta})$$

por tanto, debemos de intentar obtener un estimador, si es que existe, de entre todos los estimadores insesgados que tenga varianza mínima y éste sería el **estimador insesgado de varianza mínima.** Si además se verifica que la varianza es mínima para todos los valores posibles del parámetro entonces el estimador recibe el nombre de **estimador insesgado y uniformemente de mínima varianza** (UMVUE)[6].

Definición 2.3. Estimador insesgado uniformemente de mínima varianza.

> Diremos que el estimador insesgado $\hat{\theta}_0$, es **insesgado y uniformemente de mínima varianza** (UMVUE) para el parámetro θ, si dado cualquier otro estimador insesgado $\hat{\theta}$, de él y, se verifica que
>
> $$\mathrm{Var}(\hat{\theta}_0) \leqslant \mathrm{Var}(\hat{\theta})$$
>
> para todos los valores posibles de θ.

Para llegar a obtener el estimador insesgado uniformemente de mínima varianza, si es que éste existe, tendríamos que determinar las varianzas de todos los estimadores insesgados de θ y seleccionar el estimador que tenga la varianza más pequeña para todos los valores de θ.

Con el fin de facilitar la obtención de un estimador insesgado y uniformemente de mínima varianza (UMVUE) daremos la **desigualdad o cota de Frechet-Cramer-Rao,** la cual nos permitirá obtener una cota inferior de la varianza.

[6] Uniformly minimum-variance unbiased estimators.

2.3.2.1. Cota de Frechet-Cramer-Rao

Sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria simple de tamaño n, obtenida de una población cuya función de densidad o de cuantía es $f(x; \theta)$. Designamos la función de densidad conjunta de la muestra[7] por:

$$L(x_1, ..., x_n; \theta) = d\,F_n(x_1, ..., x_n; \theta) = f_n(x_1, x_2, ..., x_n; \theta)$$

verificándose que

$$\int_{\mathbb{R}^n} d\,F_n(x_1, x_2, ..., x_n; \theta) =$$

$$= \int_{\mathbb{R}^n} f_n(x_1, x_2, ..., x_n; \theta) dx_1 dx_2 ... dx_n = 1$$

y sea $\hat{\theta} = g(X_1, X_2, ..., X_n)$ un estimador insesgado del parámetro θ.

Entones si se verifican las **condiciones de regularidad de Wolfowitz**[8] la varianza del estimador está acotada inferiormente:

$$\operatorname{Var}(\hat{\theta}) \geqslant \frac{1}{E\left[\left(\frac{\partial \ln d\,F_n}{\partial \theta}\right)^2\right]} \qquad [2.8]$$

o bien, si las variables aleatorias son independientes e idénticamente distribuidas con función de densidad o de cuantía $f(x; \theta)$, entonces:

$$\operatorname{Var}(\hat{\theta}) \geqslant \frac{1}{n\,E\left[\left(\frac{\partial \ln f(x;\ \theta)}{\partial \theta}\right)^2\right]} \qquad [2.9]$$

o incluso

$$\operatorname{Var}(\hat{\theta}) \geqslant \frac{1}{-nE\left[\frac{\partial^2 \ln f(x;\ \theta)}{\partial \theta^2}\right]}$$

[7] También se llama **función de verosimilitud** de la muestra y se representa como:

$$L(x_1, ..., x_n; \theta) = dF_n(x_1, ..., x_n; \theta) = f_n(x_1, ..., x_n; \theta)$$

En el caso discreto la función de verosimilitud de la muestra será:

$$L(x_1, ..., x_n; \theta) = P_\theta(X_1 = x_1, ..., X_n = x_n) = P(X_1 = x_1, ..., X_n = x_n; \theta)$$

[8] Existen otras condiciones de regularidad como, por ejemplo, las dadas por Cramer o por Fisz pero son más complicadas.

Las **condiciones de regularidad de Wolfowitz** son:

a) El campo de variación del parámetro θ es un intervalo abierto D del eje real, que puede ser infinito o semi-infinito pero nunca se reduce a un punto.

b) El campo de variación de la variable aleatoria X que define la población no depende del parámetro θ.

c) Para casi todo x y todo $\theta \in D$, existe[9]

$$\frac{\partial \ln d\,F_n(x_1,\ x_2,\ \ldots,\ x_n;\ \theta)}{\partial \theta}$$

d) Se pueden diferenciar, bajo el signo integral, las expresiones $E[1]$ y $E[\hat{\theta}]$[10].

e) Se verifica que

$$E\left[\left(\frac{\partial \ln d\,F_n(x_1,\ x_2,\ \ldots,\ x_n;\ \theta)}{\partial \theta}\right)^2\right] > 0, \quad \text{para} \quad \forall \theta \in D$$

Veamos que en efecto se obtiene la expresión [2.8].

Admitimos que el estimador $\hat{\theta}(x_1, \ldots, x_n)$ es insesgado, y por tanto, se verifica:

$$E[\hat{\theta}] = \int_{\mathbb{R}^n} \hat{\theta}(x_1, \ldots, x_n)\, d\,F_n(x_1,\ x_2,\ \ldots,\ x_n;\ \theta) = \theta$$

de donde, escribiendo en lo sucesivo $\hat{\theta}$ en lugar de $\hat{\theta}(x_1, \ldots, x_n)$ y $d\,F_n$ en lugar de $d\,F_n(x_1, \ldots, x_n)$:

$$\int_{\mathbb{R}^n} (\hat{\theta} - \theta) d\,F_n = 0$$

derivando respecto de θ:

$$0 = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} (\hat{\theta} - \theta) d\,F_n = (\text{por } d) = \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} [(\hat{\theta} - \theta) d\,F_n] =$$

$$= \int_{\mathbb{R}^n} -d\,F_n + \int_{\mathbb{R}^n} (\hat{\theta} - \theta) \frac{\partial}{\partial \theta} d\,F_n$$

[9] Al decir para casi todo x, queremos decir para todo x excepto para un conjunto cuya probabilidad sea nula.

[10] Se pueden intercambiar la operación de derivación respecto de θ y la integración (o suma en el caso discreto) respecto de x.

$$= \int_{\mathbb{R}^n} -d F_n + \int_{\mathbb{R}^n} (\hat{\theta} - \theta) \frac{\frac{\partial}{\partial \theta} d F_n}{d F_n} d F_n$$

$$= -1 + \int_{\mathbb{R}^n} (\hat{\theta} - \theta) \frac{\partial}{\partial \theta} (\ln d F_n) d F_n$$

de donde se tiene:

$$1 = \int_{\mathbb{R}^n} (\hat{\theta} - \theta) \frac{\partial}{\partial \theta} (\ln d F_n) d F_n$$

elevando al cuadrado ambos miembros y teniendo en cuenta la desigualdad de Schwarz[11], tendremos:

$$1 = \left[\int_{\mathbb{R}^n} (\hat{\theta} - \theta) \frac{\partial}{\partial \theta} (\ln d F_n) dF \right]^2 \leqslant \int_{\mathbb{R}^n} (\hat{\theta} - \theta)^2 d F_n \cdot \int_{\mathbb{R}^n} \left(\frac{\partial}{\partial \theta} (\ln d F_n) \right)^2 d F_n$$

y teniendo en cuenta la definición de $\operatorname{Var}(\hat{\theta})$ y la definición de valor esperado, resulta:

$$1 \leqslant \operatorname{Var}(\hat{\theta}) \cdot E\left[\frac{\partial \ln d F_n}{\partial \theta} \right]^2$$

de donde

$$\operatorname{Var}(\hat{\theta}) \geqslant \frac{1}{E\left[\dfrac{\partial \ln d F_n(x_1, x_2, ..., x_n; \theta)}{\partial \theta} \right]^2} \qquad \text{c.q.d.}$$

La otra expresión [2.9] de la cota de Frechet-Cramer-Rao se tiene como consecuencia de que la muestra es aleatoria simple, pues entonces la función de densidad o de cuantía de la muestra es igual al producto de las funciones de densidad marginales:

$$d F_n(x_1, x_2, ..., x_n; \theta) = f_n(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} dF(x_i; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

---

[11] Desigualdad de Schwarz en un intervalo (a, b)

$$\int_a^b [f(x)g(x)dF(x)]^2 \leqslant \int_a^b [f(x)]^2 dF(x) \cdot \int_a^b [g(x)]^2 dF(x)$$

y tomando logaritmos neperianos, resulta:

$$\ln d F_n = \sum_{i=1}^{n} \ln f(x_i; \theta)$$

$$\frac{\partial \ln d F_n}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln f(x_i; \theta)$$

elevando al cuadrado ambos miembros:

$$\left(\frac{\partial \ln d F_n}{\partial \theta}\right)^2 = \left(\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln f(x_i; \theta)\right)^2 =$$

$$= \sum_{i=1}^{n} \left(\frac{\partial}{\partial \theta} \ln f(x_i; \theta)\right)^2 + \sum_{i \neq j} \left(\frac{\partial}{\partial \theta} \ln f(x_i; \theta)\right) \cdot \left(\frac{\partial}{\partial \theta} \ln f(x_j; \theta)\right)$$

Tomando valor esperado:

$$E\left[\left(\frac{\partial \ln d F_n}{\partial \theta}\right)^2\right] = E\left[\sum_{i=1}^{n} \left(\frac{\partial}{\partial \theta} \ln f(x_i; \theta)\right)^2\right] +$$

$$+ E\left[\sum_{i \neq j} \left(\frac{\partial}{\partial \theta} \ln f(x_i; \theta)\right) \cdot \left(\frac{\partial}{\partial \theta} \ln f(x_j; \theta)\right)\right] \quad [12]$$

pero teniendo en cuenta que las variables son independientes:

$$E\left[\sum_{i \neq j} \left(\frac{\partial}{\partial \theta} \ln f(x_i; \theta)\right) \cdot \left(\frac{\partial}{\partial \theta} \ln f(x_j; \theta)\right)\right] =$$

$$= \sum_{i \neq j} E\left[\frac{\partial}{\partial \theta} \ln f(x_i; \theta)\right] \cdot E\left[\frac{\partial}{\partial \theta} \ln f(x_j; \theta)\right] = 0$$

---

[12] Sabemos que $\int_{\mathbb{R}^n} dF_n = 1$, derivando respecto de θ:

$$\int_{\mathbb{R}^n} \frac{\partial \mathrm{d}F_n}{\partial \theta} = 0 \quad ; \quad \int_{\mathbb{R}^n} \frac{\frac{\partial dF_n}{\partial \theta}}{dF_n} \cdot dF_n = \int_{\mathbb{R}^n} \frac{\partial \ln dF_n}{\partial \theta} \cdot dF_n = E\left[\frac{\partial \ln dF_n}{\partial \theta}\right] = 0$$

análogamente, sabemos que $\int_{\mathbb{R}} f(x; \theta) dx = 1$, derivando respecto de θ:

$$\int_{\mathbb{R}} \frac{\partial f(x; \theta)}{\partial \theta} dx = 0 \quad ; \quad \int_{\mathbb{R}} \frac{\frac{\partial f(x; \theta)}{\partial \theta}}{f(x; \theta)} dx \cdot f(x; \theta) dx = \int_{\mathbb{R}} \frac{\partial \ln f(x; \theta)}{\partial \theta} \cdot f(x; \theta) dx = E\left[\frac{\partial \ln f(x; \theta)}{\partial \theta}\right] = 0$$

$$E\left[\left(\frac{\partial \ln d\, F_n}{\partial\theta}\right)^2\right] = \sum_{i=i}^{n} E\left[\left(\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right)^2\right] = nE\left[\left(\frac{\partial \ln f(x;\ \theta)}{\partial\theta}\right)^2\right]$$

Luego la **cota de Frechet-Crawer-Rao** tiene esta forma:

$$\mathrm{Var}(\hat{\theta}) \geqslant \frac{1}{n\,E\left[\left(\frac{\partial \ln f(x;\ \theta)}{\partial\theta}\right)^2\right]}$$

o bien esta otra:

$$\mathrm{Var}(\hat{\theta}) \geqslant \frac{1}{n\,E\left[\left(\frac{\partial \ln dF(x;\ \theta)}{\partial\theta}\right)^2\right]} \qquad [2.10]$$

Fisher llamó al denominador de la cota de F-C-R, **cantidad de información contenida en la muestra de tamaño *n***, es decir, la cantidad de información que la muestra proporciona sobre el parámetro:

$$I = E\left[\left(\frac{\partial \ln d\, F_n}{\partial\theta}\right)^2\right] = n\,E\left[\left(\frac{\partial \ln f(x;\ \theta)}{\partial\theta}\right)^2\right] \qquad [2.11]$$

Si el estimador $\hat{\theta}$ hubiera sido **sesgado**, es decir:

$$E[\hat{\theta}] = \theta + b(\hat{\theta})$$

en donde $b(\hat{\theta})$ es el sesgo del estimador, entonces la cota de Frechet-Cramer-Rao tiene la forma:

$$\mathrm{Var}(\hat{\theta}) \geqslant \frac{[1 + b'(\hat{\theta})]^2}{E\left[\frac{\partial \ln d\, F_n(x_1,\ x_2,\ \ldots,\ x_n;\ \theta)}{\partial\theta}\right]^2} \qquad [2.12]$$

siendo $b'(\hat{\theta})$ la derivada respecto de θ del sesgo del estimador.

En el supuesto de haber considerado una población de tipo discreto, bastaría sustituir la función de densidad por la correspondiente función de cuantía, obteniendo resultados análogos.

La cota o desigualdad de Frechet-Cramer-Rao nos da un límite inferior para la varianza del estimador $\hat{\theta}$, pero esto no implica que la varianza de un estimador UMVUE tenga que ser igual al límite inferior de la varianza dado por la cota de F-C-R. Es decir, se puede obtener un estimador insesgado $\hat{\theta}$ que tenga su varianza más pequeña que la de todos los demás estimadores insesgados de θ, pero mayor que el límite inferior dado por la cota de F-C-R. Un estimador que verifique lo anterior seguirá siendo un estimador UMVUE del parámetro θ.

2.3.3. ESTIMADOR EFICIENTE

Observando la definición de estimador insesgado se pone de manifiesto que la insesgadez presenta una debilidad, pues únicamente requiere que el valor esperado del estimador $\hat{\theta}$ sea igual al parámetro poblacional θ, y no requiere que muchos, o incluso algunos, de los valores del estimador (es decir, estimaciones) tomen valores próximos al parámetro poblacional, como sería deseable en un buen estimador. Por eso, la propiedad de eficiencia es importante, ya que exige algún requisito más. Así pues, cuando se quiere estimar un parámetro poblacional considerando diferentes muestras de tamaño n, es deseable que el estimador tome, para las diferentes muestras, valores próximos unos de otros, de tal manera que resulte una varianza pequeña para los diferentes valores del estimador, pues cuanto menor sea la varianza mejor será el estimador, es decir, la propiedad de eficiencia implicara que la varianza del estimador sea pequeña. Sin embargo, el hecho de que la varianza del estimador sea pequeña, por sí solo no es suficiente para tener un buen estimador, sino que el estimador tendría que ser también insesgado. Por ejemplo, si para diferentes muestras aleatorias el estimador siempre toma un mismo valor especificado 150, entonces la varianza del estimador será cero, pero el estimador será sesgado excepto que el verdadero valor del parámetro poblacional sea también 150. Luego será deseable que la varianza sea mínima y que el estimador sea insesgado.

La propiedad de eficiencia de un estimador la definiremos comparando su varianza con la varianza de los demás estimadores insesgados. Así pues:

> «el **estimador más eficiente** entre un grupo de estimadores insesgados será el que tenga menor varianza».

Supongamos que tenemos una población con función de densidad $f(x; \theta)$, en donde θ es el parámetro desconocido y consideramos tres estimadores $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_3$ del parámetro θ, basados en muestras aleatorias del mismo tamaño, siendo las distribuciones de los tres estimadores las que aparecen en el gráfico 2.7, en donde se observa que las distribuciones correspondientes a $\hat{\theta}_1$ y $\hat{\theta}_3$ tienen como media el parámetro poblacional θ, es decir, ambos estimadores $\hat{\theta}_1$ y $\hat{\theta}_3$ son insesgados, sin embargo, la distribución correspondiente a $\hat{\theta}_2$ es sesgada, tiene un sesgo positivo pues su media es mayor que el parámetro poblacional. En cuanto a la varianza de los tres estimadores se observa que la más pequeña es la correspondiente a $\hat{\theta}_2$ y sin embargo este estimador no es más eficiente ya que no es insesgado.

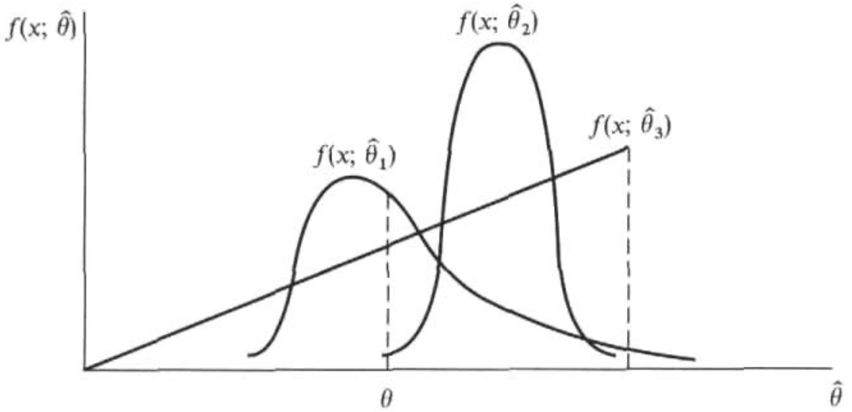


GRÁFICO 2.7. *Representación gráfica de las funciones de densidad $f(x;\, \hat\theta_1)$, $f(x;\, \hat\theta_2)$ y $f(x;\, \hat\theta_3)$ de tres estimadores $\hat\theta_1$, $\hat\theta_2$ y $\hat\theta_3$.*

Luego, para que un estimador sea el más eficiente será necesario que sea insesgado y que tenga menor varianza que cualquier otro estimador insesgado, así pues, del gráfico 2.7 se deduce que el estimador $\hat\theta_1$ es el más eficiente de los tres, pues es insesgado y tiene menor varianza que el estimador $\hat\theta_3$.

Anteriormente ya indicábamos la importancia que tenía la varianza de un estimador y aquí se pone de manifiesto otra vez que la varianza de un estimador insesgado es una medida muy importante para decidir sobre si es o no apto para estimar un parámetro θ.

Definición 2.4. Estimador eficiente.

Diremos que un estimador $\hat\theta$ del parámetro poblacional θ, es **eficiente** si es insesgado y además su varianza alcanza la cota de Frechet-Cramer-Rao. Esto es equivalente a decir que un estimador $\hat\theta$ es **eficiente** si su varianza coincide con la cota de Frechet-Cramer-Rao[13]:

$$\operatorname{Var}(\hat\theta) = \frac{1}{E\left[\left(\dfrac{\partial \ln d F_n}{\partial \theta}\right)^2\right]} \qquad [2.13]$$

o bien

$$\operatorname{Var}(\hat\theta) = \frac{1}{n\,E\left[\left(\dfrac{\partial \ln f(x;\ \theta)}{\partial \theta}\right)^2\right]}$$

[13] Pues esta cota se obtiene cuando el estimador es insesgado, y posible nos da el valor mínimo de la varianza.

Luego un estimador eficiente será un estimador insesgado y uniformemente de mínima de varianza (UMVUE), cuya varianza coincide con el límite inferior de la cota de Frechet-Cramer-Rao; pero un estimador UMVUE puede que no sea eficiente puesto que su varianza, siendo mínima, no alcance la cota de F.C.R.

Este tipo de estimadores serán de bastante utilidad en toda la inferencia estadística, siendo por ello el que se intentara obtener, siempre que exista.

Definición 2.5. Eficiencia de un estimador.

Se define la **eficiencia de un estimador insesgado,** $\hat{\theta}$, del parámetro θ como:

$$\text{eff. } (\hat{\theta}) = \frac{\text{Cota F.C.R.}}{\text{Var}(\hat{\theta})} \qquad [2.14]$$

verificándose que eff. $(\hat{\theta}) \leqslant 1$.

De aquí que si tenemos dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$ del parámetro θ, diremos que el estimador $\hat{\theta}_1$ es **más eficiente** que el estimador $\hat{\theta}_2$, si se verifica:

$$\text{eff. } (\hat{\theta}_1) \geqslant \text{eff. } (\hat{\theta}_2) \qquad [2.15]$$

es decir, si se verifica que:

$$\text{Var}(\hat{\theta}_1) \leqslant \text{Var}(\hat{\theta}_2)$$

en donde la desigualdad en sentido estricto se debe cumplir para algún valor de θ.

Pero en general nos podemos encontrar con varios estimadores insesgados, no siendo nada fácil el probar que uno de esos estimadores insesgados es el mejor de todos ellos. Para resolver esta situación de manera fácil lo que se hace es introducir el concepto e **eficiencia relativa** de dos estimadores.

Definición 2.6. Eficiencia relativa.

Dados dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$ del parámetro θ, definimos la **eficiencia relativa de $\hat{\theta}_1$ a $\hat{\theta}_2$** como:

$$\text{eff. relativa } (\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} = \frac{\text{eff. } (\hat{\theta}_1)}{\text{eff. } (\hat{\theta}_2)} \qquad [2.16]$$

Si este cociente es *menor*, *igual* o *mayor* que la unidad, diremos que $\hat{\theta}_1$ es *menos*, *igual* o *más* eficiente que $\hat{\theta}_2$[14].

[14] Algunos autores utilizan, para la eficiencia relativa, la notación eff. $(\hat{\theta}_1/\hat{\theta}_2)$.

El gráfico 2.8 representa las distribuciones muestrales de dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$. Observando la representación gráfica se deduce que la $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$, luego la eficiencia relativa de $\hat{\theta}_1$ a $\hat{\theta}_2$ será:

$$\text{eff. relativa } (\theta_1, \theta_2) = \frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)} > 1$$

y diremos que el estimador $\hat{\theta}_1$ es **más eficiente** que $\hat{\theta}_2$.

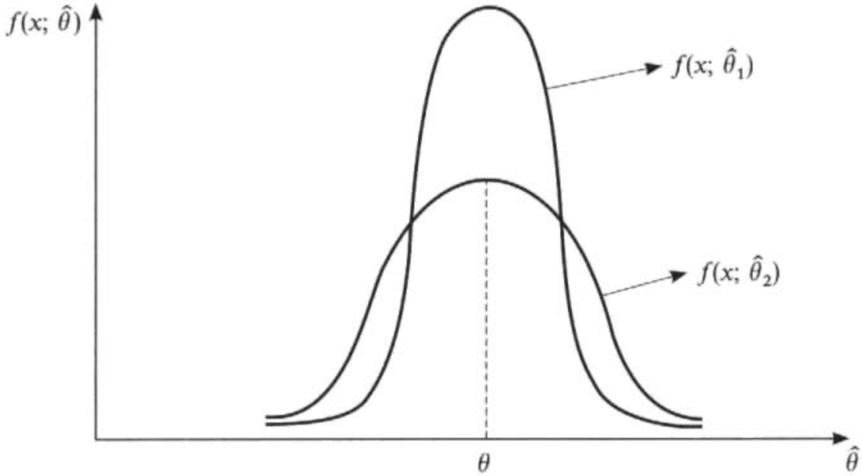


GRÁFICO 2.8. *Representación gráfica de las funciones de densidad de dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$, donde $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$.*

Ejemplo 2.5

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple procedente de una población $N(\mu, \sigma)$. Utilizando la media muestral $\bar{X}$ y la mediana muestral X_m como estimadores de la media poblacional μ. Estudiar su eficiencia relativa.

Solución:

Sabemos que los estadísticos media muestral y mediana muestral son estimadores insesgados de la media poblacional, pues la población de partida es normal y, por tanto, simétrica, coincidiendo la media, la mediana y la moda[15].

[15] Se demuestra que la mediana muestral X_m tiende a distribuirse según una distribución normal de media μ y varianza $\frac{\pi}{2}\frac{\sigma^2}{n}$, es decir, $X_m \to N\left(\mu, \sqrt{\frac{\pi}{2}\cdot\frac{\sigma^2}{n}}\right)$.

La varianza de ambos estimadores es:

$$\operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\operatorname{Var}(X_m) = \frac{\pi}{2} \cdot \frac{\sigma^2}{n} \simeq 1{,}57 \frac{\sigma^2}{n}$$

La eficiencia relativa del estimador mediana muestral X_m al estimador media muestral $\bar{X}$ será:

$$\text{eff. relativa } (X_m, \bar{X}) = \frac{\operatorname{Var}(\bar{X})}{\operatorname{Var}(X_m)} = \frac{\sigma^2/n}{1{,}57\sigma^2/n} = \frac{1}{1{,}57}$$

de donde

$$\operatorname{Var}(X_m) = 1{,}57 \operatorname{Var}(\bar{X}) = 1{,}57 \frac{\sigma^2}{n}$$

Lo cual implica que la mediana muestral X_m es 1,57 veces **menos eficiente** que la media muestral $\bar{X}$ para estimar la media de la población μ. Es decir un estimador basado en la mediana de una muestra de 157 observaciones tiene la misma varianza que un estimador basado en la media de una muestra de 100 observaciones, admitiendo que la población es normal. También podríamos decir que la varianza de la mediana muestral es superior a la varianza de la media muestral en un 57 % de ésta. Así pues, para que ambos estimadores tuvieran la misma varianza sería necesario utilizar un 57 % más de observaciones en la mediana muestral que en la media[16].

Podemos concluir que la media muestral es un estimador más eficiente de la media poblacional μ, que la mediana muestral. El gráfico 2.9 ilustra la relación entre ambas distribuciones muestrales e indica que la distribución muestral de la media tiene menor varianza que la distribución muestral de la mediana.

---

[16] Cuando estudiábamos las medidas de posición central decíamos que una ventaja de la mediana sobre la media es que daba bastante menos importancia a los valores u observaciones extremas que la media, sin embargo ahora, en la eficiencia relativa, vemos que la mediana presenta el inconveniente de necesitar mayor número de observaciones que la media.

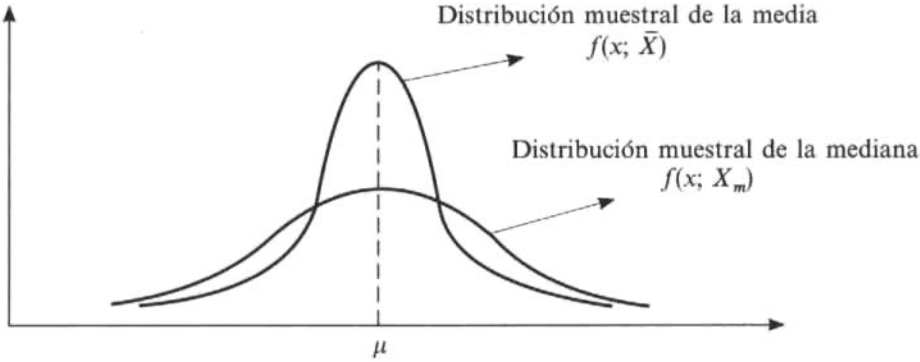


GRÁFICO 2.9. *Representación gráfica de las distribuciones muestrales de los estimadores media $\bar{X}$ y mediana X_m del parámetro μ, media poblacional.*

Proposición 2.2

Dada una población $N(\mu, \sigma)$ se verifica que la media muestral $\bar{X}$ es un estimador eficiente de la media poblacional μ.

Demostración:

Sabemos que la función de densidad de una distribución $N(\mu, \sigma)$, de parámetro μ, desconocido, es:

$$f(x; \mu) = \frac{1}{\sigma\sqrt{2n}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Para que el estadístico, $\bar{X}$, media muestral sea un estimador eficiente del parámetro μ, media poblacional, se tiene que verificar la expresión [2.13], es decir, que su varianza coincida con la cota de Frechet-Cramer-Rao:

$$\operatorname{Var}(\bar{X}) = \frac{1}{n\,E\left[\left(\frac{\partial \ln f(x;\ \mu)}{\partial\mu}\right)^2\right]}$$

En efecto:

$$\ln f(x; \mu) = \ln \frac{1}{\sigma\sqrt{2\pi}} + \ln e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

$$= \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$$\frac{\partial \ln f(x;\ \mu)}{\partial \mu} = \frac{x - \mu}{\sigma^2}$$

$$n\,E\left[\left(\frac{\partial \ln f(x;\ \mu)}{\partial \mu}\right)^2\right] = n\,E\left[\left(\frac{X - \mu}{\sigma^2}\right)^2\right] = \frac{n}{\sigma^4}\,E[(X - \mu)^2]$$

$$= \frac{n}{\sigma^4}\,\mathrm{Var}(X) = \frac{n}{\sigma^4}\cdot\sigma^2$$

$$= \frac{n}{\sigma^2}$$

Luego sustituyendo en la expresión [2.13], resultaría:

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}(\bar{X}) = \frac{1}{n\,E\left[\left(\dfrac{\partial \ln f(x;\theta)}{\partial \theta}\right)^2\right]} = \frac{\sigma^2}{n}$$

que coincide con la cota de Frechet-Cramer-Rao, además sabemos que la varianza del estadístico media muestral es $\frac{\sigma^2}{n}$ y que el estadístico media muestral es un estimador insesgado de la media población μ.

Resultando que, efectivamente, la media muestral es un estimador eficiente de la media poblacional, cuando la población es $N(\mu, \sigma)$.

Ejemplo 2.6

Dada una población $N(\mu, 7)$, y los estimadores de la media poblacional μ, para muestras aleatorias simples de tamaño $n = 3$

$$\hat{\theta}_1 = \frac{1}{3}(X_1 + X_2 + X_3)$$

$$\hat{\theta}_2 = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{4}X_3$$

Se pide:

1. Comprobar que los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ son o no insesgados.
2. Calcular la varianza de ambos estimadores.
3. ¿Son ambos estimadores eficientes?

Solución:

1. Sabemos que un estimador $\hat{\theta}$ es insesgado si se verifica que:

$$E[\hat{\theta}] = \theta$$

En este ejemplo, se conoce que:

$$E[X_i] = \mu \quad , \quad i = 1, 2, 3$$

Luego

$$\begin{aligned} E[\hat{\theta}_1] &= E\left[\frac{1}{3}(X_1 + X_2 + X_3)\right] \\ &= \frac{1}{3}(E[X_1] + E[X_2] + E[X_3]) \\ &= \frac{1}{3}(\mu + \mu + \mu) = \mu \end{aligned}$$

y por tanto el estimador $\hat{\theta}$, es un estimador lineal insesgado para μ.

Para el estimador $\hat{\theta}_2$ se tiene:

$$\begin{aligned} E(\hat{\theta}_2] &= E\left[\frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{4}X_3\right] \\ &= \frac{1}{2}E[X_1] + \frac{1}{3}E[X_2] + \frac{1}{4}E[X_3] \\ &= \frac{1}{2}\mu + \frac{1}{3}\mu + \frac{1}{4}\mu \\ &= \frac{26}{24}\mu = \frac{13}{12}\mu \end{aligned}$$

Luego este estimador lineal es sesgado para μ.

2. Veamos la varianza de ambos estimadores:

Se sabe que:

$$\operatorname{Var}(X_i) = 49 \quad , \quad i = 1, 2, 3$$

Luego:

$$\operatorname{Var}(\hat{\theta}_1) = \operatorname{Var}\left(\frac{1}{3}(X_1 + X_2 + X_3)\right)$$

$$= \frac{1}{9}(\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3))$$

$$= \frac{1}{9}(3 \cdot 49) = \frac{49}{3}$$

$$\text{Var}(\hat{\theta}_2) = \text{Var}\left(\frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{4}X_3\right)$$

$$= \frac{1}{4}\text{Var}(X_1) + \frac{1}{9}\text{Var}(X_2) + \frac{1}{16}\text{Var}(X_3)$$

$$= \frac{1}{4} \cdot 49 + \frac{1}{9} \cdot 49 + \frac{1}{16} \cdot 49$$

$$= \frac{244}{576} \cdot 49 = \frac{61}{144} \cdot 49 = \frac{2.989}{144}$$

3. Para ver si son eficientes tendremos que tener en cuenta la definición 2.4, es decir tendrán que ser insesgados y su varianza alcance la cota de F.C.R.

Ahora bien, en nuestro caso el estimador $\hat{\theta}_2$ no es insesgado y por tanto no será eficiente.

Para el estimador $\hat{\theta}_1$, que sí que es insesgado, bastará tener en cuenta la proposición 2.2, pues resulta que el estimador

$$\hat{\theta}_1 = \frac{1}{3}(X_1 + X_2 + X_3)$$

coincide exactamente con la media muestral $\bar{X}$, y según hemos visto el estadístico media muestral, $\bar{X}$, en una población $N(\mu, \sigma)$ es un estimador eficiente de la media poblacional μ.

Luego el estimador $\hat{\theta}_1$, es un estimador eficiente de la media poblacional μ.

Teorema 2.1

Si un estimador $\hat{\theta}$ es insesgado, su varianza alcanza la cota de F.C.R. si se verifica:

$$\frac{\partial \ln d F_n}{\partial \theta} = A(\theta)(\hat{\theta} - \theta)$$

Siendo $A(\theta)$ una expresión que no depende de θ y entonces el estimador $\hat{\theta}$ será eficiente.

Teorema 2.2

Si $\hat{\theta}$ es un estimador eficiente, entonces se verifica que

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{A(\theta)}$$

Demostración:

Como el estimador $\hat{\theta}$ es eficiente, entonces la $\operatorname{Var}(\hat{\theta})$ coincide con la cota de Frechet-Cramer-Rao, expresión [2.13] y teniendo en cuenta el teorema 2.1 resulta:

$$\begin{aligned}\operatorname{Var}(\hat{\theta}) &= \frac{1}{E\left[\dfrac{\partial \ln d F_n}{\partial \theta}\right]^2} = \frac{1}{E[A(\theta)(\hat{\theta} - \theta)]^2} \\ &= \frac{1}{A^2(\theta)E[\hat{\theta} - \theta]^2} \\ &= \frac{1}{A^2(\theta)\operatorname{Var}(\hat{\theta})}\end{aligned}$$

de donde se deduce que

$$(\operatorname{Var}(\hat{\theta}))^2 = \frac{1}{A^2(\theta)}$$

Luego

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{A(\theta)}$$

Definición 2.7. Estimador asintóticamente eficiente.

Diremos que un estimador $\hat{\theta}$ es **asintóticamente eficiente** si se verifica:

$$\lim_{n \to \infty} \operatorname{Var}(\hat{\theta}) = \text{Cota de Frechet-Cramer-Rao}^{17} \qquad [2.17]$$

[17] No obstante debemos tener en cuenta que la cota también depende del tamaño muestral, lo cual puede ocasionar algún problema en algún caso aislado (como podrían ser el caso de los estimadores súper-eficientes).

2.3.4. ESTIMADOR CONSISTENTE

Hasta ahora hemos considerado propiedades de los estimadores puntuales basados en muestras aleatorias de tamaño n, pero parece lógico esperar que un estimador será tanto mejor cuanto mayor sea el tamaño de la muestra. Así pues cuando el tamaño de la muestra aumenta y por tanto la información que nos proporciona esa muestra es más completa, resulta que la varianza del estimador suele ser menor y la distribución muestral de ese estimador tenderá a encontrarse más concentrada alrededor del parámetro que pretendemos estimar. Además teniendo en cuenta el teorema de Glivenko-Cantelli, resulta que cuando el tamaño de la muestra es suficientemente grande entonces la muestra puede llegar a proporcionar una información casi exacta de la población y en consecuencia el valor del estimador tiende a coincidir con el valor del parámetro. Por esto, en este apartado nos vamos a referir a las propiedades asintóticas de los estimadores, es decir a su comportamiento cuando el tamaño de la muestra se hace muy grande ($n \to \infty$). La más importante de estas propiedades asintóticas es la **consistencia.**

Sean $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n$ una sucesión de estimadores del parámetro θ, obtenidos a partir de muestras de tamaño 1, 2, ..., n, respectivamente, es decir:

$$\begin{aligned} \hat{\theta}_1 &= g(X_1) \\ \hat{\theta}_2 &= g(X_1, X_2) \\ &\vdots \\ \hat{\theta}_n &= g(X_1, X_2, ..., X_n) \end{aligned}$$

de manera que el estimador basado en la muestra de tamaño n lo notaremos por $\hat{\theta}_n$, en donde el subíndice n lo empleamos para hacer más evidente la dependencia del tamaño muestral. En general esta sucesión de estimadores se representa por $\{\hat{\theta}_n\}$.

Definición 2.8. Estimador consistente.

Diremos que una sucesión de estimadores $\{\hat{\theta}_n\}$ es **consistente,** si la sucesión converge en probabilidad hacia el parámetro θ. Es decir, si $\forall \varepsilon > 0$, se verifica:

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \quad , \quad \forall \theta \qquad [2.18]$$

y cada elemento de la sucesión se dirá que es un **estimador consistente.**

Esta consistencia que hemos definido es una **consistencia simple** o **consistencia en probabilidad** ya que se basa en la convergencia en probabilidad, por eso también se suele decir que una sucesión de estimadores $\{\hat{\theta}_n\}$ es consistente si converge en probabilidad hacia el valor del parámetro θ, a medida que el tamaño de la muestra aumenta. Lo cual implica que la distribución del estimador consistente estará más concentrada entorno al valor del parámetro θ y, por tanto, la varianza del estimador consistente debe disminuir cuando n aumenta, tendiendo a cero cuando $n \to \infty$. Situación que representamos en el gráfico 2.10.

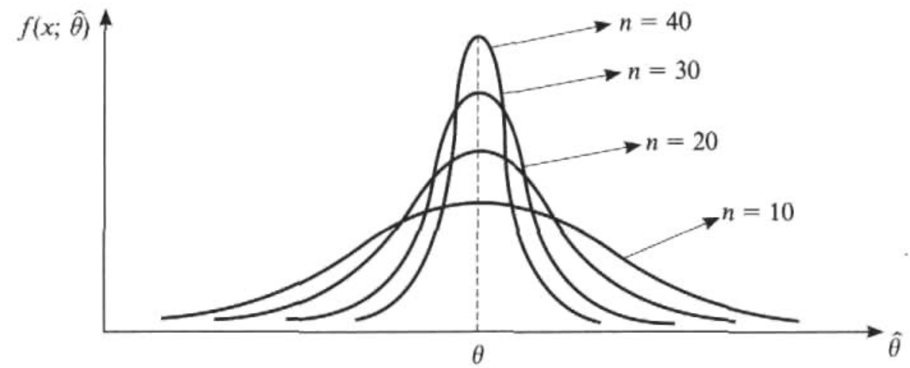


GRÁFICO 2.10. *Representación gráfica ilustrativa de la consistencia de un estimador $\hat{\theta}_n$. A medida que crece el tamaño de la muestra la distribución del estimador está más concentrada alrededor del valor del parámetro θ.*

También se puede definir un estimador consistente basándose en la convergencia en media cuadrática.

Definición 2.9. Consistencia en media cuadrática.

Diremos que una sucesión de estimadores $\{\hat{\theta}_n\}$ es **consistente en media cuadrática** para el parámetro θ cuando se verifica:

$$\lim_{n \to \infty} E[\hat{\theta}_n - \theta]^2 = 0 \qquad [2.19]$$

y cada elemento de la sucesión se dirá que es un **estimador consistente en media cuadrática.**

Análogamente a la expresión [2.2], aquí tenemos que el **error cuadrático medio** del estimador $\hat{\theta}_n$ será:

$$\mathrm{ECM}(\hat{\theta}_n) = E[(\hat{\theta}_n - \theta)^2] = \mathrm{Var}(\hat{\theta}_n) + (\text{sesgo}\ (\hat{\theta}_n))^2 \qquad [2.20]$$

que tenderá a cero si ambos sumandos tienden a cero, pues ambos sumandos son no negativos.

Luego para ver si un estimador es consistente en media cuadrática bastará con demostrar que la varianza y el sesgo del estimador tienden a cero cuando $n \to \infty$.

Teorema 2.3

> Si un estimador es consistente en media cuadrática también es consistente en probabilidad, pero no necesariamente se verifica al revés.

Demostración:

Para demostrar este teorema tendremos que demostrar que si:

$$E[\hat{\theta}_n - \theta]^2 \xrightarrow[n \to \infty]{} 0, \quad \forall \theta$$

entonces

$$P(|\hat{\theta}_n - \theta| < \varepsilon) \xrightarrow[n \to \infty]{} 1, \quad \forall \varepsilon > 0 \quad \text{y} \quad \forall \theta$$

En efecto, si tenemos en cuenta la desigualdad de Chebychev[18] escrita en la forma:

$$P[|\hat{\theta}_n - \theta| < \varepsilon] = P[|\hat{\theta}_n - \theta|^2 < \varepsilon^2] \geqslant 1 - \frac{E[\hat{\theta}_n - \theta]^2}{\varepsilon^2} \quad , \quad \varepsilon > 0 \quad [2.21]$$

y cuando $n \to \infty$, según la hipótesis de partida, el estimador era consistente en media cuadrática:

$$E[\hat{\theta}_n - \theta]^2 \to 0$$

[18] Ver CASAS Y SANTOS (1995). Capítulo 10.

y sustituyendo en la expresión [2.21], se tiene:

$$P[|\hat{\theta}_n - \theta| < \varepsilon] \to 1$$

Luego el estimador es consistente en probabilidad[19].

Definición 2.10. Consistencia casi segura.

Diremos que una sucesión de estimadores $\{\hat{\theta}_n\}$ es **consistente casi seguro** para θ cuando se verifica:

$$P\left(\lim_{n\to\infty} \hat{\theta}_n = \theta\right) = 1 \qquad [2.22]$$

y cada elemento de la sucesión se dirá que es un **estimador consistente casi seguro.**

En consecuencia, si el estimador es consistente casi seguro también lo será en probabilidad.

Ejemplo 2.7

Sea $(X_1, ..., X_n)$ una muestra aleatoria de tamaño n procedente de una población $N(\mu, \sigma)$. Demostrar que la media muestral, $\bar{X}$, y la varianza muestral, S^2, son estimadores consistentes de la media y varianza poblacional, respectivamente.

Solución:

En efecto, la media muestral $\bar{X}$ es un estimador consistente de la media poblacional μ, pues es un estimador insesgado

$$E[\bar{X}] = \mu$$

siendo el sesgo nulo para cualquier tamaño de muestra, y además la varianza del estimador media muestral, $\bar{X}$, es:

$$\operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$$

[19] También podíamos haber dicho que la demostración es inmediata ya que como sabemos la convergencia en media cuadrática implica la convergencia en probabilidad.

Luego como el sesgo la varianza del estimador, $\bar{X}$, tienden a cero cuando $n \to \infty$, resulta que se trata de un estimador consistente en media cuadrática y por tanto también en probabilidad.

Otra forma de hacerlo sería:

Sabemos que

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Tipificando tenemos:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \to N(0, 1)$$

y teniendo en cuenta la expresión [2.18] asociada a la definición de estimador consistente, tendremos:

$$P(|\bar{X} - \mu| < \varepsilon) = P\left(\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu| < \frac{\sqrt{n}}{\sigma}\varepsilon\right)$$

$$= P\left(-\frac{\sqrt{n}}{\sigma}\varepsilon < \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) < \frac{\sqrt{n}}{\sigma}\varepsilon\right) \xrightarrow[n \to \infty]{} 1$$

Lo cual demuestra que efectivamente $\bar{X}$ es un estimador consistente del parámetro media poblacional μ.

El estimador varianza muestral, S^2, es un estimador insesgado pues:

$$E[S^2] = \sigma^2$$

siendo nulo el sesgo para cualquier tamaño muestral.

La varianza del estimador varianza muestral cuando la muestra procede de una población $N(\mu, \sigma)$, según se vio en el Teorema 1.2, viene dada por:

$$\mathrm{Var}(S^2) = \frac{2\sigma^4}{n - 1} \xrightarrow[n \to \infty]{} 0$$

Luego como el sesgo y la varianza del estimador, S^2, tiende a cero cuando $n \to \infty$, resulta que el estimador S^2 es un estimador consistente de la varianza poblacional σ^2.

Definición 2.10. Estimador óptimo asintóticamente normal.

Diremos que una sucesión de estimadores $\{\hat{\theta}_n\}$ del parámetro θ da lugar a un **estimador óptimo asintóticamente normal,** $\hat{\theta}$, del parámetro θ, si se verifican las siguientes condiciones:

1. La distribución del estadístico

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\mathrm{Var}(\theta_n)}}\sqrt{n} \to N(0, 1)$$

tiende a una distribución $N(0, 1)$, cuando $n \to \infty$.

2. La sucesión de estimadores es consistente, es decir, $\forall \varepsilon > 0$, se verifica:

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1, \quad \forall\theta$$

lo cual equivale a decir que el estimador $\hat{\theta}$ es consistente para todos los valores de θ.

3. No existe ninguna otra sucesión de estimadores $\{\hat{\theta}'_n\}$ que verifique las dos condiciones anteriores y que además:

$$\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}'_n) \leqslant \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n), \quad \forall\theta$$

2.3.5. SUFICIENCIA

Hasta ahora, y como indicábamos al final del apartado 2.2, la elección de los estimadores la hacemos basándonos en la intuición y en la analogía de los parámetros poblacionales con sus correspondientes valores muestrales. También, en algunas ocasiones nos interesa que el estimador tenga alguna propiedad concreta, por ejemplo, que sea insesgado, o que cumpla cualquier otra propiedad. Pero como el estimador era simplemente un estadístico y por tanto función de las observaciones muestrales, resulta que utilizamos las observaciones muestrales para obtener los estimadores de los parámetros poblacionales, de tal manera que se resume la información que proporciona la muestra sobre los parámetros poblacionales en los valores (o estimaciones) que toman sus estimadores, pudiendo producirse una posible pérdida de la información que contiene la muestra cuado se sustituyen las observaciones individuales por el valor del estadístico. Así pues, supongamos que queremos estimar los parámetros media, μ, y varianza, σ^2, poblacional con la ayuda de una muestra aleatoria, utilizando para ello los estimadores insesgados media muestral, $\bar{X}$, y varianza muestral, S^2. Las estimaciones correspondientes, de los parámetros poblacionales serán los valores que toman los estimadores $\bar{X}$ y S^2 para las n observaciones de la muestra aleatoria, resultando que la información de las n observaciones muestrales se resume o se reduce a los dos valores de los estimadores $\bar{X}$ y S^2. En consecuencia, nos surge la pregunta: ¿en este proceso de resumen o reducción de la información (pues pasamos a tener sólo los valores de $\bar{X}$ y S^2), que nos proporcionan las n-observaciones muestrales sobre los parámetros poblacionales μ y σ, se mantiene o se ha perdido información respecto a los parámetros μ y σ?

En este apartado daremos algunos métodos para obtener estadísticos o estimadores tales que utilicen toda la información contenida en la muestra con respecto al parámetro poblacional a estimar. Tales estadísticos o estimadores los llamaremos **suficientes,** pues contienen toda la información relevante contenida en la muestra con respecto al parámetro que nos interesa[20].

2.3.5.1. Estimador suficiente

De manera intuitiva, diremos que un estadístico es **suficiente** para un parámetro θ cuando utiliza toda la información relevante contenida en la muestra aleatoria, con respecto a θ, y ningún otro estadístico puede proporcionar más información adicional sobre el parámetro poblacional θ. Es decir, mediante un estadístico suficiente tenemos una manera de resumir toda la información contenida en la muestra acerca del parámetro θ. Por ejemplo, consideremos una muestra de n repeticiones independientes de un experimento binomial, $(X_1, ..., X_n)$, con probabilidad de éxito p, y definimos el estadístico T como el número de éxitos en las n repeticiones, es decir

$$T = \sum_{i=1}^{n} X_i$$

en donde

$$X_i = \begin{cases} 1, & \text{si la } i\text{-ésima repetición es éxito, con probabilidad } p \\ 0, & \text{si la } i\text{-ésima repetición es fracaso con probabilidad } 1 - p \end{cases}$$

Como estamos interesados en el parámetro poblacional p, al tomar la muestra de n-repeticiones del experimento binomial tendremos un valor del estadístico:

$$T = \sum_{i=1}^{n} X_i = \text{número de éxitos en las } n\text{-pruebas}$$

[20] Sabemos que el estimador es una función de las observaciones muestrales y, por tanto, será un estadístico, de aquí que algunos autores utilizan de manera indiferente los términos estimador y estadístico.

y entonces nos surge la duda de si este estadístico contiene toda la información sobre el parámetro p o por el contrario se podría obtener más información sobre p considerando otros estadísticos o funciones de $(X_1, ..., X_n)$.

Para resolver esta duda obtenemos la distribución condicionada de $X_1, ..., X_n$ dado el valor del estadístico $T = t$, es decir:

$$P(X_1 = x_1, ..., X_n = x_n/T = t) = \frac{P(X_1 = x_1, ..., X_n = x_n, T = t)}{P(T = t)}$$

$$= \frac{P(X_1 = x_1, ..., X_n = x_n) \cdot P(T = t/X_1 = x_1, ..., X_n = x_n)}{P(T = t)}$$

$$= \frac{p^{x_1} \cdot (1-p)^{1-x} ... p^{x_n} \cdot (1-p)^{1-x_n} \cdot P(T = t/X_1 = x_1, ..., X_n = x_n)}{\binom{n}{t} p^t \cdot (1-p)^{n-t}}$$

$$= \frac{p^t \cdot (1-p)^{n-t} \cdot P(T = t/X_1 = x_1, ..., X_n = x_n)}{\binom{n}{t} p^t \cdot (1-p)^{n-t}}$$

$$= \begin{cases} \dfrac{1}{\binom{n}{t}} & \text{si} \quad \sum_{i=1}^{n} x_i = t \\ 0 & \text{si} \quad \sum_{i=1}^{n} x_i \neq t \end{cases} \qquad [2.22]$$

Observamos que la distribución condicionada de $X_1, ..., X_n$ dado el valor del estadístico $T = t$ no depende del parámetro p, es decir, la distribución condicionada para la muestra de n repeticiones, dado el número de éxitos, no depende de la probabilidad p de obtener un éxito, entonces conociendo el número total de éxitos en la muestra tendremos toda la información que la muestra puede proporcionar sobre el valor del parámetro p, siendo por tanto, el estadístico T suficiente para el parámetro p[21].

[21] En este ejemplo también son suficientes para el parámetro p los estadísticos $\left(X_1, \sum_{i=2}^{n} X_i\right)$, $\left(X_1, X_2, \sum_{i=3}^{n} X_i\right)$, ...

Definición 2.12. Estimador suficiente.

> Sea $(X_1, ..., X_n)$ una muestra aleatoria de una población cuya distribución depende de un parámetro θ desconocido. Diremos que el estadístico o estimador $T = T(X_1, ..., X_n)$ es **suficiente** para el parámetro θ si la distribución condicionada de $X_1, ..., X_n$ dado el valor del estadístico $T = t$, no depende del parámetro θ.

Ejemplo 2.8

Sea una muestra aleatoria (X_1, X_2, X_3) procedente de una distribución $B(1, p)$, y sean los estadísticos:

$$T_1 = X_1 + X_2 + X_3$$

$$T_2 = X_1 + 2X_2 + X_3$$

tales que para la muestra de tamaño $n = 3$ toman los valores $T_1 = 2$ y $T_2 = 2$. Comprobar que T_1 es suficiente y que T_2 no es suficiente.

Solución:

El estadístico

$$T_1 = X_1 + X_2 + X_3$$

es suficiente, pues es un caso particular del ejemplo anterior, así pues, sustituyendo en la expresión [2.23] tenemos:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3/T = 2) = \frac{1}{\binom{3}{2}}$$

y esta probabilidad no depende del parámetro p, con lo cual es el estadístico T_1 es suficiente.

Análogamente, para el estadístico:

$$T_2 = X_1 + 2X_2 + X_3$$

si obtenemos la probabilidad condicionada, por ejemplo, para la muestra $(x_1, x_2, x_3) \equiv (1, 0, 1)$ tendremos que:

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, X_3 = x_3/T_2 = 2) = \\ &= P(X_1 = 1, X_2 = 0, X_3 = 1/X_1 + 2X_2 + X_3 = 2) \\ &= \frac{P(X_1 = 1,\ X_2 = 0,\ X_3 = 1)}{P(X_1 + 2X_2 + X_3 = 2)} \end{aligned}$$

$$= \frac{p(1-p)^0 \cdot p^0(1-p) \cdot p(1-p)^0}{P(X_1 = 1,\ X_2 = 0,\ X_3 = 1) + P(X_1 = 0,\ X_2 = 1,\ X_3 = 0)}$$

$$= \frac{p^2(1-p)}{p^2(1-p) + p(1-p)^2} = \frac{p}{p + (1-p)} = p$$

la cual depende del parámetro p, y por tanto, el estadístico $T_2 = X_1 + 2X_2 + X_3$ no es suficiente.

Si observamos el estadístico:

$$T_1 = X_1 + X_2 + X_3$$

toma los valores 0, 1, 2, 3 sin pérdida de ninguna información sobre el parámetro p. Sin embargo, el estadístico

$$T_2 = X_1 + 2X_2 + X_3$$

toma los valores 0, 1, 2, 3, 4 perdiendo información sobre el parámetro p.

Esta definición de estadístico suficiente nos permite comprobar si efectivamente el estadístico o estimador T es o no suficiente pero no nos dice cómo se puede encontrar un estadístico o estimador suficiente.

Un método que, además de decirnos si un estadístico es o no suficiente, nos permite también obtener un estadístico suficiente, es el teorema de factorización de Fisher-Neyman.

2.3.5.2. Teorema de Factorización de Fisher-Neyman

Sea $(X_1, ..., X_n)$ una muestra aleatoria simple de una población con función de distribución $F(x; \theta)$ y sea la función de cuantía de la muestra:

$$P(x_1, ..., x_n; \theta) = P_\theta(X_1 = x_1, ..., X_n = x_n)$$

o la función de densidad de la muestra:

$$f(x_1, ..., x_n; \theta)$$

Entonces el estadístico $T = T(X_1, ..., X_n)$ es *suficiente* para el parámetro θ si y solamente si podemos descomponer la función de probabilidad o la función de densidad de la muestra en productos de dos factores no negativos:

$$P_\theta(x_1, ..., x_n) = g(T(x_1, ..., x_n); \theta) \cdot h(x_1, ..., x_n) \qquad [2.24]$$

o

$$f(x_1, ..., x_n; \theta) = g(T(x_1, ..., x_n); \theta) \cdot h(x_1, ..., x_n) \qquad [2.25]$$

en donde $g(T, \theta)$ es una función que depende solamente de θ y de la muestra a través del estadístico $T(X_1, ..., X_n)$, y $h(x_1, ..., x_n)$ no depende de θ.

Demostración:

Vamos a realizar la demostración para el caso discreto[22]:

Si admitimos que T es un estadístico suficiente para θ, entonces la distribución condicionada:

$$P_\theta(X_1=x_1, ..., X_n=x_n/T(x_1, ..., x_n=t))=\frac{P_\theta(X_1=x_1, ..., X_n=x_n, T(x_1, ..., x_n)=t)}{P_\theta(T(x_1, ..., x_n)=t)}$$

es independiente del parámetro θ, y podemos escribir

$$\begin{aligned} P_\theta(X_1 = x_1, ..., X_n = x_n) &= P_\theta(X_1 = x_1, ..., X_n = x_n, T(x_1, ..., x_n) = t) \\ &= P_\theta(T(x_1, ..., x_n) = t) \cdot P(X_1 = x_1, ..., X_n = x_n/T(x_1, ..., x_n) = t) \end{aligned}$$

siempre y cuando la probabilidad condicionada

$$P(X_1 = x_1, ..., X_n = x_n/T(x_1, ..., x_n) = t)$$

esté bien definida.

Y haciendo:

$$h(x_1, ..., x_n) = P(X_1 = x_1, ..., X_n = x_n/T(x_1, ..., x_n) = t)$$

que como vemos no depende de θ, y

$$g(T(x_1, ..., x_n); \theta) = P_\theta(T(x_1, ..., x_n) = t)$$

se verifica el teorema, pues:

$$P_\theta(X = x_1, ..., X_n = x_n) = g(T(x_1, ..., x_n); \theta) \cdot h(x_1, ..., x_n)$$

Veamos ahora la situación inversa, es decir, si se verifica el criterio de factorización entonces el estadístico T será suficiente. En efecto:

$$P_\theta(X_1 = x_1, ..., X_n = x_n/T(x_1, ..., x_n) = t) =$$

$$= \begin{cases} 0 & \text{si} \quad T(x_1, ..., x_n) \neq t \\ \dfrac{P_\theta(X_1 = x_1, ..., X_n = x_n, T(x_1, ..., x_n) = t)}{P_\theta(T(x_1, ..., x_n) = t)}, & \text{si} \quad T(x_1, ..., x_n) = t \end{cases}$$

[22] Para una demostración general ver LEHMANN (1986).

$$= \begin{cases} 0 & \text{si} \quad T(x_1, ..., x_n) \neq t \\ \dfrac{P_\theta(X_1 = x_1, ..., X_n = x_n)}{P_\theta(T(x_1, ..., x_n) = t)}, & \text{si} \quad T(x_1, ..., x_n) = t \end{cases}$$

Evidentemente si

$$T(x_1, ..., x_n) \neq t$$

entonces la probabilidad condicionada

$$P_\theta(X_1 = x_1, ..., X_n = x_n / T(x_1, ..., x_n) = t) = 0$$

no depende del parámetro θ.

Si $T(x_1, ..., x_n) = t$ entonces teniendo en cuenta que se verifica el criterio de factorización podemos escribir:

$$\begin{aligned} P_\theta(T(x_1, ..., x_n) = t) &= \sum_{T(x_1, ..., x_n) = t} P_\theta(X_1 = x_1, ..., X_n = x_n) \\ &= \sum_{T(x_1, ..., x_n) = t} g(T(x_1, ..., x_n); \theta) \cdot h(x_1, ..., x_n) \\ &= g(t, \theta) \cdot \sum_{T(x_1, ..., x_n) = t} h(x_1, ..., x_n) \end{aligned}$$

Luego

$$\begin{aligned} P_\theta(X_1 = x_1, ..., X_n = x_n / T(x_1, ..., x_n) = t) &= \frac{P_\theta(X_1 = x_1, ..., X_n = x_n)}{P_\theta(T(x_1, ..., x_n) = t)} \\ &= \frac{g(t; \theta) \cdot h(x_1, ..., x_n)}{g(t; \theta) \displaystyle\sum_{T(x_1, ..., x_n) = t} h(x_1, ..., x_n)} \\ &= \frac{h(x_1, ..., x_n)}{\displaystyle\sum_{T(x_1, ..., x_n) = t} h(x_1, ..., x_n)} \end{aligned}$$

que no depende de θ, y por definición se deduce que el estadístico $T = T(X_1, ..., X_n)$ es suficiente, como queríamos demostrar.

Ejemplo 2.9

Sea una muestra aleatoria $(X_1, ..., X_n)$ de una distribución $B(1; p)$. Comprobar utilizando el teorema de factorización que el estadístico $T = \sum_{i=1}^{n} X_i$ es suficiente para el parámetro p.

Solución:

La función de probabilidad conjunta de la muestra será:

$$\begin{aligned}
P_p(x_1, ..., x_n) &= P_p(X_1 = x_1, ..., X_n = x_n) \\
&= p^{x_1}(1-p)^{1-x_1}...p^{x_n}(1-p)^{1-x_n} \\
&= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \\
&= p^t(1-p)^{n-t} \\
&= \left(\frac{p}{1-p}\right)^t (1-p)^n
\end{aligned}$$

Haciendo

$$g(T(x_1, ..., x_n); p) = \left(\frac{p}{1-p}\right)^t (1-p)^n$$

$$h(x_1, ..., x_n) = 1$$

entonces resulta la siguiente factorización:

$$P_p(x_1, ..., x_n) = g(t; p) \cdot h(x_1, ..., x_n) = \left(\frac{p}{1-p}\right)^t (1-p)^n \cdot 1$$

Por tanto, el número de éxitos es un estadístico suficiente para el parámetro p (probabilidad de éxito en una distribución binomial).

Ejemplo 2.10

Sea $(X_1, ..., X_n)$ una muestra aleatoria procedente de una distribución $\Gamma\left(1, \frac{1}{a}\right)$, cuya función de densidad es:

$$f(x) = \begin{cases} \frac{1}{a} e^{-\frac{x}{a}} & , \quad x > 0 \\ 0 & , \quad x \leqslant 0 \end{cases}$$

Obtener un estadístico suficiente para el parámetro a.

Solución:

La función de densidad conjunta de la muestra es:

$$f(x_1, ..., x_n; a) = f(x_1; a)...f(x_n; a)$$
$$= \frac{1}{a} e^{-\frac{x_1}{a}} \cdot \ldots \cdot \frac{1}{a} e^{-\frac{x_n}{a}}$$
$$= \frac{1}{a^n} e^{-\frac{1}{a}(x_1 + \cdots + x_n)}$$
$$= \frac{1}{a^n} e^{-\frac{1}{a}\sum_{i=1}^{n} x_i} \cdot 1$$

Por tanto, si hacemos:

$$t = T(x_1, ..., x_n) = \sum_{i=1}^{n} x_i$$

entonces se tiene:

$$g(T(x_1, ..., x_n); \theta) = \frac{1}{a^n} \cdot e^{-\frac{1}{a}\sum_{i=1}^{n} x_i} = \frac{1}{a^n} e^{-\frac{t}{a}}$$
$$h(x_1, ..., x_n) = 1$$

Luego tendremos la siguiente factorización:

$$f(x_1, ..., x_n; a) = g(T(x_1, ..., x_n); a) \cdot h(x_1, ..., x_n)$$
$$= \frac{1}{a^n} e^{-\frac{1}{a}\sum_{i=1}^{n} x_i} \cdot 1 = \frac{1}{a^n} e^{-\frac{t}{a}} \cdot 1$$

Y podemos decir que el estadístico

$$T = \sum_{i=1}^{n} X_i$$

es un estadístico suficiente para el parámetro a.

Observemos que el estadístico media muestral $\bar{X}$ es también un estadístico suficiente para el parámetro a. En efecto, haciendo

$$T = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Rightarrow \quad nT = \sum_{i=1}^{n} X_i$$

tendríamos la siguiente factorización:

$$\begin{aligned} f(x_1, \ldots, x_n; a) &= \frac{1}{a^n} e^{-\frac{1}{a}\sum_{i=1}^{n} x_i} \cdot 1 \\ &= \frac{1}{a^n} e^{-\frac{1}{a} nt} \cdot 1 \\ &= g(T(x_1, \ldots, x_n); a) \cdot h(x_1, \ldots, x_n) \end{aligned}$$

Lo cual indica que pueden existir varios estadísticos suficientes para un mismo parámetro.

Otro resultado interesante que se ha puesto de manifiesto en el ejemplo anterior, lo recogemos en el siguiente Teorema, que es una consecuencia inmediata del teorema de factorización de Fisher-Neyman.

Teorema 2.4

Si el estadístico T_1 es suficiente y es función con inversa única del estadístico T_2, $T_1 = f(T_2)$, entonces el estadístico T_2 es también suficiente.

Demostración:

Sea $T_1 = f(T_2)$ donde f es inyectiva. Entonces existe la inversa $T_2 = f^{-1}(T_1)$ con lo cual, por ser T_1, suficiente, tenemos según la expresión [2.25] que:

$$\begin{aligned} f(x_1, \ldots, x_n; \theta) &= g(T_1; \theta) \cdot h(x_1, \ldots, x_n) \\ &= g(f(T_2); \theta) \cdot h(x_1, \ldots, x_n) \\ &= g^*(T_2; \theta) \cdot h(x_1, \ldots, x_n) \end{aligned}$$

lo cual demuestra que el estadístico T_2 también es suficiente.

Intuitivamente también se puede entender, pues si el estadístico T_1 puede calcularse a partir del estadístico T_2, entonces el conocimiento de T_2, debe de ser al menos tan bueno como el de T_1.

Esto es equivalente a decir: que si un estadístico no es suficiente ninguna reducción suya puede ser suficiente.

El recíproco del teorema 2.4, que no demostraremos, también se verifica y lo podemos enumerar mediante el siguiente teorema.

Teorema 2.5

Si los estadísticos T_1 y T_2 son suficientes para el parámetro θ entonces T_1 y T_2 están relacionados funcionalmente.

Cuando la distribución de la población depende de dos parámetros, como es el caso de la distribución normal, es interesante determinar dos estadísticos que sean **conjuntamente suficientes** para los dos parámetros. En estas situaciones el teorema de factorización se puede enunciar de la siguiente forma.

Teorema 2.6

Los estadísticos $T_1 = T_1(X_1, ..., X_n)$ y $T_2 = T_2(X_1, ..., X_n)$ son **conjuntamente suficientes** para los parámetros θ_1, y θ_2 si y solamente si la función de probabilidad o la función de densidad de la muestra se puede descomponer factorialmente de la siguiente forma:

$$f(x_1, ..., x_n; \theta_1, \theta_2) =$$
$$= g(T_1(x_1, ..., x_n), T_2\,(x_1, ..., x_n); \theta_1, \theta_2) \cdot h(x_1, ..., x_n) \qquad [2.26]$$

Ejemplo 2.11

Sea una muestra aleatoria $(X_1, ..., X_n)$ de una población $N(\mu, \sigma)$. Obtener dos estadísticos que sean conjuntamente suficientes para los parámetros poblacionales μ y σ.

Solución:

La función de densidad conjunta de la muestra será:

$$f(x_1, ..., x_n; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n}\, e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

$$= (\sigma\sqrt{2\pi})^{-n}\, e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2\right)} \cdot 1$$

Siguiendo la notación utilizada en la expresión [2.26], tenemos que:

$$T_1(x_1, ..., x_n) = \sum_{i=1}^{n} x_i$$

$$T_2(x_1, ..., x_n) = \sum_{i=1}^{n} x_i^2$$

$$g(T_1(x_1, ..., x_n), T_2(x_1, ..., x_n); \mu, \sigma) = (\sigma\sqrt{2\pi})^{-n} \cdot e^{-\frac{1}{2\sigma^2}(T_2(x_1, ..., x_n) - 2\mu T_1(x_1, ..., x_n) + n\mu^2)}$$

$$h(x_1, ..., x_n) = 1$$

verificándose la factorización dada en la expresión [2.26], y por tanto los estadísticos:

$$T_1 = \sum_{i=1}^{n} X_i$$

$$T_2 = \sum_{i=1}^{n} X_i^2$$

son conjuntamente suficientes para los parámetros μ y σ.

2.3.5.3. Estadístico mínimal suficiente

Hemos introducido el concepto de suficiencia, y decíamos que el objetivo era mediante el estadístico suficiente *condensar o resumir los datos sin que se produzca pérdida de información sobre el parámetro*. Ahora lo que pretendemos es obtener otro estadístico suficiente que reduzca o resuma los datos lo más posible, pero sin pérdida de información sobre el parámetro, y éste será el **estadístico mínimal suficiente.**

Supongamos que un estadístico T_1 es suficiente para el parámetro poblacional θ, y que además existe otro estadístico T_2, tal que $T_1 = f(T_2)$, entonces sabemos, por el teorema 2.4, que el estadístico T_2 es también suficiente para el parámetro θ. Ahora bien, salvo que la función $f(\cdot)$ sea biyectiva, el estadístico T_1 proporcionará una mayor reducción de los datos originales que el estadístico T_2.

En efecto, si volvemos al ejemplo 2.11 vemos que el estadístico:

$$T_1 = \sum_{i=1}^{n} X_i$$

es suficiente para el parámetro a.

Haciendo

$$f(X_1, ..., X_n) = \sum_{i=1}^{n} X_i$$

vemos que $(X_1, ..., X_n)$ es también un estadístico suficiente para el parámetro a.

Pero es evidente que el estadístico T_1 produce una reducción bastante mayor en los datos que si consideramos simplemente los datos de la muestra original. Debido a esto es deseable determinar, si es posible, el estadístico suficiente que produce la mayor reducción de datos, siendo éste el **estadístico mínimal suficiente.**

El hecho de que $T_1 = f(T_2)$ nos asegura que el estadístico T_1 siempre nos dará una reducción de los datos que al menos es tan buena como la dada por el estadístico T_2, si es que sigue siendo suficiente.

Definición 2.13. Estadístico mínimal suficiente.

> Diremos que un estadístico es **mínimal suficiente,** si es suficiente y cualquier reducción de la información definida por él ya no es suficiente, es decir desprecia información que está contenida en la muestra, acerca del parámetro θ.

Método de Lehmann y Scheffé[23] para obtener un estadístico mínimal suficiente

Este método parte de la existencia de dos muestras aleatorias simples de tamaño:

$$(X_1, ..., X_n) \quad \text{e} \quad (Y_1, ..., Y_n)$$

cuyas respectivas funciones de verosimilitud son:

$$L(x_1, ..., x_n; \theta) = f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

$$L(y_1, ..., y_n; \theta) = f(y_1, ..., y_n; \theta) = \prod_{i=1}^{n} f(y_i; \theta)$$

[23] Ver LINGREN, B. (1968), pág. 235.

Se obtiene el cociente de funciones de verosimilitud:

$$\frac{L(x_1, ..., x_n; \theta)}{L(y_1, ..., y_n; \theta)}$$

y si podemos encontrar una función $g(x_1, ..., x_n)$ tal que la razón de funciones de verosimilitud no dependa de θ si y solamente si

$$g(x_1, ..., x_n) = g(y_1, ..., y_n)$$

entonces decimos que

$$g(x_1, ..., x_n)$$

será el estadístico mínimal suficiente para el parámetro θ.

Si en lugar de existir un solo parámetro θ, existieran k parámetros, entonces tendríamos que obtener k funciones

$$g_1(x_1, ..., x_n), ..., g_k(x_1, ..., x_n)$$

tales que el cociente de funciones de verosimilitud no depende de $\theta_1, ..., \theta_k$, si y solamente si

$$g_i(x_1, ..., x_n) = g_i(y_1, ..., y_n), \quad \text{para} \quad i = 1, ..., k$$

y entonces decimos que

$$g_i(x_1, ..., x_n), \quad \text{para} \quad i = 1, ..., k$$

serán los estadísticos conjuntamente mínimal suficientes para los parámetros $\theta_1, ..., \theta_k$.

Ejemplo 2.12

Sea una muestra aleatoria $(X_1, ..., X_n)$ procedente de una población binomial, $B(1, p)$. Obtener, si existe, un estadístico mínimal suficiente para el parámetro θ.

Solución:

En el ejemplo 2.9 ya se había obtenido un estadístico suficiente para el parámetro p, y veíamos que, efectivamente, el estadístico $T = \sum_{i=1}^{n} X_i$ era suficiente.

Ahora vamos a tratar de obtener un estadístico mínimal suficiente, para ello consideramos dos muestras de tamaño n

$$(X_1, \ldots, X_n) \quad \text{e} \quad (Y_1, \ldots, Y_n)$$

y obtenemos la razón de funciones de verosimilitud:

$$\frac{L(x_1, \ldots, x_n); p)}{L(y_1, \ldots, y_n; p)} = \frac{p^{x_1}(1-p)^{1-x_1} \cdot \ldots \cdot p^{x_n}(1-p)^{1-x_n}}{p^{y_1}(1-p)^{1-y_1} \cdot \ldots \cdot p^{y_n}(1-p)^{1-y_n}}$$

$$= \frac{p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}}{p^{\sum_{i=1}^{n} y_i}(1-p)^{n-\sum_{i=1}^{n} y_i}}$$

$$= \left(\frac{p}{1-p}\right)^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}$$

que como vemos depende del parámetro, y únicamente no dependerá del parámetro p si y sólo si

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

Ejemplo 2.13

Sea una muestra aleatoria $(X_1, \ldots, X_n)$ procedente de una distribución $N(\mu, 1)$. Obtener un estimador mínimal suficiente del parámetro μ.

Resultando que efectivamente el estadístico $\sum_{i=1}^{n} X_i$ será minimal suficiente para el parámetro p.

Solución:

Considerando dos muestras de tamaño n

$$(X_1, \ldots, X_n) \quad \text{e} \quad (Y_1, \ldots, Y_n)$$

podemos obtener la razón de funciones de verosimilitud:

$$\frac{L(x_1, ..., x_n; \mu)}{L(y_1, ..., y_n; \mu)} = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)e^{-\frac{(x_1-\mu)^2}{2}} \cdots \left(\frac{1}{\sqrt{2\pi}}\right)e^{-\frac{x_n-\mu)^2}{2}}}{\left(\frac{1}{\sqrt{2\pi}}\right)e^{-\frac{(y_1-\mu)^2}{2}} \cdots \left(\frac{1}{\sqrt{2\pi}}\right)e^{-\frac{y_n-\mu)^2}{2}}}$$

$$= e^{-\frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\mu)^2 - \sum_{i=1}^{n}(y_i-\mu)^2\right)}$$

$$= e^{-\frac{1}{2}\left(\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}y_i^2\right) + \mu\left(\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}y_i\right)}$$

Esta función no dependerá de μ si y solamente si

$$\sum_{i=1}^{n} \mathrm{x}_i = \sum_{i=1}^{n} \mathrm{y}_i$$

Por tanto, el estadístico

$$\sum_{i=1}^{n} X_i$$

es minimal suficiente. Y puesto que $\bar{X}$ es una función inyectiva de $\sum_{i=1}^{n} X_1$, resulta que $\bar{X}$ es también un estadístico mínimal suficiente.

Ejemplo 2.14

Sea una muestra aleatoria $(X_1, ..., X_n)$ procedente de una población cuya función de densidad es:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Obtener dos estadísticos para los parámetros μ y σ que sean conjuntamente mínimal suficientes.

Solución:

En el ejemplo 2.11 ya habíamos obtenido dos estadísticos conjuntamente suficientes para los parámetros μ y σ. Veamos ahora si existen dos estadísticos que sean conjuntamente mínimal suficientes.

Consideramos dos muestras de tamaño n

$$(X_1, ..., X_n) \quad \text{e} \quad (Y_1, ..., Y_n)$$

y obtenemos la razón de funciones de verosimilitud.

$$\frac{L(x_1, ..., x_n; \mu, \sigma)}{L(y_1, ..., y_n; \mu, \sigma)} = \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}}{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_n-\mu)^2}{2\sigma^2}}}$$

$$= \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}}}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^n \frac{(y_i-\mu)^2}{2\sigma^2}}}$$

$$= e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i-\mu)^2 - \sum_{i=1}^n (y_i-\mu)^2\right)} \quad [24]$$

$$= e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right) + \frac{\mu}{\sigma^2}\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right)}$$

que como vemos depende de los parámetros μ y σ, únicamente no dependerá de estos parámetros μ y σ si y sólo si:

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2 \quad \text{y} \quad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

Resultando que los estadísticos

$$\sum_{i=1}^n X_i$$

$$\sum_{i=1}^n X_i^2$$

[24] Desarrollando y simplificando resulta:

$$-\frac{1}{2\sigma^2}\left[\left(\sum_{i=1}^n x_i^2 + n\mu^2 - 2\mu\sum_{i=1}^n x_i\right) - \left(\sum_{i=1}^n y_i^2 + n\mu^2 - 2\mu\sum_{i=1}^n y_i\right)\right] =$$

$$= -\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right) + \frac{\mu}{\sigma^2}\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right)$$

que ya habíamos visto que eran conjuntamente suficientes, resultan ser conjuntamente mínimal suficientes para los parámetros μ y σ.

2.3.5.4. Relación entre el estimador eficiente y suficiente

Si un estimador $\hat{\theta}$ es suficiente ha de verificarse por el teorema 2.1 que:

$$\frac{\partial \ln d\,F_n(x_1, \ldots, x_n;\ \theta)}{\partial \theta} = A(\theta)(\hat{\theta} - \theta)$$

o bien, sustituyendo $A(\theta)(\hat{\theta} - \theta)$ por $\dfrac{\partial \ln g(\hat{\theta}, \theta)}{\partial \theta}$, tendremos:

$$\frac{\partial \ln d\,F_n(x_1, \ldots, x_n;\ \theta)}{\partial \theta} = \frac{\partial \ln g(\hat{\theta}, \theta)}{\partial \theta}$$

integrando respecto de θ, y expresando la constante de integración como $\ln h(x_1, \ldots, x_n)$ resulta:

$$\ln d\,F_n(x_1, \ldots, x_n; \theta) = \ln g(\hat{\theta}, \theta) + \ln h(x_1, \ldots, x_n)$$

de donde:

$$d\,F_n(x_1, \ldots, x_n;\ \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, \ldots, x_n)$$

que por el criterio de factorización de Fisher-Neyman resulta que el estimador $\hat{\theta}$ es suficiente.

Luego si el estimador $\hat{\theta}$ es eficiente, también es suficiente.

Ejemplo 2.15

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple procedente de una población con distribución de Poisson de parámetro λ, en donde el parámetro λ se estima a partir de la media $\bar{X}$ de la muestra aleatoria del tamaño n. Obtener:

1.° Un estimador eficiente.

2.° Un estimador suficiente.

Solución:

1. La función de probabilidad de Poisson viene dada por:

$$P(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad ; \quad x = 0, 1, 2, \ldots$$

Según la definición 2.4 para que un estimador $\hat{\lambda}$ sea eficiente se tiene que verificar que la varianza del estimador coincida con la cota de Frechet-Cramer-Rao.

$$\operatorname{Var}(\hat{\lambda}) = \frac{1}{n\, E\left[\dfrac{\partial \ln P(x; \lambda)}{\partial \lambda}\right]^2}$$

Aplicando esta expresión a la distribución de Poisson, resulta:

$$\ln P(x; \lambda) = x \ln \lambda - \ln (x!) - \lambda$$

$$\frac{\partial \ln P(x; \lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1 = \frac{x - \lambda}{\lambda}$$

$$E\left[\frac{\partial \ln P(X; \lambda)}{\partial \lambda}\right]^2 = E\left[\frac{X - \lambda}{\lambda}\right]^2 = \frac{1}{\lambda^2} E[X - \lambda]^2$$

$$= \frac{1}{\lambda^2} \operatorname{Var}(X) = \frac{1}{\lambda^2}\, \lambda$$

$$= \frac{1}{\lambda}$$

pues en la distribución de Poisson sabemos que

$$E[X] = \lambda$$
$$\operatorname{Var}(X) = \lambda$$

Pero sabemos que en la distribución de Poisson el parámetro λ se estima mediante la media $\bar{X}$ de una muestra aleatoria; siendo la media $\bar{X}$ muestral un estimador insesgado del parámetro λ

$$E[\bar{X}] = \lambda$$

y como:

$$\operatorname{Var}(\bar{X}) = \frac{\lambda}{n}$$

Sustituyendo en la expresión de la Cota de Frechet-Cramer-Rao, resulta:

$$\frac{1}{n\,E\left[\frac{\partial \ln P(x;\lambda)}{\partial \lambda}\right]^2} = \frac{\lambda}{n}$$

y como

$$\begin{aligned}
\mathrm{Var}(\hat{\lambda}) = \mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) \\
&= \frac{1}{n^2}\, n\,\lambda \\
&= \frac{\lambda}{n}
\end{aligned}$$

resulta que la $\mathrm{Var}(\hat{\lambda})$ coincide con la cota de Frechet-Cramer-Rao.

Luego la media muestral $\bar{X}$ es un estimador eficiente del parámetro λ de Poisson.

2. Obtengamos ahora un estimador suficiente para el parámetro λ. Aplicando el criterio de factorización de Fisher-Neyman, tendremos que probar:

$$P_\lambda(x_1, \ldots, x_n) = g(T(x_1, \ldots, x_n); \lambda) \cdot h(x_1, \ldots, x_n)$$

La función de probabilidad conjunta de la muestra será:

$$\begin{aligned}
P_\lambda(x_1, \ldots, x_n) = P(x_1, \ldots, x_n; \lambda) &= P(x_1; \lambda) \ldots P(x_n; \lambda) \\
&= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdots \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} \\
&= \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda} \\
&= \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \cdot \frac{1}{\prod_{i=1}^{n} x_i!}
\end{aligned}$$

$$= g\left(\sum_{i=1}^{n} x_i; \lambda\right) \cdot h(x_1, ..., x_n)$$

y el estadístico $\sum_{i=1}^{n} X_i$ es un estimador suficiente para el parámetro λ. Pero como el estadístico $\sum_{i=1}^{n} X_i$ es función biyectiva del estadístico $\bar{X}$, pues $\sum_{i=1}^{n} X_i = n\bar{X}$, y $\sum_{i=1}^{n} X$ es suficiente, entonces por el teorema 2.4 resulta que el estadístico $\bar{X}$ también es suficiente para el parámetro λ.

Luego el estadístico media muestral es un estimador suficiente y eficiente del parámetro λ.

2.3.5.5. El papel de la suficiencia en la obtención de estimadores de mínima varianza

La suficiencia juega un papel importante en la obtención de estimadores insesgados uniformemente de mínima varianza (UMVUE)[25] como se pone de manifiesto a continuación.

Teorema de Rao-Blackwell

Sea una población con función de densidad o de cuantía representada por $f(x; \theta)$ y sea $\hat{\theta}$ un estimador insesgado para el parámetro θ y T un estadístico suficiente del mismo parámetro θ. Entonces si hacemos:

$$g(T) = E[\hat{\theta}/T]$$

se verifica:

1. $g(T)$ es un estadístico y es función del estadístico suficiente.
2. $E[g(T)] = \theta$.
3. $\mathrm{Var}(g(T)) \leqslant \mathrm{Var}(\hat{\theta})$.

Es decir, el estadístico $g(T)$ es función del estadístico suficiente, es un estimador insesgado de θ y su varianza es menor que la del estimador insesgado $\hat{\theta}$.

[25] Uniformly minimum-variance unbiased estimators (UMVUE). Estimador insesgado uniformemente de mínima varianza. Si existe un estimador UMVUE ésta será preferible a cualquier otro estimador insesgado de θ, ya que sus valores presentan menos varianza que la de cualquier otro estimador insesgado.

Este teorema, que no demostraremos aquí, nos indica que dado un estimador insesgado y un estadístico suficiente, este estadístico suficiente lo podemos utilizar para encontrar otro estimador $g(T)$ insesgado y de menor varianza que el primero. Ahora bien, no se puede asegurar que el estimador $g(T)$ sea de mínima varianza, es decir, UMVUE. Para ello recurrimos al teorema de Lehmann-Scheffé que veremos posteriormente.

Corolario

Si existe un estimador $\hat{\theta}$ UMVUE, entonces debe ser función del estadístico mínimal suficiente para el parámetro θ, el cual es UMVUE.

2.3.6. COMPLETITUD

En la sección anterior hemos estudiado la suficiencia y veíamos que mediante este concepto podíamos resumir la información contenida en la muestra sobre un parámetro desconocido de manera más eficiente y sin pérdida de información sobre el parámetro. Ahora mediante el nuevo concepto de **completitud,** veremos que cuando se verifica para un estadístico suficiente entonces obtenemos mejores estimadores.

Definición 2.14. Familia completa.

Una familia de distribuciones $\{F(x;\theta)\}$ es **completa** si para cualquier función $h(x)$ la identidad:

$$E[h(x)] = 0$$

implica que:

$$P(h(x) = 0) = 1$$

en todos los puntos para los cuales $f(x;\theta) > 0$ para algún θ.

Esta definición nos indica que una familia de distribuciones es completa si el único estimador insesgado de cero es el mismo cero.

Un **estadístico *T* es completo** si la correspondiente familia de distribuciones de T es completa. Así pues se pone de manifiesto que la propiedad de completitud es una propiedad de la familia de distribuciones.

Ejemplo 2.16

Dada la familia de distribuciones binomiales $\{B(n, p)\}$ comprobar si es completa.

Solución:

Teniendo en cuenta la definición 2.14, vemos que para cualquier real $h(x)$ de una variable aleatoria $X \to B(n, p)$ las identidad

$$E[h(X)] = \sum_{x=0}^{n} \binom{n}{x} h(x) p^x (1 - p)^{n-x} = 0, \quad \forall\, p \in (0, 1)$$

implica necesariamente que

$$h(x) = 0 \quad , \quad \forall\, x = 0, 1, ..., n$$

ya que la expresión

$$\sum_{x=0}^{n} \binom{n}{x} h(x) p^x (1 - p)^{n-x}$$

es un polinomio en p de grado n, y para que tome el valor cero para todo valor del parámetro p es necesario que todos sus coeficientes, $h(x)$, sean nulos.

Luego

$$P(h(x) = 0) = 1 \quad , \quad \forall\, p \in (0, 1)$$

y la familia es completa.

Ejemplo 2.17

Supongamos una muestra aleatoria $(X_1, ..., X_n)$ procedente de una población $B(1, p)$, y sea el estadístico

$$T = X_1 + \cdots + X_n$$

Comprobar si el estadístico T es completo.

Solución:

Sabemos que la distribución binomial es reproductiva respecto al parámetro n, y hemos visto en el ejemplo 2.16 que la familia de distribuciones binomiales $\{B(n,p)\}$ es completa, luego el estadístico $T = \sum_{i=1}^{n} X_i$ es completa pues, sigue una distribución $B(n,p)$, por la reproductividad de la distribución $B(1,p)$.

Definición 2.15. Estadístico suficiente completo.

> Diremos que un **estadístico suficiente T es completo,** si la familia de distribuciones del estadístico suficiente T es completa.

Teorema de Lehmann-Scheffé

> Si T es un estadístico suficiente y completo para θ, y si existe un estimador insesgado $\hat{\theta}$, del parámetro θ, entonces existe un único estimador UMVUE dado por
>
> $$g(T) = E[\hat{\theta}/T]$$

Luego el problema de encontrar un estimador UMVUE ha quedado reducido a la obtención de un estimador insesgado $\hat{\theta}$ y a calcular el valor esperado

$$g(T) = E[\hat{\theta}/T)$$

en donde T es un estadístico suficiente completo.

Para finalizar con la completitud daremos el concepto de estadístico complementario y un teorema que pone de manifiesto la independencia del estadístico complementario con el estadístico suficiente completo.

Definición 2.16. Estadístico complementario.

> Diremos que un estadístico U es un **estadístico complementario** para el parámetro θ, si la distribución de U es independiente de θ.

Teorema 2.7

> Sea el estadístico T suficiente y completo para el parámetro θ, y sea U un estadístico complementario para θ. Entonces los estadísticos T y U son variables aleatorias independientes[26].

[26] Este teorema debido a Basu nos facilita la demostración de la independencia de los estadísticos media y varianza muestral de una distribución normal.

2.4. LA FAMILIA EXPONENCIAL DE DISTRIBUCIONES Y LA SUFICIENCIA

Existe una clase o familia de distribuciones en la que todos los parámetros de las distribuciones que la integran tienen estadísticos suficientes. Este grupo de distribuciones recibe el nombre de **familia exponencial de distribuciones,** y como veremos será bastante fácil obtener estadísticos suficientes para conseguir información acerca del parámetro correspondiente.

Definición 2.17. Familia exponencial de distribuciones uniparamétrica.

Diremos que una **familia de distribuciones es exponencial uniparamétrica** si está formada por todas aquellas distribuciones cuyas funciones de cuantía o de densidad se expresan de la siguiente forma:

$$f(x;\theta) = B(\theta)h(x)e^{Q(\theta)R(x)} \qquad [2.27]$$

donde

1. $B(\theta)$ y $Q(\theta)$ son funciones reales de θ.
2. $h(x)$ y $R(x)$ son funciones reales de x.

En la tabla de la página siguiente aparecen algunas distribuciones pertenecientes a la familia exponencial.

Veamos ahora que utilizando el método de Lehmann-Scheffé podemos obtener un estadístico minimal suficiente para la familia exponencial de distribuciones. En efecto, si consideramos dos muestras aleatorias simples:

$$(X_1, \ldots, X_n) \quad \text{e} \quad (Y_1, \ldots, Y_n)$$

cuyas respectivas funciones de verosimilitud son:

$$\begin{aligned} L(x_1, \ldots, x_n;\theta) &= f(x_1, \ldots, x_n;\theta) = B^n(\theta)\cdot \prod_{i=1}^{n} h(x_i)e^{Q(\theta)\sum_{i=1}^{n} R(x_i)} \\ L(y_1, \ldots, y_n;\theta) &= f(y_1, \ldots, y_n;\theta) = B^n(\theta)\cdot \prod_{i=1}^{n} h(y_i)e^{Q(\theta)\sum_{i=1}^{n} R(y_i)} \end{aligned} \qquad [2.28]$$

y podemos obtener la razón de funciones de verosimilitud:

$$\frac{L(x_1, \ldots, x_n;\ \theta)}{L(y_1, \ldots, y_n;\ \theta)} = \prod_{i=1}^{n} \frac{h(x_i)}{h(y_i)}\ e^{Q(\theta)\left(\sum_{i=1}^{n} R(x_i) - \sum_{i=1}^{n} R(y_i)\right)}$$

Distribución	$f(x;\theta)$	$B(\theta)$	$h(x)$	$Q(\theta)$	$R(x)$
Binomial $(1, p)$	$p^x(1-p)^{1-x}$	$1-p$	1	$\ln \frac{p}{1-p}$	x
Binomial (n, p)	$\binom{n}{x} p^x(1-p)^{n-x}$	$(1-p)^n$	$\binom{n}{x}$	$\ln \frac{p}{1-p}$	x
Geométrica	$p(1-p)^x$	p	1	$\ln(1-p)$	x
Binomial negativa	$\binom{r+x-1}{x} p^r(1-p)^x$	p^r	$\binom{r+x-1}{x}$	$\ln(1-p)$	x
Poisson	$e^{-\lambda}\frac{\lambda^x}{x!}$	$e^{-\lambda}$	$\frac{1}{x!}$	$\ln\lambda$	x
Normal $(0, \sigma)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$	$\frac{1}{\sigma\sqrt{2\pi}}$	1	$-\frac{1}{2\sigma^2}$	x^2
Normal $(\mu, 1)$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2}}$	$e^{-\frac{x^2}{2}}$	μ	x
Gamma	$\frac{a^p}{\Gamma(p)} x^{p-1}e^{-ax}$	$\frac{a^p}{\Gamma(p)}$	x^{p-1}	$-a$	x
Exponencial	ae^{-ax}	a	1	$-a$	x

que será independiente de θ, si y solamente si:

$$\sum_{i=1}^{n} R(x_i) = \sum_{i=1}^{n} R(y_i)$$

y por tanto el estadístico

$$T(x_1, ..., x_n) = \sum_{i=1}^{n} R(x_i)$$

será un estadístico mínimal suficiente.

Veamos ahora que este estadístico mínimal suficiente, $\sum_{i=1}^{n} R(x_i)$, tiene una distribución que pertenece a la familia exponencial. La demostración la haremos sólo para el caso discreto, pues en el caso continuo se tendría que hacer una transformación de una integral múltiple.

La función de probabilidad para el estadístico

$$T(x_1, ..., x_n) = \sum_{i=1}^{n} R(x_i)$$

será:

$$P(T; \theta) = P\left(\sum_{i=1}^{n} R(x_i) = t\right) = \sum_{\Sigma R(x_i)} P(X_1 = x_1, ..., X_n = x_n)$$

$$= \sum_{\Sigma R(x_i)=t} B^n(\theta) \prod_{i=1}^{n} h(x_i) \cdot e^{Q(\theta) \cdot R(x_i) \sum_{i=1}^{n} R(x_i)}$$

$$= b(\theta)H(t)e^{tQ(\theta)}$$

en donde

$$b(\theta) = B^n(\theta) \quad \text{y} \quad H(t) = \sum_{\Sigma R(x_i)=t} \prod_{i=1}^{n} h(x_i)$$

Luego $P(T; \theta)$ pertenece a la familia exponencial de distribuciones.

Análogamente podemos hacer la extensión al caso de κ-parámetros.

Definición 2.18. Familia exponencial de distribuciones κ-paramétrica.

Diremos que una familia de distribuciones es exponencial κ-paramétrica si está formada por todas aquellas distribuciones cuyas funciones de cuantía o de densidad, se expresan de la siguiente forma:

$$f(x; \theta_1, ..., \theta_k) = B(\theta_1, ..., \theta_k)h(x)e^{Q_1(\theta_1...\theta_k)R_1(x) + \cdots + Q_k(\theta_1...\theta_k)R_k(x)} \quad [2.28]$$

donde:

1. $B(\theta_1, ..., \theta_\kappa)$ y $Q_i(\theta_1, ..., \theta_\kappa)$ son funciones reales de $\theta_1...\theta_\kappa$.
2. $h(x)$ y $R_i(x)$ son funciones reales de x.

De manera análoga al caso uniparamétrico aquí se tiene que el estadístico κ-dimensional

$$\left(\sum_{i=1}^{n} R_i(x_i), ..., \sum_{i=1}^{n} R_n(x_i)\right)$$

es mínimal suficiente para la familia.

Ejemplo 2.18

La distribución $N(\mu, \sigma)$ pertenece a la familia exponencial de distribuciones *bi*-paramétricas.

Solución:

Sabemos que la función de densidad de la $N(\mu, \sigma)$ es

$$f(x;\, \mu,\, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{\mu^2}{2\sigma^2}} e^{-\frac{x^2}{2\sigma^2} + x\frac{\mu}{\sigma^2}}$$

y haciendo:

$$B(\mu,\, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{\mu^2}{2\sigma^2}} \qquad h(x) = 1$$

$$Q_1(\mu,\, \sigma) = -\frac{1}{2\sigma^2} \qquad R_1(x) = x^2$$

$$Q_2(\mu,\, \sigma) = \frac{\mu}{\sigma^2} \qquad R_2(x) = x$$

se tiene la forma de la expresión [2.29].

2.5. ESTIMADOR INVARIANTE

Al estudiar la suficiencia y la completitud empleábamos la propiedad de insesgadez para reducir la clase de estimadores, con la esperanza de obtener un estimador de varianza mínima dentro de esa clase reducida de estimadores insesgados. Ahora introducimos, como propiedad alternativa la **invarianza,** que puede ser utilizada para restringir la clase de estimadores.

Para introducir el concepto de **invarianza** consideramos, por ejemplo, un experimento que consiste en medir la velocidad de varios coches. Entonces un estimador obtenido para la velocidad expresada en millas por hora debe de corresponderse con el estimador obtenido utilizando como unidad de medida, kilómetro por hora, y de esta manera el procedimiento estadístico de estimación puntual debe de ser tal que el estimador que se utilice sea invariante frente a cambios de escala, es decir, el estimador será independiente de la escala de medida.

Definición 2.18. Estimador invariante.

> Diremos que un estimador $\hat{\theta}$ es **invariante,** si se verifica que el estimador de una función del parámetro θ, es igual a la función del estimador del parámetro, es decir cuando se verifica que:
>
> $$\widehat{(f(\theta))} = f(\hat{\theta})$$
>
> En realidad deberíamos hablar de **método de estimación invariante,** pues realmente es el método de estimación lo que permanece invariante frente a una transformación.

Así, por ejemplo, si el estimador de la varianza poblacional σ^2 es la varianza muestral S^2, entonces si el estimador (más correctamente, si el método de estimación) fuera invariante debería de suceder que el estimador de la desviación típica poblacional σ, debería ser la desviación típica muestral S. Es decir, si el estimador de la varianza poblacional es invariante:

$$\hat{\sigma}^2 = S^2$$

entonces debería suceder que

$$\hat{\sigma} = S$$

Estudiaremos cuatro tipos de invarianzas o de estimadores invariantes:

- Estimador invariante a cambios de origen.
- Estimador invariante a cambios de escala.
- Estimador invariante a cambios de origen y de escala.
- Estimador invariante a permutaciones.

Definición 2.19. Estimador invariante a cambio de origen.

> Sea una muestra aleatoria de tamaño n, $(X_1, ..., X_n)$ y un estimador $\hat{\theta}(X_1, ..., X_n)$ del parámetro θ, entonces si realizamos un cambio de origen en los datos de la muestra, por ejemplo sumando una constante κ, la muestra se transforma en $(X_1 + \kappa, ..., X_n + \kappa)$, y diremos que el estimador $\hat{\theta}$ es **invariante a cambios de origen** o **de localización** si y solamente si se verifica que:
>
> $$\hat{\theta}(X_1 + \kappa, ..., X_n + \kappa) = \hat{\theta}(X_1, ..., X_n), \quad \forall \kappa \in \mathbb{R}$$
>
> es decir, el estimador es el mismo para los datos transformados.

Ejemplo 2.19

Estudiar si son o no invariantes frente a cambios de origen los siguientes estimadores:

1. La media muestral $\bar{X}$.
2. La varianza muestral.
3. La desviación típica muestral.
4. El estadístico $\frac{Y_1 + Y_2}{2}$, siendo

$$Y_1 = \text{mín}\,(X_1, ..., X_n)$$
$$Y_2 = \text{máx}\,(X_1, ..., X_n)$$

5. El coeficiente de correlación lineal.

Solución:

1. Se trata del estimador:

$$\hat{\theta}(X_1, ..., X_n) = \bar{X}$$

entonces:

$$\begin{aligned}\hat{\theta}(X_1 + \kappa, ..., X_n + \kappa) &= \frac{\sum_{i=1}^{n} (X_i + \kappa)}{n} = \frac{\sum_{i=1}^{n} X_i}{n} + \kappa \\ &= \bar{X} + \kappa \\ &\neq \hat{\theta}(X_1, ..., X_n) = \bar{X}\end{aligned}$$

Luego no es invariante.

2. El estimador es:

$$\hat{\theta}(X_1, ..., X_n) = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

entonces

$$\begin{aligned}\hat{\theta}(X_1 + \kappa, ..., X_n + \kappa) &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i + \kappa - \bar{X} - \kappa)^2 \\ &= S^2 \\ &= \hat{\theta}(X_1, ..., X_n)\end{aligned}$$

Luego es invariante.

3. Análogamente el estimador desviación típica muestral también es invariante, por serlo S^2.

4. *El estimador es*:

$$\hat{\theta}(X_1, ..., X_n) = \frac{\text{mín}\,(X_1, ..., X_n) + \text{máx}\,(X_1, ..., X_n)}{2}$$

entonces:

$$\hat{\theta}(X_1 + \kappa, ..., X_n + \kappa) = \frac{\text{mín}\,(X_1 + \kappa, ..., X_n + \kappa) + \text{máx}\,(X_1 + \kappa, ..., X_n + \kappa)}{2}$$

$$= \frac{\text{mín}\,(X_1, ..., X_n) + \kappa + \text{máx}\,(X_1, ..., X_n) + \kappa}{2}$$

$$= \frac{\text{mín}\,(X_1, ..., X_n) + \text{máx}\,(X_1, ..., X_n)}{2} + \kappa$$

$$= \hat{\theta}(X_1, ..., X_n) + \kappa$$

$$\neq \hat{\theta}(X_1, ..., X_n) = \frac{\text{mín}\,(X_1, ..., X_n) + \text{máx}\,(X_1, ..., X_n)}{2}$$

Luego no es invariante.

5. El coeficiente de correlación lineal si que es invariante frente a cambios de origen, en efecto:

$$\hat{\theta}(X_1 + \kappa, ..., X_n + \kappa, Y_1 + \kappa', ..., Y_n + \kappa') =$$

$$= \frac{\frac{1}{n} \sum_{i=1}^{n} [(X_i + \kappa) - (\bar{X} + \kappa)][(Y_i + \kappa') - (\bar{Y} + \kappa')]}{S_X S_Y}$$

$$= \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{S_X S_Y}$$

$$= \hat{\theta}(X_1, ..., X_n, Y_1, ..., Y_n)$$

2.6. ESTIMADOR ROBUSTO

Diremos que un procedimiento estadístico es **robusto** si su comportamiento es relativamente insensible a desviaciones de las hipótesis iniciales sobre las que se había planteado el procedimiento. En la última década ha sido muy significativo el interés mostrado por los investigadores sobre la robustez tanto de los procedimientos de estimación como de los contrastes de hipótesis.

Es frecuente considerar que una variable aleatoria X tiene una cierta función de distribución $F(x;\ \theta)$, siendo θ el parámetro que pretendemos estimar con el estimador $\hat{\theta}(X_1, ..., X_n)$, cuya correspondiente distribución muestral será $G(x;\ \hat{\theta})$. Pero en la realidad puede suceder que la función de distribución de la variable aleatoria X no sea la considerada sino que puede ser otra diferente $F_1(x;\ \theta)$ y en consecuencia la distribución muestral del estimador sería $G_1(x;\ \hat{\theta})$ distinta de la anterior. Entonces si la diferencia existente entre ambas distribuciones muestrales del estimador, $G(x;\ \hat{\theta})$ y $G_1(x;\ \hat{\theta})$, no son muy significativas y el procedimiento estadístico utilizado es insensible a estos cambios, se dice que este procedimiento estadístico es **robusto** y en consecuencia el estimador es robusto.

Definición 2.23. Estimador robusto.

> Diremos que un **estimador es robusto** cuando pequeños cambios en las hipótesis de partida del procedimiento de estimación considerado no producen variaciones significativas en los resultados obtenidos.

Por ejemplo, en una distribución $N(\mu,\ \sigma)$ al estudiar la distribución de la media muestral veíamos que si no se conoce la varianza poblacional recurríamos a la distribución t-Student, mediante el estadístico:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \to t_{n-1}$$

de manera que pequeñas variaciones en la distribución $N(\mu,\ \sigma)$ no producirán cambios sustanciales en los procedimientos estadísticos basados en el estadístico t-Student con $n - 1$ grados de libertad, cuando n es relativamente grande, ya que estos procedimientos estadísticos son robustos.

Capítulo 3
MÉTODOS DE OBTENCIÓN DE ESTIMADORES

3.1. INTRODUCCIÓN

En el capítulo anterior hemos estudiado las propiedades deseables de un buen estimador (insesgadez, consistencia, eficiencia, etc.), en el contexto de la estimación puntual, y ahora se nos presenta el problema de como obtener estimadores y además que sean buenos. Para ello, en este capítulo, daremos varios métodos de obtención de estimadores y veremos que propiedades cumplen los estimadores obtenidos por los diferentes métodos, de tal manera que la bondad o valía de un método de estimación se deduce de las propiedades que verifiquen los estimadores obtenidos por dicho método.

Los métodos que estudiaremos son:

- El método de los momentos.
- El método de la máxima verosimilitud.
- El método de la mínima χ^2.
- El método de los mínimos cuadrados.

3.2. EL MÉTODO DE LOS MOMENTOS

Fue introducido por K. Pearson y es el método general más antiguo y sencillo para la obtención de estimadores de parámetros poblacionales. En algunas ocasiones se suele utilizar para obtener una primera aproximación de los estimadores.

Este método consiste en igualar tantos momentos muestrales como parámetros haya que estimar, a los correspondientes momentos poblacionales, que son funciones de los parámetros desconocidos, y resolviendo el sistema de ecuaciones resultante tendríamos los estimadores de los parámetros[1].

Veamos ahora, de manera formal, como se desarrolla este método.

Sea una población con función de probabilidad $P(x_i; \theta_1, ..., \theta_k)$ o con función de densidad $f(x; \theta_1, ..., \theta_k)$, según que se trate de una distribución de tipo discreto o de tipo continuo, respectivamente, en las cuales aparecen k parámetros desconocidos que pretendemos estimar con la ayuda de una muestra aleatoria de tamaño n, $(X_1, ..., X_n)$. Designamos por $\alpha_1, ..., \alpha_k$ los k-primeros momentos respecto al origen de la población:

$$\alpha_j(\theta_1, ..., \theta_k) = \begin{cases} \displaystyle\sum_{i=1}^{\infty} x_i^j \, P(X = x_i) = \sum_{i=1}^{\infty} x_i^j \, P(x_i; \theta_1, ..., \theta_k), & \text{caso discreto} \\ & j = 1, ..., k, \\ \displaystyle\int_{-\infty}^{+\infty} x^j \, f(x; \theta_1, ..., \theta_k)\, dx, & \text{caso continuo} \end{cases}$$

En general α_j, será una función de los k-parámetros $\theta_1, ..., \theta_k$:

$$\alpha_j(\theta_1, ..., \theta_k), \qquad j = 1, ..., k$$

Ahora, consideremos la muestra aleatoria $(X_1, ..., X_n)$ de la población y calculemos los k-primeros momentos respecto al origen, $a_1, ..., a_j$ para estas observaciones muestrales, que son:

$$a_1 = \sum_{i=1}^{n} \frac{X_i}{n}, ..., \quad a_j = \sum_{i=1}^{n} \frac{X_i^j}{n}, ..., \quad a_k = \sum_{i=1}^{n} \frac{X_i^k}{n}$$

Igualando los k primeros momentos poblacionales, α_j, a los correspondientes momentos muestrales, a_j, tenemos un sistema de k ecuaciones con k-incógnitas $\theta_1, ..., \theta_k$:

$$\begin{aligned} \alpha_1(\theta_1, ..., \theta_k) &= a_1 \\ &\cdots\cdots\cdots \\ \alpha_j(\theta_1, ..., \theta_k) &= a_j \\ &\cdots\cdots\cdots \\ \alpha_k(\theta_1, ..., \theta_k) &= a_k \end{aligned} \qquad [3.1]$$

[1] K. Pearson, para desarrollar este método se basó en el **teorema de Khintchine,** el cual asegura, bajo condiciones muy generales, la convergencia en probabilidad de los momentos muestrales respecto a los correspondientes momentos poblacionales.

y resolviendo este sistema tendremos las soluciones:

$$\hat{\theta}_1, ..., \hat{\theta}_k$$

que son los estimadores de los parámetros $\theta_1, ..., \theta_k$.

3.2.1. PROPIEDADES DE LOS ESTIMADORES OBTENIDOS POR EL MÉTODO DE LOS MOMENTOS

I. **Insesgadez**

> Si los parámetros desconocidos y que pretendemos estimar son momentos poblacionales (la media de la distribución normal, el parámetro p de la distribución de Bernoulli, el parámetro λ de la distribución de Poisson, etc.), entonces los estimadores obtenidos por este método son **insesgados**.

Demostración:

Puesto que los parámetros a estimar son momentos poblacionales respecto al origen, α_j, tendremos para una muestra aleatoria $(X_1, ..., X_n)$ que:

$$\hat{\alpha}_j = a_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j, \qquad j = 1, ..., k$$

Tomando valores esperados resulta que:

$$\begin{aligned} E[\hat{\alpha}_j] &= E\left[\frac{1}{n} \sum_{i=1}^{n} \frac{X_i^j}{n}\right] \\ &= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i^j\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} E[X_i^j] \\ &= \frac{1}{n} \sum_{i=1}^{n} E[X^j] \\ &= \frac{1}{n} \cdot n\, E[X^j] \\ &= \alpha_j \end{aligned}$$

Luego vemos que son estimadores insesgados.

II. **Consistencia**

> Bajo condiciones bastante generales los estimadores obtenidos por este método son **consistentes.**

Aunque no haremos la demostración, daremos algún detalle más de esta propiedad; considerando el caso de dos parámetros, pero que se puede generalizar a k parámetros.

Sea una población que depende de dos parámetros desconocidos θ_1 y θ_2, entonces se demuestra[2] que:

«Los momentos muestrales a_1 y a_2 son estimadores consistentes de los respectivos momentos poblacionales α_1 y α_2».

III. **Normalidad asintótica**

> Si los parámetros desconocidos y que pretendemos estimar son los momentos poblacionales, entonces los estimadores obtenidos serán **asíntoticamente normales.**

Demostración:

Como los parámetros a estimar son los momentos poblacionales, α_j, que para una muestra aleatoria simple $(X_1, ..., X_n)$ son:

$$\hat{\alpha}_j = a_j = \sum_{i=1}^{n} \frac{X_i^j}{n}$$

$$= \frac{X_1^j}{n} + \dots + \frac{X_n^j}{n}$$

---

[2] Teniendo en cuenta el teorema de Khintchine.

También hemos de tener en cuenta el teorema de Slutsky, ya que entonces se puede demostrar que una función de los momentos muestrales es un estimador consistente de los momentos poblacionales.

Si a_1 y a_2 son los momentos muestrales, tales que convergen en probabilidad a los respectivos momentos poblacionales α_1 y α_2, es decir

$$a_1 \xrightarrow{P} \alpha_1$$

$$a_2 \xrightarrow{P} \alpha_2$$

y siendo $\hat{\theta}(a_1, a_2)$ una función continua en (α_1, α_2) entonces se cumple que $\hat{\theta}_1(a_1, a_2)$ es un estimador consistente de $\theta_1(\alpha_1, \alpha_2)$, es decir

$$\hat{\theta}_1(a_1, a_2) \xrightarrow{P} \theta_1(\alpha_1, \alpha_2)$$

Análogamente

$$\hat{\theta}_2(a_1, a_2) \xrightarrow{P} \theta_2(\alpha_1, \alpha_2)$$

resultando que el estimador $\hat{\alpha}_j = a_j$ se puede expresar como suma de n variables aleatorias $\frac{X_i^j}{n}$, independientes e idénticamente distribuidas con media y varianza:

$$E\left[\frac{X_i^j}{n}\right] = \frac{\alpha_j}{n}$$

$$\begin{aligned}\operatorname{Var}\left(\frac{X_i^j}{n}\right) &= \frac{1}{n^2}\operatorname{Var}(X_i^j)\\ &= \frac{1}{n^2}E[(X_i^j - E[X_i^j])^2]\\ &= \frac{1}{n^2}E[(X_i^j - \alpha_j)^2]\\ &= \frac{1}{n^2}(\alpha_{2j} - \alpha_j^2)\end{aligned}$$

y la media y la varianza del estimador $\hat{\alpha}_j = a_j$, será:

$$E[\hat{\alpha}_j] = E[a_j] = E\left[\sum_{i=1}^{n}\frac{X_i^j}{n}\right] = \alpha_j$$

$$\begin{aligned}\operatorname{Var}(\hat{\alpha}_j) = \operatorname{Var}(a_j) &= \operatorname{Var}\left(\sum_{i=1}^{n}\frac{X_i^j}{n}\right)\\ &= \sum_{i=1}^{n}\operatorname{Var}\left(\frac{X_i^j}{n}\right) = \frac{\alpha_{2j} - \alpha_j^2}{n}\end{aligned}$$

Luego aplicando el Teorema Central del Límite, para muestras suficientemente grandes, tenemos que el estimador $\hat{\alpha}_j = a_j$ sigue una distribución

$$\hat{\alpha}_j = a_j \to N\left(\alpha_j, \sqrt{\frac{\alpha_{2j} - \alpha_j^2}{n}}\right) \qquad [3.2]$$

o bien que la variable aleatoria

$$\frac{a_j - E[a_j]}{\sqrt{\operatorname{Var}(a_j)}} = \frac{a_j - \alpha_j}{\sqrt{\dfrac{\alpha_{2j} - \alpha_j^2}{n}}} = \frac{\sqrt{n}(a_j - \alpha_j)}{\sqrt{\alpha_{2j} - \alpha_j^2}} \xrightarrow[n \to \infty]{} N(0, 1)$$

En resumen, podemos decir que, en condiciones bastante generales, estos estimadores son:

- Consistentes.
- Asíntoticamente normales.
- En general, no son insesgados, y por tanto no son eficientes.

Fisher, estudiando estos estimadores, observó que no daban un resultado satisfactorio desde el punto de vista de la eficiencia, por lo que él era partidario de sustituir este método por el método de máxima verosimilitud, siempre que los cálculos, a que este nuevo método diese lugar, no fueran de gran dificultad. No obstante y debido a su facilidad práctica, estos estimadores se emplean como una primera aproximación, a partir de la cual, es posible, utilizando otros métodos para obtener estimadores de mayor eficiencia.

En general este método no suele proporcionar buenos estimadores, pues como hemos visto, no utiliza la distribución de la población, sino que solamente se basa en sus momentos y en consecuencia no aprovecha toda la información contenida en la muestra, sin embargo el método de la máxima verosimilitud sí que tiene en cuenta la distribución de la población, como veremos después.

Ejemplo 3.1

Sea $(X_1, ..., X_n)$ una muestra aleatoria obtenida de una población que sigue una distribución de Poisson de parámetro λ, desconocido. Obtener un estimador del parámetro λ utilizando el método de los momentos.

Solución:

Aplicando el método de los momentos igualaremos el momento de orden uno, respecto al origen, de la población α_1, al momento de orden uno de la muestra a_1.

$$\alpha_1(\lambda) = E[X] = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i)$$

$$= \sum_{i=0}^{\infty} x_i \cdot \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$= e^{-\lambda} \lambda \sum_{i=0}^{\infty} \frac{\lambda^{x_{i-1}}}{(x_{i-1})!}$$

$$= e^{-\lambda} \lambda e^{\lambda}$$

$$= \lambda$$

$$a_1 = \sum_{i=1}^{n} \frac{X_i}{n} = \bar{X}$$

Luego igualando

$$\alpha_1(\lambda) = a_1$$

resulta que el estimador por el método de los momentos de λ es:

$$\hat{\lambda} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Este estimador coincide con el que se obtiene por el método de máxima verosimilitud.

Ejemplo 3.2

Sea $(X_1, ..., X_n)$ una muestra aleatoria procedente de una $B(1, p)$. Obtener el estimador del parámetro p, utilizando el método de los momentos.

Solución:

Sabemos de la distribución $B(1, p)$ que la media o momento de orden uno respecto al origen es:

$$\begin{aligned} \alpha_1(p) = E[X] &= \sum_i x_i \cdot P(X = x_i) \\ &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ &= 0 \cdot (1 - p) + 1 \cdot p \\ &= p \end{aligned}$$

y el momento de orden uno de la muestra es:

$$a_1 = \sum_{i=1}^{n} \frac{X_i}{n}$$

Luego igualando ambos momentos resulta:

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

y si hacemos $X = \sum_{i=1}^{n} X_i \equiv$ número de éxitos en las n pruebas:

$$\hat{p} = \frac{X}{n}$$

Este estimador, como veremos después, es también el estimador obtenido por el método de la máxima verosimilitud.

Ejemplo 3.3

Sea $(X_1, ..., X_n)$ una muestra aleatoria procedente de una población con distribución $\Gamma(p, a)$. Obtener los estimadores de p y de a utilizando el método de los momentos.

Solución:

Sabemos que el momento de orden r respecto al origen en la $\Gamma(p, a)$ viene dado por:

$$\alpha_r = E[X^r] = \frac{\Gamma(p + r)}{a^r \; \Gamma(p)}$$

Luego los dos primeros momentos de la población, respecto al origen serán:

$$\alpha_1 = E[X] = \frac{\Gamma(p + 1)}{a \; \Gamma(p)} = \frac{p}{a}$$

$$\alpha_2 = E[X^2] = \frac{\Gamma(p + 2)}{a^2 \; \Gamma(p)} = \frac{(p + 1)p}{a^2}$$

y los dos primeros momentos muestrales son:

$$a_1 = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$a_2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

Igualando ambos momentos tenemos el sistema:

$$\left.\begin{aligned} \alpha_1(p, a) &= a_1 \\ \alpha_2(p, a) &= a_2 \end{aligned}\right\} \Rightarrow \left.\begin{aligned} \frac{p}{a} &= \frac{\sum_{i=1}^{n} X_i}{n} = a_1 \\ \frac{(p+1)p}{a^2} &= \frac{\sum_{i=1}^{n} X_i^2}{n} = a_2 \end{aligned}\right\}$$

Resolviendo el sistema para p y a, pero utilizando previamente a_1 y a_2, tenemos:

$$\hat{p} = \frac{a_1^2}{a_2 - a_1^2} = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}$$

$$= \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{a} = \frac{a_1}{a_2 - a_1^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}$$

$$= \frac{\bar{X}}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

que son los estimadores de p y de a obtenidos por el método de los momentos.

Ejemplo 3.4

Sea una muestra aleatoria formada por las observaciones (1,2; 2,6; 4,4; 3,4; 0,6; 2,2) procedente de una población cuya función de densidad es:

$$f(x) = \begin{cases} \theta^{-1} & , \quad 0 < x < \theta \\ 0 & , \quad \text{en el resto} \end{cases}$$

Estimar el parámetro θ por el método de los momentos.

Solución:

Para aplicar el método de los momentos tendremos que calcular los momentos de orden uno, respecto al origen, tanto para la población como para la muestra e igualarlos; con lo cual tendremos:

$$\begin{aligned} \alpha_1 = E[X] &= \int_0^\theta x \cdot \theta^{-1}\, dx \\ &= \left[\frac{x^2}{2\theta}\right]_0^\theta = \frac{\theta^2}{2\theta} \\ &= \frac{\theta}{2} \end{aligned}$$

$$\begin{aligned} a_1 = \frac{\sum_{i=1}^{n} X_i}{n} &= \frac{1,2 + 2,6 + 4,4 + 3,4 + 0,6 + 2,2}{6} \\ &= \frac{14,4}{6} \\ &= 2,4 \end{aligned}$$

Luego resolviendod la ecuación:

$$\frac{\theta}{2} = 2,4$$

tendremos el estimador $\hat{\theta}$ del parámetro θ por el método de los momentos, que será:

$$\hat{\theta} = 4,8$$

Ejemplo 3.5

Sea una población cuya distribución de probabilidad viene dada por

$$P(X = x)\begin{cases} \dfrac{1}{2}(1-\theta) & x = -1 \\ \dfrac{1}{2} & x = 0 \\ \dfrac{1}{2}\theta & x = 1 \end{cases}$$

en donde $0 < \theta < 1$.

Utilizando una muestra aleatoria simple $(X_1, ..., X_n)$.

1. Obtener un estimador del parámetro θ por el método de los momentos.
2. Comprobar si es insesgado.
3. Comprobar si es consistente.

Solución:

1. El momento de orden uno respecto al origen en la población es:

$$\begin{aligned} \alpha_1 = E[X] &= \sum_{i=1}^{3} x_i\, P(X = x_i) \\ &= (-1)\cdot\frac{1}{2}(1-\theta) + 0\cdot\frac{1}{2} + 1\cdot\frac{1}{2}\theta \\ &= \theta - \frac{1}{2} \end{aligned}$$

El momento muestral de orden uno será:

$$a_1 = \frac{\sum_{i=1}^{n} X_i}{n}$$

Igualando ambos momentos tenemos:

$$\theta - \frac{1}{2} = a_1 = \frac{\sum_{i=1}^{n} X_i}{n}$$

luego

$$\hat{\theta} = a_1 + \frac{1}{2} = \frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{2}$$

es el estimador obtenido por el método de los momentos.

2. Veamos si es insesgado

$$E[\hat{\theta}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{2}\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[X_i] + \frac{1}{2}$$

$$= \frac{1}{n} \cdot n\left(\theta - \frac{1}{2}\right) + \frac{1}{2}$$

$$= \theta$$

Luego en este caso el estimador $\hat{\theta}$ obtenido por el método de los momentos es insesgado.

3. Para ver si es consistente, tendremos en cuenta la definición 2.8 y la expresión [2.18]. Así pues probaremos que

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$$

o bien

$$P[|\hat{\theta}_n - \theta| < \varepsilon] \underset{n \to \infty}{\to} 1$$

y como el estimador es insesgado

$$\theta = E[\hat{\theta}]$$

nos queda:

$$P[|\hat{\theta} - E[\hat{\theta}]| < \varepsilon] \xrightarrow[n \to \infty]{} 1$$

Teniendo en cuenta una de las expresiones de la desigualdad de Chebychev, tenemos:

$$P[|\hat{\theta} - E[\hat{\theta}]| < \varepsilon] \geqslant 1 - \frac{\operatorname{Var}(\hat{\theta})}{\varepsilon^2}$$

Ahora bien

$$\operatorname{Var}(\hat{\theta}) = \operatorname{Var}\left(a_1 + \frac{1}{2}\right)$$

$$= \operatorname{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{2}\right)$$

$$= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n}\operatorname{Var}(X)$$

Sustituyendo en la expresión de la desigualdad de Chebycheu tenemos:

$$P[|\hat{\theta} - E[\hat{\theta}]| < \varepsilon] \geqslant 1 - \frac{\operatorname{Var}(X)}{n\ \varepsilon^2} \xrightarrow[n \to \infty]{} 1$$

ya que $\operatorname{Var}(X)$, al ser un valor fijo, no depende de n.

Luego

$$P[|\hat{\theta} - \theta| < \varepsilon] \underset{n \to \infty}{\to} 1$$

y por tanto el estimador $\hat{\theta}$ es consistente.

3.3. MÉTODO DE LA MÁXIMA VEROSIMILITUD

Es desde el punto de vista teórico, el método general de estimación más conocido. Este método ya fue utilizado por Gauss, en casos particulares, pero como método de estimación fue introducido por Fisher 1922, siendo muy importantes las contribuciones realizadas por otros autores en su desarrollo posterior.

Consideramos una muestra aleatoria simple $(X_1, ..., X_n)$ procedente de una población con función de probabilidad $P(x_i; \theta)$ o con función de densidad $f(x; \theta)$, donde θ es un parámetro desconocido que toma valores en el **espacio paramétrico** Ω, $\theta \in \Omega$.[3]

Para una muestra aleatoria $(X_1, ..., X_n)$ la función de cuantía o la función de densidad conjunta de una muestra aleatoria la indicaremos por:

$$f(x_1, ..., x_n; \theta) = f(x_1; \theta), ..., f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Definición 3.1. Función de verosimilitud.

> Definimos la **función de verosimilitud** de n variables aletorias como la función de probabilidad o la función de densidad conjunta de las n-variables.
>
> $$L(\mathbf{x}; \theta) = L(x_1, ..., x_n; \theta) = f(x_1, ..., x_n; \theta)$$
>
> Para una muestra aleatoria simple $(X_1, ..., X_n)$, al ser independientes las observaciones, la **función de verosimilitud** quedará como:
>
> $$L(\mathbf{x}; \theta) = L(x_1, ..., x_n; \theta) = f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) \qquad [3.3]$$

Vemos que la función de verosimilitud $L(x_1, ..., x_n; \theta)$ es función de la muestra observada y por tanto será una función aleatoria dependiente del parámetro θ, pues para cada muestra aleatoria tomará un valor.

El valor que toma la función de verosimilitud para una muestra dada y concreta $(x_1, ..., x_n)$ recibe el nombre de **elemento de verosimilitud** o **verosimilitud** de la muestra:

$$L(\mathbf{x}; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

y sólo depende del parámetro θ, ya que $(x_1, ..., x_n)$ son valores concretos.

---

[3] Notaremos, indistintamente, la función de probabilidad o la función de densidad por $f(x; \theta)$ de manera general, si bien cuando estemos en el caso discreto la indicaremos por $P(X = x; \theta)$ o bien $P(x; \theta)$.

Así pues cuando digamos que estamos en el caso discreto la función de verosimilitud de una muestra aleatoria será:

$$L(\mathbf{x}; \theta) = L(x_1, ..., x_n; \theta) = P(x_1, ..., x_n; \theta) = P(X_1 = x_1, ..., X_n = x_n; \theta) = \prod_{i=1}^{n} P(x_i; \theta)$$

Antes de exponer el método de la máxima verosimilitud, de manera general, veamos un ejemplo que nos ilustrará el fundamento a seguir en el método.

Ejemplo 3.6

Sea una urna que contiene bolas blancas y negras, y designamos por p la probabilidad de extraer bola blanca cuando se realiza una extracción al azar. Asociado a este experimento aleatorio tenemos la variable aleatoria X que puede tomar los valores:

$X = 1$: si la bola extraída es blanca

$X = 0$: si la bola extraída es negra

y la correspondiente distribución de probabilidad será una $B(1; p)$

$$P(X = x) = p^x(1 - p)^{1-x}$$

Seleccionamos una muestra aleatoria (con reemplazamiento) de tamaño cuatro (X_1, X_2, X_3, X_4), siendo X_i la variable aleatoria asociada a la extracción i-esima, y suponemos que ha resultado la siguiente realización (B, B, N, B).

Como el parámetro p es desconocido, pretendemos saber, entre los valores, $p = 0{,}65$, $p = 0{,}73$ y $p = 0{,}82$ qué valor hace más probable la aparición de la muestra (B, B, N, B).

Solución:

Como la muestra seleccionada ha resultado ser:

$$(B, B, N, B)$$

y la selección es aleatoria simple, es decir las extracciones son independientes, entonces la probabilidadd de aparición de esta muestra será:

$$\begin{aligned} P(B, B, N, B) &= P(B \cap B \cap N \cap B) \\ &= P(B) \cdot P(B) \cdot P(N) \cdot P(B) \\ &= p \cdot p \cdot (1 - p) \cdot p \\ &= p^3(1 - p) \end{aligned}$$

Pero como nos interesa saber, cuál de los tres valores de p considerados hace más probable la aparición de la muestra (B, B, N, B), tendremos que calcular la $P(B, B, N, B)$ para $p = 0{,}65$, $p = 0{,}73$ y $p = 0{,}82$:

para $p = 0{,}65$, $\quad P(B, B, N, B; p) = (0{,}65)^3 \cdot (0{,}35) = 0{,}0961$

para $p = 0{,}73$, $\quad P(B, B, N, B; p) = (0{,}73)^3 \cdot (0{,}27) = 0{,}1050$

para $p = 0{,}82$, $\quad P(B, B, N, B; p) = (0{,}82)^3 \cdot (0{,}18) = 0{,}0992$

Lo cual nos dice que la aparición de la muestra (B, B, N, B) es más probable cuando el parámetro poblacional $p = 0{,}73$ que para los otros dos valores, con lo cual admitimos que la población de partida es $B(1; 0{,}73)$, con más seguridad que $p = 0{,}65$ o $p = 0{,}82$. Además, observamos que este resultado está de acuerdo con el sentido común, pues si $p = 0{,}73$, nos dice que aproximadamente casi 3/4 de bolas son blancas y algo más de 1/4 son negras, siendo por tanto esta composición (este valor de p) la que hace más verosimil la aparición de la muestra (B, B, N, B), entre las consideradas.

Si ahora consideramos la muestra aleatoria simple (X_1, X_2, X_3, X_4), como las variables aleatorias X_i son independientes y toman los valores 0 o 1 con distribución $B(1; p)$, resulta que la distribución de probabilidad asociada a cada X_i serán:

$$
\begin{aligned}
P(x_1; p) &= P(X = x_1) = p^{x_1}(1 - p)^{1 - x_1} \quad ; \quad x_1 = 0, 1 \\
P(x_2; p) &= P(X = x_2) = p^{x_2}(1 - p)^{1 - x_2} \quad ; \quad x_2 = 0, 1 \\
P(x_3; p) &= P(X = x_3) = p^{x_3}(1 - p)^{1 - x_3} \quad ; \quad x_3 = 0, 1 \\
P(x_4; p) &= P(X = x_4) = p^{x_4}(1 - p)^{1 - x_4} \quad ; \quad x_4 = 0, 1
\end{aligned}
$$

y la función de verosimilitud será:

$$
\begin{aligned}
L(x_1, x_2, x_3, x_4; p) &= \prod_{i=1}^{4} P(x_i; p) \\
&= p^{x_1}(1 - p)^{1 - x_1} \cdot p^{x_2}(1 - p)^{1 - x_2} \cdot p^{x_3}(1 - p)^{1 - x_3} \cdot p^{x_4}(1 - p:^{1 - x_4} \\
&= p^{x_1 + x_2 + x_3 + x_4}(1 - p)^{4 - (x_1 + x_2 + x_3 + x_4)}
\end{aligned}
$$

$$x_i = 0{,}1 \quad ; \quad i = 1, 2, 3, 4$$

Para la muestra (B, B, N, B) el valor que toma la función de verosimilitud, es decir el elemento de verosimilitud de p, será:

$$
\begin{aligned}
L(x_1, x_2, x_3, x_4; p) &= p^{1+1+0+1}(1 - p)^{4-(1+1+0+1)} \\
&= p^3(1 - p)
\end{aligned}
$$

y hemos elegido como estimación del parámetro p aquel valor (entre los tres que estabamos considerando) que hace máximo el elemento de verosimilitud o simplemente la verosimilitud de la muestra (B, B, N, B).

Por tanto, en general poemos dar la siguiente definición:

Definición 3.2. Método de la máxima verosimilitud.

El método de la **máxima verosimilitud** consiste en elegir como estimador del parámetro desconocido θ aquel valor $\hat{\theta}(X_1, ..., X_n)$ que hace máxima la función de verosimilitud $L(x_1, ..., x_n; \theta)$. Es decir, consiste en encontrar aquel valor $\hat{\theta}(X_1, ..., X_n)$ tal que

$$L(x_1, ..., x_n; \hat{\theta}) = \max_{\theta \in \Omega} L(x_1, ..., x_n; \theta) \qquad [3.4]$$

A este estimador $\hat{\theta}(X_1, ..., X_n)$ se le llama **estimador máximo-verosímil** o **estimador de máxima verosimilitud** (EMV) **del parámetro θ**.

Continuando con la interpretación instuitiva del ejemplo 3.6, y si consideramos sólo el caso discreto, vemos que la función de verosimilitud de la muestra será:

$$L(x_1, ..., x_n; \theta) = P(X_1 = x_1, ..., X_n = x_n; \theta) = \prod_{i=1}^{n} P(x_i; \theta) \qquad [3.5]$$

y para una muestra concreta esta expresión dependerá sólo de θ, por eso también podríamos haberla notado poniendo $L(\theta)$. Entonces el método de la máxima verosimilitud lo que hace es elegir aquel valor del parámetro θ para el cual la expresión [3.5] es máxima para la muestra en cuestión, lo cual equivale a que la muestra considerada es la más probable, como sucedía en el ejemplo 3.6, y además coincide con el comportamiento lógico, siendo ese valor del parámetro θ el que se hace más verosímil con la aparición de la muestra considerada.

En resumen el valor de la función de verosimilitud $L(x_1, ..., x_n; \theta)$ para una muestra concreta nos da la **verosimilitud** o plausibilidad dę que el parámetro θ tome un cierto valor, tomando como información la proporcionada por la muestra. Así pues si $L(x_1, ..., x_n; \theta_1) > L(x_1, ..., x_n; \theta_2)$ esto nos indica que la verosimilitud de que el parámetro θ tome el valor θ_1, es mayor que la verosimilitud de que el parámetro tome el valor θ_2, dado que se ha obtenido la muestra considerada.

El razonamiento en el caso continuo es igual. A partir de ahora para todas las consideraciones teóricas que haremos nos referiremos al caso continuo, salvo que hagamos la especificación expresa del caso discreto.

Hemos dicho que el estimador de máxima verosimilitud viene dado por el valor $\hat{\theta}$ tal que:

$$L(x_1, ..., x_n; \hat{\theta}) = \max_{\theta \in \Omega} L(x_1, ..., x_n; \theta)$$

Pero con frecuencia la función de verosimilitud $L(x_1, ..., x_n; \theta)$ suele ser complicada, y al ser esta función positiva y coincidir los máximos de $L(x_1, ..., x_n; \theta)$ con los de la función $\ln L(x_1, ..., x_n; \theta)$[4], entonces lo que se hace es considerar la función:

$$\ln L(x_1, ..., x_n; \theta) = \ln f(x_1, ..., x_n; \theta) = \sum_{i=1}^{n} \ln f(x_i; \theta) \qquad [3.6]$$

y el **estimador de máxima verosimilitud**, $\hat{\theta}$, será el que verifique la expresión:

$$\ln L(x_1, ..., x_n; \hat{\theta}) = \max_{\theta \in \Omega} \ln L(x_1, ..., x_n; \theta) = \max_{\theta \in \Omega} \sum_{i=1}^{n} \ln f(x_i; \theta) \qquad [3.7]$$

que vendrá dado por la solución de la ecuación de verosimilitud[5]:

$$\frac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i; \theta)}{\partial \theta} = 0 \qquad [3.8]$$

este estimador $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ será función de las observaciones muestrales, y prescindimos de aquellas soluciones que den lugar a que el estimador fuera igual a una constante.

Si la función de densidad o de cuantía de la población depende de k parámetros, $f(x; \theta_1, ..., \theta_k)$, entonces los estimadores máximo-verosimiles de estos

[4] Pues como la función ln es una función monótona creciente, ambas funciones L y $\ln L$ tomarán sus máximos en el mismo punto.

[5] Admitimos las siguientes **condiciones de regularidad:** que el campo de variación de θ es un intervalo abierto del eje real, que el campo de variación de la variable aleatoria poblacional no depende de θ, que $f(x, \theta)$ es positiva y derivable respecto a θ y que se verifica la condición de máximo $\left.\frac{\partial^2 \ln L}{\partial \theta}\right|_{\theta < \hat{\theta}} < 0$

parámetros se obtienen resolviendo el sistema de ecuaciones de verosimilitud en $\theta_1, ..., \theta_k$.

$$\frac{\partial \ln L(x_1, ..., x_n; \theta_1, ..., \theta_k)}{\partial \theta_1} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i; \theta_1, ..., \theta_k)}{\partial \theta_1} = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad [3.9]$$

$$\frac{\partial \ln L(x_1, ..., x_n; \theta_1, ..., \theta_k)}{\partial \theta_k} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i; \theta_1, ... \theta_k)}{\partial \theta_k} = 0$$

y tendríamos:

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, ..., X_n)$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$\hat{\theta}_k = \hat{\theta}_k(X_1, ..., X_n)$$

que serían los estimadores de máxima verosimilitud de los parámetros $(\theta_1, ..., \theta_k)$.

Cualquier solución no trivial de las ecuaciones [3.8] o [3.9] será un estimador de máxima verosimilitud. Ahora bien si la solución es única diremos que se trata de un **estimador de máxima verosimilitud en sentido estricto,** dando lugar al máximo absoluto de la función de verosimilitud. Sin embargo, cuando hay más de una solución (no incluimos la trivial) entonces diremos que tenemos **estimadores de máxima verosimilitud en sentido amplio.**

Generalmente la ecuación o sistema de ecuaciones de verosimilitud se puede resolver sin grandes dificultades, no obstante en algunas ocasiones hay que recurrir a métodos iterativos de cálculo numérico.

Ejemplo 3.7

Sea una población distribuida según una $B(10, p)$. Obtener el estimador de máxima verosimilitud utilizando una muestra aleatoria (X_1, X_2, X_3, X_4).

Solución:

Obtendremos el estimador de máxima verosimilitud para el parámetro p, resolviendo la ecuación:

$$\frac{\partial \ln L(x_1, x_2, x_3, x_4; p)}{\partial p} = 0$$

y para ello calculamos:

$$L(x_1, x_2, x_3, x_4; p) = \prod_{i=1}^{4} P(x_i; p)$$

$$= \binom{10}{x_1} p^{x_1}(1-p)^{10-x_1} \cdots\cdots\cdots \binom{10}{x_4} p^{x_4}(1-p)^{10-x_4}$$

$$= \prod_{i=1}^{4} \binom{10}{x_i} p^{\sum_{i=1}^{4} x_i}(1-p)^{40-\sum_{i=1}^{4} x_i}$$

$$\ln L(x_1, x_2, x_3, x_4; p) = \sum_{i=1}^{4} \ln \binom{10}{x_i} + \sum_{i=1}^{4} x_i (\ln p) + (40 - \sum_{i=1}^{4} x_i) \ln (1-p)$$

$$\frac{\partial \ln L(x_1, x_2, x_3, x_4; p)}{\partial p} = \frac{\sum_{i=1}^{4} x_i}{p} - \frac{40 - \sum_{i=1}^{4} x_i}{1-p} = 0$$

$$(1-p) \sum_{i=1}^{4} x_i - p(40 - \sum_{i=1}^{4} x_i) = 0$$

$$\sum_{i=1}^{4} x_i - 40p = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{4} x_i}{40}$$

que será el estimador de máxima verosimilitud, o lo que es lo mismo, es el valor del parámetro p que hace máxima la función de verosimilitud para esta muestra concreta. Pero como para cualquier otra muestra llegaríamos al mismo tipo de estimación, entonces podemos considerar que el estimador será:

$$\hat{p} = \frac{\sum_{i=1}^{4} X_i}{40}$$

Ejemplo 3.8

Sea $(X_1, ..., X_n)$ una muestra aleatoria simple procedente de una población $B(1, p)$, en donde p es desconocido. Obtener el estimador de máxima verosimilitud del parámetro p.

Solución:

Sabemos que la función de probabilidad es:

$$P(x_i; p) = p^{x_i}(1 - p)^{1-x_i}, \quad x_i = 0{,}1, \quad i = 1, ..., n$$

La función de verosimilitud es:

$$L(x_1, ..., x_n; p) = P(x_1, ..., x_n; p) = \prod_{i=1}^{n} P(x_i; p)$$

$$= p^{\sum_{i=1}^{n} x_i}(1 - p)^{n - \sum_{i=1}^{n} x_i}$$

El $\ln L$ viene dado por:

$$\ln L(x_1, ..., x_n; p) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln (1 - p)$$

$$\frac{\partial \ln L(x_1, ..., x_n; p)}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1 - p} = \frac{\sum_{i=1}^{n} x_i - np}{p(1 - p)} = 0$$

$$\sum_{i=1}^{n} x_i - np = 0 \quad \Rightarrow \quad \hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{X}{n} = \bar{x}$$

Calculando la $\dfrac{\partial^2 \ln L}{\partial p^2}$ tenemos:

$$\frac{\partial^2 \ln L(x_1, ..., x_n; p)}{\partial p^2} = \frac{-\sum_{i=1}^{n} x_i}{p^2} - \frac{n - \sum_{i=1}^{n} x_i}{1 - p^2}$$

$$= \frac{-(1 - p)^2 \sum_{i=1}^{n} x_i - \left(n - \sum_{i=1}^{n} x_i\right)p^2}{p^2(1 - p)^2}$$

y particularmente para $p = \bar{x}$, se tiene:

$$\frac{\partial^2 \ln L(x_1, \ldots, x_n; p)}{\partial p^2} = -\left(\frac{n}{\bar{x}} + \frac{n}{1-\bar{x}}\right) < 0$$

con lo cual podemos decir que se trata de un máximo. Luego el estimador de máxima verosimilitud es

$$\hat{p} = \bar{x} = \frac{X}{n}$$

Ejemplo 3.9

Sea una población $N(20, \sigma)$, en donde σ es desconocida. Con la ayuda de una muestra aleatoria de tamaño n, obtener:

1. El estimador de máxima verosimilitud de σ^2.
2. El estimador de máxima verosimilitud de σ^2 para $n = 30$ y $\sum_{i=1}^{30} (x_i - 20)^2 = 3.000$

Solución:

1. Tenemos que resolver la ecuación:

$$\frac{\partial \ln L(x_1, \ldots, x_n; \sigma^2)}{\partial \sigma^2}$$

y para ello calculamos:

$$\begin{aligned} L(x_1, \ldots, x_n; \sigma^2) &= \prod_{i=1}^{n} f(x_i; \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-20)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-20)^2}{2\sigma^2}} \end{aligned}$$

$$\ln L(x_1, \ldots, x_n; \sigma^2) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{\sum_{i=1}^{n}(x_i-20)^2}{2\sigma^2}$$

$$\frac{\partial \ln L(x_1, \ldots, x_n; \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-20)^2}{2\sigma^4} = 0$$

de donde se tiene

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - 20)^2}{n}$$

pudiendo comprobarse que es un máximo y por tanto será el estimador de máxima verosimilitud[6].

2. Con la información complementaria que tenemos, el estimador de máxima verosimilitud será:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{30} (x_i - 20)^2}{30} = \frac{3.000}{30} = 100$$

Observemos que no se trata de una varianza muestral pues el valor $\mu = 20$ se refiere a la media de la población y no a la media de la muestra.

Ejemplo 3.10

Una compañía de seguros, después de analizar su fichero de siniestros sobre roturas de lunas de establecimientos comerciales, llega a la conclusión de que el número de siniestros mensuales se ajusta a una distribución de Poisson. Tomando una muestra aleatoria de 8 meses, se observó que se produjeron 310 siniestros. Obtener una estimación máximo-verosímil del parámetro λ.

---

[6] Observemos que el estimador de máxima verosimilitud de la varianza poblacional en una $N(\mu, \sigma)$ no es la varianza muestral sino que es:

$$S^{*2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

y si la media poblacional no es conocida entonces el estimador de máxima verosimilitud de la varianza σ^2 será:

$$S'^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Se comprueba que $E[S^{*2}] = \sigma^2$. En efecto:

$$E[S^{*2}] = \frac{1}{n} E\left[\sum_{i=1}^{n} (X_i - \mu)^2 \right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i - \mu]^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2$$

Solución:

La función de probabilidad de una distribución de Poisson de parámetro λ es:

$$P(x;\, \lambda) = \frac{\lambda^x}{x!}\, e^{-\lambda} \quad , \quad \lambda > 0$$

La función de verosimilitud para la muestra de tamaño $n = 8$, es

$$\begin{aligned} L(x_1, \,\dots,\, x_8;\, \lambda) &= \prod_{i=1}^{8} P(x_i;\, \lambda) \\ &= e^{-8\lambda} \prod_{i=1}^{8} \frac{\lambda^{x_i}}{x_i!} \end{aligned}$$

$$\ln L(x_1, \,\dots,\, x_8;\, \lambda) = -8\,\lambda + \ln \lambda \sum_{i=1}^{8} x_i - \sum_{i=1}^{8} \ln\,(x_i!)$$

$$\frac{\partial \ln L(x_1, \,\dots,\, x_8;\, \lambda)}{\partial \lambda} = -8 + \frac{\sum_{i=1}^{8} x_i}{\lambda} = 0$$

Luego la estimación de máxima verosimilitud es:

$$\hat{\lambda} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{310}{8} = 38{,}7$$

$$\hat{\lambda} = \bar{x}$$

En general en una distribución de Poisson $P(\lambda)$, se observa que el estimador máximo-verosímil del parámetro λ es:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

y se comprueba que efectivamente se verifica la condición de máximo, pues:

$$\left.\frac{\partial^2 \ln L(x_1, \,\dots,\, x_n;\, \lambda)}{\partial \lambda^2}\right|_{\hat{\lambda} = \bar{x}} = -\frac{\sum_{i=1}^{n} x_i}{\bar{x}^2} = -\frac{n}{\bar{x}} < 0$$

Ejemplo 3.11

Suponiendo que la cotización de una determinada acción se distribuye según una $N(\mu, \sigma)$, seleccionamos una muestra aleatoria de 20 días de cotización de esa acción, obteniendo que

$$\sum_{i=1}^{20} x_i = 35.700 \text{ ptas.}; \qquad \sum_{i=1}^{20} (x_i - \bar{x})^2 = 40.500$$

Obtener estimadores máximo-verosimiles para μ y σ, y sus correspondientes estimaciones para la muestra dada.

Solución:

Como se trata de una población $N(\mu, \sigma)$, la función de densidad es:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La función de verosimilitud para la muestra de tamaño n es:

$$L(x_1, ..., x_n; \mu, \sigma) = f(x_1, ..., x_n; \mu, \sigma) = \prod_{i=1}^{n} f(x_i; \mu, \sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\ln L(x_1, ..., x_n; \mu, \sigma) = -n \ln \sigma - \frac{n}{2} \ln (2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Derivando respecto a los dos parámetros μ y σ e igualando a cero, se tiene el siguiente sistema de dos ecuaciones:

$$\frac{\partial \ln L(x_1, ..., x_n; \mu, \sigma)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{\partial \ln L(x_1, ..., x_n; \mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

De la primera ecuación tenemos:

$$\sum_{i=1}^{n} (x_i - \mu) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} x_i = n\,\mu \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n} \sum_{i=1} x_i = \bar{x}$$

y sustituyendo en la segunda se tiene:

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0 \quad \Rightarrow \quad n\,\sigma^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$ [7]

Utilizando la información que nos proporciona la muestra resulta que las estimaciones máximo-verosimiles de los parámetros μ y σ son:

$$\hat{\mu} = \frac{\sum_{i=1}^{20} x_i}{20} = \frac{35.700}{20} = 1.785 \text{ ptas.}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{20} (x_i - \bar{x})^2}{20} = \frac{40.500}{20} = 2.025 \text{ ptas.}$$

$$\hat{\sigma} = \pm\, 45 \text{ ptas.}$$

Ejemplo 3.12

Sea $(X_1, ..., X_n)$ una muestra aleatoria de una población que se distribuye según una $\Gamma(p, a)$, con ambos parámetros desconocidos. Obtener los estimadores máximo-verosimiles.

Solución:

La función de densidad de la distribución $\Gamma(p, a)$ es:

$$f(x; p, a) = \frac{a^p}{\Gamma(p)}\, x^{p-1}\, e^{-ax} \quad , \quad x > 0, \quad a > 0, \quad p > 0$$

[7] Luego vemos que el estimador de máxima verosimilitud de la varianza poblacional no es la varianza muestral S^2, sino que hay que dividir por n, en lugar de por $n - 1$.

La función de verosimilitud viene dada por:

$$L(x_1, ..., x_n; p, a) = \prod_{i=1}^{n} f(x_i; p, a)$$

$$= \frac{a^p}{\Gamma(p)} x_1^{p-1} e^{-ax_1} \ldots\ldots \frac{a^p}{\Gamma(p)} x_n^{p-1} e^{-ax_n}$$

$$= \frac{a^{np}}{[\Gamma(p)]^n} (x_1, ..., x_n)^{p-1} e^{-a \sum_{i=1}^{n} x_i}$$

Tomando ln en la función de verosimilitud tenemos:

$$\ln L(x_1, ..., x_n; p, a) = n\,p \ln a - n \ln \Gamma(p) + (p-1) \sum_{i=1}^{n} \ln x_i - a \sum_{i=1}^{n} x_i$$

Derivando respecto de p y de a e igualando a cero, obtenemos las ecuaciones de verosimilitud:

$$\frac{\partial \ln L(x_1, ..., x_n; p, a)}{\partial p} = n \ln a - \frac{n}{\Gamma(p)} \cdot \frac{\partial \Gamma(p)}{\partial p} + \sum_{i=1}^{n} \ln x_i = 0$$

$$\frac{\partial \ln L(x_1, ..., x_n; p, a)}{\partial a} = \frac{np}{a} - \sum_{i=1}^{n} x_i = 0$$

Para resolver este sistema de ecuaciones empezamos obteniendo el parámetro a de la segunda ecuación:

$$\frac{np}{a} = \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \hat{a} = \frac{np}{\sum_{i=1}^{n} x_i} = \frac{\hat{p}}{\bar{x}}$$

y sustituyendo en la primera ecuación, se tiene:

$$n \ln \frac{p}{\bar{x}} - \frac{n}{\Gamma(p)} \cdot \frac{\partial \Gamma(p)}{\partial p} + \sum_{i=1}^{n} \ln x_i = 0$$

o bien

$$n \ln \frac{p}{\bar{x}} - n \frac{\Gamma'(p)}{\Gamma(p)} + \sum_{i=1}^{n} \ln x_i = 0$$

Pero la solución de esta ecuación hay que obtenerla de manera aproximada mediante métodos numéricos, y una vez que se tiene este estimador $\hat{p}$, el otro se obtiene fácilmente.

Ejemplo 3.13

Sea $(X_1, ..., X_n)$ una muestra aleatoria procedente de una población uniforme, $U(0, \theta)$. Obtener el estimador máximo-verosímil del parámetro θ.

Solución:

La función de densidad de la $U(0, \theta)$ es:

$$f(x) = \begin{cases} \dfrac{1}{\theta} & , \text{ para } \quad 0 \leqslant x \leqslant \theta \\ 0 & , \text{ en el resto} \end{cases}$$

Observemos que aquí no se verifica la condición de que el campo de variación de la variable X sea independiente del parámetro θ.

La función de verosimilitud será:

$$L(x_1, ..., x_n; \theta) = \sum_{i=1}^{n} f(x_i; \theta) = \frac{1}{\theta^n} \qquad \text{si} \qquad 0 \leqslant x_i \leqslant \theta$$

tomando ln se tiene:

$$\ln L(x_1, ..., x_n; \theta) = -n \ln \theta$$

derivando respecto θ e igualando cero resulta:

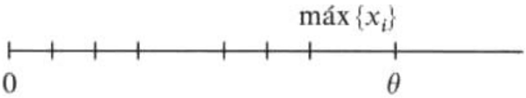
$$\frac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} = -\frac{n}{\theta} = 0$$

y no existe ningún valor de θ para el cual la derivada de la función de verosimilitud es igual a cero, pues el único valor sería $\theta =$ infinito, pero esto no es posible pues entonces $f(x) = 0$, $\forall x$.

Luego vemos que en este caso no podemos aplicar el proceso anterior de derivar el ln de la función de verosimilitud, y sin embargo si podemos encontrar el estimador de máxima verosimilitud; en efecto:

$$\text{maximizar } L(x_1, ..., x_n; \theta) \sim \text{minimizar } \theta$$

pero el mínimo valor de θ será superior al $\max_i \{x_i\}$ que será el valor de x que más se aproxime a θ



Luego el estimador máximo-verosímil de θ será: $\hat{\theta} = \max_i \{x_i\}$

Ejemplo 3.14

Dada una población cuya función de densidad es:

$$f(x;\theta) = \begin{cases} (1+\theta)\ x^{\theta} & , \quad 0 < x < 1 \\ 0 & , \quad \text{en el resto} \end{cases}$$

y una muestra aleatoria $(X_1, ..., X_n)$.

Comprobar que el estimador del parámetro θ obtenido por el método de los momentos no coincide con el estimador máximo-verosímil.

Solución:

Para obtener el estimador por el método de los momentos obtenemos el momento de orden uno respecto al origen de la población y lo igualamos al momento de orden uno de la muestra

$$\alpha_1 = E[X] = \int_0^1 x \cdot (1+\theta)\ x^{\theta}\ dx$$

$$= \int_0^1 (1+\theta)\ x^{1+\theta}\ dx$$

$$= \frac{1+\theta}{2+\theta}$$

$$a_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Igualando ambos momentos, tenemos:

$$\frac{1+\theta}{2+\theta} = \bar{X} \quad \Rightarrow \quad \hat{\theta} = \frac{1-2\bar{X}}{\bar{X}-1}$$

que es el estimador obtenido por el método de los momentos.

Para obtener el estimador máximo-verosímil procedemos como sigue

$$L(x_1, ..., x_n; \theta) = f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

$$= (1+\theta)\, x_1^{\theta} \cdots\cdots (1+\theta)\, x_n^{\theta}$$

$$= (1+\theta)^n \left(\prod_{i=1}^{n} x_i\right)^{\theta}$$

$$\ln L(x_1, ..., x_n; \theta) = n \ln (1+\theta) + \theta \sum_{i=1}^{n} \ln x_i$$

$$\frac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} = \frac{n}{1+\theta} + \sum_{i=1}^{n} \ln x_i = 0$$

$$\theta = \frac{\mathrm{n}}{-\sum_{i=1}^{n} \ln x_i} - 1$$

Luego el estimador de máxima verosimilitud será

$$\hat{\theta} = \frac{n}{-\sum_{i=1}^{n} \ln X_i} - 1$$

y como vemos no tiene porque coincidir con el estimador obtenido por el método de los momentos.

3.3.1. PROPIEDADES DE LOS ESTIMADORES DE MÁXIMA VEROSIMILITUD

Bajo condiciones de regularidad bastante generales se cumplen las siguientes propiedades:

I. Consistencia

Los estimadores de máxima verosimilitud son **consistentes,** es decir para $\forall\ \varepsilon > 0$, se verifica

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1 \qquad \forall\ \theta \tag{3.10}$$

II. Insesgadez

En general los estimadores de máxima verosimilitud no son insesgados. Pero si no son insesgados entonces son **asintoticamente insesgados**.

Si el estimador $\hat{\theta}$ de máxima verosimilitud no es insesgado, como resulta que si es consistente y verifica la expresión [3.10], entonces el estimador $\hat{\theta}$ converge al parámetro θ, y en el límite coincide con su valor medio que es θ, siendo por tanto asintoticamente insesgado.

III. Eficiencia

Si existe un estimador **eficiente** $\hat{\theta}$ del parámetro θ, entonces también es de máxima verosimilitud y es único. Pero todo estimador de máxima verosimilitud **no es eficiente**.

IV. Eficiencia asíntotica

Los estimadores de máxima verosimilitud son **asíntoticamente eficientes**.

V. Normalidad asíntotica

Los estimadores de máxima verosimilitud son **asíntoticamente normales.**

$$\hat{\theta} \to N(\theta, \sqrt{\operatorname{Var}(\hat{\theta})})$$

en donde $\operatorname{Var}(\hat{\theta})$ coincide con la cota de Frechet-Cramer-Rao, es decir:

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{nE\left[\left(\dfrac{\partial \ln f(x;\ \theta)}{\partial \theta}\right)^2\right]} \tag{3.11}$$

VI. **Suficiencia**

Si $\hat{\theta}$ es un estimador suficiente del parámetro θ, entonces el estimador de máxima verosimilitud de θ, si es único, es función del estimador suficiente $\hat{\theta}$.

VII. **Invarianza**

Los estimadores máximo-verosímiles son invariantes frente a transformaciones biunívocas. Es decir, si $\hat{\theta}$ es el estimador de máxima verosimilitud del parámetro θ y $g(\theta)$ es una función con inversa única, entonces se verifica que $g(\hat{\theta})$, es el estimador de máxima verosimilitud de $g(\theta)$.

De las propiedades I, IV y V se deduce que los estimadores de máxima v verosimilitud son estimadores **óptimos asíntoticamente normales** (O.A.N).

Ejemplo 3.15

Sea una población cuya función de densidad es:

$$f(x; \theta) = \begin{cases} \theta^{-1} \, e^{-\frac{x}{\theta}} & , \quad x > 0 \\ 0 & , \quad \text{en el resto} \end{cases}$$

y consideremos una muestra aleatoria $(X_1, ..., X_n)$. Se pide

1. Estimador máximo-verosímil del parámetro θ.
2. Comprobar si es insesgado y consistente.
3. Comprobar si el estimador máximo-verosímil es eficiente.

Solución:

1. La función de verosimilitud viende dada por:

$$\begin{aligned} L(x_1, ..., x_n; \theta) &= f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) \\ &= \theta^{-1} \, e^{-\frac{x_1}{\theta}} \cdots\cdots \theta^{-1} \, e^{-\frac{x_n}{\theta}} \\ &= \theta^{-n} \, e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}} \end{aligned}$$

El ln de la función de verosimilitud es:

$$\ln L(x_1, ..., x_n; \theta) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^{n} x_i$$

Derivando respecto a θ e igualando a cero tenemos:

$$\frac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i = 0$$

$$\theta = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Luego el estimador insesgado del parámetro θ será:

$$\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

2. Veamos que es insesgado y consistente:

Como se trata de una distribución exponencial de parámetro $\frac{1}{\theta}$, sabemos que:

$$E[X] = \theta$$

$$\mathrm{Var}(X) = \theta^2$$

Luego

$$E[\hat{\theta}] = E[\bar{X}] = E[X] = \theta$$

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{\theta^2}{n}$$

Cuando $n \to \infty$, entonces la $\mathrm{Var}(\hat{\theta}) \to 0$ y como el estimador $\hat{\theta}$ es insesgado, resulta que efectivamente el estimador de máxima verosimilitudd es consistente, pues el sesgo es nulo y la varianza tiende a cero cuando n tiende a infinito.

3. Para probar la eficiencia, tendremos que probar que la varianza del estimador coincide con la cota de Frechet-Cramer-Rao, es decir que,

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{n\ E\left[\dfrac{\partial \ln f(x;\ \theta)}{\partial \theta}\right]^2}$$

o bien

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{-n\,E\left[\dfrac{\partial \ln f(x;\ \theta)}{\partial \theta^2}\right]}$$

$$\ln f(x;\ \theta) = -\ln \theta - \frac{\mathrm{x}}{\theta}, \quad x > 0$$

$$\frac{\partial \ln f(x;\ \theta)}{\partial \theta} = -\frac{1}{\theta} + \frac{x}{\theta^2}, \quad x > 0$$

$$\frac{\partial^2 \ln f(x;\ \theta)}{\partial \theta^2} = -\frac{1}{\theta^2} + \frac{2x}{\theta^3}, \quad x > 0$$

$$E\left[\frac{\partial^2 \ln f(x;\ \theta)}{\partial \theta^2}\right] = E\left[\frac{1}{\sigma^2} - \frac{2X}{\theta^3}\right]$$

$$= \frac{1}{\theta^2} - \frac{2}{\theta^3}\,E[X]$$

$$= \frac{1}{\theta^2} - \frac{2}{\theta^3}\cdot\theta$$

$$= \frac{1}{\theta^2} - \frac{2}{\theta^2} = -\frac{1}{\theta^2}$$

Así la cota de Freschet-Cramer-Rao será:

$$\frac{1}{-\mathrm{n}\,\mathrm{E}\left[\dfrac{\partial^2 \ln f(x;\ \theta)}{\partial \theta^2}\right]} = \frac{1}{-n\left(\dfrac{-1}{\theta^2}\right)} = \frac{\theta^2}{n}$$

que coincide con la $\operatorname{Var}(\hat{\theta})$, siendo por tanto el estimador de máxima verosimilitud para este ejemplo eficiente.

3.4. MÉTODO DE LA MÍNIMA χ^2

Se trata de un método general para la obtención de estimadores puntuales, es de menos aplicación que el método de los momentos y que el método de la máxima verosimilitud, y se aplica sólo cuando hay una gran cantidad de datos tanto en distribuciones discretas como en distribuciones continuas pero con datos agrupados.

Veamos en que consiste este método de estimación:

Supongamos una población representada por la variable aleatoria X cuya función de probabilidad $p(x; \theta_1, ... \theta_k)$, depende de k parámetros $\theta_1, ..., \theta_k$ desconocidos, y el campo de variación de la variable aleatoria X de la población lo suponemos dividido en r subconjuntos excluyentes $S_1, ..., S_r$, asociando a cada uno de ellos las probabilidades $p_1, ..., p_r$ respectivamente, es decir:

$$p_i(\theta_1, ..., \theta_k) = P(S_i) = P(X \in S_i) > 0, \quad \sum_{i=1}^{r} p_i = 1$$

Para estimar los parámetros desconocidos $\theta_1, ..., \theta_k$ tomamos una muestra aleatoria de tamaño n, cuyas observaciones han sido ordenadas en forma de distribución de frecuencias, de tal manera que el número de observaciones que pertenecen a cada uno de los subconjuntos $S_1, ..., S_r$ será $n_1, ..., n_r$, siendo $\sum_{i=1}^{r} n_i = n$. Es decir tendríamos la siguiente distribución de frecuencias:

Campo de variación X	**Frecuencias absolutas** n_i	**Frecuencia relativa** n_i/n
S_1	n_1	n_1/n
S_2	n_2	n_2/n
$\vdots$	$\vdots$	$\vdots$
S_r	n_r	n_r/n
	n	

Según la distribución teórica de la población, X, a cada subconjunto S_i le corresponde la probabilidad $p_i(\theta_1, ..., \theta_k)$ es decir:

$$p_i(\theta_1, ..., \theta_k) = P(S_i) = P(X \in S_i) \qquad i = 1, ..., r$$

y según las frecuencias observadas en la muestra le corresponde una frecuencia relativa $\frac{n_i}{n}$, con lo cual existen unas desviaciones entre ambas distribuciones.

Teniendo en cuenta el principio de los mínimos cuadrados utilizaremos como medida de la desviación entre ambas distribuciones la expresión:

$$\sum_{i=1}^{r} c_i\left(\frac{n_i}{n} - p_i(\theta_1, ..., \theta_k)\right)^2 \qquad [3.12]$$

y Pearson demostró que si tomamos

$$c_i = \frac{n}{p_i(\theta_1, ..., \theta_k)}$$

entonces obtenemos una medida de la desviación cuyas propiedades son relativamente fáciles y de cierto interés para estudiar la desviación entre ambas distribuciones. Así pues sustituyendo c_i; por su valor en la expresión [3.12], y designándola por χ^2, tenemos:

$$\chi^2 = \sum_{i=1}^{r} \frac{n}{p_i(\theta_1, ..., \theta_k)}\left(\frac{n_i}{n} - p_i(\theta_1, ..., \theta_k\right)^2$$

$$= \sum_{i=1}^{r} \frac{(n_i - n\ p_i(\theta_1, ..., \theta_k))^2}{n\ p_i\ (\theta_1, ..., \theta_k)} \qquad [3.13]$$

$$= \sum_{i=1}^{r} (n_i - n\ p_i(\theta_1, ..., \theta_k))^2\ \frac{1}{n\ p_i(\theta_1, ..., \theta_k)}$$

y se demuestra[8] que este estadístico sigue una distribución χ^2_{r-k-1} pues hay k parámetros desconocidos.

Entonces el **método de la mínima** χ^2 escoge los estimadores de los parámetros θ_i de modo que el estadístico χ^2 dado por la expresión [3.13] sea mínimo.

Para minimizar el estadístico χ^2, tendremos que derivar respecto de $\theta_j(j = 1, ..., k)$ e igualar a cero:

$$\frac{\partial \chi^2}{\partial \theta_j} = \frac{\partial \chi^2}{\partial p_i} \cdot \frac{\partial p_i(\theta_1, ..., \theta_k)}{\partial \theta_j}$$

$$= \sum_{i=1}^{r}\left[-2(n_i - np_i(\theta_1,...,\theta_k))n \cdot \frac{1}{np_i(\theta_1,...,\theta_k)} - \frac{(n_i - np_i(\theta_1,...,\theta_k))^2 n}{n^2 p_i^2(\theta_1,...,\theta_k)}\right]\frac{\partial p_i(\theta_1,...,\theta_k)}{\partial \theta_j}$$

[8] Ver CRAMER H. (1963).

$$= -2\sum_{i=1}^{r}\left[\frac{n_i - n\,p_i(\theta_1, ..., \theta_k)}{p_i(\theta_1, ..., \theta_k)} + \frac{(n_i - n\,p_i(\theta_1, ..., \theta_k))^2}{2n\,p_i^2(\theta_1, ..., \theta_k)}\right]\frac{\partial z p_i(\theta_1, ..., \theta_k)}{\partial \theta_j} = 0$$

$$j = 1, ..., k \qquad [3.14]$$

y resolviendo este sistema tendríamos los estimadores de mínima χ^2, $\hat{\theta}_1, ..., \theta_k$, de los parámetros desconocidos $\theta_1, ..., \theta_k$.

Minimizar la expresión [3.13] presenta dificultades análogas a las que se presentan en el método de máxima verosimilitud, por las complicaciones que se pueden presentar en la resolución del sistema.

Los estimadores de mínima χ^2 son asíntoticamente equivalentes a los estimadores de máxima verosimilitud. Sin embargo cuando n es pequeño no se puede asegurar nada, pues el estimador de mínima χ^2 no tiene porque ser función de estimador suficiente si existe. Generalmente, el estimador de mínima χ^2 es sesgado, y no eficiente.

El sistema [3.14] suele ser complicado de resolver aún en casos sencillos, sin embargo se puede demostrar que para valores grandes de n la influencia del segundo término se hace despreciable, quedando el sistema reducido a:

$$\sum_{i=1}^{r}\frac{n_i - n\,p_i(\theta_1, ..., \theta_k)}{p_i(\theta_1, ... \theta_k)}\cdot\frac{\partial p_i(\theta_1, ..., \theta_k)}{\partial \theta_j} = 0 \qquad j = 1, ..., k \qquad [3.15]$$

facilitándose su resolución.

Este nuevo método de estimación obtenido con esta simplificación recibe el nombre de **método modificado de la mínima** χ^2, y por sencillez será el que utilizaremos.

Los estimadores obtenidos por ambos métodos son asíntoticamente equivalentes, y coincide con el estimador obtenido por el método de la máxima verosimilitud. En efecto:

El sistema [3.15] equivale a:

$$\sum_{i=1}^{r}\frac{n_i}{p_i(\theta_1, ..., \theta_k)}\cdot\frac{\partial p_i(\theta_1, ..., \theta_k)}{\partial \theta_j} = 0 \qquad j = 1, ..., k \qquad [3.16]$$

ya que

$$n\sum_{i=1}^{r}\frac{\partial p_i(\theta_1, ..., \theta_k)}{\partial \theta_j} = n\frac{\partial}{\partial \theta_j}\sum_{i=1}^{r} p_i(\theta_1, ..., \theta_k) = 0$$

puesto que

$$\sum_{i=1}^{r} p_i(\theta_1, ..., \theta_k) = 1$$

Por tanto el sistema [3.16] es equivalente a escribir la ecuación de verosimilitud:

$$\frac{\partial \ln L(x;\ \theta_1,\ ...,\ \theta_k)}{\partial \theta_j} = 0$$

en donde

$$L(x;\, \theta_1, ..., \theta_k) = p_1^{n_1}(\theta_1, ..., \theta_k) \cdots p_r^{n_r}(\theta_1, ..., \theta_k)$$

$$\ln L(x;\, \theta_1, ..., \theta_k) = n_1 \ln p_1(\theta_1, ..., \theta_k) + \cdots + n_r \ln p_r(\theta_1, ..., \theta_k)$$

$$\frac{\partial \ln L(x;\ \theta_1,\ ...,\ \theta_k)}{\partial \theta_j} = n_1 \frac{1}{p_1(\theta_1,\ ...,\ \theta_k)} \frac{\partial p_1(\theta_1,\ ...,\ \theta_k)}{\partial \theta_j} + \cdots +$$

$$+ n_r \frac{1}{p_r(\theta_1,\ ...,\ \theta_k)} \frac{\partial p_r(\theta_1,\ ...,\ \theta_k)}{\partial \theta_j}$$

$$= \sum_{i=1}^{r} \frac{n_i}{p_i(\theta_1,\ ...,\ \theta_k)} \cdot \frac{\partial p_i(\theta_1,\ ...,\ \theta_k)}{\partial \theta_j} = 0$$

$$j = 1, ..., k$$

y de aquí obtendríamos los estimadores máximo-verosímiles.

3.5. ESTIMADORES LINEALES INSESGADOS

Diremos que un **estimados es lineal** si tiene la forma

$$\hat{\theta} = a_1 X_1 + \cdots + a_n X_n$$

donde los coeficientes a_i toman diferentes valores según los parámetros a estimar.

Llamaremos **clase de los estimadores lineales insesgados** de una cierta función $g(\theta)$ a la familia de todos los estimadores insesgados de $g(\theta)$ que son funciones lineales de las observaciones muestrales.

Con frecuencia nos interesa saber si entre la clase de todos los estimadores lineales insesgados existe uno que tenga varianza mínima, y a ese estimador le llamaremos **estimador lineal insesgado de mínima varianza.**

En la práctica se suele llamar **estimador lineal insesgado óptimo**[9] cuando la varianza del estimador lineal es mínima con respecto a todos los demás estimadores lineales insesgados.

3.5.1. MÉTODO DE LA MÍNIMA VARIANZA

Es un método analítico, y consiste en hacer mínima la varianza del estimador. La técnica que se utiliza para encontrar este mínimo condicionado por las restricciones que le queramos imponer al estimador, p. ej. que sea lineal, insesgado, etc., es la de los multiplicadores de Lagrange.

Veamos dos aplicaciones:

1. **Estimador lineal, insesgado y de mínima varianza para la media poblacional**

Proposición 3.1

> Si $(X_1, ..., X_n)$ es una muestra aleatoria procedente de una población con media μ y varianza σ^2 entonces la media muestral $\bar{X}$ es un estimador lineal, insesgado y de varianza mínima para el parámetro poblacional μ.

Demostración:

Sea el estimador lineal

$$\hat{\mu} = a_1 X_1 + \cdots + a_n X_n \tag{3.17}$$

como ha de ser insesgado

$$\begin{aligned} E[\hat{\mu}] &= E[a_1 X_1 + \cdots + a_n X_n] \\ &= a_1 E[X_1] + \cdots + a_n E[X_n] \\ &= a_1 \mu + \cdots + a_n \mu \\ &= \mu \sum_{i=1}^{n} a_i \\ &= \mu \end{aligned}$$

entonces $\sum_{i=1}^{n} a_i = 1$, para que el estimador $\hat{\mu}$ sea insesgado.

[9] BLUE: Best Linear Unbiased Estimator.

La varianza del estimador $\hat{\mu}$ será:

$$\begin{aligned}\operatorname{Var}(\hat{\mu}) &= \operatorname{Var}(a_1X_1 + \cdots + a_nX_n)\\ &= a_1^2 \operatorname{Var}(X_1) + \cdots + a_n^2 \operatorname{Var}(X_n)\\ &= a_1^2\sigma^2 + \cdots + a_n^2\sigma^2\\ &= \sigma^2 \sum_{i=1}^{n} a_i^2\end{aligned}$$

pero como la $\operatorname{Var}(\hat{\mu})$ ha de ser mínima, entonces los valores $a_1, ..., a_n$ deben de ser tales que sea

$$\operatorname{Var}(\hat{\mu}) = \sigma^2 \sum_{i=1}^{n} a_i^2 \qquad \text{mínima}$$

con la restricción

$$\sum_{i=1}^{n} a_i = 1$$

Para ello aplicamos el método de los multiplicadores de Lagrange, siendo la función

$$\phi = \sigma^2 \sum_{i=1}^{n} a_i^2 + \lambda\left(\sum_{i=1}^{n} a_i - 1\right); \qquad \lambda \rightarrow \text{multiplicados de Lagrange}$$

$$\frac{\partial \phi}{\partial a_i} = 2\, a_i\sigma^2 + \lambda = 0, \qquad i = 1, ..., n$$

$$2\sigma^2 \sum_{i=1}^{n} a_i + n\,\lambda = 0$$

pero $\sum_{i=1}^{n} a_i = 1$, resulta:

$$2\sigma^2 + n\,\lambda = 0$$

$$\lambda = -\frac{2\sigma^2}{n}$$

de donde

$$2\, a_i\sigma^2 - \frac{2\sigma^2}{n} = 0$$

$$a_i = \frac{1}{n}$$

Luego, sustituyendo en la expresión [3.17], se tiene:

$$\hat{\mu} = \frac{1}{n} X_1 + \cdots + \frac{1}{n} X_n$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$= \bar{X}$$

y por tanto hemos llegado a probar que efectivamente la media muestral es un estimador lineal, insesgado y de varianza mínima para la media poblacional μ.

Proposición 3.2

Si $Y_1, ..., Y_n$ son n variables aleatorias independientes con media

$$E[Y_i] = a + bX_i$$

y varianza σ^2, entonces un estimador lineal insesgado y de mínima varianza para el coeficiente de regresión viende dado por

$$\hat{b} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \qquad [3.18]$$

Demostración:

Como el estimador $\hat{b}$ debe de ser lineal tiene la forma:

$$\hat{b} = c_1 Y_1 + \cdots + c_n Y_n = \sum_{i=1}^{n} c_i Y_i \qquad [3.19]$$

y para que sea insesgado se tiene que verificar:

$$E[\hat{b}] = E[c_1 Y_1 + \cdots + c_n Y_n]$$

$$= c_1 E[Y_1] + \cdots + c_n E[Y_n]$$

$$= \sum_{i=1}^{n} c_i \, E[Y_i]$$

$$= \sum_{i=1}^{n} c_i(a + bX_i)$$

$$= a \sum_{i=1}^{n} c_i + b \sum_{i=1}^{n} c_i X_i$$

$$= b$$

de donde se deduce que hemos de imponer las condiciones

$$\left.\begin{aligned} &\sum_{i=1}^{n} c_i = 0 \\ &\sum_{i=1}^{n} c_i X_i = 1 \end{aligned}\right. \qquad [3.20]$$

La varianza del estimador $\hat{b}$ será:

$$\begin{aligned} \operatorname{Var}(\hat{b}) &= \operatorname{Var}(c_1 Y_1 + \cdots + c_n Y_n) \\ &= c_1^2 \operatorname{Var}(Y_1) + \cdots + c_n^2 \operatorname{Var}(Y_n) \\ &= c_1^2 \sigma^2 + \cdots + c_n^2 \sigma^2 \\ &= \sigma^2 \sum_{i=1}^{n} c_i^2 \end{aligned}$$

pero como la $\operatorname{Var}(\hat{b})$ tiene que ser mínima, entonces tenemos que hacer mínima la expresión:

$$\sigma^2 \sum_{i=1}^{n} c_i^2$$

pero esto es equivalente a hacer mínima la expresión:

$$\sum_{i=1}^{n} c_i^2$$

con las restricciones dadas en [3.20]. Es decir, la correspondiente función de Lagrange será:

$$\phi = \sum_{i=1}^{n} c_i^2 - 2\,\lambda_1 \sum_{i=1}^{n} c_i - 2\,\lambda_2\left(\sum_{i=1}^{n} c_i X_i - 1\right)$$

$$\frac{\partial\,\phi}{\partial\,c_i} = 2\,c_i - 2\,\lambda_1 - 2\,\lambda_2 X_i = 0$$

$$c_i = \lambda + \lambda_2 X_i\,, \qquad i = 1, ..., n$$

$$0 = \sum_{i=1}^{n} c_i = n\,\lambda_1 + \lambda_2 \sum_{i=1}^{n} X_{-1}$$

$$\lambda_1 = -\lambda_2 \frac{\sum_{i=1}^{n} X_i}{n} = -\lambda_2 \bar{X}$$

de donde:

$$c_i = -\lambda_2 \bar{X} + \lambda_2 X_i = \lambda_2 (X_i - \bar{X})$$

$$1 = \sum_{i=1}^{n} c_i X_i = \lambda_2 \sum_{i=1}^{n} (X_i - \bar{X})\,X_i$$

$$= \lambda_2 \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

$$= \lambda_2 \sum_{i=1}^{n} (X_i - \bar{X})^2$$

y despejando λ_2:

$$\lambda_2 = \frac{1}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

y sustituyendo en la expresión de c_i tenemos:

$$c_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Luego sustituyendo en la expresión [3.19] el estimador $\hat{b}$ será[10]:

$$\hat{b} = \sum_{i=1}^{n} c_i Y_i = \frac{\sum_{i=1}^{n} (X_i - \bar{X})\, Y_i}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$= \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

que coincide con el estimador de máxima verosimilitud y con el de mínimos cuadrados.

Ejemplo 3.16

Sea (X_1, X_2, X_3) una muestra aleatoria simple procedente de una población con media μ y varianza σ^2, y sean

$$\hat{\theta}_1 = 3X_1 - X_2 - X_3$$

$$\hat{\theta}_2 = \frac{1}{3} X_1 + \frac{1}{3} X_2 + \frac{1}{3} X_3$$

estimadores lineales de la media poblacional. Determinar:

1. Si ambos estimadores son insesgados para la media poblacional.
2. Cuál de los dos estimadores es el lineal de varianza mínima.

---

[10] $\sum_{i=1}^{n} (X_i - \bar{X})\, Y_i = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$. En efecto

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - \bar{Y} \sum_{i=1}^{n} X_i - \bar{X} \sum_{i=1}^{n} Y_i + n\bar{X}\bar{Y}$$

$$= \sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$$

$$\sum_{i=1}^{n} (X_i - \bar{X})\, Y_i = \sum_{i=1}^{n} X_i Y_i - \bar{X} \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}.$$

Solución:

1. Observando los coeficientes de ambos estimadores se verifica que:

$$\sum_{i=1}^{3} a_1 = 1$$

y en efecto ambos son insesgados para la media poblacional:

$$\begin{aligned} E[\hat{\theta}_1] &= E[3X_1 - X_2 - X_3] \\ &= 3E[X_1] - [X_2] - [X_3] \\ &= 3\mu - \mu - \mu \\ &= \mu \end{aligned}$$

$$\begin{aligned} E[\hat{\theta}_2] &= E\left[\frac{1}{3} X_1 + \frac{1}{3} X_2 + \frac{1}{3} X_3\right] \\ &= \frac{1}{3} E[X_1] + \frac{1}{3} E[X_2] + \frac{1}{3} E[X_3] \\ &= \frac{1}{3} \mu + \frac{1}{3} \mu + \frac{1}{3} \mu \\ &= \mu \end{aligned}$$

2. Las varianzas de ambos estimadores son:

$$\begin{aligned} \operatorname{Var}(\hat{\theta}_1) &= \operatorname{Var}(3X_1 - X_2 - X_3) \\ &= 9 \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \operatorname{Var}(X_3) \\ &= 11\sigma^2 \end{aligned}$$

$$\begin{aligned} \operatorname{Var}(\hat{\theta}_2) &= \operatorname{Var}\left(\frac{1}{3} X_1 + \frac{1}{3} X_2 + \frac{1}{3} X_3\right) \\ &= \frac{1}{9} \operatorname{Var}(X_1) + \frac{1}{9} \operatorname{Var}(X_2) + \frac{1}{9} \operatorname{Var}(X_3) \\ &= \frac{1}{3} \sigma^2 \end{aligned}$$

Lo cual nos indica que el estimador $\hat{\theta}_2$ es de varianza mínima, pues se verifican las dos condiciones:

$$a_i = \frac{1}{3} \quad , \quad \sum_{i=1}^{3} a_i = 1$$

También podemos decir que se deduce directamente a partir de la proposición 3.1.

3.6. MÉTODO DE LOS MÍNIMOS CUADRADOS

Nos limitaremos a hacer unas consideraciones muy elementales, pues se estudiará con detalle cuando se realice el estudio de los modelos lineales.

Este método de los mínimos cuadrados fue introducido por Gauss, y generalmente se utiliza para estimar los parámetros de un modelo lineal.

Sea el modelo

$$Y = a + bX + e \tag{3.17}$$

en donde Y es una variable aleatoria cuyo valor esperado es

$$E[Y] = a + bX$$

X es una variable observable que tomará valores conocidos, a y b son parámetros desconocidos, y e una variable aleatoria o error[11].

Se toman n-valores X_i de la variable X y para cada valor observado de X_i tendremos el correspondiente valor observado Y_i y el correspondiente valor teórico Y_i proporcionado por la función que se pretende ajustar. El error cometido será:

$$e_i = Y_i - Y_i' = Y_i - a - bX_i \quad , \quad i = 1, ..., n$$

y se admite:

— Que los errores e_i están incorrelacionados, $\mathrm{cov}(e_i, e_j) = 0$.

— $E[e_i] = 0$.

— $\mathrm{Var}(e_i) = \sigma^2$.

[11] En el método de los mínimos cuadrados se parte de un conjunto de puntos observados y se pretende que la función teórica ajustada pase lo más cerca posible de todos esos puntos. Para ello hay que hacer mínima la distancia global de todos los puntos observados a la función teórica ajustada, la cual nos permitirá conocer los parámetros desconocidos.

Para estimar los parámetros a y b, haremos mínima la distancia global de todos los puntos observados a la función teórica ajustada, pero esta distancia viene dada por $\sum_{i=1}^{n} e_i^2$. Luego haremos mínima la función:

$$\phi(a, b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - a - bX_i)^2 \qquad [3.18]$$

Derivando respecto de los parámetros a y b e igualando a cero tenemos las ecuaciones normales que nos permiten conocer los valores de a y b:

$$\frac{\partial \phi}{\partial a} = -2 \sum_{i=1}^{n} (Y_i - a - bX_i) = 0$$

$$\frac{\partial \phi}{\partial b} = -2 \sum_{i=1}^{n} X_i(Y_i - a - bX_i) = 0$$

de donde se tiene:

$$\sum_{i=1}^{n} Y_i = na + b \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_iY_i = a \sum_{i=1}^{n} X_i + b \sum_{i=1}^{n} X_i^2$$

Dividiendo por n la primera ecuación obtenemos:

$$\hat{a} = \frac{\sum_{i=1}^{n} Y_i}{n} - \hat{b} \frac{\sum_{i=1}^{n} X_i}{n} = \bar{Y} - \hat{b}\bar{X}$$

sustituyendo en la segunda ecuación y despejando resulta:

$$\hat{b} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

que serían los estimadores mínimo cuadráticos.

Los estimadores obtenidos por este método son funciones lineales de las Y_i y coinciden con los estimadores lineales insesgados de mínima varianza.

Capítulo 4
ESTIMACIÓN POR INTERVALOS DE CONFIANZA

4.1. INTRODUCCIÓN

En los capítulos anteriores, nos hemos ocupado de las propiedades y de la obtención de estimadores puntuales de los parámetros poblacionales. Veíamos que los *estimadores* eran funciones de las observaciones muestrales, y cuando se calcula el valor del estimador $\hat{\theta}$ para una muestra concreta entonces se tiene la *estimación puntual*; valor que generalmente difiere del verdadero valor del parámetro θ y, en consecuencia, no nos proporciona suficiente información sobre el parámetro, siendo entonces deseable el acompañar a la estimación del parámetro θ, de alguna medida del posible error asociado a esta estimación. Es decir, asociado a cada estimación del parámetro daremos un intervalo:

$$[\underline{\hat{\theta}}(X_1, ..., X_n), \overline{\theta}(X_1, ..., X_n)]$$

y una medida que nos refleje la confianza que tenemos acerca de que el verdadero valor del parámetro θ se encuentre dentro del intervalo.

Observemos que los extremos del intervalo variarán de manera aleatoria de una muestra a otra, pues dependen de las observaciones de la muestra, luego tanto los extremos del intervalo como la longitud del intervalo serán cantidades aleatorias y, por tanto, no podremos saber con seguridad si el valor del parámetro θ se encuentre dentro del intervalo obtenido cuando se selecciona una sola muestra. El objetivo que se pretende con los intervalos de confianza es obtener un intervalo de poca amplitud y con una alta probabilidad de que el parámetro θ se encuentra en su interior. Así pues, elegiremos probabilidades cercanas a la unidad, que se representan por $1 - \alpha$ y cuyos valores más frecuentes suelen ser 0,90, 0,95 y 0,99.

Luego si deseamos obtener una estimación por intervalo del parámetro poblacional θ desconocido, tendremos que obtener dos estadísticos $\underline{\theta}(X_1, ..., X_n)$ y $\overline{\theta}(X_1, ..., X_n)$ que nos darán los valores extremos del intervalo, tales que

$$P[\underline{\theta}(X_1, ..., X_n) \leqslant \theta \leqslant \overline{\theta}(X_1, ..., X_n)] = 1 - \alpha \qquad [4.1]$$

Al valor $1 - \alpha$ se le llama **coeficiente de confianza,** y

Al valor $100(1 - \alpha)\%$ se le llama **nivel de confianza.**

Observando el intervalo dado en la expresión [4.1] se pone de manifiesto:

1.° Que se trata de un *intervalo aleatorio*, pues los extremos dependen de la muestra seleccionada y, por tanto, $\underline{\theta}$ y $\overline{\theta}$son variables aleatorias.

2.° Que el parámetro θ es desconocido.

3.° En consecuencia y antes de seleccionar una muestra no podemos decir que *la probabilidad de que el parámetro θ tome algún valor en el intervalo $(\underline{\theta}, \overline{\theta})$ es igual a* $1 - \alpha$, afirmación que no sería correcta después de seleccionar la muestra.

Para una muestra concreta se tendrían unos valores:

$$\underline{\theta}(x_1, ..., x_n) = a \quad \text{y} \quad \overline{\theta}(x_1, ..., x_n) = b$$

y no podemos afirmar que

$$P[a \leqslant \theta \leqslant b] = 1 - \alpha$$

ya que no tiene sentido alguno, pues a, b y θ son tres valores constantes. Sin embargo, una vez seleccionada la muestra y calculados, los valores:

$$a = \underline{\theta}(x_1, ..., x_n) \quad \text{y} \quad b = \overline{\theta}(x_1, ..., x_n)$$

si tiene sentido decir que:

— la probabilidad es 1 si $\theta \in [a, b]$

— la probabilidad es 0 si $\theta \notin$ [a, b]

Luego, no podemos referirnos a la probabilidad del intervalo numérico sino que nos referiremos al coeficiente de confianza del intervalo, y en consecuencia al nivel de confianza del intervalo, pues la probabilidad ya hemos indicado que, después de extraída la muestra, será 1 o cero[1].

Para precisar más sobre la interpretación del intervalo de confianza, consideramos un número grande de muestras del mismo tamaño y calculamos los límites inferior y superior para cada muestra:

$$a = \bar{\theta}(x_1, ..., x_n) \quad \text{y} \quad b = \bar{\theta}(x_1, ..., x_n)$$

entonces se obtendrá que aproximadamente en el $100(1 - \alpha)\%$ de los intervalos resultantes estará en su interior el valor del parámetro θ, y en el $100\alpha\%$ restante no estará en su interés el valor del parámetro θ, y en consecuencia al intervalo (a, b) se le llama **intervalo de confianza al nivel de confianza del 100(1 − α) %.**

Una ilustración gráfica la tenemos en el gráfico 4.1 que nos muestra gráficamente la obtención de los 100 intervalos construidos para la media μ de una población $N(\mu, \sigma)$, con σ conocida, y que como posteriormente veremos tiene la forma

$$I_\mu = \left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad ; \quad \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

en donde

$$\underline{\theta}(X_1, ..., X_n) = \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\bar{\theta}(X_1, ..., X_n) = \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Considerando como *coeficiente de confianza*

$$1 - \alpha = 0{,}95$$

tendremos la siguiente interpretación:

---

[1] Véase ARNAIZ, pág. 581, aparece un ejemplo muy sencillo en donde se pone de manifiesto la diferencia existente entre coeficiente de confianza y probabilidad.

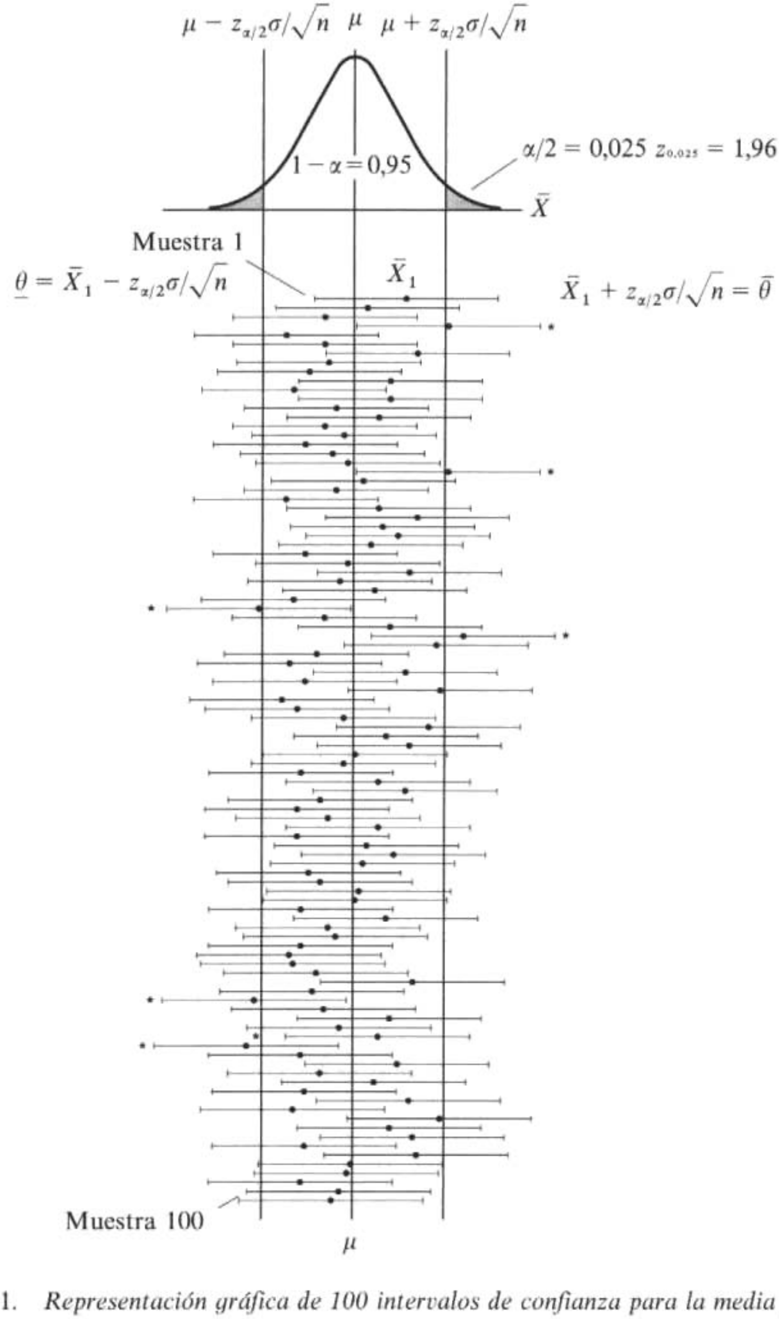


GRÁFICO 4.1. *Representación gráfica de 100 intervalos de confianza para la media μ de una población $N(\mu, \sigma)$, con muestras del mismo tamaño n y coeficiente de confianza del 0,95.*

«Si tomamos 100 muestras aleatorias de tamaño n de la misma población y calculamos los límites de confianza $\underline{\theta}$ y $\overline{\theta}$ para cada muestra, entonces esperamos que aproximadamente el 95 % de los intervalos contendrán en su interior el verdadero valor del parámetro μ, y el 5 % restante no lo contendrán. Pero como nosotros, en la práctica, sólo tomamos una muestra aleatoria y, por tanto, sólo tendremos un intervalo de confianza, no conocemos si nuestro intervalo es uno del 95 % o uno del 5 %, y por eso hablamos de que tenemos un **nivel de confianza** del 95 %.»

En el gráfico 4.1 tenemos representados los 100 intervalos del parámetro media poblacional μ, correspondientes a 100 muestras aleatorias del mismo tamaño n, y podemos observar que exactamente 94 intervalos contienen en su interior el parámetro μ, resultado que concuerda con nuestra esperanza o confianza que era de aproximadamente 95.

Hasta ahora solo hemos hablado de intervalos de confianza **bilaterales,** pero en la práctica, nos pueden interesar intervalos **unilaterales,** es decir intervalos de la forma:

$$[\underline{\theta}(X_1, ..., X_n); +\infty]$$

$$(-\infty; \overline{\theta}(X_1, ..., X_n)]$$

La **precisión** de la estimación por intervalos vendrá caracterizada por el **coeficiente de confianza** $1 - \alpha$ y por la **amplitud del intervalo.** Así pues, para un coeficiente de confianza fijo, cuanto más pequeños sea el intervalo de confianza más precisa será la estimación, o bien para una misma amplitud del intervalo, cuanto mayor sea el coeficiente de confianza mayor será la precisión. En el ejemplo 4.4 veremos cómo se producen estas variaciones.

4.2. MÉTODOS DE CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA

Básicamente daremos dos métodos para la obtención de intervalos de confianza de parámetros. El primero, el **método pivotal** o **método del pivote** basado en la posibilidad de obtener una función del parámetro desconocido y cuya distribución muestral no dependa del parámetro. El segundo, el **método general de Neyman,** está basado en la distribución de un estimador puntual del parámetro.

También veremos cómo se obtiene un intervalo de confianza cuando no se conoce la distribución de la población de partida. Por último, basándonos en las propiedades asintóticas de los estimadores, obtendremos intervalos de confianza para muestras grandes.

4.2.1. MÉTODO PIVOTAL

Sea una población con función de distribución $F(x; \theta)$ en donde θ es un parámetro desconocido, que toma valores en el espacio paramétrico Ω.

Este método básicamente consiste en la obtención de una **cantidad pivotal** o simplemente **pivote** que verifique las siguientes condiciones:

1. La cantidad pivotal o pivote, $T(X_1, ..., X_n; \theta)$, es una función de las observaciones muestrales y del parámetro θ, de tal manera que para cada muestra solo dependerá de θ.

2. La distribución muestral de la cantidad pivotal o pivote $T(X_1, ..., X_n; \theta)$ no depende del parámetro θ.

Veremos la aplicación de este método con un ejemplo.

Ejemplo 4.1

Sea $(X_1, ..., X_n)$ una muestra aleatoria procedente de una población $N(\mu, \sigma)$, con σ conocida. Deseamos obtener un intervalo de confianza al nivel del $100(1 - \alpha)\%$ para el parámetro poblacional μ.

Solución:

Sabemos que un buen estimador de la media poblacional μ es la media muestral $\bar{X}$, la cual sigue una distribución $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, en donde el parámetro μ es desconocido.

Pero el estadístico

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

se distribuye según una $N(0, 1)$, la cual no depende de μ.

Este estadístico Z, podemos considerar que sería el **pivote**

$$T(X_1, ..., X_n; \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad [4.2]$$

pues reúne las condiciones impuestas.

Podemos encontrar dos valores simétricos $-z_{\alpha/2}$ y $z_{\alpha/2}$, tales que nos proporcionen el siguiente intervalo:

$$
\begin{aligned}
1 - \alpha &= P(z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}) \\
&= P\left(-z_{\alpha/2} \leqslant \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \leqslant z_{\alpha/2} \right) \\
&= P\left(-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leqslant \bar{X} - \mu \leqslant z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \\
&= P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + z_{\alpha/2} \frac{\sigma}{n} \right) \qquad [4.3]
\end{aligned}
$$

en donde el extremo inferior es:

$$\underline{\theta}(X_1, ..., X_n) = \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Luego el intervalo de confianza con un nivel de confianza del $100(1 - \alpha)\%$ para el parámetro poblacional μ es:

$$I_\mu = \left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}};\ \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Observamos que la cantidad pivotal o pivote es un estadístico que verifica:

1. La cantidad pivotal o pivote

$$T(X_1, ..., X_n; \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

depende de las observaciones muestrales y del parámetro μ, o lo que es lo mismo, del estimador y del parámetro.

2. La distribución de la cantidad pivotal o pivote se puede obtener y no depende del parámetro μ, pues

$$T(X_1, \ldots, X_n; \mu) = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \to N(0, 1)$$

Una variable aleatoria o estadístico que satisface estas condiciones se llama **cantidad pivotal o pivote,** y nos permite obtener el intervalo de la forma:

$$P\left(-z_{\alpha/2} \leqslant \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqslant z_{\alpha/2} \right) = 1 - \alpha$$

Posteriormente y a lo largo de este capítulo utilizaremos este método pivotal junto con las distribuciones muestrales para obtener, en algunos casos, intervalos de confianza.

La dificultad de este método surge porque no siempre es posible obtener una cantidad pivotal con esas condiciones, que dé lugar a un intervalo de confianza[2].

4.2.2. MÉTODO GENERAL DE NEYMAN DE CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA

Ahora daremos un método general, debido a Neyman para la obtención de intervalos de confianza, el cual tiene menos limitaciones que el método pivotal, pues el método anterior requería encontrar una función de la muestra y del parámetro, cuya distribución fuese independiente del parámetro. No obstante, aplicando el método general, podremos obtener intervalos de confianza, sin necesidad de que exista tal función distribuida independientemente del parámetro.

Sea una población cuya función de densidad o de cuantía $f(x; \theta)$, en donde θ es un parámetro desconocido. Con la ayuda de una muestra aleatoria $(X_1, \ldots, X_n)$, obtenemos el estimador $\hat{\theta}(X_1, \ldots, X_n)$ (generalmente por el método de máxima verosimilitud) cuya función de densidad representamos por $g(\hat{\theta}, \theta)$ y pretendemos obtener un intervalo de confianza, del parámetro θ, al nivel de confianza del $100(1 - \alpha)\%$.

---

[2] Véase S. RÍOS (1977).

Para ese coeficiente de confianza $1 - \alpha$, determinaremos los extremos del intervalo $h_1(\theta_1)$ y $h_2(\theta)$, tales que:

$$P[h_1(\theta) \leqslant \hat{\theta} \leqslant h_2(\theta)] = \int_{h_1(\theta)}^{h_2(\theta)} g(\hat{\theta}, \theta)\, d\hat{\theta} = 1 - \alpha \qquad [4.4]$$

en donde suponemos que las funciones $h_1(\theta)$ y $h_2(\theta)$ son funciones continuas y monótonas de θ.

También se pueden determinar $h_1(\theta)$ y $h_2(\theta)$ de manera que:

$$\int_{-\infty}^{h_1(\theta)} g(\hat{\theta}, \theta)\, d\hat{\theta} = \alpha_1 \qquad [4.5]$$

$$\int_{h_2(\theta)}^{+\infty} g(\hat{\theta}, \theta)\, d\hat{\theta} = \alpha_2 \qquad [4.6]$$

en donde α_1 y α_2 son dos números cualesquiera[3], tales que

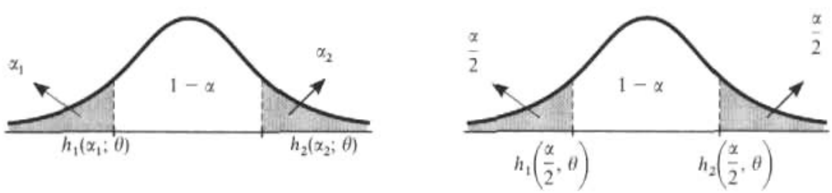
$$\alpha_1 + \alpha_2 = \alpha$$

Luego los valores de las funciones $h_1(\theta)$ y $h_2(\theta)$ para cualquier valor de θ se obtienen a partir de las expresiones [4.5] y [4.6], haciendo:

$$h_1(\theta) = \hat{\theta}$$
$$h_2(\theta) = \hat{\theta}$$

Y para una realización de una muestra concreta si el estimador toma el valor $\hat{\theta}_0$ y dado que las funciones $h_1(\theta)$ y $h_2(\theta)$, eran continuas y monótonas en θ,

[3] Habitualmente se hace $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, pues en muchas ocasiones coincide con el intervalo de menor longitud.



En todo el razonamiento hemos utilizado $h_1(\theta)$ y $h_2(\theta)$ en lugar de $h_1(\alpha_1; \theta)$ y $h_2(\alpha_2; \theta)$, pues facilita la notación.

podremos obtener los extremos inferior y superior del intervalo buscado, es decir:

$$h_1(\theta) = \hat{\theta}_0 \quad \Rightarrow \quad \underline{\theta}(x_1, ..., x_n) = h_1^{-1}(\hat{\theta}_0)$$
$$h_2(\theta) = \hat{\theta}_0 \quad \Rightarrow \quad \bar{\theta}(x_1, ..., x_n) = h_2^{-1}(\hat{\theta}_0)$$

y tendríamos el intervalo de confinza $(\underline{\theta}, \bar{\theta})$ al nivel de confianza del $100(1 - \alpha)\%$.

Veamos qué ocurre gráficamente. Una vez obtenidas las funciones

$$h_1(\theta) \quad \text{y} \quad h_2(\theta)$$

las representamos gráficamente como se indica en el Gráfico 4.2, y supongamos que para una muestra de tamaño n el valor que toma el estimador $\hat{\theta}$ es $\hat{\theta}_0$; por este punto $\hat{\theta}_0$ de ordenadas trazamos una paralela al eje de abscisas que cortará a las curvas $h_1(\theta)$ y $h_2(\theta)$ en los puntos A y B que pueden proyectarse sobre el eje de abscisas θ obteniendo los valores θ_1 y θ_2 que serán los extremos del intervalo de confianza $[\theta_1, \theta_2]$ al nivel de confianza del $100(1 - \alpha)\%$.

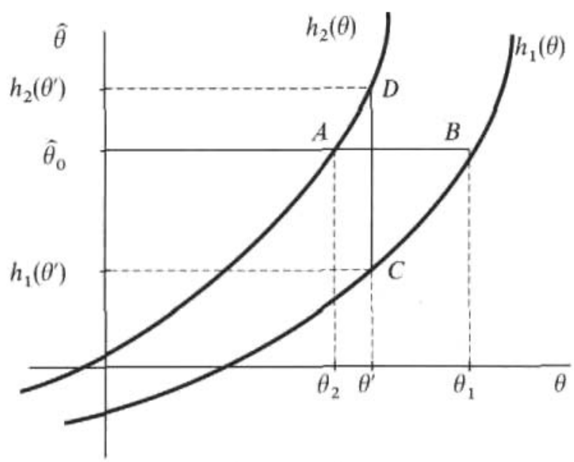


GRÁFICO 4.2.

Supongamos ahora que la muestra extraída procede de una población en que el verdadero valor del parámetro θ es θ', entonces la probabilidad de que la estimación $\hat{\theta}_0$, para esa muestra, esté comprendida entre $h_1(\theta')$ y $h_2(\theta')$ será $1 - \alpha$:

$$P[h_1(\theta') \leqslant \hat{\theta}_0 \leqslant h_2(\theta')] = 1 - \alpha$$

pero si la estimación $\hat{\theta}_0$ no cae entre $h_1(\theta')$ y $h_2(\theta')$, entonces la horizontal, AB, trazada por el correspondiente valor de $\hat{\theta}_0$ no cortará a la vertical CD, entre las curvas, y entonces el intervalo correspondiente $[\theta_1, \theta_2]$ no incluirá a θ', es decir si a cada valor del estimador $\hat{\theta}_0$ le hacemos corresponder una recta horizontal trazada por ese valor del estimador veremos que siempre que

$$h_1(\theta') < \hat{\theta}_0 < h_2(\theta')$$

la recta horizontal AB trazada por el valor del estimador $\hat{\theta}_0$ cortará a la vertical CD, trazada por θ', entre los puntos C y D, limitadas por ambas curvas, de manera que el segmento aleatorio AB cuya proyección es $[\theta_1, \theta_2]$, incluye el verdadero valor del parámetro θ'. En consecuencia, la confianza que tenemos de que el intervalo $[\theta_1, \theta_2]$, construido por este método, incluya a θ', será $1 - \alpha$.

Del gráfico 4.2 deducimos que los extremos del intervalo para el parámetro θ serán los puntos θ_1 y θ_2 tales que

$$h_1(\theta_1) = \hat{\theta}_0$$

$$h_2(\theta_2) = \hat{\theta}_0$$

y teniendo en cuenta las expresiones [4.5] y [4.6] diremos que θ_1 es el valor de θ para el cual

$$\int_{-\infty}^{\hat{\theta}_0} g(\hat{\theta}, \theta)\, d\hat{\theta} = \alpha_1$$

y θ_2 es el valor de θ para el cual

$$\int_{\hat{\theta}_0}^{\infty} g(\hat{\theta}, \theta)\, d\hat{\theta} = \alpha_2$$

Luego resolviendo estas ecuaciones resultará que las raíces serán los extremos del intervalo de confianza $[\theta_1, \theta_2]$ con un coeficiente de confianza del $1 - \alpha$.

Ejemplo 4.2

Dada una población $N(\mu, \sigma)$, con σ conocida. Obtener, aplicando el método general de Neyman, un intervalo de confianza para la media poblacional μ, con la ayuda de una muestra aleatoria de tamaño n, al nivel de confianza del $100(1-\alpha)\%$.

Solución:

El estimador insesgado de la media poblacional μ es la media muestral $\bar{X}$, es decir $\hat{\mu} = \bar{X}$.

Sabemos que este estimador $\bar{X}$ sigue una distribución normal

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

cuya función de densidad será:

$$g(\bar{X}, \mu) = \frac{1}{(\sigma/\sqrt{n})\ \sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(\bar{X}-\mu)^2}{\sigma^2/n}}$$

Aplicando el método general de Neyman, tendremos que obtener dos funciones $h_1(\mu)$ y $h_2(\mu)$ tales que

$$P[h_1(\mu) \leqslant \bar{X} \leqslant h_2(\mu)] = 1 - \alpha$$

o bien

$$\int_{-\infty}^{h_1(\mu)} \frac{1}{(\sigma/\sqrt{n})\ \sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(\bar{X}-\mu)^2}{\sigma^2/n}}\, d\bar{X} = \frac{\alpha}{2}$$

$$\int_{h_2(\mu)}^{+\infty} \frac{1}{(\sigma/\sqrt{n})\ \sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(\bar{X}-\mu)^2}{\sigma^2/n}}\, d\bar{X} = \frac{\alpha}{2}$$

Haciendo el cambio:

$$Y = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad , \quad dY = \frac{d\bar{X}}{\sigma/\sqrt{n}}$$

resultará que

$$Y \to N(0, 1)$$

y designando por:

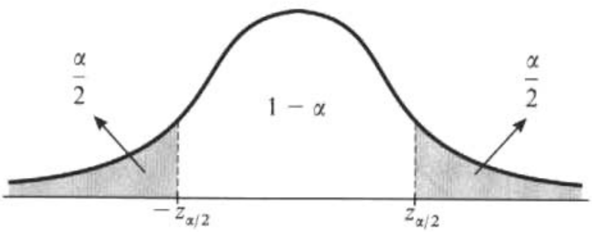
$$\lambda_1 = \frac{h_1(\mu) - \mu}{\dfrac{\sigma}{\sqrt{n}}} \quad , \quad \lambda_2 = \frac{h_2(\mu) - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

tendremos:

$$\int_{-\infty}^{\lambda_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \frac{\alpha}{2}$$

$$\int_{\lambda_2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \frac{\alpha}{2}$$

Teniendo en cuenta la simetría de la distribución normal, obtenemos un valor $z_{\alpha/2}$ tal que



en donde

$$\lambda_1 = -z_{\alpha/2}$$

$$\lambda_2 = \quad z_{\alpha/2}$$

y sustituyendo en las expresiones de $h_1(\mu)$ y $h_2(\mu)$:

$$h_1(\mu) = \mu + \lambda_1 \frac{\sigma}{\sqrt{n}} = \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$h_2(\mu) = \mu + \lambda_2 \frac{\sigma}{\sqrt{n}} = \mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Y considerando una muestra aleatoria de tamaño n, el estimador $\bar{x}_0$ del parámetro μ, tomara un valor, por ejemplo, $\bar{X}_0$, luego tenemos las ecuaciones:

$$\bar{x}_0 = \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\bar{x}_0 = \mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

que representan dos rectas paralelas a la bisectriz del primer cuadrante, siendo el intervalo de confianza:

$$I_\mu = \left[\bar{x}_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x}_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

En el Gráfico 4.3, tenemos la correspondiente representación.

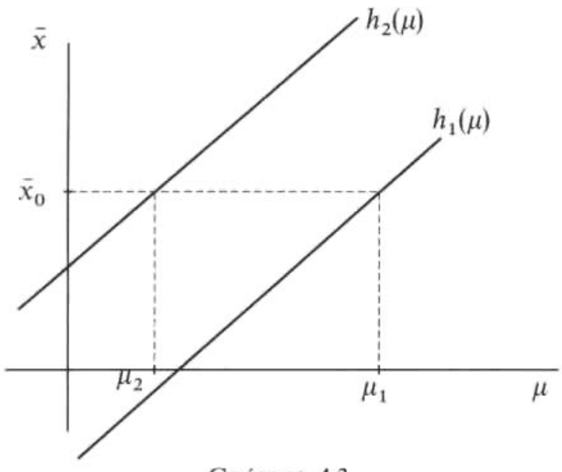


GRÁFICO 4.3.

Ejemplo 4.3

Sea una población cuya función de densidad viene dada por:

$$f(x; \theta) = \begin{cases} \frac{2}{\theta^2} (\theta - x) & , \quad 0 < x < \theta \\ 0 & , \quad \text{en el resto} \end{cases}$$

Utilizando el método general de Neyman y con una muestra de tamaño uno, obtener un intervalo de confianza para el parámetro poblacional θ al nivel de confianza del 95 %.

Solución:

Como el tamaño de muestra es uno, la función de verosimilitud coincide con la función de densidad, y entonces el estimador de máxima verosimilitud del parámetro θ será:

$$L(x;\, \theta) = \frac{2}{\theta^2}(\theta - x)$$

$$\ln L(x;\, \theta) = \ln 2 - \ln \theta^2 + \ln (\theta - x)$$

$$\frac{\partial \ln L(x;\ \theta)}{\partial \theta} = -\frac{2\theta}{\theta^2} + \frac{1}{\theta - x} = 0$$

$$-2\theta(\theta - x) + \theta^2 = 0$$

$$-\theta(\theta - 2x) = 0$$

Luego el estimador de máxima verosimilitud será:

$$\hat{\theta} = 2X$$

La función de densidad, $g(\hat{\theta}, \theta)$, del estimador se obtiene haciendo el cambio de variable:

$$x = \frac{1}{2}\hat{\theta}$$

y tendremos:

$$g(\hat{\theta}, \theta) = \frac{1}{2\theta^2}(2\theta - \hat{\theta}) \quad , \quad 0 < \hat{\theta} < 2\theta$$

Para obtener el intervalo de confianza al nivel de confianza del 95 %, obtendrermos $h_1(\theta)$ y $h_2(\theta)$ tales que:

$$\int_0^{h_1(\theta)} \frac{1}{2\theta^2}(2\theta - \hat{\theta})\, d\hat{\theta} = 0{,}025$$

$$\int_{h_2(\theta)}^{2\theta} \frac{1}{2\theta^2}(2\theta - \hat{\theta})\, d\hat{\theta} = 0{,}025$$

Integrando ambas expresiones:

$$\left[\frac{1}{2\theta^2}\left(2\theta\hat{\theta} - \frac{\hat{\theta}^2}{2}\right)\right]_0^{h_1(\theta)} = 0{,}025$$

$$\left[\frac{1}{2\theta^2}\left(2\theta\hat{\theta} - \frac{\hat{\theta}^2}{2}\right)\right]_{h_2(\theta)}^{2\theta} = 0{,}025$$

y resolviendo estas ecuaciones de segundo grado tenemos:

$$h_1(\theta) = 2(1 - \sqrt{0{,}975})\theta$$

$$h_2(\theta) = 2(1 - \sqrt{0{,}025})\theta$$

Como la muestra que consideramos es de tamaño $n = 1$, supongamos que la observación muestral ha sido, por ejemplo, $x = 3$, entonces:

$$x = 3 \quad \Rightarrow \quad \hat{\theta}_0 = 6$$

Sustituyendo en las expresiones de $h_1(\theta)$ y en la de $h_2(\theta)$ tendremos:

$$2(1 - \sqrt{0{,}975})\theta = 6 \quad \Rightarrow \quad \theta = \frac{3}{1 - \sqrt{0{,}975}}$$

$$2(1 - \sqrt{0{,}025})\theta = 6 \quad \Rightarrow \quad \theta = \frac{3}{1 - \sqrt{0{,}025}}$$

y el intervalo de confianza para el parámetro θ será:

$$\left[\frac{3}{1 - \sqrt{0{,}025}}; \frac{3}{1 - \sqrt{0{,}975}}\right]$$

Si hacemos la representación gráfica como aparece en el Gráfico 4.4 el intervalo de confianza se obtiene fácilmente.

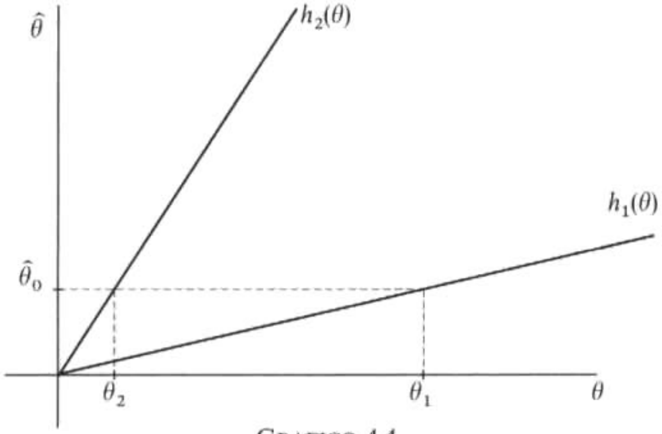


GRAFICO 4.4.

4.3. INTERVALOS DE CONFIANZA EN POBLACIONES NORMALES

En este apartado consideramos que la población de partida será normal y obtendremos intervalos de confianza para los parámetros poblaciones en el caso de una muestra y de dos muestras. Aplicaremos el método pivotal, pues en estos casos no existe gran dificultad para obtener una función del parámetro desconocido cuya distribución muestral no dependa del parámetro.

4.3.1. INTERVALO DE CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL

I. **Desviación típica σ conocida**[4].

Sea una población $N(\mu, \sigma)$, en donde el parámetro μ es desconocido y deseamos obtener un intervalo de confianza para el parámetro μ al nivel de confianza del $100(1 - \alpha)\%$.

Para ello tomamos una muestra aleatoria de tamaño n, $(X_1, ..., X_n)$, y buscaremos un estadístico (cantidad pivotal o pivote) que dependa del parámetro μ y de su estimador y cuya distribución muestral no dependa del parámetro μ. En este caso el estadístico será[5]:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to N(0, 1)$$

[4] Ya habiamos considerado este caso en el ejemplo 4.1, pero aquí lo veremos de forma más completa.

[5] El estadístico Z como función de μ, es monótona.

que se distribuye según una $N(0, 1)$, pues sabemos que el estadístico media muestral

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Como ya hemos indicado, utilizando la tabla de la distribución $N(0, 1)$, podemos encontrar dos valores λ_1 y λ_2 tales que:

$$P\left[\lambda_1 \leqslant \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \leqslant \lambda_2\right] = 1 - \alpha \qquad [4.7]$$

de donde se deduce:

$$P\left[\lambda_1 \frac{\sigma}{\sqrt{n}} \leqslant \bar{X} - \mu \leqslant \lambda_2 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[-\bar{X} + \lambda_1 \frac{\sigma}{\sqrt{n}} \leqslant -\mu \leqslant -\bar{X} + \lambda_2 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

multiplicando por -1

$$P\left[\bar{X} - \lambda_1 \frac{\sigma}{\sqrt{n}} \geqslant \mu \geqslant \bar{X} - \lambda_2 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

que es equivalente a

$$P\left[\bar{X} - \lambda_2 \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} - \lambda_1 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

y en consecuencia al intervalo:

$$\left[\bar{X} - \lambda_2 \frac{\sigma}{\sqrt{n}}; \bar{X} - \lambda_1 \frac{\sigma}{\sqrt{n}}\right] \qquad [4.8]$$

Pero la expresión [4.7] no quiere decir que λ_1 y λ_2 estén unívocamente determinado, sino que existen una infinidad de ellos. Entonces tendremos que elegir aquellos valores de λ_1 y λ_2 que hagan mínima la longitud del intervalo dado en la expresión [4.8], que será:

$$L = \left(\bar{X} - \lambda_1 \frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - \lambda_2 \frac{\sigma}{\sqrt{n}}\right) = \frac{\sigma}{\sqrt{n}} (\lambda_2 - \lambda_1)$$

sujeto a la condición dada en la expresión [4.7], que será:

$$P[\lambda_1 \leqslant Z \leqslant \lambda_2] = \int_{\lambda_1}^{\lambda_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1 - \alpha \qquad [4.9]$$

Teniendo en cuenta el método de los multiplicadores de Lagrange, tendremos que hacer mínima la expresión:

$$\phi = \frac{\sigma}{\sqrt{n}} (\lambda_2 - \lambda_1) + \gamma \left[\int_{\lambda_1}^{\lambda_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx - (1 - \alpha) \right] \qquad [4.10]$$

derivando respecto a λ_1 y λ_2 tendremos[6]:

$$\frac{\partial \phi}{\partial \lambda_1} = -\frac{\sigma}{\sqrt{n}} - \gamma \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda_1^2} = 0$$

$$\frac{\partial \phi}{\partial \lambda_2} = \frac{\sigma}{\sqrt{n}} + \gamma \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda_2^2} = 0$$

de donde se deduce:

$$e^{\frac{1}{2}\lambda_1^2} = e^{\frac{1}{2}\lambda_2^2} \quad \Rightarrow \quad \lambda_1^2 = \lambda_2^2$$

siendo las posibles soluciones:

$\lambda_1 = \lambda_2$, es inadmisible pues entonces la longitud del intervalo sería cero,

$\lambda_1 = -\lambda_2$, luego el intervalo de longitud mínima será simétrico en la $N(0, 1)$ y como según la expresión [4.9]

$$\int_{\lambda_1}^{\lambda_2} f(x)\,dx = 1 - \alpha \quad \Rightarrow \quad \begin{cases} \lambda_2 = z_{\alpha/2} \\ \lambda_1 = -z_{\alpha/2} \end{cases}$$

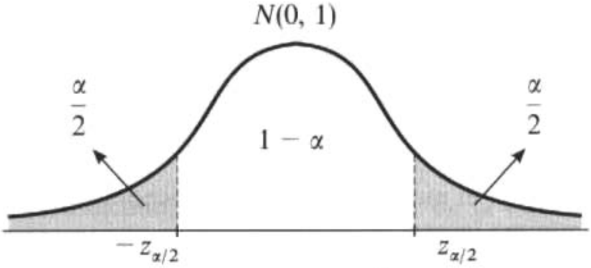
[6] Para derivar una integral respecto al límite superior de integración tendremos en cuenta:

$$\frac{\partial \int_{\lambda_1}^{\lambda_2} f(x)\,dx}{\partial \lambda_2} = f(\lambda_2)$$

Para el límite inferior λ_1, se cambian los límites cambiando de signo la integral, pues

$$\int_{\lambda_1}^{\lambda_2} f(x)\,dx = -\int_{\lambda_2}^{\lambda_1} f(x)\,dx$$

y se procede de la misma forma.



$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$$

Sustituyendo en el intervalo dado por la expresión [4.8], tendremos el intervalo de confianza para la media μ de una población $N(\mu, \sigma)$ con σ conocida:

$$I_\mu = \left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}};\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \qquad [4.11]$$

en donde los valores $z_{\alpha/2}$ se obtienen[7] de la $N(0, 1)$.

Intervalo de confianza para la media de una población normal, siendo σ conocida

Supongamos que tenemos una muestra aleatoria de n observaciones de una distribución $N(\mu, \sigma)$. Si σ es conocida, y la media muestral observada es $\bar{x}$, entonces el intervalo de confianza para la media poblacional μ, al nivel de confianza del $100(1 - \alpha)\%$ viene dado por[8]

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \qquad [4.12]$$

donde $z_{\alpha/2}$ es tal que

$$P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$$

y la variable aleatoria $Z \to N(0, 1)$.

[7] Los *intervalos unilaterales* vienen dadas por

$$\left(-\infty;\ \bar{X} + z_\alpha\frac{\sigma}{\sqrt{n}}\right] \quad \text{y} \quad \left[\bar{X} - z_\alpha\frac{\sigma}{\sqrt{n}};\ +\infty\right)$$

[8] En general los intervalos de confianza se deberían expresar siempre en la forma de la expresión [4.11] pero en algunas ocasiones y por dar mayor claridad se puede utilizar también la forma de la expresión [4.12].

Ejemplo 4.4

De una población $N(\mu, 6)$ se selecciona una muestra aleatoria cuya media es 25. Obtener un intervalo de confianza para la media poblacional μ.

1.° Cuando el tamaño de la muestra es $n = 16$ y el nivel de confianza es del 90 %.

2.° Igual pero con tamaño de muestra $n = 64$.

3.° Con tamaño de muestra $n = 16$, $1 - \alpha = 0{,}90$ pero $\sigma = 10$.

4.° Con tamaño de muestra $n = 16$, $1 - \alpha = 0{,}95$ y $\sigma = 6$.

Solución:

1.° La expresión [4.12] nos da el intervalo de confianza que nos piden:

$$\left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad ; \quad \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

siendo $\bar{x} = 25$, $\sigma = 6$, $n = 16$, $1 - \alpha = 0{,}90$

$$\frac{\alpha}{2} = 0{,}05 \quad , \quad z_{0{,}05} = 1{,}645$$

Luego el intervalo será:

$$\left[25 - 1{,}64 \frac{6}{\sqrt{16}} \quad ; \quad 25 + 1{,}64 \frac{6}{\sqrt{16}}\right]$$

$$[22{,}54 \quad ; \quad 27{,}46]$$

2.° Para: $\bar{x} = 25$, $\sigma = 6$, $n = 64$ y $1 - \alpha = 0{,}90$

$$\left[25 - 1{,}645 \frac{6}{\sqrt{64}} \quad ; \quad 25 + 1{,}645 \frac{6}{\sqrt{64}}\right]$$

$$[23{,}77 \quad 26{,}23]$$

3.º Para $\bar{x} = 25$, $\sigma = 10$, $n = 16$ y $1 - \alpha = 0{,}90$

$$\left[25 - 1{,}645\,\frac{10}{\sqrt{16}} \quad ; \quad 25 + 1{,}645\,\frac{10}{\sqrt{16}}\right]$$

$$[20{,}88 \quad ; \quad 29{,}11]$$

4.º Para $\bar{x} = 25$, $\sigma = 6$, $n = 16$ y $1 - \alpha = 0{,}95$

$$\left[25 - 1{,}96\,\frac{6}{\sqrt{16}} \quad ; \quad 25 + 1{,}96\,\frac{6}{\sqrt{16}}\right]$$

$$[22{,}06 \quad ; \quad 27{,}94]$$

Si representamos gráficamente los cuatro intervalos, Gráfico 4.5, se observa:

a) Cuando aumenta el tamaño de la muestra, disminuye la amplitud del intervalo y, por tanto, aumenta la precisión de la estimación por intervalo de confianza.

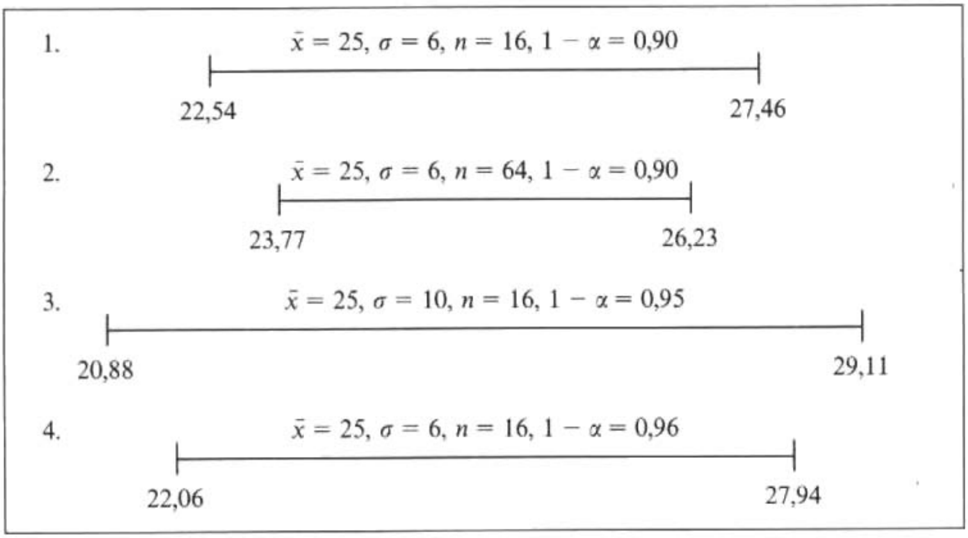


GRÁFICO 4.5. *Representación gráfica del efecto sobre la amplitud del intervalo de σ, n y $1 - \alpha$*

b) Cuando aumenta la desviación típica σ, aumenta la amplitud del intervalo y, por tanto, disminuye la precisión.

c) Cuando aumenta el nivel de confianza, aumenta la amplitud del intervalo y, por tanto, disminuye la precisión.

II. Desviación típica σ desconocida

Supongamos una población $N(\mu, \sigma)$, en donde μ y σ son desconocidos y deseamos obtener un intervalo de confianza para el parámetro μ, al nivel de confianza del $100(1 - \alpha)\%$.

Para ello se dispone de una muestra aleatoria de tamaño n, $(X_1, ..., X_n)$ y buscaremos un estadístico (cantidad pivotal o pivote) que dependa del parámetro μ y de su estimador, y cuya distribución muestral no dependa μ. Ese estadístico será[9]:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \to t_{n-1} \qquad [4.13]$$

que se distribuye según una t-Student con $n - 1$ grados de libertad, siendo S^2 la varianza muestral.

Utilizando la Tabla A.10, distribuión t-Student, del Anexo A de tablas estadísticas, podemos encontrar parejas de valores t_1 y t_2, tales que:

$$P\left[t_1 \leqslant \frac{\bar{X} - \mu}{S/\sqrt{n}} \leqslant t_2\right] = 1 - \alpha \qquad [4.14]$$

de donde se deduce que:

$$P\left[t_1 \frac{S}{\sqrt{n}} \leqslant \bar{X} - \mu \leqslant t_2 \frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[-\bar{X} + t_1 \frac{S}{\sqrt{n}} \leqslant -\mu \leqslant -\bar{X} + t_2 \frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[\bar{X} - t_2 \frac{S}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} - t_1 \frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

[9] Como se vio en el capítulo 1:

$$T = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} \to t_{n-1}$$

y

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

la varianza muestral.

e inicialmente resulta el intervalo:

$$\left[\bar{X} - t_2 \frac{S}{\sqrt{n}} \quad , \quad \bar{X} - t_1 \frac{S}{\sqrt{n}}\right] \qquad [4.15]$$

Pero igual que sucedía en el caso anterior, la expresión [4.14] no quiere decir que t_1 y t_2 estén unívocamente determinados, sino que existen una infinidad de ellos. Por tanto, tendremos que elegir aquellos valores de t_1 y t_2 que hagan mínima la longitud del intervalo dado en la expresión [4.15], que será:

$$L = \left(\bar{X} - t_1 \frac{S}{\sqrt{n}}\right) - \left(\bar{X} - t_2 \frac{S}{\sqrt{n}}\right) = (t_2 - t_1) \frac{S}{\sqrt{n}}$$

sujeto a la condición dada por [4.14] que también podemos expresarla, teniendo en cuenta la función de densidad de una t-Student con $n - 1$ grados de libertad, como:

$$P[t_1 \leqslant T \leqslant t_2] = \int_{t_1}^{t_2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{(n-1)\pi}} \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt$$

$$= k \int_{t_1}^{t_2} \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt = 1 - \alpha \qquad [4.16]$$

Aplicando el método de los multiplicadores de Lagrange, tendremos que hacer mínima la expresión:

$$\phi = (t_2 - t_1) \frac{S}{\sqrt{n}} + \gamma\left[k \int_{t_1}^{t_2} \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt - (1 - \alpha)\right] \qquad [4.17]$$

derivando respecto a t_1 y t_2 tenemos:

$$\frac{\partial \phi}{\partial t_1} = -\frac{S}{\sqrt{n}} - \gamma k\left(1 + \frac{t_1^2}{n-1}\right)^{-\frac{n}{2}} = 0$$

$$\frac{\partial \phi}{\partial t_2} = \frac{S}{\sqrt{n}} + \gamma k\left(1 + \frac{t_2^2}{n-1}\right)^{-\frac{n}{2}} = 0$$

de donde se deduce que:

$$\left(1 + \frac{t_1^2}{n-1}\right)^{-\frac{n}{2}} = \left(1 + \frac{t_2^2}{n-1}\right)^{-\frac{n}{2}}$$

$$t_1^2 = t_2^2$$

siendo las posibles soluciones:

$t_1 = t_2$, es inadmisible, pues, entonces el intervalo sería de longitud nula.

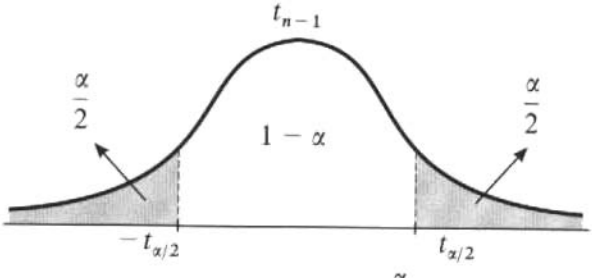
$t_1 = -t_2$, el intervalo de longitud mínima será simétrico en la t_{n-1}.

Luego haciendo:

$$t_2 = t_{\alpha/2}$$

tendremos:

$$t_1 = -t_{\alpha/2} \quad \text{y} \quad t_2 = t_{\alpha/2}$$



$$P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$$

Sustituyendo en el intervalo dado por la expresión [4.15], tenemos el intervalo de confianza para la media μ de una población $N(\mu, \sigma)$ con σ desconocida:

$$I_\mu = \left[\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \quad ; \quad \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad [4.18]$$

en donde los valores $t_{\alpha/2}$ se obtienen en la distribución t-Student con $n - 1$ grados de libertad[10].

Intervalo de confianza para la media de una población normal, siendo σ desconocida

Supongamos que tenemos una muestra aleatoria de n observaciones de una distribución $N(\mu, \sigma)$. Si σ es desconocida, y la media y la desviación típica muestral observadas son $\bar{x}$ y s, respectivamente, entonces el intervalo de confianza para la media poblacional μ, al nivel de confianza del $100(1 - \alpha)\%$ viene dado por:

$$\left[\bar{x} - t_{\alpha/2} \frac{S}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2} \frac{S}{\sqrt{n}}\right] \qquad [4.19]$$

donde $t_{\alpha/2}$ es tal que

$$P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$$

y la variable t_{n-1} sigue una distribución t-Student con $n - 1$ grados de libertad.

Ejemplo 4.5

Un fabricante de una determinada marca de vehículos de lujo sabe que el consumo de gasolina de sus vehículos se distribuye normalmente. Se selecciona una muestra aleatoria de 6 coches y se observa el consumo cada 100 km, obteniendo las siguientes observaciones

19,2 , 19,4 , 18,4 , 18,6 , 20,5 , 20,8

Obtener los intervalos de confianza para el consumo medio de gasolina de todos los vehículos de esa marca, a los niveles de confianza del 90, 95 y 99 %.

[10] Los *intervalos unilaterales* nos vienen dados por:

$$\left(-\infty; \bar{X} + t_{\alpha} \frac{S}{\sqrt{n}}\right] \text{ y } \left[\bar{X} - t_{\alpha} \frac{S}{\sqrt{n}}; +\infty\right)$$

Solución:

Con los datos de la muestra obtendremos la media y la varianza muestral

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}(19{,}2 + 19{,}4 + 18{,}4 + 18{,}6 + 20{,}5 + 20{,}8)$$

$$= \frac{116{,}9}{6} = 19{,}48$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} \bar{x}_i^2 - n\bar{x}^2\right)$$

$$= \frac{1}{5}[2.282{,}4 - 6\cdot(19{,}48)^2]$$

$$= 1{,}12$$

$$s = \sqrt{1{,}12} \simeq 1{,}06$$

El intervalo de confianza para la media poblacional cuando σ es desconocida tiene la forma dada por la expresión [4.19]

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

en donde $t_{\alpha/2}$ es tal que en la t-student con 5 grados de libertad se verifica:

$$P[t_5 > t_{\alpha/2}] = \frac{\alpha}{2}$$

Para $1 - \alpha = 0{,}90$, utilizando la Tabla A.10 del Anexo A de tablas estadísticas, que corresponde a la t-Student, tenemos:

$$P[t_5 > t_{0,05}] = 0{,}05 \quad \Rightarrow \quad t_{0,05} = 2{,}015$$

Para $1 - \alpha = 0{,}95$

$$P[t_5 > t_{0,025}] = 0{,}025 \quad \Rightarrow \quad t_{0,025} = 2{,}571$$

Para $1 - \alpha = 0{,}99$

$$P[t_5 > t_{0,005}] = 0{,}005 \quad \Rightarrow \quad t_{0,005} = 4{,}032$$

Y los intervalos de confianza serán:

Para $1 - \alpha = 0{,}90$

$$\left[19{,}48 - 2{,}015\,\frac{1{,}06}{\sqrt{6}} \quad ; \quad 19{,}48 + 2{,}015\,\frac{1{,}06}{\sqrt{6}}\right]$$

$$[18{,}61 \quad ; \quad 20{,}35]$$

Para $1 - \alpha = 0{,}95$

$$\left[19{,}48 - 2{,}571\,\frac{1{,}06}{\sqrt{6}} \quad ; \quad 19{,}48 + 2{,}571\,\frac{1{,}06}{\sqrt{6}}\right]$$

$$[18{,}37 \quad ; \quad 20{,}59]$$

Para $1 - \alpha = 0{,}99$

$$\left[19{,}48 - 4{,}032\,\frac{1{,}06}{\sqrt{6}} \quad ; \quad 19{,}48 + 4{,}032\,\frac{1{,}06}{\sqrt{6}}\right]$$

$$[17{,}74 \quad ; \quad 21{,}22]$$

Si representamos gráficamente los tres intervalos, Gráfico 4.6, vemos como, efectivamente, cuando aumenta el nivel de confianza aumenta la amplitud del intervalo.

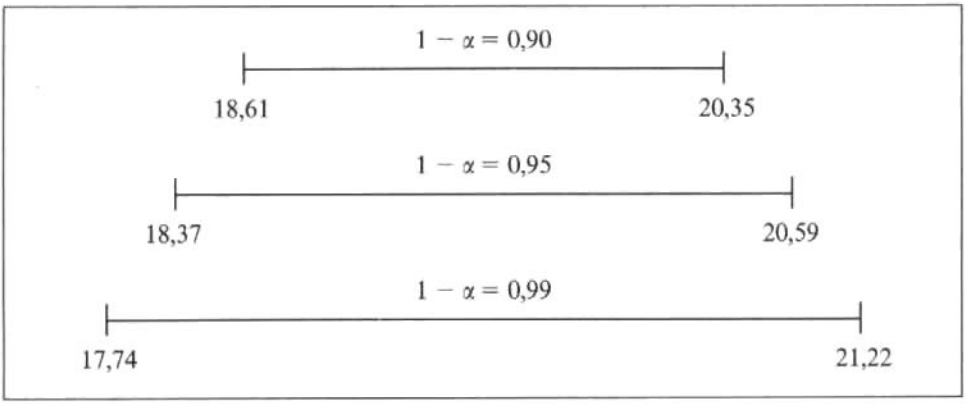


GRÁFICO 4.6. *Representación gráfica de los intervalos de confianza del ejemplo 4.5.*

4.3.2. INTERVALO DE CONFIANZA PARA LA VARIANZA DE UNA POBLACIÓN NORMAL

I.˙ Media poblacional μ desconocida

Supongamos una población $N(\mu, \sigma)$, en donde μ y σ son desconocidos y deseamos obtener un intervalo de confianza para la varianza poblacional σ^2 al nivel de confianza del $100(1 - \alpha)\%$. Para ello tomamos una muestra aleatoria de tamaño n, $(X_1, ..., X_n)$ y utilizaremos un estadístico (cantidad pivotal o pivote) que dependa del parámetro σ^2 y de su estimador y cuya distribución muestral no dependa de σ^2. Ese estadístico será[11]:

$$\frac{(n-1)S^2}{\sigma^2} \to \chi^2_{n-1}$$

que se distribuye según una χ^2 de Pearson con $n - 1$ grados de libertad, siendo S^2 la varianza muestral.

En la distribución χ^2_{n-1} podemos obtener parejas de valores k_1 y k_2 tales que

$$P\left[k_1 \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant k_2\right] = 1 - \alpha$$

pero estos valores k_1 y k_2 tendríamos que determinarlos de manera que el intervalo fuera de longitud mínima, pero como la distribución χ^2 no es simétrica, resulta que los extremos del intervalo dependerán de los grados de libertad, y con el fin de simplificar y poder llegar a un intervalo único adoptamos el criterio de considerar la misma probabilidad en los dos extremos, es decir:

$$\alpha_1 = \frac{\alpha}{2} \qquad \alpha_2 = \frac{\alpha}{2}$$

[11] En el capítulo 1 habíamos estudiado que:

$$\frac{(n-1)S^2}{\sigma^2} \to \chi^2_{n-1}$$

siendo

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

la varianza muestral.

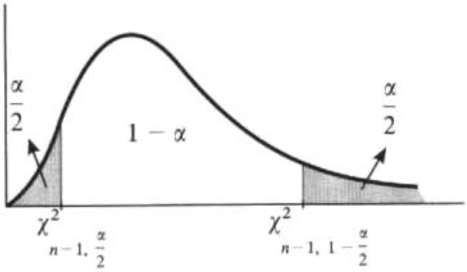


GRÁFICO 4.7. *Distribución* χ^2_{n-1}.

Teniendo en cuenta la forma de la distribución χ^2, Gráfico 4.7, y la tabla A.9 del Anexo A de tablas estadísticas, en donde nos aparece tabulada la distribución χ^2, podemos considerar que el intervalo de confianza del estadístico será:

$$P\left[\chi^2_{n-1,\ \alpha/2} \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant \chi^2_{n-1,\ 1-\alpha/2}\right] = 1-\alpha \qquad [4.20]$$

o bien

$$P\left[\frac{\chi^2_{n-1,\ \alpha/2}}{(n-1)S^2} \leqslant \frac{1}{\sigma^2} \leqslant \frac{\chi^2_{n-1,\ 1-\alpha/2}}{(n-1)S^2}\right] = 1-\alpha$$

Reordenando esta expresión, se tiene:

$$P\left[\frac{(n-1)S^2}{\chi^2_{n-1,\ 1-\alpha/2}} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{\chi^2_{n-1,\ \alpha/2}}\right] = 1-\alpha$$

y el intervalo de confianza para σ^2 al nivel de confianza del $(1-\alpha)\%$ sería:

$$I_{\sigma^2} = \left[\frac{(n-1)S^2}{\chi^2_{n-1,\ 1-\alpha/2}}\ ;\ \frac{(n-1)S^2}{\chi^2_{n-1,\ \alpha/2}}\right] \qquad [4.21]$$

en donde los valores $\chi^2_{n-1,\ 1-\alpha/2}$ y $\chi^2_{n-1,\ \alpha/2}$ se obtienen en la distribución χ^2 con $n-1$ grados de libertad[12].

[12] Los *intervalos unilaterales* vienen dados por

$$\left(0, \frac{(n-1)S^2}{\chi^2_{n-1,\ \alpha}}\right] \quad \text{y} \quad \left[\frac{(n-1)S^2}{\chi^2_{n-1,\ 1-\alpha}}, \infty\right)$$

Intervalo de confianza para la varianza de una población normal

Supongamos una muestra aleatoria de n observaciones de una distribución $N(\mu, \sigma)$. Si σ es desconocida y la varianza muestral observada es s^2, entonces el intervalo de confianza para la varianza poblacional σ^2, al nivel de confianza del $100(1 - \alpha)\%$ viene dado por:

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}} \leqslant \sigma^2 \leqslant \frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}}\right] \qquad [4.22]$$

donde $\chi^2_{n-1,\,1-\alpha/2}$ es tal que:

$$P[\chi^2_{n-1} \leqslant \chi^2_{n-1,\,1-\alpha/2}] = 1 - \frac{\alpha}{2}$$

y $\chi^2_{n-1,\,\alpha/2}$ es tal que:

$$P[\chi^2_{n-1} \leqslant \chi^2_{n-1,\,\alpha/2}] = \frac{\alpha}{2}$$

y la variable aleatoria χ^2_{n-1} sigue una distribución χ^2 de Pearson con $n - 1$ grados de libertad.

II. **Media poblacional μ conocida**

En este caso tal estadístico (cantidad pivotal o pivote) que dependa del parámetro σ^2 y cuya distribución muestral no dependa de σ^2 será:

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} \to \chi^2_n$$

que para cada valor fijo de σ^2 sigue una distribución χ^2 de Pearson con n grados de libertad, pues al ser la media μ conocida no hay que estimarla y el número de grados de libertad es n.

Razonando análogamente al caso anterior, en donde μ era desconocida, llegamos a obtener el intervalo de confianza:

$$I_{\sigma^2} = \left[\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{n,\,1-\alpha/2}} \; ; \; \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{n-1,\,\alpha/2}}\right] \qquad [4.23]$$

Ejemplo 4.6

El precio de un determinado artículo perecedero en los comercios de alimentación de una ciudad sigue una distribución normal. Se toma una muestra aleatoria de 8 comercios y se observa el precio de ese artículo, obteniendo las siguientes observaciones:

132 , 125 , 130 , 139 , 126 , 138 , 124 , 140

Obtener al nivel de confianza del 95 %.

1. Un intervalo de confianza para la media poblacional.
2. Un intervalo de confianza para la varianza poblacional.

Solución:

A partir de las observaciones muestrales podemos calcular la media y la varianza muestral:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1.054}{8} = 131,75$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{301,5}{7} = 43,07$$

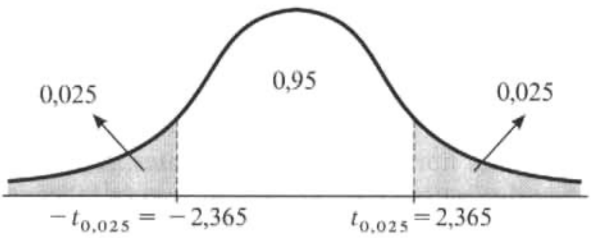
$$s = \sqrt{43,07} = 6,56$$

1. El intervalo de confianza para la media poblacional cuando σ es desconocido y $1 - \alpha = 0,95$ viene dado por:

$$\left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \quad ; \quad \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

$$\left[131,75 - t_{0,025}\frac{6,56}{\sqrt{8}} \quad ; \quad 131,75 + t_{0,025}\frac{6,56}{\sqrt{8}}\right]$$

En la tabla A.10 de la distribución t-Student para 7 grados de libertad, obtenemos los valores $t_{\alpha/2}$:



Sustituyendo en la expresión anterior:

$$\left[131,75 - 2,365\,\frac{6,56}{\sqrt{8}} \quad ; \quad 131,75 + 2,365\,\frac{6,56}{\sqrt{8}}\right]$$

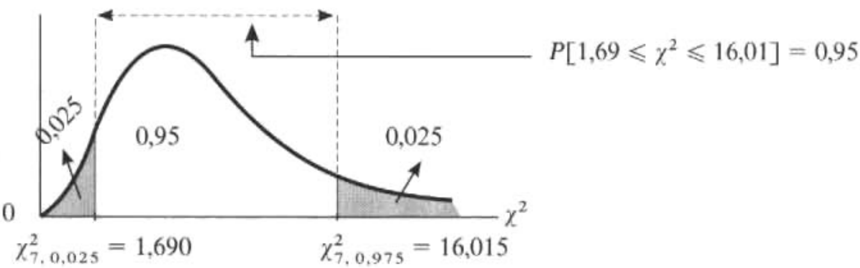
$$[126,25 \quad ; \quad 137,23]$$

2. El intervalo de confianza para la varianza σ^2 es:

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}} \quad ; \quad \frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}}\right]$$

En la tabla A.9 de la distribución χ^2 con 7 grados de libertad obtenemos los valores:

$$\chi^2_{n-1,\,\alpha/2} \equiv \chi^2_{7,\,0,025} \quad \text{y} \quad \chi^2_{n-1,\,1-\alpha/2} \equiv \chi^2_{7,\,0,975}$$



Sustituyendo en la expresión del intervalo:

$$\left[\frac{7 \cdot 43,07}{16,01} \quad ; \quad \frac{7 \cdot 43,07}{1,690}\right]$$

$$[18,83 \quad ; \quad 178,39]$$

4.3.3. INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS EN POBLACIONES NORMALES: MUESTRAS INDEPENDIENTES

Sean dos poblaciones normales e independientes, $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$, dfe las cuales se extraen dos muestras aleatorias, que serán independientes entre sí, $(X_1, ..., X_{n_x})$ e $(Y_1, ..., Y_{n_y})$, respectivamente. Siendo los estadísticos muestrales correspondientes:

$$\bar{X} = \frac{1}{n_x} \sum_{i=1}^{n_x} X_i \qquad S_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (X_i - \bar{X})^2$$

$$\bar{Y} = \frac{1}{n_y} \sum_{i=1}^{n_y} Y_i \qquad S_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (Y_i - \bar{Y})^2$$

Consideremos diferentes situaciones:

Medias desconocidas y desviaciones típicas diferentes pero conocidas

Para obtener un intervalo de confianza para la diferencia de medias poblacionales $\mu_x - \mu_y$ al nivel de confianza del $100(1 - \alpha)\%$ se toman dos muestras independientes de tamaños n_x y n_y de las poblaciones $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$, respectivamente. Por el teorema 1.7 sabemos que el estadístico:

$$\bar{X} - \bar{Y} \to N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)$$

entonces el estadístico (cantidad pivotal o pivote) que depende de los parámetros μ_x y μ_y y de sus estimadores y cuya distribución muestral no depende de los parámetros será:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \to N(0, 1)$$

Razonando de manera análoga al apartado 4.4.1, tendríamos el intervalo:

$$I_{\mu_x - \mu_y} = \left[(\bar{X} - \bar{Y}) - z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \ ; \ (\bar{X} - \bar{Y}) + z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right] \qquad [4.24]$$

en donde los valores $z_{\alpha/2}$ se obtienen en la tabla de $N(0, 1)$, de manera que

$$P(z > z_{\alpha/2}] = \frac{\alpha}{2}$$

Si los tamaños muestrales n_x y n_y son grandes, entonces una buena aproximación al intervalo de confianza para $\mu_x - \mu_y$ se obtiene reemplazando las varianzas poblacionales en la expresión [4.24] por las correspondientes varianzas muestrales observadas S_x^2 y S_y^2. Resultando que para muestras grandes, $n > 30$, esta aproximación será adecuada incluso si las distribuciones poblacionales no son normales.

Medias desconocidas y desviaciones típicas iguales y conocidas

El razonamiento es el mismo, únicamente tendremos que:

$$\sigma_x = \sigma_y = \sigma \text{ conocida}$$

con lo cual el intervalo resultante será:

$$I_{\mu_x - \mu_y} = \left[(\bar{X} - \bar{Y}) - z_{\alpha/2}\sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \quad ; \quad (\bar{X} - \bar{Y}) + z_{\alpha/2}\sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \right] \quad [4.25]$$

Medias desconocidas y desviaciones típicas iguales y desconocidas

En este caso las poblaciones normales de partida son:

$$N(\mu_x, \sigma) \quad \text{y} \quad N(\mu_y, \sigma)$$

es decir, $\sigma_x = \sigma_y = \sigma$.

Teniendo en cuenta el apartado 1.7.5, en donde estudiábamos la distribución de la diferencia de medias muestrales cuando no se conoce la varianza poblacional, expresión [1.30], aquí podemos utilizar como estadístico (cantidad pivotal o pivote) que dependa de los parámetros μ_x y μ_y y de sus estimadores cuya distribución muestral no dependa de ellos, el estadístico:

$$T = \frac{\dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}}}{\sqrt{\dfrac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{\sigma^2} \Big/ (n_x - n_y - 2)}} =$$

$$= \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \to t_{n_x + n_y - 2} \qquad [4.26]$$

que sigue una distribución t-Student con $n_x + n_y - 2$ grados de libertad.

Utilizando este estadístico T podemos escribir:

$$P[-t_{\alpha/2} \leqslant T \leqslant t_{\alpha/2}] = 1 - \alpha$$

$$P\left[-t_{\alpha/2} \leqslant \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \leqslant t_{\alpha/2}\right] = 1 - \alpha$$

Multiplicando cada término de la desigualdad por el denominador del término intermedio, restando $(\bar{X} - \bar{Y})$ a cada término y multiplicando por -1, se tiene:

$$P\left[(\bar{X} - \bar{Y}) - t_{\alpha/2} \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$$

$$\left. \leqslant (\bar{X} - \bar{Y}) + t_{\alpha/2} \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\right] = 1 - \alpha \qquad [4.27]$$

Por tanto el intervalo de confianza al nivel de confianza del $100(1 - \alpha)\%$ como:

$$\left[(\bar{X} - \bar{Y}) - t_{\alpha/2} \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}} \right. ;$$

$$\left. (\bar{X} - \bar{Y}) + t_{\alpha/2} \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}} \right] \qquad [4.28]$$

siendo $t_{\alpha/2}$ tal que

$$P[t_{n_x + n_y - 2} > t_{\alpha/2}] = \frac{\alpha}{2}$$

Medias desconocidas y desviaciones típicas distintas y desconocidas

Las poblaciones normales que consideramos serán:

$$N(\mu_x, \sigma_x) \quad \text{y} \quad N(\mu_y, \sigma_y)$$

siendo $\sigma_x \neq \sigma_y$.

El estadístico (cantidad pivotal o pivote) que dependa de los parámetros μ_x y μ_y y de sus estimadores y cuya distribución muestral no dependa de ellos, será el estadístico

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \to t_v \qquad [4.29]$$

que según la expresión [1.32] y [1.33] sigue una distribución t-Student con v grados de libertad, siendo

$$v = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{(S_x^2/n_x)^2}{n_x - 1} + \frac{(S_y^2/n_y)^2}{n_x - 1}} \qquad [4.30]$$

y tomaremos como valor de v el valor anterior más próximo.

Utilizando este estadístico T y procediendo igual que en el caso anterior, llegaríamos a obtener el intervalo de confianza al nivel del $100(1 - \alpha)\%$ para la diferencia de medias $\mu_x - \mu_y$:

$$I_{\mu_x - \mu_y} = \left[(\bar{X} - \bar{Y}) - t_{\alpha/2}\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} \;\; ; \;\; (\bar{X} - \bar{Y}) + t_{\alpha/2}\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}\right] \qquad [4.31]$$

siendo $t_{\alpha/2}$ el valor tal que:

$$P[t_v > t_{\alpha/2}] = \frac{\alpha}{2}$$

Intervalo de confianza para la diferencia de medias en poblaciones normales independientes

Desviaciones típicas diferentes pero conocidas: $\sigma_x \neq \sigma_y$ conocidas

Supongamos dos muestras independientes de tamaño n_x y n_y procedentes de poblaciones normales $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$, respectivamente. Si las medias para las muestras observadas son $\bar{x}$ e $\bar{y}$, entonces un intervalo de confianza, al nivel de confianza del $100(1 - \alpha)\%$, para la diferencia de medias poblacionales $\mu_x - \mu_y$ viene dado por:

$$\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right] \quad [4.32]$$

en donde $z_{\alpha/2}$ es el número tal que: $P[Z > z_{\alpha/2}] = \dfrac{\alpha}{2}$ y la variable aleatoria Z sigue una $N(0, 1)$.

Desviaciones típicas iguales y conocidas: $\sigma_x = \sigma_y = \sigma$ conocida

$$\left[[\bar{x} - \bar{y}) - z_{\alpha/2}\sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \leqslant \mu_x - \mu_y \leqslant (\bar{x} - \bar{y}) + z_{\alpha/2}\sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_x}}\right] \quad [4.33]$$

Desviaciones típicas iguales y desconocidas: $\sigma_x = \sigma_y = \sigma$ desconocida

$$\left[(\bar{x} - \bar{y}) - t_{\alpha/2}\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}\sqrt{\frac{n_x + n_y}{n_x n_y}} \leqslant \mu_x - \mu_y \leqslant\right.$$

$$\left.\leqslant (\bar{x} - \bar{y}) + t_{\alpha/2}\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}\sqrt{\frac{n_x + n_y}{n_x n_y}}\right] \quad [4.34]$$

siendo $t_{\alpha/2}$ el número tal que: $P[t_{n_x + n_y - 2} > t_{\alpha/2}] = \dfrac{\alpha}{2}$.

Desviaciones típicas distintas y desconocidas: $\sigma_x \neq \sigma_y$ desconocida

$$\left[(\bar{x} - \bar{y}) - t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant (\bar{x} - \bar{y}) + t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right] \quad [4.35]$$

en donde $t_{\alpha/2}$ es el número tal que: $P[t_v > t_{\alpha/2}] = \dfrac{\alpha}{2}$ y v viene dado por [4.30].

Ejemplo 4.7

En un estudio sobre los préstamos realizados por dos entidades financieras a sus clientes se toma una muestra aleatoria de 6 préstamos de la primera entidad observando que el importe medio es de 9.972.000 ptas. y una desviación típica de 7.470.000 ptas. Una muestra aleatoria, independiente de la anterior, de tamaño 9, de préstamos de la segunda entidad muestra un importe medio de 2.098.000 ptas. y una desviación típica de 10.834.000 ptas. Admitiendo que las dos distribuciones poblacionales de préstamos son normales con la misma varianza, obtener al nivel del 95 % un intervalo de confianza para la diferencia entre sus medias poblacionales.

Solución:

Se trata de obtener un intervalo de confianza para la diferencia de medias poblacionales cuando las varianzas poblaciones son iguales pero desconocidas. Utilizaremos para ello la expresión [4.33]:

$$\left[(\bar{x} - \bar{y}) - t_{\alpha/2} \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$$

$$\left. \leqslant (\bar{x} - \bar{y}) + t_{\alpha/2} \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}} \right]$$

$$n_x = 6 \qquad \bar{x} = 9.972.000 \qquad s_x = 7.470.000$$

$$n_y = 9 \qquad \bar{y} = 2.098.000 \qquad s_y = 10.834.000$$

$$n_x + n_y - 2 = 6 + 9 - 2 = 13$$

Utilizando la tabla A.10 del Anexo de tablas estadísticas, se tiene:

$$P[t_{13} > t_{0,025}] = 0{,}025$$

$$t_{0,025} = 2{,}160$$

Para simplificar los cálculos utilizaremos las cantidades en miles de pesetas.

Sustituyendo en la expresión que nos da el intervalo de confianza tenemos:

$$\left[(9.972 - 2.098) - (2{,}160) \sqrt{\frac{(6-1)(7.470)^2 + (9-1)(10.834)^2}{6+9-2}} \sqrt{\frac{6+9}{54}} \right. ;$$

$$\left. (9.972 - 2.098) + (2{,}160) \sqrt{\frac{(6-1)(7.470)^2 + (9-1)(10.834)^2}{6+9-2}} \sqrt{\frac{6+9}{54}} \right]$$

$$\left[7.874 - (2{,}160)(9.680) \sqrt{\frac{6+9}{54}} \quad ; \quad 7.874 + (2{,}160)(9.680) \sqrt{\frac{6+9}{54}} \right]$$

$$[-3.146 \quad ; \quad 18.893]$$

que sería el intervalo de confianza al nivel de confianza del 95 %.

Observemos que este intervalo incluye el cero, lo cual podemos interpretarlo como que no existe diferencia significativa entre los importes medios de los préstamos de ambas entidades financieras al 95 % de confianza.

Ejemplo 4.8

Supongamos que una máquina automática de envasado de un producto químico está preparada para depositar 8 c.c. de producto en cada frasco de la cadena de envasado. Antes de proceder a una revisión y ajuste de la máquina se toma una muestra aleatoria de 4 frascos observando que la cantidad, medica en c.c., depositada de producto químico en cada frasco ha sido:

$$6{,}88 \quad , \quad 7{,}34 \quad , \quad 6{,}92 \quad , \quad 7{,}26$$

Después de revisada y ajustada la máquina se vuelve a tomar otra muestra aleatoria de 5 frascos, observando que las cantidades depositadas de producto químico han sido:

$$7{,}10 \quad , \quad 7{,}65 \quad , \quad 7{,}33 \quad , \quad 7{,}93 \quad , \quad 7{,}49$$

Suponemos que las distribuciones del contenido de producto químico en los frascos son normales y que la varianza poblacional varía cuando la máquina se revisa y se ajusta.

Obtener un intervalo de confianza al nivel de confianza del 98 % para la diferencia de las medias poblacionales.

Solución:

Se trata de obtener un intervalo de confianza para la diferencia de medias poblacionales cuando las varianzas son distintas y desconocidas. Utilizaremos la expresión [4.35]:

$$\left[(\bar{x} - \bar{y}) - t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant (\bar{x} - \bar{y}) + t_{\alpha/2}\sqrt{\frac{s^2}{n_x} + \frac{s_y^2}{n_y}}\right]$$

siento $t_{\alpha/2}$ tal que

$$P[t_v > t_{\alpha/2}] = \frac{\alpha}{2}$$

en donde v viene dado por la expresión [4.30].

A partir de las muestras podemos obtener:

$$\bar{x} = \frac{1}{3}(6{,}88 + 7{,}34 + 6{,}92 + 7{,}26) = 7{,}1$$

$$s_x^2 = \frac{1}{4}[(6{,}88 - 7{,}10)^2 + (7{,}34 - 7{,}10)^2 + (6{,}92 - 7{,}10)^2 + (7{,}26 - 7{,}10)^2] =$$

$$= 0{,}0546$$

$$\bar{y} = \frac{1}{5}(7{,}10 + 7{,}65 + 7{,}33 + 7{,}93 + 7{,}49) = 7{,}5$$

$$s_y^2 = \frac{1}{4}[(7{,}10 - 7{,}5)^2 + (7{,}65 - 7{,}5)^2 + (7{,}33 - 7{,}5)^2 + (7{,}93 - 7{,}5)^2 +$$

$$+ (7{,}49 - 7{,}5)^2] = 0{,}099$$

$$\frac{s_x^2}{n_x} = \frac{0{,}0546}{4} = 0{,}0137$$

$$\frac{s_y^2}{n_y} = \frac{0{,}099}{5} = 0{,}0198$$

El número de grados de libertad v se obtiene sustituyendo en la expresión [4.29]:

$$v = \frac{(0{,}0137 + 0{,}0198)^2}{\dfrac{(0{,}0137)^2}{3} + \dfrac{(0{,}0198)^2}{4}} = 7{,}7 \simeq 8$$

Utilizando la tabla A.10, del Anexo A de tablas estadísticas, tendremos el correspondiente valor de la t-Student:

$$P[t_8 > t_{0,01}] = 0{,}01$$

$$t_{0,01} = 2{,}896$$

El intervalo de confianza será:

$$\left[(7{,}1 - 7{,}5) - (2{,}896)\sqrt{\frac{0{,}0546}{4} + \frac{0{,}099}{5}};\ (7{,}1 - 7{,}5) + (2{,}896)\sqrt{\frac{0{,}0546}{4} + \frac{0{,}099}{5}}\right]$$

$$[-0{,}929 \quad ; \quad 0{,}129]$$

4.3.4. INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS EN POBLACIONES NORMALES: MUESTRAS APAREADAS

Ahora tratamos de obtener un intervalo de confianza para la diferencia de dos medias cuando las muestras extraídas de las poblaciones normales no son independientes y las varianzas poblacionales no tienen porqué ser iguales. Es decir, supongamos que obtenemos una muestra aleatoria de n pares de observaciones $(X_1, Y_1) \cdots (X_n, Y_n)$ de poblaciones normales[13] con medias μ_x y μ_y, en donde $(X_1, ..., X_n)$ indica la muestra de la población con media μ_x y $(Y_1, ..., X_n)$ indica la muestra de la población con media μ_y.

En este caso podemos reducir la información a una sola muestra $(D_1, ..., D_n)$ en donde:

$$D_i = X_i - Y_i \quad , \quad i = 1, ..., n$$

[13] Se admite que (X, Y) sigue una distribución normal bivariante.

y por las propiedades de la distribución normal, esta muestra $(D_1, ..., D_n)$ procederá también de una población normal de media:

$$\mu_D = E[D] = E[X_i - Y_i] = E[X_i] - E[Y_i] = \mu_x - \mu_y$$

y varianza desconocida, σ_D^2.

La varianza poblacional, σ_D^2, se puede estimar por la varianza muestral S_d^2 que sería la varianza de las diferencias que constituyen la muestra:

$$S_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2$$

siendo

$$\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i$$

Un estimador puntual de la media poblacional de las diferencias:

$$\mu_D = \mu_x - \mu_y$$

viene dado por $\bar{D}$.

Como la varianza poblacional es desconocida y pretendemos obtener un intervalo de confianza, al nivel de confianza del $100(1 - \alpha)\%$, para μ_D procederemos de manera análoga al apartado 4.4.1 referente al intervalo de confianza para la media de una población normal cuando σ era desconocida. Así pues, buscaremos un estadístico (cantidad pivotal o pivote) que depende del parámetro μ_D y de su estimador y cuya distribución muestral no depende de μ_0. Ese estadístico será:

$$T = \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}} \to t_{n-1}$$

que se distribuye según una t-Student con $n - 1$ grados de libertad.

El intervalo de confianza se obtendrá como sigue:

$$P\left[-t_{\alpha/2} \leqslant \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}} \leqslant t_{\alpha/2}\right] = 1 - \alpha$$

de donde se tiene:

$$P\left[\bar{D} - t_{\alpha/2}\frac{S_d}{\sqrt{n}} \leqslant \mu_D \leqslant \bar{D} + t_{\alpha/2}\frac{S_d}{\sqrt{n}}\right] = 1 - \alpha$$

o bien

$$\left[\bar{D} - t_{\alpha/2}\frac{S_d}{\sqrt{n}} \quad ; \quad \bar{D} + t_{\alpha/2}\frac{S_d}{\sqrt{n}}\right] \qquad [4.36]$$

siendo $t_{\alpha/2}$ tal que se verifica:

$$P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$$

Intervalo de confianza para la diferencia de medias en poblaciones normales apareadas

Supongamos que tenemos una muestra de n pares de observaciones de las poblaciones con media μ_x y μ_y. Sea $\bar{d}$ y s_d la media y desviación típica muestral de las n diferencias $d_i = x_i - y_i$. Si la distribución poblacional de las diferencias es normal, entonces el intervalo de confianza al nivel de confianza del $100(1 - \alpha)\%$ para $\mu_D = \mu_x - \mu_y$ viene dado por:

$$\left[\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} \leqslant \mu_D \leqslant \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}\right] \qquad [4.37]$$

donde $t_{\alpha/2}$ es tal que

$$P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$$

y la variable t_{n-1} sigue una distribución t-Student con $n - 1$ grados de libertad.

Ejemplo 4.9

La tabla 4.1 muestra el consumo de gasolina por 1.000 km de una muestra aleatoria de 9 coches con dos carburantes X e Y. Si admitimos que los consumos de gasolina se distribuyen normalmente, obtener un intervalo de confianza al nivel de confianza del 99 % para la diferencia de las medias poblacionales.

TABLA 4.1. *Consumo de gasolina por 1.000 km, para los modelos X e Y.*

	Modelo X	**Modelo Y**	**Diferencias d_i**	d_i^2
1	132	124	8	64
2	139	141	−2	4
3	126	118	8	64
4	114	116	−2	4
5	122	114	8	64
6	132	132	0	0
7	142	145	−3	9
8	119	123	−4	16
9	126	121	5	25
			18	250

Solución:

Con la información de la Tabla 4.1 podemos obtener la media y la varianza de las diferencias d_i en el consumo de gasolina

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i = \frac{1}{9}\, 18 = 2$$

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} d_i^2 - n\bar{d}^2 \right)$$

$$= \frac{1}{8}\,(250 - 9 \cdot 4) = 26{,}75$$

$$s_d = 5{,}17$$

Según la expresión [4.37] el intervalo de confianza será:

$$I_{\mu_D} = \left[\bar{d} - t_{\alpha/2} \frac{s_d}{\sqrt{n}} \quad ; \quad \bar{d} + t_{\alpha/2} \frac{s_d}{\sqrt{n}} \right]$$

en donde $t_{\alpha/2}$ lo obtenemos utilizando la Tabla A.10 del Anexo A, correspondiente a la distribución t-Student.

$$P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$$

$$P[t_8 > t_{0,005}] = 0{,}005$$

$$t_{0,005} = 3{,}355$$

Sustituyendo en la expresión del intervalo, se tiene:

$$I_{\mu_D} = \left[2 - 3{,}355 \frac{5{,}17}{\sqrt{9}} \quad ; \quad 2 + 3{,}355 \frac{5{,}17}{\sqrt{9}} \right]$$

$$[-3{,}781 \quad ; \quad 7{,}781]$$

4.3.5. INTERVALO DE CONFIANZA PARA EL COCIENTE DE VARIANZAS EN POBLACIONES NORMALES

Varianzas desconocidas y medias desconocidas

Consideramos dos muestras aleatorias de tamaño n_x y n_y independientes, procedentes de poblaciones normales $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$, respectivamente, con medias y varianzas desconocidas, y se pretende obtener un intervalo de confianza para el cociente de varianzas $\frac{\sigma_x^2}{\sigma_y^2}$.

Teniendo en cuenta el apartado 1.7.6, en donde estudiábamos la distribución del cociente de varianzas cuando las medias poblacionales eran desconocidas, entonces, aquí podemos utilizar como estadístico (cantidad pivotal o pivote) que dependa de los parámetros desconocidos σ_x^2 y σ_y^2 y de sus estimadores y cuya distribución muestral no dependa de los parámetros, el estadístico:

$$F = \frac{\dfrac{(n_x - 1)S_x^2}{\sigma_x^2} \bigg/ (n_x - 1)}{\dfrac{(n_y - 1)S_y^2}{\sigma_y^2} \bigg/ (n_y - 1)} = \frac{S_x^2}{S_y^2} \cdot \frac{\sigma_y^2}{\sigma_x^2} \to F_{n_x - 1,\, n_y - 1}$$

que sigue una distribución F-Snedecor con $n_x - 1$ y $n_y - 1$ grados de libertad.

Utilizando este estadístico F y observando el Gráfico 4.8 podemos escribir[14]:

$$P[F_{n_x-1,\, n_y-1,\, \alpha/2} \leqslant F \leqslant F_{n_x-1,\, n_y-1,\, 1-\alpha/2}] = 1 - \alpha$$

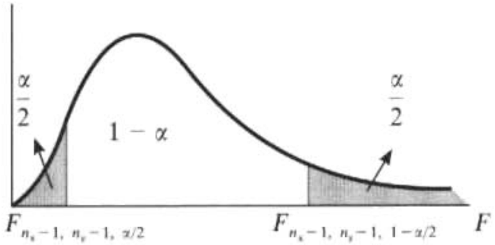


GRÁFICO 4.8. *Distribución* $F_{n_x-1,\, n_y-1}$

$$P\left[F_{n_x-1,\, n_y-1,\, \alpha/2} \leqslant \frac{S_x^2}{S_y^2}\,\frac{\sigma_y^2}{\sigma_x^2} \leqslant F_{n_x-1,\, n_y-1,\, 1-\alpha/2}\right] = 1 - \alpha$$

Multiplicando cada término de la desigualdad por $\dfrac{S_y^2}{S_x^2}$ y después al invertir cada término, cambiando el sentido de la desigualdad, nos quedará[15]:

$$P\left[\frac{S_x^2}{S_y^2}\,\frac{1}{F_{n_x-1,\, n_y-1,\, 1-\alpha/2}} \leqslant \frac{\sigma_x^2}{\sigma_y^2} \leqslant \frac{S_x^2}{S_y^2}\,\frac{1}{F_{n_x-1,\, n_y-1,\, \alpha/2}}\right] = 1 - \alpha$$

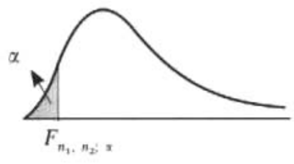
Luego el intervalo de confianza al nivel de confianza del $100(1 - \alpha)\%$ será:

$$\left[\frac{S_x^2}{S_y^2}\,\frac{1}{F_{n_x-1,\, n_y-1,\, 1-\alpha/2}}\ ;\ \frac{S_x^2}{S_y^2}\,\frac{1}{F_{n_x-1,\, n_y-1,\, \alpha/2}}\right] \qquad [4.38]$$

---

[14] La tabla A.11 correspondiente a la distribución F-snedecor nos da:

$$P(F \leqslant F_{n_1,\, n_2;\, \alpha}) = \alpha$$



[15] Teniendo en cuenta la propiedad de reciprocidad de la F-Snedecor

$$F_{n_x-1,\, n_y-1,\, \alpha/2} = \frac{1}{F_{n_y-1,\, n_x-1,\, 1-\alpha/2}} \quad \text{o bien} \quad F_{n_y-1,\, n_x-1,\, 1-\alpha/2} = \frac{1}{F_{n_x-1,\, n_y-1,\, \alpha/2}}$$

Varianzas desconocidas y medias conocidas

Teniendo en cuenta el apartado 1.7.6.a utilizaremos como cantidad pivotal o pivote el estadístico:

$$F = \frac{\dfrac{n_x S_x^{*2}}{\sigma_x^2} \Big/ n_x}{\dfrac{n_y S_y^{*2}}{\sigma_y^2} \Big/ n_y} = \frac{S_x^{*2}}{S_y^{*2}} \cdot \frac{\sigma_y^2}{\sigma_x^2} \to F_{n_x,\ n_y}$$

que sigue una distribución F-Snedecor con n_x, n_y grados de libertad.

Procediendo de manera análoga al caso anterior llegaríamos a obtener el siguiente intervalo de confianza:

$$\left[\frac{S_x^{*2}}{S_y^{*2}} \cdot \frac{1}{F_{n_x,\ n_y,\ 1-\alpha/2}} \quad ; \quad \frac{S_x^{*2}}{S_y^{*2}} \frac{1}{F_{n_x,\ n_y,\ \alpha/2}}\right] \qquad [4.38]$$

siendo

$$S_x^{*2} = \frac{1}{n_x} \sum_{i=1}^{n_x} (X_i - \mu_x)^2 \quad ; \quad S_y^{*2} = \frac{1}{n_y} \sum_{i=1}^{n_y} (Y_i - \mu_y)^2$$

Intervalo de confianza para el cociente de varianzas en poblaciones normales

Varianzas desconocidas y medias desconocidas

Supongamos dos muestras aleatorias independientes de tamaños n_x y n_y seleccionadas de dos poblaciones normales. Si las varianzas para las muestras observadas son s_x^2 y s_y^2, entonces un intervalo de confianza, al nivel de confianza del $100(1-\alpha)\%$ para el cociente de varianzas $\dfrac{\sigma_x^2}{\sigma_y^2}$ viene dado por:

$$\left[\frac{s_x^2}{s_y^2} \cdot \frac{1}{F_{n_x-1,\ n_y-1,\ 1-\alpha/2}} \leqslant \frac{\sigma_x^2}{\sigma_y^2} \leqslant \frac{s_x^2}{s_y^2} \frac{1}{F_{n_x-1,\ n_y-1,\ \alpha/2}}\right] \qquad [4.40]$$

Varianzas desconocidas y medias conocidas

$$\left[\frac{s_x'^2}{s_y'^2} \cdot \frac{1}{F_{n_x,\ n_y,\ 1-\alpha/2}} \leqslant \frac{\sigma_x^2}{\sigma_y^2} \leqslant \frac{s_x'^2}{s_y'^2} \frac{1}{F_{n_x,\ n_y,\ \alpha/2}}\right] \qquad [4.41]$$

siendo

$$s_x^{*2} = \frac{1}{n_x} \sum_{i=1}^{n_x} (x_i - \mu_x)^2 \quad ; \quad s_y^{*2} = \frac{1}{n_y} \sum_{i=1}^{n_y} (y_i - \mu_y)^2$$

Ejemplo 4.10

Supongamos que la distribución de las notas en la asignatura de estadística en segundo curso de la licenciatura de económicas sigue una distribución normal en los dos grupos existentes. Seleccionada una muestra aleatoria de 21 alumnos del primer grupo y otra de 26 alumnos del segundo grupo, ambas independientes, se obtiene como varianzas 1.250 y 900, respectivamente. Obtener un intervalo de confianza para el cociente de las varianzas poblacionales al nivel de confianza del 90 %.

Solución:

Como las medias poblacionales son desconocidas utilizaremos la expresión [4.40] para el intervalo:

$$\left[\frac{s_x^2}{s_y^2}\,\frac{1}{F_{n_x-1,\;n_y-1,\;1-\alpha/2}} \leqslant \frac{\sigma_x^2}{\sigma_y^2} \leqslant \frac{s_x^2}{s_y^2}\,\frac{1}{F_{n_x-1,\;n_y-1,\;\alpha/2}}\right]$$

$$n_x = 21 \qquad s_x^2 = 1.250$$

$$n_y = 26 \qquad s_y^2 = 900$$

$$P[F \leqslant F_{20,\;25,\;0{,}95}] = 0{,}95 \quad \Rightarrow \quad F_{20,\;25,\;0{,}95} = 2{,}01$$

$$P[F \leqslant F_{20,\;25,\;0{,}05}] = 0{,}05$$

Como este valor no viene en las tablas aplicamos la propiedad de reciprocidad y tenemos:

$$F_{20,\;25,\;0{,}05} = \frac{1}{F_{25,\;20,\;0{,}95}} = \frac{1}{2{,}07} = 0{,}48$$

Sustituyendo en la expresión del intervalo, se tiene:

$$\left[\frac{1.250}{900}\,\frac{1}{2{,}01} \quad ; \quad \frac{1.250}{900}\,\frac{1}{0{,}48}\right]$$

$$[0{,}69 \quad ; \quad 2{,}89]$$

4.4. INTERVALOS DE CONFIANZA EN POBLACIONES NO NECESARIAMENTE NORMALES

Hasta ahora hemos considerado que las poblaciones de partida eran normales y se han obtenido intervalos de confianza para distintos parámetros. Pero en muchas situaciones nos podemos encontrar con poblaciones cuya distribución no es conocida, no siendo de aplicación los criterios dados anteriormente[16], y por eso daremos aquí otros métodos de obtención de intervalos de confianza que, en general, serán aplicables a cualquier tipo de población, aunque no se conozca su distribución.

4.4.1. APLICACIÓN DE LA DESIGUALDAD DE CHEBYCHEV PARA LA OBTENCIÓN DE INTERVALOS DE CONFIANZA

Si no se conoce la distribución de la población, entonces podemos utilizar la desigualdad de Chebychev[17] para obtener un intervalo de confianza para la media μ de cualquier distribución con varianza σ^2 conocida.

En efecto, si $(X_1, ..., X_n)$ es una muestra aleatoria simple procedente de una distribución con varianza σ^2 conocida, sabemos que un buen estimador de la media poblacional μ es la media muestral $\bar{X}$, que evidentemente es una variable aleatoria, cuya media y varianza son:

$$E[\bar{X}] = \mu$$

$$\operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Aplicando la desigualdad de chebychev:

$$P[|\bar{X} - E[\bar{X}]| \leqslant k] \geqslant 1 - \frac{\operatorname{Var}(\bar{X})}{k^2} = 1 - \frac{\sigma^2}{nk^2}$$

$$P[|\bar{X} - \mu| \leqslant k] \geqslant 1 - \frac{\sigma^2}{nk^2} \qquad [4.42]$$

[16] En el apartado 4.2 dábamos el *método pivotal* y *método general de Neyman* para la obtención de intervalos, en donde era necesario conocer la función de distribución de la población, pero no es necesario que las distribuciones sean normales.

[17] La *desigualdad de Chebychev* para cualquier variable aleatoria X se puede expresar como:

$$P[|X - \mu| \leqslant k] \geqslant 1 - \frac{\sigma^2}{k^2}$$

y si queremos un nivel de confianza mayor o igual al $100(1 - \alpha)\%$, haremos:

$$1 - \frac{\sigma^2}{nk^2} = 1 - \alpha$$

de donde

$$\frac{\sigma^2}{nk^2} = \alpha$$

$$nk^2 = \frac{\sigma^2}{\alpha}$$

$$k = \frac{\sigma}{\sqrt{n\alpha}}$$

y sustituyendo en la desigualdad de Chebychev, tenemos:

$$P\left[|\bar{X} - \mu| \leqslant \frac{\sigma}{\sqrt{n\alpha}}\right] \geqslant 1 - \alpha$$

de aquí que:

$$P\left[-\frac{\sigma}{\sqrt{n\alpha}} \leqslant \bar{X} - \mu \leqslant \frac{\sigma}{\sqrt{n\alpha}}\right] \geqslant 1 - \alpha$$

$$P\left[\bar{X} - \frac{\sigma}{\sqrt{n\alpha}} \leqslant \mu \leqslant \bar{X} + \frac{\sigma}{\sqrt{n\alpha}}\right] \geqslant 1 - \alpha$$

Luego el intervalo de confianza al nivel del $100(1 - \alpha)\%$ o superior para μ será:

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n\alpha}} \quad ; \quad \bar{X} + \frac{\sigma}{\sqrt{n\alpha}}\right] \qquad [4.43]$$

Ejemplo 4.11

En una central telefónica se seleccionan 150 llamadas telefónicas, observándose que el tiempo medio que tardan en descolgar el teléfono los receptores de esas llamadas era de 2 segundos, con una desviación típica de 0,6 segundos. Se pide, para un nivel de confianza del 99 %:

1. Sin hacer ninguna hipótesis sobre la población de llamadas telefónicas, obtener un intervalo de confianza para el tiempo medio que tardan los usuarios en descolgar el teléfono, suponiendo que la desviación típica poblacional es 0,6.
2. Suponiendo que la población de llamadas telefónicas sigue una distribución normal con desviación típica 0,6, obtener un intervalo de confianza para el tiempo medio que tardan los usuarios en descolgar el teléfono.
3. Ídem al caso anterior pero sin conocer la desviación típica de la población.

Solución:

1. Como no conocemos la distribución de la población tendremos que utilizar la desigualdad de Chebychev para obtener el intervalo de confianza. Así pues, según la desigualdad de Chebychev, expresión [4.42]:

$$P[|\bar{X} - \mu| \leqslant k] \geqslant 1 - \frac{\sigma^2}{nk^2}$$

Como utilizamos un nivel de confianza del 99 %, tendremos:

$$1 - \frac{(0,6)^2}{150k^2} = 0,99$$

$$\frac{0,36}{150k^2} = 0,01$$

$$k^2 = \frac{0,36}{150 \cdot (0,01)} \quad ; \quad k = 0,489$$

$$[|2 - \mu| \leqslant 0,489]$$

$$[2 - 0,489 \leqslant \mu \leqslant \; + 0,489]$$

Luego el intervalo de confianza será:

$$[1,511 \quad ; \quad 2,489]$$

Evidentemente llegaríamos al mismo resultado sustituyendo directamente en la expresión [4.43] del intervalo de confianza.

2. Como admitimos que la población es $N(\mu, 0{,}6)$ con desvición típica conocida, aplicaremos la expresión [4.12]

$$\left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

en donde $z_{\alpha/2}$ es tal que

$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$$

y $Z \to N(0, 1)$.

Sustituyendo en la expresión del intervalo:

$$\left[2 - z_{0,005} \frac{0{,}6}{\sqrt{150}} \leqslant \mu \leqslant 2 + z_{0,005} \frac{0{,}6}{\sqrt{150}}\right]$$

De la tabla A.7 correspondiente a la distribución $N(0, 1)$ se tiene que:

$$P[Z > z_{0,005}] = 0{,}005 \quad \Rightarrow \quad Z_{0,005} = 2{,}575$$

$$\left[2 - 2{,}575 \frac{0{,}6}{\sqrt{150}} \quad ; \quad 2 + 2{,}575 \frac{0{,}6}{\sqrt{150}}\right]$$

$$[2 - 0{,}126 \quad ; \quad 2 + 0{,}126]$$

$$[1{,}874 \quad ; \quad 2{,}126]$$

3. Como la distribución de partida es normal y no conocemos la desviación típica de la población, entonces utilizaremos la expresión [4.19]

$$\left[\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right]$$

en donde $t_{\alpha/2}$ es tal que

$$P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$$

y la variable t_{n-1} sigue una distribución t-Student con $n + - 1$ grados de libertad.

Sustituyendo en la expresión del intervalo:

$$\left[2 - t_{0,005} \frac{0,6}{\sqrt{150}} \quad ; \quad 2 + t_{0,005} \frac{0,6}{\sqrt{150}}\right]$$

De la Tabla A.10 correspondiente a la distribución t-Student con 149 grados de libertad, se tiene, aproximadamente, puesto que utilizamos el valor 150, que:

$$P[t_{149} > t_{0,005}] = 0,005 \quad \Rightarrow \quad t_{0,005} = 2,609$$

$$\left[2 - 2,609 \frac{0,6}{\sqrt{150}} \quad ; \quad 2 + 2,609 \frac{0,6}{\sqrt{150}}\right]$$

$$[2 - 0,128 \quad ; \quad 2 + 0,128]$$

$$[1,872 \quad ; \quad 2,128]$$

4.4.2. INTERVALOS DE CONFIANZA PARA MUESTRAS GRANDES

Los métodos descritos para la obtención de intervalos de confianza (método pivotal y método general de Neyman) presentan algunas deficiencias prácticas. Así pues, el método pivotal depende de la posibilidad de encontrar un estadístico o cantidad pivotal que contenga el parámetro y su estimador y tal que su distribución muestral no dependa del parámetro; pero tal estadístico puede no existir. Por otro lado, el método general de Neyman puede implicar cálculos pesados. Por estas razones daremos otros métodos que pueden ser utilizados si tenemos una muestra grande, y que se basan en la distribución asintótica del estimador de máxima verosimilitud, si es que existe, o en el Teorema Central del Límite.

4.4.2.1. Intervalos de confianza para muestras grandes a partir de un estimador de máxima verosimilitud

Sabemos que si $\hat{\theta}$ es un estimador de máxima verosimilitud[18] del parámetro θ, entonces para muestras grandes es asintóticamente eficiente y asintóticamente normal y según la expresión [3.11] se tiene:

$$\hat{\theta} \underset{n \to \infty}{\longrightarrow} N(\theta, \sqrt{\operatorname{Var}(\hat{\theta})})$$

[18] Véase apartado 3.3.1 sobre propiedades de los estimadores de máxima verosimilitud.

en donde la $\mathrm{Var}(\hat{\theta})$ coincide con la cota de Frechet-Cramer-Rao:

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{nE\left[\dfrac{\partial \ln f(x;\ \theta)}{\partial \theta}\right]^2}$$

Lo cual nos permite llegar a que:

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\mathrm{Var}(\hat{\theta})}} \underset{n \to \infty}{\longrightarrow} N(0,\ 1) \qquad [4.44]$$

es decir, cuando $n \to \infty$ el estadístico Z se aproxima a una distribución $N(0,\ 1)$.

En consecuencia el estadístico

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\mathrm{Var}(\hat{\theta})}}$$

lo podemos utilizar como cantidad pivotal o pivote, pues depende del parámetro y de su estimador y su distribución es independiente del parámetro, pues para n grande es aproximadamente $N(0,\ 1)$.

En consecuencia para el nivel de confianza del $100(1 - \alpha)\%$, se puede obtener un intervalo de confianza aproximada para el parámetro θ a partir de la expresión:

$$P\left[-z_{\alpha/2} \leqslant \frac{\hat{\theta} - \theta}{\sqrt{\mathrm{Var}(\hat{\theta})}} \leqslant z_{\alpha/2}\right] = 1 - \alpha$$

$$P[\hat{\theta} - z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\theta})} \leqslant \theta \leqslant \hat{\theta} + z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\theta})}] = 1 - \alpha$$

Luego el intervalo de confianza para el parámetro θ será:

$$I_\theta = [\hat{\theta} - z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\theta})}\ \ ;\ \ \hat{\theta} + z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\theta})}] \qquad [4.44]$$

siendo $z_{\alpha/2}$ tal que

$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$$

se puede comprobar que los intervalos de confianza construidos para muestras grandes a partir de un estimador de máxima verosimilitud son de menos amplitud, es decir, son más precisos que los obtenidos a partir de cualquier otro estimador.

Este procedimiento general para construir intervalos de confianza para muestras grandes lo podemos resumir en los siguientes pasos:

1. Se determina el estimador de máxima verosimilitud $\hat{\theta}$ del parámetro θ.
2. Se obtiene la $\operatorname{Var}(\hat{\theta})$, directamente o utilizando la cota de Frechet-Cramer-Rao.
3. En la expresión de la $\operatorname{Var}(\hat{\theta})$ si aparece el parámetro θ se sustituye por su estimador de máxima verosimilitud, $\hat{\theta}$.
4. Se construye el intervalo de confianza al nivel de confianza del $100(1-\alpha)\%$ utilizando la expresión:

$$I_\theta = [\hat{\theta} - z_{\alpha/2}\sqrt{\operatorname{Var}(\hat{\theta})} \quad ; \quad \hat{\theta} + z_{\alpha/2}\sqrt{\operatorname{Var}(\hat{\theta})}]$$

Ejemplo 4.12

Obtener el intervalo de confianza al nivel de confianza del 90 % para el parámetro a utizando una muestra de tamaño 144, procedente de una distribución $\Gamma\left(1, \frac{1}{a}\right)$, sabiendo que la media de la muestra es 5.

Solución:

La función de densidad de la $\Gamma\left(1, \frac{1}{a}\right)$ será:

$$f(x; a) = \frac{1}{a} e^{-\frac{x}{a}} \quad , \quad x > 0 \quad , \quad a > 0$$

Seguiremos los cuatro pasos indicados.

1. En el ejemplo 3.11 se obtenían los estimadores de máxima verosimilitud de una $\Gamma(p, a)$, luego fácilmente se observa que en nuestro caso de una $\Gamma\left(1, \frac{1}{\mathrm{a}}\right)$ el estimador de máxima verosimilitud del parámetro a será:

$$\hat{a} = \bar{X}$$

2. La $\mathrm{Var}(\hat{a})$ será:

$$\mathrm{Var}(\hat{a}) = \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{a^2}{n}$$

pues en la $\Gamma(p, a)$ habíamos obtenido que

$$\mathrm{Var}(X) = \frac{p}{a^2}$$

y aquí estamos considerando una $\Gamma\left(1, \frac{1}{a}\right)$.

3. Ya que la $\mathrm{Var}(\hat{a})$ depende del parámetro a, lo reemplazaremos por su estimador $\bar{X}$, luego

$$\mathrm{Var}(\hat{a}) = \frac{\bar{X}^2}{n} = \frac{25}{144}$$

4. Aplicando la expresión [4.45] tendremos el intervalo:

$$I_a = [\hat{a} - Z_{0,05}\sqrt{\mathrm{Var}(\hat{a})} \quad ; \quad \hat{a} + Z_{0,05}\sqrt{\mathrm{Var}(\hat{a})}]$$

De la Tabla A.7 correspondiente a la $N(0, 1)$, tenemos:

$$P[Z > z_{0,05}] = 0{,}05 \quad \Rightarrow \quad z_{0,05} = 1{,}645$$

$$\left[5 - 1{,}645\sqrt{\frac{25}{144}} \quad ; \quad 5 + 1{,}645\sqrt{\frac{25}{144}}\right]$$

$$[5 - 0{,}685 \quad ; \quad 5 + 0{,}685]$$

$$[4{,}314 \quad ; \quad 5{,}685]$$

que será el intervalo de confianza para el parámetro a al nivel de confianza del 90 %.

4.4.2.2. Intervalos de confianza para muestras grandes aplicando el Teorema Central del Límite

Supongamos una muestra aleatoria simple $(X_1, ..., X_n)$ suficientemente grande procedente de una población con distribución desconocida y varianza σ^2 finita conocida y deseamos obtener un intervalo de confianza al nivel del $100(1 - \alpha)\%$ para la media, desconocida, μ de la población.

Puesto que se cumplen la hipótesis sabemos por el Teorema Central del Límite que cuando n es suficientemente grande:

$$Z = \frac{S_n - E[S_n]}{\sigma(S_n)} = \frac{\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]}{\sqrt{\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right)}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,\, 1)$$

Luego podemos utilizar el estadístico

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

como cantidad pivotal o pivote, y tendríamos:

$$P\left[-z_{\alpha/2} \leqslant \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leqslant z_{\alpha/2}\right] \simeq 1 - \alpha$$

y de manera análoga a como procedíamos anteriormente, llegaremos a que el intervalo de confianza al nivel del $(1 - \alpha)\%$ será:

$$I_\mu = \left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad ; \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \qquad [4.46]$$

siendo $z_{\alpha/2}$ tal que

$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$$

y Z se distribuye aproximadamente según una $N(0, 1)$.

La diferencia con los intervalos obtenidos anteriormente en que aquellos eran exactos y ahora son aproximados y sólo son válidos para muestras grandes, $n > 30$.

Cuando σ^2 es desconocida se toma como valor aproximado la varianza muestral S^2, y se obtendría como intervalo de confianza:

$$I_\mu = \left[\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \quad ; \quad \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad [4.47]$$

expresión equivalente a la [4.18] pero ahora cuando n es grande:

$$t_{n-1} \underset{n \to \infty}{\longrightarrow} N(0, 1)$$

es decir la t-Student tiende a la distribución $N(0, 1)$.

Expresión análoga a la [4.47], obtenida anteriormente, se tendrá para el caso de la diferencia de medias poblacionales, llegando a obtener:

$$I_{\mu_x - \mu_y} = \left[(\bar{X} - \bar{Y}) + - z_{\alpha/2} \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} \ ; \ (\bar{X} - \bar{Y}) + z_{\alpha/2} \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} \right] \quad [4.48]$$

Ejemplo 4.13

De los exámenes correspondientes a una oposición realizada a nivel nacional, se extrae una muestra de 75 ejercicios correspondientes a mujeres y otra de 50 ejercicios correspondientes a hombres, siendo la calificación media de la muestra de mujeres 82 puntos con una desviación típica muestral de 8, mientras que para los hombres la calificación media fue de 78 con una desviación típica de 6. Obtener el intervalo de confianza al nivel de confianza del 95 % para la diferencia de la puntuación media μ_x de las mujeres y la puntuación media μ_y de los hombres.

Solución:

Como las muestras son suficientemente grandes, pues son mayores que 30 y las poblaciones no son normales podemos obtener un intervalo de confianza aproximado utilizando la expresión [4.48] en donde sustituimos las varianzas poblacionales por las varianzas muestrales, obteniendo el intervalo:

$$I_{\mu_x - \mu_y} = \left[(\bar{x} - \bar{y}) - z_{\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \ ; \ (\bar{x} - \bar{y}) + z_{\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \right]$$

Así pues del enunciado se tiene:

$$\bar{x} = 82 \qquad s_x = 8 \qquad n_x = 75$$

$$\bar{y} = 78 \qquad s_y = 6 \qquad n_y = 50$$

$$P[Z > z_{0,025}] = 0{,}025$$

$$z_{0,025} = 1{,}96$$

Sustituyendo en la expresión del intervalo tenemos:

$$\left[(82-78)-1{,}96\sqrt{\frac{64}{75}+\frac{36}{50}}\quad ; \quad (82-78)+1{,}96\sqrt{\frac{64}{75}+\frac{36}{50}}\right]$$

$$[4-(1{,}96)(1{,}25)\quad ; \quad 4+(1{,}96)(1{,}25)]$$

$$[1{,}55\quad ; \quad 6{,}45]$$

4.5. INTERVALO DE CONFIANZA DE UNA PROPORCIÓN

Sea una población binomial $B(1, p)$ y una muestra aleatoria de tamaño n de esa población, es decir realizamos n repeticiones del experimento de Bernoulli que estamos considerando, y deseamos obtener un intervalo de confianza al nivel del $100(1-\alpha)\%$ para el parámetro poblacional p.

Consideramos dos situaciones según que el tamaño de la muestra sea pequeño o sea grande.

4.5.1. INTERVALO DE CONFIANZA DE UNA PROPORCIÓN PARA MUESTRAS PEQUEÑAS

En esta situación no podemos encontrar un estadístico (cantidad pivotal o pivote) que reúna los requisitos necesarios para aplicar el método pivotal y entonces tendremos que recurrir al método general de Neyman, que desarrollamos en el apartado 4.2.2.

Para aplicar este método necesitamos obtener un estimador del parámetro p, y en este caso consideramos el estimador de máxima verosimilitud que para la $B(1, p)$ sabemos que viene dado por:

$$\hat{p}=\frac{X}{n}=\frac{\text{número de éxitos en } n \text{ pruebas}}{\text{número de pruebas}}$$

que para una muestra concreta sería $\hat{p}=\dfrac{x}{n}$.

Ahora tendríamos que obtener la función de probabilidad[19] del estimador $\hat{p} = \frac{X}{n}$, que será:

$$\binom{n}{nx} p^{nx}(1-p)^{n-nx} \quad , \quad x = 0, \frac{1}{n}, \frac{2}{n}, \ldots, 1$$

Si admitimos que $\hat{p} = p_0$, entonces según el método general tenemos que obtener dos valores p_i y p_s, tales que

$$P(\hat{p} \leqslant p_i) = \sum_{x=0}^{p_i} \binom{n}{nx} p^{nx}(1-p)^{n-nx} \leqslant \frac{\alpha}{2}$$

$$P(\hat{p} \geqslant p_s) = \sum_{x=p_s}^{1} \binom{n}{nx} p_i^{nx}(1-p)^{n-nx} \leqslant \frac{\alpha}{2}$$

y si $\hat{p} = p_0$, bastará sustituir en ambas expresiones p por p_0

Estas ecuaciones se pueden resolver para obtener los valores aproximados de p_i y p_s, pero los cálculos son bastante pesados incluso para valores moderados de n, pues se tendrían que hacer consecutivamente dando valores a p_i y p_s y observando el error. De esta forma obtendríamos al nivel de confianza del $100(1-\alpha)\%$ un intervalo para el parámetro p:

$$I_p = [p_i \leqslant p \leqslant p_s]$$

Ante la dificultad de resolución de estas ecuaciones, Clopper y Pearson decidieron acudir a un método gráfico, o mediante *ábacos* que obtuvieron resolviendo ecuaciones del tipo anterior para distintos valores de x con n y α fijos. Así pues, para cada valor de x, es decir para cada muestra, y, en consecuencia, para cada valor de $\hat{p} = \frac{x}{n}$, se tienen dos soluciones, que en el plano cartesiano $(\hat{p}, p)$ vendrán representadas por dos puntos uno inferior p_i y otro superior p_s. Uniendo los puntos inferiores, por una parte, y los puntos superio-

[19] Cuando queremos estimar el parámetro p en la $B(1, p)$, utilizamos la cantidad $\hat{p} = Y = \frac{X}{n}$, siendo X el número de éxitos en n pruebas, que sigue una $B(n, p)$. Ahora para obtener la distribución de Y consideramos $Y = g(X) = \frac{X}{n}$ que es una aplicación de uno en uno, luego tenemos:

$$P_y(x) = P_X(nx) = \begin{cases} \binom{n}{nx} p^{nx}(1-p)^{n-nx}, & x = 0, \frac{1}{n}, \frac{2}{n}, \ldots, 1 \\ 0 & \text{en el resto} \end{cases}$$

res correspondientes, por otra, se obtiene una curva inferior y otra superior para cada n y $1 - \alpha$ considerados. Así pues, para un nivel de confianza del 95 % y para distintos valores de n tenemos el conjunto de curvas o ábacos que aparecen en el gráfico 4.9. Existen tablas en las cuales aparecen gráficos de este tipo para cada valor del nivel de confianza. En la Tabla A.13 del Anexo A de tablas estadísticas se dan gráficas para los niveles de confianza del 95 y 99 %.

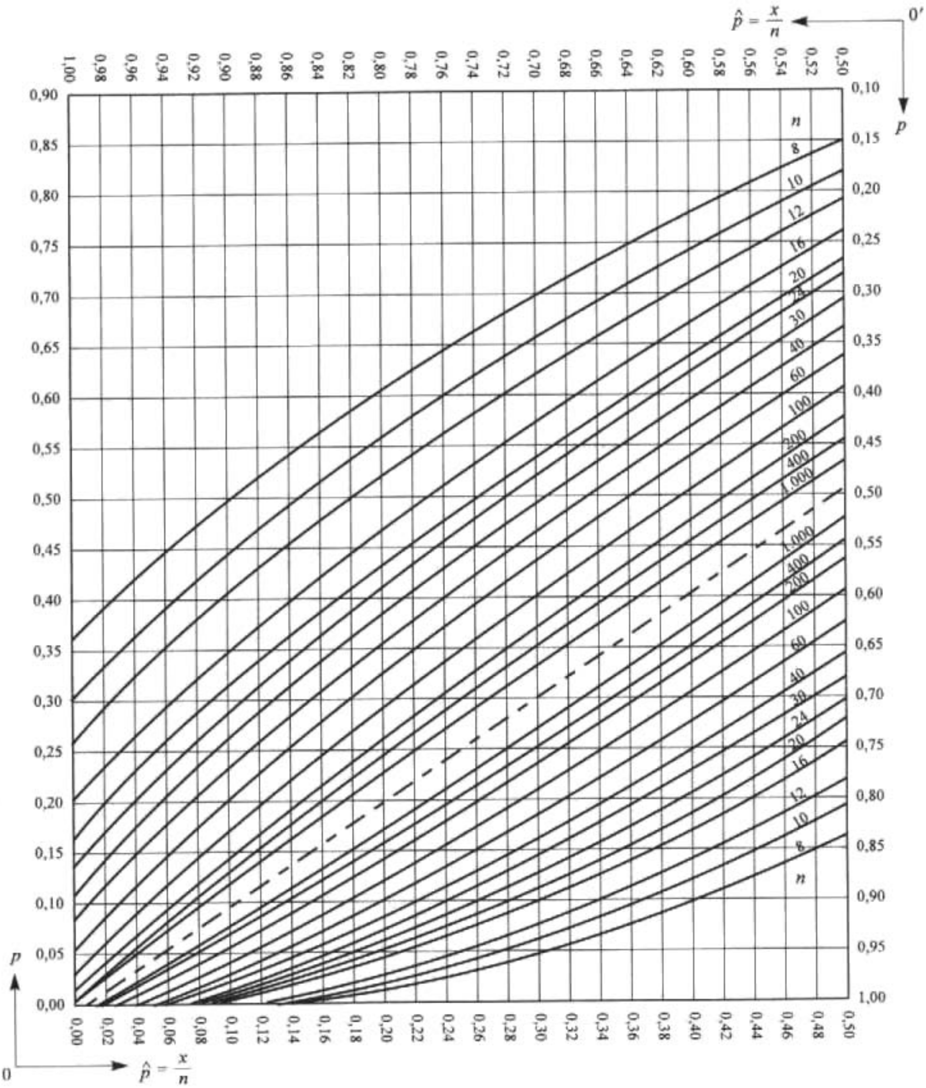


GRÁFICO 4.9. *Gráficas de intervalos de confianza del parámetro p de una distribución binomial al nivel de confianza del* 95 %.(Fuente: Tables for Statisticians, *Biometrika*, vol. 1, 1966.)

Cada gráfico presenta valores de $\hat{p}$ en la parte inferior y la parte superior y valores de p a la izquierda y a la derecha. La escala inferior de valores de $\hat{p}$ entre 0,00 y 0,50 y la superior entre 0,50 y 1,00. Así pues cuando $\hat{p}$ toma valores en la escala inferior los límites del intervalo de confianza para p se tienen en la escala de la izquierda, y cuando $\hat{p}$ toma valores en la escala superior, los límites del intervalo se tienen en la escala de la derecha.

Ejemplo 4.14

De una muestra aleatoria de 20 votantes para la elección de un candidato A, resulta que 6 tienen intención de votar al candidato A. Obtener un intervalo de confianza para el parámetro p, proporción de votantes del candidato A, al nivel de confianza del 95 %.

Solución:

Con la muestra de tamaño $n = 20$ se obtiene que $x = 6$ y, por tanto, la estimación será:

$$\hat{p} = \frac{x}{n} = \frac{6}{20} = 0,30$$

Para obtener los límites de confianza para p al nivel del 95 % nos situamos en el punto $\hat{p} = 0,30$ del eje de abscisas (eje $O\hat{p}$) del Gráfico 4.9, se levanta una vertical hasta cortar las curvas correspondientes a $n = 20$, y se proyectan los dos puntos resultantes sobre el eje de ordenadas (eje Op) que nos darían los valores $p_i = 0,11$ y $p_s = 0,54$.

Luego el intervalo para el parámetro p será:

$$[0,11 \quad ; \quad 0,54]$$

Si el úmero de personas que tienen intención de votar al candidato A hubiera sido de 14, entonces $x = 14$ y

$$\hat{p} = \frac{14}{20} = 0,70$$

entonces este valor de $\hat{p} = 0,70$ lo buscaríamos en el eje $O'\hat{p}$ superior y proyectaríamos sobre el eje $O'p$ de la derecha, así pues obtendríamos el intervalo

$$[0,46 \quad ; \quad 0,89]$$

4.5.2. INTERVALO DE CONFIANZA DE UNA PROPORCIÓN PARA MUESTRAS GRANDES

Supongamos una población $B(1, p)$ y consideramos una muestra aleatoria de tamaño n suficientemente grande, es decir, realizamos un número grande de repeticiones independientes del experimento de Bernoulli que estamos considerando y queremos obtener un intervalo de confianza al nivel del $100(1-\alpha)\%$ para el parámetro p.

Sabemos que el estimador de máxima verosimilitud del parámetro p de una $B(1, p)$ viene dado por:

$$\hat{p} = \frac{X}{n} = \frac{\text{número de éxitos en } n \text{ pruebas}}{\text{número de pruebas}}$$

y para una muestra concreta de tamaño n la estimación será:

$$\hat{p} = \frac{x}{n}$$

Teniendo en cuenta el apartado 4.5.2.1 y recordando las propiedades de los estimadores de máxima verosimilitud, diremos que el estimador $\hat{p}$ es asintóticamente normal, es decir:

$$\hat{p} \underset{n \to \infty}{\longrightarrow} N(E[\hat{p}], \sqrt{\operatorname{Var}(\hat{p})})$$

en donde:

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{np}{n} = p$$

$$\operatorname{Var}(\hat{p}) = \operatorname{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\operatorname{Var}(X) = \frac{1}{n^2}\,np(1-p) = \frac{p(1-p)}{n} = \frac{pq}{n}$$

Luego[20]

$$\hat{p} \to N\left(p, \sqrt{\frac{pq}{n}}\right)$$

[20] También veíamos al estudiar la distribución de la proporción muestral, apartado 1.8, que según el Teorema Central del Límite:

$$\hat{p} \to N\left(p, \sqrt{\frac{pq}{n}}\right)$$

Lo cual nos permite decir que el estadístico

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \to N(0,\ 1)$$

se aproxima a una distribución $N(0,\ 1)$ cuando n es suficientemente grande, $n \to \infty$.

En consecuencia este estadístico Z lo podemos utilizar como cantidad pivotal o pivote, pues depende del parámetro y de su estimador y su distribución es independiente del parámetro p, pues se trata de una $N(0,\ 1)$. Por tanto, podremos obtener un intervalo de confianza para el parámetro p al nivel del $100(1 - \alpha)\%$ a partir de la expresión

$$P\left[-z_{\alpha/2} \leqslant \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}} \leqslant z_{\alpha/2}\right] = 1 - \alpha$$

Multiplicando cada término de la desigualdad por $\sqrt{\frac{pq}{n}}$, restado después a cada término $\hat{p}$ y multiplicando por -1, se tiene:

$$P\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{pq}{n}} \leqslant p \leqslant \hat{p} + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right] = 1 - \alpha \qquad [4.49]$$

Pero los límites de la expresión [4.49] dependen del parámetro desconocido p. Si n es grande una solución satisfactoria se obtiene sustituyendo p por su estimación $\hat{p}$ en el límite interior y en el límite superior, resultando:

$$P\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leqslant p \leqslant \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right] \simeq 1 - \alpha \qquad [4.50]$$

Luego el intervalo de confianza al nivel de confianza del $100(1 - \alpha)\%$ para el parámetro p será:

$$I_p = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\ ;\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right] \qquad [4.51]$$

en donde $z_{\alpha/2}$ es tal que

$$P[Z \geqslant z_{\alpha/2}] = \frac{\alpha}{2}$$

y la variable aleatoria Z sigue una $N(0, 1)$ cuando n es suficientemente grande[21]:

Intervalo de confianza para la proporción poblacional para muestras grandes[22]

Sea una población $B(1, p)$ y si $\hat{p}$ representa la proporción de éxitos en una muestra aleatoria de tamaño n suficientemente grande y $\hat{q} = 1 - \hat{p}$, entonces un intervalo de confianza aproximado para la proporción poblacional p al nivel de confianza del $100(1 - \alpha)\%$ viene dado por:

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leqslant p \leqslant \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right] \qquad [4.52]$$

en donde $z_{\alpha/2}$ es tal que

$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$$

y la variable aleatoria Z sigue una distribución $N(0, 1)$.

Observando la expresión [4.52] podemos decir que si la estimación $\hat{p}$ ocupa el lugar central o punto medio del intervalo de confianza, entonces $\hat{p}$ estima puntualmente, sin error, el valor del parámetro proporción poblacional p, pero generalmente esto no sucederá y se cometerá un **error en la estimación** que vendrá dado por la diferencia positiva entre el verdadero valor del parámetro p y la estimación $\hat{p}$, y además tendremos la confianza del $100(1 - \alpha)\%$ de que este error a lo sumo será

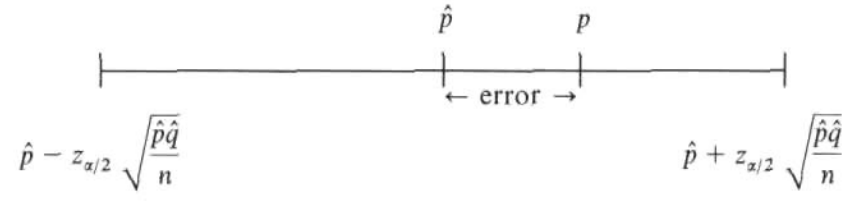
$$z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

[21] Algunos autores consideran que la aproximación es buena cuando $np > 5$ y $n > 30$.

[22] Los *intervalos unilaterales* vienen dados por

$$\left(-\infty, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right] \quad \text{y} \quad \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}, +\infty\right)$$

Gráficamente el error que se comete al estimar el parámetro p mediante $\hat{p}$ será:



Ejemplo 4.15

Se selecciona una muestra aleatoria de 600 familias, a las cuales se les pregunta si poseen o no ordenador personal en casa, resultando que 240 de esas familias contestaron afirmativamente. Obtener un intervalo de confianza al nivel del 95 % para estimar la proporción real de familias que poseen ordenador personal en casa.

Solución:

El estimador puntuar de p sabemos que es:

$$\hat{p} = \frac{X}{n}$$

y para la muestra concreta de 600 familias la estimación correspondiente será:

$$\hat{p} = \frac{x}{n} = \frac{240}{600} = 0{,}40$$

Utilizando la Tabla A.7, correspondiente a la $N(0, 1)$:

$$P[Z > z_{0,025}] = 0{,}025 \quad \Rightarrow \quad z_{0,025} = 1{,}96$$

Sustituyendo en la expresión [4.52] tendremos el intervalo de confianza pedido:

$$\left[0{,}40 - 1{,}96\sqrt{\frac{(0{,}40)(0{,}60)}{600}} \quad ; \quad 0{,}40 + 1{,}96\sqrt{\frac{(0{,}40)(0{,}60)}{600}}\right]$$

$$[0{,}40 - 0{,}04 \quad ; \quad 0{,}40 + 0{,}04]$$

$$[0{,}36 \quad ; \quad 0{,}44]$$

y diremos que con un nivel de confianza del 95 % la estimación $\hat{p} = 0{,}40$ difiere del parámetro p a lo sumo en la cantidad 0,04, es decir el error máximo a este nivel de confianza será de 0,04.

4.6. INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE PROPORCIONES

Ahora estamos interesados en estimar la diferencia entre dos parámetros poblacionales p_X y p_Y, es decir queremos obtener un intervalo de confianza para la diferencia $(p_X - p_Y)$ de los dos parámetros poblacionales. Para ello se seleccionan dos muestras aleatorias independientes de tamaño n_x y n_y de cada una de las dos poblaciones $B(1, p_X)$ y $B(1, p_Y)$, respectivamente.

Los estimadores puntuales de los parámetros p_X y p_Y serán:

$$\hat{p}_X = \frac{X}{n_x} \quad , \quad \hat{p}_Y = \frac{Y}{n_y}$$

y las estimaciones para unas muestras concretas serán:

$$\hat{p}_x = \frac{x}{n_x} \quad , \quad \hat{p}_y = \frac{y}{n_y}$$

Pero a nosotros nos interesa el intervalo de confianza para la diferencia $(p_X - p_Y)$, luego utilizaremos como estimador de esta diferencia, el estadístico:

$$\hat{p}_X - \hat{p}_Y = \frac{X}{n_x} - \frac{Y}{n_y} \qquad [4.53]$$

y teniendo en cuenta que las muestras se toman independientemente, entonces la media y la varianza serán:

$$E[\hat{p}_X - \hat{p}_Y] = E[\hat{p}_X] - E[\hat{p}_Y] = p_X - p_Y \qquad [4.54]$$

$$\begin{aligned}\operatorname{Var}(\hat{p}_X - \hat{p}_Y) &= \operatorname{Var}(\hat{p}_X) + \operatorname{Var}(\hat{p}_Y) \\ &= \frac{p_X(1 - p_X)}{n_x} + \frac{p_Y(1 - p_Y)}{n_y} \\ &= \frac{p_X q_X}{n_x} + \frac{p_Y q_Y}{n_y} \qquad [4.55]\end{aligned}$$

Además, sabemos que si los tamaños de muestras son grandes, la distribución de esta variable aleatoria $(\hat{p}_X - \hat{p}_Y)$ es aproximadamente normal, es decir:

$$\hat{p}_X - \hat{p}_Y \to N\left(p_X - p_Y, \sqrt{\frac{p_X q_X}{n_x} + \frac{p_Y q_Y}{n_y}}\right)$$

Procediendo de manera análoga a como lo hacíamos en el apartado anterior, resulta que:

$$Z = \frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\dfrac{p_X q_X}{n_x} + \dfrac{p_Y q_Y}{n_y}}} \to N(0, 1)$$

Por tanto, también podemos escribir:

$$P\left[-z_{\alpha/2} \leqslant \frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\dfrac{p_X q_X}{n_x} + \dfrac{p_Y q_Y}{n_y}}} \leqslant z_{\alpha/2}\right] = 1 - \alpha$$

de donde llegaremos a:

$$P\left[(\hat{p}_X - \hat{p}_Y) - z_{\alpha/2}\sqrt{\frac{p_X q_X}{n_x} + \frac{p_Y q_Y}{n_y}} \leqslant p_X - p_Y \leqslant \right.$$

$$\left. \leqslant (\hat{p}_X - \hat{p}_Y) + z_{\alpha/2}\sqrt{\frac{p_X q_X}{n_x} + \frac{p_Y q_Y}{n_y}}\right] = 1 - \alpha$$

y como los extremos de esta expresión dependen de los parámetros desconocidos p_X y p_Y los reemplazaremos por sus estimaciones que para unas muestras concretas serán:

$$\hat{p}_x = \frac{x}{n} \quad , \quad \hat{p}_y = \frac{y}{n}$$

y nos quedará:

$$P\left[(\hat{p}_x - \hat{p}_y) - z_{\alpha/2}\sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y q_y}{n_y}} \leqslant p_X - p_Y \leqslant \hat{p}_x - \hat{p}_y) + z_{\alpha/2}\sqrt{\frac{\hat{p}_x q_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}}\right] = 1 - \alpha$$

Luego el intervalo de confianza al nivel del $100(1 - \alpha)\%$ para la diferencia de los parámetros poblacionales $p_X - p_Y$ sefa:

$$\left[(\hat{p}_x - \hat{p}_y) - z_{\alpha/2}\sqrt{\frac{\hat{p}_x\hat{q}_x}{n_x} + \frac{\hat{p}_y\hat{q}_y}{n_y}} \leqslant (\hat{p}_x - \hat{p}_y) + z_{\alpha/2}\sqrt{\frac{\hat{p}_x\hat{q}_x}{n_x} + \frac{\hat{p}_y\hat{q}_y}{n_y}}\right] \qquad [4.56]$$

Intervalo de confianza para la diferencia de proporciones

Sea $\hat{p}_x$ la proporción de éxitos observados en una muestra aleatoria de tamaño n_x de una población $B(1, p_X)$, y sea $\hat{p}_y$ la proporción de éxitos observados en una muestra aleatoria de tamaño n_y de la población $B(1, p_Y)$. Entonces si las muestras son independientes y los tamaños son grandes el intervalo de confianza al nivel del $100(1 - \alpha)\%$ para la diferencia de los parámetros $p_X - p_Y$ será:

$$\left[(\hat{p}_x - \hat{p}_y) - z_{\alpha/2}\sqrt{\frac{\hat{p}_x\hat{q}_x}{n_x} + \frac{\hat{p}_y\hat{q}_y}{n_y}} \leqslant p_X - p_Y \leqslant z_{\alpha/2}\sqrt{\frac{p_x q_x}{n_x} + \frac{p_y q_y}{n_y}}\right] \qquad [4.57]$$

donde $z_{\alpha/2}$ es tal que

$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$$

y la variable aleatoria Z sigue una distribución $N(0, 1)$.

Ejemplo 4.16

En una ciudad A se toma una muestra aleatoria de 98 cabezas de familia, de los cuales 48 han sido poseedores de acciones de Telefónica. Mientras que en otra ciudad B se selecciona otra muestra aleatoria de tamaño 127 cabezas de familia, de los cuales 21 han sido poseedores de acciones de Telefónica. Obtener un intervalo de confianza al nivel del 95 % para la diferencia entre las proporciones de cabezas de familia que han sido poseedores de ese tipo de acciones en ambas ciudades.

Solución:

De la información del enunciado se deduce:

$$n_x = 98 \qquad x = 48 \qquad \hat{p}_x = \frac{48}{98} = 0{,}490$$

$$n_y = 127 \qquad y = 21 \qquad \hat{p}_y = \frac{21}{127} = 0{,}165$$

Para el nivel de confianza del 95 %, $\alpha = 0{,}05$, luego

$$P[Z > z_{0,025}] = 0{,}025 \quad \Rightarrow \quad z_{0,025} = 1{,}96$$

Sustituyendo en la expresión [4.57]:

$$\left[(0{,}490 - 0{,}510) - 1{,}96\sqrt{\frac{(0{,}490)(0{,}165)}{98} + \frac{(0{,}165)(0{,}835)}{127}}\ ;\right.$$

$$\left.(0{,}490 - 0{,}165) + 1{,}96\sqrt{\frac{(0{,}490)(0{,}510)}{98} + \frac{(0{,}165)(0{,}835)}{127}}\right]$$

$$[0{,}208 \ ; \ 0{,}443]$$

Como el 0 está fuera del rango del intervalo, esto nos indica que es bastante más probable que un cabeza de familia de la ciudad *A* haya tenido acciones de Telefónica que un cabeza de familia de la ciudad *B*.

4.7. ESTIMACIÓN DEL TAMAÑO MUESTRAL

Hasta ahora hemos dado métodos para obtener intervalos de confianza de parámetros de una población, basándonos en la información contenido en una muestra dada. Siguiendo tal proceso, un investigador puede pensar que el intervalo de confianza resultante es demasiado amplio, reflejando una importante incertidumbre sobre el parámetro estimado. La única manera de obtener un intervalo más preciso (de menor amplitud), con un nivel de confianza dado, es aumentando el tamaño de la muestra.

En algunas circunstancias, se puede fijar de antemano la amplitud del intervalo de confianza, eligiendo un tamaño de muestra bastante grande para garantizar la amplitud. A continuación veremos cómo se puede determinar el tamaño de la muestra para estimar algunos parámetros poblacionales.

4.7.1. TAMAÑO DE MUESTRA PARA ESTIMAR LA MEDIA μ DE UNA POBLACIÓN NORMAL CON σ CONOCIDA

Sabemos que si tomamos una muestra aleatoria simple de tamaño n procedente de una población $N(\mu, \sigma)$, siendo σ conocida, el intervalo de confianza al nivel del $100(1 - \alpha)\%$ para la media poblacional μ venía dado por:

$$I_\mu = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad ; \quad \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{4.58}$$

Siendo la amplitud del intervalo

$$L = \left(\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{4.59}$$

Si, previamente, se fija la longitud del intervalo L y deseamos conocer el tamaño de la muestra para obtener ese intervalo al nivel de confianza del $100(1 - \alpha)\%$, bastará despejar n de la expresión [4.59], pues L, $z_{\alpha/2}$ y σ son conocidos, y tendremos que el tamaño de la muestra será:

$$n = 4z^2_{\alpha/2} \cdot \frac{\sigma^2}{L^2} \tag{4.60}$$

el cual nos permitirá construir un intervalo al nivel de confianza del $100(1 - \alpha)\%$ y de amplitud L para la media de una población normal con σ conocida.

Ejemplo 4.17

La longitud de los tornillos fabricados por una determinada máquina se distribuye según una distribución normal con desviación típica $\sigma = 2$ mm. Con el fin de obtener un intervalo al 99 % de confianza para la longitud media de los tornillos producidos durante un día determinado se toma una muestra aleatoria de 10 tornillos cuya longitud media resultó ser de 96 mm. Calcular el correspondiente intervalo de confianza con estos datos y determinar el tamaño de muestra necesario para construir un intervalo al 99 % de confianza para la longitud media de esos tornillos, con una longitud de 2 mm.

Solución:

El intervalo de confianza para la media de una población normal se obtiene a partir de la expresión [4.58], pues sustituyendo:

$$\bar{x} = 96 \text{ mm} \quad , \quad z_{\alpha/2} = z_{0,005} = 2.575 \quad , \quad \sigma = 2 \quad , \quad n = 10$$

$$\left[96 - 2.575 \frac{2}{\sqrt{10}} \quad ; \quad 96 + 2.575 \frac{2}{\sqrt{10}}\right]$$

$$[94,37 \quad ; \quad 97,63]$$

y el tamaño de muestra necesario para un intervalo de longitud 2 mm se obtendrá sustituyendo en la expresión [4.60]

$$n = 4 \frac{2.575^2 \cdot 2^2}{2^2} = 26,52 \simeq 27$$

Luego se necesita una muestra de tamaño 27 tornillos para la obtención de un intervalo al nivel de confianza del 99 % y con una longitud de 2 mm.

Pero esta situación no suele ser real, ya que si no conocemos la media de la población, y por ello queremos obtener un intervalo de confianza para la media poblacional, probablemente tampoco conoceremos la varianza σ^2 de la población, de tal manera que no podremos aplicar la expresión [4.60], ya que previamente tendríamos que estimar, con la ayuda de una muestra, la varianza poblacional, utilizando para ello la varianza muestral S^2 y obtendríamos una expresión distinta a la dada en [4.60].

4.7.2. TAMAÑO DE MUESTRA PARA ESTIMAR LA MEDIA μ DE UNA POBLACIÓN NORMAL CON σ DESCONOCIDA

El intervalo de confianza al nivel del $100(1 - \alpha)\%$ para la media de una población normal con σ desconocida, según la expresión [4.19], viene dado por:

$$\left[\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right] \qquad [4.61]$$

en donde $t_{\alpha/2}$ es tal que

$$P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$$

y t_{n-1} sigue una t-Student con $n - 1$ grados de libertad.

La amplitud del intervalo es:

$$L = \left(\bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right) - \left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}\right) = 2t_{\alpha/2}\frac{s}{\sqrt{n}} \qquad [4.62]$$

pero si, previamente, fijamos la longitud del intervalo L, y deseamos conocer el tamaño n de la muestra para obtener el correspondiente intervalo al nivel de confianza del $100(1 - \alpha)\%$, no tendremos nada más que despejar n de la expresión [4.62] y resultará que

$$n = 4\,\frac{t_{\alpha/2}^2 s^2}{L^2} \qquad [4.63]$$

en donde s^2 sino se conoce se estimará de una muestra piloto o con información indirecta.

Ejemplo 4.18

Una empresa dedicada al transporte de viajeros en autobuses desea obtener unn intervalo al 90 % de confianza para el tiempo medio μ que tarda el autobús en realizar el recorrido entre Madrid y Granada. La longitud del intervalo se quiere que sea de 10 minutos, es decir de ± 5 minutos por encima y por debajo de la media. Se toma una muestra de 12 viajes observando los tiempos invertidos en realizar cada uno, resultando que $\bar{x} = 310$ minutos y la desviación típica muestral $s = 20$ minutos. Determinar el tamaño de la muestra que tendríamos que tomar para poder obtener el intervalo indicado.

Solución:

El intervalo de confianza al nivel del 90 %, según la expresión [4.61], será:

$$\left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

de Tabla A.10, correspondiente a la distribución t-Student, tenemos que

$$P[t_{11} > t_{0,05}] = 0{,}05 \quad \Rightarrow \quad t_{0,05} = 1{,}796$$

$$\left[310 - 1{,}796 \frac{20}{\sqrt{12}} \quad ; \quad 310 + 1{,}796 \frac{20}{\sqrt{12}}\right]$$

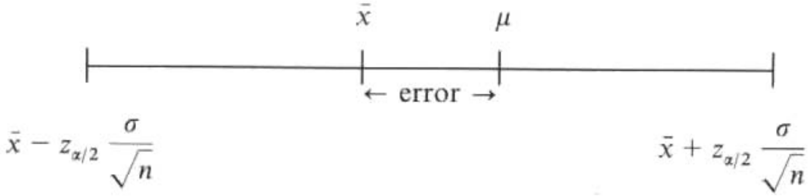
$$[299{,}6 \quad ; \quad 320{,}38]$$

Para determinar el tamaño de la muestra que tendríamos que tomar para obtener un intervalo de longitud 10 minutos, sustituimos en la expresión [4.63]

$$n = 4 \frac{(1{,}796)^2 20^2}{10^2} = 51{,}6 \simeq 52$$

resultando que necesitaríamos una muestra de 52 viajes para obtener el intervalo indicado, es decir hay que tomar 40 observaciones (viajes) aleatorias para completar la muestra previa de tamaño 12.

También podríamos hacer el siguiente razonamiento, cuando σ sea conocido, como lo hacen algunos autores, si la media μ fuera el valor central del intervalo, entonces $\bar{x}$ estimaría puntualmente a μ sin error alguno,



Pero generalmente $\bar{x}$ no será exactamente igual a μ y entonces se comete un error, $e = |\bar{x} - \mu|$, y como máximo será:

$$e = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad [4.64]$$

entonces si queremos determinar el tamaño de muestra necesario para obtener un intervalo de confianza para la media poblacional μ, admitiendo un error e, tendremos que despejando de la expresión anterior:

$$n = z^2_{\alpha/2} \frac{\sigma^2}{e^2} \qquad [4.65]$$

Análogamente ocurre para el caso en que σ no es conocida, y tendríamos:

$$n = \frac{t_{\alpha/2}^2 s^2}{e^2} \qquad (4.66]$$

Observemos que el **error** e es la mitad de la **amplitud o precisión** del intervalo L, luego las expresiones [4.60] y [4.63] son equivalentes a las expresiones [4.65] y [4.66], respectivamente.

4.7.3. TAMAÑO DE MUESTRA PARA ESTIMAR LA PROPORCIÓN p DE UNA POBLACIÓN

Sabemos que el intervalo al nivel de confianza del $100(1 - \alpha)\%$ para la proporción poblacional p es:

$$I_p = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad ; \quad \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right]$$

La longitud del intervalo es:

$$L = 2z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

y despejando el valor de n, tendremos:

$$n = 4\, \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{L^2} \qquad [4.67]$$

Expresión que utilizaremos para determinar el tamaño de la muestra necesario para obtener un intervalo de confianza para la proporción poblacional p al nivel de confianza del $100(1 - \alpha)\%$ y, con una longitud L.

Si en lugar de utilizar la amplitud L del intervalo utilizamos el *error* $e = |\hat{p} - p|$, el cual como máximo será:

$$e = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

y entonces el tamaño de muestra es:

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2} \qquad [4.68]$$

que es equivalente a [4.67], pues allí está multiplicado por 4, como veremos en el ejemplo 4.19.

El valor del estimador $\hat{p}$ se puede obtener de varias maneras:

1.ª A partir de una muestra previa.

2.ª Utilizando el valor máximo que puede tomar $\hat{p}\hat{q} = \hat{p}(1 - \hat{p})$, que se alcanzaría en:

$$\hat{p} = 0,5$$

y entonces el valor máximo de $\hat{p}\hat{q}$ será:

$$\hat{p} \cdot (1 - \hat{p}) = \hat{p}\hat{q} = 0,25$$

En efecto si consideramos los valores posibles de $\hat{p}$ tendremos los valores de $\hat{q}$ y los de $\hat{p}\hat{q}$ en la siguiente tabla:

$\hat{p}$	$\hat{q} = (1 - \hat{p})$	$\hat{p}\hat{q} = \hat{p}(1 - \hat{p})$
0,1	0,9	0,09
0,2	0,8	0,16
0,3	0,7	0,21
0,4	0,6	0,24
0,5*	0,5*	0,25*
0,6	0,4	0,24
0,7	0,3	0,21
0,8	0,2	0,16
0,9	0,1	0,09
1,0	0	0

Luego sustituyendo en la expresión [4.67] tenemos:

$$n = 4 \cdot \frac{z_{\alpha/2}^2 0,25}{L^2} \qquad [4.69]$$

que será el tamaño muestral lo suficientemente grande para garantizarnos un intervalo de confianza de longitud L.

Ejemplo 4.19

El Departamento de Estadística de una Universidad pretende estimar la proporción de licenciados matriculados en los estudios de doctorado con un error máximo del 0,05 y un nivel de confianza del 90 %. Determinar:

1. El tamaño de la muestra necesario si se tiene como información complementaria que la proporción como máximo es 0,40.

2. El tamaño de la muestra en la misma situación anterior pero con una precisión de 0,1.

3. El tamaño de la muestra cuando no se tiene información alguna acerca del valor de la proporción p y admitimos una precisión de 0,1.

Solución:

1. Aplicando la expresión [4.68]

$$n = z^2_{\alpha/2} \frac{\hat{p}\hat{q}}{e^2}$$

$$n = \frac{(1{,}645)^2 \cdot (040) \cdot (0{,}60)}{(0{,}05)^2} = 259{,}7 \simeq 260$$

2. Como la precisión es equivalente a la amplitud del intervalo, tendremos que aplicar la expresión [4.67]

$$n = 4z^2_{\alpha/2} \frac{\hat{p}\hat{q}}{L^2}$$

$$n = 4 \cdot (1{,}645)^2 \cdot \frac{(0{,}60) \cdot (0{,}40)}{(0{,}1)^2} = 259{,}7 \simeq 260$$

y vemos que efectivamente coincide con la solución anterior, como ya indicábamos en el apartado 4.7.2.

3. Como no se tiene información alguna sobre el parámetro p tomaremos el valor más desfavorable, es decir el valor de p que nos dé máximo tamaño de muestra n, y ese será el valor de p que hace máximo el producto $\hat{p}\hat{q}$, luego aplicando la expresión [4.64] o directamente la expresión [4.69] tendremos:

$$n = 4z^2_{\alpha/2} \frac{\hat{p}\hat{q}}{L^2} = 4 \cdot z^2_{\alpha/2} \frac{(0{,}5)(0{,}5)}{L^2} = 4z^2_{\alpha/2} \frac{0{,}25}{L^2}$$

$$n = 4 \cdot (1{,}645)^2 \frac{(0{,}5)(0{,}5)}{(0{,}1)^2} = 270{,}6 \simeq 271$$

4.8. REGIONES DE CONFIANZA

En algunas ocasiones podemos estar interesados en construir intervalos de confianza para más de un parámetro, es decir, en lugar de obtener un intervalo para un parámetro unidimensional lo que deseamos es que el parámetro sea k-dimensional y entonces no nos aparecerá un intervalo de confianza al nivel del $100(1-\alpha)\%$ sino que nos aparecerá una región R del espacio k-dimensional a la que llamaremos **región de confianza** al nivel de confianza del $100(1-\alpha)\%$.

Consideremos una población con función de distribución $F(x; \theta_1, ..., \theta_k)$, que depende de k parámetros desconocidos $\theta_1, ..., \theta_k$, entonces dado un nivel de confianza del $100(1-\alpha)\%$, o simplemente un coeficiente de confianza $1-\alpha$, obtendremos una región aleatoria[23] $R(X_1, ..., X_n)$ tal que

$$P[(\theta_1, ..., \theta_k) \in R(X_1, ..., X_n)] = 1 - \alpha$$

la cual nos indica que después de seleccionar la muestra aleatoria y construir la región R, tenemos una confianza del $100(1-\alpha)\%$ de que la región incluya en su interior a los parámetros $\theta_1, ..., \theta_k$.

No necesariamente la región de confianza se tiene que referir a todos los parámetros, pues en alguna situación nos puede interesar construir la región de confianza para h parámetros, $h < k$, no considerando los restantes parámetros $(\theta_{h+1}, ..., \theta_k)$ que algunos autores les llaman *parámetros perturbadores* y otros *parámetros accesorios.*

4.8.1. REGIÓN DE CONFIANZA PARA LA MEDIA μ Y VARIANZA σ^2 DE UNA POBLACIÓN NORMAL

En una primera aproximación para obtener una región de confianza para μ y σ^2 podríamos pensar en utilizar los correspondientes intervalos de confianza, obteniendo la región dada en el Gráfico 4.10, para una muestra aleatoria concreta.

- Intervalo de confianza para la media de una población normal con σ desconocida:

$$I_\mu = \left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \quad ; \quad \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right] \qquad [4.70]$$

[23] La región es aleatoria puesto que depende de la muestra aleatoria seleccionada, y para una muestra concreta obtendremos una región de confianza al nivel de $100(1-\alpha)\%$ en el espacio $\mathbb{R}^k$.

- Intervalo de confianza para la varianza de una población normal

$$I_{\sigma^2} = \left[\frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}} \quad ; \quad \frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}} \right] \qquad [4.61]$$

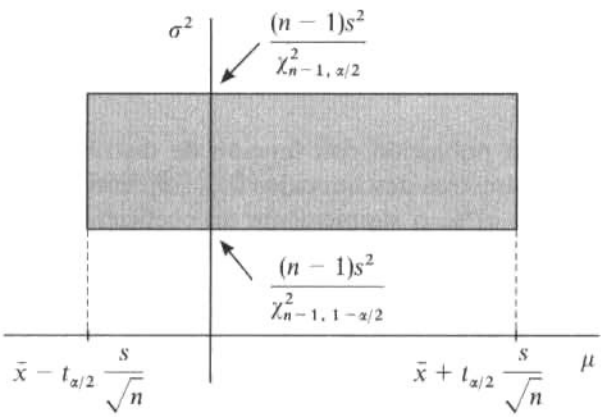


GRÁFICO 4.10. *Representación gráfica de la supuesta región delimitada por las expresiones [4.70] y [4.71].*

Ahora bien, la probabilidad de que se verifiquen simultáneamente ambos sucesos (los correspondientes a las expresiones [4.70] y [4.71], sería $(1-\alpha)^2$, es decir, suponiendo que la probabilidad conjunta de ambos sucesos fuera igual al producto de las probabilidades de cada uno. Pero esto no es correcto ya que los estadísticos utilizados, t y χ^2 no son independientes, y en consecuencia la probabilidad conjunta de ambos sucesos no es igual al producto de las probabilidades correspondientes; y por tanto no podemos decir que tenemos una confianza del $100(1-\alpha)^2$ % de que el rectángulo del Gráfico 4.10 incluya en su interior el valor de los parámetros (μ, σ^2).

Sin embargo, una forma correcta de obtener una región de confianza para los parámetros (μ, σ^2), es decir una región R tal que:

$$P[(\mu, \sigma^2) \in R(X_1, ..., X_n)] = 1 - \alpha$$

sería a partir de la distribución conjunta de los estadísticos $\bar{X}$ y $(n-1)S^2$, pues por el teorema de Fisher, son independientes. En efecto, sabemos que:

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to N(0, 1)$$

$$\frac{(n-1)S^2}{\sigma^2} \to \chi^2_{n-1}$$

Podemos obtener dos valores $-z_{\alpha'/2}$ y $z_{\alpha/2}$ tales que:

$$P\left[-z_{\alpha'/2} \leqslant \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqslant z_{\alpha'/2}\right] = \sqrt{1 - \alpha} \qquad [4.72]$$

siendo $\alpha' = 1 - \sqrt{1 - \alpha}$.

y mediante la Tabla A.9, correspondiente a la distribución χ^2, podemos encontrar dos valores[24] c_1 y c_2 tales que

$$P\left[c_1 \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant c_2\right] = \sqrt{1 - \alpha} \qquad [4.73]$$

Por otra parte, como los dos sucesos son independientes, la probabilidad conjunta será:

$$P\left[-z_{\alpha'/2} \leqslant \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqslant z_{\alpha'/2} \quad ; \quad c_1 \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant c_2\right] =$$

$$= P\left[-z_{\alpha'/2} \leqslant \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqslant z_{\alpha'/2}\right] \cdot P\left[c_1 \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant c_2\right] = 1 - \alpha \qquad [4.74]$$

De la expresión [4.72] se deduce (reemplazando los signos desiguales por iguales) que se tiene una región bidimensional en el plano (μ, σ^2), determinada a partir de las expresiones:

$$z^2_{\alpha'/2} = \frac{(\bar{X} - \mu)^2}{\sigma^2/n} = \frac{n(\bar{X} - \mu)^2}{\sigma^2} \quad \Rightarrow \quad \sigma^2 = \frac{n(\bar{X} - \mu)^2}{z^2_{\alpha'/2}}$$

$$\sigma^2 = \frac{(n-1)S^2}{c_1} \quad , \quad \sigma^2 = \frac{(n-1)S^2}{c_2}$$

[24] Los valores c_1 y c_2 son los que aparecen en la expresión [4.20], es decir:

$$c_1 = \chi^2_{n-1,\ \alpha'/2} \quad , \quad c_2 = \chi^2_{n-1,\ 1-\alpha'/2}$$

pero haciendo las correspondientes correcciones en $\frac{\alpha}{2}$ y $1 - \frac{\alpha}{2}$ como se verá en el ejemplo 4.20, es decir haciendo

$$\alpha' = 1 - \sqrt{1 - \alpha}$$

Es decir la región, Gráfico 4.11, está limitada por el arco de parábola, cuya ecuación es:

$$\sigma^2 = \frac{n(\bar{X} - \mu)^2}{z^2_{\alpha'/2}} \qquad [4.75]$$

y las rectas paralelas al eje de abscisas:

$$\sigma^2 = \frac{(n-1)S^2}{c_1} \qquad [4.76]$$

$$\sigma^2 = \frac{(n-1)S^2}{c_2} \qquad [4.77]$$

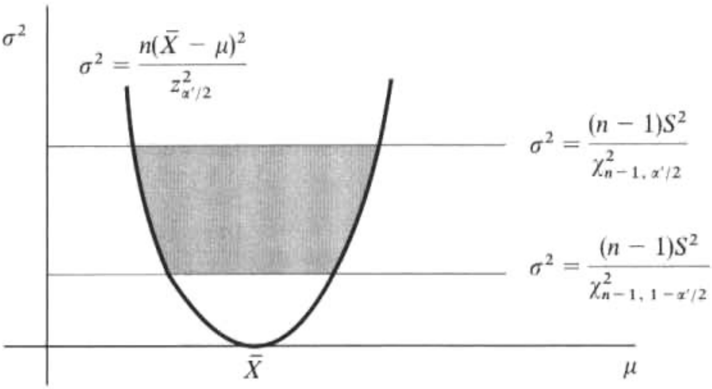


GRÁFICO 4.11. *Representación gráfica de la región de confianza para los parámetros* (μ, σ^2) *de una población normal al nivel del* $100(1 - \alpha)\%$.

Ejemplo 4.20

Construir la región de confianza al nivel del 90 % para los parámetros μ y σ^2 de una población normal, con la ayuda de una muestra aleatoria tamaño $n = 30$, en la cual $\bar{x} = 10$ y la varianza muestral $s^2 = 9$.

Solución:

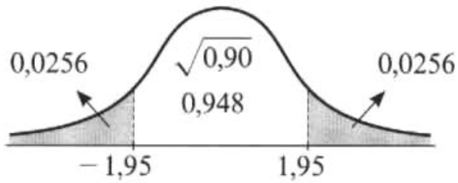
A partir de la expresión [4.75]

$$\sigma^2 = \frac{n(\bar{X} - \mu)^2}{z^2_{\alpha/2}}$$

y teniendo en cuenta la expresión [4.72], podemos calcular $z_{\alpha/2}$ ya que

$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$$

$$P[Z > z_{0,0256}] = 0,0256$$



$$z_{0,0256} = 1,95$$

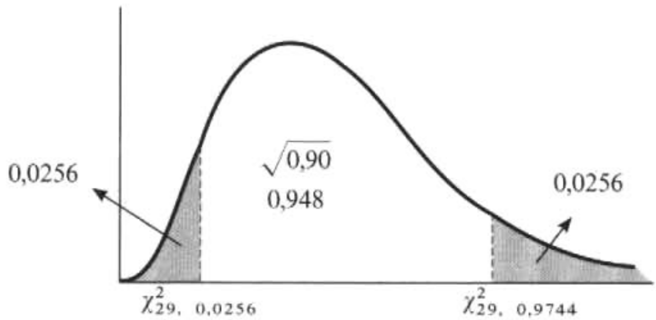
Luego la parábola será:

$$\sigma^2 = \frac{30(10 - \mu)^2}{(1,95)^2} \quad ; \quad \sigma^2 = 7,88(10 - \mu)^2$$

Las expresiones [4.76] y [4.77] nos dan:

$$\sigma^2 = \frac{(n-1)S^2}{\chi^2_{n-1,\ \alpha/2}} \quad ; \quad \sigma^2 = \frac{(n-1)S^2}{\chi^2_{n-1,\ 1-\alpha/2}}$$

y para calcular $\chi^2_{n-1,\ \alpha/2}$ y $\chi^2_{n-1,\ 1-\alpha/2}$, tendremos en cuenta la expresión [4.73], según la cual gráficamente equivale a:



$$\chi^2_{n-1,\ \alpha/2} \equiv \chi^2_{29,\ 0,0256} = 16,047$$

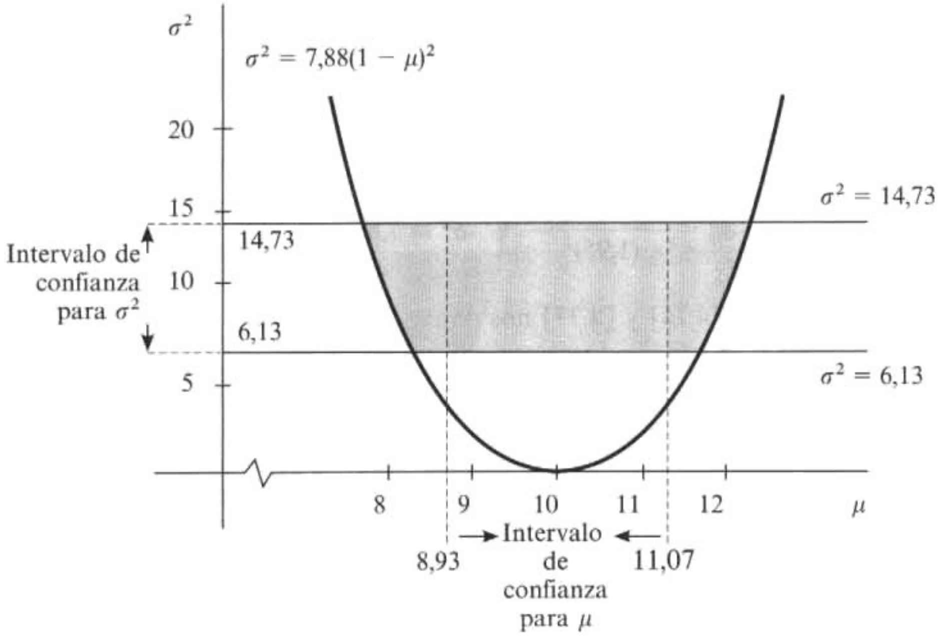
$$\chi^2_{n-1,\ 1-\alpha/2} \equiv \chi^2_{29,\ 0,9744} = 45,72$$

Luego las ecuaciones de las rectas son:

$$\sigma^2 = \frac{29 \cdot 9}{16,047} \quad ; \quad \sigma^2 = 16,26$$

$$\sigma^2 = \frac{29 \cdot 9}{45,72} \quad ; \quad \sigma^2 = 5,70$$

Gráficamente la región de confianza será:



Calculando el intervalo de confianza para la media de la población μ, utilizando S en lugar de σ resulta:

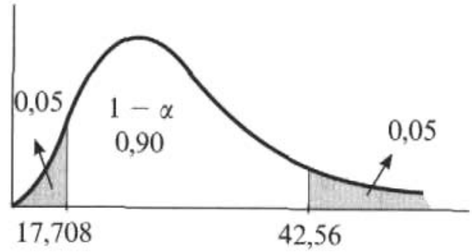
$$I_\mu = \left[\bar{X} - z_{\alpha/2} \frac{s}{\sqrt{n}} \quad ; \quad \bar{X} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

$$\left[10 - 1,95 \frac{3}{\sqrt{30}} \quad ; \quad 10 + 1,95 \frac{3}{\sqrt{30}} \right]$$

$$[8,93 \quad ; \quad 11,07]$$

Análogamente el intervalo de confianza para la varianza σ^2, será:

$$I_{\sigma^2} = \left[\frac{(n-1)S^2}{\chi^2_{n-1,\ 1-\alpha/2}} \quad ; \quad \frac{(n-1)S^2}{\chi^2_{n-1,\ \alpha/2}} \right]$$



resultando el intervalo de confianza para σ^2:

$$\left[\frac{29{,}9}{42{,}56} \quad ; \quad \frac{29{,}9}{17{,}708} \right]$$

$$[6{,}13 \quad ; \quad 14{,}73]$$

que podemos representarlo en el gráfico.

4.9. CUADRO RESUMEN DE LOS INTERVALOS DE CONFIANZA

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
$N(\mu, \sigma)$	μ $\boldsymbol{\mu}$	$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ $Z \to N(0,1)$, $P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu, \sigma)$	μ, σ $\boldsymbol{\mu}$	$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	n pequeña $\left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ $t \to t_{n-1}$, $P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$
Desconocida	μ, σ $\boldsymbol{\mu}$	$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	n grande $\left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ $Z \to N(0, 1)$, $P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu\ \sigma)$	μ, σ $\boldsymbol{\sigma}$	$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$	n pequeña $\left[\frac{(n-1)s^2}{\chi^2_{n-1,\ 1-\alpha/2}} \leqslant \sigma^2 \leqslant \frac{(n-1)s^2}{\chi^2_{n-1,\ \alpha/2}}\right]$ $P[\chi^2_{n-1} \leqslant \chi^2_{n-1,\ 1-\alpha/2}] = 1 - \frac{\alpha}{2}$, $P[\chi^2_{n-1} \leqslant \chi^2_{n-1,\ \alpha/2}] = \frac{\alpha}{2}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
$N(\mu, \sigma)$	μ, σ σ	$\hat{\sigma}^2 = s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$	n grande $\left[\frac{s^2}{1 + z_{\alpha/2} \sqrt{\frac{2}{n-1}}} \leqslant \sigma^2 \leqslant \frac{s^2}{1 - z_{\alpha/2} \sqrt{\frac{2}{n-1}}} \right]$ $Z_1 \to N(0,1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu, \sigma)$	σ σ	$\hat{\sigma}^2 = s^{*2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$	n pequeña $\left[\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\chi^2_{n,\ 1-\alpha/2}} \leqslant \sigma^2 \leqslant \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\chi^2_{n,\ \alpha/2}} \right]$ $P[\chi^2_n \leqslant \chi^2_{n,\ 1-\alpha/2}] = 1 - \frac{\alpha}{2} \quad ; \quad P[\chi^2_n \leqslant \chi^2_{n,\ \alpha/2}] = \frac{\alpha}{2}$
$N(\mu, \sigma)$	σ σ	$\hat{\sigma}^2 = s^{*2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$	n grande $\left[\frac{s^{*2}}{1 + z_{\alpha/2} \sqrt{\frac{2}{n}}} \leqslant \sigma^2 \leqslant \frac{s^{*2}}{1 - z_{\alpha/2} \sqrt{\frac{2}{n}}} \right]$ $Z \to N(0,1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y $\boldsymbol{\mu_x - \mu_y}$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$	$\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$ $\left. \leqslant (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right]$ $Z \to N(0, 1)$, $P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y σ_x, σ_y $\sigma_x = \sigma_y$ $\boldsymbol{\mu_x - \mu_y}$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$	n_x, n_y pequeñas $\left[(\bar{x} - \bar{y}) - t_{\alpha/2} \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$ $\left. \leqslant (\bar{x} - \bar{y}) + t_{\alpha/2} \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}}\right]$ $t \to t_{n_x + n_y - 2}$, $P[t_{n_x + n_y - 2} > t_{\alpha/2}] = \frac{\alpha}{2}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
Desconocidas	μ_x, μ_y σ_x, σ_y $\sigma_x = \sigma_y$ $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$	n_x, n_y grandes $\left[(\bar{x} - \bar{y}) - z_{\alpha/2} \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}} \leqslant \mu_x - \mu_y \leqslant\right.$ $\left.\leqslant (\bar{x} - \bar{y}) + z_{\alpha/2} \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \sqrt{\frac{n_x + n_y}{n_x n_y}}\right]$ $Z \to N(0, 1) \quad , \quad P[Z > z_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y σ_x, σ_y $\sigma_x \neq \sigma_y$ $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} - \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$	n_x, n_y pequeñas $\left[(\bar{x} - \bar{y}) - t_{\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant (\bar{x} - \bar{y}) + t_{\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right]$ $t \to t_\nu$ (t-Student) $\quad \nu = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y - 1}}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
Desconocidas	μ_x, μ_y σ_x, σ_y $\sigma_x \neq \sigma_y$ $\mu_x - \mu_y$	$\hat{\mu}_x = \bar{x} = \frac{1}{n_x}\sum_{i=1}^{n_x} x_i$ $\hat{\mu}_y = \bar{y} = \frac{1}{n_y}\sum_{i=1}^{n_y} y_i$ $\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x}(x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{n_y}(y_i - \bar{y})^2$	n_x, n_y grandes $\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \leqslant \mu_x - \mu_y \leqslant \right.$ $\left. \leqslant (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right]$ $Z \to N(0, 1)$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$ Apareadas	$\mu_D = \mu_x - \mu_y$ μ_D	$\hat{\bar{D}} = \bar{d} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)$ $\hat{\sigma}_D^2 = s_d^2 = \frac{1}{n - 1}\sum_{i=1}^{n}(d_i - \bar{d})^2$	n pequeño $\left[\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} \leqslant \mu_D \leqslant \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}\right]$ $t \to t_{n-1}$, $P[t_{n-1} > t_{\alpha/2}] = \frac{\alpha}{2}$
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	μ_x, μ_y σ_x, σ_y $\frac{\sigma_x^2}{\sigma_y^2}$	$\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x}(x_i - \bar{x})^2$ $\hat{\sigma}_y^2 = s_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{n_y}(y_i - \bar{y})^2$	$\left[\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{n_x-1,\, n_y-1,\, 1-\alpha/2}} \leqslant \frac{\sigma_x^2}{\sigma_y^2} \leqslant \frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{n_x-1,\, n_y-1,\, \alpha/2}}\right]$ $F_{n_x-1,\, n_y-1,\, 1-\alpha/2} = \frac{1}{F_{n_x-1,\, n_y-1,\, \alpha/2}}$

Población	Parámetros desconocidos	Estimaciones puntuales	Intervalos de confianza
$N(\mu_x, \sigma_x)$ $N(\mu_y, \sigma_y)$	σ_x, σ_y $\dfrac{\sigma_x^2}{\sigma_y^2}$	$\hat{\sigma}_x^2 = s_x^{*2} = \dfrac{1}{n_x}\sum_{i=1}^{n_x}(x_i - \mu_x)^2$ $\hat{\sigma}_y^2 = s_y^{*2} = \dfrac{1}{n_y}\sum_{i=1}^{n_y}(y_i - \mu_y)^2$	$\left[\dfrac{s_x^{*2}}{s_y^{*2}} \cdot \dfrac{1}{F_{n_x, n_y, 1-\alpha/2}} \leqslant \dfrac{\sigma_x^2}{\sigma_y^2} \leqslant \dfrac{s_x^{*2}}{s_y^{*2}} \cdot \dfrac{1}{F_{n_y, n_x, \alpha/2}}\right]$
$B(1, p)$	p p	$\hat{p} = \dfrac{x}{n} = \dfrac{\text{n.º de éxitos en } n \text{ pruebas}}{\text{n.º de pruebas}}$	n pequeño Gráficos: Tabla A.13 n grande $\left[\hat{p} - z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \leqslant p \leqslant \hat{p} + z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\right]$ $Z \to N(0, 1)$, $P[Z > z_{\alpha/2}] = \dfrac{\alpha}{2}$
$B(1, p_X)$ $B(1, p_Y)$	p_X, p_Y $p_X - p_Y$	$\hat{p}_x = \dfrac{x}{n}$ $\hat{p}_y = \dfrac{y}{n}$	$\left[(\hat{p}_x - \hat{p}_y) - z_{\alpha/2}\sqrt{\dfrac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \dfrac{\hat{p}_y(1-\hat{p}_y)}{n_y}} \leqslant p_X - p_Y \leqslant\right.$ $\left.\leqslant (\hat{p}_x - \hat{p}_y) + z_{\alpha/2}\sqrt{\dfrac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \dfrac{\hat{p}_y(1-\hat{p}_y)}{n_y}}\right]$ $Z \to N(0, 1)$, $P[Z > z_{\alpha/2}] = \dfrac{\alpha}{2}$