

ESTADÍSTICA

• Objeto de la estadística \Rightarrow análisis de datos obtenidos experimentalmente. Por ejemplo se quiere medir el valor de una magnitud con un instrumento científico. Los valores que nos da el aparato no siempre serán iguales (P. ej. medida del peso con una balanza) debido a variables externas y desconocidas que actúan sobre el aparato de medida.

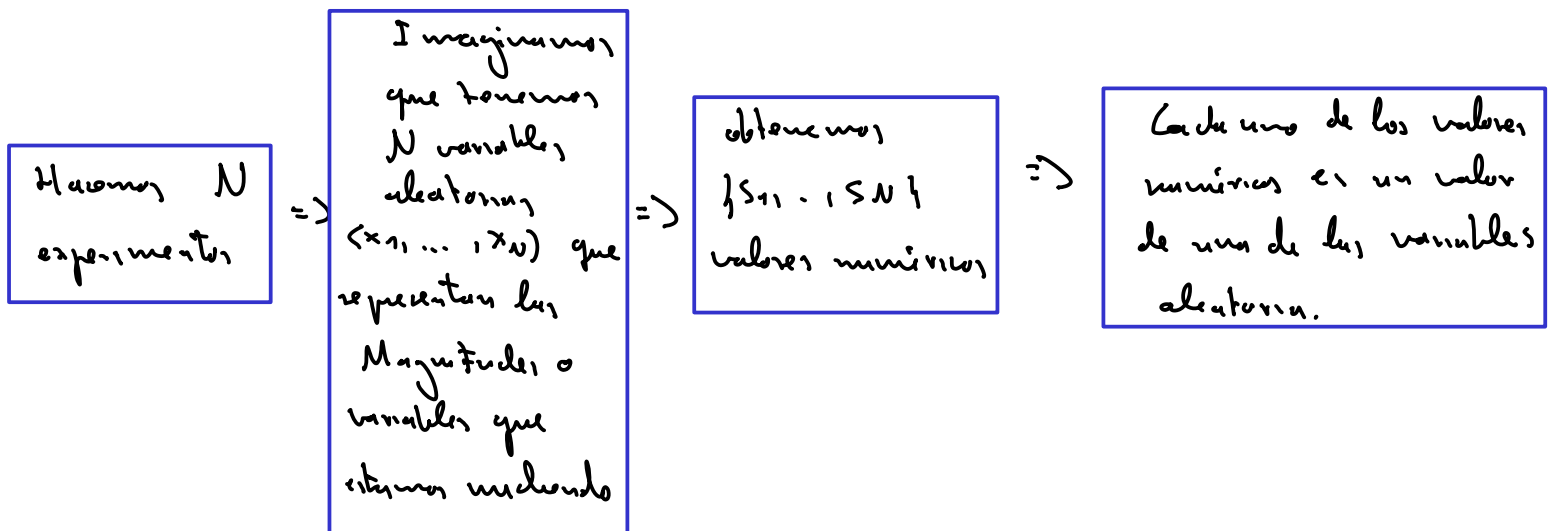
Más ejemplos de experimentos {

- * lanzar piedras y medir las coordenadas del punto de impacto.
- * Medir el consumo de combustible de diversos vehículos cuando hacen el viaje de Bilbao a Madrid.
- * Llamar por teléfono a los habitantes de un país y preguntarle su edad.

• En la práctica podemos considerar que los experimentos son experimentos aleatorios porque o bien no conocemos las leyes subyacentes, o no nos interesan.

• Experimento \Rightarrow N datos numéricos $\Rightarrow \{s_1, s_2, \dots, s_N\}$

• Podemos considerar los valores numéricos obtenidos en los experimentos como valores de una variable aleatoria



Terminología básica:

Población: Es el recorrido de todas las variables aleatorias (los valores que pueden tomar las magnitudes que estamos midiendo)

$$x_1: I_1, x_2: I_2 \dots x_N: I_N \Rightarrow \text{Población} = I_1 \times I_2 \times \dots \times I_N$$

Población es lo análogo al espacio muestral del cálculo de probabilidades

- Ejemplo de las piedras $\Rightarrow \text{Población} = \mathbb{R}^N$ (suponiendo que las piedras caen en una recta)
- Ejemplo de los coches $\Rightarrow \text{Población} = [0, 1000]^N = \overbrace{[0, 1000] \times [0, 1000] \times \dots \times [0, 1000]}^N$
- Ejemplo de las edades $\Rightarrow \text{Población} = [1, 120 \text{ años}]^N = [1, 120 \text{ años}] \times [1, 120 \text{ años}] \times \dots \times [1, 120 \text{ años}]$

En general la población depende del número de veces que repetimos el experimento.

Muestra de la población: Valores concretos de los resultados de N mediciones.

Un proceso de una o más mediciones se conoce como un muestreo de la población.

* Ejemplo de las piedras $\{ 0.3 \text{ m}, 1.38 \text{ m}, 0.68 \text{ m}, 1.01 \text{ m} \}$ ($N=4$)

* Ejemplo de los coches $\{ 19 \text{ l}, 25 \text{ l}, 43 \text{ l}, 27 \text{ l} \}$ ($N=4$)

* Ejemplo de las edades $\{ 5 \text{ años}, 101 \text{ años}, 25 \text{ años}, 39 \text{ años} \}$ ($N=4$)

* Si tenemos N experimentos sucesivos tenemos N variables aleatorias (x_1, x_2, \dots, x_N)

¿Cuál será la distribución de probabilidad conjunta $P(x_1, x_2, \dots, x_N)$?

* La distribución de probabilidad P se conoce como distribución de población.

* En general la distribución P es desconocida. Uno de los objetivos fundamentales de la estadística es deducir propiedades de P a partir de los valores de un conjunto de medidas mediante un proceso conocido como inferencia estadística.

* Si las medidas son independientes unas de otras entonces las variables aleatorias $\{x_1, \dots, x_N\}$ serán independientes y la distribución de probabilidad conjunta

P se factorizará:

$$P(x_1, \dots, x_N) = P^1(x_1) P^2(x_2) \dots P^N(x_N) \Rightarrow \left\{ \begin{array}{l} \text{Las } P^i(x_i) \text{ son distribuciones} \\ \text{de probabilidad que en general} \\ \text{serán diferentes.} \end{array} \right.$$

* Ejemplo: hacemos N medidas de una magnitud física con aparatos diferentes o bajo condiciones experimentales diferentes: $\Rightarrow P^1 \neq P^2 \neq \dots \neq P^N$

* Ejemplos de la peder, del consumo de combustible: los experimentos son independientes

Ejemplo de la edad: los experimentos son dependientes porque la misma persona podría contestar a llamadas diferentes (aunque los números sean diferentes) **¡¡ comparable!!**

Si $P^1 = P^2 = \dots = P^N$ entonces $P(x_1, \dots, x_N) = P(x_1) P(x_2) \dots P(x_N)$

Esto sucederá por ejemplo para un proceso de medida de una cierta magnitud efectuado bajo las mismas condiciones experimentales o con los mismos aparatos.

Esta propiedad se cumplirá en todos los ejemplos anteriores. Si ahora hacemos el experimento de la medida de los consumos de combustible y hacemos las 3 primeras medidas con un aparato las 2 siguientes con otro y en las dos últimas cambiamos el recorrido Bilbao - Madrid por Bilbao - Barcelona, entonces se tiene que:

$$P = P^1(x^1) P^1(x^2) P^1(x^3) \cdot P^2(x^4) P^2(x^5) P^3(x^6) P^3(x^7) \quad P^1 \neq P^2 \neq P^3$$

• En el caso de que las distribuciones de probabilidad de las variables aleatorias $\{x_1, x_2, \dots, x_N\}$ asociadas a las medidas coincidan entonces podemos trabajar con una única variable aleatoria x . En este caso particular se dice que las medidas $\{s_1, s_2, \dots, s_N\}$ constituyen una muestra aleatoria de N elementos de una población, que es el recorrido de la variable aleatoria x . Este será el caso del que nos ocuparemos.

Estadística descriptiva

• Estadísticos muestrales

Dada una muestra aleatoria de una población $\{s_1, \dots, s_N\}$, un estadístico es una función de los valores de la muestra que no contiene valores desconocidos. La idea es utilizar diferentes estadísticos para obtener propiedades de los datos.

• Media aritmética: Para una muestra $\{s_1, \dots, s_N\}$, la media aritmética \bar{s} es un estadístico que se define así

$$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i$$

► Table 31.1 gives eight values for the round trip time in milliseconds for a computer 'packet' to travel from Cambridge UK to Cambridge MA. Find the sample mean.

188.7	204.7	193.2	169.0
168.1	189.8	166.3	200.0

table 31.1

Para aquellos que usen Linux y MAC (¿También Windows?), esta información se puede obtener usando el comando "ping".

$$\begin{aligned} \bar{s} &= \frac{1}{8} (188.7 + 204.7 + 193.2 + 169.0 + 168.1 + 189.8 + 166.3 + 200.0) \\ &= \frac{1479.8}{8} = 184.975 \Rightarrow \text{redondear hasta dejar un decimal, que es lo} \\ &\quad \text{que traían los datos originales,} \\ &\quad 185.0 \end{aligned}$$

Además de la media aritmética hay más tipos de medias:

$$\text{Media geométrica: } \bar{s}_g = \left(\prod_{i=1}^N s_i \right)^{\frac{1}{N}}$$

$$\text{Media armónica: } \bar{s}_h = \frac{N}{\sum_{i=1}^N \frac{1}{s_i}}$$

$$\text{¿Valor eficaz? : } \bar{s}_e = \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{\frac{1}{2}}$$

⇒ Moda: el valor que aparece con más frecuencia en la muestra $\{s_1, \dots, s_N\}$

⇒ Mediana: Si $s_1 \leq s_2 \leq \dots \leq s_N$ entonces la mediana es $\frac{s_{N+1}}{2}$ (N impar)

$$\frac{1}{2}(s_{\frac{N}{2}} + s_{\frac{N}{2}+1}) \quad (N \text{ par})$$

⇒ En el ejemplo: cada valor solo aparece una vez, así que todos los valores cumplen con la definición de moda

$$\text{mediana: } \frac{1}{2}(s_4 + s_5) = 189.3$$

⇒ Varianza y desviación estándar ⇒ ahora tenemos que usar $\{x_1, \dots, x_N\}$ para los valores de la muestra:

$$\bullet \text{ Varianza: } s_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2, \quad \text{Desviación típica } s_x = \left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^{\frac{1}{2}}$$

$$\left(\begin{aligned} s_x^2 &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N (x_i^2 + \bar{x}^2 - 2x_i\bar{x}) = \frac{1}{N} \sum_{i=1}^N x_i^2 + \bar{x}^2 - 2\bar{x}^2 = \overline{x^2} - \bar{x}^2 \end{aligned} \right)$$

$$\text{Con la notación: } \overline{x^2} = \frac{1}{N} \sum_{i=1}^N x_i^2$$

• Momentos de la muestra: $m_r = \frac{1}{N} \sum_{i=1}^N x_i^r$, $r = (1, 2, \dots)$

• Momentos centrales de la muestra $\Rightarrow n_r = \frac{1}{N} \sum_{i=1}^N (x_i - m_1)^r$, $r = (1, 2, \dots)$

Se tiene que

$$m_1 = \bar{x} \quad s_x^2 = \bar{x^2} - \bar{x}^2 \Rightarrow n_2 = m_2 - m_1^2$$

$$m_2 = \overline{x^2}$$

$$s_x^2 = n_2$$

Es posible expresar los momentos centrales en términos de los momentos de la muestra.

$$\begin{aligned} \text{Ej. } n_3 &= \frac{1}{N} \sum_{i=1}^N (x_i - m_1)^3 = \frac{1}{N} \sum_{i=1}^N (x_i^3 - 3x_i^2 m_1 + 3x_i m_1^2 - m_1^3) = m_3 - 3m_2 m_1 + 3m_1^3 - m_1^3 = \\ &= m_3 - 3m_2 m_1 + 2m_1^3 \end{aligned}$$

• Descripción de la muestra en términos de cantidades adimensionales

$$g_1 = \frac{n_1}{n_2^{1/2}}, \quad g_2 = \frac{n_2}{n_2} = 1, \quad g_3 = \frac{n_3}{n_2^{3/2}} = \text{coeficiente de asimetría}$$

$$g_4 = \frac{n_4}{n_2^2} = \text{curtosis}$$

$$g_4 - 3 = \text{exceso de curtosis}$$

• Covarianza y correlación. Hasta ahora hemos supuesto que los datos son cantidades unidimensionales $\{x_1, \dots, x_N\}$. Supongamos ahora que los datos vienen dados por pares de números

$\{(x_1, y_1), \dots, (x_N, y_N)\}$ (muestra bidimensional). La covarianza de la muestra se define así

$$\begin{aligned} V_{xy} &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) = \overline{(x - \bar{x})(y - \bar{y})} = \frac{1}{N} \left(\sum_{i=1}^N x_i y_i - x \bar{y} - \bar{x} y + \bar{x} \bar{y} \right) = \\ &= \overline{xy} - \bar{x} \bar{y} - \bar{x} \bar{y} + \bar{x} \bar{y} = \overline{xy} - \bar{x} \bar{y} = \frac{1}{N} \sum_{i=1}^N x_i y_i - \frac{1}{N^2} \left(\sum_{i=1}^N x_i \right) \left(\sum_{i=1}^N y_i \right) \end{aligned}$$

* Correlación de la muestra: $r_{xy} \equiv \frac{V_{xy}}{s_x s_y} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2\right)^{\frac{1}{2}} \left(\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2\right)^{\frac{1}{2}}} =$

$= \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^N (y_i - \bar{y})^2\right)^{\frac{1}{2}}} = \frac{\vec{X} \cdot \vec{Y}}{\|\vec{X}\| \|\vec{Y}\|}$

$\vec{X} = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_N - \bar{x} \end{pmatrix} \quad \vec{Y} = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_N - \bar{y} \end{pmatrix}$

$|r_{xy}| = \frac{|\vec{X} \cdot \vec{Y}|}{\|\vec{X}\| \|\vec{Y}\|} \leq \frac{\|\vec{X}\| \|\vec{Y}\|}{\|\vec{X}\| \|\vec{Y}\|} = 1 \Rightarrow -1 \leq r_{xy} \leq 1$

← construyamos vectores con los datos

desigualdad
de Cauchy-Schwarz

* Se tiene que $r_{xy} = 0 \Leftrightarrow V_{xy} = 0 \Leftrightarrow \overline{xy} = \bar{x} \cdot \bar{y}$

$\vec{e} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

* Si $y_i = a x_i + b \Rightarrow \bar{y} = a \bar{x} + b \quad \vec{Y} = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_N - \bar{y} \end{pmatrix} = \begin{pmatrix} a x_1 + b - a \bar{x} - b \\ \vdots \\ a x_N + b - a \bar{x} - b \end{pmatrix} = a \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_N - \bar{x} \end{pmatrix} = a \vec{X}$

así que $r_{xy} = \frac{\vec{X} \cdot \vec{Y}}{\|\vec{X}\| \|\vec{Y}\|} = \frac{a \vec{X} \cdot \vec{X}}{a \|\vec{X}\| \|\vec{X}\|} = \frac{a}{|a|} \frac{\|\vec{X}\|^2}{\|\vec{X}\|^2} = \text{sign}(a) = \pm 1$

La correlación alcanza sus valores extremos si existe una relación afín (lineal) entre el 2º y el 1º por.

$\text{sign}(a) = +1 \Rightarrow r_{xy} = 1$

$\text{sign}(a) = -1 \Rightarrow r_{xy} = -1$

La otra implicación también es cierta: $r_{xy} = \pm 1 \Rightarrow \vec{X} \cdot \vec{Y} = \pm \|\vec{X}\| \|\vec{Y}\| \Rightarrow$ los vectores \vec{X}, \vec{Y} son proporcionales $\Rightarrow y_i - \bar{y} = a(x_i - \bar{x}) \Rightarrow y_i = a x_i + \bar{y} - a \bar{x}$

→ Si aplicamos una transformación afín a los datos bidimensionales entonces se tiene que

$$(x_i, y_i) \mapsto (x'_i, y'_i) = (ax_i + b, cy_i + d) \Rightarrow \bar{x}' = a\bar{x} + b, \bar{y}' = c\bar{y} + d$$

$$\vec{X}' = \begin{pmatrix} x'_1 - \bar{x}' \\ \vdots \\ x'_N - \bar{x}' \end{pmatrix} = \begin{pmatrix} ax_1 + b - a\bar{x} - b \\ \vdots \\ ax_N + b - a\bar{x} - b \end{pmatrix} = a \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_N - \bar{x} \end{pmatrix} = a\vec{X}, \quad \vec{Y}' = c\vec{Y}$$

$$\text{así que } r_{x'y'} = \frac{\vec{X}' \cdot \vec{Y}'}{\|\vec{X}'\| \|\vec{Y}'\|} = \frac{ac \vec{X} \cdot \vec{Y}}{|a| |\vec{X}| |c| |\vec{Y}|} = \text{Sign}(ac) r_{xy}$$

El valor de la correlación no se altera si efectuamos un cambio de origen o escala con signos positivos en las magnitudes x o y .

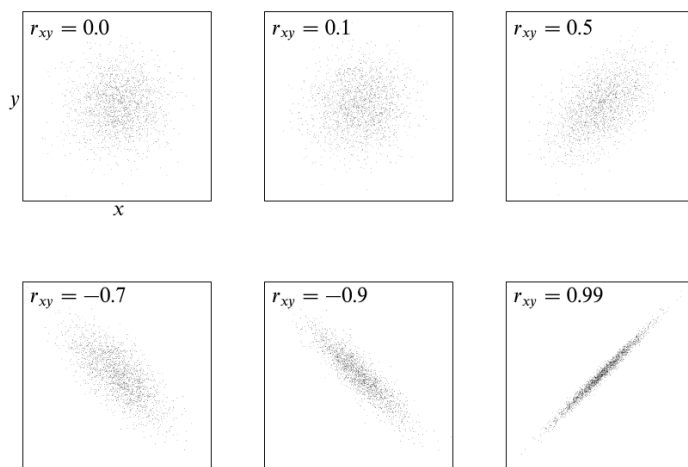


Figure 31.1 Scatter plots for two-dimensional data samples of size $N = 1000$, with various values of the correlation r . No scales are plotted, since the value of r is unaffected by shifts of origin or changes of scale in x and y .

► Ten UK citizens are selected at random and their heights and weights are found to be as follows (to the nearest cm or kg respectively):

Person	A	B	C	D	E	F	G	H	I	J
Height (cm)	194	168	177	180	171	190	151	169	175	182
Weight (kg)	75	53	72	80	75	75	57	67	46	68

Calculate the sample correlation between the heights and weights.

$$r_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{s_x s_y}$$

$$s_x^2 = \overline{x^2} - \bar{x}^2$$

$$s_y^2 = \overline{y^2} - \bar{y}^2$$

$$\sum_{i=1}^{10} x_i = 1757, \quad \sum_{i=1}^{10} y_i = 668$$

$$\Downarrow$$

$$\bar{x} = 175.7, \quad \bar{y} = 66.8$$

$$\sum_{i=1}^{10} x_i y_i = 118079$$

$$\Downarrow$$

$$\overline{xy} = 11807.9$$

$$\sum_{i=1}^{10} x_i^2 = 310041, \quad \sum_{i=1}^{10} y_i^2 = 45746$$

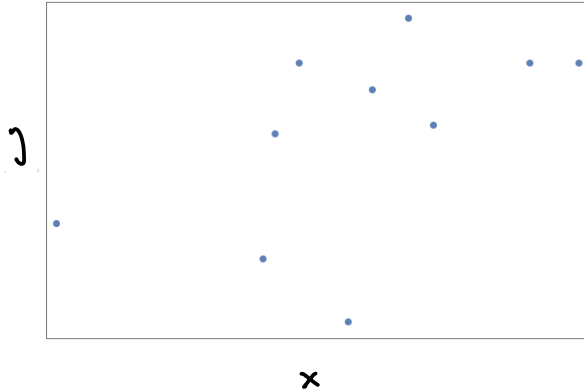
$$\Downarrow$$

$$\overline{x^2} = 31004.1, \quad \overline{y^2} = 4574.6$$

$$\left. \begin{aligned} S_x^2 &= \overline{x^2} - \bar{x}^2 = 31004.1 - 175.7^2 = 137.61 \Rightarrow S_x = 11.6 \\ S_y^2 &= \overline{y^2} - \bar{y}^2 = 4574.6 - 66.8^2 = 112.36 \Rightarrow S_y = 10.6 \end{aligned} \right\} r_{xy} = \frac{11807.9 - 175.7 \times 66.8}{11.6 \times 10.6} = 0.54$$

$$\Downarrow$$

Representación
gráfica de
los puntos.



hay una correlación positiva
entre la altura y el
peso.

• Esto se puede generalizar para muestras multidimensionales,

n dimensión
de la muestra
 N número
de elementos
de la muestra

$$\left(\begin{array}{c} x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(n)} \\ x_c^{(1)}, x_c^{(2)}, \dots, x_c^{(n)} \\ x_N^{(1)}, x_N^{(2)}, \dots, x_N^{(n)} \end{array} \right) \Rightarrow$$

$$V_{kl} = \overline{x^{(k)} x^{(l)}} - \overline{x^{(k)}} \overline{x^{(l)}}$$

Matriz de
covarianza de la muestra

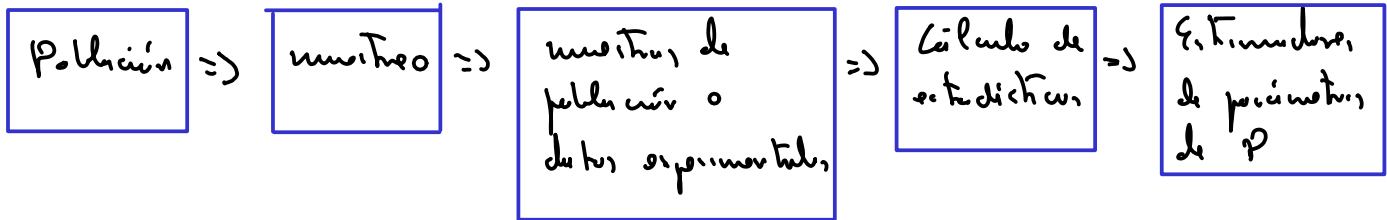
$$r_{kl} = \frac{\overline{x^{(k)} x^{(l)}} - \overline{x^{(k)}} \overline{x^{(l)}}}{S_k S_l}$$

Matriz de correlación de la
muestra.

V_{kl} y S_{kl} son matrices simétricas pero no tienen por qué ser definidas positivas.

Inferencia estadística

Ya hemos dicho que la distribución de población P es, en general, desconocida. El objetivo de la inferencia estadística es obtener o estimar propiedades de P a partir de muestras aleatorias de la población.



La distribución de población P es una distribución de probabilidad que depende de parámetros

$$P = P(x | \vec{a}) = P(x | a_1, \dots, a_n) \quad \left(\text{Recuérdense las expresiones de las distribuciones de probabilidad que hemos estudiado} \right)$$

En el caso que nos ocupa P es unidimensional, porque los experimentos/medidas son independientes y además se realizan bajo las mismas condiciones. Así que

$$P(x^1, \dots, x^N | \vec{a}) = P(x^1 | \vec{a}) \cdot \dots \cdot P(x^N | \vec{a})$$

* **Estimadores:** un estimador es una función $\hat{a} = \hat{a}(x_1, \dots, x_N)$ de los valores de una muestra $\{x_1, \dots, x_N\}$ que se emplea para estimar uno o varios parámetros de la distribución de probabilidad P . La función \hat{a} puede ser escalar si sólo queremos estimar un parámetro o vectorial si estimamos estimando más de un parámetro.

Como los valores $\{x_1, \dots, x_N\}$ se consideran valores de una variable aleatoria, se sigue que la transformación $\hat{a} = \hat{a}(x_1, \dots, x_N)$ se puede considerar como una

Transformación de variables aleatorias, siendo la propia \hat{a} una variable aleatoria con una distribución de probabilidad asociada. Esta distribución se conoce como "distribución de la muestra" y se representa así

$$P(\hat{a}|a)$$

↙ recordemos que \hat{a} sirve para estimar a

$$\int_I P(\hat{a}|a) d\hat{a} = \int_S P(x_1|a) P(x_2|a) \dots P(x_N|a) dx^1 \dots dx^N$$

$$\hat{a}(S) = I$$



De esta fórmula se deduce que $P(\hat{a}|a)$ dependerá de la relación $\hat{a} = \hat{a}(x_1, \dots, x_N)$, de la distribución de la población $P(x|a)$ y del número N de experimentos que hagamos ($N = \text{tamaño de la muestra}$)

Ejemplo:

Se sabe que la distribución de población asociada a una muestra $\{x_1, \dots, x_N\}$ es la distribución normal $N(\mu, \sigma)$. Supongamos que escogemos la media \bar{x} de la muestra como estimador $\hat{\mu}$ de la media de la distribución de población. ¿Cuál es la distribución muestral del estimador $\hat{\mu}$?

$$\text{Distribución normal: } N(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$\text{Estimador } \hat{\mu} = \hat{\mu}(x_1, \dots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i$$

Apliquemos la fórmula

$$\int_I P(\hat{\mu}|\mu, \sigma) d\hat{\mu} = \int_S P(x_1|\mu, \sigma) \dots P(x_N|\mu, \sigma) dx^1 \dots dx^N$$

$I = \hat{\mu}(S)$

$$\bar{I} = [\hat{\mu}_0, \hat{\mu}_0 + h], \quad S = S_1 \times \dots \times S_N$$

$$\int_{\hat{\mu}_0}^{\hat{\mu}_0 + h} P(\hat{\mu} | \mu, \sigma) d\hat{\mu} = \int_S \left(\frac{1}{\sigma\sqrt{N}} \right)^N \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right] dx_1 \dots dx_N$$

La región S viene dada por:

$$S = \left\{ \hat{\mu}_0 \leq \frac{1}{N} \sum_{i=1}^N x_i \leq \hat{\mu}_0 + h, \quad -\infty < x_1 < \infty, \dots, -\infty < x_{N-1} < \infty \right\}$$

$$\hat{\mu}_0 \leq \frac{1}{N} \left(\sum_{i=1}^{N-1} x_i \right) + \frac{x_N}{N} \leq \hat{\mu}_0 + h \Rightarrow N\hat{\mu}_0 - \sum_{i=1}^{N-1} x_i \leq x_N \leq N(\hat{\mu}_0 + h) - \sum_{i=1}^{N-1} x_i$$

Así que:

$$\int_{\hat{\mu}_0}^{\hat{\mu}_0 + h} P(\hat{\mu} | \mu, \sigma) d\hat{\mu} = \left(\frac{1}{\sigma\sqrt{N}} \right)^N \times$$

$$\int_{-\infty}^{\infty} dx_1 e^{-\frac{1}{2\sigma^2}(x_1 - \mu)^2} \dots \int_{-\infty}^{\infty} dx_{N-1} \exp \left[-\frac{1}{2\sigma^2} (x_{N-1} - \mu)^2 \right] \int_{N\hat{\mu}_0 - \sum_{i=1}^{N-1} x_i}^{(\hat{\mu}_0 + h)N - \sum_{i=1}^{N-1} x_i} \exp \left[-\frac{1}{2\sigma^2} (x_N - \mu)^2 \right] dx_N$$

Derivamos respecto a h y hacemos $h=0$

$$\hat{P}(\hat{\mu}_0 | \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{N}} \right)^N \int_{-\infty}^{\infty} dx_1 e^{-\frac{1}{2\sigma^2}(x_1 - \mu)^2} \dots \int_{-\infty}^{\infty} dx_{N-1} \exp \left[-\frac{1}{2\sigma^2} (x_{N-1} - \mu)^2 \right] \times$$

$$N \exp \left[-\frac{1}{2\sigma^2} \left(\hat{\mu}_0 N - \sum_{i=1}^{N-1} x_i - \mu \right)^2 \right]$$

Definimos:

$$I(x_1, \dots, x_{N-2}) = \int_{-\infty}^{\infty} dx_{N-1} \exp \left[-\frac{1}{2\sigma^2} (x_{N-1} - \mu)^2 \right] \exp \left[-\frac{1}{2\sigma^2} \left(\hat{\mu}_0 N - \sum_{i=1}^{N-2} x_i - x_{N-1} - \mu \right)^2 \right]$$

Operando:

$$I(x_1, \dots, x_{N-2}) = \int_{-\infty}^{\infty} dx_{N-1} \exp \left[-\frac{1}{2\sigma^2} \left((x_{N-1} - \mu)^2 + \left(x_{N-1} - \hat{\mu}_0 N + \mu + \sum_{i=1}^{N-2} x_i \right)^2 \right) \right]$$

Esta integral se puede resolver si tenemos en cuenta la siguiente identidad:

$$\int_{-\infty}^{\infty} \exp[-A(px-a)^2 - B(qx-b)^2] dx = \sqrt{\frac{\pi}{Ap^2 + Bq^2}} \exp\left[-\frac{AB(bp-aq)^2}{Ap^2 + Bq^2}\right]$$

Así que

$$I(x_1, \dots, x_{N-2}) = \sqrt{\frac{\pi}{2 \frac{1}{2\sigma^2}}} \exp \left[-\frac{\left(\frac{1}{2\sigma^2}\right) \left(\mu - \hat{\mu}_0 N + \mu + \sum_{i=1}^{N-2} x_i \right)^2}{2 \frac{1}{2\sigma^2}} \right] =$$

$$= \sigma \sqrt{\pi} \exp \left[-\frac{1}{2} \frac{1}{2\sigma^2} \left(\mu - \hat{\mu}_0 N + \mu + \sum_{i=1}^{N-2} x_i \right)^2 \right] = \sigma \sqrt{\pi} \exp \left[-\frac{1}{4\sigma^2} \left(2\mu - \hat{\mu}_0 N + \sum_{i=1}^{N-2} x_i \right)^2 \right]$$

Sustituimos en la integral

$$\left(\frac{1}{\sigma\sqrt{\pi}}\right)^N N \sigma \sqrt{\pi} \int_{-\infty}^{\infty} dx_1 e^{\frac{1}{2\sigma^2}(x_1 - \mu)^2} \dots \int_{-\infty}^{\infty} dx_{N-2} \exp \left[-\frac{1}{2\sigma^2} (x_{N-2} - \mu)^2 \right] \times$$

$$\exp \left[-\frac{1}{4\sigma^2} \left(2\mu - \hat{\mu}_0 N + \sum_{i=1}^{N-2} x_i \right)^2 \right]$$

$$\Rightarrow \text{Un cálculo largo da } P(\hat{\mu}_0 | \mu, \sigma) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{N}}} \exp \left[-\frac{(\hat{\mu}_0 - \mu)^2}{\frac{2\sigma^2}{N}} \right] \text{ y como } \hat{\mu}_0 \text{ es}$$

$$\text{arbitrario la distribución muestral es } P(\hat{\mu} | \mu, \sigma) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{N}}} e^{-\frac{(\hat{\mu} - \mu)^2}{\frac{2\sigma^2}{N}}} \Rightarrow N\left(\mu, \frac{\sigma}{\sqrt{N}}\right)$$

Propiedades básicas de los estimadores

* Para cada parámetro α de la distribución de población $P(x|\alpha)$ es posible definir un estimador $\hat{\alpha}$ que no es único. Podemos analizar la calidad de un estimador $\hat{\alpha}$ estudiando ciertas propiedades de su distribución muestral $P(\hat{\alpha}|\alpha)$. Existen 3 criterios básicos para juzgar la calidad de un estimador:

* **Consistencia:** si el tamaño N de la muestra tiende a ∞ , entonces el estimador $\hat{\alpha}$ debe coincidir con la cantidad α que está estimando.

$$\lim_{N \rightarrow \infty} \hat{\alpha} = \alpha$$

* **Sesgo de un estimador:** El sesgo de un estimador se define así: calculamos primero el valor esperado $E(\hat{\alpha})$ con respecto a la distribución muestral

$$E(\hat{\alpha}) = \int \hat{\alpha} P(\hat{\alpha}|\alpha) d\hat{\alpha} \Rightarrow \text{El sesgo } b(\alpha) \text{ se define así } b(\alpha) = E(\hat{\alpha}) - \alpha$$

Dado que $E(\hat{\alpha})$ depende del tamaño de la muestra (porque $\hat{\alpha}$ depende en general del tamaño de la muestra) se deduce que el sesgo $b(\alpha)$ dependerá también del tamaño de la muestra. Se dice que un estimador no tiene sesgo si $b(\alpha) = 0$

Un estimador $\hat{\alpha}$ posee un sesgo $b(\alpha)$ dado por $b(\alpha) = (b_1 - 1)\alpha + b_2$ siendo b_1 y b_2 constantes constantes. Definir un estimador sin sesgo para α .

Por definición:

$$E(\hat{\alpha}) = \alpha + b(\alpha)$$

$$\text{Definimos el estimador de } \alpha : \hat{\hat{\alpha}} = \alpha \hat{\alpha} + \beta \Rightarrow$$

$$E(\hat{\hat{\alpha}}) = \alpha + (b_1 - 1)\alpha + b_2 = \\ = \alpha b_1 + b_2$$

$$E(\hat{\hat{\alpha}}) = \alpha E(\hat{\alpha}) + \beta = \alpha(\alpha b_1 + b_2) + \beta =$$

$$\text{Sesgo de } \hat{\hat{\alpha}} \Rightarrow b(\hat{\hat{\alpha}}) = \alpha(\alpha b_1 + b_2) + \beta - \alpha = \alpha^2 b_1 + \alpha b_2 + \beta - \alpha$$

$$\text{Si el sesgo es cero entonces : } \alpha(\alpha b_1 - 1) + \alpha b_2 + \beta = 0 \Rightarrow \text{Tomar } \alpha = \frac{1}{b_1} \quad \beta = -\frac{b_2}{b_1}$$

$$\text{El estimador no sesgado será } \hat{\hat{\alpha}} = \frac{\hat{\alpha}}{b_1} - \frac{b_2}{b_1}$$

2 Eficiencia de un estimador.

Considerare la varianza de la distribución muestral asociada a un estimador \hat{a}

$$V(\hat{a}) = \int (\hat{a} - E(\hat{a}))^2 P(\hat{a}|a) d\hat{a} \Rightarrow V(\hat{a}) \text{ es decir la extensión que está } P(\hat{a}|a) \text{ en torno al valor medio } E(\hat{a})$$

Un estimador \hat{a} cuya varianza $V(\hat{a})$ es pequeña medio $E(\hat{a})$ es más eficiente que otro estimador \hat{a}' con una varianza $V(\hat{a}')$ mayor.

* Desigualdad de Fisher - Cramér-Rao

Si $P(x|a)$ es la distribución de población original entonces la desigualdad de Fisher - Cramér-Rao nos permite obtener una cota inferior para $V[\hat{a}]$, en términos del sesgo $b(a)$ y la distribución de población $P(x|a)$

$$V[\hat{a}] \geq \frac{(1 + \frac{\partial}{\partial a} b(a))^2}{E\left[-\frac{\partial^2}{\partial a^2} \ln P(x|a)\right]} = V_{\min}(a)$$

Se define la eficiencia o precisión e del estimador $V[\hat{a}]$ así: $e \equiv \frac{V_{\min}}{V[\hat{a}]} \leq 1$

Si $e=1$ entonces \hat{a} es un estimador de varianza mínima o eficiente. En caso contrario $e < 1$ y hablamos de un estimador no eficiente o ineficiente.

Sea $\{x_1, x_2, \dots, x_N\}$ una muestra aleatoria de una distribución Gaussiana $N(\mu, \sigma)$. Probar que la media de la muestra \bar{x} es un estimador de μ que es consistente, no sesgado y de varianza mínima.

En este caso sabemos que la distribución muestral es

$$P(\bar{\mu} | \mu, \sigma) = \frac{1}{\sqrt{\frac{\pi \sigma^2}{N}}} e^{-\frac{(\bar{\mu} - \mu)^2}{\frac{\sigma^2}{N}}} \Rightarrow N(\bar{\mu}, \frac{\sigma}{\sqrt{N}})$$

De aquí se deduce que $E(\bar{\mu}) = \mu$ y $V(\bar{\mu}) = \frac{\sigma^2}{N}$. Lo primero implica que $\bar{\mu}$ no tiene sesgo. De lo segundo se deduce que $\lim_{N \rightarrow \infty} V(\bar{\mu}) = 0$ así que $\bar{\mu}$ es un estimador consistente (si la varianza es 0 en el límite entonces $\bar{\mu} \rightarrow \mu$ si $N \rightarrow \infty$)

Para calcular la eficiencia del estimador calculamos, $V_{\text{min}}(\mu)$

$$V_{\text{min}}(\mu) = \frac{(1 + \frac{\partial}{\partial \mu} b(\mu))}{E \left[-\frac{\partial^2}{\partial \mu^2} \ln P(x|\mu, \sigma) \right]} = \frac{1}{E \left[-\frac{\partial^2}{\partial \mu^2} \ln P(x|\mu, \sigma) \right]}$$

$$P(x_1, \dots, x_N | \mu, \sigma) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^N \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right]$$

$$\ln P(x|\mu, \sigma) = -N \ln[\sigma \sqrt{2\pi}] - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2$$

$$\frac{\partial \ln P}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu), \quad \frac{\partial^2 \ln P}{\partial \mu^2} = -\frac{N}{\sigma^2} \Rightarrow E \left[-\frac{\partial^2 \ln P}{\partial \mu^2} \right] = \frac{N}{\sigma^2}$$

$$\Rightarrow V_{\text{min}}(\mu) = \frac{\sigma^2}{N} \Rightarrow e = \frac{\frac{\sigma^2}{N}}{V(\bar{\mu})} = 1$$

Para cualquier otro estimador $\hat{\mu}$ de μ no sesgado la desigualdad de Fisher nos dice que

$$V(\hat{\mu}) \geq \frac{\sigma^2}{N}$$

* Error estándar de los estimadores

Si $\hat{\alpha}$ es un estimador de un parámetro α con distribución normal $P(\hat{\alpha}|\alpha)$ entonces, el error estándar del estimador $\sigma_{\hat{\alpha}}$ viene dado por:

$$\sigma_{\hat{\alpha}} = V(\hat{\alpha})^{\frac{1}{2}}$$

• σ_a^2 se calcula a partir de $V(\hat{a})$ que a su vez se calcula a partir de la distribución muestral $P(\hat{a}|a)$. Esto quiere decir que $\sigma_a = \sigma_a(\hat{a}, a)$ y por tanto dependerá de un parámetro o parámetros a que en general serán desconocidos. Si \hat{a} es no sesgado entonces podemos interpretar σ_a como el "error estadístico" de a :

$$a = \hat{a} \pm \sigma_a$$

Ejemplo:

Se hace un muestreo de 10 valores de una distribución de población Gaussiana cuya desviación estándar es $\sigma = 1$. Los valores de la muestra son (con 2 decimales de exactitud)

$$x_1 = 2.22, x_2 = 2.56, x_3 = 1.07, x_4 = 0.24, x_5 = 0.18, x_6 = 0.95, x_7 = 0.73$$

$$x_8 = -0.74, x_9 = 2.09, x_{10} = 1.81$$

Estimar la media μ de la distribución de población y dar el error estándar del resultado.

En el ejemplo anterior se demostró que $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ es un estimador de μ que es consistente, sin sesgo, y de varianza mínima. Dicha varianza viene dada por

$$V(\bar{\mu}) = \frac{\sigma^2}{N} \Rightarrow \sigma_{\bar{\mu}} = \frac{\sigma}{\sqrt{N}}$$

$$\text{Por tanto } \mu = \bar{\mu} \pm \frac{\sigma}{\sqrt{N}} : \bar{x} \pm \frac{\sigma}{\sqrt{N}} : 1.11 \pm \frac{1}{\sqrt{10}} = 1.11 \pm 0.32$$

Si no hubiésemos conocido el valor verdadero de σ entonces habría que haber utilizado una estimación para $\sigma_{\bar{\mu}}$.

* Si trabajásemos con un estimador sesgado \hat{a} , entonces utilizaríamos la desviación cuadrática media E_a^2 como error del parámetro a :

$$\begin{aligned}
 E^2_{\hat{a}} &\equiv E[(\hat{a} - a)^2] = E[(\hat{a} - E(\hat{a}) + (E(\hat{a}) - a))^2] = \\
 &= E[(\hat{a} - E(\hat{a}))^2] + (E(\hat{a}) - a)^2 + 2E[(\hat{a} - E(\hat{a})) (E(\hat{a}) - a)] = \left[\begin{array}{l} \text{Recordar que} \\ b(a) \equiv E(\hat{a}) - a \end{array} \right] \\
 &= V(\hat{a}) + b(a)^2 + 2b(a)E[(\hat{a} - E(\hat{a}))] = \sigma_{\hat{a}}^2 + b(a)^2
 \end{aligned}$$

$\begin{array}{ccc} E(\hat{a}) - E(\hat{a}) & \uparrow & \uparrow \\ 0 & \text{error} & \text{error} \\ & \text{estadístico} & \text{experimental} \end{array}$

En este caso damos el resultado así $a = \hat{a} \pm E_{\hat{a}}$

Intervalos de confianza

* Un intervalo de confianza es otra manera de dar el error de uno de los parámetros de la distribución de población. Supongamos que tenemos una distribución de población $P(x|a)$ con sólo un parámetro a desconocido. Hacemos un muestreo de la población y definimos la distribución muestral de la manera usual

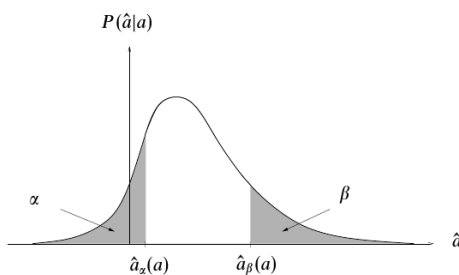
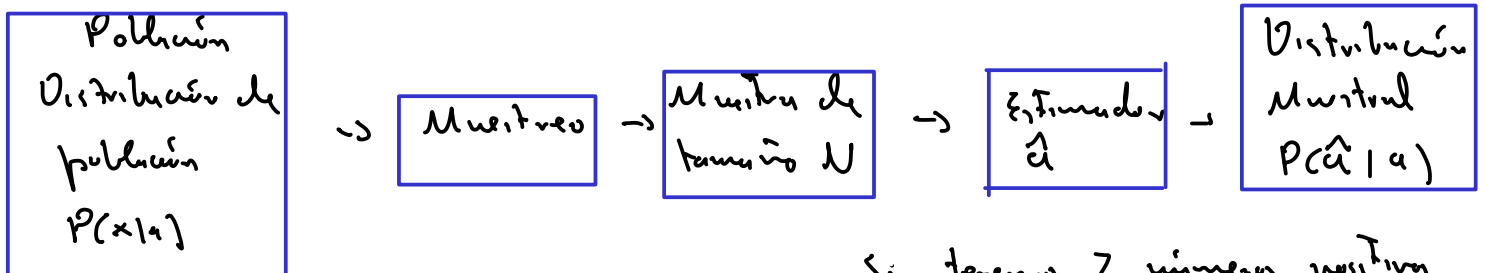


Figure 31.2 The sampling distribution $P(\hat{a}|a)$ of some estimator \hat{a} for a given value of a . The shaded regions indicate the two probabilities $\Pr(\hat{a} < \hat{a}_\alpha(a)) = \alpha$ and $\Pr(\hat{a} > \hat{a}_\beta(a)) = \beta$.

Si tenemos 2 números positivos

$\alpha, \beta < 1$, definimos las funciones

\hat{a}_α y \hat{a}_β así

$$a \mapsto \hat{a}_\alpha(a) : \int_{-\infty}^{\hat{a}_\alpha(a)} P(\hat{a}|a) d\hat{a} = \alpha$$

$$a \mapsto \hat{a}_\beta(a) = \int_{\hat{a}_\beta(a)}^{\infty} P(\hat{a}|a) d\hat{a} = \beta$$

$$\left. \begin{aligned} P_r(\hat{a} < \hat{a}_o(a)) &= \int_{-\infty}^{\hat{a}_o(a)} P(\hat{a}|a) d\hat{a} = \alpha \\ P_r(\hat{a}_\beta(a) < \hat{a}) &= \int_{\hat{a}_\beta(a)}^{\infty} P(\hat{a}|a) d\hat{a} = \beta \end{aligned} \right\} \begin{array}{l} \text{Suponemos ahora que las} \\ \text{funciones } \hat{a}_o \text{ y } \hat{a}_\beta \text{ son invertibles,} \\ \text{(lo cual sería el caso si } \hat{a} \text{ es un} \\ \text{buen estimador)} \end{array}$$

Se puede pensar que las inversas \hat{a}_o^{-1} y \hat{a}_β^{-1} son transformaciones de variables aleatorias

Si ahora tenemos una muestra y calculamos el valor correspondiente \hat{a}_{obs} a dicha muestra del estimador, entonces se tiene que

$$\left. \begin{aligned} P_r(\hat{a} < \hat{a}_{obs} = \hat{a}_o(a_+)) &= \alpha = \int_{-\infty}^{\hat{a}_o(a_+)} P(\hat{a}|a_+) d\hat{a} \\ P_r(\hat{a}_{obs} = \hat{a}_\beta(a_-) < \hat{a}) &= \beta = \int_{\hat{a}_\beta(a_-)}^{\infty} P(\hat{a}|a_-) d\hat{a} = \beta \end{aligned} \right\} \begin{array}{l} a_+ \text{ y } a_- \text{ son valores que} \\ \text{quedan determinados por las} \\ \text{condiciones} \end{array}$$

$$\hat{a}_{obs} = \hat{a}_o(a_+)$$

$$\hat{a}_{obs} = \hat{a}_\beta(a_-)$$

|| utilizar que \hat{a}_o^{-1} , \hat{a}_β^{-1} son transformaciones de variables aleatorias

$$\left. \begin{aligned} P_r(a < a_+ = \hat{a}_o^{-1}(\hat{a}_{obs})) &= \alpha \\ P_r(a_- = \hat{a}_\beta^{-1}(\hat{a}_{obs}) < a) &= \beta \end{aligned} \right\} \Rightarrow 1 - \alpha - \beta = P_r(a_- < a < a_+)$$

El intervalo $[a_-, a_+]$ se conoce como

límite superior de confianza $1-\alpha-\beta$

intervalo de confianza del parámetro a

límite inferior de confianza $1-\alpha-\beta$

con nivel de confianza $1-\alpha-\beta$.

Si $c = a_+ - \hat{a}_{obs}$, $d = \hat{a}_{obs} - a_-$ entonces tenemos que $a = \hat{a}_{obs} \pm c$

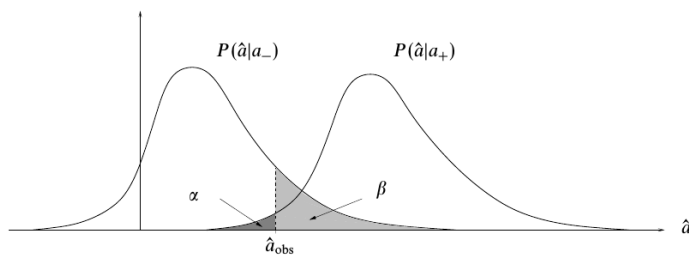


Figure 31.3 An illustration of how the observed value of the estimator, \hat{a}_{obs} , and the given values α and β determine the two confidence limits a_- and a_+ , which are such that $\hat{a}_\alpha(a_+) = \hat{a}_{obs} = \hat{a}_\beta(a_-)$.

• Ejemplo: el caso gaussiano. Supongamos que la distribución muestral de un estimador es una gaussiana

$$P(\hat{a} | a, \sigma_a^2) = \frac{1}{\sqrt{2\pi} \sigma_a^2} \exp\left[-\frac{(\hat{a} - a)^2}{2 \sigma_a^2}\right]$$

El teorema del límite central implica que para casi cualquier estimador consistente su distribución muestral se podrá aproximar bien por una gaussiana si el tamaño de la muestra N tiende a ∞ .

Para valores dados α y β definamos las funciones $\hat{a}_\alpha(a)$ y $\hat{a}_\beta(a)$. Si tenemos un valor \hat{a}_{obs} para una muestra entonces

$$a_+ \mapsto \hat{a}_\alpha(a_+) = \hat{a}_{obs} \quad \int_{-\infty}^{\hat{a}_{obs}} \frac{d\hat{a}}{\sqrt{2\pi} \sigma_a^2} \exp\left[-\frac{(\hat{a} - a_+)^2}{2 \sigma_a^2}\right] = \alpha$$

$$a_- \mapsto \hat{a}_\beta(a_-) = \hat{a}_{obs} \quad \int_{\hat{a}_{obs}}^{\infty} \frac{d\hat{a}}{\sqrt{2\pi} \sigma_a^2} \exp\left[-\frac{(\hat{a} - a_-)^2}{2 \sigma_a^2}\right] = \beta$$

Recordemos ahora la definición de la función error:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z du e^{-\frac{1}{2}u^2} = \left[\frac{u}{\sqrt{2}} = x \atop du = \sqrt{2} dx \right] = \frac{\sqrt{2}}{\sqrt{2\pi}} \int_{-\infty}^{\frac{z}{\sqrt{2}}} dx e^{-x^2} = \frac{1}{2} \operatorname{Erf}\left[\frac{z}{\sqrt{2}}\right]$$

$$\int_{-\infty}^{\hat{a}_{obs}} \frac{d\hat{a}}{\sqrt{2\pi}\sigma_{\hat{a}}^2} \exp\left[-\frac{(\hat{a}-a_+)^2}{2\sigma_{\hat{a}}^2}\right] = \left[\frac{\hat{a}-a_+}{\sigma_{\hat{a}}} = u \right] = \frac{\sigma_{\hat{a}}}{\sigma_{\hat{a}}} \int_{-\infty}^{\frac{\hat{a}_{obs}-a_+}{\sigma_{\hat{a}}}} \frac{du}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

$$= \Phi\left[\frac{\hat{a}_{obs}-a_+}{\sigma_{\hat{a}}}\right] \Rightarrow \Phi\left[\frac{\hat{a}_{obs}-a_+}{\sigma_{\hat{a}}}\right] = \alpha$$

$$\int_{\hat{a}_{obs}}^{\infty} \frac{d\hat{a}}{\sqrt{2\pi}\sigma_{\hat{a}}^2} \exp\left[-\frac{(\hat{a}-a_-)^2}{2\sigma_{\hat{a}}^2}\right] = 1 - \int_{-\infty}^{\hat{a}_{obs}} \frac{d\hat{a}}{\sqrt{2\pi}\sigma_{\hat{a}}^2} \exp\left[-\frac{(\hat{a}-a_-)^2}{2\sigma_{\hat{a}}^2}\right] :$$

$$= 1 - \Phi\left[\frac{\hat{a}_{obs}-a_-}{\sigma_{\hat{a}}}\right] = \beta \quad \text{elegamos a los } \Rightarrow \begin{cases} \Phi\left[\frac{\hat{a}_{obs}-a_+}{\sigma_{\hat{a}}}\right] = \alpha \\ \Phi\left[\frac{\hat{a}_{obs}-a_-}{\sigma_{\hat{a}}}\right] = 1-\beta \end{cases}$$

Usamos estas ecuaciones para despejar a_+ y a_- .

usar la propiedad $\Phi^{-1}[z] = -\Phi^{-1}[1-z]$

$$\left. \begin{aligned} a_+ &= \hat{a}_{obs} - \sigma_{\hat{a}} \Phi^{-1}[\alpha] = \hat{a}_{obs} + \sigma_{\hat{a}} \Phi^{-1}[1-\alpha] \\ a_- &= \hat{a}_{obs} - \sigma_{\hat{a}} \Phi^{-1}[1-\beta] = \hat{a}_{obs} - \sigma_{\hat{a}} \Phi^{-1}[1-\beta] \end{aligned} \right\}$$

El caso más común sucede cuando $\alpha = \beta$ (intervalo central de confianza). Podemos tratar de hallar el valor de α para que $\Phi^{-1}[1-\alpha] = 1 \Rightarrow 1-\alpha = \Phi(1) \Rightarrow \alpha = 1-\Phi(1)$,

$\Phi(1) = 0.8413 \Rightarrow \alpha = 0.1587$ Nivel de confianza = $1-\alpha-\beta = 1-2\alpha = 0.683 \Rightarrow$

Tenemos un nivel de confianza del 68% en \hat{a}_{obs}

Ejemplo:

Se tiene un número de 10 valores de una distribución de población Gaussiana cuya desviación estándar es $\sigma = 1$. Los valores de la muestra son (con 2 decimales de exactitud)

$$x_1 = 2.22, x_2 = 2.56, x_3 = 1.07, x_4 = 0.24, x_5 = 0.18, x_6 = 0.95, x_7 = 0.73$$

$$x_8 = -0.74, x_9 = 2.09, x_{10} = 1.81$$

Hallar el intervalo que representa un nivel de confianza del 90% alrededor de la media

En el ejemplo anterior se demostró que $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ es un estimador de μ que es consistente, sin sesgo, y de varianza mínima. Dicha varianza viene dada por

$$V(\bar{\mu}) = \frac{\sigma^2}{N} \Rightarrow \sigma_{\bar{\mu}} = \frac{\sigma}{\sqrt{N}}$$

$$\text{Por tanto } \mu = \bar{\mu} \pm \frac{\sigma}{\sqrt{N}} : \bar{x} \pm \frac{\sigma}{\sqrt{N}} = 1.11 \pm \frac{1}{\sqrt{10}} = 1.11 \pm 0.32$$

Nivel de confianza $1 - 2\alpha = 0.9 \Rightarrow \alpha = 0.05$ $1 - \alpha = 0.95$

$$a_+ = \hat{a}_{obs} + \sigma_{\hat{a}} \Phi^{-1}[1 - \alpha] = \bar{x} + \sigma_{\bar{\mu}} \Phi^{-1}[0.95] = 1.11 + 0.32 \times \overset{1.65}{\Phi^{-1}[0.95]} = 1.64$$

$$a_- = \hat{a}_{obs} - \sigma_{\hat{a}} \Phi^{-1}[1 - \alpha] = \bar{x} - \sigma_{\bar{\mu}} \Phi^{-1}[0.95] = 1.11 - 0.32 \times 1.65 = 0.58$$

El intervalo pedido es:

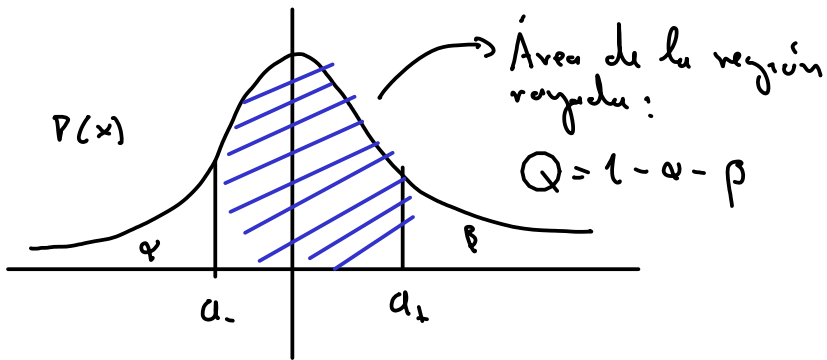
$$[a_-, a_+] = [0.58, 1.64]$$

Podemos pensar en un nivel de confianza como en una probabilidad de que el parámetro que estamos estimando esté en un cierto intervalo (el intervalo de confianza).

$P(x) \Rightarrow$ distribución de probabilidad:

$[a_-, a_+] =$ Intervalo de confianza
para la variable aleatoria x
con un nivel de confianza de $Q\%$

$$\Leftrightarrow \frac{Q}{100} = P_v(a_- \leq x \leq a_+)$$



a_- = nivel de confianza inferior
 a_+ = nivel de confianza superior.
 $\alpha = \beta \Rightarrow$ nivel central de confianza para un intervalo central de confianza

Por ejemplo si $P(x)$ es una normal $\Rightarrow P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$

$x \sim N(\mu, \sigma) \Rightarrow$ se tiene la fórmula:

$$P_v(\mu - n\sigma \leq x \leq \mu + n\sigma) = 2\Phi(n) - 1$$

Así que podemos utilizar σ para definir el intervalo central de confianza $[\mu - n\sigma, \mu + n\sigma]$. Se habla entonces de una confianza de " n sigmas".

Para un nivel de confianza del 95% tenemos que $2\Phi(n) - 1 = 0.95 \Rightarrow n = 1.9$