Inferencia Estadística

Tema 2:

28 EJERCICIOS RESUELTOS

Distribuciones en el muestreo

2.1 MUESTREO ALEATORIO SIMPLE

Siendo "X" una variable aleatoria unidimensional (discreta o continua), se llama **población** de "X" al conjunto de individuos (ya sean personas, animales, plantaciones de patatas, etc.) en los que está definida la variable aleatoria "X". Se llama **muestra de tamaño "n"** a todo conjunto de "n" observaciones de la variable "X", y se llama **muestreo** al procedimiento mediante el que obtenemos una muestra de la población. Asociada al muestreo de tamaño "n" de la variable "X" hay una variable aleatoria n-dimensional $(X_1; \ldots ; X_n)$ que expresa los resultados que pueden presentarse al seleccionar una muestra de tamaño "n"; de ella diremos que es la **variable muestral.** De las variables unidimensionales $X_1, \ldots, X_n$ diremos que son las variables **componentes muestrales.**

Por ejemplo, si "X" expresa el aleatorio peso de los ciudadanos y tomamos una muestra tamaño 3, la variable muestral $(X_1; X_2; X_3)$ es tridimensional, expresando X_i el aleatorio peso del i-ésimo ciudadano seleccionado $(i = 1,2,3)$. Así, si la muestra seleccionada es la $(60;76;90)$, resulta obvio que las variables X_1, X_2 y X_3 se han concretado respectivamente en los números 60, 76 y 90.

Como la selección de las muestras tiene como fin el estudio de las propiedades de la población, conviene que la muestra seleccionada sea representativa de ésta. El **muestreo aleatorio simple** se apoya en los siguientes supuestos:

1) La muestra se obtiene al azar, de modo que todos los individuos de la población tienen la misma probabilidad de ser seleccionados, y las sucesivas elecciones de los "n" individuos que forman la muestra son independientes.
2) La distribución de probabilidad "X" no se altera debido al muestreo.
3) Las variables componentes muestrales $X_1, \ldots, X_n$ son independientes y tienen la misma distribución de probabilidad que la variable "X". Por tanto, la función de cuantía/densidad de la variable muestral es el producto de las respectivas funciones de cuantía/densidad de las variables $X_1, \ldots, X_n$.

2.2 ESTADÍSTICO

Llamaremos **estadístico** a cualquier función real de la variable muestral $(X_1; \ldots ; X_n)$ que a su vez sea una variable aleatoria.

Nada de asustarse: un estadístico "T" es una variable aleatoria unidimensional corriente y moliente cuyo valor depende de los valores que tomen las variables $X_1, \ldots, X_n$; por tanto, para el estadístico "T" nos plantearemos las historias de siempre: ¿cuáles son las funciones de distribución, generatriz de momentos y de cuantía/densidad de "T"?, ¿cuánto valen E(T) y V(T)?, ¿cuál es la probabilidad del suceso $T = t$?, ¿cuál es la probabilidad del suceso $a < T \leq b$?

2.3 ESTADÍSTICOS FAMOSOS

Momentos muestrales respecto al origen

El estadístico llamado **momento muestral de orden "k" respecto al origen** se denota a_k, siendo:

$$a_k = \frac{1}{n}.\sum_{i=1}^{n} X_i^k,\ k = 0,1,2,.....$$

El más famoso es el de primer orden, que llamaremos **media muestral** y denotaremos $\overline{X}$:

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

Momentos muestrales centrales

El estadístico llamado **momento muestral central de orden "k"** se denota m_k, siendo:

$$m_k = \frac{1}{n}.\sum_{i=1}^{n} (X_i - \overline{X})^k,\ k = 1,2,.....$$

El más famoso es el de segundo orden, que llamaremos **varianza muestral** y denotaremos S^2:

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

Observa: el momento central de primer orden es el estadístico nulo:

$$m_1 = \frac{1}{n}.\sum_{i=1}^{n} (X_i - \overline{X})^k \underset{\boxed{\text{si } k=1}}{=} \frac{1}{n}.\sum_{i=1}^{n} (X_i - \overline{X}) =$$

$$= \left(\frac{1}{n}.\sum_{i=1}^{n} X_i\right) - \left(\frac{1}{n}.n.\overline{X}\right) = \overline{X} - \overline{X} = 0$$

El estadístico **cuasivarianza muestral** se denota S_c^2, siendo:

$$S_c^2 = \frac{1}{n-1}.\sum_{i=1}^{n} (X_i - \overline{X})^2$$

Obviamente, es $S_c^2 = \frac{n}{n-1}.S^2$ y $S^2 = \frac{n-1}{n}.S_c^2$.

Recuerda que para toda variable "X" es $V(X) = E(X^2) - (E(X))^2$. Pues bien, esa misma relación hay entre la varianza muestral (S^2), el momento muestral de orden 2 respecto al origen (a_2) y la media muestral $(\overline{X})$; o sea: $S^2 = a_2 - \overline{X}^2$:

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

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

$$\boxed{\sum_{i=1}^{n}X_i=n.\overline{X}}$$

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

$$=\left(\frac{1}{n}.\sum_{i=1}^{n}X_i^2\right)-\overline{X}^2=a_2-\overline{X}^2$$

Caso bidimensional

Si de una variable aleatoria bidimensional $(X;Y)$ se toma muestra de tamaño "n" (imagina que "X" expresa la aleatoria altura de los ciudadanos e "Y" su aleatorio peso), la variable muestral será $\big((X_1;Y_1);(X_2;Y_2);\ldots;(X_n;Y_n)\big)$; así:

- Las medias muestrales son:

$$\overline{X}=\frac{1}{n}.\sum_{i=1}^{n}X_i\;;\;\overline{Y}=\frac{1}{n}.\sum_{i=1}^{n}Y_i$$

- Los momentos muestrales de segundo orden son:

$$a_{2;0}=\frac{1}{n}.\sum_{i=1}^{n}X_i^2\;;\;a_{0;2}=\frac{1}{n}.\sum_{i=1}^{n}Y_i^2$$

$$S_1^2=\frac{1}{n}.\sum_{i=1}^{n}(X_i-\overline{X})^2\;;\;S_2^2=\frac{1}{n}.\sum_{i=1}^{n}(Y_i-\overline{Y})^2$$

- Los momentos muestrales orden $p;q$ son:

$$a_{p;q}=\frac{1}{n}.\sum_{i=1}^{n}X_i^p.Y_i^q\;;\;m_{p;q}=\frac{1}{n}.\sum_{i=1}^{n}(X_i-\overline{X})^p.(Y_i-\overline{Y})^q$$

- La **covarianza muestral,** que se denota S_{11}, es:

$$S_{11}=m_{1;1}=\frac{1}{n}.\sum_{i=1}^{n}(X_i-\overline{X}).(Y_i-\overline{Y})=\frac{1}{n}.\left(\sum_{i=1}^{n}X_i.Y_i\right)-\overline{X}.\overline{Y}$$

- El **coeficiente de correlación muestral,** que se denota "r", es

$$r=\frac{S_{11}}{S_1.S_2}$$

FONEMATO 2.3.1

Se toma m.a.s de tamaño 2 de una población "X" cuya función de cuantía es

x	0	1	2	3
P(X = x)	1/8	3/8	3/8	1/8

Determínese la distribución de probabilidad de los siguientes estadísticos:

$$1)\ T_1 = \overline{X}\ ;\ 2)\ T_2 = X_1 + X_2\ ;\ 3)\ T_3 = S^2\ ;\ 4)\ T_4 = S_c^2$$

$$5)\ T_5 = X_1 - X_2\ ;\ 6)\ T_2 = 2.X_1 - X_2\ ;\ T_7 = \text{máx.}(X_1;X_2)$$

SOLUCIÓN

Las variables muestrales X_1 y X_2 son independientes y tienen igual distribución de probabilidad que la población "X":

x_1	0	1	2	3
$P(X_1 = x_1)$	1/8	3/8	3/8	1/8

;

x_2	0	1	2	3
$P(X_2 = x_2)$	1/8	3/8	3/8	1/8

La función de cuantía de la variable bidimensional $(X_1;X_2)$ es:

$$P(X_1 = x_1;X_2 = x_2) \underset{\uparrow}{=} P(X_1 = x_1).P(X_2 = x_2) \Rightarrow$$

$\boxed{\text{pues } X_1 \text{ y } X_2 \text{ son independientes}}$

$\Rightarrow$

X_1 / X_2	0	1	2	3
0	1/64	3/64	3/64	1/64
1	3/64	9/64	9/64	3/64
2	3/64	9/64	9/64	3/64
3	1/64	3/64	3/64	1/64

La siguiente tabla recoge toda la información necesaria para determinar la distribución de probabilidad de cada estadístico:

$(x_1;x_2)$	$P(X_1 = x_1;X_2 = x_2)$	T_1	T_2	T_3	T_4	T_5	T_6	T_7
(0;0)	1/64	0	0	0	0	0	0	0
(0;1)	3/64	0'5	1	0'25	0'5	−1	−1	1
(0;2)	3/64	1	2	1	2	−2	−2	2
(0;3)	1/64	1'5	3	2'25	4'5	−3	−3	3
(1;0)	3/64	0'5	1	0'25	0'5	1	2	1
(1;1)	9/64	1	2	0	0	0	1	1
(1;2)	9/64	1'5	3	0'25	0'5	−1	0	2
(1;3)	3/64	2	4	1	2	2	−1	3
(2;0)	3/64	1	2	1	2	2	4	2
(2;1)	9/64	1'5	3	0'25	0'5	1	3	2
(2;2)	9/64	2	4	0	0	0	2	2
(2;3)	3/64	2'5	5	0'25	0'5	−1	1	3
(3;0)	1/64	1'5	3	2'25	4'5	3	6	3
(3;1)	3/64	2	4	1	2	2	5	3
(3;2)	3/64	2'5	5	0'25	0'5	1	4	3
(3;3)	1/64	3	6	0	0	0	3	3

- Veamos cómo se ha calculado el valor que toma cada estadístico en una muestra concreta; por ejemplo, en la muestra $(2;3)$:

1) $$T_1 = \overline{X} = (X_1 + X_2)/2 \Rightarrow T_1(2;3) = (2+3)/2 = 2'5$$

2) $$T_2 = X_1 + X_2 \Rightarrow T_2(2;3) = 2+3 = 5$$

3) $$T_3 = S^2 = \frac{1}{2}.\sum_{i=1}^{2}(X_i - \overline{X})^2 = \frac{1}{2}.\left((X_1-\overline{X})^2 + (X_2-\overline{X})^2\right) \Rightarrow$$
$$\Rightarrow T_3(2;3) = \frac{1}{2}.\left((2-2'5)^2 + (3-2'5)^2\right) = 0'25$$

4) $$T_4 = S_c^2 = \frac{1}{2-1}.\sum_{i=1}^{2}(X_i - \overline{X})^2 = (X_1-\overline{X})^2 + (X_2-\overline{X})^2 \Rightarrow$$
$$\Rightarrow T_4(2;3) = (2-2'5)^2 + (3-2'5)^2 = 0'5$$

5) $$T_5 = X_1 - X_2 \Rightarrow T_5(2;3) = 2-3 = -1$$

6) $$T_6 = 2.X_1 - X_2 \Rightarrow T_6(2;3) = 2.2-3 = 1$$

7) $$T_7 = \text{máx.}(X_1;X_2) \Rightarrow T_7(2;3) = \text{máx.}(2;3) = 3$$

Lo mismo que con el punto $(2;3)$ se hace con los restantes 15 puntos que forman la distribución de masa de probabilidad asociada a la variable muestral $(X_1;X_2)$ que expresa los resultados que pueden presentarse al tomar m.a.s. de tamaño 2 de la población "X"

1) La distribución de probabilidad del estadístico $T_1 = \overline{X}$ es:

t	0	0'5	1	1'5	2	2'5	3
$P(T_1 = t)$	1/64	6/64	15/64	20/64	15/64	6/64	1/64

Veamos cómo hemos obtenido que $P(T_1 = 2) = 15/64$: la probabilidad del suceso $T_1 = 2$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras $(1;3)$, $(2;2)$ y $(3;1)$ en que T_1 toma el valor 2:

$$P(T_1 = 2) = P(X_1 = 1; X_2 = 3) + P(X_1 = 2; X_2 = 2) + P(X_1 = 3; X_2 = 1) =$$
$$= \frac{3}{64} + \frac{9}{64} + \frac{3}{64} = \frac{15}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_1.

Es:

$$E(T_1) = E\left(\frac{1}{2}.(X_1 + X_2)\right) = \frac{1}{2}.\left(E(X_1) + E(X_2)\right) = \frac{1}{2}.(1'5 + 1'5) = 1'5$$

$$\boxed{E(X_i) = E(X) = 0.\frac{1}{8} + 1.\frac{3}{8} + 2.\frac{3}{8} + 3.\frac{1}{8} = 1'5,\ \forall\ i = 1,2}$$

$$V(T_1) = V\left(\frac{1}{2}.(X_1 + X_2)\right) = \frac{1}{4}.\left(V(X_1) + V(X_2)\right) = \frac{1}{4}.(0'75 + 0'75) = \frac{3}{8}$$

$$\boxed{V(X_i) = V(X) = E(X^2) - (E(X))^2 = 3 - 1'5^2 = 0'75,\ \forall\ i = 1,2}$$

$$\boxed{E(X^2) = 0^2.\frac{1}{8} + 1^2.\frac{3}{8} + 2^2.\frac{3}{8} + 3^2.\frac{1}{8} = 3}$$

2) La distribución de probabilidad del estadístico $T_2 = X_1 + X_2$ es:

t	0	1	2	3	4	5	6
$P(T_2 = t)$	1/64	6/64	15/64	20/64	15/64	6/64	1/64

Veamos cómo hemos obtenido que $P(T_2 = 5) = 6/64$: la probabilidad del suceso $T_2 = 5$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras $(2;3)$ y $(3;2)$ en que T_2 toma el valor 5:

$$P(T_2 = 5) = P(X_1 = 2; X_2 = 3) + P(X_1 = 3; X_2 = 2) = \frac{3}{64} + \frac{3}{64} = \frac{6}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_2.

Es:

$$E(T_2) = E(2.T_1) = 2.E(T_1) = 2.1'5 = 3$$

$$\boxed{T_2 = 2.T_1}$$

$$V(T_2) = V(2.T_1) = 4.V(T_1) = 4.\frac{3}{8} = 1'5$$

3) La distribución de probabilidad del estadístico $T_3 = S^2$ es:

t	0	0'25	1	2'25
$P(T_3 = t)$	20/64	30/64	12/64	2/64

Veamos cómo hemos obtenido que $P(T_3 = 1) = 12/64$: la probabilidad del suceso $T_3 = 1$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras $(0;2)$, $(1;3)$, $(2;0)$ y $(3;1)$ en que T_3 toma el valor 1:

$$P(T_3 = 1) = P(X_1 = 0; X_2 = 2) + P(X_1 = 1; X_2 = 3) +$$
$$+ P(X_1 = 2; X_2 = 0) + P(X_1 = 3; X_2 = 1) = \frac{3}{64} + \frac{3}{64} + \frac{3}{64} + \frac{3}{64} = \frac{12}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_3.

Es:

$$E(T_3) = 0.\frac{20}{64} + 0'25.\frac{30}{64} + 1.\frac{12}{64} + 2'25.\frac{2}{64} = \frac{24}{64}$$

$$V(T_3) = E(T_3^2) - (E(T_3))^2 = \frac{24}{64} - (\frac{24}{64})^2 = \frac{15}{64}$$

$$\boxed{E(T_3^2) = 0^2.\frac{20}{64} + 0'25^2.\frac{30}{64} + 1^2.\frac{12}{64} + 2'25^2.\frac{2}{64} = \frac{24}{64}}$$

4) La distribución de probabilidad del estadístico $T_4 = S_c^2$ es:

t	0	0'5	2	4'5
$P(T_4 = t)$	20/64	30/64	12/64	2/64

Veamos cómo hemos obtenido que $P(T_4 = 2) = 12/64$: la probabilidad del suceso $T_4 = 2$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras $(0;2), (1;3), (2;0)$ y $(3;1)$ en que T_4 toma el valor 2:

$$P(T_4 = 2) = P(X_1 = 0; X_2 = 2) + P(X_1 = 1; X_2 = 3) +$$
$$+P(X_1 = 2; X_2 = 0) + P(X_1 = 3; X_2 = 1) = \frac{3}{64} + \frac{3}{64} + \frac{3}{64} + \frac{3}{64} = \frac{12}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_4.

Para una muestra aleatoria simple de tamaño "n" es $S_c^2 = n.S^2/(n-1)$; así, si $n = 2$, es $T_4 = S_c^2 = 2.S^2 = 2.T_3$. En consecuencia:

$$E(T_4) = E(2.T_3) = 2.E(T_3) = 2.(24/64) = 48/64$$
$$V(T_4) = V(2.T_3) = 4.V(T_3) = 4.(15/64) = 60/64$$

5) La distribución de probabilidad del estadístico $T_5 = X_1 - X_2$ es:

t	-3	-2	-1	0	1	2	3
$P(T_5 = t)$	1/64	6/64	15/64	20/64	15/64	6/64	1/64

Veamos cómo hemos obtenido que $P(T_5 = -2) = 6/64$: la probabilidad del suceso $T_5 = -2$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras $(0;2)$ y $(1;3)$ en que T_5 toma el valor -2:

$$P(T_5 = -2) = P(X_1 = 0; X_2 = 2) + P(X_1 = 1; X_2 = 3) = \frac{3}{64} + \frac{3}{64} = \frac{6}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_5.

Es:

$$E(T_5) = E(X_1 - X_2) \underset{\uparrow}{=} E(X_1) - E(X_2) = 0$$

$\boxed{\text{pues } E(X_1) = E(X_2)}$

$$V(T_5) = V(X_1 - X_2) \underset{\uparrow}{=} V(X_1) + V(X_2) = 1'5$$

$\boxed{V(X_i) = V(X) = 0'75,\ \forall\ i = 1,2}$

6) La distribución de probabilidad del estadístico $T_6 = 2.X_1 - X_2$ es:

t	-3	-2	-1	0	1	2	3	4	5	6
$P(T_6 = t)$	1/64	3/64	6/64	10/64	12/64	12/64	10/64	6/64	3/64	1/64

Veamos cómo hemos obtenido que $P(T_6 = -1) = 6/64$: la probabilidad del suceso $T_6 = -1$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras $(0;1)$ y $(1;3)$ en que T_6 toma el valor -1:

$$P(T_6 = -2) = P(X_1 = 0; X_2 = 1) + P(X_1 = 1; X_2 = 3) = \frac{3}{64} + \frac{3}{64} = \frac{6}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_6.

Es:

$$E(T_6)=E(2.X_1-X_2)=2.E(X_1)-E(X_2)=2.1'5-1'5=1'5$$

$$\boxed{E(X_1)=E(X_2)=1'5}$$

$$V(T_6)=V(2.X_1-X_2)=4.V(X_1)+V(X_2)=3'75$$

$$\boxed{V(X_i)=V(X)=0'75,\ \forall\ i=1,2}$$

7) La distribución de probabilidad del estadístico $T_7=\text{máx.}(X_1;X_2)$ es:

t	0	1	2	3
$P(T_7=t)$	1/64	15/64	33/64	15/64

Veamos cómo hemos obtenido que $P(T_7=1)=15/64$: la probabilidad del suceso $T_7=1$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (0;1), (1;0) y (1;1) en que T_7 toma el valor 1:

$$P(T_7=1)=P\big(X_1=0;X_2=1\big)+P\big(X_1=1;X_2=0\big)+P\big(X_1=1;X_2=1\big)=$$
$$=\frac{3}{64}+\frac{3}{64}+\frac{9}{64}=\frac{15}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_7.

Es:

$$E(T_7)=0.\frac{1}{64}+1.\frac{15}{64}+2.\frac{33}{64}+3.\frac{15}{64}=\frac{126}{64}$$

$$V(T_7)=E(T_7^2)-(E(T_7))^2=\frac{282}{64}-(\frac{126}{64})^2$$

$$\boxed{E(T_7^2)=0^2.\frac{1}{64}+1^2.\frac{15}{64}+2^2.\frac{33}{64}+3^2.\frac{15}{64}=\frac{282}{64}}$$

¡Que quede claro!

Siendo discreta la población "X", con muestra de tamaño 2, para calcular la función de cuantía del estadístico "Pepe" calculamos el valor que toma "Pepe" en cada punto de la distribución bidimensional discreta asociada a la variable muestral (X_1;X_2); y después obramos en consecuencia.

FONEMATO 2.3.2

Sea "X" una población continua de la que se toma m.a.s de tamaño "n".

1) Determínese la distribución del mínimo valor muestral "Z".
2) Determínese la distribución del máximo valor muestral "Y".
3) Determínese la distribución de recorrido muestral $R = Y - Z$.
4) Aplíquese al caso $n = 3$ si la densidad de "X" es $f(x) = 2.x,\ x \in [0;1]$.
5) En las condiciones de 4), determínense las siguientes probabilidades

$$P(Y \leq 0'1)\ ;\ P(Y > 0'9)\ ;\ P(Z \leq 0'1)\ ;\ P(Z > 0'9)\ ;\ P(R \leq 0'5)$$

SOLUCIÓN

Siendo "f" y "F" las respectivas funciones de densidad y de distribución de "X", sea $(X_1; \ldots ; X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño "n"; por tanto, $X_1, \ldots, X_n$ son independientes y tienen igual distribución de probabilidad que la población "X".

1) Siendo $Z = \text{mín.}(X_1; \ldots ; X_n)$, su función de distribución F_Z es:

$$F_Z(z) = P(Z \leq z) = P\big(\text{mín.}(X_1; \ldots ; X_n) \leq z\big) =$$

$$= 1 - P\big(\text{mín.}(X_1; \ldots ; X_n) > z\big) = 1 - P\big(X_1 > z; \ldots ; X_n > z\big) =$$

$\boxed{\text{pues } X_1, \ldots, X_n \text{ son independientes}}$

$$= 1 - \prod_{i=1}^{n} P(X_i > z) = 1 - \big(1 - F(z)\big)^n$$

$\boxed{P(X_i > z) = P(X > z) = 1 - F(z),\ \forall\ i = 1, 2, \ldots, n}$

- Siendo $f_Z(z)$ la densidad de probabilidad de "Z", es:

$$f_Z(z) = \frac{dF_Z(z)}{dz} = n.\big(1 - F(z)\big)^{n-1}.\frac{dF(z)}{dz} = n.\big(1 - F(z)\big)^{n-1}.f(z)$$

$\boxed{F_Z(z) = 1 - \big(1 - F(z)\big)^n}$ $\boxed{dF(z)/dz = f(z)}$

- También podemos calcular la densidad $f_Z(z)$ de "Z" así:

$$P(z < Z < z + dz) = f_Z(z).dz = \underbrace{n}_{a}.\underbrace{f(z).dz}_{b}.\underbrace{\left(\int_{z+dz}^{+\infty} f(x).dx\right)^{n-1}}_{c} \Rightarrow$$

a	El "papel" de "mínimo valor muestral" puede hacerlo cualquiera de las "n" variables componentes muestrales.
b	Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(z; z + dz)$.
c	Probabilidad de las restantes $n - 1$ variables componentes muestrales tomen valores mayores que $z + dz$.

$$\Rightarrow f_Z(z) = n.f(z).\left(\int_z^{+\infty} f(x).dx\right)^{n-1} = n.(1 - F(z))^{n-1}.f(z)$$

$$\boxed{\int_z^{+\infty} f(x).dx = 1 - F(z)}$$

dividimos los dos miembros por "dz", y consideramos que, por ser "dz" un infinitésimo, es $\int_{z+dz}^{+\infty} f(x).dx = \int_z^{+\infty} f(x).dx$

2) Siendo $Y = \text{máx.}(X_1; \ldots ; X_n)$, su función de distribución F_Y es:

$$F_Y(y) = P(Y \le y) = P(\text{máx.}(X_1; \ldots ; X_n) \le y) =$$

pues $X_1, \ldots, X_n$ son indpendientes

$$= P(X_1 \le y; \ldots ; X_n \le y) = \prod_{i=1}^{n} P(X_i \le y) = (F(y))^n$$

$$\boxed{P(X_i \le y) = P(X \le y) = F(y),\ \forall\, i = 1, 2, \ldots, n}$$

- Siendo $f_Y(y)$ la densidad de probabilidad de "Y", es:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = n.(F(y))^{n-1}.\frac{dF(y)}{dy} = n.(F(y))^{n-1}.f(y)$$

$$\boxed{F_Y(y) = (F(y))^n} \quad \boxed{dF(y)/dy = f(y)}$$

- También podemos calcular la densidad $f_Y(y)$ de "Y" así:

$$P(y < Y < y + dy) = f_Y(y).dy = \underbrace{n}_{a}.\underbrace{f(y).dy}_{b}.\underbrace{\left(\int_{-\infty}^{y} f(x).dx\right)^{n-1}}_{c} \Rightarrow$$

a El "papel" de "máximo valor muestral" puede hacerlo cualquiera de las "n" variables componentes muestrales.

b Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(y; y + dy)$.

c Probabilidad de las restantes $n - 1$ variables componentes muestrales tomen valores menores que "y".

$$\Rightarrow f_Y(y) = n.f(y).\left(\int_{-\infty}^{y} f(x).dx\right)^{n-1} = n.(F(y))^{n-1}.f(y)$$

es $\int_{-\infty}^{y} f(x).dx = F(y)$

Dividimos los dos miembros por "dy"

3) Para determinar la distribución de probabilidad de $R = Y - Z$ debemos calcular previamente la densidad $g(z;y)$ de la variable bidimensional $(Z;Y)$; es:

$$P(z < Z < z + dz; y < Y < y + dy) = g(z;y).dz.dy =$$

$$= \underbrace{n.(n-1)}_{\boxed{a}}.\underbrace{f(z).dz}_{\boxed{b}}.\underbrace{f(y).dy}_{\boxed{c}}.\underbrace{\left(\int_z^y f(x).dx\right)^{n-2}}_{\boxed{d}} \Rightarrow$$

$\boxed{a}$ El "papel" de "mínimo valor muestral" puede hacerlo cualquiera de las "n" variables componentes muestrales, y el "papel" de "máximo valor muestral" puede hacerlo cualquiera de las restantes $n-1$ variables componentes muestrales,; por tanto, el número de pares distintos $(\text{mínimo};\text{máximo})$ en que pueden concretarse las "n" observaciones muestrales es $n.(n-1)$.

$\boxed{b}$ Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(z;z+dz)$.

$\boxed{c}$ Probabilidad de que otra de las variables componentes muestrales tome un valor en el intervalo $(y;y+dy)$.

$\boxed{d}$ Probabilidad de las restantes $n-2$ variables componentes muestrales tomen valores en el intervalo $(z+dz;y+dy)$. Como "dz" y "dy" son infinitésimos, despreciamos su influencia y consideramos el intervalo $(z;y)$.

$$\Rightarrow g(z;y) = n.(n-1).f(z).f(y).\left(\int_z^y f(x).dx\right)^{n-2}$$

dividimos los dos miembros por "dz.dy"

- Haciendo el cambio de variables $\begin{Bmatrix} S = Z \\ R = Y - Z \end{Bmatrix} \Rightarrow \begin{Bmatrix} Z = S \\ Y = S + R \end{Bmatrix}$, el jacobiano es:

$$J = \begin{vmatrix} \partial z/\partial s & \partial z/\partial r \\ \partial y/\partial s & \partial y/\partial r \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

La densidad de probabilidad $h(s;r)$ de la variable $(S;R)$ es:

$$h(s;r) = g(s;s+r).|J| = n.(n-1).f(s).f(s+r).\left(\int_s^{s+r} f(x).dx\right)^{n-2}$$

La densidad de probabilidad $f_R(r)$ del recorrido muestral "R" es:

$$f_R(r) = \int_{s=-\infty}^{s=+\infty} h(s;r).ds$$

4) Si la densidad de "X" es $f(x) = 2.x$, $x \in [0;1]$, entonces:

$$F(x) = P(X \le x) = \int_0^x 2.x.dx = x^2$$

- Para $n=3$, siendo $Z=\text{mín.}(X_1;X_2;X_3)$ e $Y=\text{máx.}(X_1;X_2;X_3)$, sus respectivas funciones de densidad de probabilidad son:

$$f_Z(z)=3.\left(1-F(z)\right)^{3-1}.f(z)=6.z.(1-z^2),\ z\in[0:1]$$

$$f_Y(y)=3.\left(F(y)\right)^{3-1}.f(y)=6.y^2,\ y\in[0:1]$$

- La densidad de probabilidad $g(z;y)$ de la variable $(Z;Y)$ es:

$$g(z;y)=3.(3-1).f(z).f(y).\left(\int_z^y f(x).dx\right)^{3-2}=$$

$$=6.(2.z).(2.y).\int_z^y 2.x.dx=24.z.y.(y^2-z^2),\ \text{siendo}\begin{cases}0\le z\le 1\\ z\le y\le 1\end{cases}$$

- Haciendo el cambio de variables $\begin{Bmatrix}S=Z\\ R=Y-Z\end{Bmatrix}\Rightarrow\begin{Bmatrix}Z=S\\ Y=S+R\end{Bmatrix}$, cuyo jacobiano es 1, la densidad de probabilidad $h(s;r)$ de la variable $(S;R)$ es:

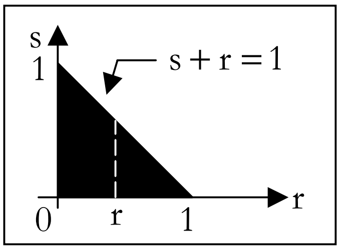
$$h(s,r)=g(s;s+r).|J|=24.s.(s+r).\left((s+r)^2-s^2\right)=$$

$\boxed{\text{en } g(z;y) \text{ sustituimos "z" por "s" e "y" por "s + r"}}$

$$=24.s.r.(s+r).(2.s+r),\ \text{siendo}\underbrace{\begin{cases}0\le s\le 1\\ 0\le r\le 1-s\end{cases}}$$

$$\boxed{\text{si}\begin{Bmatrix}0\le z\le 1\\ z\le y\le 1\end{Bmatrix}\text{ y }\begin{Bmatrix}z=s\\ y=s+r\end{Bmatrix}\Rightarrow\begin{Bmatrix}0\le s\le 1\\ s\le s+r\le 1\end{Bmatrix}\Rightarrow\begin{Bmatrix}0\le s\le 1\\ 0\le r\le 1-s\end{Bmatrix}}$$

- Densidad de probabilidad $f_R(r)$ del recorrido "R":



$$f_R(r)=\int_{s=-\infty}^{s=+\infty}h(s;r).ds=$$

$$=\int_{s=0}^{s=1-r}24.s.r.(s+r).(2.s+r).ds=\ldots.=$$

$$=12.r.(1-r^2),\ r\in[0;1]$$

5) Siendo $F_Z(z)=1-\left(1-F(z)\right)^n$ y $F_Y(y)=(F(y))^n$, si $F(x)=x^2$ y $n=3$, es:

$$F_Z(z)=1-(1-z^2)^3\ ;\ F_Y(y)=y^6$$

Por tanto:

$*\ P(Y\le 0'1)=F_Y(0'1)=0'1^6$

$*\ P(Y>0'9)=1-P(Y\le 0'9)=1-F_Y(0'9)=1-0'9^6$

$*\ P(Z\le 0'1)=F_Z(0'1)=1-(1-0'1^2)^3$

$*\ P(Z>0'9)=1-P(Z\le 0'9)=1-F_Z(0'9)=1-\left(1-(1-0'9^2)^3\right)$

$*\ P(R\le 0'5)=\int_0^{0'5}f_R(r).dr=\int_0^{0'5}12.r.(1-r^2).dr=\ldots.$

FONEMATO 2.3.3

De una población $X \approx N(0;\sigma)$ se toma m.a.s de tamaño 2. Demuéstrese que los estadísticos "Z" y "T" son incorrelados, siendo:

$$Z=\frac{1}{2}.(X_1+X_2)\ ;\ T=X_1^2+X_2^2$$

SOLUCIÓN

Sea $(X_1;X_2)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño 2; por tanto, X_1 y X_2 son independientes y tienen distribución $N(0;\sigma)$, como la población "X".

Para demostrar que los estadísticos "Z" y "T" son incorrelados debemos demostrar que su covarianza es nula.

$$\text{COV.}(Z;T)=E(Z.T)-E(Z).E(T)=0$$

$$* \ E(Z)=E\left(\frac{1}{2}.(X_1+X_2)\right)=\frac{1}{2}.(E(X_1)+E(X_2))=\frac{1}{2}.(0+0)=0$$

$$* \ E(Z.T)=E\left(\frac{1}{2}.(X_1+X_2).(X_1^2+X_2^2)\right)=$$

$$=\frac{1}{2}.E\left(X_1^3+X_1.X_2^2+X_1^2.X_2+X_2^3\right)=$$

$$=\frac{1}{2}.\left(E(X_1^3)+E(X_1.X_2^2)+E(X_1^2.X_2)+E(X_2^3)\right)=$$

pues las variables aleatorias X_1 y X_2 son independientes

$$=\frac{1}{2}.\left(E(X_1^3)+E(X_1).E(X_2^2)+E(X_1^2).E(X_2)+E(X_2^3)\right)=0$$

como en la distribución $N(0;\sigma)$ son nulos todos los momentos de orden impar respecto al origen, es $E(X_1^3)=E(X_1)=E(X_2)=E(X_2^3)=0$

FONEMATO 2.3.4

1) En el muestreo aleatorio simple de tamaño 5 de una población exponencial de parámetro "a", determínese la distribución del recorrido muestral.
2) Calcúlese la probabilidad de que el recorrido muestral sea mayor que 1.

SOLUCIÓN

1) Sea $X \approx \text{Exp.}(a)$ y $(X_1;\ldots;X_5)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s. de tamaño 5; por tanto, $X_1,\ldots,X_5$ son independientes y tienen distribución $\text{Exp.}(a)$, como la población "X".

   Si $Z=\text{mín.}(X_1;\ldots;X_n)$ e $Y=\text{máx.}(X_1;\ldots;X_n)$, el recorrido muestral "R" es $R=Y-Z$, y para determinar la distribución de probabilidad de "R" debemos calcular antes la densidad $g(z;y)$ de la variable bidimensional $(Z;Y)$.

Siendo "f" la función de densidad de probabilidad de la población "X", es:

$$P(z<Z<z+dz;y<Y<y+dy)=g(z;y).dz.dy=$$

$$=\underbrace{5.(5-1)}_{a}.\underbrace{f(z).dz}_{b}.\underbrace{f(y).dy}_{c}.\underbrace{\left(\int_z^y f(x).dx\right)^{5-2}}_{d}\Rightarrow$$

$\boxed{a}$ El "papel" de "mínimo valor muestral" puede hacerlo cualquiera de las 5 variables componentes muestrales, y el "papel" de "máximo valor muestral" puede hacerlo cualquiera de las restantes $5-1$ variables componentes muestrales,; por tanto, el número de pares distintos (mínimo; máximo) en que pueden concretarse las "n" observaciones muestrales es $5.(5-1)$.

$\boxed{b}$ Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(z;z+dz)$.

$\boxed{c}$ Probabilidad de que otra de las variables componentes muestrales tome un valor en el intervalo $(y;y+dy)$.

$\boxed{d}$ Probabilidad de las restantes $5-2$ variables componentes muestrales tomen valores en el intervalo $(z+dz;y+dy)$. Como "dz" y "dy" son infinitésimos, despreciamos su influencia y consideramos el intervalo $(z;y)$.

dividimos los dos miembros por "dz.dy"

$$\Rightarrow g(z;y)=20.f(z).f(y).\left(\int_z^y f(x).dx\right)^3\Rightarrow$$

siendo $X\approx \text{Exp.}(a)$, es $f(x)=a.e^{-a.x}$, $x>0$

$$\Rightarrow g(z;y)=20.\left(a.e^{-a.z}\right).\left(a.e^{-a.y}\right).\left(\int_z^y a.e^{-a.x}.dx\right)^3=$$

$$=20.\left(a.e^{-a.z}\right).\left(a.e^{-a.y}\right).\left(-\left(e^{-a.x}\right)_z^y\right)^3=$$

$$=20.a^2.e^{-a.(z+y)}.\left(e^{-a.z}-e^{-a.y}\right)^3,\ \text{siendo } 0<z\leq y$$

- Haciendo el cambio de variables

$$\left\{\begin{matrix} S=Z \\ R=Y-Z\end{matrix}\right\}\Rightarrow\left\{\begin{matrix} Z=S \\ Y=S+R\end{matrix}\right\}$$

el jacobiano es:

$$J=\begin{vmatrix} \partial z/\partial s & \partial z/\partial r \\ \partial y/\partial s & \partial y/\partial r\end{vmatrix}=\begin{vmatrix} 1 & 0 \\ 1 & 1\end{vmatrix}=1$$

La densidad de probabilidad $h(s;r)$ de la variable $(S;R)$ es:

$$h(s;r)=g(s;s+r).|J| \underset{\uparrow}{=} 20.a^2.e^{-a.(s+s+r)}.\left(e^{-a.s}-e^{-a.(s+r)}\right)^3=$$

$$\boxed{\text{en } g(z;y) \text{ sustituimos "z" por "s" e "y" por "s + r"}}$$

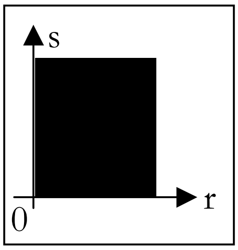
$$=20.a^2.e^{-a.r}.e^{-5.a.s}.\left(1-e^{-a.r}\right)^3, \text{ siendo } \underbrace{\begin{cases} s>0 \\ r\ge 0\end{cases}}$$

$$\boxed{\text{si } \begin{Bmatrix} z>0 \\ z\le y\end{Bmatrix} \text{ y } \begin{Bmatrix} z=s \\ y=s+r\end{Bmatrix} \Rightarrow \begin{Bmatrix} s>0 \\ s\le s+r\end{Bmatrix} \Rightarrow \begin{Bmatrix} s>0 \\ r\ge 0\end{Bmatrix}}$$

- La densidad de probabilidad $f_R(r)$ del recorrido muestral "R" es:

$$f_R(r)=\int_{s=-\infty}^{s=+\infty} h(s;r).ds=\int_{s=0}^{s=+\infty} h(s;r).ds=$$

$$=20.a^2.e^{-a.r}.\left(1-e^{-a.r}\right)^3.\int_{s=0}^{s=+\infty} e^{-5.a.s}.ds \underset{\uparrow}{=}$$



$$\boxed{\int_{s=0}^{s=+\infty} e^{-5.a.s}.ds=\frac{1}{5.a}.\int_{s=0}^{s=+\infty} e^{-5.a.s}.d(5.a.s)=\frac{\Gamma(1)}{5.a}=\frac{1}{5.a}}$$

$$=4.a.e^{-a.r}.\left(1-e^{-a.r}\right)^3;\ r\ge 0$$

2) Es: $\quad P(R>1)=\int_1^{+\infty} f_R(r).dr=\int_1^{+\infty} 4.a.e^{-a.r}.\left(1-e^{-a.r}\right)^3.dr=$

$$=\left(\left(1-e^{-a.r}\right)^4\right)_1^{+\infty}=1-\left(1-e^{-a}\right)^4$$

FONEMATO 2.3.5

Sea "X" una población continua de la que se toma m.a.s de tamaño impar $2.n+1$.

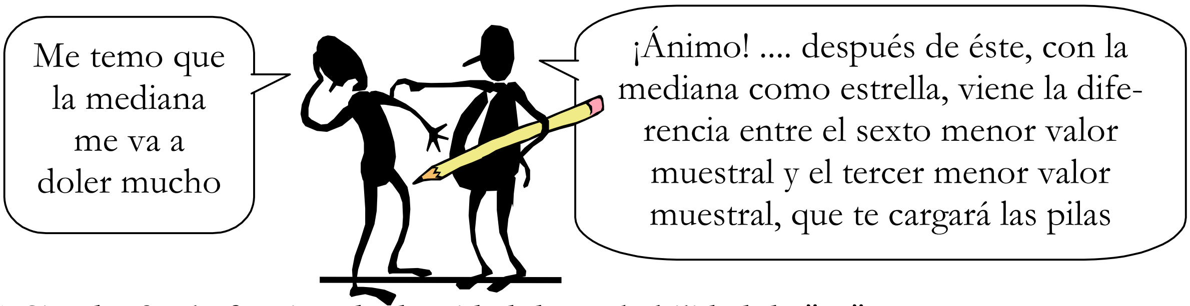
1) Determínese la distribución de la mediana muestral "M".
2) Particularícese en el caso $n=2$ si la densidad de "X" es $f(x)=2.x,\ x\in[0;1]$.
3) En las condiciones de 2), calcúlense la media y la varianza de "M".

SOLUCIÓN

Si de una población continua "X" se toma m.a.s de tamaño 9 (impar) y se obtiene la $(13;16;9;7;18;6;8;5,3)$, ordenando de menor a mayor los valores observados, resulta:

$$\boxed{3<5<6<7}<8<\boxed{9<13<16<18}$$

Por tanto, en esa muestra la aleatoria mediana muestral "M" se concreta en el número 8, que es la observación inferior a la mitad de las restantes observaciones y superior a la otra mitad.



1) Siendo f_M la función de densidad de probabilidad de "M", es:

$$f_M(m).dm = P(m < M < m + dm) =$$

$$= \frac{(2.n+1)!}{n!.1!.n!}.\left(\int_{-\infty}^{m} f(x).dx\right)^n.(f(m).dm).\left(\int_{m+dm}^{+\infty} f(x).dx\right)^n \Rightarrow$$

En el experimento consistente en observar una vez el valor que toma la variable "X", por fuerza ha de ocurrir alguno de los siguientes sucesos:

* $A_1 \equiv$ la variable "X" toma un valor no superior a "m"
* $A_2 \equiv$ la variable "X" toma un valor en el intervalo $(m; m + dm)$
* $A_3 \equiv$ la variable "X" toma un valor no inferior a "$m + dm$"

Siendo "f" la función de densidad de "X", es:

$$P(A_1) = \int_{-\infty}^{m} f(x).dx \;;\; P(A_2) = f(m).dm \;;\; P(A_3) = \int_{m+dm}^{+\infty} f(x).dx$$

Si el experimento se repite $2.n+1$ veces, la mediana muestral "M" toma un valor en el intervalo $(m; m + dm)$ sólo si el suceso A_1 ocurre "n" veces, el suceso A_2 ocurre una vez y el suceso A_3 ocurre "n" veces .... y la probabilidad de tal contingencia es:

$$\underbrace{PR_{n;1;n}^{2n+1}.(P(A_1))^n.P(A_2).(P(A_3))^n}$$

La probabilidad de que ocurra

$$\underbrace{A_1, A_1, \ldots, A_1}_{\text{"n" veces}}, \underbrace{A_2}_{\text{una vez}}, \underbrace{A_3, A_3, \ldots, A_3}_{\text{"n" veces}}$$

es $(P(A_1))^n.P(A_2).(P(A_3))^n$, y debe multiplicarse por $PR_{n;1;n}^{2n+1}$, que es el número de permutaciones con repetición (alineaciones) que pueden formarse con A_1, A_2 y A_3 de modo que A_1 y A_3 estén presentes "n" veces y A_2 una vez, siendo:

$$PR_{n;1;n}^{2n+1} = \frac{(2.n+1)!}{n!.1!.n!}$$

$$\Rightarrow f_M(m) = \frac{(2.n+1)!}{n!.1!.n!}.\left(\int_{-\infty}^{m} f(x).dx\right)^n.f(m).\left(\int_{m}^{+\infty} f(x).dx\right)^n$$

dividimos por "dm" y pasamos del "dm" de la última integral

2) Si $n=2$ y la densidad de "X" es $f(x)=2.x,\ x\in[0;1]$, la densidad de "M" es:

$$f_M(m)=\frac{5!}{2!.1!.2!}.\left(\int_{-\infty}^{m} f(x).dx\right)^2.f(m).\left(\int_{m}^{+\infty} f(x).dx\right)^2=$$

$$=30.\left(\int_0^m 2.x.dx\right)^2.(2.m).\left(\int_m^1 2.x.dx\right)^2=....=60.m^5.(1-m^2)^2,\quad m\in[0:1]$$

$$\boxed{f(x)=2.x,\ x\in[0;1]}$$

2) Es:

- $$E(M)=\int_{-\infty}^{+\infty} m.f_m(m).dm=\int_0^1 60.m^6.(1-m^2)^2.dm=$$

$m=\sqrt{u}\Rightarrow dm=du/(2.\sqrt{u})$
los límites de integración no cambian

$$=60.\int_0^1 u^3.(1-u)^2.\frac{du}{2.\sqrt{u}}=30.\int_0^1 u^{5/2}.(1-u)^2.du=30.\beta(7/2;3)=$$

$\beta(r;s)=\int_0^1 u^{r-1}.(1-u)^{s-1}.du$

En nuestro caso es $\begin{cases} r-1=5/2 \\ s-1=3\end{cases}\Rightarrow\begin{cases} r=7/2 \\ s=4\end{cases}$

$$=30.\frac{\Gamma(7/2).\Gamma(3)}{\Gamma(13/2)}=\frac{30.\Gamma(7/2).\Gamma(3)}{(11/2).(9/2).(7/2).\Gamma(7/2)}=\frac{30.\Gamma(3)}{(11/2).(9/2).(7/2)}=\frac{160}{231}$$

- $$V(M)=E(M^2)-(E(M))^2=\frac{1}{2}-\left(\frac{160}{231}\right)^2$$

$$E(M^2)=\int_{-\infty}^{+\infty} m^2.f_m(m).dm=\int_0^1 60.m^7.(1-m^2)^2.dm=$$

$m=\sqrt{u}\Rightarrow dm=du/(2.\sqrt{u})$
los límites de integración no cambian

$$=60.\int_0^1 u^{7/2}.(1-u)^2.\frac{du}{2.\sqrt{u}}=30.\int_0^1 u^3.(1-u)^2.du=$$

$$=30.\beta(4;3)=30.\frac{\Gamma(4).\Gamma(3)}{\Gamma(7)}=30.\frac{3!.2!}{6!}=\frac{1}{2}$$

FONEMATO 2.3.6

Si se toma m.a.s de tamaño 9 de una población "X" con distribución uniforme en el intervalo (0;1), determínese la distribución de probabilidad de la diferencia entre el sexto menor valor muestral y el tercer menor valor muestral.

SOLUCIÓN

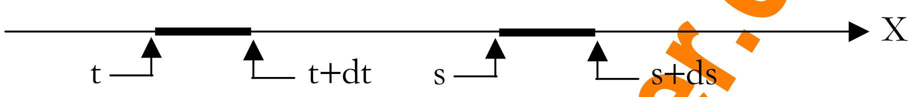
Siendo "S" el sexto menor valor muestral y "T" el tercer menor valor muestral, sea $Z=S-T$ su diferencia. Para determinar la distribución de probabilidad de la variable aleatoria $Z=S-T$ debemos calcular previamente la función de densidad de probabilidad $g(t;s)$ de la variable bidimensional $(T;S)$; es:

$$P\left(t<T<t+dt;s<S<s+ds\right)=g(t;s).dt.ds=$$

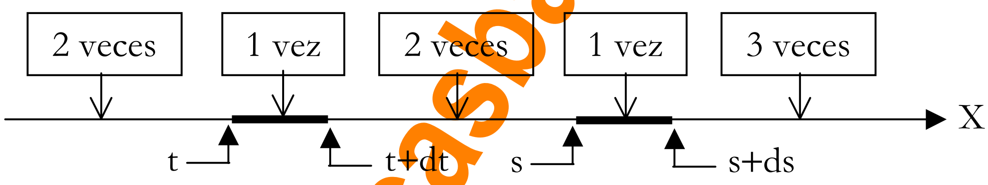
$$=\frac{9!}{2!.1!.2!.1!.3!}.\left(P(A_1)\right)^2.P(A_2).\left(P(A_3)\right)^2.P(A_4).\left(P(A_5)\right)^3=$$

En el experimento consistente en observar <u>una vez</u> el valor que toma la variable "X", por fuerza ha de ocurrir alguno de los siguientes sucesos:

* $A_1 \equiv$ la variable "X" toma un valor no superior a "t"
* $A_2 \equiv$ la variable "X" toma un valor en el intervalo $(t;t+dt)$
* $A_3 \equiv$ la variable "X" toma un valor en el intervalo $[t+dt;s]$
* $A_4 \equiv$ la variable "X" toma un valor en el intervalo $(s;s+ds)$
* $A_5 \equiv$ la variable "X" toma un valor no inferior a "s + ds"



Si el experimento se repite 9 veces, el tercer menor valor muestral "T" toma un valor en el intervalo $(t;t+dt)$ y el sexto menor valor muestral toma un valor en el intervalo $(s;s+ds)$ sólo si el suceso A_1 ocurre 2 veces, el suceso A_2 ocurre 1 vez, el suceso A_3 ocurre 2 veces, el suceso A_4 ocurre 1 vez y el suceso A_5 ocurre 3 veces.



La probabilidad de tal contingencia es:

$$\underbrace{PR^{9}_{2;1;2;1;3}.\left(P(A_1)\right)^2.P(A_2).\left(P(A_3)\right)^2.P(A_4).\left(P(A_5)\right)^3}$$

La probabilidad de que ocurra

$$\underbrace{A_1,A_1}_{2\text{ veces}},\underbrace{A_2}_{\text{una vez}},\underbrace{A_3,A_3}_{2\text{ veces}},\underbrace{A_4}_{\text{una vez}},\underbrace{A_5,A_5,A_5}_{3\text{ veces}}$$

es

$$\left(P(A_1)\right)^2.P(A_2).\left(P(A_3)\right)^2.P(A_4).\left(P(A_5)\right)^3$$

y debe multiplicarse por $PR^{9}_{2;1;2;1;3}$, que es el número de permutaciones con repetición que pueden formarse con A_1,A_2,A_3,A_4 y A_5 de modo que A_1 y A_3 estén presentes 2 veces, A_2 y A_4 estén presentes 1 vez y A_5 esté presente 3 veces, siendo:

$$PR^{9}_{2;1;2;1;3}=\frac{9!}{2!.1!.2!.1!.3!}$$

$$=\frac{9!}{2!.2!.3!}.(t)^2.(dt).(s-t)^2.(ds).(1-t)^3$$

Si $X \approx U(0;1)$, su función de densidad es $f(x)=1,\ x\in(0;1)$; por tanto:

$$P(A_1)=P(X\le t)=\int_0^t dx=t$$
$$P(A_2)=P(t<X<t+dt)=f(t).dt=dt$$
$$P(A_3)=P(t+dt\le X\le s)=P(t\le X\le s)=\int_t^s dx=(s-t)$$
$$P(A_4)=P(s<X\le s+ds)=f(s).ds=ds$$
$$P(A_5)=P(X\ge s+ds)=P(X\ge s)=\int_s^1 dx=(1-s)$$

Por tanto:

$$g(t;s)=\frac{9!}{2!.2!.3!}.(t)^2.(s-t)^2.(1-t)^3$$

siendo

$$0<t<1\ ;\ 0<s<1\ ;\ t\le s$$

Haciendo el cambio de variables $\begin{Bmatrix} Z=S-T \\ W=S \end{Bmatrix} \Rightarrow \begin{Bmatrix} T=W-Z \\ S=W \end{Bmatrix}$, el jacobiano es:

$$J=\begin{vmatrix} \partial t/\partial w & \partial t/\partial z \\ \partial s/\partial w & \partial s/\partial z \end{vmatrix}=\begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix}=1$$

La densidad de probabilidad $h(w;z)$ de la variable $(W;Z)$ es:

$$h(w;z)=g(w-z;w).|J|=$$

en $g(t;s)$ sustituimos "t" por "$w-z$" y "s" por "w"

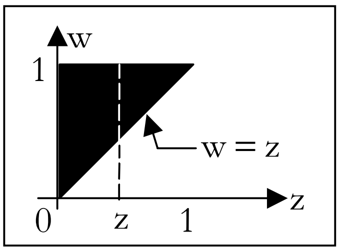
$$=\frac{9!}{2!.2!.3!}.(w-z)^2.(w-(w-z))^2.(1-w)^3=$$

$$=\frac{9!}{2!.2!.3!}.(w-z)^2.z^2.(1-w)^3$$

siendo: $\underbrace{0<w-z<1\ ;\ 0<w<1\ ;\ z\ge 0}$

si $\begin{Bmatrix} 0<t<1 \\ 0<s<1 \\ t\le s \end{Bmatrix}$ y $\begin{Bmatrix} t=w-z \\ s=w \end{Bmatrix} \Rightarrow \begin{Bmatrix} 0<w-z<1 \\ 0<w<1 \\ w-z\le w \end{Bmatrix} \Rightarrow \begin{Bmatrix} 0<w-z<1 \\ 0<w<1 \\ z\ge 0 \end{Bmatrix}$

La densidad de probabilidad $f_Z(z)$ de "Z" $(Z=S-T)$ es:



$$f_Z(z)=\int_{w=-\infty}^{w=+\infty} h(w;z).dw=$$

$$=\int_{w=z}^{w=1} \frac{9!}{2!.2!.3!}.(w-z)^2.z^2.(1-w)^3.dw=....$$

siendo $z\in[0;1)$.

Siendo $\overline{X}$ el estadístico media muestral en el muestreo aleatorio simple de tamaño "n" de una población "X" de media μ y varianza σ^2, es:

$$E(\overline{X}) = E\left(\frac{1}{n}.\sum_{i=1}^{n} X_i\right) = \frac{1}{n}.\sum_{i=1}^{n} E(X_i) \underset{\uparrow}{=} \frac{1}{n}.n.\mu = \mu$$

$$\boxed{E(X_i) = \mu \ \ ,\forall\ i = 1,2, ... ,n}$$

$$\boxed{V(X_i) = \sigma^2 \ \ ,\forall\ i = 1,2, ... ,n}$$

$$V(\overline{X}) = V\left(\frac{1}{n}.\sum_{i=1}^{n} X_i\right) \underset{\uparrow}{=} \frac{1}{n^2}.\sum_{i=1}^{n} V(X_i) \overset{\downarrow}{=} \frac{1}{n^2}.n.\sigma^2 = \frac{\sigma^2}{n}$$

$$\boxed{\text{pues } X_1, X_2, ... , X_n \text{ son independientes}}$$

O sea, el centro de gravedad de la distribución de masa de probabilidad asociada al estadístico $\overline{X}$ coincide con el centro de gravedad de la distribución de masa de probabilidad asociada a "X", pero la variabilidad de $\overline{X}$, expresada por $V(\overline{X})$, es la n-ésima parte de la variabilidad de "X", expresada por $V(X)$. ***Por ejemplo,*** si $n = 100$, la varianza de $\overline{X}$ es la centésima parte de la varianza de "X".

En el Tema 3, para expresar que el valor esperado del estadístico media muestral $\overline{X}$ siempre coincide con el valor esperado μ de la población "X", diremos que $\overline{X}$ es un **estimador insesgado** de la media poblacional μ.

Función generatriz de momentos del estadístico media muestral

Si $\Phi_X(t)$ es la f.g.m. de la variable "X", la f.g.m. de $\overline{X}$ es $\Phi(t) = \left(\Phi_X(t/n)\right)^n$:

$$\Phi(t) = E\left(e^{t.\overline{X}}\right) = E\left(e^{t.(X_1 + X_2 + ... + X_n)/n}\right) = E\left(\prod_{i=1}^{n} e^{t.X_i/n}\right) \underset{\uparrow}{=}$$

$$\boxed{\text{pues } X_1, X_2, ... , X_n \text{ son independientes}}$$

$$\boxed{\begin{array}{c}\text{como } X_1, X_2, ... , X_n \text{ tienen la misma distribución de probabilidad} \\ \text{que "X", es } E\left(e^{t.X_i/n}\right) = E\left(e^{t.X/n}\right),\ \forall\ i = 1,2, ... ,n\end{array}}$$

$$= \prod_{i=1}^{n} E\left(e^{t.X_i/n}\right) \overset{\downarrow}{=} \left(E\left(e^{t.X/n}\right)\right)^n \underset{\uparrow}{=} \left(\Phi_X(t/n)\right)^n$$

$$\boxed{\text{pues } E\left(e^{\text{Pepe}.X}\right) = \Phi_X(\text{Pepe})}$$

Por ejemplo, siendo $X \approx G(p;a)$ es $\Phi_X(t) = \left(1 - (t/a)\right)^{-p}$, y la f.g.m. de $\overline{X}$ es:

$$\Phi(t) = \left(\Phi_X(t/n)\right)^{-n} = \left(\left(1 - \frac{t}{a.n}\right)^{-p}\right)^n = \left(1 - \frac{t}{a.n}\right)^{-p.n}$$

que corresponde a una variable con distribución $G(p.n;a.n)$.

Función de cuantía/densidad del estadístico media muestral

La determinación de la función de cuantía/densidad de la media muestral $\overline{X}$ puede ser cosa fácil o menos fácil.

- Si el tamaño muestral n" es grande (≥ 30), el teorema Central del Límite nos permitirá aproximar la distribución de $\overline{X}$ mediante la normal con media y varianza las de $\overline{X}$; es decir, mediante la $N(\mu;\sigma/\sqrt{n})$.

- **Para muchas variables famosas, con independencia del tamaño muestral, podremos determinar la función de cuantía/densidad del estadístico media muestral a través del estadístico "suma total".**

  Ejemplo 1: si la variable "X" tiene distribución de Bernouilli de parámetro "p", siendo $T = X_1 + ... + X_n \approx B(n;p)$ y $\overline{X} = (X_1 + ... + X_n)/n = T/n$, es:

$$P(\overline{X} = x) = P\big(T/n = x\big) = P(T = n.x) =$$

$$\boxed{T \approx B(n;p) \Rightarrow P(T = t) = \binom{n}{t}.p^t.(1-p)^{n-t}, \ t = 0,1,2,.....,n}$$

$$= \binom{n}{n.x}.p^{n.x}.(1-p)^{n-n.x}, \ x = 0,\frac{1}{n},\frac{2}{n},....,1$$

  Ejemplo 2: si la variable "X" tiene distribución $B(k;p)$, siendo

$$T = X_1 + ... + X_n \approx B(n.k;p) \ ; \ \overline{X} = (X_1 + ... + X_n)/n = T/n$$

  es:

$$P(\overline{X} = x) = P\big(T/n = x\big) = P(T = n.x) =$$

$$\boxed{T \approx B(n.k;p) \Rightarrow P(T = t) = \binom{n.k}{t}.p^t.(1-p)^{n.k-t}, \ t = 0,1,2,.....,n.k}$$

$$= \binom{n.k}{n.x}.p^{n.x}.(1-p)^{n.k-n.x}, \ x = 0,\frac{1}{n},\frac{2}{n},....,k$$

  Ejemplo 3: si la variable "X" tiene distribución $P(\theta)$, siendo

$$T = X_1 + ... + X_n \approx P(n.\theta) \ ; \ \overline{X} = (X_1 + ... + X_n)/n = T/n$$

  es:

$$P(\overline{X} = x) = P\big(T/n = x\big) = P(T = n.x) =$$

$$\boxed{T \approx P(n.\theta) \Rightarrow P(T = t) = \frac{e^{-n.\theta}.(n.\theta)^t}{t!}, \ t = 0,1,2,.....}$$

$$= \frac{e^{-n.\theta}.(n.\theta)^{n.x}}{(n.x)!}, \ x = 0,\frac{1}{n},\frac{2}{n},.......$$

  Ejemplo 4: si la variable "X" tiene distribución $G(p)$, siendo

$$T = X_1 + ... + X_n \approx BN(n;p) \ ; \ \overline{X} = (X_1 + ... + X_n)/n = T/n$$

es: $$P(\overline{X}=x)=P(T/n=x)=P(T=n.x)=$$

$$\boxed{T\approx BN(n;p)\Rightarrow P(T=t)=\binom{t+n-1}{t}.(1-p)^t.p^n,\ t=0,1,2,.....}$$

$$=\binom{n.x+n-1}{n.x}.(1-p)^{n.x}.p^n,\ \ x=0,\frac{1}{n},\frac{2}{n},.......$$

Ejemplo 5: si la variable "X" tiene distribución $BN(k;p)$, siendo

$$T=X_1+....+X_n\approx BN(n.k;p)\ ;\ \overline{X}=(X_1+....+X_n)/n=T/n$$

es:

$$P(\overline{X}=x)=P(T/n=x)=P(T=n.x)=$$

$$\boxed{T\approx BN(n.k;p)\Rightarrow P(T=t)=\binom{t+n.k-1}{t}.(1-p)^t.p^{n.k},\ t=0,1,2,.....}$$

$$=\binom{n.x+n.k-1}{n.x}.(1-p)^{n.x}.p^{n.k},\ \ x=0,\frac{1}{n},\frac{2}{n},.......$$

Ejemplo 6: si la variable "X" tiene distribución $Exp.(a)\equiv G(1;a)$, siendo

$$T=X_1+....+X_n\approx G(n;a)$$

la función de densidad de probabilidad de "T" es:

$$f_T(t)=\frac{a^n}{\Gamma(n)}.t^{n-1}.e^{-a.t},\ \ t>0$$

Así, siendo $\overline{X}=T/n\ \ (\Rightarrow T=n.\overline{X})$ la aleatoria media muestral y denotando $g(x)$ a la función densidad de probabilidad de $\overline{X}$, es:

$$g(x)=f_T(n.x).\left|\frac{d(n.x)}{dx}\right|=\frac{a^n}{\Gamma(n)}.(n.x)^{n-1}.e^{-a.n.x}.n=$$

$$=\frac{(a.n)^n}{\Gamma(n)}.x^{n-1}.e^{-(a.n).x},\ \ x>0$$

que es la densidad de una variable con distribución $G(n;a.n)$.

Ejemplo 7: si la variable "X" tiene distribución $G(p;a)$, siendo

$$T=X_1+....+X_n\approx G(p.n;a)$$

la función de densidad de probabilidad de "T" es:

$$f_T(t)=\frac{a^{p.n}}{\Gamma(p.n)}.t^{p.n-1}.e^{-a.t},\ \ t>0$$

Por tanto, siendo $\overline{X}=T/n\ \ (\Rightarrow T=n.\overline{X})$ la aleatoria media muestral y denotando $g(x)$ a la función densidad de probabilidad de $\overline{X}$, es:

$$g(x)=f_T(n.x).\left|\frac{d(n.x)}{dx}\right|=\frac{a^{p.n}}{\Gamma(p.n)}.(n.x)^{p.n-1}.e^{-a.n.x}.n=$$

$$=\frac{(a.n)^{p.n}}{\Gamma(p.n)}.x^{p.n-1}.e^{-a.n.x},\ \ x>0$$

que es la densidad de una variable con distribución $G(p.n;a.n)$.

¡No estamos descubriendo la pólvora!

Los 7 ejemplos precedentes son auténticas gilipolleces que aprendimos a lidiar cuando en el Tema 2 estudiamos el asunto de los cambios de variable unidimensionales: se conoce la función de cuantía/densidad de una variable "T" (la suma total) y nos planteamos el cálculo de la función de cuantía/densidad de otra variable $\overline{X}$ (la media muestral) que está ligada con "T" por la relación monótona $T = n.\overline{X}$

Ejemplo 8: si se toma una muestra de tamaño 2 de una variable "X" con distribución $U(0;3)$, el problema de determinar la distribución de probabilidad del estadístico media muestral es distinto a los anteriores, pues **al contrario que en los ejemplos precedentes, ahora no tenemos ni puta idea de la distribución de probabilidad del estadístico suma total.**

Siendo "X" continua y muestra de tamaño 2, para calcular la densidad de la media muestral "W" (la denotamos "W" en vez de $\overline{X}$, por comodidad), trabajaremos como aprendimos en el Tema 3: partiendo de la densidad $f(x_1;x_2)$ de la variable muestral $(X_1;X_2)$, calcularemos la densidad $g(w;u)$ de la variable bidimensional $(W;U)$, siendo $U = X_1$ ó $U = X_2$, y después calcularemos la densidad de "W" como marginal de la bidimensional $(W;U)$.

Vamos al tajo: como las variables X_1 y X_2 son independientes, la función de densidad $f(x_1;x_2)$ de la variable bidimensional $(X_1;X_2)$ es el producto de las respectivas funciones de densidad de X_1 y X_2; por tanto, como X_1 y X_2 tienen distribución $U(0;3)$, es:

$$f(x_1;x_2) = f_1(x_1).f_2(x_2) \underset{\uparrow}{=} 1/9,\ \ 0 < x_1 < 3,\ 0 < x_2 < 3$$

$$\boxed{X_i \approx U(0;3) \Rightarrow f_i(x_i) = 1/3,\ \ 0 < x_i < 3,\ i = 1,2}$$

Recuerda: si $f(x_1;x_2)$ es la densidad de la variable bidimensional continua $(X_1;X_2)$ definida en un dominio $D \subseteq \Re^2$, siendo $X_1 = h_1(W;U)$ y $X_2 = h_2(W;U)$, la densidad $g(w;u)$ de la variable bidimensional $(W;U)$, es:

$$g(w;u) = f(h_1(w;u);h_2(w;u)).\left|J\right| \ ;\ (w;u) \in D^*$$

donde D^* es el dominio en que se transforma "D" como consecuencia del cambio de variables, y "J" es el jacobiano asociado a dicho cambio de variables.

* Siendo $W = (X_1 + X_2)/2$ y $U = X_1$, al despejar X_1 y X_2, resulta:

$$X_1 = h_1(W;U) = U \ ; \ X_2 = h_2(W;U) = 2.W - U$$

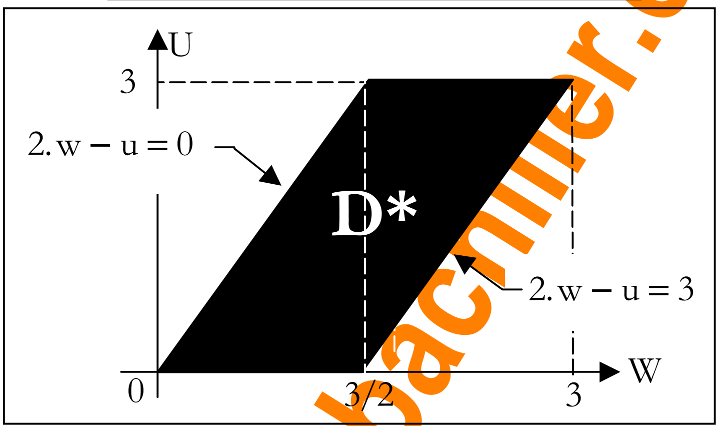
* El jacobiano es $J = \begin{vmatrix} \partial x_1/\partial w & \partial x_1/\partial u \\ \partial x_2/\partial w & \partial x_2/\partial u \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 2 & -1 \end{vmatrix} = -2 \Rightarrow |J| = 2$

* El dominio $D = \left\{(x_1;x_2) \in \Re^2 \ / \ 0 < x_1 < 3, \ 0 < x_2 < 3\right\}$ se transforma en:

$$D^* = \left\{(w;u) \in \Re^2 \ / \ 0 < u < 3, 0 < 2.w - u < 3\right\}$$

$$\left.\begin{matrix} 0 < x_1 < 3 \\ 0 < x_2 < 3 \end{matrix}\right\} \Rightarrow \left\{\begin{matrix} 0 < u < 3 \\ 0 < 2.w - u < 3 \end{matrix}\right.$$

$$\boxed{x_1 = u \ ; \ x_2 = 2.w - u}$$



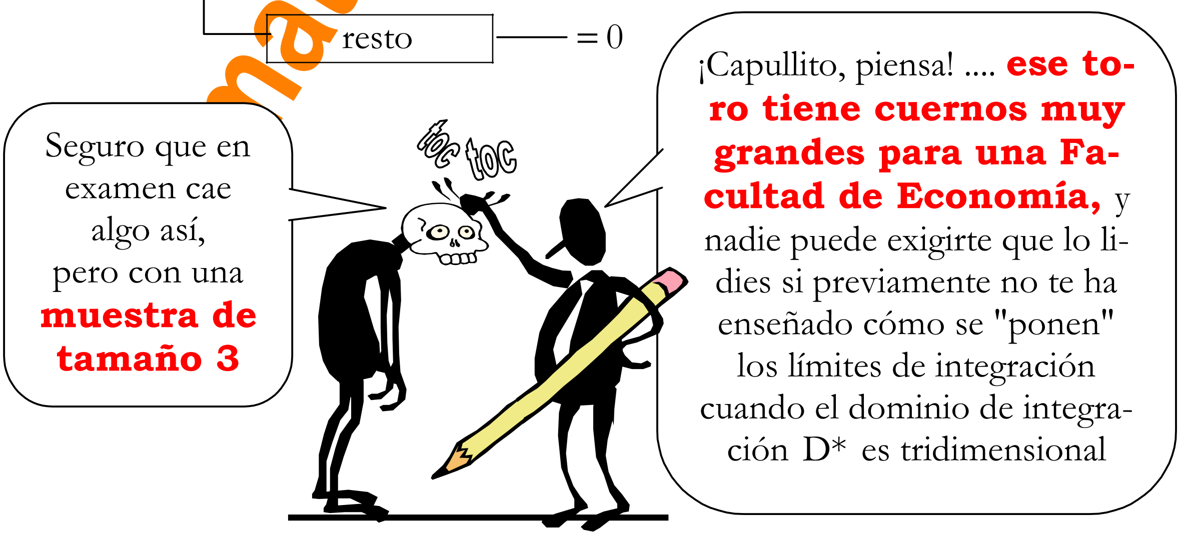
* Función de densidad $g(w;u)$ de la variable $(W;U)$:

$$g(w;u) = f(h_1(w;u);h_2(w;u)).|J| = 2/9 \ ; \ (w;u) \in D^*$$

* Como queda dicho, calculamos la densidad $g_1(w)$ del estadístico media muestral "W" como marginal de la bidimensional $(W;U)$:

$$g_1(w) = \int_{u=-\infty}^{u=+\infty} g(w;u).du =$$

$$= \begin{cases} \boxed{\text{si } 0 < w < 3/2} \longrightarrow = \int_{u=0}^{u=2.w} 2.du/9 = 4.w/9 \\ \boxed{\text{si } 3/2 \le w < 3} \longrightarrow = \int_{u=0}^{u=2.w-3} 2.du/9 = (4.w-6)/9 \\ \boxed{\text{resto}} \longrightarrow = 0 \end{cases}$$



Naturalmente, **una vez conocida la distribución de probabilidad del estadístico media muestral, puede marearse la perdiz con cualquiera de las puñetas típicas de una variable aleatoria unidimensional;** es decir, pueden pedirte la media o la varianza de "W", la función generatriz de momentos de "W", la función de distribución de "W", los coeficientes de asimetría y de curtosis de "W", probabilidades de sucesos relacionados con "W", etc. .... y nada nuevo bajo el sol:

$$* \ E(W)=\int_0^3 w.g_1(w).dw=\int_0^{3/2}\frac{4.w^2}{9}.dw+\int_{3/2}^3\frac{4.w^2-6.w}{9}.dw=....$$

$$* \ V(W)=E\left(W^2\right)-\left(E(W)^2\right)=......$$

$$\boxed{E(W^2)=\int_0^3 w^2.g_1(w).dw=\int_0^{3/2}\frac{4.w^3}{9}.dw+\int_{3/2}^3\frac{4.w^3-6.w^2}{9}.dw}$$

$$* \ \Phi(t)=E\left(e^{t.W}\right)=\int_0^3 e^{t.w}.g_1(w).dw=$$

$$=\int_0^{3/2}\frac{4.w}{9}.e^{t.w}.dw+\int_{3/2}^3\frac{4.w-6}{9}.e^{t.w}.dw=....$$

$$* \ P(0'3<W<1'7)=\int_{0'3}^{1'7} g_1(w).dw=\int_{0'3}^{3/2}\frac{4.w}{9}.dw+\int_{3/2}^{1'7}\frac{4.w-6}{9}.dw=....$$

Observa: una vez conocida la función de densidad de probabilidad del estadístico media muestral "W", el cálculo de la función de densidad de probabilidad $r(t)$ del estadístico suma total $T=2.W$ $(\Rightarrow W=T/2)$ se reduce a un cambio de variable unidimensional; es:

$$r(t)=g_1(t/2).\left|\frac{d(t/2)}{dt}\right|=\frac{1}{2}.g_1(t/2)=$$

$$=\begin{cases}\boxed{\text{si } 0<t<3} & =\dfrac{4}{9}.\dfrac{t}{2}.\dfrac{1}{2}\\ \boxed{\text{si } 3\le t<6} & =\dfrac{1}{9}.\left(4.\dfrac{t}{2}-6\right).\dfrac{1}{2}\\ \boxed{\text{resto}} & =0\end{cases}$$

Que quede claro: el método seguido en este ejemplo para calcular la distribución de probabilidad de la media muestral (siendo continua la población "X", con muestra de tamaño 2), puedes emplearlo para cualquier estadístico.

Por ejemplo, para calcular la función de densidad del estadístico $Z=X_1.X_2$, partiendo de la densidad $f(x_1;x_2)$ de la variable muestral $(X_1;X_2)$, calcularemos la densidad $s(z;v)$ de la variable bidimensional $(Z;V)$, siendo $V=X_1$ ó $V=X_2$, y después calcularemos la densidad de "Z" como marginal de la bidimensional $(Z;V)$. Vamos al tajo:

* Siendo $Z=X_1.X_2$ y $V=X_1$, al despejar X_1 y X_2, resulta:

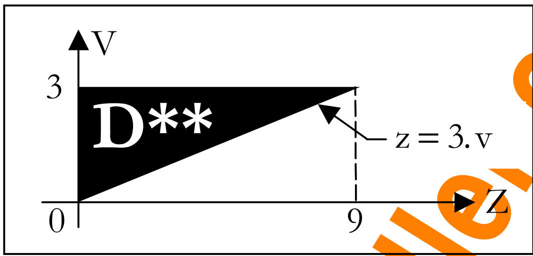
$$X_1=\gamma_1(Z;V)=V \ ; \ X_2=\gamma_2(Z;V)=Z/V$$

* El jacobiano es $J = \begin{vmatrix} \partial x_1/\partial z & \partial x_1/\partial v \\ \partial x_2/\partial z & \partial x_2/\partial v \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1/v & -z/v^2 \end{vmatrix} = \dfrac{1}{v}$

* El dominio $D = \left\{(x_1;x_2) \in \Re^2 \;/\; 0 < x_1 < 3,\; 0 < x_2 < 3\right\}$ se transforma en:

$$D^{**} = \left\{(z;v) \in \Re^2 \;/\; 0 < v < 3, 0 < z < 3.v\right\}$$

$$\left.\begin{matrix} 0 < x_1 < 3 \\ 0 < x_2 < 3 \end{matrix}\right\} \underset{\boxed{x_1 = v\,;\, x_2 = z/v}}{\Rightarrow} \left\{\begin{matrix} 0 < v < 3 \\ 0 < z/v < 3 \Rightarrow 0 < z < 3.v \end{matrix}\right.$$



* Función de densidad $s(z;v)$ de la variable $(Z;V)$:

$$s(z;v) = f(\gamma_1(z;v);\gamma_2(z;v)).\left|J\right| = \frac{1}{9.\left|v\right|} \quad ; \quad (z;v) \in D^{**}$$

* Como queda dicho, calculamos la densidad $s_1(z)$ de "Z" como marginal de la bidimensional $(Z;V)$:

$$s_1(z) = \int_{v=-\infty}^{v=+\infty} s(z;v).dv = \int_{v=z/3}^{v=3} \frac{1}{9.v}.dv = \frac{1}{9}.\mathrm{Ln}\,\frac{9}{z},\;\; 0 < z < 9$$

$$\boxed{\text{en } D^{**} \text{ es } v > 0; \text{ por tanto: } \frac{1}{9.\left|v\right|} = \frac{1}{9.v}}$$

Ejemplo 9: considera que se toma una muestra de tamaño 2 de una variable discreta "X" cuya función de cuantía es

x	0	1	2
$P(X = x)$	0'5	0'2	0'3

El problema de determinar la función de cuantía del estadístico media muestral es distinto de los anteriores, pues ahora, además de desconocer la distribución de probabilidad del estadístico suma total, hay una **putadita adicional: la función de cuantía de la variable discreta "X" no tiene una expresión genérica para todo valor que puede tomar "X".**

Siendo "X" discreta y muestra de tamaño 2, la determinación de la función de cuantía de la media muestral (variable aleatoria discreta) es una gilipollez que aprendimos a lidiar en el Tema 3: calcularemos el valor que toma la media muestral en cada posible muestra (o sea, en cada punto de la distribución bidimensional discreta asociada a la variable muestral $(X_1;X_2)$), y después obraremos en consecuencia.

Vamos al tajo: determinamos la función de probabilidad o cuantía de la variable muestral $(X_1;X_2)$: como X_1 y X_2 tienen la misma distribución que "X", es:

$$\begin{array}{c|ccc} x_1 & 0 & 1 & 2 \\ \hline P(X_1 = x_1) & 0'5 & 0'2 & 0'3 \end{array} \; ; \; \begin{array}{c|ccc} x_2 & 0 & 1 & 2 \\ \hline P(X_2 = x_2) & 0'5 & 0'2 & 0'3 \end{array}$$

Como las variables X_1 y X_2 son independientes, la función de cuantía de la variable bidimensional $(X_1;X_2)$ es el producto de las respectivas funciones de cuantía de X_1 y X_2; por tanto:

$$P\left(X_1 = x_1; X_2 = x_2\right) = P(X_1 = x_1).P(X_2 = x_2) \Rightarrow$$

$$\Rightarrow \begin{array}{|c|ccc|} \hline X_1 / X_2 & 0 & 1 & 2 \\ \hline 0 & 0'25 & 0'10 & 0'15 \\ 1 & 0'10 & 0'04 & 0'06 \\ 2 & 0'15 & 0'06 & 0'09 \\ \hline \end{array}$$

Ahora calculamos el valor que toma $\overline{X} = (X_1 + X_2)/2$ en cada punto de la distribución de probabilidad de la variable bidimensional $(X_1;X_2)$:

$(x_1;x_2)$	$P\left(X_1 = x_1; X_2 = x_2\right)$	$\overline{X} = (X_1 + X_2)/2$
(0;0)	0'25	0
(0;1)	0'10	0'5
(0;2)	0'15	1
(1;0)	0'10	0'5
(1;1)	0'04	1
(1;2)	0'06	1'5
(2;0)	0'15	1
(2;1)	0'06	1'5
(2;2)	0'09	2

A la vista del valor que toma $\overline{X} = (X_1 + X_2)/2$ en cada posible muestra y de la probabilidad que corresponde a esa muestra, con la gorra determinamos la función de cuantía del estadístico $\overline{X}$; es:

$$\left\{\begin{array}{l} P(\overline{X} = 0) = P\left(0;0\right) = 0'25 \\ P(\overline{X} = 0'5) = P\left(0;1\right) + P\left(1;0\right) = 0'20 \\ P(\overline{X} = 1) = P\left(0;2\right) + P\left(1;1\right) + P\left(2;0\right) = 0'34 \\ P(\overline{X} = 1'5) = P\left(1;2\right) + P\left(2;1\right) = 0'12 \\ P(\overline{X} = 2) = P\left(2;2\right) = 0'09 \end{array}\right\} \Rightarrow$$

$$\Rightarrow \begin{array}{c|ccccc} x & 0 & 0'5 & 1 & 1'5 & 2 \\ \hline P(\overline{X} = x) & 0'25 & 0'20 & 0'34 & 0'12 & 0'09 \end{array}$$

Como queda dicho en el ejemplo anterior, una vez conocida la distribución de probabilidad del estadístico media muestral $\overline{X}$, puede marearse la perdiz no cualquiera de las puñetas típicas de una variable aleatoria unidimensional:

$$* \ E(\overline{X}) = 0.0'25 + 0'5.0'20 + 1.0'34 + 1'5.0'12 + 2.0'09 = ...$$

$$* \ V(\overline{X}) = E\left(\overline{X}^2\right) - \left(E(\overline{X})^2\right) = ...$$

$$\boxed{E(\overline{X}^2) = 0^2.0'25 + 0'5^2.0'20 + 1^2.0'34 + 1'5^2.0'12 + 2^2.0'09}$$

$$* \ \Phi(t) = E\left(e^{t.\overline{X}}\right) =$$

$$= e^{0.t}.0'25 + e^{0'5.t}.0'20 + e^{1.t}.0'34 + e^{1'5.t}.0'12 + e^{2.t}.0'09 = ...$$

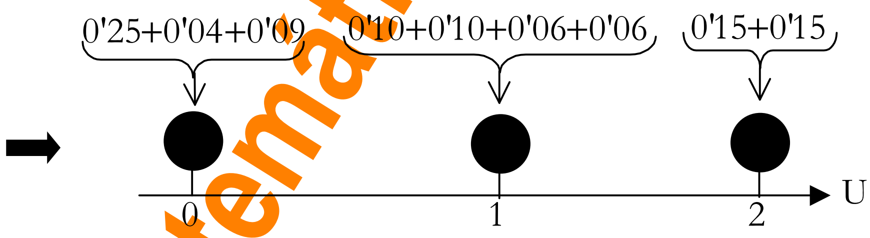
$$* \ P\left(\frac{1}{\overline{X} - 0'7} \le 2\right) = P\left(\overline{X} - 0'7 \ge \frac{1}{2}\right) = P(\overline{X} \ge 1'2) = 0'12 + 0'09$$

¡Que quede claro!

Siendo "X" discreta y muestra de tamaño 2, para calcular la función de cuantía de un estadístico "Pepe", haremos lo que acabamos de hacer con el estadístico media muestral: calcularemos el valor que toma "Pepe" en cada punto de la distribución bidimensional discreta asociada a la variable muestral $(X_1;X_2)$, y después obraremos en consecuencia.

Por ejemplo, para el estadístico "U" tal que $U = |X_1 - X_2|$, es:

$(x_1;x_2)$	(0;0)	(0;1)	(0;2)	(1;0)	(1;1)	(1;2)	(2;0)	(2;1)	(2;2)
$P(x_1;x_2)$	0'25	0'10	0'15	0'10	0'04	0'06	0'15	0'06	0'09
$U = \lvert X_1 - X_2 \rvert$	0	1	2	1	0	1	2	1	0



Una vez conocida la distribución de probabilidad del estadístico "U", puede marearse la perdiz con cualquier puñeta típica de una variable unidimensional:

$$* \ E(U) = 0.0'38 + 1.0'32 + 2.0'30 = ...$$

$$* \ V(U) = E\left(U^2\right) - \left(E(U)^2\right) = ...$$

$$\boxed{E(U^2) = 0^2.0'38 + 1^2.0'32 + 2^2.0'30}$$

$$* \ \Phi(t) = E\left(e^{t.U}\right) = e^{0.t}.0'38 + e^{1.t}.0'32 + e^{2.t}.0'30 = ...$$

$$* \ P(1/(U - 0'7) \le 2) = P(U - 0'7 \ge 1/2) = P(U \ge 1'2) = 0'30$$

FONEMATO 2.4.1

1) Si el 70 % de los estudiantes se chupan el dedo y seleccionamos 100 estudiantes al azar, determínese la probabilidad de que la proporción de estudiantes que se chupan el dedo esté entre 0'65 y 0'8.
2) Determínese un intervalo en el que con probabilidad 0'95 puede esperarse que se encuentre la proporción muestral de chupadedos.

SOLUCIÓN

1) Si chuparse el dedo es un "éxito", la población "X" objeto de estudio tiene distribución de Bernouilli de parámetro 0'7, pues dicen que el 70 % de los estudiantes se lo chupan. Así, siendo $(X_1;\ldots;X_{100})$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño 100, el estadístico media muestral $\overline{X}=(X_1+\ldots+X_{100})/100$ expresa la proporción de estudiantes que se chupan el dedo en la muestra; y como el tamaño muestral es grande, podemos aproximar la distribución de probabilidad de $\overline{X}$ mediante la normal con media y varianza las de $\overline{X}$:

$$E(\overline{X})=E(X)=0'7$$

$$\boxed{X\approx B(p)\Rightarrow E(X)=p \text{ y } V(X)=p.(1-p)}$$

$$V(\overline{X})=\frac{V(X)}{100}=\frac{0'7.0'3}{100}=0'0021$$

Por tanto:

$$P(0'65<\overline{X}<0'8)\cong P\left(0'65<N(0'7;\sqrt{0'0021})<0'8\right)=$$

$$=P\left(N(0;1)<\frac{0'8-0'7}{\sqrt{0'0021}}\right)-P\left(N(0;1)<\frac{0'65-0'7}{\sqrt{0'0021}}\right)=\ldots$$

2) En la distribución de probabilidad de la variable $N(0'7;\sqrt{0'0021})$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, como la $N(0'7;\sqrt{0'0021})$ es simétrica respecto al punto 0'7, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto (puede demostrase que éste es el de menor amplitud). Por tanto, debemos determinar $c>0$ de modo que $P\left(\left|\overline{X}-0'7\right|<c\right)=0'95$:

$$P\left(\left|\overline{X}-0'7\right|<c\right)=0'95\Rightarrow P\left(\left|N(0;1)\right|<\frac{c}{\sqrt{0'0021}}\right)=0'95\Rightarrow$$

$$\Rightarrow 2.P\left(N(0;1)<\frac{c}{\sqrt{0'0021}}\right)-1=0'95\Rightarrow P\left(N(0;1)<\frac{c}{\sqrt{0'0021}}\right)=0'975\Rightarrow$$

$$\Rightarrow\frac{c}{\sqrt{0'0021}}=1'96\Rightarrow c=1'96.\sqrt{0'0021}=0'0898$$

O sea, si tomamos sucesivas muestras de tamaño 100 de $X\approx B(0'7)$, el 95 % de las veces sucederá que el estadístico media muestral $\overline{X}\approx N(0'7;\sqrt{0'0021})$ toma un valor en el intervalo $(0'7-0'0898;0'7+0'0898)$.

FONEMATO 2.4.2

De una población "X" se toma muestra aleatoria simple de tamaño 81. En los siguientes casos, determínese un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre el estadístico media muestral.

1) Población de Poisson de parámetro 5.
2) Población $B(4;0'3)$.
3) Población $Exp.(0'02)$.
4) Población $U(-1;1)$.
5) Población con densidad $f(x) = 2.x,\ 0 \leq x \leq 1$.
6) Población geométrica de parámetro 0'1.
7) Población $G(20;2)$.

SOLUCIÓN

Como es sabido, en el muestreo aleatorio simple de tamaño $n \geq 30$ de una población "X" con media μ y varianza σ^2, Lindeberg-Levy permiten aproximar la distribución de probabilidad del estadístico media muestral $\overline{X}$ mediante la normal con media y varianza las de $\overline{X}$; o sea, mediante la $N(\mu;\sigma/\sqrt{n})$.

En la distribución de probabilidad de la variable $N(\mu;\sigma/\sqrt{n})$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, como la $N(\mu;\sigma/\sqrt{n})$ es simétrica respecto al punto μ, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto (puede demostrase que éste es el de menor amplitud).

Por tanto, debemos determinar $c > 0$ de modo que $P(|\overline{X} - \mu| < c) = 0'95$:

$$P(|\overline{X} - \mu| < c) = 0'95 \Rightarrow P\left(|N(0;1)| < \frac{c}{\sigma/\sqrt{n}}\right) = 0'95 \Rightarrow$$

$$\Rightarrow 2.P\left(N(0;1) < \frac{c}{\sigma/\sqrt{n}}\right) - 1 = 0'95 \Rightarrow P\left(N(0;1) < \frac{c}{\sigma/\sqrt{n}}\right) = 0'975 \Rightarrow$$

$$\Rightarrow \frac{c}{\sigma/\sqrt{n}} = 1'96 \Rightarrow c = 0'217.\sigma$$

$$\boxed{\text{si } n = 81}$$

1) Si $X \approx P(5)$ es $E(X) = \mu = 5$ y $V(X) = \sigma^2 = 5$.
   Así, es $P(|\overline{X} - 5| < c) = 0'99$ si $c = 0'217.\sqrt{5} = 0'48$; o sea, tomando sucesivas muestras de tamaño 81 de $X \approx P(5)$, el 95 % de las veces sucederá que la media muestral toma un valor en el intervalo $(5 - 0'48; 5 + 0'48)$.

2) Si $X \approx B(4;0'3)$ es $E(X) = \mu = 4.0'3 = 1'2$ y $V(X) = \sigma^2 = 4.0'3.0'7 = 0'84$.
   Así, es $P(|\overline{X} - 1'2| < c) = 0'99$ si $c = 0'217.\sqrt{0'84} = 0'19$; o sea, tomando sucesivas muestras de tamaño 81 de $X \approx B(4;0'3)$, el 95 % de las veces sucederá

que la media muestral $\overline{X} \approx N(1'2;\sqrt{0'84}/9)$ se concreta en algún punto del intervalo $(1'2-0'19;1'2+0'19)$.

3) Si $X \approx Exp.(0'02)$ es $E(X)=\mu=1/0'02=50$ y $V(X)=\sigma^2=1/0'02^2=2500$.
Así, es $P(|\overline{X}-50|<c)=0'99$ si $c=0'217.\sqrt{2500}=10'95$; o sea, si tomamos sucesivas muestras de tamaño 81 de $X \approx Exp.(0'02)$, el 95 % de las veces sucederá que la media muestral $\overline{X} \approx N(50;\sqrt{2500}/9)$ toma un valor en el intervalo $(50-10'95;50+10'95)$.

4) Si $X \approx U(-1;1)$ es $E(X)=\mu=(-1+1)/2=0$ y $V(X)=\sigma^2=(1-(-1))^2/12=1/3$.
Así, es $P(|\overline{X}-0|<c)=0'99$ si $c=0'217.\sqrt{1/3}=0'12$; o sea, si tomamos sucesivas muestras de tamaño 81 de $X \approx U(-1;1)$, el 95 % de las veces sucederá que la media muestral $\overline{X} \approx N(0;\sqrt{1/243})$ toma un valor en el intervalo $(-0'12;0'12)$.

5) Si la densidad de la población "X" es $f(x)=2.x$, $0 \le x \le 1$, entonces:

$$E(X)=\int_0^1 x.f(x).dx=\int_0^1 2.x^2.dx=2/3$$

$$V(X)=\sigma^2=E(X^2)-(E(X))^2=1/18$$

$$\boxed{E(X^2)=\int_0^1 x^2.f(x).dx=\int_0^1 2.x^3.dx=1/2}$$

Así, es $P(|\overline{X}-2/3|<c)=0'99$ si $c=0'217.\sqrt{1/18}=0'05$; o sea, si tomamos sucesivas muestras de tamaño 81 de "X", el 95 % de las veces sucederá que la media muestral $\overline{X} \approx N(2/3;\sqrt{1/1458})$ se concreta en algún punto del intervalo $((2/3)-0'05;(2/3)+0'05)$.

6) Si $X \approx G(0'1)$ es $E(X)=\mu=(1-0'1)/0'1=9$ y $V(X)=\sigma^2=(1-0'1)/0'1^2=90$.
Así, es $P(|\overline{X}-9|<c)=0'99$ si $c=0'217.\sqrt{90}=2'05$; o sea, tomando sucesivas muestras de tamaño 81 de $X \approx G(0'1)$, el 95 % de las veces sucederá que la media muestral toma un valor en el intervalo $(9-2'05;9+2'05)$.

7) Si $X \approx G(20;2)$ es $E(X)=\mu=20/2=10$ y $V(X)=\sigma^2=20/2^2=5$.
Así, es $P(|\overline{X}-10|<c)=0'99$ si $c=0'217.\sqrt{5}=0'48$; o sea, tomando sucesivas muestras de tamaño 81 de $X \approx G(20;2)$, el 95 % de las veces sucederá que $\overline{X} \approx N(10;\sqrt{5}/9)$ toma un valor en el intervalo $(10-0'48;10+0'48)$.

FONEMATO 2.4.3

La producción diaria de tomate de una empresa oscila entre 6000 y 10000 kg.

Determínese la probabilidad de que la producción media durante 320 días supere los 8000 kg.

SOLUCIÓN

Supuesto que la variable "X" que expresa la producción diaria tiene distribución $U(6000;10000)$, y siendo $(X_1;\ldots;X_{320})$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s. de tamaño 320, el estadístico media muestral es $\overline{X}=(X_1+\ldots+X_{320})/320$.

Como el tamaño muestral es grande, podemos aproximar la distribución de probabilidad de $\overline{X}$ mediante la normal con media y varianza las de $\overline{X}$, que son:

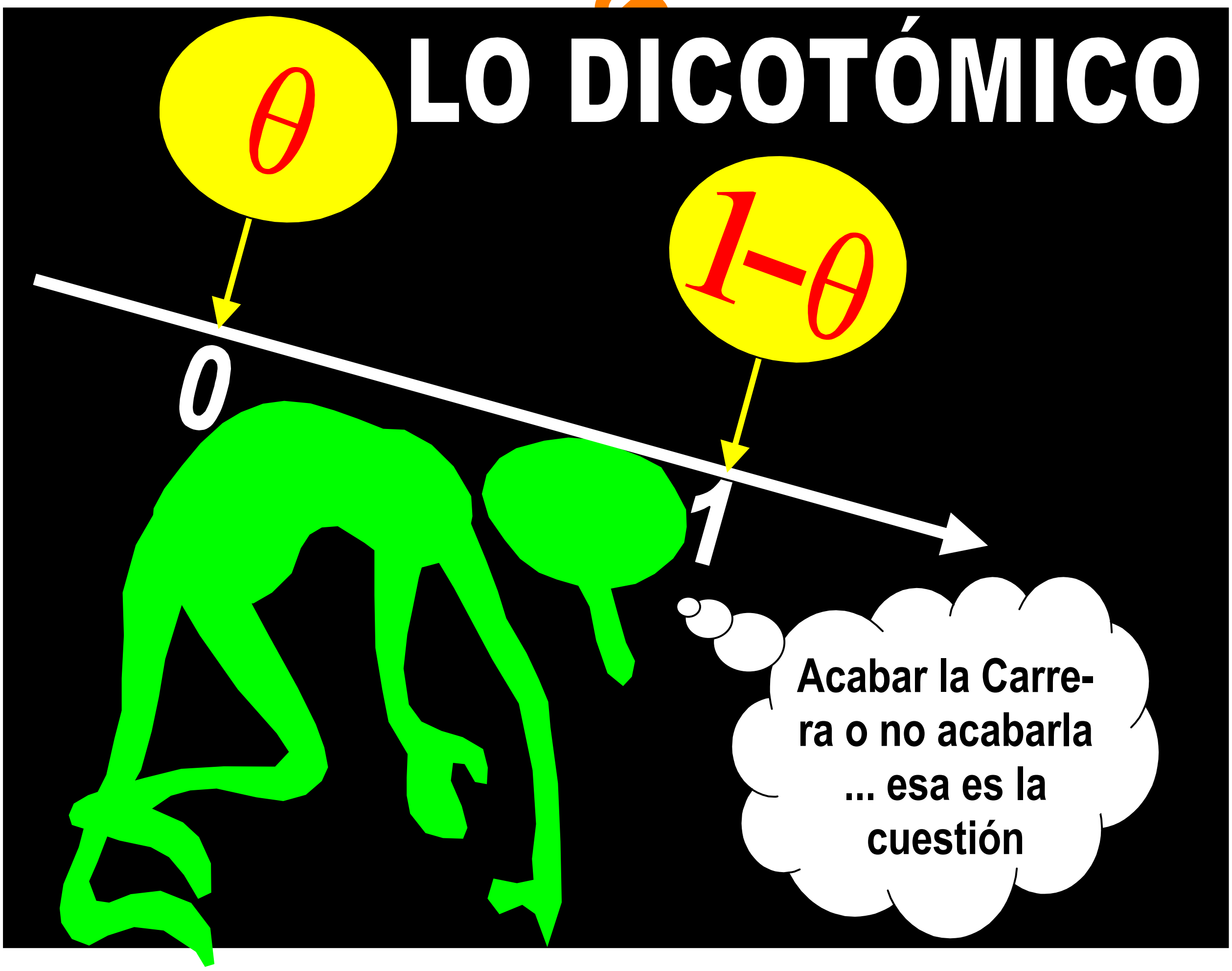
$$E(\overline{X})=E(X)=\frac{6000+10000}{2}=8000$$

$$\boxed{X\approx U(a;b)\Rightarrow E(X)=\frac{a+b}{2}\ \text{ y }\ V(X)=\frac{(b-a)^2}{12}}$$

$$V(\overline{X})=\frac{V(X)}{320}=\frac{4000^2}{12}$$

Por tanto:

$$P(\overline{X}>8000)\cong P\left(N(8000;4000/\sqrt{12})>8000\right)=P\left(N(0;1)>0\right)=0'5$$



2.5 LA VARIANZA MUESTRAL

Siendo S^2 el estadístico varianza muestral en el muestreo aleatorio simple de tamaño "n" de una población "X" de media μ y varianza σ^2, es:

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

$$\boxed{\begin{aligned}&\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i + (\mu-\mu) - \overline{X})^2 = \sum_{i=1}^{n}((X_i-\mu)-(\overline{X}-\mu))^2 = \\ &= \sum_{i=1}^{n}\left((X_i-\mu)^2 - 2.(X_i-\mu).(\overline{X}-\mu) + (\overline{X}-\mu)^2\right) = \\ &= \left(\sum_{i=1}^{n}(X_i-\mu)^2\right) - 2.(\overline{X}-\mu).\left(\sum_{i=1}^{n}(X_i-\mu)\right) + n.(\overline{X}-\mu)^2 = \\ &\boxed{\sum_{i=1}^{n}(X_i-\mu) = \left(\sum_{i=1}^{n}X_i\right) - n.\mu = n.\overline{X} - n.\mu \qquad \boxed{\sum_{i=1}^{n}X_i = n.\overline{X}}} \\ &= \left(\sum_{i=1}^{n}(X_i-\mu)^2\right) - 2.(\overline{X}-\mu).(n.\overline{X}-n.\mu) + n.(\overline{X}-\mu)^2 = \\ &= \left(\sum_{i=1}^{n}(X_i-\mu)^2\right) - n.(\overline{X}-\mu)^2\end{aligned}}$$

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

$$\boxed{\begin{aligned}&* \; E\left(\sum_{i=1}^{n}(X_i-\mu)^2\right) = \sum_{i=1}^{n}E\left((X_i-\mu)^2\right) = \sum_{i=1}^{n}V(X_i) = n.\sigma^2 \\ &\boxed{E(X_i)=\mu \Rightarrow E\left((X_i-\mu)^2\right) = V(X_i),\ \forall\, i=1,2,\dots,n} \\ &* \; E\left((\overline{X}-\mu)^2\right) = E\left((\overline{X}-E(\overline{X}))^2\right) = V(\overline{X}) = \sigma^2/n \\ &\boxed{\mu = E(\overline{X})}\end{aligned}}$$

Observa: siendo $S_c^2 = n.S^2/(n-1)$ el estadístico cuasivarianza muestral, es:

$$E(S_c^2) = E\left(\frac{n}{n-1}.S^2\right) = \frac{n}{n-1}.E(S^2) = \frac{n}{n-1}.\frac{n-1}{n}.\sigma^2 = \sigma^2$$

En el Tema 3, para expresar que el valor esperado del estadístico cuasivarianza muestral S_c^2 siempre coincide con la varianza σ^2 de la población "X", diremos que S_c^2 es un **estimador insesgado** de la varianza poblacional σ^2. Como $E(S^2) = (n-1).\sigma^2/n \neq \sigma^2$, el estadístico varianza muestral S^2 no es un estimador insesgado de σ^2; diremos que S^2 es un estimador **asintóticamente insesgado** de σ^2 para expresar que $\lim\limits_{n\to\infty} E(S^2) = \lim\limits_{n\to\infty} (n-1).\sigma^2/n = \sigma^2$.

FONEMATO 2.5.1

Determínese la distribución de probabilidad de la varianza muestral en el muestreo aleatorio simple de tamaño 2 de una población $Z \approx \text{Exp.}(1)$.

SOLUCIÓN

Siendo $Z \approx \text{Exp.}(1)$, su función de densidad es $f(z) = e^{-z}$, $z > 0$. La función de densidad "g" de la variable muestral $(X;Y)$ que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño 2 de la población $Z \approx \text{Exp.}(1)$ es el producto de las densidades de "X" e "Y", pues éstas son independientes:

$$g(x;y) = f(x).f(y) = e^{-(x+y)},\ x > 0,\ y > 0$$

$$\boxed{\text{pues } X \approx \text{Exp.}(1) \text{ e } Y \approx \text{Exp.}(1)}$$

Siendo $U = (X+Y)/2$ la media muestral, el estadístico varianza muestral "T" es:

$$T = \frac{1}{2}.\left((X-U)^2 + (Y-U)^2\right) = \frac{1}{2}.\left(\left(X - \frac{X+Y}{2}\right)^2 + \left(Y - \frac{X+Y}{2}\right)^2\right) =$$

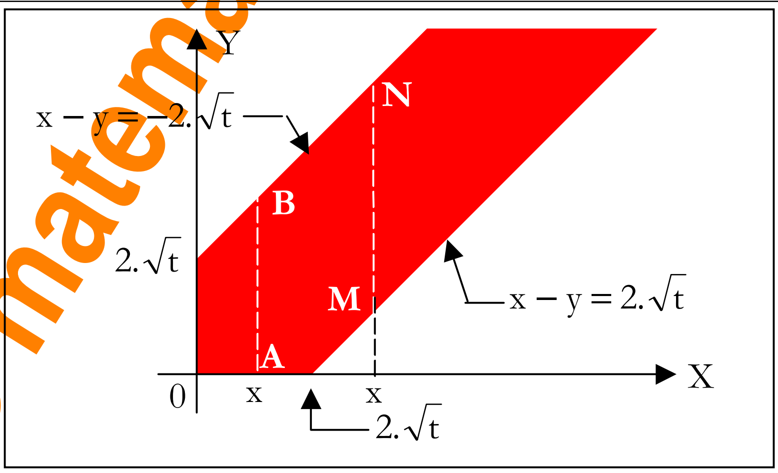
$$= \frac{1}{2}.\left(\left(\frac{X-Y}{2}\right)^2 + \left(\frac{Y-X}{2}\right)^2\right) = \frac{1}{4}.(X-Y)^2$$

- Función de distribución de "T":

$$F_T(t) = P(T \leq t) = P\left(\frac{1}{4}.(X-Y)^2 \leq t\right) = P\left((X-Y)^2 \leq 4.t\right) =$$

$$= P\left(-2.\sqrt{t} \leq X - Y \leq 2.\sqrt{t}\right) = \iint_D g(x;y).dx.dy =$$

$$\boxed{\begin{array}{c}\text{siendo "g" la función de densidad de la variable muestral } (X;Y) \text{ y} \\ D = \{(x;y) \in \Re^2 \ /\ -2.\sqrt{t} \leq x - y \leq 2.\sqrt{t},\ x > 0,\ y > 0\}\end{array}}$$



$$= \int_{x=0}^{x=2.\sqrt{t}} \left(\int_{y_A=0}^{y_B=x+2.\sqrt{t}} e^{-(x+y)}.dy\right).dx + \int_{x=2.\sqrt{t}}^{x=+\infty} \left(\int_{y_M=x-2.\sqrt{t}}^{y_N=x+2.\sqrt{t}} e^{-(x+y)}.dy\right).dx =$$

$$= \int_{x=0}^{x=2.\sqrt{t}} e^{-x}.\left(-e^{-y}\right)_{y=0}^{y=x+2.\sqrt{t}}.dx + \int_{x=2.\sqrt{t}}^{x=+\infty} e^{-x}.\left(-e^{-y}\right)_{y=x-2.\sqrt{t}}^{y=x+2.\sqrt{t}}.dx =$$

$$= \int_{x=0}^{x=2.\sqrt{t}} e^{-x}.\left(1-e^{-x-2.\sqrt{t}}\right).dx + \int_{x=2.\sqrt{t}}^{x=+\infty} e^{-x}.\left(e^{-x+2.\sqrt{t}} - e^{-x-2.\sqrt{t}}\right).dx =$$

$$= \int_{x=0}^{x=2.\sqrt{t}} e^{-x}.dx - \int_{x=0}^{x=2.\sqrt{t}} e^{-2.x-2.\sqrt{t}}.dx$$

$$+\int_{x=2.\sqrt{t}}^{x=+\infty} e^{-2.x+2.\sqrt{t}}.dx - \int_{x=2.\sqrt{t}}^{x=+\infty} e^{-2.x-2.\sqrt{t}}.dx =$$

$$= \left(-e^{-x}\right)_0^{2.\sqrt{t}} + \frac{1}{2}.\left(e^{-2.x-2.\sqrt{t}}\right)_0^{2.\sqrt{t}} -$$

$$-\frac{1}{2}.\left(e^{-2.x+2.\sqrt{t}}\right)_{2.\sqrt{t}}^{+\infty} + \frac{1}{2}.\left(e^{-2.x-2.\sqrt{t}}\right)_{2.\sqrt{t}}^{+\infty} =$$

$$= \left(1-e^{-2.\sqrt{t}}\right) + \frac{e^{-6.\sqrt{t}} - e^{-2.\sqrt{t}}}{2} - \frac{e^{-\infty} - e^{-2.\sqrt{t}}}{2} + \frac{e^{-\infty} - e^{-6.\sqrt{t}}}{2} =$$

$$= 1 - e^{-2.\sqrt{t}}, \ t > 0$$

- Función de densidad de "T": $f_T(t) = dF_T(t)/dt = \dfrac{1}{\sqrt{t}}.e^{-2.\sqrt{t}}, \ t > 0$

FONEMATO 2.5.2

Determínese la distribución de probabilidad de la varianza muestral en el muestreo aleatorio simple de tamaño 2 de una población $Z \approx U(0;1)$.

SOLUCIÓN

Siendo $Z \approx U(0;1)$, su función de densidad es $f(z) = 1, \ z \in (0;1)$. La función de densidad "g" de la variable muestral $(X;Y)$ que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño 2 de la población $Z \approx U(0;1)$ es el producto de las densidades de "X" e "Y", pues éstas son independientes:

$$g(x;y) \underset{\boxed{\text{pues } X \approx U(0;1) \text{ e } Y \approx U(0;1)}}{=} f(x).f(y) = 1, \ x \in (0;1), \ y \in (0;1)$$

Siendo $U = (X+Y)/2$ la media muestral, el estadístico varianza muestral "T" es:

$$T = \frac{1}{2}.\left((X-U)^2 + (Y-U)^2\right) = \frac{1}{2}.\left(\left(X - \frac{X+Y}{2}\right)^2 + \left(Y - \frac{X+Y}{2}\right)^2\right) =$$

$$= \frac{1}{2}.\left(\left(\frac{X-Y}{2}\right)^2 + \left(\frac{Y-X}{2}\right)^2\right) = \frac{1}{4}.(X-Y)^2$$

- Función de distribución de "T":

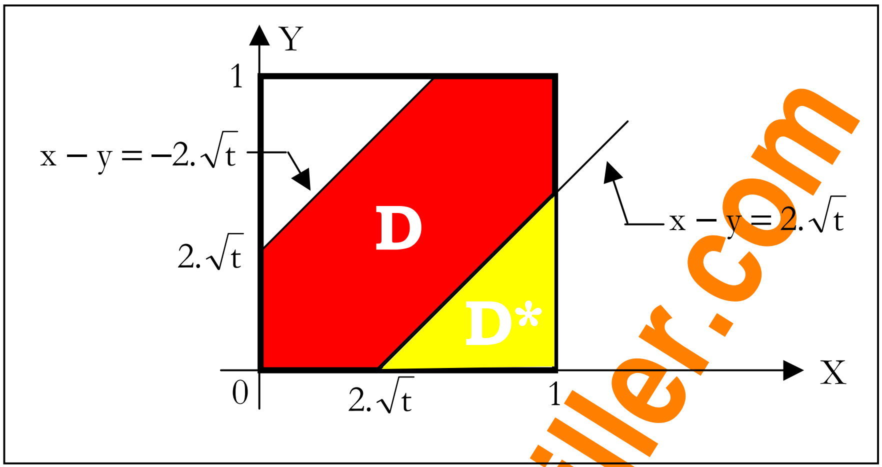
$$F_T(t) = P(T \le t) = P\left(\frac{1}{4}.(X-Y)^2 \le t\right) = P\left((X-Y)^2 \le 4.t\right) =$$

$$= P\left(-2.\sqrt{t} \le X - Y \le 2.\sqrt{t}\right) \underset{\uparrow}{=} \iint_D g(x;y).dx.dy =$$

$$\boxed{\begin{array}{c}\text{siendo "g" la función de densidad de la variable muestral } (X;Y) \text{ y} \\ D = \{(x;y) \in \Re^2 \ / \ -2.\sqrt{t} \le x - y \le 2.\sqrt{t}, \ 0 < x < 1, \ 0 < y < 1\}\end{array}}$$

siempre hay algún capullito de alhelí que no se da cuenta de que al ser constante la densidad g(x ; y) es innecesario lidiar la integral doble, pues el asunto se reduce a calcular áreas

$$= 1 - 2.(\text{Área de } D^*) = 1 - 2.\frac{(1-2.\sqrt{t})^2}{2} = 4.\sqrt{t} - 4.t$$



- Función de densidad de "T":

$$f_T(t) = \frac{dF_T(t)}{dt} = \frac{2}{\sqrt{t}} - 4, \quad 0 < t < 1/4$$

2.6 MUESTREO DE POBLACIÓN NORMAL

Siendo $X \approx N(\mu;\sigma)$, sea $(X_1, \ldots ; X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s. simple de tamaño "n".

1) Distribución de la media muestral si se conoce σ

Siendo $\overline{X} = (X_1 + \ldots + X_n)/n$ una combinación lineal de variables normales independientes, la distribución de $\overline{X}$ es normal con media y varianza las de $\overline{X}$; es decir, $\overline{X}$ tiene distribución $N(\mu;\sigma/\sqrt{n})$. Por tanto, la f.g.m. de $\overline{X}$ es

$$\Phi(t) = e^{\mu.t + \frac{\sigma^2.t^2}{2.n}}$$

2) Teorema de Fisher

En el muestreo aleatorio simple de una variable $X \approx N(\mu;\sigma)$, los estadísticos (variables aleatorias unidimensionales) media muestral $\overline{X}$ y varianza muestral S^2 son independientes. En efecto, la f.g.m. de la variable (n+1)-dimensional $(\overline{X};X_1 - \overline{X}; \ldots ;X_n - \overline{X})$ es:

$$\Phi(t;t_1;\ldots;t_n) = E\left(\exp\left(t.\overline{X} + t_1.(X_1 - \overline{X}) + \ldots + t_1.(X_n - \overline{X})\right)\right) =$$

$$= E\left(\exp\left(t.\overline{X} + \sum_{i=1}^{n} t_i.(X_i - \overline{X})\right)\right) =$$

$$= E\left(\exp\left(\frac{1}{n}.\sum_{i=1}^{n} t.X_i + \sum_{i=1}^{n} t_i.(X_i - \overline{X})\right)\right) =$$

$$= E\left(\exp\left(\sum_{i=1}^{n} X_i.\left(\frac{t}{n} + t_i - \frac{1}{n}.\sum_{i=1}^{n} t_i\right)\right)\right) =$$

$$= E\left(\exp\left(\sum_{i=1}^{n} X_i.\left(t_i - \frac{t_1 + t_2 + \ldots + t_n - t}{n}\right)\right)\right) =$$

por comodidad en la notación, hacemos $\delta = \frac{1}{n}.\sum_{i=1}^{n} t_i$

$$= E\left(\exp\left(\sum_{i=1}^{n} X_i.\left(t_i - \frac{n.\delta - t}{n}\right)\right)\right) = E\left(\prod_{i=1}^{n} \exp\left(X_i.\left(t_i - \frac{n.\delta - t}{n}\right)\right)\right) =$$

pues $E(\exp(Pepe.X_i)) = \Phi_i(Pepe)$, siendo Φ_i la f.g.m. de X_i

$$= \prod_{i=1}^{n} E\left(\exp\left(X_i.\left(t_i - \frac{n.\delta - t}{n}\right)\right)\right) = \prod_{i=1}^{n} \Phi_i\left(t_i - \frac{n.\delta - t}{n}\right) =$$

la f.g.m de la variable $X_i \approx N(\mu;\sigma)$ es $\Phi_i(\lambda) = \exp\left(\mu.\lambda + \frac{1}{2}.\lambda^2.\sigma^2\right)$

$$=\prod_{i=1}^{n} \exp\left(\left(t_i - \frac{n.\delta - t}{n}\right).\mu + \frac{1}{2}.\left(t_i - \frac{n.\delta - t}{n}\right)^2.\sigma^2\right) =$$

$$= \exp\left(\left(t_i - \frac{n.\delta - t}{n}\right).\mu + \frac{1}{2}.\sigma^2.\sum_{i=1}^{n}\left(t_i - \frac{n.\delta - t}{n}\right)^2\right) =$$

$$= \exp\left(\left(\left(\sum_{i=1}^{n} t_i\right) - n.\delta + t\right).\mu + \frac{1}{2}.\sigma^2.\sum_{i=1}^{n}\left(t_i - \frac{n.\delta - t}{n}\right)^2\right) =$$

$$\boxed{\delta = \frac{1}{n}.\sum_{i=1}^{n} t_i \Rightarrow \sum_{i=1}^{n} t_i = \delta.n}$$

$$= \exp\left(\mu.t + \frac{\sigma^2}{2.n^2}.\sum_{i=1}^{n}\left(t + n.(t_i - \delta)\right)^2\right) =$$

$$= \exp\left(\mu.t + \frac{\sigma^2}{2.n^2}.\sum_{i=1}^{n}\left(t^2 + n^2.(t_i - \delta)^2 + 2.t.n.(t_i - \delta)\right)\right) =$$

$$= \exp\left(\mu.t + \frac{\sigma^2}{2.n^2}.\left(n.t^2 + n^2.\left(\sum_{i=1}^{n}(t_i - \delta)^2\right) + 2.t.n.\left(\sum_{i=1}^{n}(t_i - \delta)\right)\right)\right) =$$

$$= \exp\left(\mu.t + \frac{\sigma^2}{2.n^2}.\left(n.t^2 + n^2.\left(\sum_{i=1}^{n}(t_i - \delta)^2\right) + 2.t.n.\left(\left(\sum_{i=1}^{n} t_i\right) - n.\delta\right)\right)\right) =$$

$$\boxed{\delta = \frac{1}{n}.\sum_{i=1}^{n} t_i \Rightarrow \left(\sum_{i=1}^{n} t_i\right) - n.\delta = 0}$$

$$= \exp\left(\mu.t + \frac{\sigma^2}{2.n^2}.\left(n.t^2 + n^2.\left(\sum_{i=1}^{n}(t_i - \delta)^2\right)\right)\right) =$$

$$= \exp\left(\mu.t + \frac{1}{2}.\frac{\sigma^2}{n}.t^2\right).\exp\left(\frac{\sigma^2}{2}.\left(\sum_{i=1}^{n}(t_i - \delta)^2\right)\right) = \Phi_1(t).\Phi_2(t_1; .... ; t_n)$$

$$\boxed{\Phi_1(t) = \exp\left(\mu.t + \frac{1}{2}.\frac{\sigma^2}{n}.t^2\right) \; ; \; \Phi_2(t_1; .... ; t_n) = \exp\left(\frac{\sigma^2}{2}.\left(\sum_{i=1}^{n}(t_i - \delta)^2\right)\right)}$$

Como $\Phi_1(t)$ es la f.g.m de $\overline{X} \approx N(\mu; \sigma/\sqrt{n})$, entonces $\Phi_2(t_1; .... ; t_n)$ es necesariamente la f.g.m de la variable n-dimensional $(X_1 - \overline{X}; .... ; X_1 - \overline{X})$. Como la f.g.m. de la variable $(\overline{X}; X_1 - \overline{X}; .... ; X_n - \overline{X})$ es el producto de las respectivas f.g.m de las variables $\overline{X}$ y $(X_1 - \overline{X}; .... ; X_n - \overline{X})$, éstas son independientes; por tanto, también son independientes $\overline{X}$ y S^2, siendo:

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

Nadie nace sabiendo de números ..... y nadie espabila sin sentirse gilipollas multitud de veces

3) Distribución de $n.S^2/\sigma^2$

En el muestreo de tamaño "n" de una variable $X \approx N(\mu;\sigma)$, el estadístico $n.S^2/\sigma^2$ tiene distribución χ^2_{n-1}. En efecto, al calcular la varianza del estadístico S^2 vimos que:

$$\sum_{i=1}^{n}(X_i-\overline{X})^2=\left(\sum_{i=1}^{n}(X_i-\mu)^2\right)-n.(\overline{X}-\mu)^2$$

Por tanto, es $\displaystyle\sum_{i=1}^{n}(X_i-\mu)^2=\left(\sum_{i=1}^{n}(X_i-\overline{X})^2\right)+n.(\overline{X}-\mu)^2$; así:

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\left(\frac{1}{\sigma^2}.\sum_{i=1}^{n}(X_i-\overline{X})^2\right)+\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2\Rightarrow$$

$$\boxed{S^2=\frac{1}{n}.\sum_{i=1}^{n}(X_i-\overline{X})^2\Rightarrow\sum_{i=1}^{n}(X_i-\overline{X})^2=n.S^2}$$

$$\Rightarrow\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\frac{n.S^2}{\sigma^2}+\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2\Rightarrow\frac{n.S^2}{\sigma^2}\approx\chi^2_{n-1}$$

$$* \; X_i\approx N(\mu;\sigma)\Rightarrow\frac{X_i-\mu}{\sigma}\approx N(0;1)\Rightarrow\left(\frac{X_i-\mu}{\sigma}\right)^2\approx\chi^2_1\Rightarrow$$

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

pues $X_1,\ldots,X_n$ son independientes

$$* \; \overline{X}\approx N(\mu;\sigma/\sqrt{n})\Rightarrow\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\approx N(0;1)\Rightarrow\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2\approx\chi^2_1$$

* Como $\dfrac{n.S^2}{\sigma^2}$ y $\left(\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2\approx\chi^2_1$ son independientes (por serlo S^2 y $\overline{X}$), si su suma tiene distribución χ^2_n, ha de ser $\dfrac{n.S^2}{\sigma^2}\approx\chi^2_{n-1}$

4) Distribución de la media muestral si se desconoce σ

Siendo independientes $\overline{X} \approx N(\mu;\sigma/\sqrt{n})$ y S^2, también son independientes las variables $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx N(0;1)$ y $\dfrac{n.S^2}{\sigma^2} \approx \chi^2_{n-1}$; por tanto, recordando que la variable t-Student con "k" grados de libertad es el cociente entre una variable $Pepa \approx N(0;1)$ y la raíz cuadrada de una variable $Juana \approx \chi^2_k$ dividida por sus grados de libertad (siendo independientes "Pepa" y "Juana"), resulta que:

$$Z = \frac{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{n.S^2/\sigma^2}{n-1}}} = \frac{\overline{X}-\mu}{S/\sqrt{n-1}} \approx t_{n-1}$$

Siendo $S_c^2 = n.S^2/(n-1)$, es $S/\sqrt{n-1} = S_c/\sqrt{n}$; por tanto, también es:

$$Z = \frac{\overline{X}-\mu}{S_c/\sqrt{n}} \approx t_{n-1}$$

5) Distribución de la diferencia entre las medias muestrales de sendas poblaciones normales

Considera que de las variables independientes $N(\mu_1;\sigma_1)$ y $N(\mu_2;\sigma_2)$ se toman muestras de tamaños respectivos "n" y "m", siendo $\overline{X}_1$ y $\overline{X}_2$ las respectivas medias muestrales y S_1^2 y S_2^2 las varianzas muestrales

a) Varianzas poblacionales conocidas

Siendo $\overline{X}_1 \approx N(\mu_1;\sigma_1/\sqrt{n})$ y $\overline{X}_2 \approx N(\mu_2;\sigma_2/\sqrt{m})$, es:

$$\overline{X}_1 - \overline{X}_2 \approx N\left(\mu_1-\mu_2;\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)$$

b) Varianzas poblacionales desconocidas pero iguales

Siendo $\overline{X}_1 \approx N(\mu_1;\sigma/\sqrt{n})$ y $\overline{X}_2 \approx N(\mu_2;\sigma/\sqrt{m})$, es:

$$\overline{X}_1 - \overline{X}_2 \approx N\left(\mu_1-\mu_2;\sigma.\sqrt{\frac{1}{n}+\frac{1}{m}}\right)$$

Además $n.S_1^2/\sigma^2 \approx \chi^2_{n-1}$ y $m.S_2^2/\sigma^2 \approx \chi^2_{m-1}$, y siendo éstas independientes, es:

$$\frac{n.S_1^2}{\sigma^2}+\frac{m.S_2^2}{\sigma^2} = \frac{n.S_1^2+m.S_2^2}{\sigma^2} \approx \chi^2_{n+m-2}$$

Recordando que la variable t-Student con "k" grados de libertad es el cociente entre una variable $\text{Pepa} \approx N(0;1)$ y la raíz cuadrada de una variable $\text{Juana} \approx \chi^2_k$ dividida por sus grados de libertad (siendo independientes "Pepa" y "Juana"), si consideramos que

$$\text{Pepa} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma.\sqrt{\frac{1}{n} + \frac{1}{m}}} \approx N(0;1)$$

$$\text{Juana} = \frac{n.S_1^2 + m.S_2^2}{\sigma^2} \approx \chi^2_{n+m-2}$$

es:

$$Z = \frac{\dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma.\sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\dfrac{(n.S_1^2 + m.S_2^2)/\sigma^2}{n+m-2}}} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{n.S_1^2 + m.S_2^2}{n+m-2}}.\sqrt{\frac{1}{n} + \frac{1}{m}}} \approx t_{n+m-2}$$

c) Varianzas poblacionales desconocidas y distintas, con muestras grandes

Si las muestras son grandes (tamaño ≥ 30), las varianzas muestrales son muy buenas estimaciones de las varianzas poblacionales; por tanto, considerando que $\sigma_1^2 = S_1^2$ y que $\sigma_2^2 = S_2^2$, sucede que:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n} + \dfrac{S_2^2}{m}}} \approx N(0;1)$$

d) Varianzas poblacionales desconocidas y distintas, con muestras pequeñas

Para muestra pequeñas, es:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n} + \dfrac{S_2^2}{m}}} \approx t_{n*}$$

donde $n*$ es el entero más cercano a la solución de la ecuación

$$\frac{1}{n*} = \frac{c^2}{n-1} + \frac{(1-c)^2}{m-1}$$

siendo:

$$c = \frac{S_1^2/(n-1)}{\dfrac{S_1^2}{n-1} + \dfrac{S_2^2}{m-1}}$$

6) Distribución del cociente entre las varianzas muestrales de sendas poblaciones normales

Sean las variables independientes $X \approx N(\mu_1;\sigma_1)$ e $Y \approx N(\mu_2;\sigma_2)$ de las que se toman muestras de tamaños respectivos "n" y "m", siendo S_1^2 y S_2^2 las respectivas varianzas muestrales.

Recordando que la variable F-Snedecor con r;s grados de libertad es el cociente entre una variable $Pepa \approx \chi_r^2$ y una variable $Juana \approx \chi_s^2$ dividida cada una por sus grados de libertad (siendo independientes "Pepa" y "Juana"), si consideramos que

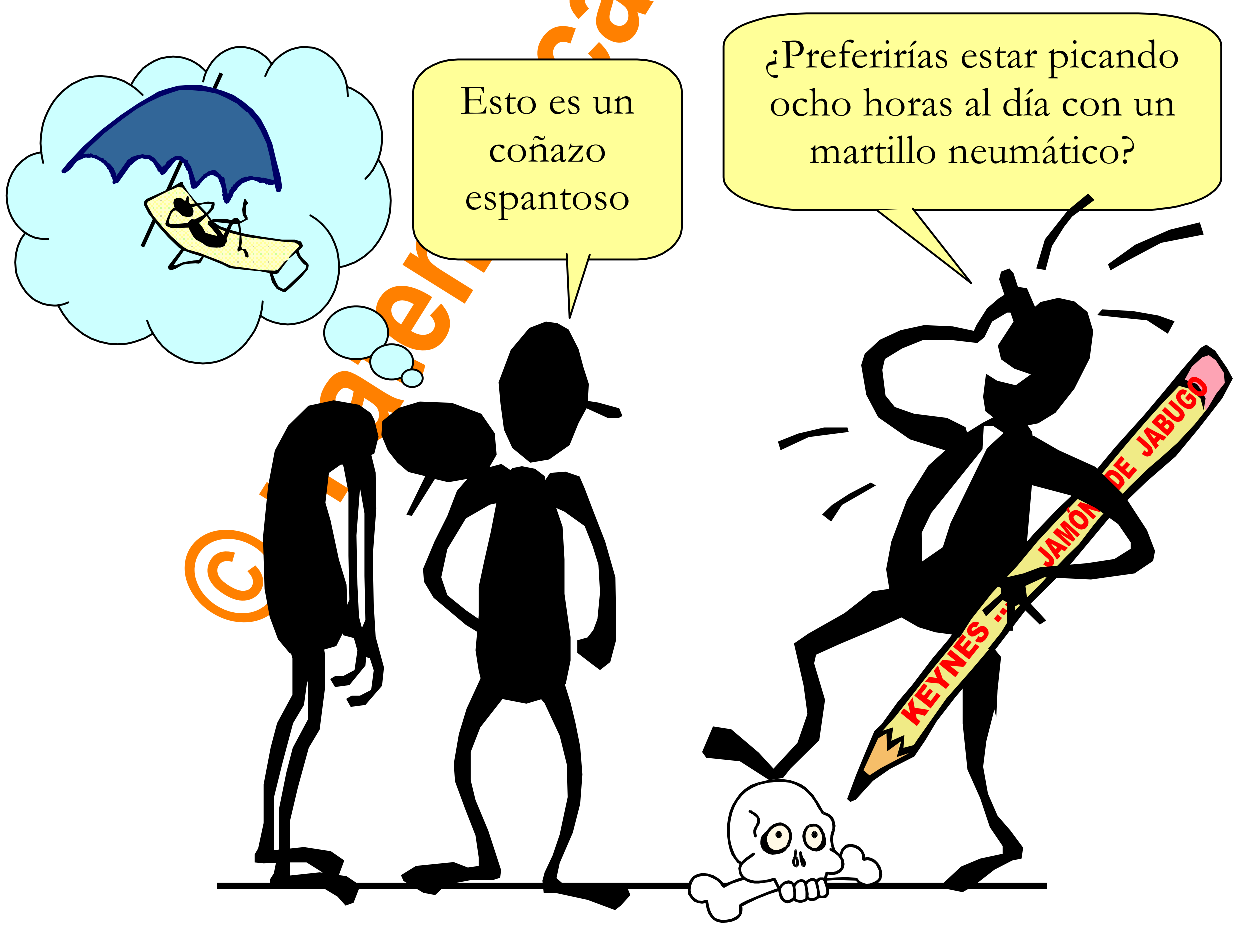
$$Pepa = n.S_1^2/\sigma_1^2 \approx \chi_{n-1}^2 \ ; \ Juana = m.S_2^2/\sigma_2^2 \approx \chi_{m-1}^2$$

entonces:

$$Z = \frac{\dfrac{n.S_1^2/\sigma_1^2}{n-1}}{\dfrac{m.S_2^2/\sigma_2^2}{m-1}} = \frac{n.(m-1).\sigma_2^2.S_1^2}{m.(n-1).\sigma_1^2.S_2^2} \approx F_{n-1;m-1} \Rightarrow$$

$$\boxed{S_{c1}^2 = \frac{n.S_1^2}{n-1} \ ; \ S_{c2}^2 = \frac{m.S_2^2}{m-1}}$$

$$\Rightarrow Z = \frac{\sigma_2^2.S_{c1}^2}{\sigma_1^2.S_{c2}^2} \approx F_{n-1;m-1}$$



FONEMATO 2.6.1

Una máquina de refrescos está regulada de modo que la cantidad de bebida que suministra tiene distribución normal con media 22'5 dl. y desviación típica 1'5 dl.

Determínese la probabilidad de que al obtener una muestra de 36 bebidas, el contenido medio supere los 24 dl.

SOLUCIÓN

Si $X \approx N(22'5;1'5)$ expresa la cantidad de bebida que suministra la máquina, siendo $(X_1;....;X_{36})$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño 36, el estadístico media muestral es $\overline{X} = (X_1 + ... + X_{36})/36$, que tiene distribución normal, por ser combinación lineal de normales independientes, siendo:

$$E(\overline{X}) = E(X) = 22'5 \;;\; V(\overline{X}) = V(X)/36 = 1'5^2/36$$

Por tanto:

$$P(\overline{X} > 24) = P\big(N(22'5;1'6/6) > 24\big) = P\left(N(0;1) > \frac{24 - 22'5}{1'6/6}\right) = ...$$

FONEMATO 2.6.2

De una población normal de varianza 1000 se toma m.a.s. de tamaño 100. Determínese la probabilidad de que la media muestral supere a la media poblacional en no menos de 0'2.

SOLUCIÓN

Siendo $X \approx N(\mu;10)$ y $(X_1;....;X_{100})$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño 100, el estadístico media muestral es $\overline{X} = (X_1 + ... + X_{100})/100$. La distribución de probabilidad de $\overline{X}$ es normal, pues $\overline{X}$ es combinación lineal de normales independientes, siendo $E(\overline{X}) = E(X) = \mu$ y $V(\overline{X}) = V(X)/100 = 1$. Así:

$$P(\overline{X} - \mu \geq 0'2) = P\left(N(0;1) \geq \frac{0'2}{1}\right) = ...$$

FONEMATO 2.6.3

Siendo $X \approx N(100;12)$, determínese el tamaño muestral mínimo que debe tomarse para garantizar, con probabilidad no inferior a 0'95, que la media muestral no difiera de la media poblacional en más de 2 unidades.

SOLUCIÓN

Siendo $X \approx N(100;12)$ y $(X_1;....;X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño "n", el estadístico media muestral es $\overline{X} = (X_1 + ... + X_n)/n$. La distribución de probabilidad de $\overline{X}$ es normal, por ser combinación lineal de normales independientes, siendo $E(\overline{X})=E(X)=100$ y $V(\overline{X})=V(X)/n=144/n$. Debemos determinar "n" de modo que $P(|\overline{X}-100|\leq 2)\geq 0'95$:

$$P(|\overline{X}-100|\leq 2)\geq 0'95 \Rightarrow P\left(|N(0;1)|\leq \frac{2}{12/\sqrt{n}}\right)\geq 0'95 \Rightarrow$$

$$\Rightarrow 2.P\left(N(0;1)\leq \frac{2}{12/\sqrt{n}}\right)-1\geq 0'95 \Rightarrow P\left(N(0;1)\leq \frac{2}{12/\sqrt{n}}\right)\geq 0'975 \Rightarrow$$

$$\Rightarrow \frac{2}{12/\sqrt{n}}\geq 1'96 \Rightarrow n\geq 138'29 \Rightarrow n\geq 139$$

FONEMATO 2.6.4

Siendo $X\approx N(\mu;1)$, determínese el tamaño muestral mínimo que debe tomarse para garantizar, con probabilidad no inferior a 0'95, que la media muestral no diferirá de la media poblacional en más de 0'5 unidades.

SOLUCIÓN

Siendo $X\approx N(\mu;1)$ y $(X_1;....;X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño "n", el estadístico media muestral es $\overline{X}=(X_1+....+X_n)/n$. La distribución de probabilidad de $\overline{X}$ es normal, pues $\overline{X}$ es combinación lineal de normales independientes, siendo $E(\overline{X})=E(X)=\mu$ y $V(\overline{X})=V(X)/n=1/n$.

Debemos determinar "n" de modo que $P(|\overline{X}-\mu|\leq 0'5)\geq 0'95$:

$$P(|\overline{X}-\mu|\leq 0'5)\geq 0'95 \Rightarrow P\left(|N(0;1)|\leq \frac{0'5}{1/\sqrt{n}}\right)\geq 0'95 \Rightarrow$$

$$\Rightarrow 2.P\left(N(0;1)\leq \frac{0'5}{1/\sqrt{n}}\right)-1\geq 0'95 \Rightarrow P\left(N(0;1)\leq \frac{0'5}{1/\sqrt{n}}\right)\geq 0'975 \Rightarrow$$

$$\Rightarrow \frac{0'5}{1/\sqrt{n}}\geq 1'96 \Rightarrow n\geq 15'21 \Rightarrow n\geq 16$$

FONEMATO 2.6.5

En horas, la duración de un tipo de bombillas tiene distribución $N(1000;100)$. Se desea enviar un muestra de modo que la duración media muestral no difiera de la poblacional en más de 50 horas con probabilidad no inferior a 0'95.

1) Determínese el tamaño de la muestra.
2) Determínese el tamaño de la muestra si de "X" sólo se sabe que tiene media 1000 y desviación típica 100.

SOLUCIÓN

1) Siendo $X\approx N(1000;100)$ y $(X_1;....;X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño "n", el estadístico media muestral es $\overline{X}=(X_1+....+X_n)/n$. La distribu-

ción de probabilidad de $\overline{X}$ es normal, pues $\overline{X}$ es combinación lineal de normales independientes, siendo $E(\overline{X}) = E(X) = 1000$ y $V(\overline{X}) = V(X)/n = 100^2/n$. Hay que determinar "n" de modo que $P(|\overline{X} - 1000| \leq 50) \geq 0'95$:

$$P(|\overline{X} - 1000| \leq 50) \geq 0'95 \Rightarrow P\left(|N(0;1)| \leq \frac{50}{100/\sqrt{n}}\right) \geq 0'95 \Rightarrow$$

$$\boxed{P(|N(0;1)| \leq a) = 2.P(N(0;1) \leq a) - 1}$$

$$\Rightarrow 2.P\left(N(0;1) \leq \frac{50}{100/\sqrt{n}}\right) - 1 \geq 0'95 \Rightarrow P\left(N(0;1) \leq \frac{50}{100/\sqrt{n}}\right) \geq 0'975 \Rightarrow$$

$$\Rightarrow \frac{50}{100/\sqrt{n}} \geq 1'96 \Rightarrow n \geq 15'36 \Rightarrow n \geq 16$$

2) Si $E(\overline{X}) = E(X) = 1000$ y $V(\overline{X}) = V(X)/n = 100^2/n$, desconociendo la distribución de probabilidad de "X", lidiamos mediante la acotación de Tchebychef:

$$P(|\overline{X} - 1000| \leq 50) = P(|\overline{X} - E(\overline{X})| \leq 50) \geq 1 - \frac{V(\overline{X})}{50^2} =$$

$$= 1 - \frac{100^2/n}{50^2} \geq 0'95 \Rightarrow n \geq 80$$

FONEMATO 2.6.6

De una población $X \approx N(13;6)$ se toma m.a.s de tamaño 9.

1) Determínese un intervalo en el que puede esperarse, con probabilidad 0'99, que se encuentre la media muestral.
2) Repita el apartado 1) si $n = 100$ y si $n = 10000$.
3) ¿Qué diría si en una m.a.s. de tamaño 10000 la media muestral es 15?

SOLUCIÓN

1) Es sabido que en el muestreo aleatorio simple de tamaño "n" de una población $X \approx N(\mu;\sigma)$, el estadístico media muestral $\overline{X}$ tiene distribución $N(\mu;\sigma/\sqrt{n})$. Por tanto, si $n = 9$ y $X \approx N(13;6)$, es $\overline{X} \approx N(13;6/\sqrt{9}) \equiv N(13;2)$.

En la distribución de probabilidad de $\overline{X} \approx N(13;2)$ hay infinidad de intervalos que "encierran" masa 0'99, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, como la distribución $N(13;2)$ es simétrica respecto al punto 13, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto (puede demostrase que éste es el de menor amplitud). Por tanto, debemos determinar $c > 0$ de modo que $P(|\overline{X} - 13| < c) = 0'99$

$$P\left(\left|\overline{X}-13\right|<c\right)=0'99 \Rightarrow P\left(\left|N(0;1)\right|<\frac{c}{2}\right)=0'99 \Rightarrow$$

$$\Rightarrow 2.P\left(N(0;1)<\frac{c}{2}\right)-1=0'99 \Rightarrow P\left(N(0;1)<\frac{c}{2}\right)=0'995 \Rightarrow$$

$$\Rightarrow \frac{c}{2}=2'57 \Rightarrow c=5'14$$

Así, tomando muestras de tamaño 9 de la población $X \approx N(13;6)$, el 99 % de las veces ocurrirá que el estadístico media muestral $\overline{X}$ se concretará en un punto del intervalo $(13-5'14;13+5'14) \equiv (7'86;18'14)$.

2) Si $n=100$ entonces $\overline{X} \approx N(13;6/\sqrt{100}) \equiv N(13;0'6)$; por tanto:

$$P\left(\left|\overline{X}-13\right|<c\right)=0'99 \Rightarrow P\left(\left|N(0;1)\right|<\frac{c}{0'6}\right)=0'99 \Rightarrow$$

$$\Rightarrow \frac{c}{0'6}=2'57 \Rightarrow c=1'54$$

Así, tomando muestras de tamaño 100 de la población $X \approx N(13;6)$, el 99 % de las veces ocurrirá que el estadístico muestral $\overline{X}$ se concretará en un punto del intervalo $(13-1'54;13+1'54) \equiv (11'46;14'54)$.

- Si $n=10000$ entonces $\overline{X} \approx N(13;6/\sqrt{10000}) \equiv N(13;0'06)$; por tanto:

$$P\left(\left|\overline{X}-13\right|<c\right)=0'99 \Rightarrow P\left(\left|N(0;1)\right|<\frac{c}{0'06}\right)=0'99 \Rightarrow \frac{c}{0'06}=2'57 \Rightarrow c=0'154$$

Así, tomando muestras de tamaño 10000 de la población $X \approx N(13;6)$, el 99 % de las veces ocurrirá que el estadístico muestral $\overline{X}$ se concretará en un punto del intervalo $(13-0'154;13+0'154) \equiv (12'846;13'154)$.

3) ***Pregunta:*** ¿qué dirías si en una m.a.s. de tamaño 10000 se obtiene $\overline{X}=15$?
Respuesta: depende
Pregunta: ¿de qué pende?
Respuesta: entre el blanco y el negro que describimos a continuación hay grises que sin duda el lector puede imaginar.

El blanco: tranquilidad total

Si estamos podridos de dinero y podemos tomar muestras de tamaño 10000 constantemente (por ejemplo, una muestra cada cinco minutos), estaremos felices mientras sea mayor o igual que 0'99 la proporción de muestras en que la media muestral $\overline{X}$ se concreta en algún punto del intervalo $(12'846;13'154)$. Mientras estemos felices, el que en una muestra concreta (la de las 13'35) suceda que $\overline{X}=15 \notin (12'846;13'154)$ no nos preocupará en absoluto: lo achacaremos a causas aleatorias por las que no merece la pena obsesionarse, pues es lógico y normal que, si tomamos muchas muestras, algunas sean "raras" (en este ejemplo, si $n=10000$, las muestras "raras" son aquéllas en que la media muestral no toma un valor en el intervalo $(12'846;13'154)$) .... pero estaremos tranquilos mientras la proporción de muestras "raras" no supere el 1 %.

El negro: mosqueo total

Si somos unos pobretones y para pagar la m.a.s. de tamaño 10000 hemos tenido que romper nuestra hucha cerdito de barro y empeñar el oso de peluche, entonces, si ocurre que $\overline{X} = 15 \notin (12'846;13'154)$ nos pegamos un tiro, pues el que haya sucedido tal cosa indica que algo falla estrepitosamente o que somos cenizos.

- **Falla la tracción:** los que han tomado la muestra estaban borrachos, o son tan inútiles que no saben tomar una m.a.s, o están comprados por la "competencia" y deliberadamente nos suministran datos falsos, o ....
- **Falla la información:** el responsable del desastre es el sinvergüenza que nos ha dicho que $X \approx N(13;6)$ ...... no es cierto que $X \approx N(13;6)$.
- **Somos megacenizos:** la tracción hace bien su trabajo y es cierto que $X \approx N(13;6)$, pero hemos tenido la desgracia de seleccionar una muestra "rara".

Debes saber que este tipo de triángulos carecen de solución: si nos obsesionamos en investigar qué ha pasado, los encargados de la "tracción" dirán que "ellos" no han sido, el que nos ha dicho que $X \approx N(13;6)$ nos jurará por Snoopy que "su" información es buena ..... y como la muestra no puede hablar, no habrá más remedio que pedir un crédito al banco y seleccionar otra muestra, o suicidarnos.

Observa: si en la muestra aleatoria de tamaño 10000 seleccionada se obtuviera $\overline{X} = 12'83 \notin (12'846;13'154)$, diríamos lo mismo que si $\overline{X} = 15 \notin (12'846;13'154)$, pues $\overline{X} = 12'83$ también corresponde a una muestra "rara". No obstante, la media muestral 12'83 deja de ser "rara" si sólo consideramos "raras" las muestras en que $\overline{X} \notin (11'46;14'54)$, que es el intervalo que corresponde a $n = 100$; lo que no sucede con una media muestral 15, que sigue siendo "rara" para $n = 100$.

FONEMATO 2.6.7

De una población $X \approx N(5;0'1)$ se toma m.a.s de tamaño 16.

Determínese la probabilidad del suceso $5 < \overline{X} < 5'2$, la del suceso $S^2 \leq 0'019125$, la del suceso $S > 0'1$ y la del suceso $5 < \overline{X} < 5'2 \,/\, S^2 \leq 0'019125$.

SOLUCIÓN

En el muestreo aleatorio simple de tamaño "n" de una población $N(\mu;\sigma)$, siendo

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

es $\overline{X} \approx N(\mu;\sigma/\sqrt{n})$ y $n.S^2/\sigma^2 \approx \chi^2_{n-1}$. Así, si $n = 16$, $\mu = 5$ y $\sigma = 0'1$, es:

$$\overline{X} \approx N(5;0'1/4) \;;\; 16.S^2/0'1^2 \approx \chi^2_{15}$$

- $$P\left(5 < \overline{X} < 5'2\right) = P\left(5 < N(5;0'1/4) < 5'2\right) = ...$$
- $P\left(S^2 \leq 0'019125\right) = P\left(16.S^2/0'1^2 \leq 16.0'019125/0'1^2\right) = P\left(\chi^2_{15} \leq 30'6\right) = 0'99$

- $P(S > 0'1) = P\left(S^2 > 0'1^2\right) = P\left(16.S^2/0'1^2 > 16.0'1^2/0'1^2\right) = P\left(\chi^2_{15} > 16\right) = ...$
- $$P\left(5 < \overline{X} < 5'2 \,/\, S^2 \le 0'019125\right) = P\left(5 < \overline{X} < 5'2\right)$$

en el muestreo de una población normal, los estadísticos media muestral y varianza muestral son independientes

FONEMATO 2.6.8

Si de una población $X \approx N(1'7;4)$ se toma m.a.s de tamaño 10, determínese la cota que superará la desviación típica muestral con probabilidad 0'99.

SOLUCIÓN

En el m.a.s de tamaño "n" de una población $X \approx N(\mu;\sigma)$, siendo S^2 el estadístico varianza muestral, es $n.S^2/\sigma^2 \approx \chi^2_{n-1}$. Por tanto, siendo $n = 10$ y $\sigma = 4$, es $10.S^2/4^2 \approx \chi^2_9$, y buscamos "c" $(c > 0)$ tal que $P(S > c) = 0'99$:

$$P(S > c) = 0'99 \Rightarrow P(S^2 > c^2) = 0'99 \Rightarrow P\left(10.S^2/4^2 > 10.c^2/4^2\right) = 0'99 \Rightarrow$$
$$\Rightarrow P\left(\chi^2_9 > 10.c^2/4^2\right) = 0'99 \Rightarrow 10.c^2/4^2 = 2'09 \Rightarrow c = 1'82$$

FONEMATO 2.6.9

En el muestreo aleatorio simple de tamaño "n" de una población normal, determínese la función de densidad de probabilidad del estadístico varianza muestral.

SOLUCIÓN

Siendo "Z" el estadístico varianza muestral, en el m.a.s de tamaño "n" de una población $N(\mu;\sigma)$, es $W = n.Z/\sigma^2 \approx \chi^2_{n-1} \equiv G\left((n-1)/2;1/2\right)$.

- Función de distribución de la varianza muestral "Z":

$$F_Z(z) = P(Z \le z) = P\left(\sigma^2.W/n \le z\right) = P\left(W \le n.z/\sigma^2\right) = F_W(n.z/\sigma^2)$$

$W = n.Z/\sigma^2 \Rightarrow Z = \sigma^2.W/n$

siendo F_W la función de distribución de $W \approx \chi^2_{n-1}$

- Función de densidad de probabilidad de la varianza muestral "Z":

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \frac{dF_W(n.z/\sigma^2)}{dz} = \frac{dF_W(n.z/\sigma^2)}{d(n.z/\sigma^2)} \cdot \frac{d(n.z/\sigma^2)}{dz} =$$

siendo f_W la función de distribución de $W \approx \chi^2_{n-1}$, es:

$$\frac{dF_W(n.z/\sigma^2)}{d(n.z/\sigma^2)} = f_W(n.z/\sigma^2)$$

$$=\frac{n}{\sigma^2}.f_W(n.z/\sigma^2)=\frac{n}{\sigma^2}.\frac{(1/2)^{(n-1)/2}}{\Gamma((n-1)/2)}.(n.z/\sigma^2)^{(n-3)/2}.e^{-n.z/2.\sigma^2}=$$

$$\boxed{W\approx\chi^2_{n-1}\equiv G\left(\frac{n-1}{2};\frac{1}{2}\right)\Rightarrow f_W(w)=\frac{(1/2)^{(n-1)/2}}{\Gamma((n-1)/2)}.w^{(n-3)/2}.e^{-w/2},\quad w>0}$$

$$=\frac{n}{\sigma^2}.\frac{(1/2.\sigma^2)^{(n-1)/2}}{\Gamma((n-1)/2)}.z^{(n-1)/2}.e^{-n.z/2.\sigma^2},\quad z>0$$

que corresponde a una variable con distribución $G((n-1)/2;n/2.\sigma^2)$.

FONEMATO 2.6.10

En una m.a.s. $X_1,\dots,X_n$ de una población $N(\mu;\sigma)$, siendo $\overline{X}$ y S^2 las correspondientes media y varianza muestrales, y siendo $X_{n+1}\approx N(\mu;\sigma)$ independiente de $X_1,\dots,X_n$, determínese la distribución de probabilidad de "T", siendo:

$$T=\frac{X_{n+1}-\overline{X}}{S}.\sqrt{\frac{n-1}{n+1}}$$

SOLUCIÓN

En el m.a.s de tamaño "n" de una población $N(\mu;\sigma)$ sucede que:

$$\overline{X}\approx N(\mu;\sigma/\sqrt{n})\ ;\ n.S^2/\sigma^2\approx\chi^2_{n-1}$$

Como las variables X_{n+1} y $\overline{X}$ son independientes (pues X_{n+1} es independiente de $X_1,\dots,X_n$), la distribución de probabilidad de $X_{n+1}-\overline{X}$ es normal, siendo:

$$E(X_{n+1}-\overline{X})=E(X_{n+1})-E(\overline{X})=\mu-\mu=0$$

$$V(X_{n+1}-\overline{X})=V(X_{n+1})+V(\overline{X})=\sigma^2+\frac{\sigma^2}{n}=\frac{n+1}{n}.\sigma^2$$

Por tanto, siendo $X_{n+1}-\overline{X}\approx N(0;\sigma.\sqrt{(n+1)/n})$, es:

$$\frac{X_{n+1}-\overline{X}}{\sigma.\sqrt{(n+1)/n}}\approx N(0;1)$$

Recordando que siendo independientes $Y\approx N(0;1)$ y $U\approx\chi^2_k$ es $\dfrac{Y}{\sqrt{U/k}}\approx t_k$, entonces, tomando como $N(0;1)$ la variable

$$\frac{X_{n+1}-\overline{X}}{\sigma.\sqrt{(n+1)/n}}$$

y recordando que $n.S^2/\sigma^2\approx\chi^2_{n-1}$, resulta evidente que:

$$T=\frac{\dfrac{X_{n+1}-\overline{X}}{\sigma.\sqrt{(n+1)/n}}}{\sqrt{\dfrac{n.S^2/\sigma^2}{n-1}}}=\frac{X_{n+1}-\overline{X}}{S}.\sqrt{\frac{n-1}{n+1}}\approx t_{n-1}$$

FONEMATO 2.6.11

De una población $X \approx N(\mu;2)$ se toma m.a.s de tamaño 9. Determínese un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre la varianza muestral. Repita lo anterior si $n = 101$.

SOLUCIÓN

En el m.a.s de tamaño "n" de una población $N(\mu;\sigma)$ es $n.S^2/\sigma^2 \approx \chi^2_{n-1}$.

- Siendo $n = 9$ y $\sigma = 2$, es $9.S^2/4 \approx \chi^2_8$. En la distribución de probabilidad de la variable $9.S^2/4 \approx \chi^2_8$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. Como la distribución χ^2_8 no es simétrica respecto a ningún punto, por comodidad, elegimos el intervalo que reparte la probabilidad $1 - 0'95 = 0'05$ en partes iguales entre las dos colas (izquierda y derecha) de dicha χ^2_8 (puede demostrase que éste no es el de menor amplitud, pero del de menor amplitud es tan petardo de calcular que pasamos de él); la tabla de la función de distribución de $9.S^2/4 \approx \chi^2_8$ indica que:

$$P\left(2'18 \le 9.S^2/4 \approx \chi^2_8 \le 17'53\right) = 0'95$$

  Por tanto, es $P\left(\frac{2'18.4}{9} \le S^2 \le \frac{17'53.4}{9}\right) = 0'95$; o sea: $P\left(0'96 \le S^2 \le 7'79\right) = 0'95$

  En definitiva, tomando sucesivas muestras de tamaño 9 de una población $X \approx N(\mu;2)$, el 95 % de las veces sucederá que el estadístico varianza muestral se concreta en algún punto del intervalo $(0'96;7'79)$.

- Siendo $n = 101$ y $\sigma = 2$, es $101.S^2/4 \approx \chi^2_{100}$. Procediendo como en el caso anterior, la tabla de la función de distribución de $101.S^2/4 \approx \chi^2_{100}$ indica que:

$$P\left(74'2 \le 101.S^2/4 \approx \chi^2_{100} \le 129'6\right) = 0'95$$

  Por tanto, es $P\left(\frac{74'2.4}{101} \le S^2 \le \frac{129'6.4}{101}\right) = 0'95$; o sea: $P\left(2'93 \le S^2 \le 5'13\right) = 0'95$.

  En definitiva, tomando sucesivas muestras de tamaño 101 de una población $X \approx N(\mu;2)$, el 95 % de las veces sucederá que el estadístico varianza muestral se concreta en algún punto del intervalo $(2'93;5'13)$.

FONEMATO 2.6.12

De una población normal de varianza 4'5 se toman dos muestras independientes de tamaño "n". Determínese "n" de modo que sea al menos 0'95 la probabilidad de que las medias muestrales difieran en no más de 2 unidades.

SOLUCIÓN

Siendo $X \approx N(\mu; \sqrt{4'5})$, si $(X_1; \ldots; X_n)$ e $(Y_1; \ldots; Y_n)$ son las respectivas variables muestrales que expresan los resultados que pueden presentarse al tomar las dos muestras de tamaño "n" indicadas, es:

$$\overline{X} = \frac{1}{n}.\sum_{i=1}^{n} X_i \approx N(\mu; \sqrt{4'5/n}) \; ; \; \overline{Y} = \frac{1}{n}.\sum_{i=1}^{n} Y_i \approx N(\mu; \sqrt{4'5/n})$$

Por tanto, siendo independientes los estadísticos medias muestrales $\overline{X}$ e $\overline{Y}$, es:

$$\overline{X} - \overline{Y} \approx N(0; \sqrt{9/n})$$

$$\boxed{\begin{array}{c} E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu - \mu = 0 \\ V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y}) = 9/n \end{array}}$$

Debemos determinar "n" de modo que $P\left(\left|\overline{X} - \overline{Y}\right| \leq 2\right) \geq 0'95$:

$$P\left(\left|\overline{X} - \overline{Y}\right| \leq 2\right) = 0'95 \Rightarrow P\left(\left|N(0;1)\right| \leq \frac{2}{\sqrt{9/n}}\right) \geq 0'95 \Rightarrow$$

$$\Rightarrow 2.P\left(N(0;1) \leq \frac{2}{\sqrt{9/n}}\right) - 1 \geq 0'95 \Rightarrow P\left(N(0;1) \leq \frac{2}{\sqrt{9/n}}\right) \geq 0'975 \Rightarrow$$

$$\Rightarrow \frac{2}{\sqrt{9/n}} \geq 1'96 \Rightarrow n \geq 9$$

FONEMATO 2.6.13

De dos poblaciones independientes $N(\mu_1;\sigma_1)$ y $N(\mu_2;\sigma_2)$ se toman muestras de tamaños respectivos "n" y "m", siendo $\overline{X}$ e $\overline{Y}$ las respectivas medias muestrales y S_{c1}^2 y S_{c2}^2 las cuasivarianzas muestrales. Determínese la distribución de probabilidad de los estadísticos T_1 y T_2:

$$T_1 = \frac{n-1}{\sigma_1^2}.S_{c1}^2 + \frac{m-1}{\sigma_2^2}.S_{c2}^2$$

$$T_2 = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2/n)+(\sigma_2^2/m)}}.\frac{\sqrt{n+m-2}}{\sqrt{T_1}}$$

SOLUCIÓN

Para la muestra de tamaño "n" de la población $N(\mu_1;\sigma_1)$, siendo S_1^2 la correspondiente varianza muestral, es $(n-1).S_{c1}^2/\sigma_1^2 \approx \chi_{n-1}^2$, pues $n.S_1^2/\sigma_1^2 \approx \chi_{n-1}^2$ y $S_1^2 = (n-1).S_{c1}^2/n$. Para la muestra de tamaño "m" de la población $N(\mu_2;\sigma_2)$, siendo S_2^2 la correspondiente varianza muestral, es $(m-1).S_{c2}^2/\sigma_2^2 \approx \chi_{m-1}^2$, pues $m.S_2^2/\sigma_2^2 \approx \chi_{m-1}^2$ y $S_2^2 = (m-1).S_{c2}^2/m$.

En consecuencia, debido a la reproductividad de la distribución ji-dos, es:

$$T_1 = \frac{n-1}{\sigma_1^2}.S_{c1}^2 + \frac{m-1}{\sigma_2^2}.S_{c2}^2 \approx \chi_{n+m-2}^2$$

Siendo $\overline{X} \approx N(\mu_1;\sigma_1/\sqrt{n})$ e $\overline{Y} \approx N(\mu_2;\sigma_2/\sqrt{m})$ las respectivas medias muestrales, es $\overline{X}-\overline{Y} \approx N\left(\mu_1-\mu_2;\sqrt{(\sigma_1^2/n)+(\sigma_2^2/m)}\right)$; por tanto:

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2/n)+(\sigma_2^2/m)}} \approx N(0;1)$$

En consecuencia, recordando que la variable t-Student con "k" grados de libertad es el cociente entre una variable $Pepa \approx N(0;1)$ y la raíz cuadrada de una variable $Juana \approx \chi_k^2$ dividida por sus grados de libertad (siendo independientes "Pepa" y "Juana"), si consideramos que

$$Pepa = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2/n)+(\sigma_2^2/m)}} \approx N(0;1) \ ; \ Juana = T_1 \approx \chi_{n+m-2}^2$$

es:

$$T_2 = \frac{\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2/n)+(\sigma_2^2/m)}}}{\sqrt{\dfrac{T_1}{n+m-2}}} = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2/n)+(\sigma_2^2/m)}}.\frac{\sqrt{n+m-2}}{\sqrt{T_1}} \approx t_{n+m-2}$$

FONEMATO 2.6.14

De dos poblaciones independientes con la misma varianza σ^2 se toman muestras de tamaños respectivos "n" y "m", siendo $\overline{X}$ e $\overline{Y}$ las respectivas medias muestrales. Si $\overline{W}$ es la media aritmética de todas las observaciones realizadas, demuéstrese que $V(\overline{W}-\overline{X}) = m.\sigma^2/(n^2+n.m)$.

SOLUCIÓN

Siendo $(X_1; \ldots ; X_n)$ e $(Y_1; \ldots ; Y_m)$ las respectivas variables muestrales que expresan los resultados que pueden presentarse al tomar las dos muestras indicadas, sean:

$$\overline{X} = \frac{1}{n}.\sum_{i=1}^{n} X_i \;\; ; \;\; \overline{Y} = \frac{1}{m}.\sum_{j=1}^{m} Y_j$$

Siendo $\overline{W}$ la media aritmética de todas las observaciones realizadas, es:

$$\overline{W} = \frac{1}{n+m}.\left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j\right)$$

Por tanto:

$$V(\overline{W}-\overline{X}) = V\left(\frac{1}{n+m}.\left(\sum_{j=1}^{m} Y_j - \frac{m}{n}.\sum_{i=1}^{n} X_i\right)\right) =$$

$$\boxed{\begin{array}{c} \overline{W}-\overline{X} = \dfrac{1}{n+m}.\left(\displaystyle\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j\right) - \dfrac{1}{n}.\displaystyle\sum_{i=1}^{n} X_i = \\ = \left(\dfrac{1}{n+m} - \dfrac{1}{n}\right).\displaystyle\sum_{i=1}^{n} X_i + \dfrac{1}{n+m}.\displaystyle\sum_{j=1}^{m} Y_j = \dfrac{1}{n+m}.\left(\displaystyle\sum_{j=1}^{m} Y_j - \dfrac{m}{n}.\sum_{i=1}^{n} X_i\right) \end{array}}$$

$$= \frac{1}{(n+m)^2}.\left(\sum_{j=1}^{m} V(Y_j) + \frac{m^2}{n^2}.\sum_{i=1}^{n} V(X_i)\right) =$$

$$\boxed{\text{pues } X_1, \ldots, X_n, Y_1, \ldots, Y_m \text{ son independientes}}$$

$$= \frac{1}{(n+m)^2}.\left(m.\sigma^2 + \frac{m^2}{n^2}.n.\sigma^2\right) = \frac{m.\sigma^2}{n^2+n.m}$$

$$\boxed{\begin{array}{c}\text{las variables } X_1, \ldots, X_n, Y_1, \ldots, Y_m \text{ tienen varianza } \sigma^2 \text{, como} \\ \text{las dos poblaciones muestreadas}\end{array}}$$

FONEMATO 2.6.15

La duración media de las bombillas tipo A es de 2500 horas con desviación típica 600 horas, y la duración media de las bombillas tipo B es de 2300 horas con desviación típica 800 horas. Se toman 300 bombillas tipo A y 200 tipo B.

1) Calcúlese la probabilidad de que la duración media de las bombillas tipo A no supere en más de 100 horas a la duración media de la del tipo B.
2) Calcúlese la probabilidad de que la duración media de las bombillas tipo A supere en más de 200 horas a la duración media de la del tipo B.
3) Determínese un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre la diferencia de medias muestrales.

SOLUCIÓN

Siendo "X" la duración de las bombillas tipo A e "Y" la duración de las del tipo B, se nos dice que:

$$E(X) = 2500 \; ; \; V(X) = 600^2 \; ; \; E(Y) = 2300 \; ; \; V(Y) = 800^2$$

Siendo $\overline{X}$ e $\overline{Y}$ las respectivas medias muestrales correspondientes a las dos muestras indicadas, por ser grandes los tamaños muestrales 300 y 200, podemos aproximar (Lindeberg-Levy) la distribución de probabilidad de $\overline{X}$ mediante la normal con media y varianza las de $\overline{X}$, y la distribución de probabilidad de $\overline{Y}$ mediante la normal con media y varianza las de $\overline{Y}$, que son:

$$E(\overline{X}) = E(X) = 2500 \; ; \; V(\overline{X}) = V(X)/300 = 600^2/300$$
$$E(\overline{Y}) = E(Y) = 2300 \; ; \; V(\overline{Y}) = V(Y)/200 = 800^2/200$$

Por tanto, como los estadísticos $\overline{X}$ e $\overline{Y}$ son independientes, la distribución de probabilidad de $\overline{X} - \overline{Y}$ es normal, siendo:

$$E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = 2500 - 2300 = 200$$
$$V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y}) = (600^2/300) + (800^2/200) = 4400$$

1) Es:

$$P\left(\overline{X} - \overline{Y} \le 100\right) = P\left(N(200;\sqrt{4400}) \le 100\right) = P\left(N(0;1) \le \frac{100-200}{\sqrt{4400}}\right) = \ldots.$$

2) Es:

$$P\left(\overline{X} - \overline{Y} > 200\right) = P\left(N(200;\sqrt{4400}) > 200\right) = P\left(N(0;1) > \frac{200-200}{\sqrt{4400}}\right) = 0'5$$

3) En la distribución de probabilidad del estadístico $\overline{X} - \overline{Y} \approx N(200;\sqrt{4400})$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, como la variable $N(200;\sqrt{4400})$ es simétrica respecto al punto 200, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto, que es el de menor amplitud. Así, debemos determinar $c > 0$ de modo que $P\left(\left|\overline{X} - \overline{Y} - 200\right| < c\right) = 0'95$:

$$P\left(\left|\overline{X}-\overline{Y}-200\right|<c\right)=0'95 \Rightarrow P\left(\left|N(0;1)\right|<\frac{c}{\sqrt{4400}}\right)=0'95 \Rightarrow$$

$$\Rightarrow \frac{c}{\sqrt{4400}}=1'96 \Rightarrow 130$$

Por tanto, si tomamos sucesivas muestras de tamaño 300 de las bombillas tipo A y de tamaño 200 de las del tipo B, la diferencia entre las medias muestrales se concretará en el intervalo $(200-130;200+130)$ el 95 % de las veces.

FONEMATO 2.6.16

Sean S_1^2 y S_2^2 las respectivas varianzas muestrales correspondientes a dos muestras independientes de tamaños 5 y 4 de dos poblaciones normales con igual varianza.

1) Determínese la probabilidad del suceso $S_2^2/S_1^2 > 5'2$.
2) Con los mismos tamaños muestrales, si las poblaciones son $N(3;2)$ y $N(7;8)$, determínese un intervalo en el que, con probabilidad 0'95, se encuentre el aleatorio cociente de varianzas muestrales.

SOLUCIÓN

Es sabido que en el m.a.s de tamaño "n" de una población $N(\mu;\sigma)$, el estadístico $n.S^2/\sigma^2$ tiene distribución χ^2_{n-1}. Por tanto, siendo S_1^2 y S_2^2 las respectivas varianzas muestrales correspondientes a dos muestras independientes de tamaños 5 y 4 de las poblaciones $X \approx N(\mu_1;\sigma_1)$ e $Y \approx N(\mu_2;\sigma_2)$, es:

$$5.S_1^2/\sigma_1^2 \approx \chi^2_{5-1}=\chi^2_4 \ ; \ 4.S_2^2/\sigma_2^2 \approx \chi^2_{4-1}=\chi^2_3$$

En consecuencia, siendo $\sigma_1=\sigma_2$, es: $\dfrac{\dfrac{4.S_2^2/\sigma_2^2}{3}}{\dfrac{5.S_1^2/\sigma_1^2}{4}}=\dfrac{16.S_2^2}{15.S_1^2}\approx F_{3;4}$.

1)
$$P\left(S_2^2/S_1^2>5'2\right)=P\left(\frac{16.S_2^2}{15.S_1^2}>\frac{16}{15}.5'2\right)=P\left(F_{3;4}>\frac{16}{15}.5'2\right)=$$

$$=1-P\left(F_{3;4}\le\frac{16}{15}.5'2\right)=1-0'928$$

$$\boxed{\text{interpolando entre } P\left(F_{3;4}\le 4'19\right)=0'9 \text{ y } P\left(F_{3;4}\le 6'59\right)=0'95}$$

2) Si $X \approx N(3;2)$ e $Y \approx N(7;8)$, para tamaños muestrales respectivos 5 y 4, es:

$$5.S_1^2/2^2 \approx \chi^2_{5-1}=\chi^2_4 \ ; \ 4.S_2^2/8^2 \approx \chi^2_{4-1}=\chi^2_3$$

Por tanto:
$$\frac{\dfrac{5.S_1^2/2^2}{4}}{\dfrac{4.S_2^2/8^2}{3}}=\frac{15.S_1^2}{S_2^2}\approx F_{4;3}$$

En la distribución de probabilidad de la variable $15.S_1^2/S_2^2 \approx F_{4;3}$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. Como la distribución $F_{4;3}$ no es simétrica respecto a ningún punto, por comodidad, elegimos el intervalo que reparte la probabilidad $1-0'95=0'05$ en partes iguales entre las dos colas (izquierda y derecha) de dicha $F_{4;3}$ (puede demostrase que éste no es el de menor amplitud, pero del de menor amplitud es tan petardo que pasamos de él). Según la tabla de la función de distribución de la variable $F_{4;3}$ es $P(F_{4;3} \leq 15'1)=0'975$; además:

$$P(F_{4;3} \leq c)=0'025 \Rightarrow P(1/F_{4;3} \geq 1/c)=0'025 \Rightarrow P(F_{3;4} \geq 1/c)=0'025 \Rightarrow$$
$$\Rightarrow 1-P(F_{3;4} \leq 1/c)=0'025 \Rightarrow P(F_{3;4} \leq 1/c)=0'975 \Rightarrow 1/c=9'98 \Rightarrow c=0'1$$

Así, siendo $P(0'1 \leq 15.S_1^2/S_2^2 \approx F_{4;3} \leq 15'1)=0'95$, es:

$$P(0'1/15 \leq .S_1^2/S_2^2 \leq 15'1/15)=0'95$$

Por tanto, si tomamos sucesivas muestras de tamaños 5 y 4 de las poblaciones $N(3;2)$ y $N(7;8)$, el 95 % de las veces ocurrirá que estadístico S_1^2/S_2^2 se concreta en el intervalo $(0'1/15;15'1/15)$.