10 MUESTREO

1 ¿Cómo variará la desviación típica muestral si se multiplica por cuatro el tamaño de la muestra? ¿Y si se aumenta el tamaño de la muestra de 16 a 144?

Si μ y σ son la media y la desviación típica poblacionales, y se extraen muestras de tamaño n, entonces la distribución muestral de la media tiene una media $\mu_{\overline{X}} = \mu$ y una desviación típica $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

Si se toman muestras de tamaño $n' = 4n$, entonces $\sigma_{\overline{X}'} = \dfrac{\sigma}{\sqrt{4n}} = \dfrac{\sigma}{2\sqrt{n}} = \dfrac{1}{2}\sigma_{\overline{X}}$.

Si el tamaño muestral pasa de $n = 16$ a $n' = 144 = 9 \cdot n$, la desviación típica pasa de $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$ a $\sigma_{\overline{X}'} = \dfrac{\sigma}{\sqrt{9n}} = \dfrac{\sigma}{3\sqrt{n}} = \dfrac{1}{3}\sigma_{\overline{X}}$

2 Consideremos la población P = {1, 2, 3, 4}.

a) Calcular la media y la desviación típica de la población.
b) Formar todas las muestras <u>con reemplazamiento</u> de tamaño 2, y calcular la media y la desviación típica de las medias muestrales.
c) Comparar los resultados de a) y b).

a) $\mu = \dfrac{1+2+3+4}{4} = \dfrac{10}{4} = \dfrac{5}{2}$; $\sigma^2 = \dfrac{1+4+9+16}{4} - \left(\dfrac{5}{2}\right)^2 = \dfrac{5}{4}$; $\sigma = \dfrac{\sqrt{5}}{2}$

b) Las muestras de tamaño n = 2 que pueden formarse <u>con reemplazamiento</u> a partir de la población P = {1,2,3,4}, con #P = N = 4, son las variaciones binarias <u>con repetición</u> de los elementos de P, en número de $VR_{4,2} = 4^2 = 16$. Explícitamente, estas muestras y sus medias son:

M_i	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)
$\overline{x}_i$	1	3/2	2	5/2	3/2	2	5/2	3

(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)
2	5/2	3	7/2	5/2	3	7/2	4

La distribución muestral de las medias es:

$\overline{x}_i$	1	3/2	2	5/2	3	7/2	4	
n_i	1	2	3	4	3	2	1	16
$n_i\overline{x}_i$	1	3	6	10	9	7	5	40
$\overline{x}_i^2$	1	9/4	4	25/4	9	49/4	16	
$n_i\overline{x}_i^2$	1	9/2	12	25	27	49/2	16	110

La media y desviación típica de la distribución muestral de las medias son:

$$\mu_{\bar{X}} = \bar{\bar{X}} = \frac{\sum n_i\, \bar{x}_i}{\sum n_i} = \frac{40}{16} = \frac{5}{2}$$, que <u>coincide</u> con la media poblacional, $\mu = \frac{5}{2}$

$$\sigma^2_{\bar{X}} = \frac{\sum n_i\, x_i^2}{\sum n_i} - (\mu_{\bar{X}})^2 = \frac{110}{16} - \left(\frac{5}{2}\right)^2 = \frac{55}{8} - \frac{25}{4} = \frac{5}{8}$$

$$\sigma_{\bar{X}} = \sqrt{\frac{5}{8}} = \frac{\sqrt{5}}{2\sqrt{2}} = \frac{\sqrt{5}/2}{\sqrt{2}}$$, que <u>no coincide</u> con $\sigma = \frac{\sqrt{5}}{2}$, sino que es $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

c) La comparación está hecha.

3 Las calificaciones obtenidas por los alumnos de Estadística de Segundo de Bachillerato de la provincia de Granada están distribuidas según una ley desconocida de media 6,4 y desviación típica 1,8. Si se elige una muestra aleatoria simple de 121 alumnos, calcular:

a) La probabilidad de que la media muestral sea a lo sumo 6,6.
b) La probabilidad de que la media muestral esté entre 6,4 y 7,4.

Aunque X siga una ley desconocida, de media $\mu = 6{,}4$ y desviación típica $\sigma = 1{,}8$, como el tamaño de la muestra $n = 121 > 30$ es suficientemente grande, por el Teorema Central del Límite, la media muestral sigue aproximadamente una ley normal

$$\bar{X} \sim \mathcal{N}\left(\mu_{\bar{X}} = \mu = 6{,}4\,;\ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1{,}8}{11} = 0{,}163636\right)$$

a) Tipificamos la variable $\bar{X}$ y $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0;\ 1)$, luego:

$$P(\bar{X} \le 6{,}6) = P\left(\frac{\bar{X} - 6{,}4}{0{,}163636} \le \frac{6{,}6 - 6{,}4}{0{,}163636}\right) = P(Z \le 1{,}2222) = 0{,}8892$$

b) $$P(6{,}4 < \bar{X} \le 7{,}4) = P\left(\frac{6{,}4 - 6{,}4}{0{,}163636} < Z \le \frac{7{,}4 - 6{,}4}{0{,}163636}\right) = P(0 < Z \le 6{,}1111) = \Phi(6{,}1111) - \Phi(0) =$$

$$= 1{,}0000 - 0{,}5000 = 0{,}5$$

4 Supongamos que la longitud de las truchas de Riofrío se distribuye normalmente con media 24,5 cm y desviación típica 4 cm. Se toma una muestra de 25 truchas.

a) Calcular la probabilidad de que la media muestral esté comprendida entre 22 cm y la media poblacional.
b) Calcular la probabilidad de que la media muestral sea mayor o igual que 26 cm.
c) ¿Qué tamaño han de tener las muestras si se sabe que la probabilidad de que la media muestral sea a lo sumo 25 es 0,6915?

a) Sea X la variable aleatoria “longitud”, con X $\sim \mathcal{N}(\mu = 24{,}5;\ \sigma = 4)$, y sea $\overline{X}$ la variable aleatoria media muestral en el muestreo, con tamaño muestral $n = 25$. Entonces, por el Teorema Central del Límite, $\overline{X}$ sigue una ley normal:

$$\mathcal{N}\left(\mu_{\overline{X}} = \mu = 24{,}5;\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{5} = 0{,}8\right)$$

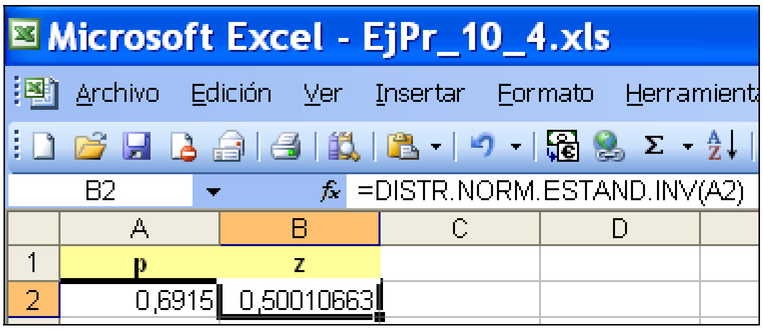
Tipificamos la variable $\overline{X}$, y $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0;\ 1)$. Entonces:

$$P(22 < \overline{X} \le 24{,}5) = P\left(\frac{22 - 24{,}5}{0{,}8} < Z \le 0\right) = P(-3{,}125 < Z \le 0) = \Phi(0) - \Phi(-3{,}125) =$$
$$= 0'5000 - 0{,}00088903 = 0{,}4991$$

b) $$P(\overline{X} \ge 26) = 1 - P(\overline{X} < 26) = 1 - P\left(Z < \frac{26 - 24{,}5}{0{,}8}\right) = 1 - P(Z < 1{,}875) =$$
$$= 1 - 0{,}9696 = 0{,}0304$$

c) Se trata de hallar el tamaño muestral n sabiendo que $P(\overline{X} \le 25) = 0{,}6915$.

$$P\left(Z \le \frac{25 - 24{,}5}{\frac{4}{\sqrt{n}}}\right) = 0{,}6915 \Rightarrow P\left(0{,}125\sqrt{n}\right) = 0{,}6915 \Rightarrow 0{,}125\sqrt{n} \approx 0{,}5000 \Rightarrow \sqrt{n} = 4 \Rightarrow n = 16$$



5 Una máquina fabrica bolas cuyo diámetro X es una variable aleatoria de ley desconocida de media 20 mm y desviación típica 1 mm. Se extraen a intervalos regulares muestras de tamaño 100.

a) ¿Cuál es la ley de probabilidad de la media muestral, $\overline{X}$?
b) Calcular la probabilidad de que la media muestral sea inferior o igual a 19,5 mm

a) Aunque X siga una ley desconocida, de media $\mu = 20$ y desviación típica $\sigma = 1$, como el tamaño de la muestra $n = 100 > 30$ es suficientemente grande, por el Teorema Central del Límite, la media muestral sigue aproximadamente una ley normal

$$\overline{X} \sim \mathcal{N}(\mu_{\overline{X}} = \mu = 20;\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{10} = 0{,}1)$$

b) Tipificamos la variable $\overline{X}$ y $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0;\ 1)$, luego:

$$P(\overline{X} \leq 19{,}5) = P\left(\frac{\overline{X} - 20}{0{,}1} \leq \frac{19{,}5 - 20}{0{,}1}\right) = P(Z \leq -5) = 1 - P(X < 5) \approx 0$$

6 Una máquina fabrica bolas cuyo diámetro X es una variable aleatoria de ley normal de media 20 mm y desviación típica 1 mm. Se extraen a intervalos regulares muestras de tamaño 4.

a) ¿Cuál es la ley de probabilidad de la media muestral, $\overline{X}$?
b) Calcular la probabilidad de que la media muestral sea mayor o igual a 20,5 mm.

a) Ahora X sigue una ley normal, de media $\mu = 20$ y desviación típica $\sigma = 1$, y aunque el tamaño de la muestra $n = 4$ sea pequeño, por el Teorema Central del Límite, la media muestral sigue una ley normal

$$\overline{X} \sim \mathcal{N}(\mu_{\overline{X}} = \mu = 20;\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{2} = 0{,}5)$$

b) Tipificamos la variable $\overline{X}$ y $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0;\ 1)$, luego:

$$P(\overline{X} \geq 20{,}5) = P\left(\frac{\overline{X} - 20}{0{,}5} \geq \frac{20{,}5 - 20}{0{,}5}\right) = P(Z \geq 1) = 1 - P(X < 1) = 1 - \Phi(1) =$$
$$= 1 - 0{,}84134474 = 0{,}15865526$$

7 En una clínica maternal se ha estimado que la proporción de varones nacidos en un cierto período largo de tiempo es del 45%. Calcular la probabilidad de que en los próximos 100 nacimientos:

a) Sean varones más del 50%.
b) Sean niñas menos del 50%.
c) La proporción de niñas esté comprendida entre el 50% y el 60%.

Si el valor estimado de la proporción de varones es $\hat{p} = 0,45$ y el tamaño muestral es $n = 100$, la variable aleatoria P, proporción de varones en la población, sigue una ley normal

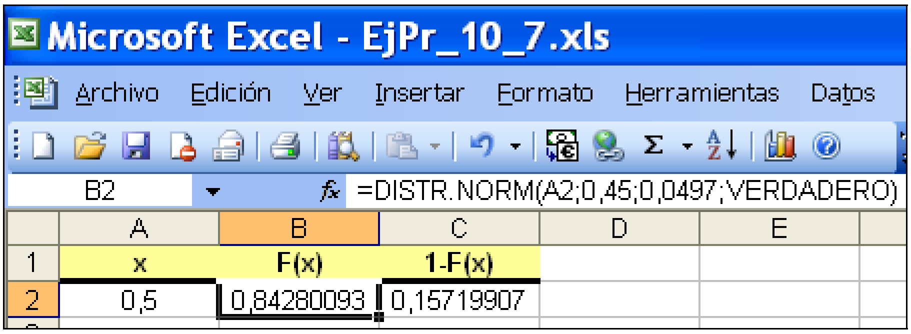
$$P \sim \mathcal{N}\left(\mu_P = \hat{p} = 0,45 \,;\ \sigma_P = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{0,45\cdot 0,55}{100}} = 0,0497 \right)$$

a) $Pr(P > 0,50) = 1 - Pr(P \leq 0,50) = 1 - F(0,50) = 1 - 0,8428 = 0,1572$,

donde $F(x)$ es la función de distribución acumulativa de la

$$\mathcal{N}\left(\mu_P = \hat{p} = 0,45 \,;\ \sigma_P = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = 0,0497 \right)$$

no tipificada, ya que EXCEL posibilita su uso:


Microsoft Excel - EjPr_10_7.xls

Archivo Edición Ver Insertar Formato Herramientas Datos

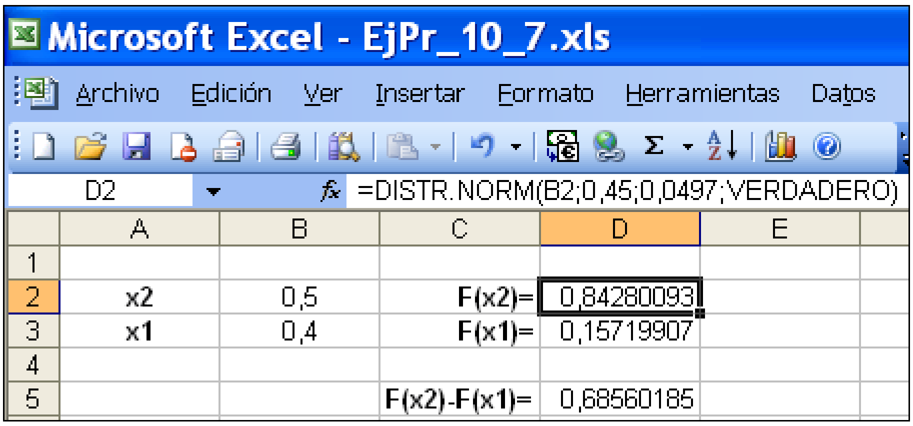
B2 =DISTR.NORM(A2;0,45;0,0497;VERDADERO)

	A	B	C	D	E
1	x	F(x)	1-F(x)		
2	0,5	0,84280093	0,15719907		

b) Suceso equivalente al anterior:

$$Pr(1 - P < 0,50) = Pr(P > 0,50) = 0,1572$$

c) $Pr(0,5 < 1 - P < 0,6) = Pr(-0,6 < P - 1 < -0,5) = Pr(0,4 < P < 0,5) =$
$= F(0,5) - F(0,4) = 0,8428 - 0,1572 = 0,6856$


Microsoft Excel - EjPr_10_7.xls

Archivo Edición Ver Insertar Formato Herramientas Datos

D2 =DISTR.NORM(B2;0,45;0,0497;VERDADERO)

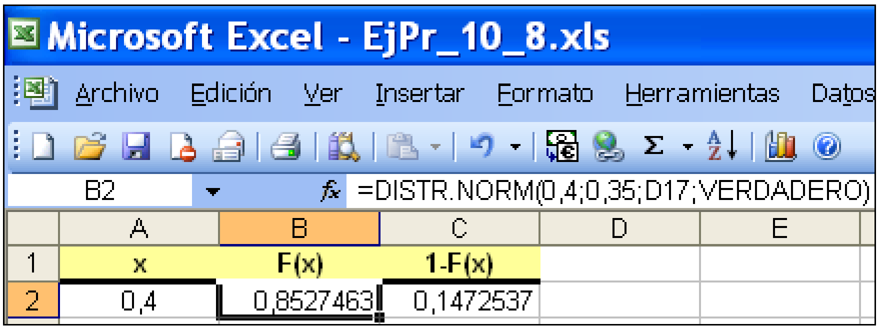
	A	B	C	D	E
1					
2	x2	0,5	F(x2)=	0,84280093	
3	x1	0,4	F(x1)=	0,15719907	
4					
5			F(x2)-F(x1)=	0,68560185	

8 Se sabe que el 35% de los adultos de cierta ciudad es usuario habitual del autobús. Si se elige al azar una muestra de 100 adultos de dicha ciudad, ¿cuál es la probabilidad de que más de 40 sean usuarios habituales del autobús?

Si el valor estimado de la proporción de usuarios es $\hat{p} = 0{,}35$ y el tamaño muestral es $n = 100$, la variable aleatoria P, proporción de usuarios en la población, sigue una ley normal

$$P \sim \mathcal{N}\left(\mu_P = \hat{p} = 0{,}35\ ;\ \sigma_P = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{0{,}35 \cdot 0{,}65}{100}} = 0{,}0477 \right)$$

$$Pr(P > 0{,}40) = 1 - Pr(P \le 0{,}40) = 1 - F(0{,}40) = 1 - 0{,}8527 = 0{,}1473\ .$$



Microsoft Excel - EjPr_10_8.xls

Archivo Edición Ver Insertar Formato Herramientas Datos

B2 =DISTR.NORM(0,4;0,35;D17;VERDADERO)

	A	B	C	D	E
1	x	F(x)	1-F(x)		
2	0,4	0,8527463	0,1472537		

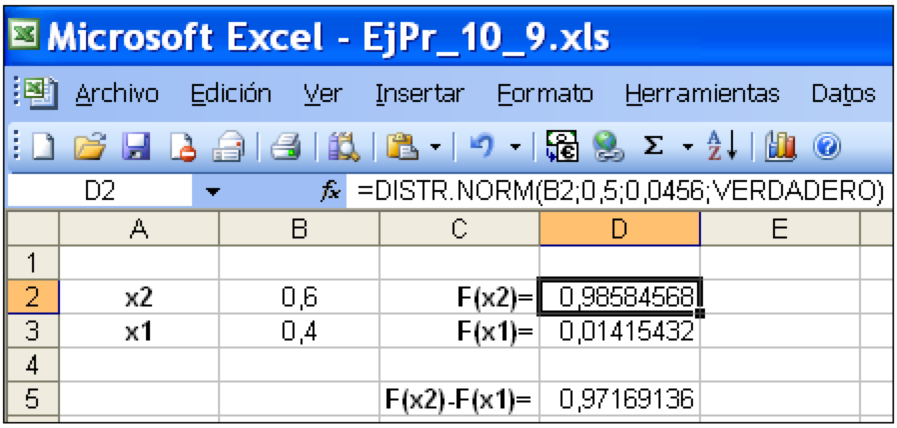
9 Hallar la probabilidad de que en 120 lanzamientos de una moneda no trucada la proporción de caras resultante esté comprendida entre el 40% y el 60%.

Ahora $\hat{p} = 0{,}50$ y el tamaño muestral es $n = 120$, la variable aleatoria P, proporción de caras, sigue una ley normal

$$P \sim \mathcal{N}\left(\mu_P = \hat{p} = 0{,}50\ ;\ \sigma_P = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{0{,}50 \cdot 0{,}50}{120}} = 0{,}0456 \right)$$

Entonces:

$$Pr(0{,}40 < P < 0{,}60) = F(0{,}60) - F(0{,}40) = 0{,}9717$$



Microsoft Excel - EjPr_10_9.xls

Archivo Edición Ver Insertar Formato Herramientas Datos

D2 =DISTR.NORM(B2;0,5;0,0456;VERDADERO)

	A	B	C	D	E
1					
2	x2	0,6	F(x2)=	0,98584568	
3	x1	0,4	F(x1)=	0,01415432	
4					
5			F(x2)-F(x1)=	0,97169136	

10 Los resultados de unas elecciones municipales muestran que un cierto candidato obtuvo el 46% de los votos. Determinar la probabilidad de que, de 1000 individuos elegidos al azar de entre la población votante, se hubiese obtenido una mayoría de votos para dicho candidato.

Ahora $\hat{p} = 0{,}46$ y el tamaño muestral es $n = 1000$, la variable aleatoria P, proporción de votantes al candidato en la población, sigue una ley normal

$$P \sim \mathcal{N}\left(\mu_P = \hat{p} = 0{,}46 \; ; \; \sigma_P = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{0{,}46\cdot 0{,}54}{1000}} = 0{,}0158 \right)$$

$$Pr(P > 0{,}50) = 1 - Pr(P \leq 0{,}50) = 1 - F(0{,}50) = 1 - 0{,}8527 = 0{,}1473 \,.$$