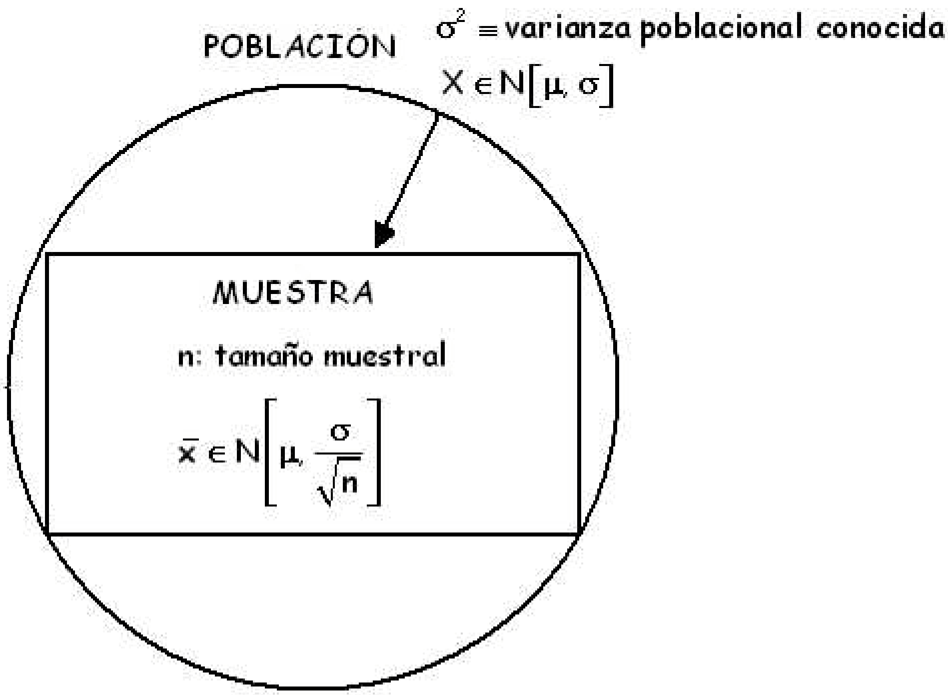
DISTRIBUCIONES EN EL MUESTREO

TEOREMA CENTRAL DEL LIMITE (T.C.L.)

Si $\left[\, X_1, X_2, \,\cdots\, , X_n \right]$ son n variables aleatorias independientes e idénticamente distribuidas con media μ y desviación típica σ, entonces si $n \geq 30$,

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \sim N\left[\mu,\ \frac{\sigma}{\sqrt{n}}\right] \quad \mapsto \quad \sum_{i=1}^{n} x_i \sim N\left[n.\mu,\ \sqrt{n}.\sigma\right]$$

- **DISTRIBUCIÓN DE LA MEDIA MUESTRAL CON VARIANZA POBLACIONAL CONOCIDA**



$$P\left[\bar{x} \geq k\right] \overset{\text{tipificando } N\left[\mu,\ \frac{\sigma}{\sqrt{n}}\right]}{=} P\left[\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{k-\mu}{\frac{\sigma}{\sqrt{n}}}\right] = P\left[z \geq \frac{k-\mu}{\frac{\sigma}{\sqrt{n}}}\right]$$

EJEMPLO.- Se sabe que el peso medio de los jóvenes entre 14 y 18 años sigue una distribución normal con media 50 kg y desviación típica 25 kg.
Para llevar a cabo un estudio del control del peso se seleccionan aleatoriamente 100 jóvenes con edades comprendidas en el intervalo señalado. Si el peso medio muestral está entre 45 y 70 kg se considera que están dentro de los límites normales.
¿Cuál es la probabilidad de que el peso esté fuera de control?

Solución:
Sea la variable aleatoria X = "peso de los jóvenes"

Se trata de una nuestra de una población normal $N\left[50\,,25\right]$, con varianza conocida.

En consecuencia, por el T.C.L., en una muestra de tamaño $n = 100$ la media muestral $\overline{x}$ sigue una distribución normal $N\left[50\,,\frac{25}{\sqrt{100}}\right] \equiv N\left[50\,,2{,}5\right]$

Para que el peso medio muestral se encuentre fuera de control: $\overline{x} < 45$ o $\overline{x} > 70$

$$P\left[\,(\overline{x} < 45) \cup (\overline{x} > 70)\,\right] = P\left[\overline{x} < 45\right] + P\left[\overline{x} > 70\right] = P\left[\frac{\overline{x} - 50}{2{,}5} < \frac{45 - 50}{2{,}5}\right] + P\left[\frac{\overline{x} - 50}{2{,}5} > \frac{70 - 50}{2{,}5}\right] =$$

$$= P\left[z < -2\right] + P\left[z > 8\right] = P\left[z > 2\right] + P\left[z > 8\right] = 0{,}0228 + 0 = 0{,}0228$$

- **DISTRIBUCIÓN DE LA MEDIA MUESTRAL CON VARIANZA POBLACIONAL DESCONOCIDA**

Se define como variable t de Student con n grados de libertad: $t_n = \dfrac{X}{\sqrt{\dfrac{\sum_{i=1}^{n}[X_i]^2}{n}}}$

donde $\left[X_1, X_2, X_3, \cdots, X_n, X\right]$ son $(n+1)$ variables aleatorias $N(0,\sigma)$ independientes entre sí.

En el muestreo, si en una población $N(\mu,\sigma)$ se toman muestras de media $\bar{x}$ y desviación típica σ_x, o desviación estándar (cuasidesviación típica) s_x, la variable $t_{n-1} = \dfrac{\bar{x}-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}}$ es una t de Student con $(n-1)$ grados de libertad.

Adviértase de la relación:

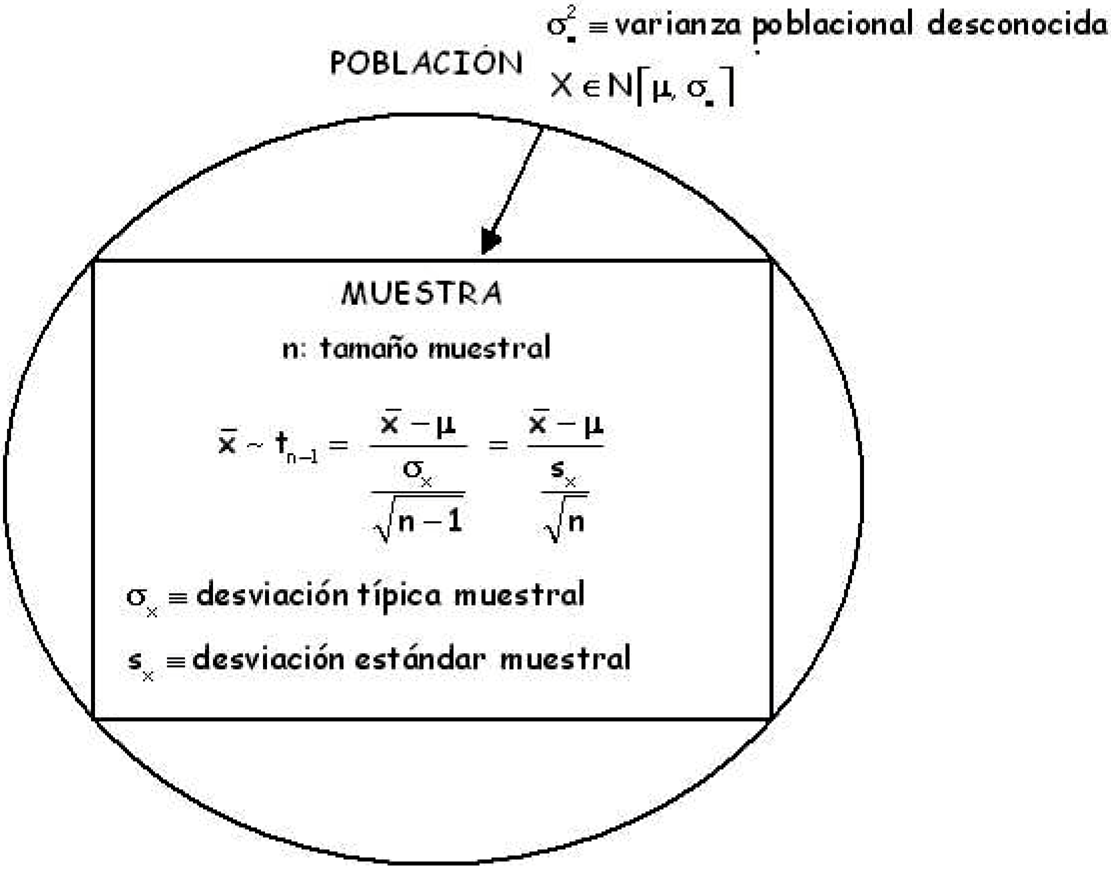
Varianza muestral: $\sigma_x^2 = \dfrac{\sum_{i=1}^{k}(x_i-\bar{x})^2}{n} \Rightarrow \sum_{i=1}^{k}(x_i-\bar{x})^2 = \sigma_x^2 \,.\, n$

Cuasivarianza muestral: $s_x^2 = \dfrac{\sum_{i=1}^{k}(x_i-\bar{x})^2}{n-1} \Rightarrow \sum_{i=1}^{k}(x_i-\bar{x})^2 = s_x^2 \,.\, (n-1)$

$$\sigma_x^2 \,.\, n = s_x^2 \,.\, (n-1) \quad \mapsto \quad \frac{\sigma_x^2}{(n-1)} = \frac{s_x^2}{n}$$

con lo que, $t_{n-1} = \dfrac{\bar{x}-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}} = \dfrac{\bar{x}-\mu}{\dfrac{s_x}{\sqrt{n}}}$

Esta propiedad es muy utilizada en la estimación y en el contraste de hipótesis sobre la media de la población μ



$$P\left[\bar{x} \geq k\right] \overset{t_{n-1}}{=} P\left[\frac{\bar{x}-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}} \geq \frac{k-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}}\right] = P\left[t_{(n-1)} \geq \frac{k-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}}\right] \xrightarrow{n>30} P\left[z \geq \frac{k-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}}\right]$$

EJEMPLO.- La empresa Grano Sol vende galletas ecológicas en paquetes de 60 unidades. Los dueños saben que el peso de cada galleta es una variable aleatoria que tienen una media de 71 gr y una dispersión, medida a través de la desviación típica, de 10 gr. Se pide:

a) ¿Cuál es la probabilidad de que en un paquete de 60 galletas escogidas aleatoriamente, el peso medio de las galletas sea superior a 70 gramos?

b) ¿Cuál sería el resultado si la varianza poblacional fuera desconocida?, suponiendo entonces que la desviación típica muestral es de 5 kg, y una cuasidesviación típica de 5,04.

Solución

a) Sea la variable aleatoria X = "peso de la galleta"

Se trata de una nuestra de una población normal $N\left[71\,,10\right]$, con varianza conocida. En consecuencia, en una muestra de tamaño $n=60$, el peso medio muestral $\bar{x}$ sigue una distribución normal $N\left[71\,,\frac{10}{\sqrt{60}}\right]\equiv N\left[71\,,\,1{,}29\right]$

$$P\left[\bar{x}>70\right]=P\left[\frac{\bar{x}-71}{1{,}29}>\frac{70-71}{1{,}29}\right]=P\left[z>-0{,}77\right]=P\left[z<0{,}77\right]=1-P\left[z>0{,}77\right]=$$
$$=1-0{,}2206=0{,}7794$$

b) En este caso, se trata de una población normal $N\left[71\,,\sigma\right]$ con varianza desconocida, en una muestra de tamaño 60, el peso medio muestral $\bar{x}$ sigue una t de Student con $(n-1)\equiv(60-1)$ grados de libertad, es decir:

$$\bar{x}\sim t_{n-1}=\frac{\bar{x}-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}}\qquad \text{donde } \sigma_x\equiv \text{desviación típica muestral}$$

$$P\left[\bar{x}>70\right]\overset{t_{60-1}}{=}P\left[\frac{\bar{x}-71}{\dfrac{5}{\sqrt{60-1}}}>\frac{70-71}{\dfrac{5}{\sqrt{60-1}}}\right]=P\left[t_{60-1}>\frac{-1}{\dfrac{5}{\sqrt{59}}}\right]=P\left[t_{59}>-1{,}53\right]=$$

$$=P\left[t_{59}<1{,}53\right]\xrightarrow{n>30}P\left[z<1{,}53\right]=1-P\left[z>1{,}53\right]=1-0{,}063=0{,}937$$

- **DISTRIBUCIÓN DE LA VARIANZA MUESTRAL**

La variable χ^2 de Pearson se define: $\chi^2 = X_1^2 + X_2^2 + X_3^2 + \cdots + X_n^2$
donde $\left[X_1, X_2, X_3, \cdots, X_n\right]$ son variables aleatorias $N(0,1)$ independientes entre sí.

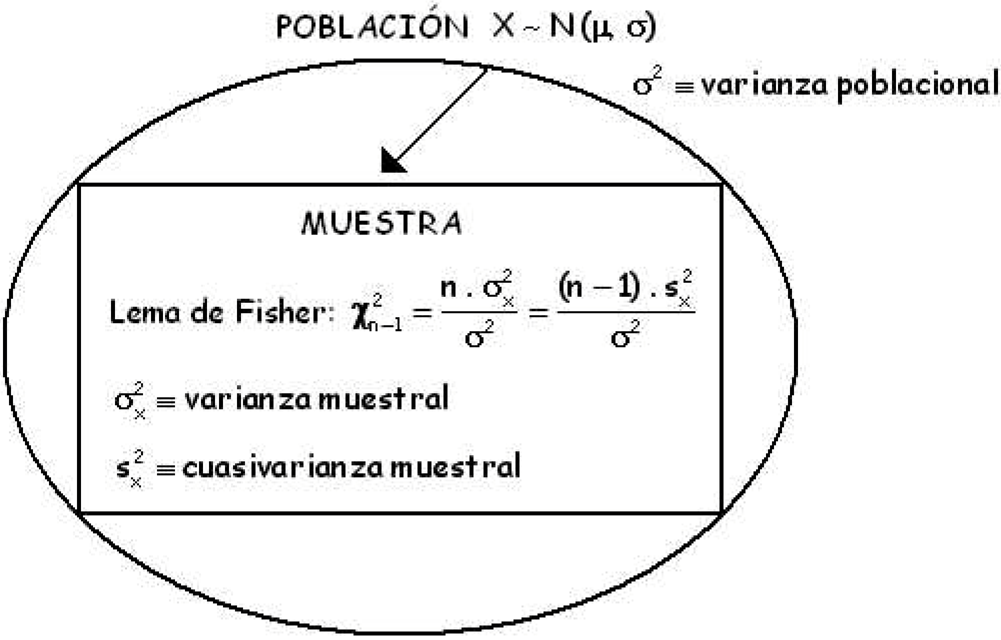
LEMA DE FISHER: En el muestreo, si en una población $N(\mu, \sigma)$ se toman muestras de media $\overline{x}$ y desviación típica σ_x, o desviación estándar (cuasidesviación típica) s_x, la variable $\chi^2_{n-1} = \dfrac{n \, . \, \sigma_x^2}{\sigma^2}$ es una χ^2 de Pearson con $(n-1)$ grados de libertad.

Adviértase de la relación:

Varianza muestral: $\sigma_x^2 = \dfrac{\sum_{i=1}^{k} (x_i - \overline{x})^2}{n} \Rightarrow \sum_{i=1}^{k} (x_i - \overline{x})^2 = \sigma_x^2 \, . \, n$

Cuasivarianza muestral: $s_x^2 = \dfrac{\sum_{i=1}^{k} (x_i - \overline{x})^2}{n-1} \Rightarrow \sum_{i=1}^{k} (x_i - \overline{x})^2 = s_x^2 \, . \, (n-1)$

$\sigma_x^2 \, . \, n = s_x^2 \, . \, (n-1)$, con lo cual: $\chi^2_{n-1} = \dfrac{n \, . \, \sigma_x^2}{\sigma^2} = \dfrac{(n-1) \, . \, s_x^2}{\sigma^2}$



EJEMPLO.- Por datos censales se conoce que la variabilidad en la puntuación en un test mediante la varianza es de 16. No obstante, para analizar la variabilidad en el muestreo se decide tomar una m.a.s. (muestra aleatoria simple) de 20 personas. ¿Cuál es la probabilidad de que la varianza muestral sea mayor que 18?
Nota: Se supone que la calificación del test es una variable normalmente distribuida

Solución

Para el análisis de la varianza muestral se utiliza el estadístico χ^2_{n-1} de Pearson con $(n-1)$ grados de libertad.

La variable aleatoria X = "calificación del test" sigue una distribución normal con varianza poblacional $\sigma^2 = 16$

En la muestra $n = 20$ por el Lema de Fisher: $\chi^2_{n-1} = \dfrac{n\,.\,\sigma^2_x}{\sigma^2}$, se tiene que $\chi^2_{19} = \dfrac{20\,.\,\sigma^2_x}{16}$

con lo cual, $P\left[\sigma^2_x > 18\right] = P\left[\dfrac{20\,.\,\sigma^2_x}{16} > \dfrac{20\,.\,18}{16}\right] = P\left[\chi^2_{19} > 22{,}5\right] = x$ (hay que interpolar)

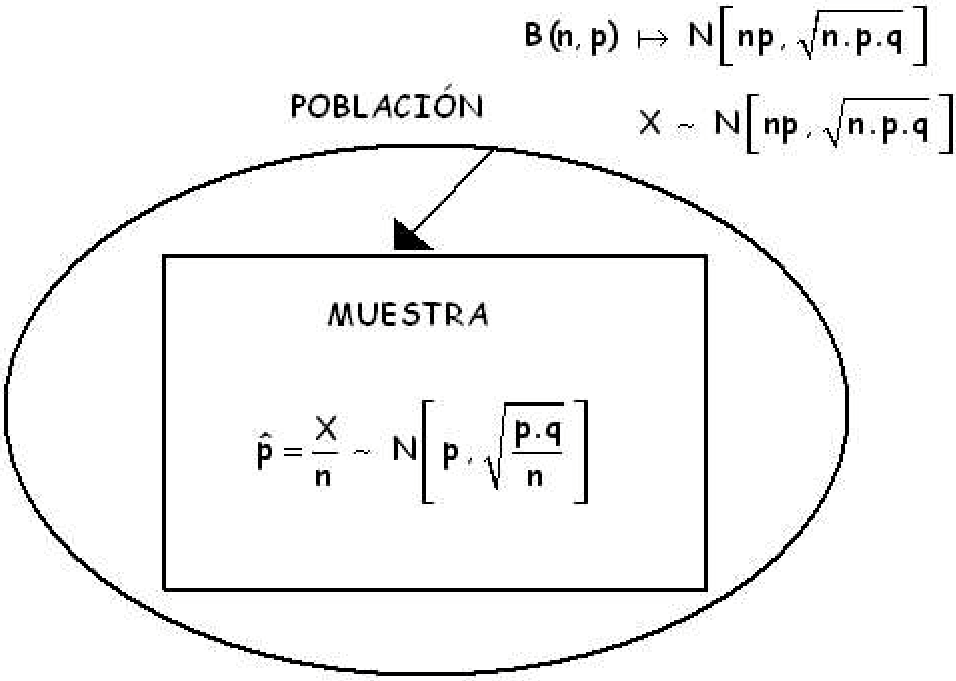
Observando en la tabla de la χ^2 de Pearson con 19 grados de libertad:

$$\frac{10{,}865 - 25{,}989}{0{,}90 - 0{,}10} = \frac{22{,}5 - 25{,}989}{x - 0{,}10} \mapsto \frac{-15{,}124}{0{,}80} = \frac{-3{,}489}{x - 0{,}10} \mapsto 15{,}124\,.\,(x - 0{,}10) = 3{,}489\,.\,0{,}80$$

$$x = 0{,}10 + \frac{3{,}489\,.\,0{,}80}{15{,}124} = 0{,}10 + 0{,}1846 = 0{,}2846$$

con lo que, $P\left[\sigma^2_x > 18\right] = P\left[\dfrac{20\,.\,\sigma^2_x}{16} > \dfrac{20\,.\,18}{16}\right] = P\left[\chi^2_{19} > 22{,}5\right] = 0{,}2846$

- **DISTRIBUCIÓN DE LA PROPORCIÓN MUESTRAL**



EJEMPLO.- Un concesionario vende dos tipos de vehículos, unos de gama alta y otros de gama media. Los coches de gama alta suponen el 30% del total de los coches vendidos. ¿Cuál es la probabilidad de que entre los 100 últimos vehículos vendidos más del 35% sean de gama alta?

Solución

La variable poblacional X = "venta de coches de gama alta" es una variable binomial $b(100, 0{,}3)$ que sigue aproximadamente una distribución normal $X \sim N\left[np, \sqrt{n.p.q}\,\right]$

$$\hat{p} = \frac{X}{n} \xrightarrow{\text{Teorema Central Límite}} \hat{p} \sim N\left[p, \sqrt{\frac{p.q}{n}}\right]$$

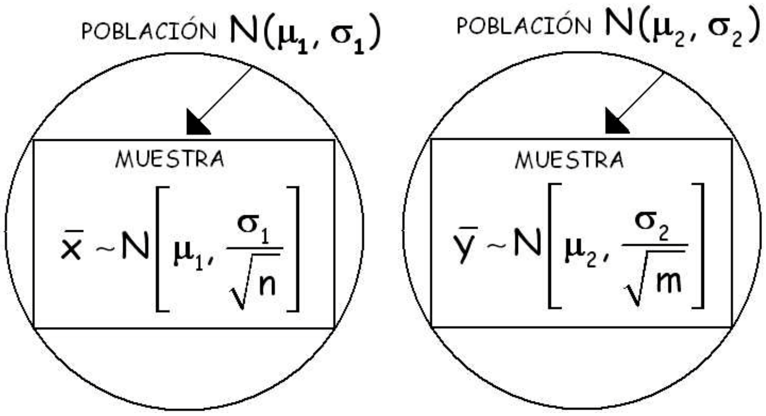
Con los datos del ejercicio:

$$(n = 100,\ p = 0{,}3,\ q = 0{,}7): \quad \hat{p} \sim N\left[0{,}3, \sqrt{\frac{0{,}3 \,.\, 0{,}7}{100}}\right] = \hat{p} \sim N\left[0{,}3,\ 0{,}0458\right]$$

$$P\left[\hat{p} > 0{,}35\right] \overset{T.C.L}{=} P\left[\frac{\hat{p} - 0{,}3}{0{,}0458} > \frac{0{,}35 - 0{,}3}{0{,}0458}\right] = P\left[z > 1{,}09\right] = 0{,}1379$$

- **DISTRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES DE DISTINTAS POBLACIONES**

Cuando se consideran dos poblaciones $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$, y de cada una de ellas se extrae una muestra aleatoria simple, de la primera población de tamaño n $(x_1, x_2, \cdots, x_n)$ y de la segunda de tamaño m $(y_1, y_2, \cdots, y_m)$ independiente de la primera:



Se considera como estadístico la diferencia de las medias muestrales $\left[\bar{x} - \bar{y}\right]$, combinación lineal de $(n + m)$ variables aleatorias normales e independientes, por lo que su distribución será normal,

media: $E\left[\bar{x} - \bar{y}\right] = E\left[\bar{x}\right] - E\left[\bar{y}\right] = \mu_1 - \mu_2$

varianza: $V\left[\bar{x} - \bar{y}\right] = V\left[\bar{x}\right] + V\left[\bar{y}\right] = \dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}$

con lo cual, $(\bar{x} - \bar{y}) \sim N\left[\mu_1 - \mu_2 \, , \sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}\right]$

EJEMPLO.- Se desea analizar las diferencias de las calificaciones entre dos grupos de alumnos. Unos proceden del Grupo b1 y otros del Grupo b2. Para analizar la distribución en el muestreo de la diferencia de medias se toman muestras aleatorias independientes de ambas poblaciones, obteniéndose la siguiente tabla:

	Tamaño población	Tamaño muestra	Media población	Media muestra	Desviación típica población	Desviación típica muestra
Grupo b1	200	100	4,10	4,2153	1,55	1,5635
Grupo b2	150	75	5,18	5,3247	1,95	1,8238

¿Cuál es la probabilidad de que la diferencia de medias muestrales sea mayor que uno?

Solución:

v.a. X = "calificación Grupo b1" $\bar{x} \sim N\left[\mu_1 \; ; \; \dfrac{\sigma_1}{\sqrt{n}}\right] \equiv N\left[4,10 \; ; \; \dfrac{1,55}{\sqrt{100}}\right] \equiv N\left[4,10 \; ; \; 0,155\right]$

v.a. Y = "calificación Grupo b2" $\bar{y} \sim N\left[\mu_2 \; ; \; \dfrac{\sigma_2}{\sqrt{m}}\right] \equiv N\left[5,18 \; ; \; \dfrac{1,95}{\sqrt{75}}\right] \equiv N\left[5,18 \; ; \; 0,225\right]$

El estadístico $(\bar{x} - \bar{y}) \sim N\left[\mu_1 - \mu_2 \, , \, \sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}\right]$, con lo cual

$$(\bar{x} - \bar{y}) \sim N\left[4,10 - 5,18 \; ; \; \sqrt{0,155^2 + 0,225^2}\right] \equiv N\left[-1,08 \; ; \; 0,2732\right]$$

- $$P\left[\,|\bar{x} - \bar{y}| > 1\,\right] = P\left[\big((\bar{x} - \bar{y}) > 1\big) \cup \big((\bar{x} - \bar{y}) < -1\big)\right] = P\left[(\bar{x} - \bar{y}) > 1\right] + P\left[(\bar{x} - \bar{y}) < -1\right] =$$

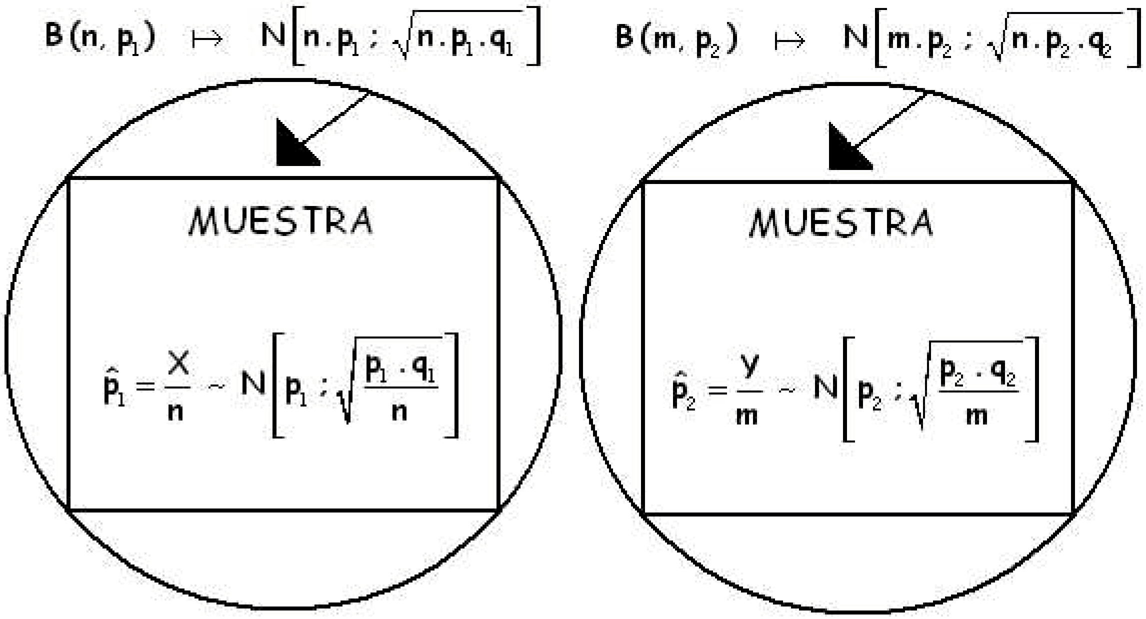
$$= P\left[\frac{(\bar{x} - \bar{y}) + 1,08}{0,2732} > \frac{1 + 1,08}{0,2732}\right] + P\left[\frac{(\bar{x} - \bar{y}) + 1,08}{0,2732} < \frac{-1 + 1,08}{0,2732}\right] =$$

$$= P\left[z > 7,61\right] + P\left[z < 0,2928\right] = 0 + \big(1 - P\left[z > 0,2928\right]\big) = 1 - 0,3859 = 0,6141$$

- $$P\left[\,|\bar{x} - \bar{y}| > 1\,\right] = 1 - P\left[\,|\bar{x} - \bar{y}| < 1\,\right] = 1 - P\left[-1 < (\bar{x} - \bar{y}) < 1\right] =$$

$$= 1 - P\left[\frac{-1 + 1,08}{0,2732} < \frac{(\bar{x} - \bar{y}) + 1,08}{0,2732} < \frac{1 + 1,08}{0,2732}\right] = 1 - P\left[0,2928 < z < 7,61\right] = 0,6141$$

- **DISTRIBUCIÓN DE LA DIFERENCIA DE PROPORCIONES DE DISTINTAS POBLACIONES**



El estadístico diferencia de las proporciones muestrales $\left[\hat{p}_1 - \hat{p}_2\right]$, combinación lineal de $(n+m)$ variables aleatorias normales e independientes

media: $E\left[\hat{p}_1 - \hat{p}_2\right] = E\left[\hat{p}_1\right] - E\left[\hat{p}_2\right] = p_1 - p_2$

varianza: $V\left[\hat{p}_1 - \hat{p}_2\right] = V\left[\hat{p}_1\right] + V\left[\hat{p}_2\right] = \dfrac{p_1 \cdot q_1}{n} + \dfrac{p_2 \cdot q_2}{m}$

con lo cual, $(\hat{p}_1 - \hat{p}_2) \sim N\left[(p_1 - p_2)\,,\, \sqrt{\dfrac{p_1 \cdot q_1}{n} + \dfrac{p_2 \cdot q_2}{m}}\right]$

EJEMPLO.- Según los resultados de un estudio de la población, un 80% de las mujeres entrevistadas afirman utilizar algún producto cosmético todos los días, mientras que en el caso de los hombres este porcentaje es del 55%.
Una pequeña firma de cosmética se plantea sacar al mercado una crema hidratante de uso específico para hombres. Antes de crear esta nueva línea de negocio decide realizar su propia encuesta sobre una pequeña muestra aleatoria, para ello selecciona a 50 mujeres y a 60 hombres y les pregunta sobre sus hábitos cosméticos.
Calcule la probabilidad de que la diferencia entre la proporción de mujeres que utiliza cosméticos respecto a la proporción de hombres que los utiliza sea inferior al 20%.

Solución:

Sean las variables aleatorias:

- X = "mujeres que utilizan algún producto cosmético" $p_1 = 0,8 \quad q_1 = 0,2$

  tamaño muestra $n = 50$: $\hat{p}_1 = \dfrac{X}{n} \sim N\left[0,8\ ;\ \sqrt{\dfrac{0,8\ .\ 0,2}{50}}\right]$

- Y = "hombres que utilizan algún producto cosmético" $p_2 = 0,55 \quad q_2 = 0,45$

  tamaño muestra $m = 60$: $\hat{p}_2 = \dfrac{Y}{m} \sim N\left[0,55\ ;\ \sqrt{\dfrac{0,55\ .\ 0,45}{60}}\right]$

- El estadístico diferencia proporciones $(\hat{p}_1 - \hat{p}_2) \sim N\left[(p_1 - p_2)\ ,\ \sqrt{\dfrac{p_1 . q_1}{n} + \dfrac{p_2 . q_2}{n}}\right]$

con lo cual,

$$(\hat{p}_1 - \hat{p}_2) \sim N\left[(0,8 - 0,55)\ ,\ \sqrt{\dfrac{0,8\ .\ 0,2}{50} + \dfrac{0,55\ .\ 0,45}{60}}\right] \equiv N\left[0,25\ ;\ 0,0856\right]$$

$$P\left[(\hat{p}_1 - \hat{p}_2) < 0,20\right] = P\left[\dfrac{(\hat{p}_1 - \hat{p}_2) - 0,25}{0,0856} < \dfrac{0,20 - 0,25}{0,0856}\right] = P\left[z < -0,58\right] =$$

$$= P\left[z > 0,58\right] = 0,2810$$

Como la $P\left[(\hat{p}_1 - \hat{p}_2) > 0,20\right] = 1 - P\left[(\hat{p}_1 - \hat{p}_2) < 0,20\right] = 1 - 0,2810 = 0,719$ es probable que se aconsejara sacar el producto del mercado.

EJERCICIOS DE DISTRIBUCIONES MUESTRALES

Ejercicio 1.- Una fábrica de pasteles elabora, en su producción habitual, un 3% de pasteles defectuosos. Un cliente recibe un pedido de 500 pasteles de la fábrica. Calcular la probabilidad de que encuentre más de un 5% de pasteles defectuosos.

Solución:

Considerando que la distribución muestral de una proporción se ajusta a una distribución normal $N\left(p\,,\sqrt{\dfrac{pq}{n}}\right)$: $\quad \hat{p} \sim N\left(0{,}03\,;\sqrt{\dfrac{0{,}03\,.\,0{,}97}{500}}\right) \equiv N\big(0{,}03\,;\ 0{,}0076\big)$

$$P\big[\hat{p} > 0{,}05\big] \overset{T.C.L}{=} P\left[\frac{\hat{p}-0{,}03}{0{,}0076} > \frac{0{,}05-0{,}03}{0{,}0076}\right] = P\big[z > 2{,}63\big] = 0{,}00427$$

Ejercicio 2.- Un ascensor limita el peso de sus cuatro ocupantes a 300 kilogramos. Si el peso de una persona sigue una distribución normal $N\big(71\,;\ 7\big)$, calcular la probabilidad de que el peso 4 personas supere los 300 kilogramos.

Solución:

Si el peso de una persona sigue una distribución normal $N\big(71, 7\big)$, la muestra de 4 personas sigue una distribución normal $N\left(71\,;\ \dfrac{71}{\sqrt{4}}\right) \equiv N\big(71\,;\ 3{,}5\big)$

$$P\big[(X_1+X_2+X_3+X_4) > 300\big] = P\left[\frac{X_1+X_2+X_3+X_4}{4} > \frac{300}{4}\right] = P\big[\,\bar{x} > 75\big] =$$

$$= P\big[\,\bar{x} > 75\big] = P\left[\frac{\bar{x}-71}{3{,}5} > \frac{75-71}{3{,}5}\right] = P\big[\,z > 1{,}143\big] = 0{,}1265$$

Interpolando: $\dfrac{0{,}1271-0{,}1251}{1{,}14-1{,}15} = \dfrac{x-0{,}1251}{1{,}143-1{,}15} \quad \mapsto \quad \dfrac{0{,}002}{-0{,}01} = \dfrac{x-0{,}1251}{-0{,}007}$

$$x = 0{,}1251 + \frac{0{,}002\,.\,0{,}007}{0{,}01} = 0{,}1265$$

Ejercicio 3.- Se prueba el rendimiento, en kilómetros por litro, de dos tipos de gasolina, encontrándose una desviación de 1,23 kilómetros por litro para la primera gasolina y una desviación de 1,37 kilómetros por litro para la segunda gasolina. La primera gasolina se prueba en 35 vehículos y la segunda gasolina en 42 vehículos.

a) ¿Cuál es la probabilidad de que la primera gasolina dé un rendimiento promedio mayor de 0,45 kilómetros por litro que la segunda gasolina?

b) ¿Cuál es la probabilidad de que la diferencia en rendimiento promedio se encuentre entre 0,65 y 0,83 kilómetros por litro a favor de la primera gasolina?

Solución:

a) Es una situación de dos poblaciones normales, donde se supone igual media poblacional, $N(\mu\,;\,1{,}23)$ y $N(\mu\,;\,1{,}37)$

Considerando como estadístico la diferencia de medias muestrales, siendo el tamaño de la primera muestra $n = 35$ y de la segunda muestra $m = 42$

$$(\bar{x} - \bar{y}) \sim N\left[\mu - \mu\,;\,\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right] \equiv N\left[0\,;\,\sqrt{\frac{1{,}23^2}{35} + \frac{1{,}37^2}{42}}\right] \equiv N\left[0\,;\,0{,}296\right]$$

$$P\left[(\bar{x} - \bar{y}) > 0{,}45\right] = P\left[\frac{(\bar{x} - \bar{y}) - 0}{0{,}296} > \frac{0{,}45 - 0}{0{,}296}\right] = P\left[z > 1{,}52\right] = 0{,}0643$$

b) $$P\left[0{,}65 < (\bar{x} - \bar{y}) < 0{,}83\right] = P\left[\frac{0{,}65 - 0}{0{,}296} < \frac{(\bar{x} - \bar{y}) - 0}{0{,}296} < \frac{0{,}83 - 0}{0{,}296}\right] = P\left[2{,}19 < z < 2{,}80\right] =$$

$$= P\left[z > 2{,}19\right] - P\left[z > 2{,}80\right] = 0{,}0143 - 0{,}00256 = 0{,}01174$$

Ejercicio 4.- De una población normal $N(10\,;\,4)$ se toman muestras aleatorias simples de tamaño 100, calcular la probabilidad que la media muestral sea mayor que 10,1.

Solución:

La media muestral sigue una distribución normal $N\left(10\,;\,\dfrac{4}{\sqrt{100}}\right) \equiv N(10\,;\,0{,}4)$

$$P(\bar{x} > 10{,}1) = P\left(\frac{\bar{x} - 10}{0{,}4} > \frac{10{,}1 - 10}{0{,}4}\right) = P(z > 0{,}25) = 0{,}4013$$

Ejercicio 5.- De la distribución $N(5\ ;\ 2)$ se toma una muestra aleatoria simple de tamaño 20, y de la distribución $N(-4\ ;\ 10)$ otra muestra aleatoria simple de tamaño 25 e independiente de la primera. Calcular $P(\overline{x}-\overline{y}\leq 15)$

Solución:

Considerando como estadístico la diferencia de medias muestrales, siendo el tamaño de la primera muestra $n=20$ y de la segunda muestra $m=25$

$$(\overline{x}-\overline{y}) \sim N\left[\mu_1-\mu_2\ ;\ \sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right] \equiv N\left[5-(-4)\ ;\ \sqrt{\frac{2^2}{20}+\frac{10^2}{25}}\right] \equiv N\left[9\ ;\ 2{,}05\right]$$

$$P\left[(\overline{x}-\overline{y})<15\right]=P\left[\frac{(\overline{x}-\overline{y})-9}{2{,}05}<\frac{15-9}{2{,}05}\right]=P\left[z<2{,}93\right]=1-P\left[z>2{,}93\right]=1-0{,}00169=0{,}9983$$

Ejercicio 6.- En cierta población, en una muestra de tamaño 100, la media muestral $\overline{x}$ de una característica sigue una distribución normal. Se conoce que $P(\overline{x}\leq 75)=0{,}58$ y $P(\overline{x}>80)=0{,}04$. Hallar la media y la desviación típica poblacional.

Solución:

La media muestral $\overline{x}\sim N\left(\mu\ ;\ \frac{\sigma}{\sqrt{n}}\right)\equiv N\left(\mu\ ;\ \frac{\sigma}{\sqrt{100}}\right)\equiv N\left(\mu\ ;\ \frac{\sigma}{10}\right)$

Se desconocen los parámetros poblacionales μ y σ, por lo que se necesitan dos ecuaciones:

$$P(\overline{x}\leq 75)\overset{TCL}{=}P\left(\frac{\overline{x}-\mu}{\frac{\sigma}{10}}\leq\frac{75-\mu}{\frac{\sigma}{10}}\right)=0{,}58 \quad\mapsto\quad P\left(\frac{\overline{x}-\mu}{\frac{\sigma}{10}}\geq\frac{75-\mu}{\frac{\sigma}{10}}\right)=0{,}42 \quad\mapsto\quad \frac{75-\mu}{\frac{\sigma}{10}}=0{,}20$$

$$P(\overline{x}>80)\overset{TCL}{=}P\left(\frac{\overline{x}-\mu}{\frac{\sigma}{10}}>\frac{80-\mu}{\frac{\sigma}{10}}\right)=0{,}04 \quad\mapsto\quad \frac{80-\mu}{\frac{\sigma}{10}}=1{,}75$$

De las ecuaciones $\begin{cases} 750 - 10\,\mu = 0{,}20\ .\ \sigma \\ 800 - 10\ \mu = 1{,}75\ .\ \sigma \end{cases} \Rightarrow \quad 50 = 1{,}55\ .\ \sigma \quad \mapsto \quad \begin{bmatrix} \sigma = 32{,}26 \\ \mu = 74{,}35 \end{bmatrix}$

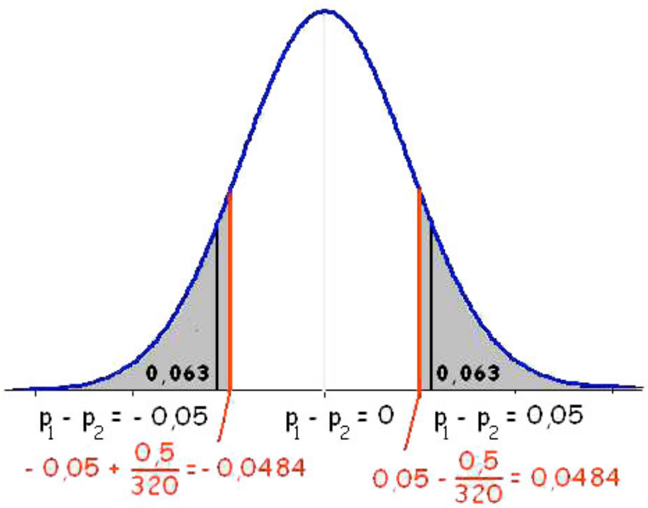
Ejercicio 7.- Una encuesta del periódico constó de 320 trabajadores de una multinacional que fueron despedidos entre 2010-2013, encontrando que el 20% habían estado sin trabajo por lo menos durante dos años.
Si el periódico tuviera que seleccionar otra muestra aleatoria de 320 empleados despedidos entre 2010-2013. ¿Cuál sería la probabilidad de que el porcentaje muestral de trabajadores sin empleo durante por lo menos dos años, difiera del porcentaje obtenido en la primera encuesta, en un 5% o más?.

Solución:

En la situación descrita hay una sola población, de la que se extraen dos muestras, queriendo conocer la diferencia de porcentajes de esas dos muestras, por lo que se debe utilizar la distribución muestral de diferencias de proporciones con $p_1 = p_2$ al tratarse de una misma población.

Otra cuestión a plantearse es que se desconoce la proporción poblacional $p_1 = p_2$, sólo se conoce que la proporción muestral en 320 trabajadores es $\hat{p}_1 = 0{,}20$.
Para ello, se utiliza la proporción muestral $\hat{p}_1 = 0{,}20$ como una estimación puntual de la proporción poblacional p_1, es decir, $p_1 = E\left[\hat{p}_1\right] = 0{,}20$

Analizando la pregunta: ¿Cuál sería la probabilidad de que el porcentaje muestral de trabajadores sin empleo durante por lo menos dos años, difiera del porcentaje obtenido en la primera encuesta, en un 5% o más?. Con diferir debe entenderse que la diferencia puede ser a favor de la muestra uno, o a favor de la muestra dos, por lo que se deben calcular dos áreas en la distribución y sumarlas.



$n = m = 320 \qquad p_1 - p_2 = 0$

$$(\hat{p}_1 - \hat{p}_2) \sim N\left[(p_1 - p_2)\,,\ \sqrt{\frac{p_1 \cdot q_1}{n} + \frac{p_2 \cdot q_2}{m}}\right]$$

$$\sqrt{\frac{p_1 \cdot q_1}{n} + \frac{p_2 \cdot q_2}{m}} = \sqrt{\frac{0{,}2\ .\ 0{,}8}{320} + \frac{0{,}2\ .\ 0{,}8}{320}} = 0{,}0316$$

$$(\hat{p}_1 - \hat{p}_2) \sim N\left[0\ ;\ 0{,}316\right]$$

$$P\left[\left|\hat{p}_1-\hat{p}_2\right|\geq 0,05\right]=P\left[\left(\hat{p}_1-\hat{p}_2\geq 0,05\right)\cup\left(\hat{p}_1-\hat{p}_2\leq -0,5\right)\right]=$$

$$=P\left[\left(\hat{p}_1-\hat{p}_2\right)\geq 0,05\right]+P\left[\left(\hat{p}_1-\hat{p}_2\right)\leq -0,05\right]=\text{(factor corrección)}=$$

$$=P\left[\left(\hat{p}_1-\hat{p}_2\right)\geq 0,05-\frac{0,5}{320}\right]+P\left[\left(\hat{p}_1-\hat{p}_2\right)\leq -0,05+\frac{0,5}{320}\right]=$$

$$=P\left[\left(\hat{p}_1-\hat{p}_2\right)\leq -0,0484\right]+P\left[\left(\hat{p}_1-\hat{p}_2\right)\geq 0,0484\right]=$$

$$=P\left[\frac{\left(\hat{p}_1-\hat{p}_2\right)-0}{0,0316}\leq\frac{-0,0484-0}{0,0316}\right]+P\left[\frac{\left(\hat{p}_1-\hat{p}_2\right)-0}{0,0316}\geq\frac{0,0484-0}{0,0316}\right]=$$

$$=P\left[\frac{0,0484-0}{0,0316}\leq\frac{\left(\hat{p}_1-\hat{p}_2\right)-0}{0,0316}\leq\frac{-0,0484-0}{0,0316}\right]=P\left[1,53\leq z\right]+P\left[z\leq -1,53\right]=$$

$$=2\ P\left[z\geq 1,53\right]=2\,.\,0,0630=0,1260$$