DISTRIBUCIÓN MUESTRAL DE MEDIAS.

Si tomamos una muestra de tamaño “n” de una población, obtendremos una media para esa muestra y una desviación típica que no tienen por qué ser exactamente iguales a las de la población completa pero se parecerán, tanto más cuanto mayor sea el tamaño de la muestra.

Cada muestra de tamaño “n” que podemos tomar de una población proporciona una media. Si tomamos varias muestras diferentes y consideramos cada una sus medias como una nueva variable, podemos estudiar su denominada **distribución muestral de medias**.

Si la población sigue una distribución normal **N (μ, σ)** y extraemos de ella muestras de tamaño n, la distribución muestral de medias sigue una distribución también normal (independientemente del tamaño de la muestra) de la forma: $\boxed{N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$

TEOREMA CENTRAL DEL LÍMITE.

Aunque la población no siga una distribución Normal, si el tamaño de la muestra es suficientemente grande podemos utilizar este Teorema que dice:

Si tomamos una muestra aleatoria **de tamaño n>30** de una población con una distribución cualquiera de media μ y desviación típica σ, la **distribución muestral de medias** se aproxima a una distribución normal $\boxed{N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$

Como se puede apreciar:

-La media de la distribución muestral es la misma que la de la población.

-La desviación típica de la distribución muestral es la de la muestra dividida por la raíz del número de datos.

INTERVALOS DE CONFIANZA.

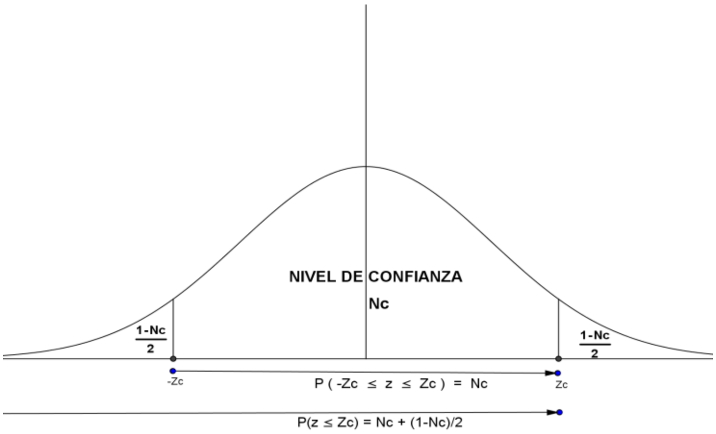
Llamamos intervalo de confianza al intervalo que, con una cierta probabilidad, contenga al parámetro que se está estimando (normalmente la media). A esta probabilidad se le llama nivel de confianza $\mathbf{N_c}$.

El intervalo debe tener el centro en la media de la muestra y una amplitud tal que la probabilidad de que la media poblacional esté en ese intervalo sea el Nivel de confianza requerido.

1º.- Utilizando la curva normal estándar N(0,1), calcularemos el valor de la variable tipificada **z-crítica (z_c)** de tal forma que:

$$P(-z_c \le z \le z_c) = N_c$$

$$P(z \le z_c) = N_c + \frac{1-N_c}{2}$$



2º.- Obtenemos un intervalo en la variable z en el cual tenemos una confianza N_c de que estará la media buscada.

$$z \in (-z_c\ ,\ z_c)$$

3º.- Traspasamos los extremos del intervalo tipificado (con "z") a la variable buscada μ .

$$\left.\begin{array}{l} \dfrac{\mu-\bar{x}}{\sigma/\sqrt{n}} = -z_c \quad \Rightarrow \quad \mu = \bar{x} - z_c \dfrac{\sigma}{\sqrt{n}} \\ \dfrac{\mu-\bar{x}}{\sigma/\sqrt{n}} = z_c \quad \Rightarrow \quad \mu = \bar{x} + z_c \dfrac{\sigma}{\sqrt{n}} \end{array}\right\} \quad \Rightarrow \quad \boxed{\mu \in \left(\bar{x} - z_c \dfrac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_c \dfrac{\sigma}{\sqrt{n}}\right)}$$

Error máximo cometido.

La media de la población se encontrará (con nivel de confianza N_c) dentro de ese intervalo pero no podemos asegurar dónde, lo que si podemos asegurar es que lo más alejada que puede estar del centro del intervalo es cuando esté en uno de sus extremos.

Si decimos que la media es justo el centro del intervalo el error máximo que podemos cometer será **la distancia del centro a cada uno de los extremos**, es decir:

$$E = z_c \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \boxed{E < z_c \frac{\sigma}{\sqrt{n}}}$$

Tamaño de la muestra.

Si en un intervalo de confianza para la media del tipo $\left(\bar{x} - z_c \dfrac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_c \dfrac{\sigma}{\sqrt{n}}\right)$ queremos calcular el tamaño de la muestra para obtener ese intervalo, podemos hacerlo sin mas que despejar en el Error.

$$E = z_c \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad n = \frac{z_c^{\ 2} \cdot \sigma^2}{E^2} \quad \Rightarrow \quad \boxed{n > \frac{z_c^{\ 2} \cdot \sigma^2}{E^2}}$$

No hace falta que te aprendas de memoria el cálculo de "n", despeja en la fórmula del error.

EJEMPLO1. Se selecciona al azar una muestra de 100 personas cuya media de ingresos anual fue de 10675€ y la desviación típica de 2000€. Halla el intervalo de confianza para la media de la población con un nivel de confianza del 92%.

Los datos a tener en cuenta son:

Tamaño de la muestra: n=100

Media y desviación típica de la muestra : $\bar{x} = 10675$; $\sigma = 2000$

Nivel de confianza: N_c = 92%

Para calcular la z_c que corresponde al 92% de probabilidad $P(-z_c \le z \le z_c) = 0{,}92$

Esto es lo mismo que $P(z \le z_c) = 0{,}96$ lo que corresponde a una z_c=1,75

El intervalo que buscamos es $\left(\bar{x} - z_c \dfrac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_c \dfrac{\sigma}{\sqrt{n}}\right) = \left(10675 - 1{,}75\dfrac{2000}{10}\ ,\ 10675 + 1{,}75\dfrac{2000}{10}\right)$

Esto es: $\mu \in (10325\ ,\ 11025)$

Sol: Los ingresos medios de la población están entre 10325 € y 11025 € con una confianza del 92%

EJEMPLO 2. Una fábrica de neumáticos asegura que su modelo "Gomático MXV" tiene una duración que se distribuye según una distribución normal, con una desviación típica de 5000 km. La OCU realiza un estudio de estos neumáticos tomando muestras de tamaño 25.

a) ¿Cuál es la probabilidad de que la duración media de una muestra diste de la media poblacional menos de 500 km?

$$E = z_c \frac{\sigma}{\sqrt{n}} \Rightarrow 500 = z_c \frac{5000}{\sqrt{25}} \Rightarrow z_c = 0{,}5$$

Necesito averiguar por lo tanto la probabilidad de que $z \in (-0{,}5\ ,\ +0{,}5)$

$$P(-0{,}5 \leq z \leq 0{,}5) = P(z \leq 0{,}5) - P(z \leq -0{,}5) = 0{,}6915 - [1 - 0{,}6915] = 0{,}3830$$

b) ¿Cuál es la probabilidad de que la duración media de una muestra diste de la media poblacional menos de 2000 km?

La solución debe ser 0,9544, que se puede apreciar es mucho mayor que en el apartado a). Lógico porque el margen de error es mayor

c) ¿Cuánto dista la media muestral y la media de la población en el 95% de todas las muestras posibles de tamaño 25?

$$P(-z_c \leq z \leq z_c) = 0{,}95 \Rightarrow P(z \leq z_c) = 0{,}975 \Rightarrow z_c = 1{,}96$$

La mitad de la anchura del intervalo de confianza al 95% será el máximo de separación entre las medias poblacional y de la muestra, esto es el Error máximo admitido. $E < 1{,}96 \frac{5000}{\sqrt{25}} \Rightarrow E < 1960\ km.$ que es el valor que buscábamos.

Obsérvese que en este ejercicio ni siquiera conocemos la media de la población ni la media de la muestra. No nos ha hecho falta para responder a las preguntas anteriores, pero si la hubiéramos necesitado para averiguar un intervalo de confianza para la media de la población (duración media de un neumático de esta marca).

DISTRIBUCIÓN MUESTRAL DE PROPORCIONES. LA BINOMIAL

En numerosas ocasiones se plantea estimar una proporción, probabilidad o porcentaje. Esto ocurre en los casos en los que la variable aleatoria de estudio pueda tomar solamente dos valores diferentes: sí (éxito) o no (fracaso) ; Votantes a un partido o No votantes a ese partido; mayor de 18 años o menor de 18 años; etc.

En estos casos afirmamos que la respectiva población sigue una **DISTRIBUCIÓN BINOMIAL**.

Cuando la población es grande (que será lo habitual) la distribución se aproxima a una **NORMAL** que ya conocemos. Supondremos que esto es válido para n>30.

Llamamos "p" al parámetro poblacional que es la proporción o probabilidad de uno de los valores que determina la variable, por tanto el otro parámetro que llamaremos "q" tendrá que cumplir que:

q=1-p

Las relaciones que existen entre los parámetros de la población y los estadísticos de la distribución muestral de proporciones son los siguientes:

Media muestral $\Rightarrow$ $\mu = p$

Desviación típica $\Rightarrow$ $\sigma = \sqrt{pq}$

Para muestras de tamaño n>30 , la distribución muestral de proporciones sigue una distribución normal $N\left(p\ ,\ \dfrac{\sqrt{pq}}{\sqrt{n}}\right)$

A partir de este momento todo el estudio puede hacerse exactamente igual que en la distribución muestral de medias.

EJEMPLO 3. Para estimar la proporción de estudiantes de una universidad que está a favor de la reinserción social del delincuente, se entrevistó a 500 estudiantes. El 58% estaba a favor. Halla el intervalo de confianza, con un nivel del 95%, en el cual se hallará la población que se encuentra a favor.

Los datos a tener en cuenta son:

Tamaño de la muestra: n=500

Proporciones de la muestra: $\begin{cases} Favorables\ \ p = 0{,}58\ \ ;\ \ En\ contra\ \ \ q = 0{,}42 \\ \sigma = \sqrt{pq} = \sqrt{0{,}58 \cdot 0{,}42} = 0{,}4936 \end{cases}$

Nivel de confianza: $N_c = 95\%$

Para calcular la z_c que corresponde al 95% de probabilidad $P(-z_c \le z \le z_c) = 0{,}95$

Esto es lo mismo que $P(z \le z_c) = 0{,}975$ lo que corresponde a una z_c=1,96

El intervalo que buscamos es:

$$\left(\bar{x} - z_c\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_c\frac{\sigma}{\sqrt{n}}\right) = \left(p - z_c\frac{\sqrt{pq}}{\sqrt{n}}\ ,\ p + z_c\frac{\sqrt{pq}}{\sqrt{n}}\right) = \left(0{,}58 - 1{,}96\frac{0{,}4936}{\sqrt{500}}\ ,\ 0{,}58 + 1{,}96\frac{0{,}4936}{\sqrt{500}}\right)$$

El intervalo buscado para la proporción que está a favor es: $(0.537\ ,\ \ 0.623)$

Sol: Entre un 53,7% y un 62,35 de la población está a favor de la reinserción social con un nivel de confianza del 95%

CUADRO RESUMEN PARA INTERVALOS DE CONFIANZA.

Para muestras n>30	Distribución muestral de medias.	Distribución muestral de proporciones
Distribución Normal	$N\left(\mu\ ,\ \frac{\sigma}{\sqrt{n}}\right)$	$N\left(p\ ,\ \frac{\sqrt{pq}}{\sqrt{n}}\right)$
Tipificación de las variables	$z = \frac{\bar{x}-\mu}{\left(\sigma/\sqrt{n}\right)}$	$z = \frac{p-\mu}{\left(\sqrt{pq}/\sqrt{n}\right)}$
Intervalos de confianza	$\left(\bar{x}-z_c\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x}+z_c\frac{\sigma}{\sqrt{n}}\right)$	$\left(p-z_c\frac{\sqrt{pq}}{\sqrt{n}}\ ,\ p+z_c\frac{\sqrt{pq}}{\sqrt{n}}\right)$
Error máximo cometido	$E = z_c\frac{\sigma}{\sqrt{n}}$	$E = z_c\frac{\sqrt{pq}}{\sqrt{n}}$
Tamaño de la muestra (Despejando "n")	$n = \frac{z_c{}^2\cdot\sigma^2}{E^2}$	$n = \frac{z_c{}^2\cdot pq}{E^2}$