INTERVALOS DE CONFIANZA

Índice:

1. ***Estimación puntual. Propiedades de los estimadores*** -- 2
2. ***Estimación por intervalos*** -- 2
3. ***Intervalo de confianza par el parámetro p de una distribución binomial*** ------------------ 4
4. ***Intervalo de confianza para la media*** -- 5
5. ***Intervalo de confianza para la diferencia de medias poblacionales*** ------------------------ 7
6. ***Tamaño de la muestra*** --- 8

1. Estimación puntual. Propiedades de los estimadores.

Algunos conceptos que debemos de conocer antes de efectuar la estimación de algunos parámetros por medios de intervalos son:

- **Parámetro**.- Es el valor numérico que describe alguna característica de una población.
- **Estadístico**.- Es un valor numérico que describe una característica de una muestra de la población.
- **Estimador puntual**.- Es un estadístico que se usa para estimar algún parámetro poblacional. Como un estimador puntual es a su vez una variable aleatoria muestral, tiene una distribución de probabilidad, que denominamos distribución muestral.
- **Estimación puntual**.- Es el valor numérico que toma el estimador puntual para una muestra determinada.

 # Ejemplo.- El número medio de móviles que tienen los ciudadanos españoles, es un parámetro de la población, que puede ser estimado utilizando una muestra aleatoria mediante la media muestral, que es un estimador puntual.

Sin embargo, cuando elegimos un estimador para deducir un parámetro poblacional, debemos de elegir un estimador insesgado y eficiente. Donde:

Un estimador es insesgado si su media coincide con el parámetro que queremos estimar.

Un estimador es eficiente cuando su varianza es mínima

2. Estimación por intervalos.

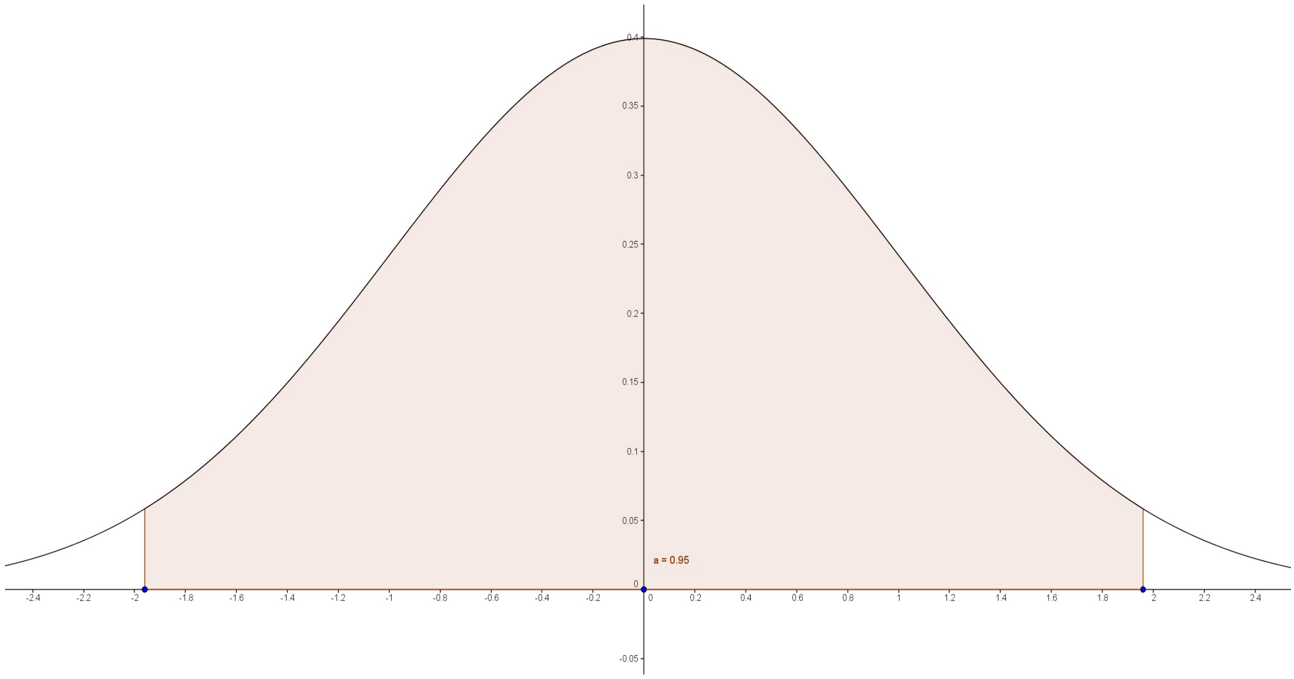
Habitualmente necesitamos estimar aproximadamente un parámetro poblacional, para lo cual utilizamos un intervalo (*denominado de confianza*) en el cual se encontrará el parámetro con un cierto valor de probabilidad.

 # Ejemplo.- Si al elegir una muestra de 1000 personas hemos obtenido una proporción $\hat{p}=0{,}52$ *de los que practican algún deporte, teniendo en cuenta, que el parámetro poblacional p (practicar algún deporte) seguirá una distribución*

$$N\left(\hat{p}, \sqrt{\frac{\hat{p}.(1-\hat{p})}{1000}}\right)=N\left(0{,}52 ; , \sqrt{\frac{0{,}52 . 0{,}48}{1000}}\right) \approx N(0{,}52 ; 0{,}016)$$

Y dado que la variable $Z=\dfrac{p-0{,}52}{0{,}16}$ *sigue una distribución* $N(0{,}1)$ *, con una probabilidad del 95%, existirá un número positivo k, tal que*

$$P(-k \leq Z \leq k) = 0{,}95$$



Que mirando en las tablas de la normal, será $k = 1{,}96$.

Es decir, con probabilidad 0,95, se cumplirá

$$-1{,}96 \leq Z \leq 1{,}96 \Rightarrow -1{,}96 \leq \frac{p-0{,}52}{0{,}016} \leq 1{,}96 \Rightarrow 0{,}052 - 0{,}016\,.\,1{,}96 \leq p \leq 0{,}052 + 0{,}016\,.\,1{,}96$$

Que equivale a que con dicha probabilidad p pertenecerá al intervalo

$$(0{,}052 - 0{,}016\,.\,1{,}96\,;\,0{,}052 + 0{,}016\,.\,1{,}96) = (0{,}489\,;\,0{,}551)$$

A la probabilidad $0{,}95$ *se el denomina coeficiente de confianza (* $(1-\alpha)$ *) , y al valor* $1 - 0{,}95 = 0{,}05$ *nivel de riesgo (* α *), y al número k se representa por* $z_{\frac{\alpha}{2}}$.

Para poder establecer un procedimiento de hallar un cierto intervalo de confianza para un determinado parámetro, con una probabilidad determinada, debemos de conocer los siguientes conceptos.

Estimador por intervalo.- Es un par de estadísticos T_1 , T_2 que se utilizan para estimar un intervalo de un parámetro poblacional de la forma (T_1, T_2) .

- **Estimación por intervalos**.- Es el valor numérico que toma un estimador por intervalo para una muestra concreta, es decir son los extremos del intervalo de confianza.
- Coeficiente de confianza ($(1-\alpha)$).- Es la probabilidad de que un estimador por intervalo contenga al valor del parámetro poblacional..

- **Nivel de significación o riesgo (α).-** Es la diferencia entre la certeza y el nivel de confianza deseado; es decir $1-(1-\alpha)=\alpha$
- **Valor crítico** $z_{\frac{\alpha}{2}}$ **.-** Es el valor de la abscisa que deja a su derecha una probabilidad o área igual a $\frac{\alpha}{2}$, siendo $1-\alpha$ el coeficiente de confianza.
- **Margen de error.-** Es la diferencia entre el extremo superior e inferior de un intervalo de confianza.

3. Intervalo de confianza para el parámetro p de una distribución binomial

Si el parámetro p de una distribución binomial $B(n,p)$ es desconocido, y para estimarlo tomamos el estimador $\hat{p}=\frac{x}{n}$. Donde x es el número de éxitos que se presentan en en pruebas, al tomar una muestra aleatoria de tamaño n. Teniendo en cuenta que

$$\hat{p}\equiv N\left(p,\sqrt{\frac{p.(1-p)}{n}}\right)$$

Tipificando, resulta que

$$Z=\frac{\hat{p}-p}{\sqrt{\frac{p.(1-p)}{n}}}$$

se cumplirá que para la probabilidad $1-\alpha$ (*coeficiente de confianza*), existirá un valor crítico $z_{\frac{\alpha}{2}}$, tal que

$$P\left[-z_{\frac{\alpha}{2}}<Z<z_{\frac{\alpha}{2}}\right]=P\left[-z_{\frac{\alpha}{2}}<\frac{\hat{p}-p}{\sqrt{\frac{p.(1-p)}{n}}}<z_{\frac{\alpha}{2}}\right]=$$

$$=P\left[-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p.(1-p)}{n}}<\hat{p}-p<z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p.(1-p)}{n}}\right]=$$

$$=P\left[\hat{p}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p.(1-p)}{n}}<p<\hat{p}+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p.(1-p)}{n}}\right]=1-\alpha$$

Luego, el intervalo de confianza para p, al nivel de confianza $1-\alpha$ es

$$\left(\hat{p}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p.(1-p)}{n}};\hat{p}+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p.(1-p)}{n}}\right)$$

Si $n\,p>5$ y $n.(1-p)>5$, podemos tomar $\hat{s}=\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}$ en vez de $\sigma=\sqrt{\frac{\hat{p}.(1-p)}{n}}$ y obtendremos el intervalo de confianza

$$\left(\hat{p}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}\,;\,\hat{p}+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}\right)$$

 # Ejemplo.- En un experimento que consiste en el lanzamiento de una moneda, se han efectuado 50 lanzamientos y se han obtenido 30 caras.

Tomando $\hat{p}=\frac{30}{50}=0{,}6$ *, como* $\hat{p}\equiv N\left(p\,,\sqrt{\frac{p.(1-p)}{n}}\right)$ *, y dado que*

$$n.\hat{p}=50.\frac{3}{5}=30>5 \quad y \quad n.(1-\hat{p})=50.\frac{2}{5}=20>5$$

Se cumplirá

$$\hat{p}\equiv N\left(p\,,\sqrt{\frac{\hat{p}.(1-\hat{p})}{50}}\right)$$

Luego, los intervalos de confianza para el parámetro de proporción número de caras al nivel de confianza $1-\alpha$ *será*

$$\left(\hat{p}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{0{,}6.0{,}4}{50}}<;\hat{p}+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{0{,}6.0{,}4}{50}}\right)$$

que para 90%, 95% y 99%, serán

$$\alpha=0{,}05 \quad \Rightarrow \quad \left(0{,}6-1{,}64.\sqrt{\frac{0{,}6.0{,}4}{50}}\,;0{,}6+1{,}64.\sqrt{\frac{0{,}6.0{,}4}{50}}\right)=(0{,}487\,;0{,}714)$$

$$\alpha=0{,}025 \quad \Rightarrow \quad \left(0{,}6-1{,}96.\sqrt{\frac{0{,}6.0{,}4}{50}}\,;0{,}6+1{,}96.\sqrt{\frac{0{,}6.0{,}4}{50}}\right)=(0{,}464\,;0{,}736)$$

$$\alpha=0{,}005 \quad \Rightarrow \quad \left(0{,}6-2{,}58.\sqrt{\frac{0{,}6.0{,}4}{50}}\,;0{,}6+2{,}58.\sqrt{\frac{0{,}6.0{,}4}{50}}\right)=(0{,}421\,;0{,}779)$$

4. Intervalo de confianza para la media

Si el parámetro μ de una distribución normal $N(\mu,\sigma)$ es desconocido, y para estimarlo tomamos el estimador $\bar{x}$, al tomar una muestra aleatoria de tamaño n.

Como la variable aleatoria $var\,X\equiv N\left(\mu\,,\frac{\sigma}{\sqrt{n}}\right)$

Tipificando, resulta que

$$Z=\frac{X-\mu}{\sqrt{\frac{\sigma}{n}}}\equiv N(0,1)$$

y se cumplirá que para la probabilidad $1-\alpha$ (*coeficiente de confianza*), existirá un valor crítico $z_{\frac{\alpha}{2}}$, tal que

$$P\left[-z_{\frac{\alpha}{2}}<Z<z_{\frac{\alpha}{2}}\right]=P\left[-z_{\frac{\alpha}{2}}<\frac{X-\mu}{\frac{\sigma}{\sqrt{n}}}<z_{\frac{\alpha}{2}}\right]=$$

$$=P\left[-z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}<\bar{x}-\mu<z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}\right]=$$

$$=P\left[\bar{x}-z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}\right]=1-\alpha$$

Luego, el intervalo de confianza para p, al nivel de confianza $1-\alpha$ es

$$\left(\bar{x}-z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}};\bar{x}+z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}\right)$$

Si σ es desconocido y $n\geq 30$, podemos tomar la desviación típica muestral $\hat{s}=\frac{1}{n-1}.\sum_{i=1}^{n}(x_i-\bar{x})^2$ en vez de la desviación típica σ y obtendremos el intervalo de confianza

$$\left(\bar{x}-z_{\frac{\alpha}{2}}.\frac{\hat{s}}{\sqrt{n}};\bar{x}+z_{\frac{\alpha}{2}}.\frac{\hat{s}}{\sqrt{n}}\right)$$

 # Ejemplo.- Se sabe que la desviación típica de las tallas de los alumnos de una Universidad es igual 5 cm. Se desea estimarla talla media de los alumnos, para lo cual se escoge una muestra de 100 estudiantes y se obtiene la media $\bar{x}=172\,cm$ *. Para hallar el intervalo de confianza al nivel 0,90, 0,95 y 0,99.*

Como $\bar{x}\equiv N\left(\mu,\frac{\sigma}{n}\right)=N\left(\mu,\frac{5}{\sqrt{100}}\right)=N(\mu;0,5)$

El intervalo de confianza para la talla media de los alumnos al nivel de confianza $1-\alpha$ *será*

$$\left(\bar{x}-z_{\frac{\alpha}{2}}\cdot\frac{5}{\sqrt{100}};\bar{x}+z_{\frac{\alpha}{2}}\cdot\frac{5}{\sqrt{100}}\right)=N(\mu;0{,}5)$$

que para 90%, 95% y 99%, serán

$\alpha=0{,}05 \quad \Rightarrow \quad (172-1{,}64\,.\,0{,}5;172+1{,}64\,.\,0{,}5)=(171{,}18;172{,}82)$

$\alpha=0{,}025 \quad \Rightarrow \quad (172-1{,}96\,.\,0{,}5;172+1{,}96\,.\,0{,}5)=(171{,}02;172{,}98)$

$\alpha=0{,}005 \quad \Rightarrow \quad (172-2{,}58\,.\,0{,}5;172+2{,}58\,.\,0{,}5)=(170{,}71;173{,}29)$

Para hallar los intervalos de confianza, tanto para la proporción como para la media, mirando el la tabla $N(0,1)$, para cada valor $1-\alpha$, obtenemos los valores más utilizados

$1-\alpha$	α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0,8	0,200	0,100	1,28
0,9	0,100	0,050	1,64
0,95	0,050	0,025	1,96
0,99	0,010	0,005	2,58

5. Intervalo de confianza para la diferencias de media poblacional

Sean dos poblaciones normales $N(\mu_1,\sigma_1)$ y $N(\mu_2,\sigma_2)$. Si de cada una de ellas tomamos una muestra de tamaño n_1 y n_2 respectivamente, y sean $\bar{x}_1\, y\, \bar{x}_2$, las respectivas medias muestrales.

Como

$$\mu_1-\mu_2\equiv N\left(\mu_1-\mu_2,\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

El intervalo de confianza de $\mu_1-\mu_2$ al nivel de confianza $1-\alpha$ es:

- Si σ_1 y σ_2 son conocidas

$$\left(\bar{x}_1-\bar{x}_2-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}};\bar{x}_1-\bar{x}_2+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

- Si σ_1 y σ_2 son desconocidas y $n_1\geq 30$ y $n_2\geq 30$, podemos utilizar las desviaciones típicas muestrales $\hat{s}_1$ y $\hat{s}_2$, y será

$$\left(\bar{x}_1-\bar{x}_2-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{s}_1^2}{n_1}+\frac{\hat{s}_2^2}{n_2}};\bar{x}_1-\bar{x}_2+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{s}_1^2}{n_1}+\frac{\hat{s}_2^2}{n_2}}\right)$$

 # Ejemplo.- Se sabe que sabe que los pesos medios de los toros de lidia se distribuyen normalmente, los de la ganadería A, con una desviación típica $s_A = 45\,kg$ *, y los de la ganadería B, con una desviación típica* $s_B = 51\,kg$ *.*

Para hallar el intervalo de confianza al nivel 0,95 de la diferencia de pesos medios, se han realizado dos muestras aleatorias de dichas ganaderías de tamaño 50 para la ganadería A y de 38 para la ganadería B.

Si las medias obtenidas han sido

$$\bar{x}_A = 490\,kg \qquad y \qquad \bar{x}_B = 475\,kg \quad .$$

El intervalo de confianza será

$$\left(\bar{x}_1 - \bar{x}_2 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}\,;\, \bar{x}_1 - \bar{x}_2 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}\right) =$$

$$= \left(490 - 475 - 1{,}96 \cdot \sqrt{\frac{45^2}{50} + \frac{51^2}{38}}\,;\, 490 - 475 + 1{,}96 \cdot \sqrt{\frac{45^2}{50} + \frac{51^2}{38}}\right) = (-5{,}46\,kg.\,;\,35{,}46\,kg.)$$

6. Tamaño de la muestra

Si utilizamos un estimador $\hat{\theta}$ al tomar una muestra de tamaño n, para estimar un parámetro poblacional θ . Y la distribución de $\hat{\theta}$ es $N(\theta, \sigma(n))$,y queremos hallar un intervalo de confianza para $\hat{\theta}$ al nivel de confianza $1-\alpha$, de la forma

$$\left(\theta - z_{\frac{\alpha}{2}} \cdot \sigma(n)\,;\, \theta + z_{\frac{\alpha}{2}} \cdot \sigma(n)\right)$$

Teniendo en cuenta que cuanto mayor sea el número de la muestra n, menor será el error cometido. Si queremos elegir el tamaño de la muestra n, para que el error máximo al tomar $\hat{\theta}$ en vez de θ sea E, es decir:

$$|\theta - \hat{\theta}| < E$$

Se deberá de cumplir la inecuación

$$z_{\frac{\alpha}{2}} \cdot \sigma(n) < E \Rightarrow \sigma(n) < \frac{E}{z_{\frac{\alpha}{2}}} \Rightarrow n < \sigma^{-1}\left(\frac{E}{z_{\frac{\alpha}{2}}}\right)$$

 # Ejemplo.- Una empresa dedicada a la venta de palomitas compra el maíz directamente a los agricultores. Antes de efectuar la compra, un agente de la compañía quiere estimar la probabilidad p de que el grano de maíz se abra al freírlo. ¿Cuántos granos deberá examinar para estar seguro al nivel del 90% de que el error máximo que cometa sea 0,01?. Para ello, se ha realizado un estudio sobre una pequeña muestra de 60 granos, en la que se obtuvo que 48 granos se abrían.

Como sabemos que

$$\hat{p}=\frac{48}{60}=0{,}8$$

y teniendo en cuenta que según la tabla normal N(0,1)

$$z_{\frac{0{,}1}{2}}=1{,}64$$

y que

$$\sigma(n)=\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}=\sqrt{\frac{0{,}8\,.\,0{,}2}{n}}=\sqrt{\frac{0{,}16}{n}}$$

si queremos que el error sea menor que 0,01, se deberá cumplir:

$$z_{\frac{0{,}1}{2}}.\sigma(n)<0{,}01 \Rightarrow 1{,}64\,.\sqrt{\frac{0{,}16}{n}}<0{,}01 \Rightarrow n>\frac{1{,}64^2.0{,}16}{0{,}01^2} \Rightarrow n>4303{,}36$$

Luego, podemos tomar

$$n=40304$$