Medidas de tendencia central en datos no agrupados

Media aritmética de datos no agrupados

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{n}$$

$i = (1,2,3...N).$

$\bar{X}$ = *Media aritmética,*

x_i = *Representa el valor de* i.

N = *Número total de datos.*

Media geométrica de datos no agrupados

$$MG = \bar{X}_G = \sqrt[N]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_i \cdot \ldots \cdot x_N}$$

$MG =$ *Media geométrica.*

$x_i =$ *Elemento del conjunto de datos.*

$N =$ *Total de elementos del conjunto.*

Media armónica de datos no agrupados

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_i} + \ldots + \frac{1}{x_N}}$$

$H =$ *Media armónica .*

$x_i =$ *Elementos del conjunto de datos.*

$N =$ *Total de elementos del conjunto.*

Media cuadrática de datos no agrupados

$$RMS = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_i + \ldots + x_N^2}{N}}$$

$RMS =$ *Media cuadrática.*

$x_i =$ *Elemento del conjunto de datos.*

$N =$ *total de elementos del conjunto.*

Mediana de datos no agrupados

Moda de datos no agrupados

Valor centra de la colección. Hay dos casos para n datos :

1. ***Impar :*** *el valor de la mediana es el valor central de la misma, dividiendo en dos segmentos iguales la colección.*

2. ***Par :*** *el valor de la mediana es el promedio aritmético de los valores centrales.*

Valor (es) con la mayor frecuencia (los que más se repiten). Una colección puede ser :

1. ***Modal :*** *Cuenta con un solo valor de mayor frecuencia.*

2. ***Bimodal :*** *Cuenta con dos valores con la misma frecuencia.*

3. ***Multimodal :*** *Cuenta con más de dos valores con la misma frecuencia.*

Medidas de dispersión de datos no agrupados

Varianza poblacional

$$S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}$$

Varianza muestral

$$S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}$$

x_i = *Término de la colección.*

$\bar{X}$ = *Media aritmética.*

n = *Número total de datos.*

Desviación estándar poblacional

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}}$$

Desviación estándar muestral

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}}$$

x_i = *Término de la colección.*

$\bar{X}$ = *Media aritmética.*

n = *Número total de datos.*

Coeficiente de variación

$$CV = \frac{S_{poblacional}}{V}$$

Medidas de Posición en datos no agrupados

$$PS_p = (n+1)\left(\frac{P}{100}\right)$$

PS_p = *Posición del percentil P*

n = *número total de datos*

Valores de P

Percentil : *divide la colección en 100 partes iguales.*

Decil : *divide la colección en 10 partes iguales.*

Cuartil : *divide la colección en 4 partes iguales.*

Agrupación de datos

Marca de clase

$$MC = \frac{L_i + L_S}{s}$$

$$c = L_{iB} - L_{iA}$$

MC = *Marca de clase*

L_i = *Límite inferior*

L_S = *Límite superior*

c = *Amplitud de clase*

L_{iB} = *Límite inferior de B*

L_{iA} = *Límite inferior de A*

Amplitud de clase

$$I = \frac{R}{k}$$

I = *Amplitud de clase o tamaño de intervalo.*

R = *Rango.*

k = *Cantidad de clases.*

Rango

$$Rango = LRS - LRI$$

$LRS =$ *Límite real superior de la última clase.*

$LRS =$ *Límite real inferior de la primera clase.*

Regla de Sturges

$$k = 1 + \frac{LogN}{Log2}$$

$k =$ *Cantidad de clases que deben construirse.*

$M =$ *Total de elementos en el conjunto de datos.*

Nota: si no es entero, entonces siempre debe redondearse al entero inmediato superior.

Medidas de tendencia central en datos agrupados

Media aritmética de datos agrupados

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i \cdot MC_i}{N}$$

$i = (1, 2, 3 \ldots n)$

$\bar{X} =$ *media aritmética.*

$f_i =$ *Frecuencia absoluta en el intervalo i.*

$MC_i =$ *Marca de clase del intervalo i.*

$N =$ *número total de datos.*

Media geométrica de datos agrupados

$$MG = M_g = 10^{\frac{\sum_{i=1}^{k} f_i \cdot log(x_i)}{N}}$$

$MG = M_g =$ *Media geométrica.*

$log(x_i) =$ *Logaritmo de la marca de clase.*

$k =$ *Total de clases que agrupan al conjunto de datos.*

$f_i =$ *Frecuencia absoluta de la clase .*

$N =$ *Total de elementos del conjunto.*

Media armónica de datos agrupados

$$H = \frac{N}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

$H =$ *Media armónica*

$x_i =$ *Marca de clase.*

$k =$ *Total de clases que agrupan al conjunto de datos.*

$f_i =$ *Frecuencia absoluta de la clase .*

$N =$ *Total de elementos del conjunto.*

Media cuadrática de datos agrupados

$$RMS = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 \cdot f_i}{N}}$$

$RMS =$ *Media cuadrática*

$x_i =$ *Marca de clase.*

$k =$ *Total de clases que agrupan al conjunto de datos.*

$f_i =$ *Frecuencia absoluta de la clase .*

$N =$ *Total de elementos del conjunto.*

Mediana en datos agrupados

$$M_d = LIR_{M_d} + \frac{\frac{N}{2} - fa_{antM_d}}{f_{M_d}} \cdot C$$

M_d = Mediana.

LIR_{M_d} = Límite inferior real de la clase mediana.

fa_{antM_d} = Frecuencia acumulada de la. clase anterior a la clase mediana.

f_{M_d} = Frecuencia de la clase mediana.

C = amplitud de clase de la clase mediana.

N = número total de datos.

Moda en datos agrupados

$$M_o = LIR_{M_0} + \frac{\left(f_{M_o} - f_1\right)}{\left(f_{M_o} - f_1\right) + \left(f_{M_o} - f_2\right)} \cdot C$$

M_o = moda.

LIR_{M_0} = límite inferior real de la clase modal.

f_{M_o} = frecuencia de la clase modal.

f_1 = frecuencia anterior a la clase modal.

f_2 = frecuencia posterior a la clase modal.

C = amplitud de clase de la clase mediana.

Medidas de dispersión de datos agrupados

Dimensión estadística	*P*	*M*
Media	μ	$\bar{X}$
Desviación estándar	σ	S
Número de elementos	N	n

$P =$ *población*

$M =$ *muestra*

Rango

$Rango = LRS - LRI$

$LRS =$ *Límite real superior de la última clase.*

$LRS =$ *Límite real inferior de la primera clase.*

Varianza poblacional

$$S^2 = \frac{\sum_{i=1}^{n} \left(MC_i - \bar{X} \right)^2 \ f_i}{n}$$

Varianza muestral

$$S^2 = \frac{\sum_{i=1}^{n} \left(MC_i - \bar{X} \right)^2 \ f_i}{n-1}$$

$MC_i =$ *Marca de clase.*

$f_i =$ *Frecuencia absoluta.*

$\bar{X} =$ *Media aritmética.*

$n =$ *Número total de datos.*

Desviación estándar poblacional

$$S=\sqrt{\frac{\sum_{i=1}^{n}\left(MC_i-\bar{X}\right)^2 f_i}{n}}$$

Desviación estándar muestral

$$S=\sqrt{\frac{\sum_{i=1}^{n}\left(MC_i-\bar{X}\right)^2 f_i}{n-1}}$$

MC_i = *Marca de clase.*

f_i = *Frecuencia absoluta.*

$\bar{X}$ = *Media aritmética.*

n = *Número total de datos.*

Coeficiente de variación

$$CV=\frac{S_{poblacional}}{V}$$

Medidas de posición en datos agrupados

$$P_p = LIR_p + \frac{\left(\frac{P}{100}\cdot N\right) - fa_{antP}}{f_p}\cdot C$$

P_p = *perceptil P.*

LIR_p = *Límite inferior real de la clase del perceptil P.*

fa_{antP} = *Frecuencia acumulada de la clase anterior a la clase percentil.*

f_p = *Frecuencia de la clase percentil.*

N = *Número total de datos.*

C = *Amplitud de clase de la clase percentil.*

Valores de P

Percentil : *Divide la colección en 100 partes iguales.*

Decil : *Divide la colección en 10 partes iguales.*

Cuartil : *Divide la colección en 4 partes iguales.*

Momentos estadísticos

Diferencia de marca de clase con respecto a la media

$Y = MC - \bar{X}$

Y = *marca de clase con respecto a la media.*
MC = *marca de clase.*
$\bar{X}$ = *media aritmética.*
ME_i = *momento estadístico.*

Momento estadístico de primer grado

$$ME_1 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i}{n} = 0$$

Momento estadístico de segundo grado

$$ME_2 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^2}{n} = S^2$$

Momento estadístico de tercer grado

$$ME_3 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^3}{n}$$

$$k_3 = \frac{ME_3}{S^3}$$

k_3 *es el coeficiente de asimetría.*
si $k_3 > 0$ *la curva de distribución tiene una asimetría derecha o sesgo positivo.*
si $k_3 = 0$ *la curva de distribución es simé – trica.*
si $k_3 < 0$ *la curva de distribución tiene una asimetría izquierda o sesgo negativo.*

Momento estadístico de cuarto grado

$$ME_4 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^4}{n}$$

$$k_4 = \frac{ME_4}{S^4}$$

k_4 *es el coeficiente de apuntamiento (o curtosis).*

si $k_4 - 3 > 0$ *la curva es leptocúrtica.*

si $k_4 - 3 = 0$ *la curva es mesocúrtica.*

si $k_4 - 3 < 0$ *la curva es platocúrtica.*

f_i = *frecuencia del intervalo i*

Y_i = *diferencia de marca de clase del intervalo i con respecto a la media.*

n = *número total de datos.*

S = *desviación estándar.*

Técnicas de conteo

Principio aditivo

$$TP = n + m + p$$

$TP =$ *Total de posibles formas de realizar tres tareas excluyentes entre sí.*
$n =$ *Cantidad de posibles formas de realizar la tarea* 1.
$m =$ *Cantidad de posibles formas de realizar la tarea* 2.
$p =$ *Cantidad de posibles formas de realizar la tarea* 3.

Principio multiplicativo

$$TP = n \cdot m \cdot p$$

$TP =$ *Total de posibles formas de realizar tres tareas excluyentes entre sí.*
$n =$ *Cantidad de posibles formas de realizar la tarea* 1.
$m =$ *Cantidad de posibles formas de realizar la tarea* 2.
$p =$ *Cantidad de posibles formas de realizar la tarea* 3.

Variaciones

Sin repetición

Importa el orden

$$V_n^r = \frac{n!}{(n-r)!}$$

$r =$ *Los elementos de cada variación.*

$n =$ *Número* de elementos.

NO entran todos los elementos.

Con repetición

Importa orden

$$VR_m^n = m^n$$

$m =$ *Los elementos* tomados de n.

$n =$ *Número* de elementos.

NO entran todos los elementos.

Permutaciones

Sin repetición

Importa el orden

$$V_n^n = P_n = n!$$

$n =$ *Número* de elementos.

Entran todos los elementos.

Con repetición

Importa orden

$$P_n^{a,b,c} = \frac{n!}{a!b!c!}$$

$a,b,c =$ *Los elementos* tomados de n.

$n =$ *Número* de elementos.

Entran todos los elementos.

Combinaciones

Sin repetición

No importa el orden

$$ {}_nC_r = C_n^r = \binom{r}{n} = \frac{n!}{r!(n-r)!} $$

$r = Los\ elementos$ tomados de n.

$n = Número$ de elementos.

Combinaciones

Con repetición

No importa el orden

$$ {}_nCR_r = CR_n^r = \binom{r}{n} = \frac{(n+r-1)!}{n!(r-1)!} $$

$r = Los\ elementos$ tomados de n.

$n = Número$ de elementos.

Probabilidad

$$\textit{Eventos independientes} \rightarrow P(A \cap B) = P(A) \cdot P(B)$$

$$\textit{Eventos dependientes} \rightarrow P(A \cap B) = P(A) \cdot P\left(\frac{B}{A}\right)$$

$$\rightarrow P(A \cap B) = P(B) \cdot P\left(\frac{A}{B}\right)$$

Probabilidad de un evento A

$$P(A) = \frac{\textit{número de resultados favorables}}{\textit{número total de resultados posibles}}$$

• *La probabilidad del evento seguro es un uno* $P(A) = 1$

•*La probabilidad del evento imposible es cero* $P(\phi) = 0$

•*La probabilidad de un evento cualquiera toma valores entre cero y uno*

•*La probabilidad del evento contrario de A* $P(\bar{A}) = 1 - P(A)$

$$P\left(\frac{A_i}{B}\right) = \frac{P(A_i) \cdot P\left(\frac{B}{A_i}\right)}{P(B)}$$

$P(A_i) =$ *probabilidad a priori*

$P\left(\frac{B}{A_i}\right) =$ *probabilidad condicional*

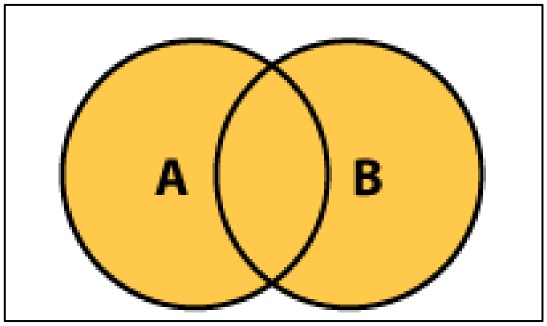
$P(B) =$ *probabilidad total*

$P\left(\frac{A_i}{B}\right) =$ *probabilidad a posteriori*

Operaciones con conjuntos

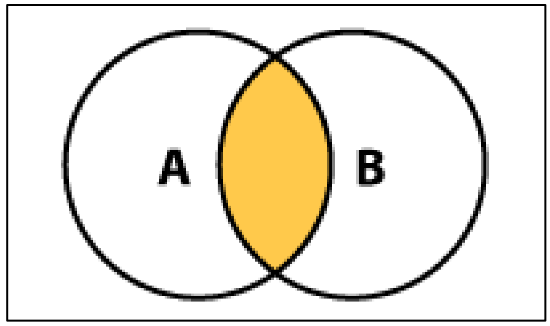
Unión $(A \cup B)$

Notación $\cup$, *o (uno u otro o ambos)*



Intersección $(A \cap B)$

Notación $\cap$, *y (uno y otro)*



Para eventos compatibles

(se pueden realizar a la vez):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Para eventos no compatibles

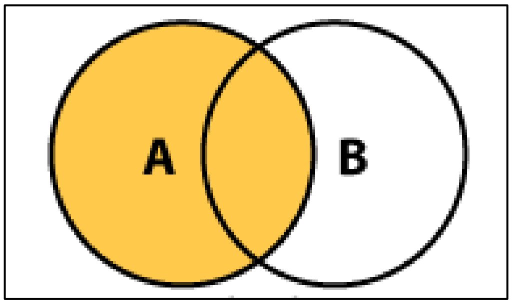
(no se pueden realizar a la vez):

$$P(A \cup B) = P(A) + P(B)$$

por lo tanto $(A \cap B) = 0$

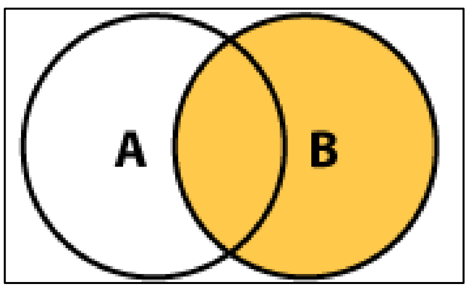
Diferencia $A - B \to A$ *y* $\bar{B}$
(cuando se realiza A y no B)

$$A - B = \left(A \cap \bar{B}\right) \to P\left(A - B\right) = P\left(A\right) - P\left(A \cap B\right)$$

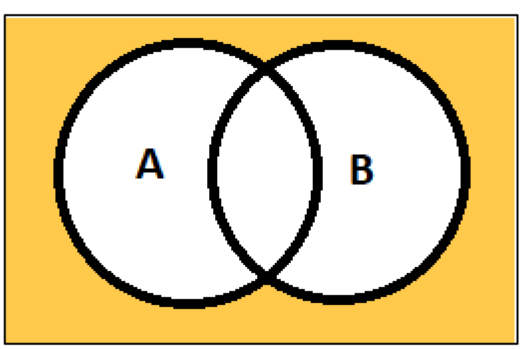


Diferencia $B - A \to \bar{A}$ *y* B
(cuando se realiza B y no A)

$$B - A = \left(\bar{A} \cap B\right) \to P\left(B - A\right) = P\left(B\right) - P\left(A \cap B\right)$$

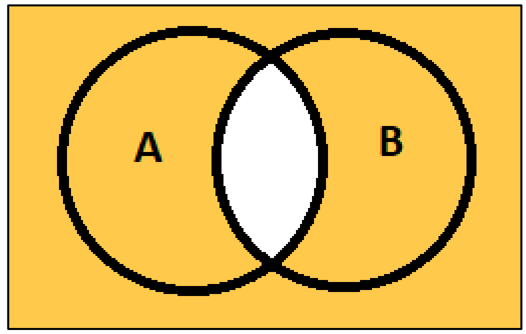


1. *El complementario de la unión de dos eventos es la intersección de los complementarios de estos* $\overline{\left(A \cup B\right)} = \bar{A} \cap \bar{B}$



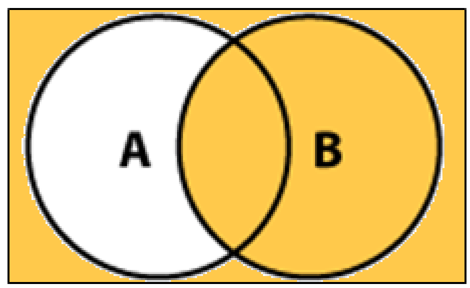
2. *El complementario de la intersección de dos eventos es la unión de los complementarios de estos.*

$$\overline{\left(A \cap B\right)} = \bar{A} \cup \bar{B}$$



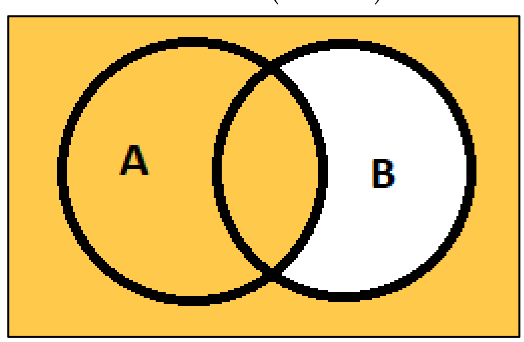
3. *El complementario de A*

$$\mathrm{A}^{C} = \left(B - A\right) \cup \overline{\left(A \cap B\right)} = \left(B - A\right) \cup \bar{A} \cup \bar{B}$$



4. *El complementario de B*

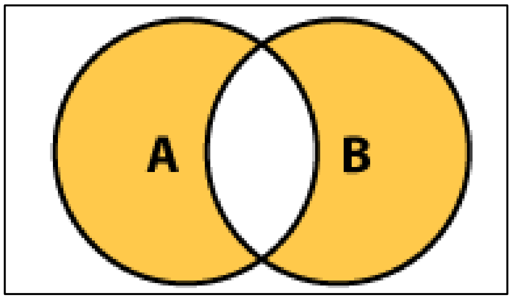
$$\mathrm{B}^{C} = \left(A - B\right) \cup \overline{\left(A \cap B\right)} = \left(A - B\right) \cup \bar{A} \cup \bar{B}$$



$Diferencia\ Simétrica\ A\Delta B \rightarrow A\ y\ \bar{B}$

$(cuando\ se\ realiza\ A\ y\ B\ sin\ intersección)$

$A\Delta B = P(A-B) \cup P(B-A) = P(A) - P(A \cap B)$



Distribución de probabilidad

Distribución binominal

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$q = 1 - p$

Valor esperado *np*

Varianza *npq*

n = número de experimentos

x = número de éxitos

p = probabilidad de tener éxito

q = probabilidad de fracaso

Distribución de Poisson

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Valor esperado λ

Varianza λ

λ *= número de éxitos en un intervalo de tiempo*

x = número de éxitos

Distribución normal

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

Valor esperado μ

Varianza σ^2

x = número de éxitos

σ *= desviación estandar*

σ^2 *= varianza*

μ *= media aritmética*

Distribución geométrica

$$P(x) = pq^{x-1}$$

$q = 1 - p$

Valor esperado $\frac{1}{p}$

Varianza $\frac{1-p}{p^2}$

x = número de intentos hasta el primer éxito

p = probabilidad de éxito

q = probabilidad de fracaso

Distribución exponencial

$$f(x) = \lambda \cdot e^{-\lambda x}$$

Valor esperado $\dfrac{1}{\lambda}$

Varianza $\dfrac{1}{\lambda^2}$

x = *lapso de tiempo*

λ = *tasa promedio de ocurrencia*

e = *base del logaritmo natural*

Distribución T de Student

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

$\bar{X}$ = *media muestral*

μ = *media poblacional*

S = *desviación estándar*

n = *tamaño de la muestra*

Distribución hipergeométrica

$$P(x) = \frac{\dfrac{k!}{n!(n-k)!} \cdot \dfrac{(N-k)!}{(n-x)!(N-k-n+x)!}}{\dfrac{N!}{n!(N-n)!}}$$

Valor esperado $\dfrac{nk}{N}$

Varianza $\dfrac{nk}{N}\left(1 - \dfrac{k}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

n = *tamaño de la muestra*

x = *número de éxitos*

N = *tamaño de la población*

k = *número de éxitos en la población*

Estimaciones

Para estimar una media poblacional

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$

$$E = \bar{X} - m = z\sigma_{\bar{X}}$$

n = *tamaño muestral*

z = *probabilidad de ocurrencia*

σ = *desviación* éstandar

E = *error muestral*

$\bar{X}$ = *meida muestral*

μ = *media poblacional*

$\sigma_{\bar{X}}$ = *error éstandar de la media*

Para estimar la proporción poblacional

$$n = \frac{z^2 \cdot P \cdot Q}{E^2}$$

$$E = p - P = z\sigma_{\bar{P}}$$

n = *tamaño muestral*

z = *probabilidad de ocurrencia*

P = *valor de la proporción en la población*

$$Q = 1 - P$$

E = *error muestral*

p = *proporción*

$\sigma_{\bar{P}}$ = *error estándar de la proporción*

Proporciones

Proporción promedio

$$\bar{P} = \frac{\sum_{i=1}^{Nm} p_i}{Nm}$$

$\bar{P}$ = *promedio muestral de la proporción.*

p_i = *proporcion de la muestra i.*

Nm = *número de muestras.*

Error éstandar de la proporción

$$\sigma_{\bar{P}} = \sqrt{\frac{PQ}{n} \cdot \frac{N-n}{N-1}}$$

$\sigma_{\bar{P}}$ = *error éstandar de la proporción.*

P = *valor de la proporción en la pobla – ción.*

$Q = 1 - P.$

N = *número de elementos de la pobla – ción.*

n = *número de elementos de la muestra.*

Probabilidad de ocurrencia de una proporción

$$z = \frac{p - P}{\sigma_{\bar{P}}}$$

$z =$ probabilidad de ocurrencia de una proporción.

$p =$ proporción.

$P =$ valor de la proporción en la población.

$\sigma_{\bar{P}} =$ error éstandar de la proporción.

Intervalo de confianza

Intervalo de confianza para la media poblacional

Valor del límite inferior : $\bar{X} - z\sigma_{\bar{X}}$

Valor del límite superior : $\bar{X} + z\sigma_{\bar{X}}$

Intervalo de confianza para la proporción poblacional

Valor del límite inferior : $\bar{P} - z\sigma_{\bar{P}}$

Valor del límite superior : $\bar{P} + z\sigma_{\bar{P}}$

$\bar{X} =$ media aritmética

$z =$ probabilidad de ocurrencia

$\sigma_{\bar{X}} =$ error éstandar de la media

$\bar{P} =$ promedio muestra de la proporción

$\sigma_{\bar{P}} =$ error éstandar de la proporción