UNIVERSIDAD CATÓLICA DE SANTA FE

Facultad de Ciencias Económicas

Cátedra: ESTADÍSTICA

TABLA DE FÓRMULAS

Año 2015

ESTADÍSTICA DESCRIPTIVA Y ANÁLISIS EXPLORATORIO DE DATOS

Nota: Si sus datos constituyen una muestra al azar utilice "n" para simbolizar el tamaño muestral en las expresiones; si sus datos constituyen una población estadística utilice "N" para simbolizar el tamaño de la población en las fórmulas

Para lotes de datos sin agrupar

Media aritmética:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{k} x_i . f_i}{n} = \sum_{i=1}^{k} x_i . h_i$$

Varianza:

$$S_n^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 . f_i}{n} = \sum_{i=1}^{k} (x_i - \overline{x})^2 . h_i$$

$$S_n^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2$$

Siendo: f_i = frecuencia absoluta del *i*-ésimo dato diferente del lote

h_i = frecuencia relativa del *i*-ésimo dato diferente del lote

k = nº de datos diferentes del lote

Medidas de posición:

Q_1 = primer cuartil — Posición $Q_1 = \frac{n+1}{4}$

Me = mediana — Posición $Me = \frac{n+1}{2}$

Q_3 = tercer cuartil — Posición $Q_3 = \frac{3(n+1)}{4}$

Amplitud o Rango o recorrido:

$$\mathrm{R} = x_{máx} - x_{mín}$$

Rango medio:

$$\mathrm{Rm} = \frac{x_{máx} + x_{mín}}{2}$$

Eje medio $= \frac{Q_1 + Q_3}{2}$

Rango intercuartílico $= Q_3 - Q_1$

Diagrama de tallos y hojas:

Fórmula Dixon y Kronmal

$$L = \left[10 \log_{10} n\right] \quad 20 \le n \le 300$$

L= cantidad máxima de líneas.

Diagrama de caja:

Intervalos de valores alejados

$$\left[Q_1 - 3(Q_3 - Q_1)\ ;\ Q_1 - 1,5(Q_3 - Q_1)\right)$$

$$\left(Q_3 + 1,5(Q_3 - Q_1)\ ;\ Q_3 + 3(Q_3 - Q_1)\right]$$

Intervalos de valores muy alejados

$$\left(-\infty\ ;\ Q_1 - 3(Q_3 - Q_1)\right)$$

$$\left(Q_3 + 3(Q_3 - Q_1)\ ;\ \infty\right)$$

Para datos agrupados en intervalos de clase

Regla empírica de Sturges

k = $\left[1+3{,}322\log_{10} n\right]$ k = nº de intervalos

Media aritmética:

$$\overline{x} = \frac{\sum_{i=1}^{k} pm_i . f_i}{n} = \sum_{i=1}^{k} pm_i . h_i$$

Varianza:

$$S_n^2 = \frac{\sum_{i=1}^{k} (pm_i - \overline{x})^2 . f_i}{n} = \sum_{i=1}^{k} (pm_i - \overline{x})^2 . h_i$$

f_i = frecuencia absoluta del *i*-ésimo intervalo de clase

pm_i = punto medio del i-ésimo intervalo de clase

h_i = frecuencia relativa del *i*-ésimo intervalo de clase

n = tamaño de la muestra

Medidas de posición:

Q_1 = primer cuartil — Posición $Q_1 = \frac{n}{4}$

Me = mediana — Posición $Me = \frac{n}{2}$

Q_3 = tercer cuartil — Posición $Q_3 = \frac{3.n}{4}$

Para lotes de datos agrupados y sin agrupar

Coeficiente de Variación: $CV = \left(\frac{S}{\overline{X}}\right).100\%$

Coeficiente de asimetría de Bowley

$$A_B = \frac{(Q_3 - Me) - (Me - Q_1)}{(Me - Q_1) + (Q_3 - Me)}$$

Siendo Me = mediana
Qi = i-ésimo cuartil , con i = 1, 3

Coeficiente de asimetría de Pearson

Sólo válido para distribuciones campaniformes y unimodales.

$$A_p = \frac{\overline{x} - Mo}{S}$$

Siendo Mo = modo o moda (puede reemplazarse por la mediana)

Regla de Bienaymé-Chebyshev

$\left(1 - \frac{1}{k^2}\right)100\%$; $k \geq 1$; k = cantidad de desviaciones estándares

Regla empírica

- Aproximadamente 67% de las observaciones pertenece al intervalo $\left(\overline{x} - s_n \ , \ \overline{x} + s_n\right)$
- Aproximadamente entre el 90 y 95% de las observaciones pertenece al intervalo $\left(\overline{x} - 2s_n \ , \ \overline{x} + 2s_n\right)$

Distribución Normal

- 68,25% de las observaciones pertenece al intervalo $\left(\overline{x} - s_n \ , \ \overline{x} + s_n\right)$
- 95,44% de las observaciones pertenece al intervalo $\left(\overline{x} - 2s_n \ , \ \overline{x} + 2s_n\right)$

PROBABILIDAD

Probabilidad de eventos complementarios:

$$P(A^c) = P\left(\bar{A}\right) = 1 - P(A)$$

Siendo A^c o $\bar{A}$ el evento complementario de A

Regla general de la adición:

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

$$P(\bigcup_{i=1}^{k} A_i) = P(A_1 \ \cup A_2 \cup ... \cup A_k)$$

$$P(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i) - \sum_{i<j=2}^{k} P(A_i \cap A_j) + \sum_{i<j<r=3}^{k} P(A_i \cap A_j \cap A_r) + ... + (-1)^{k-1} P(A_1 \cap A_2 \cap ... \cap A_k)$$

Regla de la adición para eventos mutuamente excluyentes:

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

$$P(\bigcup_{i=1}^{k} A_i) \ = \ P(A_1 \ \cup A_2 \cup A_3 \cup ... \cup A_k) \ = \ P(A_1) + P(A_2) + P(A_3) + ... + P(A_k)$$

Probabilidad de A condicionado a B:

$$P(A/B) \ = \ \frac{P(A \ \cap \ B)}{P(B)} \qquad P(B) \neq 0$$

Regla general de la multiplicación:

$$P(A \ \cap \ B) = P(A \ y \ B) = P(A/B).\ P(B)$$

Regla de la multiplicación para eventos independientes:

$$P(A \ \cap \ B) = P(A).\ P(B)$$

Probabilidad Total:

$$P(A) = P(A/B_1).\ P(B_1) + P(A/B_2).\ P(B_2) + ... + P(A/B_k).\ P(B_k)$$

Teorema de Bayes:

$$P(B_i/A) = \frac{P(A/B_i).\ P(B_i)}{P(A/B_1).\ P(B_1) + P(A/B_2).\ P(B_2) + ... + P(A/B_k).\ P(B_k)}$$

Siendo: $B_1, B_2, \ldots, B_i, \ldots B_k$ eventos mutuamente excluyentes y colectivamente exhaustivos

VARIABLES ALEATORIAS

Valor esperado de una variable aleatoria:

Variable aleatoria discreta:

$$\mu_X = E(X) = \sum_{i=1}^{n}\left[x_i . P(X = x_i)\right]$$

Variable aleatoria continua:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x . f(x) dx$$

f(x): *fdp* función de densidad de probabilidad de X

Varianza de una variable aleatoria:

$$\sigma_X^2 = V(X) = E\left[(X - \mu_X)^2\right] = E(X^2) - \mu_X^2$$

Variable aleatoria discreta:

$$\sigma_X^2 = V(X) = \sum_{i=1}^{n}\left[(x_i - \mu_X)^2 . P(X = x_i)\right]$$

Variable aleatoria continua:

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 . f(x)\, dx$$

f(x): *fdp* función de densidad de probabilidad de X

Desviación estándar de una variable aleatoria (discreta o continua):

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{V(X)}$$

Algunas propiedades del valor esperado de una variable aleatoria (discreta o continua)

$$E(k) = k$$

$$E(kX) = kE(X)$$

Siendo k una constante y X una variable aleatoria

Siendo $X_1, X_2, ..., X_n$ *n* variables aleatorias:

$$E(X_1 \pm X_2 \pm ... \pm X_n) = E(X_1) \pm E(X_2) \pm ... \pm E(X_n)$$

Además, si $X_1, X_2, ..., X_n$ son independientes:

$$E(X_1 . X_2 . \ldots . X_n) = E(X_1) . E(X_2) . \ldots . E(X_n)$$

Algunas propiedades de la varianza de una variable aleatoria (discreta o continua)

$$V(k) = 0$$

$$V(X + k) = V(X)$$

$$V(k.X) = k^2 . V(X)$$

Siendo: k una constante
X una variable aleatoria

$$V(X_1 \pm X_2 \pm ... \pm X_n) = V(X_1) + V(X_2) + ... + V(X_n)$$

Siendo $X_1, X_2, ..., X_n$ n variables aleatorias independientes

DISTRIBUCIONES DE PROBABILIDAD

Distribución Binomial

X = Número de éxitos en *n* repeticiones independientes de un experimento Bernoulli, cuya probabilidad de éxito es p

$$X \sim Bin(n; p)$$

$$P(X = x) = \binom{n}{x} p^x.(1-p)^{n-x}$$

Para x = 0, 1,…,n

$$\boldsymbol{\mu_X = E(X) = n.p}$$

$$\boldsymbol{\sigma_X^2 = V(X) = n.p.(1-p)}$$

Distribución Hipergeométrica

X = Número de éxitos en una muestra aleatoria de tamaño *n* seleccionada sin reposición de una población finita de tamaño **N** en la cual hay *A* éxitos

$$X \sim Hiperg.(n; N; A)$$

$$P(X = x) = \frac{\binom{A}{x}.\binom{N-A}{n-x}}{\binom{N}{n}}$$

Para x = 0, 1, 2 ….,n siendo $n \leq A$

$$\mu_X = E(X) = n.\frac{A}{N}$$

$$\sigma_X^2 = V(X) = n.\frac{A}{N}.(1-\frac{A}{N}).\frac{N-n}{N-1}$$

Distribución de Poisson

X = Número de éxitos en un intervalo continuo (en el cual hay, en promedio, λ éxitos por unidad)

$$X \sim Poisson(\lambda)$$

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

Para x = 0, 1, 2, …

$$\boldsymbol{\mu_X = E(X) = \lambda} \qquad \boldsymbol{\sigma_X^2 = V(X) = \lambda}$$

Distribución Normal

La función de densidad de una variable aleatoria $X \sim Normal(\mu_X; \sigma_X)$ es:

$$f(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{1}{2}\left[(X-\mu_X)/\sigma_X\right]^2} \qquad -\infty < x < \infty$$

Donde $\pi = 3{,}14159\ldots$ y $e = 2{,}71828\ldots$

$E(X) = \mu_X$ con $-\infty < \mu_X < \infty$

$V(X) = \sigma_X^2$ con $\sigma_X > 0$

Estandarización de la Distribución Normal

$$Z = \frac{X-\mu_X}{\sigma_X}$$

$$Z \sim Normal(\mu_z = 0; \sigma_z = 1)$$

APROXIMACIONES ENTRE DISTRIBUCIONES

Aproximación entre la Distribución Hipergeométrica y la Distribución Binomial

$$X \sim Hiperg.(n; N; A) \rightarrow Bin(n; p = \frac{A}{N})$$

Cuando $n \leq 0{,}05\, N$

Aproximación entre la Distribución Binomial y la Distribución de Poisson

$$X \sim Bin(n; p) \rightarrow Poisson(\lambda = n.p)$$

Cuando $n \geq 20$ y $p \leq 0{,}05$

Aproximación entre la Distribución Binomial y la Distribución Normal

$$X \sim Bin(n; p) \rightarrow Normal(\mu_X = n.p \ ; \ \sigma_X = \sqrt{n.p.(1-p)})$$

Cuando $n.p \geq 5$ y $n.(1\text{-}p) \geq 5$

Aproximación entre la Distribución Poisson y la Distribución Normal

$$X \sim Poisson(\lambda) \rightarrow Normal(\mu_X = \lambda \ ; \ \sigma_X = \sqrt{\lambda})$$

Cuando $\lambda \geq 5$

Aproximación entre la Distribución Hipergeométrica y la Distribución Normal

$$X \sim Hiperg.(n; N; A) \rightarrow Normal\left(\mu_X = n.p \ ; \ \sigma_X = \sqrt{n.\frac{A}{N}.(1-\frac{A}{N}).\frac{N-n}{N-1}}\right)$$

Cuando $n.\frac{A}{N} \geq 5$ y $n.(1-\frac{A}{N}) \geq 5$

Aproximación entre la Distribución *t-Student* y la Distribución Normal

$$T \sim t_n \ (n = \text{grados de libertad}) \Rightarrow T \rightarrow Normal(0;1)$$

Cuando $n > 30$

INTERVALOS DE CONFIANZA
(POBLACIÓN INFINITA)

Intervalo de confianza para la media de la población, con σ_X conocido

$$\left[\overline{x} - z_{\frac{\alpha}{2}}.\frac{\sigma_X}{\sqrt{n}} \; ; \overline{x} + z_{\frac{\alpha}{2}}.\frac{\sigma_X}{\sqrt{n}}\right] \text{ al } (1-\alpha)\% \text{ de confianza}$$

Siendo $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$

Intervalo de confianza para la media de la población, con σ_X desconocido

$$\left[\overline{x} - t_{n-1;\frac{\alpha}{2}}.\frac{s_{n-1}}{\sqrt{n}} \; ; \overline{x} + t_{n-1;\frac{\alpha}{2}}.\frac{s_{n-1}}{\sqrt{n}}\right] \text{ al } (1-\alpha)\% \text{ de confianza}$$

Siendo $t_{n-1;\frac{\alpha}{2}}$ es el valor de T $\sim t_{n-1}$ tal que $P(T > t_{n-1;\frac{\alpha}{2}}) = \frac{\alpha}{2}$; n= tamaño de la muestra

Tamaño de muestra para un determinado error de estimación e, con σ_X conocido

$$n = \left(\frac{z_{\frac{\alpha}{2}}.\sigma_X}{e}\right)^2$$

Intervalo de confianza para la proporción de la población

$$\left[p_s - z_{\frac{\alpha}{2}}.\sqrt{\frac{p_s.(1-p_s)}{n}} \; ; p_s + z_{\frac{\alpha}{2}}.\sqrt{\frac{p_s.(1-p_s)}{n}}\right] \text{ al } (1-\alpha)\% \text{ de confianza} \quad \left(\text{siendo } p_s = \frac{x}{n}\right)$$

Siendo $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$

Tamaño de muestra para un determinado error *e*

$$n = p_s(1-p_s)\left(\frac{z_{\frac{\alpha}{2}}}{e}\right)^2$$

INTERVALOS DE CONFIANZA
(POBLACIÓN FINITA DE TAMAÑO N)

fpc = Factor de corrección para población finita

$$fpc=\sqrt{\frac{N-n}{N-1}} \qquad \left(\text{despreciable para } \frac{n}{N}\leq 0{,}05\right)$$

Intervalo de confianza para la media de la población, con σ_X conocido

$$\left[\bar{x}-z_{\frac{\alpha}{2}}.\frac{\sigma_x}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}\ ;\ \bar{x}+z_{\frac{\alpha}{2}}.\frac{\sigma_x}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}\right] \text{ al } (1-\alpha)\% \text{ de confianza}$$

Siendo $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z>z_{\frac{\alpha}{2}})=\frac{\alpha}{2}$

Intervalo de confianza para la media de la población, con σ_X desconocido

$$\left[\bar{x}-t_{n-1;\frac{\alpha}{2}}.\frac{s_{n-1}}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}\ ;\ \bar{x}+t_{n-1;\frac{\alpha}{2}}.\frac{s_{n-1}}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}\right] \text{ al } (1-\alpha)\% \text{ de confianza}$$

Siendo $t_{n-1;\frac{\alpha}{2}}$, es el valor de $T \sim t_{n-1}$ tal que $P(T>t_{n-1;\frac{\alpha}{2}})=\frac{\alpha}{2}$; n= tamaño de la muestra

Tamaño de muestra n para un determinado error de estimación e, con σ_X conocido, cuando se ha aplicado el factor de corrección por población finita (de tamaño N)

$$n=\frac{n_0 N}{n_0+(N-1)}$$

Siendo: $n_0=\left(\frac{z_{\frac{\alpha}{2}}\sigma_X}{e}\right)^2$ y $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z>z_{\frac{\alpha}{2}})=\frac{\alpha}{2}$

Intervalo de confianza para la proporción de la población

$$\left[p_s-z_{\frac{\alpha}{2}}.\sqrt{\frac{p_s.(1-p_s)}{n}}.\sqrt{\frac{N-n}{N-1}}\ ;\ p_s+z_{\frac{\alpha}{2}}.\sqrt{\frac{p_s.(1-p_s)}{n}}.\sqrt{\frac{N-n}{N-1}}\right] \text{ al } (1-\alpha)\% \text{ de confianza}$$

Siendo: $p_s=\frac{x}{n}$ y $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z>z_{\frac{\alpha}{2}})=\frac{\alpha}{2}$

Tamaño de muestra n para un determinado error de estimación e, cuando se ha aplicado el factor de corrección por población finita (de tamaño N)

$$n=\frac{n_0 N}{n_0+(N-1)}$$

Siendo: $n_0=p_s(1-p_s)\left(\frac{z_{\frac{\alpha}{2}}}{e}\right)^2$ y $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z>z_{\frac{\alpha}{2}})=\frac{\alpha}{2}$

TEST DE HIPÓTESIS
(POBLACIÓN INFINITA)

α = Nivel de significancia del test

Test de hipótesis para la media de la población, con σ_X conocido

De dos colas	De una cola
De dos colas: $H_0: \mu_X = \mu_0$; $H_1: \mu_X \neq \mu_0$	De una cola: $H_0: \mu_X = \mu_0$; $H_1: \mu_X < \mu_0$ (o bien $\mu_X > \mu_0$)
$Z = \dfrac{\bar{X} - \mu_0}{\frac{\sigma_x}{\sqrt{n}}}$	$Z = \dfrac{\bar{X} - \mu_0}{\frac{\sigma_x}{\sqrt{n}}}$
Región de rechazo: $(-\infty\ ,\ -z_{\alpha/2}) U (z_{\alpha/2}\ ,\ \infty)$	Región de rechazo: $(-\infty\ ,\ -z_{\alpha})$ $(o\ bien\ (z_\alpha\ ,\ \infty)\ )$
Siendo $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z > z_{\frac{\alpha}{2}}) = \dfrac{\alpha}{2}$	Siendo z_α el valor de Z tal que $P(Z > z_\alpha) = \alpha$

Test de hipótesis para la media de la población, con σ_X desconocido

De dos colas	De una cola
De dos colas: $H_0: \mu_X = \mu_0$; $H_1: \mu_X \neq \mu_0$	De una cola: $H_0: \mu_X = \mu_0$; $H_1: \mu_X < \mu_0$ (o bien $\mu_X > \mu_0$)
$T_{n-1} = \dfrac{\bar{X} - \mu_0}{\frac{S_{n-1}}{\sqrt{n}}}$	$T_{n-1} = \dfrac{\bar{X} - \mu_0}{\frac{S_{n-1}}{\sqrt{n}}}$
Región de rechazo: $(-\infty\ ,\ -t_{n-1\ ;\ \alpha/2}) U (t_{n-1\ ;\ \alpha/2}\ ,\ \infty)$	Región de rechazo: $(-\infty\ ,\ -t_{n-1\ ;\ \alpha})$ (o bien $(t_{n-1\ ;\ \alpha}\ ,\ \infty)$)
Siendo: $t_{n-1;\frac{\alpha}{2}}$ es el valor de $T \sim t_{n-1}$ tal que $P(T > t_{n-1;\frac{\alpha}{2}}) = \dfrac{\alpha}{2}$	Siendo: $t_{n-1;\alpha}$ es el valor de $T \sim t_{n-1}$ tal que $P(T > t_{n-1;\alpha}) = \alpha$
n= tamaño de la muestra	n= tamaño de la muestra

Tamaño de muestra para α y β determinados, en un Test de hipótesis para la media de la población, con σ_X conocido (POBLACIONES INFINITAS)

Tamaño de muestra para α y β determinados

Test de una cola	Test de dos colas
Test de una cola: $n = \dfrac{\sigma_X^2 (z_\alpha - z_\beta)^2}{(\mu_0 - \mu_1)^2}$	Test de dos colas: $n = \dfrac{\sigma_X^2 (z_{\alpha/2} - z_\beta)^2}{(\mu_0 - \mu_1)^2}$

TEST DE HIPÓTESIS
(POBLACIÓN FINITA)

α = Nivel de significancia del test

Test de hipótesis para la media de la población, con σ_X conocido

De dos colas: $H_0: \mu_X = \mu_0$; $H_1: \mu_X \neq \mu_0$

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma_x}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Región de rechazo: $\left(-\infty\ ,\ -z_{\alpha/2}\right) U \left(z_{\alpha/2}\ ,\ \infty\right)$

Siendo $z_{\frac{\alpha}{2}}$ el valor de Z tal que $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$

De una cola: $H_0: \mu_X = \mu_0$; $H_1: \mu_X < \mu_0$ (o bien $\mu_X > \mu_0$)

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma_x}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Región de rechazo:
$\left(-\infty\ ,\ -z_\alpha\right)$ (*o bien* $\left(z_\alpha\ ,\ \infty\right)$)

Siendo z_α el valor de Z tal que $P(Z > z_\alpha) = \alpha$

Test de hipótesis para la media de la población, con σ_X desconocido

De dos colas: $H_0: \mu_X = \mu_0$; $H_1: \mu_X \neq \mu_0$

$$T_{n-1} = \frac{\bar{X} - \mu_0}{\frac{s_{n-1}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Región de rechazo:
$\left(-\infty\ ,\ -t_{n-1\ ;\ \alpha/2}\right) U \left(t_{n-1\ ;\ \alpha/2}\ ,\ \infty\right)$

Siendo:

$t_{n-1;\frac{\alpha}{2}}$, es el valor de T t_{n-1} tal que $P(T > t_{n-1;\frac{\alpha}{2}}) = \frac{\alpha}{2}$

n= tamaño de la muestra

De una cola: $H_0: \mu_X = \mu_0$; $H_1: \mu_X < \mu_0$ (o bien $\mu_X > \mu_0$)

$$T_{n-1} = \frac{\bar{X} - \mu_0}{\frac{s_{n-1}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Región de rechazo:
$\left(-\infty\ ,\ -t_{n-1\ ;\ \alpha}\right)$ (o bien $\left(t_{n-1\ ;\ \alpha}\ ,\ \infty\right)$)

Siendo:

$t_{n-1;\alpha}$ es el valor de T t_{n-1} tal que $P(T > t_{n-1;\alpha}) = \alpha$

n= tamaño de la muestra

REGRESIÓN LINEAL SIMPLE Y CORRELACIÓN

Ecuación de la recta de regresión muestral (por mínimos cuadrados):

$$\hat{Y} = a + bX$$

$$a = \overline{y} - b\overline{x}$$

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n.\overline{x}.\overline{y}}{\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2}$$

o también: $$b = \frac{n.\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n.\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

SST = Suma total de cuadrados

SSR = Suma de cuadrados debida a la regresión o explicada

SSE = Suma de cuadrados de error o no explicada.

SST = SSR + SSE

SST = $\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2$

SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i$

SSR = $\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i - n\overline{y}^2$

Error estándar muestral de la regresión *S*

S_{YX} o S_e :

$$S_{YX} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i}{n-2}}$$

Coeficiente de determinación muestral r^2:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{SSE}{SST}$$

$$r^2 = \frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i - n(\overline{y})^2}{\sum_{i=1}^{n} y_i^2 - n(\overline{y})^2} = \frac{SSR}{SST}$$

Coeficiente de correlación muestral r :

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

$$r = \frac{\sum_{i=1}^{n}(x_i y_i) - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}}}$$

$$r = \frac{(\sum_{i=1}^{n} x_i.y_i) - n.\overline{x}.\overline{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n.\overline{x}^2}.\sqrt{\sum_{i=1}^{n} y_i^2 - n.\overline{y}^2}}$$

SERIES TEMPORALES

MODELO CLÁSICO MULTIPLICATIVO

Datos registrados anualmente o en periodos mayores a un año	Datos registrados en periodos menores a un año
$Y_i = T_i \,.\, C_i \,.\, I_i$	$Y_i = T_i \,.\, S_i \,.\, C_i \,.\, I_i$

Siendo: Y_i = valor de la variable observado en el período i

T_i = valor del componente de tendencia en el período i

S_i = valor del componente estacional en el período i

C_i = valor del componente cíclico en el período i

I_i = valor del componente irregular en el período i

Tendencia lineal de la serie

Ecuación de la recta de regresión muestral (por mínimos cuadrados):

$$T_i = \hat{Y} = a + bX$$

$$a = \bar{y} - b\bar{x} \qquad b = \frac{\sum_{i=1}^{n} x_i y_i - n.\bar{x}.\bar{y}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}$$

Siendo:

x_i = código asignado al i-ésimo período observado.

y_i = valor de la variable observado en el i-ésimo período.

Variación cíclica

Porcentaje de tendencia	Residuo cíclico relativo
$\frac{Y_i}{\hat{Y}_i}.100$	$\frac{Y_i - \hat{Y}_i}{\hat{Y}_i} = \frac{Y_i}{\hat{Y}_i} - 1$

Relativas Cíclicas Irregulares

Si: $Y_i = T_i \,.\, C_i \,.\, I_i$

$$C_i \,.\, I_i = \frac{Y_i}{\hat{Y}_i} = \frac{T_i \,.\, C_i \,.\, I_i}{T_i}$$

Si: $Y_i = T_i \,.\, S_i \,.\, C_i \,.\, I_i$

$$C_i \,.\, I_i = \frac{Y_i}{\hat{Y}_i \,.\, S_i} = \frac{T_i \,.\, S_i \,.\, C_i \,.\, I_i}{T_i \,.\, S_i}$$

NÚMEROS ÍNDICES

Cambio de base

$$I_{nuevo}^{(t)} = \frac{I_{antiguo}^{(t)}}{I_{base\ deseada}^{(0)}} \times 100$$

$I_{nuevo}^{(t)} =$ nº índice en el período t con base sustituida al tiempo deseado.

$I_{antiguo}^{(t)} =$ nº índice en el período t con base antigua.

$I_{base\ deseada}^{(0)} =$ nº índice con base antigua correspondiente al tiempo que se desea como base

Índices simples

$$I_0^t(i) = \frac{y_{it}}{y_{i0}} \times 100 \qquad \boldsymbol{i}: 1,2,\ldots \boldsymbol{n}; \quad \boldsymbol{t}: 0,1,\ldots.. \boldsymbol{k}$$

y_{it} = valor de una magnitud y_i medida en el momento t (período actual)

y_{i0} = valor de una magnitud y_i medida en el momento t =0 (período base)

Índice de Precios

$$P_0^t(i) = \frac{p_{it}}{p_{i0}} \times 100$$

$\boldsymbol{i}: 1,2,\ldots \boldsymbol{n}; \quad \boldsymbol{t}: 0,1,\ldots.. \boldsymbol{k}$

p_{it} = precio de un bien i en el período t

p_{i0} = precio del mismo bien i en el período base

Índice de Cantidad

$$Q_0^t(i) = \frac{q_{it}}{q_{i0}} \times 100$$

$\boldsymbol{i}: 1,2,\ldots \boldsymbol{n}; \quad \boldsymbol{t}: 0,1,\ldots.. \boldsymbol{k}$

q_{it} = cantidad de un bien i en el período t

q_{i0} = cantidad del mismo bien i en el período base

Índice de valor

$$V_0^t(i) = \frac{v_{it}}{v_{i0}} \times 100 = \frac{p_{it}q_{it}}{p_{i0}q_{i0}} \times 100$$

$\boldsymbol{i}: 1,2,\ldots \boldsymbol{n}; \quad \boldsymbol{t}: 0,1,\ldots.. \boldsymbol{k}$

v_{it} = valor de un bien i en el período t

v_{i0} = valor del mismo bien i en el período base

Índices complejos no ponderados (precio o cantidad)

Índice media aritmética simple o Índice de Sanerbeck

$$I_{St} = \frac{\sum_{i=1}^{n} \frac{y_{it}}{y_{i0}}}{n} \times 100$$

Índice media agregativa simple o Índice de Bradstrest y Dutot

$$I_{BD} = \frac{\sum_{i=1}^{n} \frac{y_{it}}{n}}{\sum_{i=1}^{n} \frac{y_{i0}}{n}} \times 100 = \frac{\sum_{i=1}^{n} y_{it}}{\sum_{i=1}^{n} y_{i0}} \times 100$$

y_{it} = valor de la magnitud "y" (precio o cantidad) del artículo (o servicio) i-ésimo, medida en el momento t (período actual)

y_{i0} = valor de una magnitud "y" (precio o cantidad) del artículo (o servicio) i-ésimo, medida en el momento t=0 (período base)

n = número de artículos (o servicios) del compuesto objeto de estudio

Índice de precios de Laspeyres en el período *t*	**Índice de cantidades de Laspeyres en el período *t***
$$P_L^{(t)} = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \times 100$$	$$Q_L^{(t)} = \frac{\sum_{i=1}^{n} p_{i0} q_{it}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \times 100$$

Siendo:

n = número de artículos (o servicios) del compuesto objeto de estudio

p_{it} = precio del artículo (o servicio) i-ésimo en el período *t* (período actual)

p_{i0} = precio del artículo (o servicio) i-ésimo en el período base (t=0)

q_{it} = cantidad consumida del artículo (o servicio) i-ésimo en el período *t* (período actual)

q_{i0} = cantidad consumida del artículo (o servicio) i-ésimo en el período base (t=0)

Índice de precios de Paasche en el período *t*	**Índice de cantidades de Paasche en el período *t***
$$P_P^{(t)} = \frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{i0} q_{it}} \times 100$$	$$Q_P^{(t)} = \frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{it} q_{i0}} \times 100$$

Siendo:

n = número de artículos (o servicios) del compuesto objeto de estudio

p_{it} = precio del artículo (o servicio) i-ésimo en el período *t* (período actual)

p_{i0} = precio del artículo (o servicio) i-ésimo en el período base (t=0)

q_{it} = cantidad consumida del artículo (o servicio) i-ésimo en el período *t* (período actual)

q_{i0} = cantidad consumida del artículo (o servicio) i-ésimo en el período base (t=0)

Índice de precios de Fisher en el período *t*	**Índice de cantidades de Fisher en el período *t***
$$P_F^{(t)} = \sqrt{P_L^{(t)} P_P^{(t)}}$$	$$Q_F^{(t)} = \sqrt{Q_L^{(t)} Q_P^{(t)}}$$
$P_L^{(t)}$ = Índice de precios de Laspeyres en el período *t* (período actual) $P_P^{(t)}$ = Índice de precios de Paasche en el período *t* (período actual)	$Q_L^{(t)}$ = Índice de cantidades de Laspeyres en el período *t* (período actual) $Q_P^{(t)}$ = Índice de cantidades de Paasche en el período *t* (período actual)