

Pequeño diccionario de Estadística



Edición 2026

Colección Hojamat.es

© Antonio Roldán Martínez

<http://www.hojamat.es>

PRESENTACIÓN

Esta publicación sustituye a la primera versión en HTML contenida en hojamat.es. Esta página está actualmente en periodo de simplificación y eliminación de material que ya no se visita. Por eso es conveniente descargarla de enlaces y unificar los materiales más interesantes.

Se han añadido algunos términos que se echaban de menos e insertado algunas imágenes y gráficos.

Seguirá siendo un glosario de tipo elemental o medio, que era el pretendido desde su primera edición.

CONTENIDO

- Presentación 2**
- Contenido 3**
- A 5
- B 7
- C 8
- D 13
- E 17
- F 22
- G 24
- H 24
- I 25
- L 27
- M 29
- N 33

O.....	34
P.....	35
R.....	37
S.....	39
T.....	40
U.....	42
V.....	42

A

Agrupación de datos

Si la variable que se estudia es continua, o discreta con muchos valores distintos, se organizarán sus datos en forma de intervalos. Para ello se fija un valor mínimo y otro máximo, de forma que todos los datos estén comprendidos entre ellos (a veces esto no se garantiza y quedan intervalos abiertos). La diferencia entre ambos se denomina rango de los datos y posteriormente se divide en un número de intervalos mediante valores intermedios.

Aleatorio

Experimento aleatorio

Un experimento se llama aleatorio cuando repetido indefinidamente presenta siempre resultados totalmente impredecibles.

Variable aleatoria

Llamaremos **Variable aleatoria simple** (discreta) a un conjunto de valores $X_1, X_2, X_3, \dots, X_n$ (llamados también sucesos) a los que les corresponden unos números (llamados probabilidades), $p_1, p_2, p_3, \dots, p_n$ que cumplen:

- a) Todas las probabilidades son positivas o nulas.
- b) La suma de todas ellas es igual a la unidad

Amplitud

Se llama amplitud de un [intervalo](#) de datos agrupados a la diferencia entre los valores de sus [extremos](#).

Análisis

Análisis factorial

Es un procedimiento técnico para reducir datos y encontrar grupos homogéneos de variables a partir de un conjunto numeroso más numeroso

[Análisis de varianza](#)

Aplastamiento

Sinónimo de [curtosis](#).

Asimetría

Asimetría de una [distribución de frecuencias](#) es la característica por la que los datos pierden su simetría respecto a la media. Expresado de otra forma, es el mayor o menor grado de desviación que existe entre la [media](#) (reparto equitativo) y la [mediana](#) (punto medio de la distribución). Existen varios coeficientes para su medida. El más simple es el de Pearson:

$$A_p = \frac{\text{media} - \text{moda}}{\text{desviación típica}}$$

B

Bernouilli

Una distribución de Bernouilli se compone de dos sucesos contrarios A y B, a los que se les suele llamar *éxito* y *fracaso*, con probabilidades p y q respectivamente

Binomial

Distribución binomial

Esta importante distribución se aplica a pruebas repetidas de la ley de [bernouilli](#), con las siguientes condiciones:

- a) Se realizan experimentos repetidos del tipo Bernouilli, n en total.
- b) La probabilidad p permanece constante en todos ellos
- c) Cada experimento es independiente del resultado anterior.

Su función de probabilidad es

$$B(n, k) = \binom{n}{k} p^k q^{n-k}$$

Bondad de ajuste

Es una medida del grado de ajuste entre los datos observados y los proporcionados por un modelo. El más popular es el que se realiza mediante la [Chi-cuadrado](#).

C

Campana de Gauss

Nombre asignado coloquialmente a la representación gráfica de la distribución [normal](#).

Característica

Es cualquier propiedad de objetos o personas que deseamos estudiar en Estadística

Censo

Es el estudio y recuento de todos los elementos de una [población](#). Si sólo se eligen algunos, se tratará de un [muestreo](#).

Chi-cuadrado

Es la [distribución teórica](#) que representa la distribución muestral de la suma de cuadrados de los datos dividida entre la varianza de la población.

$$\chi^2 = \sum \frac{(O - T)^2}{T^2}$$

Coeficiente

[De variación](#)

[De correlación](#)

[De determinación](#)

Confiabilidad

(Ver [Nivel de Confianza](#))

Constante

Llamaremos constante a una [característica](#) que sólo admite una [modalidad](#), por ejemplo la constante de gravitación universal

Continua

Una variable se llama continua si entre cada dos valores suyos pueden existir infinitos otros, como el peso, la estatura, etc.

Contraste

Contraste de hipótesis

Sinónimo de [Test de hipótesis](#)

Correlación

Coefficiente de Pearson

Es el cociente de dividir la [covarianza](#) de una [distribución bidimensional](#) entre las [desviaciones típicas](#) de X e Y respectivamente.

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Covarianza

Es la varianza conjunta en una distribución bidimensional X-Y. Se calcula como el cociente de los productos de las diferencias de X y de Y respecto a sus medias, entre el número de pares de la distribución.

$$\rho_{xy} = \frac{\sum \left((x - m_x)(y - m_y) \right)}{N}$$

Cualitativo/a

Se aplica a la variable (o dato, o medida) que sólo admite una [medida nominal](#)

Cuantil

Diremos que un número es el cuantil de orden p en una [distribución de frecuencias](#) si el porcentaje de datos inferiores a él es igual a p (y los superiores $100-p$). Por ejemplo, el cuantil C_{85} será un punto que cumple que el 85% de los datos es inferior a él. Según el valor de p , pueden denominarse [cuartiles](#), [deciles](#), [centiles](#), ...

Cuantitativo/a

Se aplica a la variable que admite medidas de intervalo o de razón (numéricas)

Cuartil

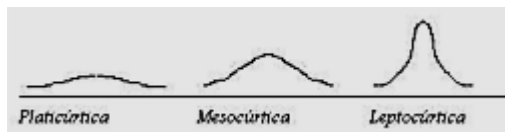
Los [cuantiles](#) que dividen a la distribución en cuatro partes iguales, es decir, C_{25} , C_{50} y C_{75} , reciben el nombre de cuartiles, y se representan por Q_1 o primer cuartil es el número que deja inferiores a él un 25% de los datos. Q_2 o segundo cuartil o mediana es el número que deja inferiores a él un 50% de los datos. Q_3 o tercer cuartil es el número que deja inferiores a él un 75% de los datos.

Cuasivarianza

Cuasivarianza o varianza insesgada es similar a la [varianza](#), pero dividiendo las sumas de cuadrados entre $n-1$. Es la varianza muestral.

Curtosis

Independientemente de su asimetría, una distribución puede presentar los datos con un reparto más uniforme, en el que las frecuencias sean muy parecidas. El gráfico aparecerá como aplastado y diremos que la distribución es **platicúrtica** o de poca curtosis. En el otro extremo, si las frecuencias cercanas al centro son mayores (con diferencia) que las alejadas, diremos que es **leptocúrtica** o con gran curtosis. Al caso intermedio lo denominaremos como distribución **mesocúrtica**



D

Dato

Es el valor cuantitativo o cualitativo que representa un atributo o medida en la [población](#). Puede ser nominal (textos) o cualitativos, o bien cuantitativo, (números).

Decil

Se suelen definir 9 deciles D1, D2, ... D9, que son los puntos que dividen al intervalo en diez partes iguales, correspondientes a los [cuantiles](#) de porcentajes 10%, 20%, ...90% respectivamente.

Desviación

Desviación media

Es una medida de la dispersión consistente en la media aritmética de las desviaciones individuales respecto a la media, tomadas en valor absoluto. También se usan desviaciones respecto a la mediana.

Desviación típica

Es la raíz cuadrada de la [varianza](#).

Desviación estándar

Es la raíz cuadrada de la [cuasivarianza](#) en la muestra.

Diagrama

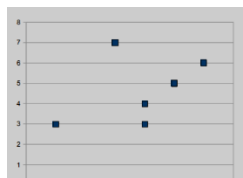
Es un gráfico dedicado a representar un conjunto de datos. Los más populares son de barras, columnas, histograma o sectores. No requieren explicación

De cajas y bigotes

Es un tipo de gráfico especial que permite visualizar los parámetros centrales y de posición, como media, mediana y percentiles.

De dispersión

Es un gráfico bidimensional en el que se representan los datos con puntos de coordenadas (X,Y)



Determinación

Coefficiente

El coeficiente de determinación es el cociente entre la [varianza explicada](#) y la [total](#) en un ajuste a la recta de regresión.

Dicotómico/a

Adjetivo que se aplica a toda medida o proceso que sólo puede presentar dos valores, como SÍ/NO, Hombre/Mujer, Encendido/Apagado.

Discreta

Si una variable solo admite un número finito de valores entre cada dos, recibirá el nombre de discreta (edades medidas en años, número de hermanos, etc.).

Distribución

De frecuencias

El conjunto formado por los valores de la variable y sus frecuencias constituye la distribución de frecuencias de la población o muestra, y se representa en las tablas de frecuencias.

Bidimensional

Si en un experimento todas las medidas que se obtienen son dobles, pertenecientes a dos variables distintas, a las que llamaremos X e Y respectivamente, se denominará distribución bidimensional a la formada por los pares X-Y de valores relacionados en ambas variables.

Muestral

Distribución muestral es la resultante de considerar, de forma teórica, todas las posibles muestras que se puedan elegir. Es una distribución teórica, construida sobre variables aleatorias, y sus elementos se obtienen mediante técnicas matemáticas.

Función de distribución

Distribución teórica

Llamaremos distribución teórica a la correspondiente distribución de probabilidades en una variable aleatoria.

Las principales distribuciones teóricas son:

Uniforme

Una distribución se llama uniforme cuando todas las probabilidades son iguales. Como todas suman 1 (caso discreto), cada una será igual a $1/n$.

De Bernouilli

Binomial

De Poisson

Normal

E

Ecuación

Ecuación de regresión lineal

Es la ecuación de la recta de regresión, del tipo $Y=aX+b$, que ayuda a encontrar los valores estimadas en una regresión lineal.

Error

Absoluto Es la diferencia entre el valor real y la medida obtenida

Relativo

Cociente entre el error absoluto y el dato real. Cuando este no se conoce se sustituye por la medida.

De predicción

Es la diferencia entre un valor de Y y su estimación Y' en una recta de regresión (o en una curva de regresión general)

Cuadrático

Similar a los anteriores, pero sumando los cuadrados de las diferencias entre valor real y la medida y extrayendo finalmente la raíz cuadrada.

Típico de estimación

Es la raíz cuadrada de la varianza residual en una operación de estimación.

Error de tipo I

Es el que se comete rechazando la hipótesis nula cuando en realidad es verdadera.

Error de tipo II

El error de aceptar la hipótesis nula cuando en realidad es falsa.

Escala

Escala de medida

Es un conjunto básico de [modalidades](#) y números (considerados como sus [medidas](#)) a partir del cual se construye un procedimiento para medir las restantes modalidades. Así, la escala centígrada de temperaturas se basa en asignar 0° a la temperatura de fusión del agua y 100° a la de ebullición

Escala nominal

Una escala se llama ***nominal*** si la única relación que tiene en cuenta es la de igualdad (y su contraria la desigualdad). Suele estar formada por nombres, códigos o números considerados como etiquetas (como el DNI).

Así, son nominales los apellidos, la Comunidad Autónoma, el distrito postal, etc.

Escala ordinal

La escala ***ordinal*** añade a la nominal la posibilidad de ordenar los datos, es decir, considera las relaciones de mayor y menor, aunque no se plantea una distancia entre unas medidas y otras. La escala de Insuficiente, Suficiente, Bien, Notable y Sobresaliente es ordinal.

Escala de intervalos

Se introduce una medida tipo (o patrón) llamada unidad y se tiene en cuenta cuantas unidades están comprendidas entre dos medidas distintas. Tienen sentido, además de la igualdad y el orden, las diferencias entre dos medidas. Podemos sumar y restar medidas, pero no tienen sentido sus cocientes. Son de intervalo la gran mayoría de las escala de las ciencias experimentales: temperatura, peso, velocidad, intensidad de la corriente eléctrica, etc.

Escala de razón

En esta escala se le da también un sentido a las razones entre dos medidas, es decir, las veces que una medida contiene a la otra. Fue la medida por excelencia de la Geometría griega y se ha trasladado a todas las Ciencias Sociales y de la Naturaleza. Se distingue también por la existencia de un cero verdadero, no convencional. Así, la escala centígrada de temperatura es sólo de intervalo y la Kelvin es de razón.

Esperanza

La esperanza matemática de una variable aleatoria discreta es la suma de los productos de sus valores por sus probabilidades. Equivale a la media en una distribución de frecuencias.

Estadístico

Se llama estadístico a todo valor numérico extraído mediante cálculos de los datos de una [muestra](#). Normalmente se usa para estimar un [parámetro](#) de la [población](#).

Estadístico de contraste

Es la expresión matemática, calculada a partir de la muestra, que nos servirá para tomar la decisión en un contraste de hipótesis.

Estimación

Es la operación por la que se asigna a un [parámetro](#) de la [población](#) el mismo valor que a un [estadístico](#) calculado a partir de una [muestra](#), que recibe el nombre de *estimador*.

Estimación puntual

Estimación por intervalos

Al ser la estimación una operación arriesgada, en lugar de apostar por una estimación puntual, se rodea esta de un intervalo de seguridad, que es el Intervalo de confianza.

Estimador

Es un estadístico calculado en una muestra que estima un parámetro de la población. Los más importantes son los que estiman la media y la varianza. Se llama *insesgado* si su media teórica coincide con el parámetro a estimar.

Extremo inferior

Es el valor mínimo que puede tener un valor incluido en un intervalo de datos agrupados.

Extremo superior

Es el valor máximo posible en un intervalo de datos agrupados. Se considera no alcanzable. Así si un intervalo comprende desde 5 hasta 10, incluiremos en el mismo los valores comprendidos entre estos dos, incluyendo el 5 y sin incluir el 10.

F

Frecuencia

Definición

El número de veces que se repite un valor concreto en una recogida de datos se llama frecuencia absoluta o simplemente frecuencia.

Frecuencia absoluta

Es sinónimo de frecuencia. Se representa por la letra n o por la f , según los distintos textos.

Frecuencia relativa o proporción

Es el cociente de dividir cada frecuencia absoluta entre el total de valores N . Se representa por f o por h .

Frecuencia acumulada

Es el número de datos del conjunto que son menores o iguales a un valor dado. Por tanto, se calculará sumando todas las frecuencias de datos menores o iguales al mismo. Podemos acumular las frecuencias absolutas y también las relativas y los porcentajes.

Frecuencias conjuntas

Son los pares de frecuencias formados en una distribución bidimensional

Frecuencia marginal

Llamaremos frecuencia marginal de un valor de X en una distribución bidimensional X-Y a la que le corresponde a ese valor si no tenemos en cuenta la existencia de Y. En la práctica coincide con la suma de todas las frecuencias contenidas en la fila correspondiente a ese valor.

Frecuencias condicionadas

Son las frecuencias que posee una variable si sólo consideramos un valor (o varios) de la otra variable en una distribución bidimensional X-Y. En la práctica se traduce a considerar sólo una fila o sólo una columna, según el valor elegido.

Función

Función de distribución

Llamaremos **función de distribución $F(x)$** de una variable aleatoria, a la formada por las probabilidades acumuladas, es decir: $F(m) = \text{Prob}(x \leq m)$ (El símbolo Prob designa a la probabilidad de que sea cierta la comparación del paréntesis)

G

Gauss

Distribución de Gauss

Sinónimo de [distribución normal](#).

Grados de libertad (g.l.)

Es el número de valores independientes que pueden variar libremente al no estar ligados entre sí. Por ejemplo, en una muestra simple de N elementos, la media y varianza se calculan a partir de $N-1$ g.l.

H

Hipótesis

Hipótesis nula

Llamaremos Hipótesis nula H_0 a la afirmación que hacemos sobre los parámetros de una población y cuya validez deseamos contrastar.

Hipótesis alternativa

Frente a la hipótesis nula podemos oponer otra, a la que llamamos hipótesis alternativa H_1 . Suele ser una desigualdad que se opone a la igualdad que afirmamos.

Test de hipótesis

Histograma

Representación gráfica de una distribución de datos agrupados en intervalos. Es similar al diagrama de barras, pero con los rectángulos adosados y de áreas proporcionales a las frecuencias de los intervalos.

|

Inferencia estadística

Es la operación de obtener conclusiones respecto a una población mediante los estimadores obtenidos en una muestra. Da nombre a la *Estadística Inferencial*.

Índice

Índice simple de base fija

Un término de la serie se identifica (convencionalmente) con el número 1, o el 100%. Diremos que este valor y_0 posee el índice 1. Para el resto de valores se define el índice como el cociente entre su propio valor y_i y el valor y_0 identificado como de índice 1.

Índice simple de base variable (o en cadena)

Tiene la misma definición que el anterior, pero en lugar de elegir un valor y_0 como base, en el cociente se toma el término anterior y_{i-1} .

Índice compuesto

Cuando se desea comparar la evolución de varios conjuntos a la vez, se definen índices compuestos, obtenidos combinando los índices simples. Una técnica sencilla es sustituir los múltiples valores de cada término por su media ponderada.

Índice de Gini

Es una medida convencional de la mayor o menor equidad en el reparto y viene representada por la concentración.

$$IG = \frac{\sum(p - q)}{\sum p}$$

Inferencia

Inferencia estadística

Es la ciencia que estudia las operaciones de [estimación](#)

Insesgado

Un estimador es insesgado cuando su media muestral coincide con el [parámetro](#)

Intervalo

Rango de datos definido por dos valores extremos. Por ejemplo: rango de edad entre 14 y 18 años.

Escala de intervalo

Intervalos en distribuciones de frecuencias

Si la variable que se estudia es continua, o discreta con muchos valores distintos, se organizarán sus datos en forma de intervalos, que son conjuntos formados por los números reales comprendidos entre un máximo y un mínimo.

Intervalo de confianza

Es el intervalo del que se rodea una estimación puntual acompañada de una probabilidad de que el parámetro estimado pertenezca a dicho intervalo.

L

Leptocúrtica

Distribución de frecuencias con gran curtosis.

Ley

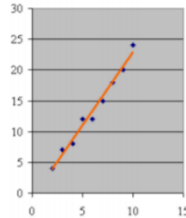
Ley de los grandes números

"Las frecuencias observadas tienen como límite las probabilidades cuando n tiende al infinito"

Línea de regresión

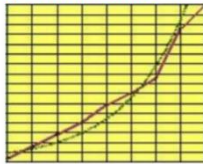
Es la línea que mejor representa una distribución de datos bidimensional, obtenida generalmente mediante un ajuste de mínimos cuadrados. Puede ser *lineal*, *polinómica*, *logarítmica* o *exponencial*, que son las más usadas.

Lineal



Exponencial

Exponencial creciente



Exponencial decreciente



Logarítmica



M

Marca de clase

Promedio entre los dos extremos (o punto medio de un intervalo de datos agrupados), que se elige como representante de todos los valores comprendidos.

Media

Media aritmética

Llamaremos media aritmética o simplemente media al valor resultante de sumar todos los datos y después dividir el resultado entre el número de ellos.

Media geométrica

Es la raíz enésima del producto de los datos. Se usa cuando el producto es más representativo que la suma, como ocurre cuando se promedian cocientes o razones.

Media armónica

Es la media diseñada para promediar cantidades inversamente proporcionales y equivale al inverso de la media de los inversos de x

Media cuadrática

Es muy usada en la teoría de errores y en estudios sobre ajustes de datos. Es la raíz cuadrada de la media de los cuadrados de los datos.

Media ponderada

En esta media se multiplica cada dato por un peso (valor numérico), se suman todos los productos se divide el resultado entre la suma e los pesos.

Mediana

Llamaremos mediana de un conjunto de datos de tipo ordinal (o de intervalo o razón) al dato que ocupa el punto medio de la distribución ordenada de datos. Es decir, es el punto que divide a la distribución en dos partes iguales: el total de frecuencias de los datos inferiores a la mediana es igual al de las frecuencias de los datos mayores.

Medida

Es la operación de asignar un número a cada una de las [modalidades](#) de una [característica](#), convirtiendo algunas relaciones entre modalidades en sus correspondientes relaciones entre los números que representan su medida.

Medida directa

Llamaremos medida directa en cualquier estudio o experimento, a aquella que se ha obtenido directamente

sobre los objetos, individuos o entidades con los instrumentos usuales de medida.

Medida diferencial

Dada una medida directa X , llamaremos **medida diferencial** x a su diferencia con la [media](#) del grupo.

Medida típica Z

Si se divide una medida diferencial entre la desviación típica del grupo, se obtiene la medida o **puntuación típica Z**.

$$Z = \frac{X - \bar{X}}{\sigma}$$

Mesocúrtica

Distribución de frecuencias con [curtosis](#) media.

Mínimos cuadrados

Técnica para ajustar datos a un modelo mediante la búsqueda del mínimo valor de la suma los cuadrados de las diferencias entre los datos reales y los proporcionados por el modelo,

Moda

Llamaremos **Moda** al valor de la distribución de datos que presente una frecuencia mayor.

Modalidad

Las distintas formas de presentarse una característica se llaman *modalidades*. Por ejemplo, 1,82 y 1,65 cm. son dos modalidades de la característica *altura*, y varón y mujer dos modalidades de la característica *sexo*

Muestra

Definición

Es un subconjunto de la población que es más fácil de estudiar que la [población](#).

Muestreo

Definición

Es un conjunto de operaciones o técnicas dirigidos a la elección de la [muestra](#) adecuada.

N

Nivel de confianza

Es la probabilidad de que un valor estimado pertenezca al intervalo de confianza que rodea a la estimación. Los más usados son 90%, 95% y 99%

Nivel de significación

La probabilidad de que unos valores caigan en la región de rechazo en un contraste de hipótesis, a pesar de que H_0 sea verdadera, se conoce con el nombre de nivel de significación α ,

Nominal

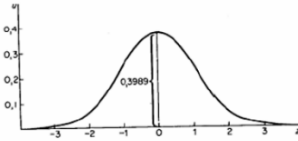
Escala Nominal

Normal

Distribución normal

La distribución Normal o ley de Gauss es la más usada de las distribuciones teóricas continuas. La popularizaron Gauss, en el estudio de los errores de las medidas, y también Laplace, pero ya la había usado Moivre como límite de la binomial.

Por su característica forma, se la conoce también como campana de Gauss.



Aquí sólo nos interesa su definición y uso dentro de la Estadística. La expresión de su función de densidad con media 0 y desviación típica 1 es

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

O

Ordenada en el origen

Su significado más usual es el del término independiente de la ecuación de la [recta de regresión](#). Se puede representar como el corte de esa recta con el eje Y.

P

Parámetro

Un número que caracterice o describa una [población](#) recibe el nombre de parámetro. La estatura media de los alumnos y alumnas de 16 años es un parámetro de esa población, o la Renta per cápita de la población española

Pendiente

Su significado más usual es el del coeficiente de la variable X en la [recta de regresión](#) lineal.

Percentil

Similares a los [deciles](#), P1, P2, P3,P99, son 99 números que dividen la distribución en 100 partes iguales.

Poisson

Esta distribución, llamada de los sucesos raros, es el caso límite de la binomial, con las siguientes condiciones:

- a) El número de intentos n debe tender a infinito.
- b) La propiedad p debe ser muy pequeña (de ahí el nombre de suceso raro)

d) El producto de $n \cdot p$ ha de ser constante, y al que llamaremos m .

Fórmula:

$$p(x) = \frac{e^{-m}}{x!} m^x$$

Platicúrtica

Distribución de frecuencias con poca [curtosis](#).

Población

Llamaremos población a un conjunto bien definido por ciertas características que deseamos estudiar: La población de una Comunidad Autónoma, los aprobados de 2º de Bachillerato en mi Centro, los profesores de E.S.O. en la Delegación Norte, etc.

Porcentaje

Equivale a la frecuencia relativa expresada como tanto por ciento o porcentaje.

Predicción

Llamaremos pronóstico o predicción para un valor de X a su imagen Y' en la recta de regresión. Esta definición se extiende a cualquier otra curva de ajuste de datos.

Proporción

Es sinónimo de [frecuencia relativa](#)

Puntual

La [estimación](#) se llama puntual cuando identificamos, sin más, el [parámetro](#) con el [estadístico](#). En ese caso añadiremos un acento circunflejo al parámetro para representar que estamos estimando.

P-valor

El p-valor de un resultado en un experimento es la probabilidad de obtener ese valor u otros menores (o mayores, según sea el experimento) si se satisface la hipótesis nula.

R

Rango

Si se fija un valor mínimo y otro máximo, de forma que todos los datos de un recuento estén comprendidos entre ellos (a veces esto no se garantiza y quedan intervalos

abiertos), la diferencia entre ambos se denomina **rango** de los datos.

Rango percentil

Es la medida inversa del percentil. Dada una medida concreta, como puede ser la calificación de una alumna en Música, su rango percentil equivale al percentil más cercano a esa calificación. Un alumno que tenga rango percentil de 78 es aquel en el que el 78% de sus compañeros tiene una puntuación inferior a él.

Razón

Se llama *razón* a la comparación entre dos medidas mediante su división.

Escala de razón

Regresión

Recta de regresión

La recta de regresión de Y sobre X es aquella que minimiza la suma de cuadrados de las diferencias entre los valores de Y y los correspondientes Y' medidos en dicha recta.

Línea de regresión

Residuos

Son las diferencias que se producen en la regresión entre el valor estimado y el real. Forman la varianza residual o no explicada.

S

Sesgo

Sinónimo de [asimetría](#)

Sumas de cuadrados

En ANOVA

Total

Es la suma de las diferencias al cuadrado entre los datos experimentales y su media.

Intragrupos

Representa la suma de cuadrados corregidos que se da dentro de los grupos, es decir, las diferencias de los datos entre la media de cada grupo.

Intergrupos

Es la suma ponderada de las diferencias al cuadrado entre las medias de los grupos y la media total.

Interacción

En un modelo con varios factores representa la influencia mutua entre ellos.

Supuesto

Es una afirmación que se hace de una población en la Estadística Inferencial: si es simétrica, normal, continua... y sobre la muestra, si es aleatoria simple, es de tamaño mayor que 30...

T

T de Student

Distribución que sigue la estimación de la desviación típica.

Tasa

Es la relación entre el número de casos de una categoría determinada y el número total de la población: tasa de

sospensos, tasa de pobreza, ... Generalmente se expresa mediante porcentajes.

Teorema

Teorema central del límite

Si las variables $x_1, x_2, x_3, \dots, x_n$, tienen todas la misma distribución, con los mismos valores m para la media y s para la desviación típica, la variable

$$\frac{x_1 + x_2 + \dots + x_n - nm}{s\sqrt{n}}$$

sigue asintóticamente la distribución normal $N(0,1)$.

Test

Test de hipótesis

Un test de hipótesis (o contraste) es un proceso, compuesto de varios pasos muy concretos, que nos permite aceptar o rechazar una hipótesis en términos estadísticos.

Tipificación

Es la operación de convertir una medida en típica restándole la media y dividiendo entre la desviación típica.

U

Uniforme

Distribución uniforme

V

Variable

Variable aleatoria

Llamaremos **Variable aleatoria simple** (discreta) a un conjunto de valores $X_1, X_2, X_3, \dots, X_n$ (llamados también *sucesos*) a los que les corresponden unos números (llamados *probabilidades*) , $p_1, p_2, p_3, \dots, p_n$ que cumplen:

- a) Todas las probabilidades son positivas o nulas.
- b) La suma de todas ellas es igual a la unidad

Variación

Coeficiente de variación

Es el cociente de dividir la desviación típica entre la media.

Varianza

Definición

Es el cociente de dividir la suma de los cuadrados de las desviaciones de los datos respecto a la media entre el número total de datos. Su raíz cuadrada es la [desviación típica](#).

Cuasivarianza

Idéntica a la anterior, pero se divide entre el número de datos menos la unidad, porque pueden ser los grados de libertad.

Explicada

Es la parte de una varianza que se considera producida por un factor determinado que influya en un experimento. En la regresión lineal es la varianza de las [predicciones](#).

Total

Es la varianza total observada en un experimento, independientemente de las variables que puedan influir en los resultados.

Residual

Es la diferencia entre la varianza total y la explicada.

Análisis de Varianza ANOVA

La técnica del Análisis de la Varianza consiste en descomponer la variabilidad de una población (representada por su varianza) en diversos sumandos según los factores que intervengan en la creación de esa variabilidad.

Análisis de covarianza

Si en el ANOVA se introduce el modelo de regresión, se denomina ANCOVA.