

GLOSARIO DE TÉRMINOS DE ESTADÍSTICA

(Tomado de: <http://www.cortland.edu/flteach/stats/glos-sp.html>)

COMPRESION Y USO DE LA ESTADÍSTICA

Fernando Valdes - Universidad Romulo Gallegos
(Robert Ponterio - Project Director and HTML)
SUNY Cortland Dept. of International Communications and Culture

GLOSARIO

Al Azar o Aleatorio:  randomness, random;  le hasard, aléatoire

Son todos aquellos eventos fortuitos o productos de la suerte.

Aleatoriamente:  randomly;  au hasard

Actividades o métodos producidos o llevados a cabo simulando un comportamiento al azar.

Clase:  class;  la classe

Subdivisión de escala de datos.

Correlación:  correlation;  la corrélation

Cuando dos fenómenos sociales, físicos o biológicos crecen o decrecen de forma simultánea y proporcional debido a factores externos, se dice que los fenómenos están positivamente correlacionados. Si uno crece en la misma proporción que el otro decrece, los dos fenómenos están negativamente correlacionados. El grado de correlación se calcula aplicando un [coeficiente](#) de correlación a los datos de ambos fenómenos. Una correlación positiva perfecta tiene un [coeficiente](#) + 1, y para una correlación negativa perfecta es -1. La ausencia de correlación da como [coeficiente](#) 0. Por ejemplo, el coeficiente 0,89 indica una correlación positiva grande, -0,76 es una correlación negativa grande y 0,13 es una correlación positiva pequeña. ([Correlación](#))

Cuartiles, Deciles y Percentiles:  quartile, decile, percentile;  le quartile, le décile, le centile/percentile

Si un conjunto de datos está ordenado por magnitud, el valor central (o la media de los dos centrales) que divide al conjunto en dos mitades iguales, es la mediana.

Extendiendo esa idea, podemos pensar en aquellos valores que dividen al conjunto de datos en cuatro partes iguales. Esos valores denotados Q_1 , Q_2 , y Q_3 , se llaman **primer cuartíl**, **segundo cuartíl** y **tercer cuartíl**, respectivamente. EL Q_2 coincide con la mediana.

Análogamente, los valores que dividen a los datos en 10 partes iguales se llaman deciles, y se le denotan $D_1, D_2, \dots, D_9$, mientras que los valores que lo dividen en 100 partes iguales se llaman percentiles, denotados por $P_1, P_2, \dots, P_{99}$. El 5º decil y el 50º percentil coinciden con la mediana. Los 25º y 75º percentiles coinciden con el primer y tercer cuartiles.

Colectivamente, cuartiles, deciles y percentiles se denominan cuantiles.

Curva Normal:  *normal, Bell curve*;  *la courbe Laplace-Gauss, normale*

La forma prevista de la [distribución normal](#).

Datos:  *data*;  *les données*

Son los valores cualitativos o cuantitativos mediante los cuales se miden las características de los objetos, sucesos o fenómenos a estudiar.

Desviación:  *deviation*;  *l'écart*

Diferencia entre un valor y otro valor medio o típico. ([Desviación Media](#))

Desviación Típica:  *standard deviation*;  *l'écart type*

La desviación de un elemento del conjunto es su diferencia con respecto a la medida; por ejemplo, en sucesión $x_1, x_2, \dots, x_n$ la desviación de x_1 es $x_1 - \bar{x}$, y el cuadrado de la desviación es $(x_1 - \bar{x})^2$. La [varianza](#) es la medida del cuadro de las desviaciones. Por último, la desviación típica, representada por la letra griega sigma (σ), es la raíz cuadrada de la [varianza](#), y se calcula de la siguiente manera:

$$\sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

Si la desviación típica es pequeña, los datos están agrupados cerca de la media; si es grande, están muy dispersos.

Es una medida de dispersión expresada por el espacio de la curva normal comprendido entre la media y $\pm \sigma$.

Entrevista y Encuesta:  *interview, survey, poll;*  *une entrevue (interviewer), une enquête, un sondage*

Son métodos de recolección de datos, la entrevista es una serie de preguntas realizadas personalmente y la encuesta es llevada a cabo generalmente a través de algún formulario que la persona debe llenar.

Estadística:  *statistics;*  *les statistiques*

La Estadística estudia los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas con tal análisis.

En un sentido menos amplio, el término estadística se usa para denotar los propios datos, o números derivados de ellos, tales como los promedios. Así se habla de estadística de empleo, estadística de accidentes.

Estadístico:  *statistic;*  *une statistique (un statisticien)*

Unidad de medida referente a la muestra. Se le llama estadístico también a la persona que trabaja con la estadística.

Frecuencia:  *frequency;*  *la fréquence*

Número de veces en que se repite un dato.

Frecuencia Acumulada:  *cumulative frequency;*  *fréquence cumulée*

Es el número de estudiantes con calificaciones iguales o menores que el rango de cada intervalo sucesivo. (Frecuencia)

Frecuencia Relativa:  *relative frequency;*  *fréquence relative*

Es la proporción entre la frecuencia de un intervalo y el número total de datos.

Histograma:  *histogram;*  *un histogramme*

Es una serie de rectángulos con bases iguales al rango de los intervalos y con área proporcional a sus frecuencias.

Inferir:  *to infer*;  *inférer*

Es emitir juicios o conclusiones basados en algún conocimiento o experiencia sobre un evento o suceso.

Inferencia Estadística:  *statistical inference*;  *une inférence statistique*

Es aplicar resultados de estudios de una muestra a la poblaciones y emitir juicios o conclusiones sobre esa población en general. (Estadística)

Ejemplo:

Estudio: Causas de la deserción estudiantil en la Universidad Rómulo Gallegos.

Población: 4000 alumnos.

Muestra: 10% de la población.

Resultados del Estudio de la Muestra: La situación económica, dificultad en el aprendizaje.

Conclusiones: Se puede inferir que las causas de deserción de los alumnos de la Universidad Rómulo Gallegos son la situación económica y la dificultad en el aprendizaje

Intervalo de Clase:  *class interval*;  *un intervalle de classe*

Pequeña sección de la escala según la cual se agrupan las puntuaciones de una distribución de frecuencia. Tamaño o rango de la **Clase**.

Límites del Intervalo:  *interval limits*;  *limites de l'intervalle*

Son los valores extremos que tiene el intervalo de clase, inferior y superior, entre los cuales van a estar los valores de los datos agrupados en ese intervalo de clase.

Media Aritmética:  *arithmetic mean*;  *la moyenne arithmétique*

La media de un conjunto de N números, $X_1, X_2, X_3, \dots, X_N$. Se define por:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N} = \frac{\sum_{j=1}^N x_j}{N} = \frac{\sum x}{N}$$

(Propiedades de la Media Aritmética)

(Relación Empírica entre Media, Mediana y Moda)

(Media Armónica H)

(Media Geométrica G)

(Relación entre las Medias Aritmética, Geométrica y Armónica)

Media Aritmética Ponderada:  *weighted arithmetic mean*;  *la moyenne arithmétique pondérée*

A veces asociamos con los números $X_1, X_2, X_3, \dots, X_k$ ciertos factores peso (o pesos) $W_1, W_2, W_3, \dots, W_k$ dependientes de la relevancia asignada a cada número. En tal caso,

$$\bar{X} = \frac{w_1 X_1 + w_2 X_2 + \dots + w_k X_k}{w_1 + w_2 + \dots + w_k} = \frac{\sum wX}{\sum w}$$

Se llama la media aritmética ponderada con pesos $f_1, f_2, \dots, f_k$.

Ejemplo: Si el examen final de un curso cuenta tres veces más que una evaluación parcial, y un estudiante tiene calificación 85 en el examen final y 70 y 90 en los dos parciales, la calificación media es:

$$\bar{X} = \frac{(1)(70) + (1)(90) + (3)(85)}{1 + 1 + 3} = 83$$

Propiedades de la Media Aritmética

1. La suma algebraica de las desviaciones de un conjunto de números respecto de su media aritmética es cero.
2. La suma de los cuadrados de las desviaciones de un conjunto de números X_j respecto de un cierto número "a" es mínima si y sólo si

$a = \bar{X}$ Si f_1 números tienen media m_1 , f_2 números tienen $m_2, \dots, f_k$ números tienen media m_k , entonces la media de todos los números es,

$$\bar{X} = \frac{f_1 m_1 + f_2 m_2 + \dots + f_k m_k}{f_1 + f_2 + \dots + f_k}$$

Es decir, una media aritmética ponderada de todas las medias.

3. Si A es una media aritmética supuesta o conjeturada (que puede ser cualquier número) y si $d_j = X_j - A$ son las desviaciones de X_j respecto de A, las ecuaciones (1) y (2) se convierten, respectivamente, en

$$\bar{X} = A + \frac{\sum_{j=1}^k f_j d_j}{\sum_{j=1}^k f_j} = A + \frac{\sum f d}{N}$$

$$\bar{X} = A + \frac{\sum_{j=1}^N d_j}{N} = A + \frac{\sum d}{N}$$

$$\text{Donde } N = \sum_{j=1}^k f_j = \sum f$$

Mediana:  *median*;  *la médiane*

La mediana de un conjunto de números ordenados en magnitud es o el valor central o la media de los dos valores centrales.

Ejemplo: El conjunto de números 3,4,4,5,6,8,8,8 y 10 tiene mediana 6.

Ejemplo: El conjunto de números 5,5,7,9,11,12,15 y 18 tiene mediana.

$$\text{Mediana} = \frac{1}{2}(9 + 11) = 10$$

Para datos agrupados, la mediana obtenida por interpolación viene dada por:

$$L_i + \left(\frac{\frac{N}{2} - (\sum f_i)}{f_{\text{mediana}}} \right) c$$

Mediana=

Donde:

L_i = frontera inferior de la clase de la mediana.

N = número de datos (frecuencia total).

$(\sum f_i)$ = suma de frecuencia de las clases inferiores a la de la mediana

f_{mediana} = frecuencia de la clase de la mediana

c = anchura del intervalo de la clase de la mediana.

Geométricamente la mediana es el valor de X (abscisa) que corresponde a la recta vertical que divide un histograma en dos partes de igual área. Ese valor de X se suele denotar por $\tilde{X}$.

(Relación Empírica entre Media, Mediana y Moda)

Medidas de la Dispersión:  *measures of dispersion*;  *measures de dispersion*

Normalmente la estadística también se ocupa de la dispersión de la distribución, es decir, si los datos aparecen sobre todo alrededor de la media o si están distribuidos por todo el rango. Una medida de la dispersión es la diferencia entre dos percentiles, por lo general entre el 25 y el 75. El percentil es un número tal que un por ciento de los

datos son menores o iguales que . En particular, los percentiles 25 y 75 se denominan cuartiles inferior y superior respectivamente. La desviación típica es otra medida de la dispersión, pero más útil que los percentiles, pues está definida en términos aritméticos como se explica a continuación.

Medidas de Tendencia Central:  *measures of central tendency*;  *les mesures de tendance centrale*

Un promedio es un valor típico o representativo de un conjunto de datos. tales valores suelen situarse hacia al centro del conjunto de datos ordenados por magnitud.

Una vez que los datos han sido reunidos y tabulados, comienza el análisis con el objeto de calcular un número único, que represente o resuma todos los datos. Dado que por lo general la frecuencia de los intervalos centrales es mayor que el resto, este número se suele denominar valor o medida de la tendencia central.

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

Sean $X_1, X_2, \dots, X_n$ los datos de un estudio estadístico. El valor utilizado mas a menudo es la media aritmética o promedio aritmético que se escribe $\bar{X}$, y que es igual a la suma de todos los valores dividida por n .

El símbolo $\sum$ o sumatorio, denota la suma de todos los datos. Si las X se agrupan en K intervalos, con puntos medios $m_1, m_2, \dots, m_k$ y frecuencias $f_1, f_2, \dots, f_k$, la media aritmética viene dada por:

$$\text{Medis}_{\text{Aritmetica}} = \frac{\sum (f_i * m_i)}{\sum f_i}$$

donde $i=1,2,\dots,k$.

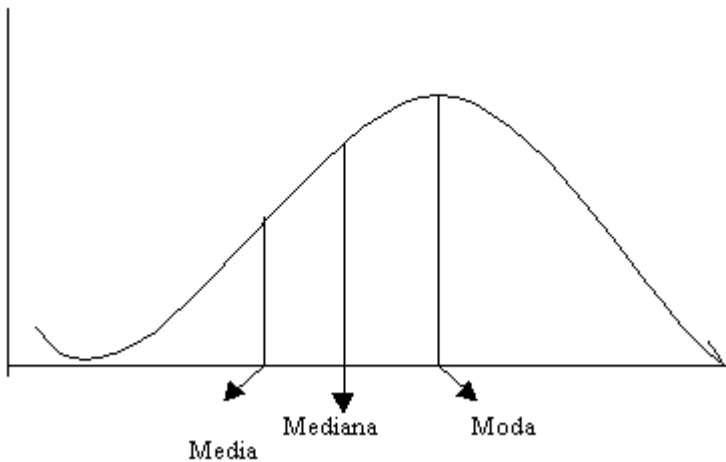
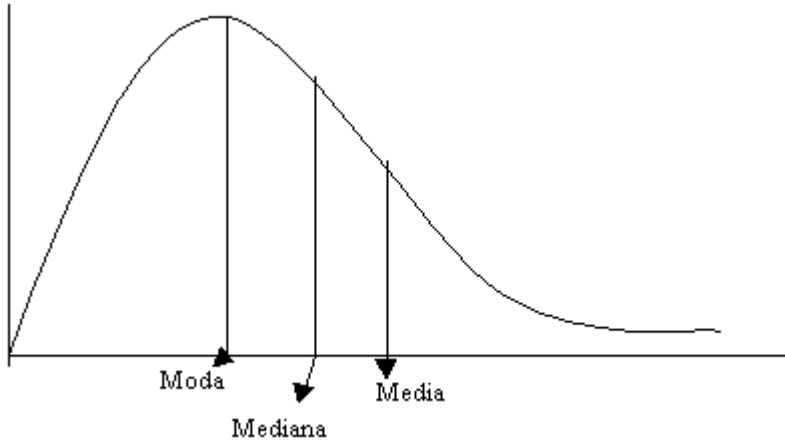
La mediana y la moda son otros dos valores de la tendencia central. Si las x se ordenan según sus valores numéricos, si n es impar la mediana es la x que ocupa la posición central y si n es par la mediana es la media o promedio de las dos x centrales. La moda es la x que aparece con mayor frecuencia. Si dos o más x aparecen con igual máxima frecuencia, se dice que el conjunto de las x no tiene moda, o es bimodal, siendo la moda las dos x que aparecen con más frecuencia, o es trimodal, con modas las tres x más frecuentes.

Relación Empírica entre Media, Mediana y Moda  *mean, median, mode*;  *la moyenne, la médiane, le mode*

Para curvas de frecuencia unimodales que sean poco asimétricas tenemos la siguiente relación empírica

$$\text{Media} - \text{Moda} = 3(\text{media} - \text{mediana})$$

Posiciones relativas de la media, la mediana y la moda para curvas de frecuencia asimétrica a derecha e izquierda, respectivamente, para curvas simétricas los tres valores coinciden.



Moda:  *mode*;  *le mode*

Es el valor que ocurre con mayor frecuencia; es decir, el valor más frecuente. La moda puede no existir, e incluso no ser única en caso de existir.

Ejemplo: El conjunto 2,2,5,7,9,9,9,10,10,11,12 y 18 tiene moda 9.

Ejemplo: El conjunto 3,5,8,10,12,15 y 16 no tiene moda.

Ejemplo: El conjunto 2,3,4,4,4,5,5,7,7,7 y 9 tiene dos modas, 4 y 7 y se llama bimodal.

Una distribución con moda única se dice *unimodal*.

En el caso de datos agrupados donde se haya construido una curva de frecuencias para ajustar los datos, la moda será el valor (o valores) de X correspondiente al máximo (o máximos) de la curva. Ese valor de X se denota por $\hat{X}$.

La moda puede deducirse de una distribución de frecuencias o de un histograma a partir de la fórmula:

$$Moda = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c$$

donde:

L_1 = frontera inferior de la clase modal (clase que contiene a la moda).

Δ_1 = exceso de la frecuencia modal sobre la de la clase inferior inmediata.

Δ_2 = exceso de la frecuencia modal sobre la de la clase superior inmediata.

c = anchura del intervalo de clase modal.

(Relación Empírica entre Media, Mediana y Moda)

Observación de Campo:  *field observation*;  *enquête sur le terrain*

Es la observación que realiza el investigador en el lugar donde ocurren los sucesos o eventos, para extraer los datos.

Población y Muestra:  *population & sample*;  *la population & un échantillon*

Al recoger datos relativos a las características de un grupo de individuos u objetos, sean alturas y pesos de estudiantes de una universidad o tuercas defectuosas producidas en una fábrica, suele ser imposible o nada práctico observar todo el grupo, en especial si es muy grande. En vez de examinar el grupo entero, llamado *población* o *universo*, se examina una pequeña parte del grupo, llamada *muestra*.

Una población puede ser finita o infinita. Por ejemplo, la población consistente en todas las tuercas producidas por una fábrica un cierto día es finita, mientras que la determinada por todos los posibles resultados (caras, cruces) de sucesivas tiradas de una moneda, es infinita.

Si una muestra es representativa de una población, es posible inferir importantes conclusiones sobre las poblaciones a partir del análisis de la muestra. La fase de la estadística que trata con las condiciones bajo las cuales tal diferencia es válida se llama estadística inductiva o inferencia estadística. Ya que dicha inferencia no es del todo exacta, el lenguaje de las probabilidades aparecerá al establecer nuestras conclusiones.

La parte de la estadística que sólo se ocupa de describir y analizar un grupo dado, sin sacar conclusiones sobre un grupo mayor, se llama estadística descriptiva o deductiva.

Polígono de frecuencia:  *frequency polygon*;  *le polygone des fréquences*

Se obtiene conectando los puntos medios de cada intervalo de un histograma de frecuencias acumuladas con segmentos rectilíneos.

Porcentaje:  *percentage*;  *le pourcentage*

Es la proporción de una cantidad de datos específicos, con respecto al total de esos datos.

Probabilidad:  *probability*;  *la probabilité*

Es el conjunto de posibilidades de que un evento ocurra o no en un momento y tiempo determinado. Dichos eventos pueden ser medibles a través de una escala de 0 a 1, donde el evento que no pueda ocurrir tiene una probabilidad de 0 y uno que ocurra con certeza es de 1.

Ejemplo: Cuando se lanza una moneda, se desea saber cual es la probabilidad de que se selle o cara, es decir existe un 0,5 (50%) de que sea cara o 0,5 (50%) de que sea sello.

Promedio:  *average*;  *le milieu*

Es una medida que caracteriza un grupo de datos bajo algún criterio. Como: la media aritmética y la media ponderada.

Rango:  *rank*;  *le rang*

Situación de un dato respecto de una distribución.

Tamaño de la Muestra:  *sample size*;  *la taille de l'échantillon*

Es la cantidad de datos que serán extraídos de la población para formar parte de la muestra.

Validez:  *validity*;  *la validité*

Importancia predictiva para los propósitos que se persiguen. (Validez)

Project developed under the LxC Select Languages Across the Curriculum grant from FIPSE. Additional funding provided by SUNY Office of Educational Technology [Languages Across the Curriculum Multimedia Development Project](#).

W3 page maintained by PonterioR@cortland.edu.
Copyright © 1998 Robert Ponterio, Fernando Valdes