



Universidad Nacional del Centro del Perú

Facultad de Economía

Informática para Economistas

2025-II

Cel. : 972370070

E-mail : cmachacuay@uncp.edu.pe

SEMANA 4

Fundamentos Básicos de la Estadística

M.Sc. Ciro Ivan Machacuay Meza
Docente

Contenido

1 Introducción a la Estadística

2 Estadística descriptiva básica

Introducción a la Estadística

Definición y objetivos de la estadística.

Es la ciencia que proporciona un conjunto de métodos, técnicas o procedimientos para recopilar, organizar (clasificar, agrupar), presentar y analizar datos con el fin de describirlos o de realizar generalizaciones válidas sobre los mismos. Así, se tiene la siguiente división : Descriptiva e inferencial

Estadística descriptiva

Conjunto de métodos estadísticos que se relacionan con el resumen y descripción de los datos, como tablas, gráficos y el análisis mediante algunos cálculos.

Estadística inferencial

Conjunto de métodos con los que se hacen las generalizaciones o la inferencia sobre una población utilizando una muestra. La inferencia puede contener conclusiones que pueden ser no ciertas en forma absoluta, por lo que es necesario que éstas sean dadas con una medida de confiabilidad, que es la probabilidad. Estas dos partes de la estadística no son mutuamente excluyentes, ya que para utilizar los métodos de la inferencia estadística, se requiere conocer los métodos de la estadística descriptiva.

Objetivo de la estadística

El objetivo principal de la estadística es extraer información significativa a partir de datos para apoyar la toma de decisiones, comprender fenómenos y predecir comportamientos.

Objetivo específicos de la estadística

Describir y resumir datos :

- (1) Organizar información compleja mediante medidas de tendencia central (media, mediana, moda) y dispersión (desviación estándar, rango).
- (2) Representar datos visualmente con gráficos (histogramas, diagramas de dispersión, etc.). Ejemplo : Calcular el promedio de ingresos en una población.

Objetivo específicos de la estadística

Inferir conclusiones sobre poblaciones :

- (1) Generalizar resultados de una muestra a una población usando técnicas como intervalos de confianza o pruebas de hipótesis. Ejemplo : Estimar la preferencia electoral nacional basándose en encuestas.

Objetivo específicos de la estadística

Identificar relaciones entre variables :

- (1) Analizar correlaciones, regresiones o patrones para entender cómo interactúan los factores. Ejemplo : Determinar si el nivel educativo está relacionado con el salario.

Objetivo específicos de la estadística

Predecir comportamientos futuros :

- (1) Usar modelos estadísticos (series temporales, machine learning) para anticipar tendencias. Ejemplo : Predecir ventas anuales de un producto.

Objetivo específicos de la estadística

Apoyar la toma de decisiones :

- (1) Reducir la incertidumbre en ámbitos como negocios, medicina, políticas públicas o ciencia. Ejemplo : Decidir si lanzar un fármaco basándose en ensayos clínicos.

Objetivo específicos de la estadística

Controlar procesos :

- (1) Monitorear la calidad en producción (gráficos de control) o evaluar impactos de intervenciones. Ejemplo : Asegurar que las botellas de agua tengan $500 \text{ ml} \pm 5 \text{ ml}$ de contenido.

Tipos de datos : cualitativos y cuantitativos.

Los datos se clasifican en dos grandes categorías según su naturaleza : cualitativos (nominales/ordinales) y cuantitativos (discretos/continuos). Esta distinción es clave para elegir el análisis estadístico adecuado.

- (1) Datos Cualitativos (o Categóricos) Representan cualidades, atributos o categorías no medibles numéricamente.
- (2) Datos Cuantitativos (o Numéricos) Representan cantidades medibles con valores numéricos.

¿Qué es una Variable?

Una variable es una característica, atributo o propiedad que puede variar (cambiar) entre individuos, objetos o situaciones, *y que es susceptible de ser medida u observada.*

- (1) Cualitativas (Categóricas) : Representan categorías o cualidades. Ejemplos : Género, Color de ojos, Nivel educativo.
- (2) Cuantitativas (Numéricas) : Representan cantidades medibles. Ejemplos : Edad, Ingresos, Peso.

¿Qué es una Escala de Medición?

Las escalas de medición definen cómo se miden las variables y qué operaciones matemáticas o estadísticas son válidas con ellas. Existen 4 tipos, ordenadas de menor a mayor precisión :

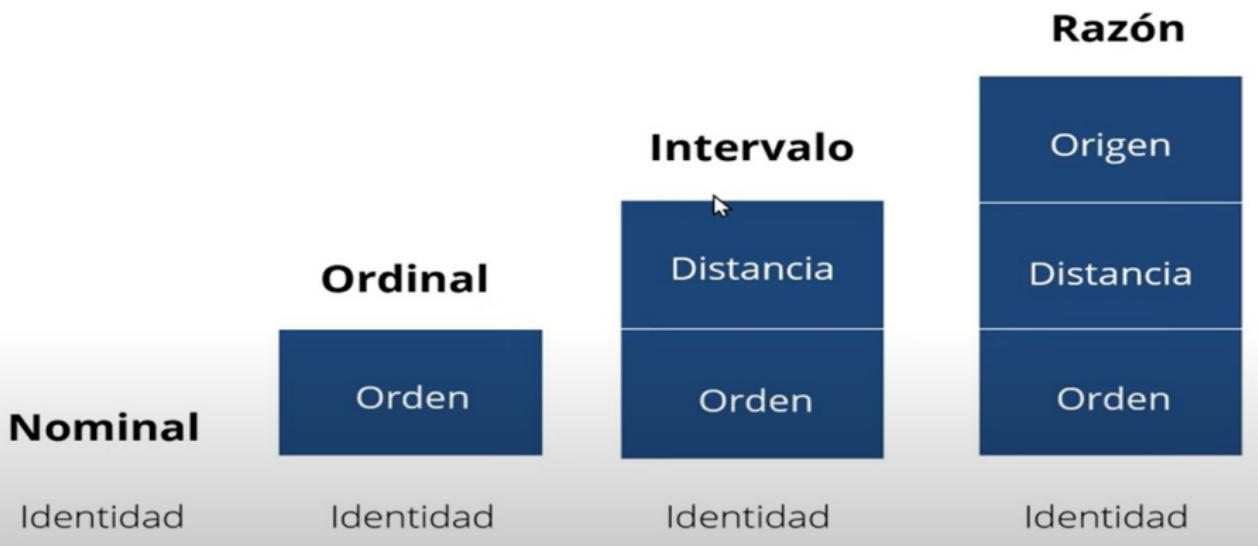
(1) Nominal

(2) Intervalo

(3) Razon

(4) Ordinal

¿Qué es una Escala de Medición?



¿Qué es una Escala de Medición?

Tipo	VARIABLES CATEGÓRICAS						VARIABLES NUMÉRICAS											
	CUALITATIVAS			CUANTITATIVAS														
Naturaleza	NOMINAL			ORDINAL			INTERVALO			RAZÓN								
	Ningún atributo			Un atributo			Dos atributos			Tres atributos								
Atributos de la escala	Orden	Distancia	Origen	Orden	Distancia	Origen	Orden	Distancia	Origen	Orden	Distancia	Origen						
Características	Posee categorías a las que se le asigna un nombre sin que exista ningún orden expreso entre ellas.			Posee categorías ordenadas, pero no permite cuantificar la distancia entre una categoría y otra.			Tiene intervalos medibles entre valores, al no tener un origen real; Puede asumir valores negativos.			Tiene intervalos medibles, con un origen real; donde el cero significa la ausencia del individuo.								
Ejemplos	Género		Estado Civil	Instrucción	Intensidad	Temperatura °C	Saldo Bancario \$		Peso (Kg)	Hijos (Unidades)								
Valores Finales	Masculino	Soltero	Primaria	Leve	-10.0	0.0	72.35	Uno	Femenino	Casado	Secundaria	Moderado	20.0	-15.0	80.24	65.45	Dos	Tres
Observaciones	Dicotómicas: Tienen solamente dos categorías Ejemplos de Ordinal Dicotómica: Nuevo - Continuador Vivo - Fallecido Sano - Enfermo						Continuas: Es el resultado de medir Se pueden expresar con números enteros o fraccionarios Entre dos valores siempre existe un número intermedio											
	Polítómicas: Tienen más de dos categorías						Discretas: Es el resultado de contar Solamente se expresan con números enteros											

Estadística descriptiva básica

Medidas de tendencia central : Medidas de tendencia central

Los datos organizados en una distribución de frecuencias destacan sus características más esenciales, como marcas de clases, centro, forma de distribución (asimétrica, simétrica), etc. Sin embargo, los indicadores que describen a los datos en forma más precisa, deben calcularse. Estos indicadores resumen los datos en medidas descriptivas que se refieren a la centralización o posición, a la dispersión o variación, a la simetría, y a la curtosis de los datos. *Las medidas de tendencia central, denominados también promedios, ubican el centro de los datos, como la media aritmética, la media geométrica, la media armónica y la mediana.*

Medidas de tendencia central : La Mediana

Es el número, que separa a la serie de datos ordenados (en forma creciente o decreciente) en dos partes de igual número de datos. *La mediana es el percentil 50 de los datos observados no agrupados o agrupados por intervalos.* Así, la mediana es la medida promedio que depende del número de orden de los datos y no de los valores de los mismos, por lo que, no afectan los valores aislados grandes o pequeños.

Medidas de tendencia central : La Mediana

Para realizar el cómputo de la mediana de n valores no agrupados de una variable cuantitativa X se siguen los siguientes pasos :

- (1) Se ordenan los datos en forma creciente.
- (2) Se ubica el valor central. Si n es impar, la mediana es el dato ordenado del centro. Pero si n es par, la mediana es la semisuma de los dos valores ordenados centrales.

Medidas de tendencia central : La Mediana

Algunas propiedades de la mediana son :

- (1) La mediana solo depende del número de datos ordenados y no del valor de los datos. Por lo tanto, no es sesgada por algún valor aislado grande o pequeño.
- (2) La mediana puede ser calculada para distribuciones de frecuencia con intervalos de diferente amplitud, siempre que se pueda determinar el límite inferior L_i del intervalo que contiene a la mediana.
- (3) La mediana puede ser calculada para variables con valores en escala ordinal.

Medidas de tendencia central : La Moda

La moda de una serie de datos es el valor M_o que se define como el dato que ocurre con mayor frecuencia. En la distribución de frecuencias por intervalos la moda se ubica en el intervalo que tiene la mayor frecuencia. La moda de una función cualquiera es el valor de la variable en el que existe un máximo absoluto (o dos o más máximos relativos iguales). La moda no siempre existe, y si existe, no siempre es única. El empleo de la moda como medida promedio puede estar justificado cuando se quiere señalar el valor más común de una serie de datos o se precise rápidamente de una medida promedio y no haya tiempo de calcular otras. Por ejemplo, los comerciantes se stockean con productos que están de moda.

Medidas de tendencia central : La Moda

Si se presenta información en datos agrupados por intervalos, el cálculo de la moda se realiza con la siguiente ecuación :

$$M_0 = L_i + \frac{d_1}{d_1 + d_2} \cdot A$$

Donde :

- L_i = Límite inferior de la clase modal
- d_1 = Diferencia entre la frecuencia de la clase modal y la clase anterior
- d_2 = Diferencia entre la frecuencia de la clase modal y la clase siguiente
- A = Amplitud del intervalo de clase

Donde : L_i es el límite inferior del intervalo modal, $d_1 = f_i - f_{i-1}$ (diferencia entre la frecuencia del intervalo modal y la frecuencia del intervalo anterior), $d_2 = f_i - f_{i+1}$ (diferencia entre la frecuencia del intervalo modal y la frecuencia del intervalo siguiente), A es la amplitud del intervalo modal

Medidas de tendencia central : La Media aritmética

Es el valor numérico que se obtiene dividiendo la suma total de los valores observados de una variable entre el número de observaciones. Para valores de una variable X observados de una muestra, la media aritmética será denotada por \bar{X} , siendo esta :

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

Donde :

- \bar{X} representa la media muestral
- n es el tamaño de la muestra
- X_i es el valor de la i -ésima observación
- \sum es el símbolo de sumatoria

Medidas de tendencia central : La Media aritmética

Algunas propiedades de la media aritmética son :

(1) La suma total de n valores cuya media es \bar{x} es igual a $n\bar{x}$:

$$\sum_{i=1}^n x_i = n\bar{x}$$

(2) Si a la variable X se le realiza una transformación lineal $Y = aX + b$, se verifica que :

$$\bar{y} = a\bar{x} + b$$

donde :

- \bar{y} es la media de la variable transformada
- \bar{x} es la media de la variable original
- a y b son constantes de la transformación

Medidas de tendencia central : La Media aritmética

Algunas propiedades de la media aritmética son :

- (1) La suma algebraica de las desviaciones de n datos x_i con respecto a su media \bar{x} es igual a cero :

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Medidas de tendencia central : La Media aritmética

- (1) La suma de los cuadrados de las desviaciones de n datos con respecto a su media es mínima. Para datos no tabulados se tiene :

$$\sum_{i=1}^n (x_i - c)^2 \text{ es mínima cuando } c = \bar{x}$$

En efecto se tiene :

$$\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - c)^2$$

$$\begin{aligned} \sum_{i=1}^n (x_i - c)^2 &= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - c)]^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + 2(\bar{x} - c) \sum_{i=1}^n (x_i - \bar{x}) + n(\bar{x} - c)^2 \end{aligned}$$

$$\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2$$

Medidas de tendencia central : La Media aritmética

- (1) Si los valores de la variable X se observaron en k grupos de tamaños respectivos n_1, n_2, \dots , entonces la media global es :

$$\bar{x} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \cdots + n_k \cdot x_k}{\sum_{i=1}^k n_i} = \frac{\sum_{i=1}^k n_i x_i}{\sum_{i=1}^k n_i}$$

Donde :

- x_1, x_2, \dots, x_k son los valores observados
- n_1, n_2, \dots, n_k son los pesos correspondientes a cada valor
- k es el número total de grupos o categorías

Medidas de tendencia central : La Media aritmética

(1) La media de los valores x_1, x_2, \dots, x_k ponderada por los pesos w_1, \dots, w_k se define por la siguiente relación :

$$\bar{x} = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_k \cdot x_k}{w_1 + w_2 + \cdots + w_k} = \frac{\sum_{i=1}^k w_i x_i}{\sum_{i=1}^k w_i}$$

Donde :

- x_i son los valores individuales ($i = 1, 2, \dots, k$)
- w_i son los pesos correspondientes a cada valor
- k es el número total de observaciones

Medidas de tendencia central : La Media aritmética

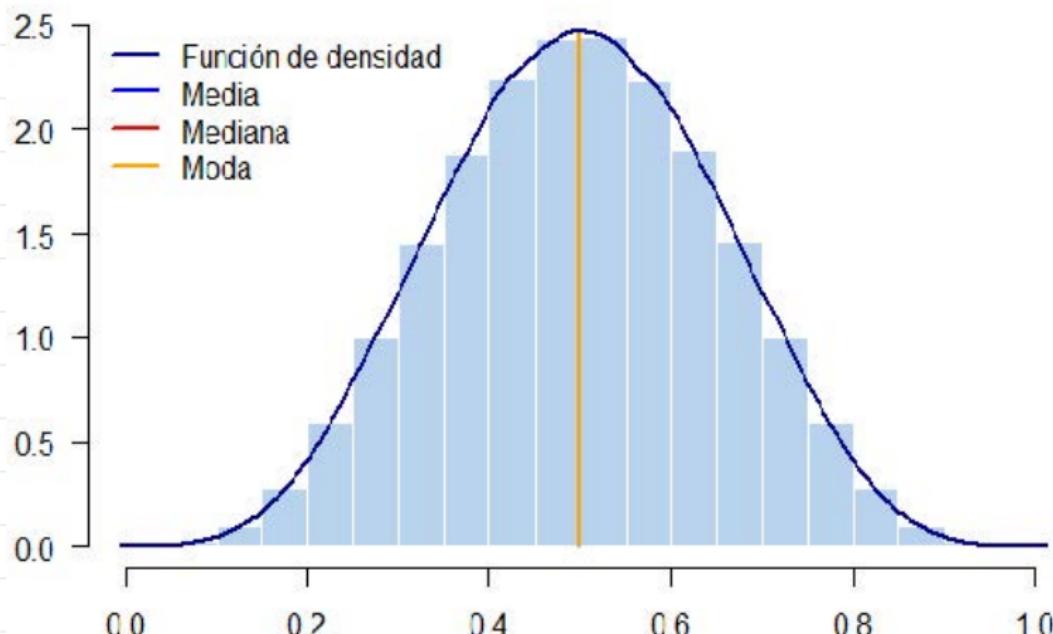
Algunas propiedades de la media aritmética son :

- (1) La media aritmética depende de todos los valores observados, en consecuencia, es afectada o sesgada por valores extremos discordantes o atípicos (extremadamente grandes o pequeños).
- (2) La media aritmética puede ser calculada también en distribución de frecuencias por intervalos de amplitud diferentes, siempre que puedan determinarse los puntos medios (marcas de clase) de los intervalos.
- (3) Existe una relación entre la media, mediana y moda.

Distribución de frecuencias

Para una distribución de frecuencias **perfectamente simétrica**, se cumple que :

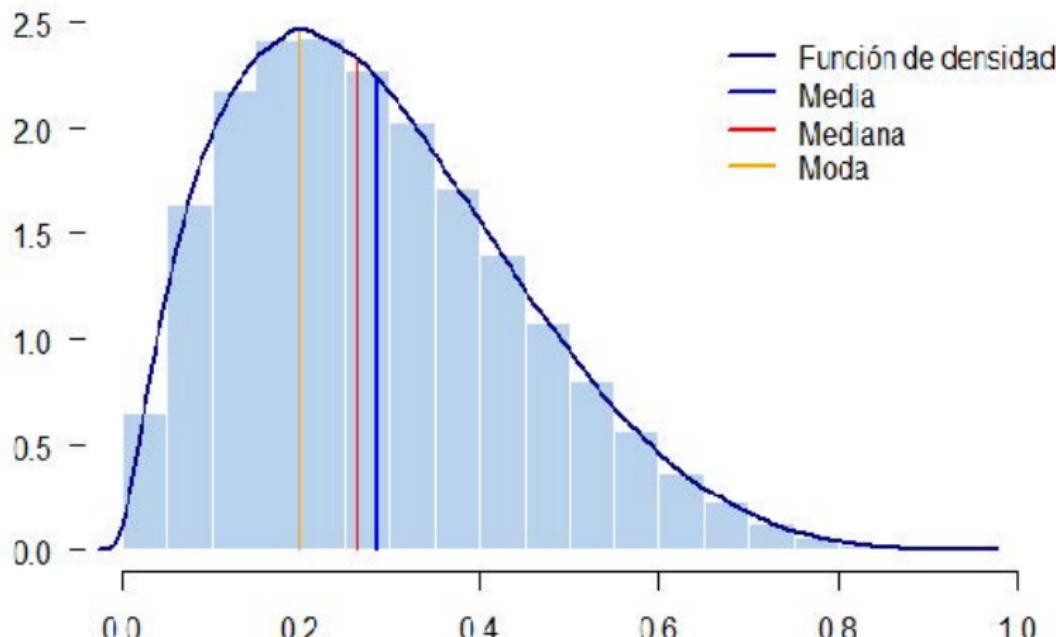
$$\bar{x} = Me = Mo$$



Distribución de frecuencias

Para una distribución de frecuencias **asimétrica positiva** (cola derecha), se cumple la siguiente relación :

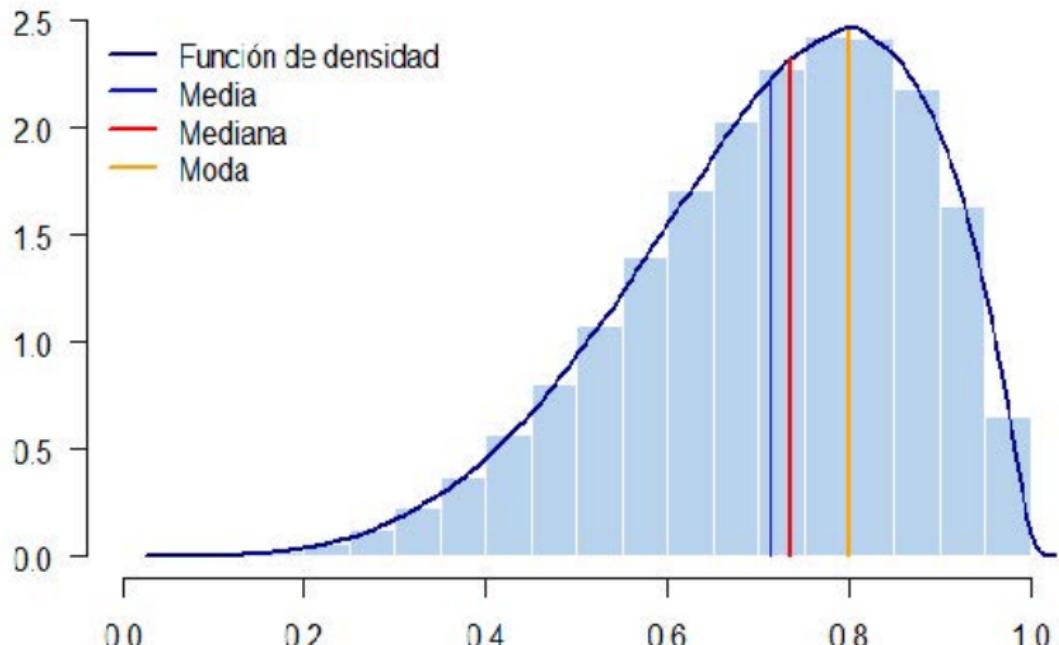
$$Mo < Me < \bar{x}$$



Distribución de frecuencias

Para una distribución de frecuencias **asimétrica negativa** (cola izquierda), se cumple la siguiente relación :

$$\bar{x} < Me < Mo$$



Medidas de Posición : Cuartil

Los cuartiles dividen la muestra tanto como sea posible en cuartos. Una muestra tiene tres de aquéllos. Existen diferentes formas de calcular cuartiles, pero todas dan aproximadamente el mismo resultado. El método más simple cuando se calcula manualmente es el siguiente : Sea n el tamaño de la muestra. Ordene los valores de la muestra del más pequeño al más grande. Para encontrar el primer cuartil, calcule el valor $0,25(n+1)$.

Medidas de Posición : Percentil

El p-ésimo percentil de una muestra, para un número p entre 0 y 100, divide a la muestra tanto como sea posible, el p% de los valores de la muestra es menor que el p-ésimo percentil y el (100p)% son mayores. Hay muchas maneras para calcular los percentiles; con todas se obtienen resultados similares. Ordene los valores de la muestra del más pequeño al más grande y después calcule la cantidad $(p/100)(n + 1)$, donde n es el tamaño de la muestra. Si esta cantidad es un entero, el valor de la muestra en esta posición es el p-ésimo percentil. Por otro lado, promedie los dos valores de la muestra en cualquier lado. Observe que el primer cuartil es el 25avo. percentil, la mediana es el 50avo. percentil y el tercer cuartil es el 75avo. percentil.

Los percentiles con frecuencia se usan para interpretar puntajes de exámenes estandarizados. Por ejemplo, si a una estudiante se le informa que su puntaje en un examen de ingreso a la universidad está en el 64avo. percentil, esto significa que 64% de los estudiantes que presentaron el examen obtuvieron puntajes inferiores.

Medidas de dispersión

Las medidas de tendencia central no son suficientes para describir un conjunto de valores de alguna variable estadística. Los promedios determinan el centro pero no indican acerca de cómo están situados los datos respecto del centro. Por ejemplo, se necesita una medida de dispersión con la finalidad de ampliar la descripción de los datos o de comparar dos o más series de datos. También se necesita una medida del grado o nivel de la asimetría o deformación en ambos lados del centro de una serie de datos, con el fin de describir la forma de la distribución de los datos. Se necesita una medida que nos permita comparar el apuntamiento o curtosis de distribuciones simétricas con respecto a la distribución simétrica normal.

Medidas de dispersión : Rango o recorrido de variable

Denotado por R es el número que resulta de la diferencia del valor máximo (X_{\max}) menos el valor mínimo (X_{\min}) de una serie de datos observados de la variable X . Es muy inestable, ya que solo depende de los valores extremos.

Medidas de dispersión : Rango intercuartílico

Denotado por RI, es el número que resulta de la diferencia del cuartil 3 menos el cuartil 1 de los datos, es decir : $RI = Q_3 - Q_1$. Es una medida que excluye el 25% superior (cuarto superior) y el 25% inferior (cuarto inferior) dando un rango dentro del cual se encuentra el 50% central de los datos observados, y a diferencia del rango, no se encuentra afectada por los valores extremos.

Medidas de dispersión : Varianza

La varianza es una medida estadística que cuantifica la dispersión de los datos respecto a su media. Indica qué tan alejados están los valores individuales del promedio.

La varianza se define como la media aritmética de los cuadrados de las diferencias de los datos con respecto a su media aritmética.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

Donde :

- s^2 es la varianza muestral
- n es el tamaño de la muestra
- x_i son los valores individuales
- \bar{x} es la media muestral ($\frac{1}{n} \sum_{i=1}^n x_i$)

Medidas de dispersión : Varianza



Medidas de dispersión : Desviación estándar

La varianza es una medida que cuantifica el nivel de dispersión o de variabilidad de los valores de una variable cuantitativa con respecto a su media aritmética. Si los datos tienden a concentrarse alrededor de su media, la varianza será pequeña. Si los valores tienden a distribuirse lejos de su media, la varianza será grande.

La desviación estándar es la raíz cuadrada de la varianza, es decir :

$$s_n = \sqrt{s_n^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2}$$

Donde :

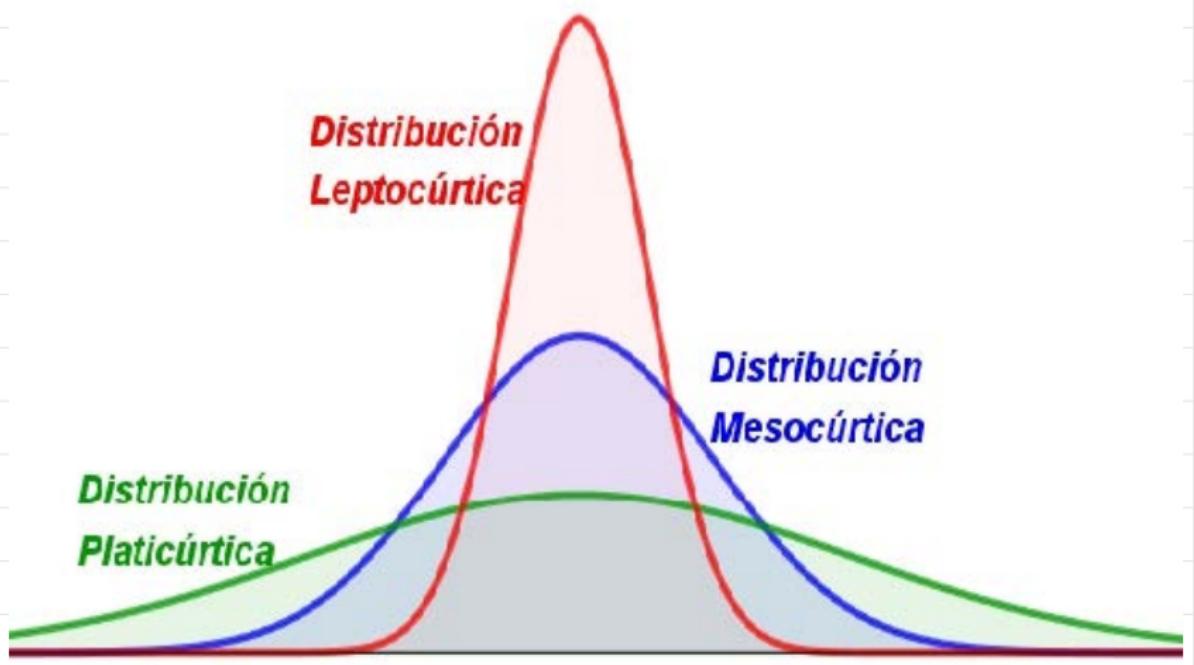
- s_n es la desviación estándar muestral para datos completos
- n es el tamaño de la muestra
- x_i son los valores individuales de la muestra
- \bar{x} es la media muestral $\left(\frac{1}{n} \sum_{i=1}^n x_i\right)$

Medidas de dispersión : Curtosis

Es la propiedad de una distribución de frecuencias por la cual se compara la dispersión de los datos observados cercanos al valor central con la dispersión de los datos cercanos a ambos extremos de la distribución.

Medidas de dispersión : Curtosis

Una curva es leptocúrtica cuando su curtosis es mayor que la normal ($K > 3$). Mientras que, cuando la curtosis es menor que la normal ($K < 3$) se denomina platicúrtica.

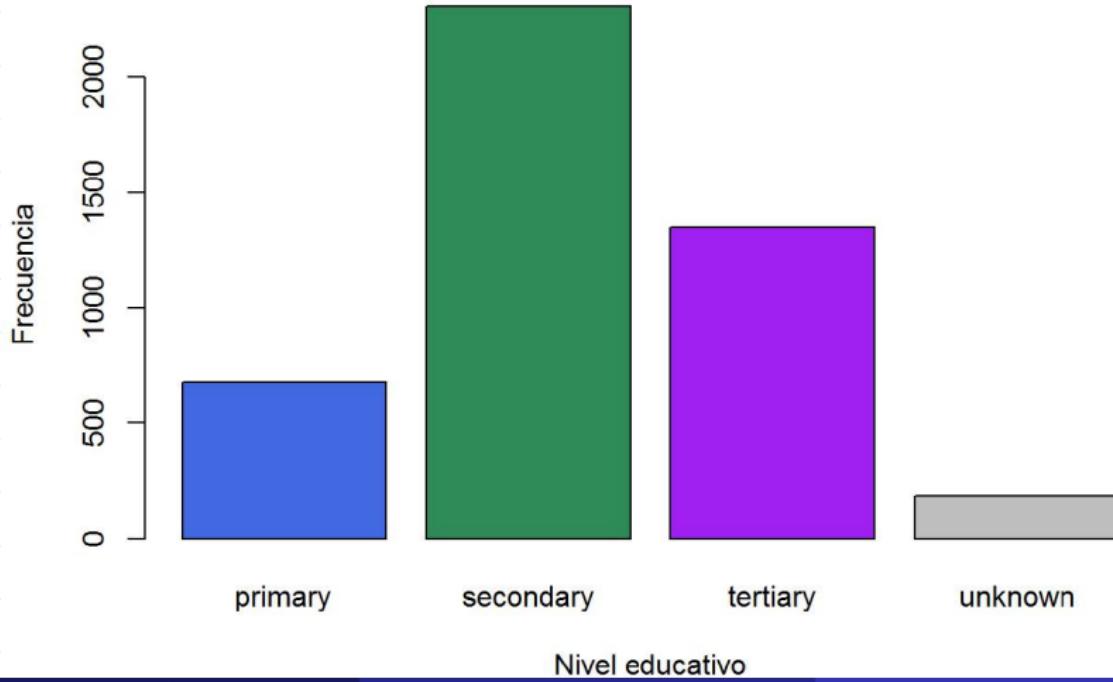


Gráficos : Barras

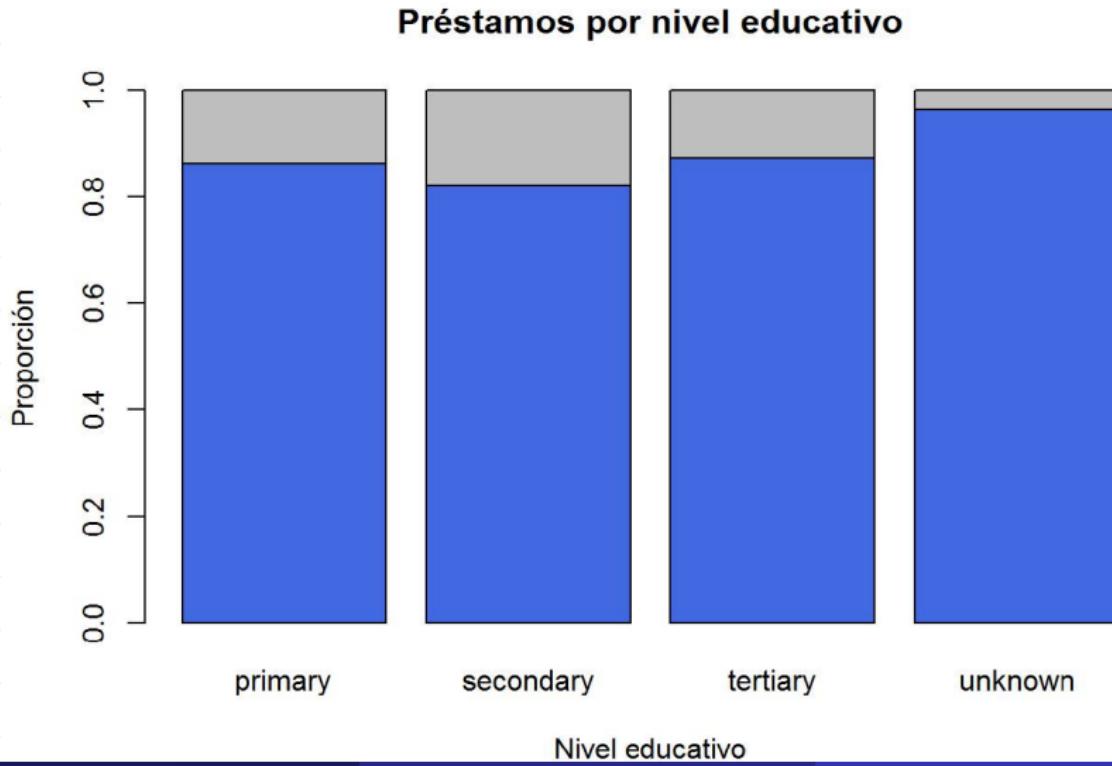
Un gráfico de barras es una forma de resumir un conjunto de datos por categorías. Muestra los datos usando varias barras de la misma anchura, cada una de las cuales representa una categoría concreta. La altura de cada barra es proporcional a una agrupación específica (por ejemplo, la suma de los valores de la categoría que representa). Las categorías podrían ser desde grupos de edad a ubicaciones geográficas. Si se aplica al crear el análisis, el gráfico de barras puede mostrar información adicional en líneas de referencia o varios tipos distintos de curvas. Estas líneas o curvas podrían, por ejemplo, mostrar si los puntos de los datos se adaptan bien a un ajuste de curva polinómica determinado, o resumir un conjunto de puntos de datos de muestra ajustándolos a un modelo que describirá los datos y mostrará una curva o una línea recta sobre la visualización. La curva normalmente cambia su aspecto en función de los valores que se hayan filtrado del análisis.

Gráficos : Barras

Gráfica de Educación

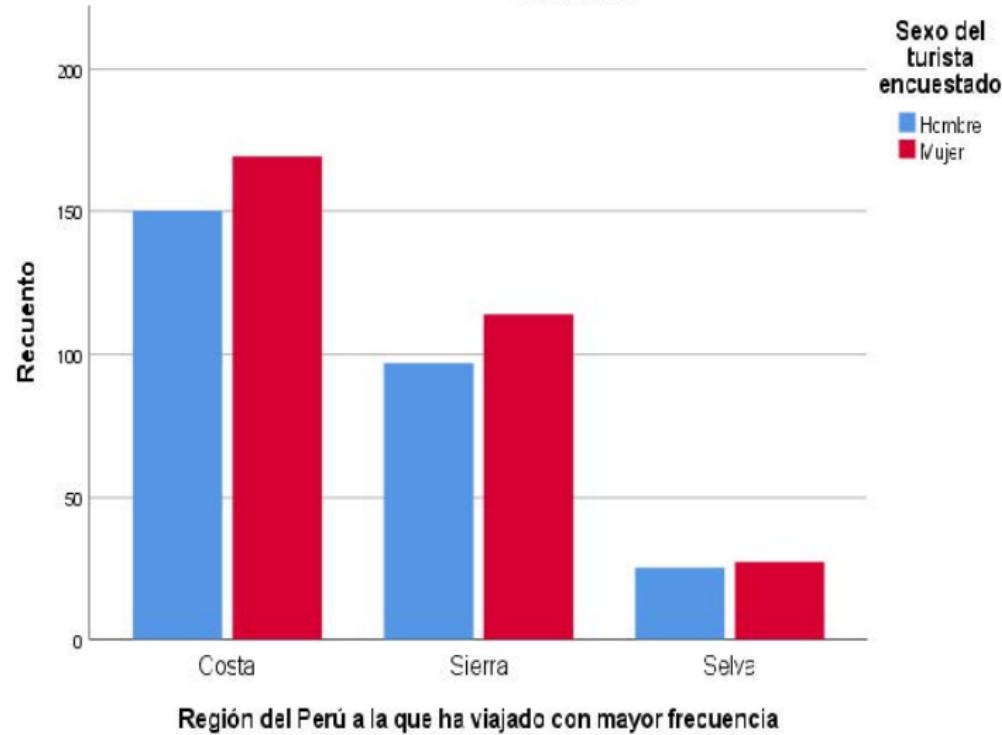


Gráficos : Barras



Gráficos : Barras

Barras agrupadas Recuento de Región del Perú a la que ha viajado con mayor frecuencia por Sexo del turista encuestado

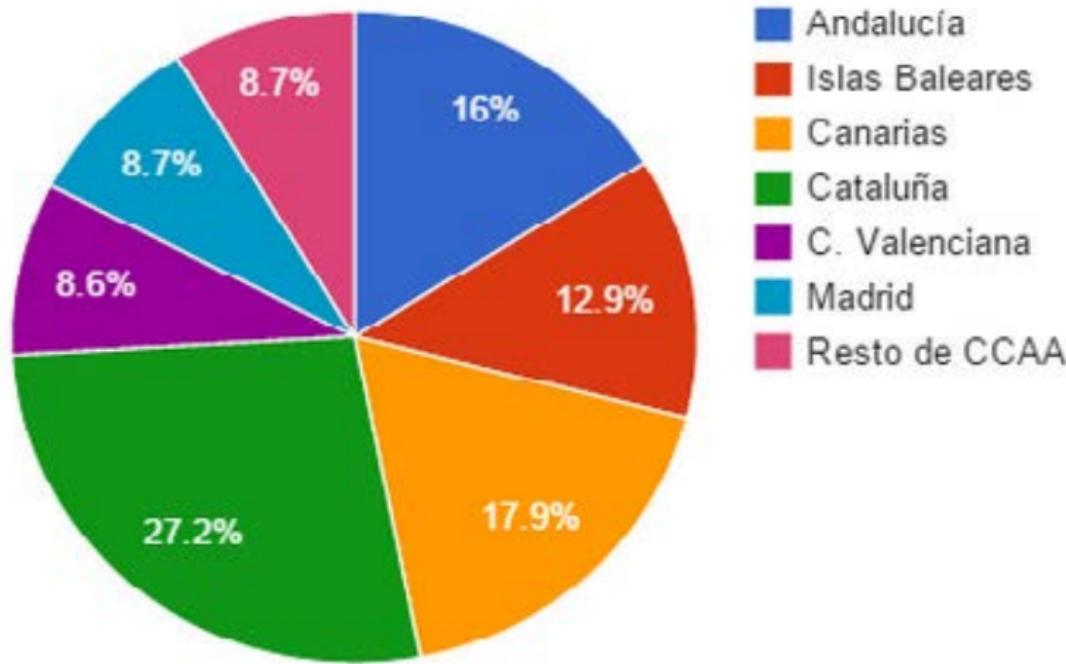


Gráficos : Circular o Pastel

Un gráfico circular o gráfica circular, también llamado gráfico de pastel, gráfica de pizza, gráfico de tarta, gráfico de torta o gráfica de 360 grados, es un recurso estadístico que se utiliza para representar porcentajes y proporciones. El número de elementos comparados dentro de una gráfica circular suele ser de más de cuatro. Se utilizan en aquellos casos donde interesa no solamente mostrar el número de veces que se dan una característica o atributo de manera tabular sino más bien de manera gráfica, de tal manera que se pueda visualizar mejor la proporción en que aparece esa característica respecto del total. A pesar de su popularidad, se trata de un tipo de gráfico poco recomendable debido a que nuestra capacidad perceptual para estimar relaciones de proporción o diferencias entre áreas de sectores circulares es mucho menor que, por ejemplo, entre longitudes o posiciones, tal y como sucede en otras gráficas.

Gráficos : Circular o Pastels

Recepcion de turistas

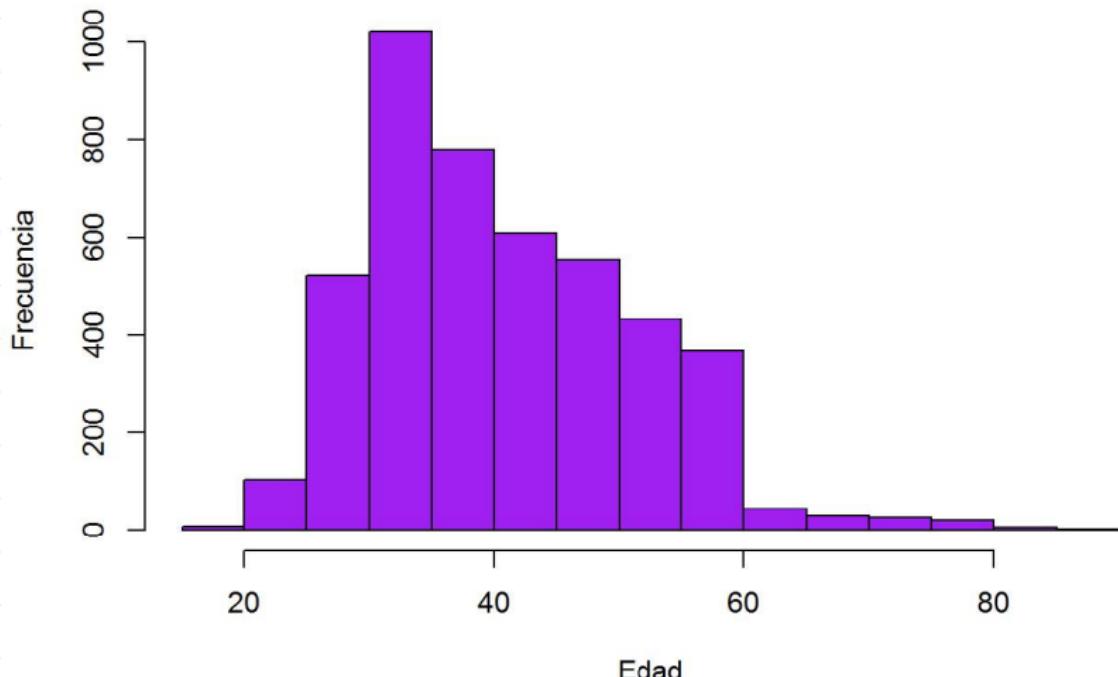


Gráficos : Histogramas

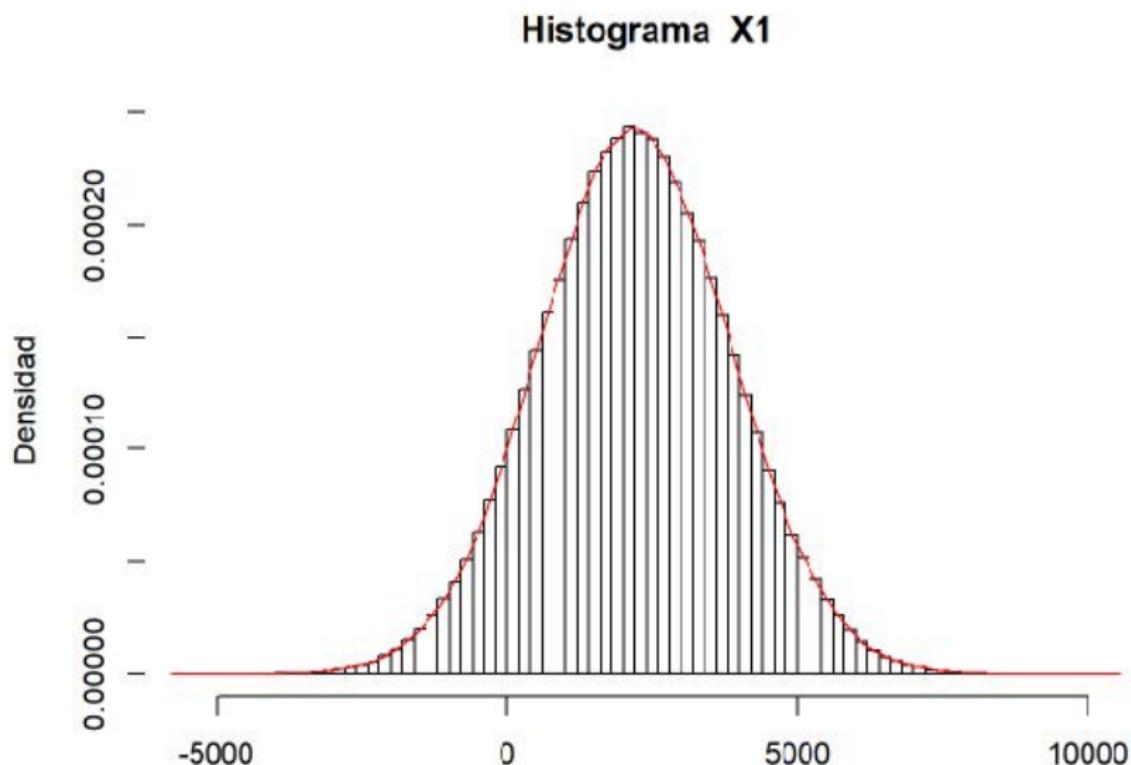
Es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. Sirven para obtener una "primera vista" general, o panorama, de la distribución de la población, o de la muestra, respecto a una característica, cuantitativa y continua (como la longitud o el peso). De esta manera ofrece una visión de grupo permitiendo observar una preferencia, o tendencia, por parte de la muestra o población por ubicarse hacia una determinada región de valores dentro del espectro de valores posibles (sean infinitos o no) que pueda adquirir la característica. Así pues, podemos evidenciar comportamientos, observar el grado de homogeneidad, acuerdo o concisión entre los valores de todas las partes que componen la población o la muestra, o, en contraposición, poder observar el grado de variabilidad, y por ende, la dispersión de todos los valores que toman las partes, también es posible no evidenciar ninguna tendencia y obtener que cada miembro de la población toma por su lado y adquiere un valor de la característica aleatoriamente sin mostrar ninguna preferencia o tendencia.

Gráficos : Histogramas

Histograma de Edad



Gráficos : Histogramas

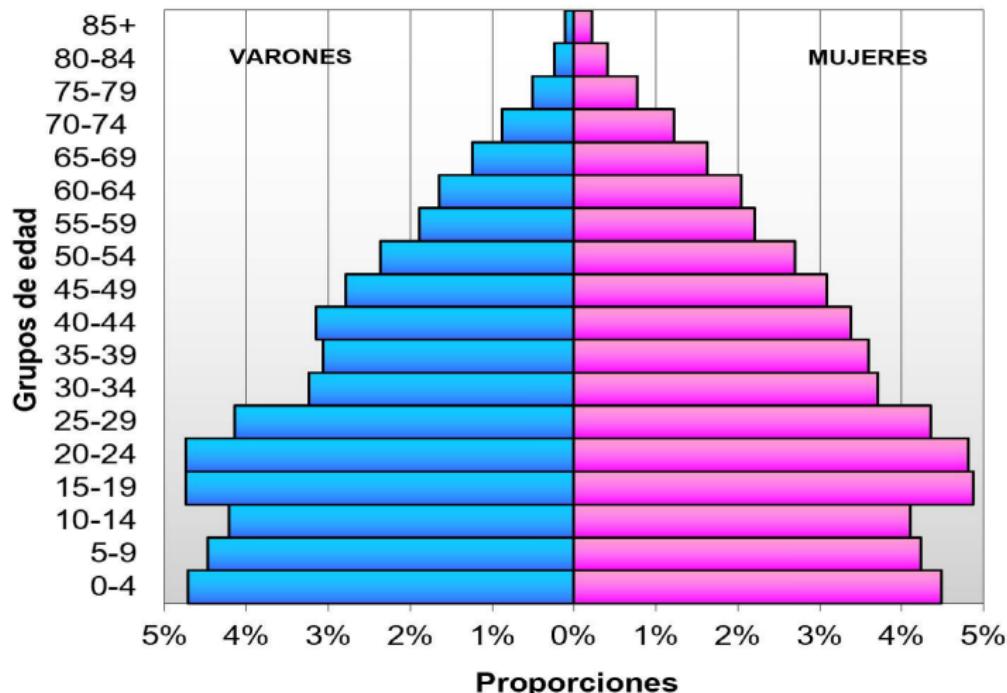


Gráficos : Pirámide poblacional

La pirámide de población es una representación gráfica de las características de una población perteneciente a una localidad, ciudad o país, en un momento en el tiempo. Se muestra con barras en posición horizontal, la longitud de cada barra tiene una relación directamente y proporcional al número de individuos de la población. Permite visualizar de manera sencilla la proporción de edad, sexo, envejecimiento, entre otros. Es decir, gracias a su construcción que es posible observar el crecimiento, la estructura la distribución, así como la movilidad de una población y simplificar su correlación con la dinámica social, económica así como ambiental que prevalece. Así mismo, permite mostrar una tendencia; facilitando así su interpretación y la toma de decisiones con respecto a la planeación territorial, así como políticas de salud pública. Es posible que se realicen comparativos en distintas ciudades o países para conocer la evolución demográfica en cada uno, con respecto al tiempo en que se lleva a cabo la construcción de la pirámide.

Gráficos : Pirámide poblacional

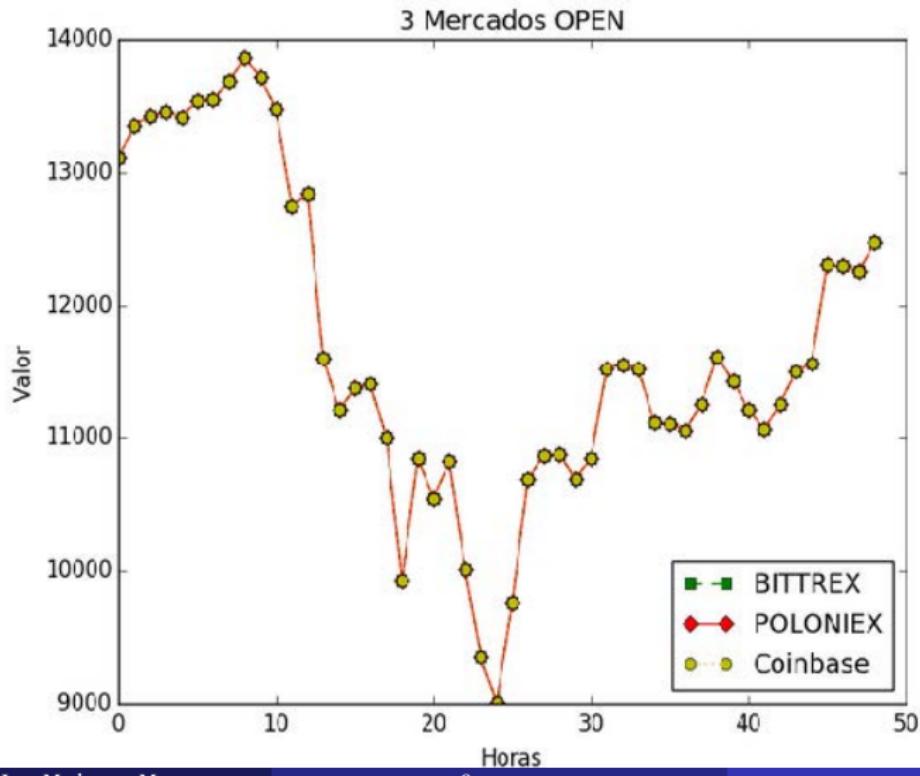
Pirámide de población de España, año 1950



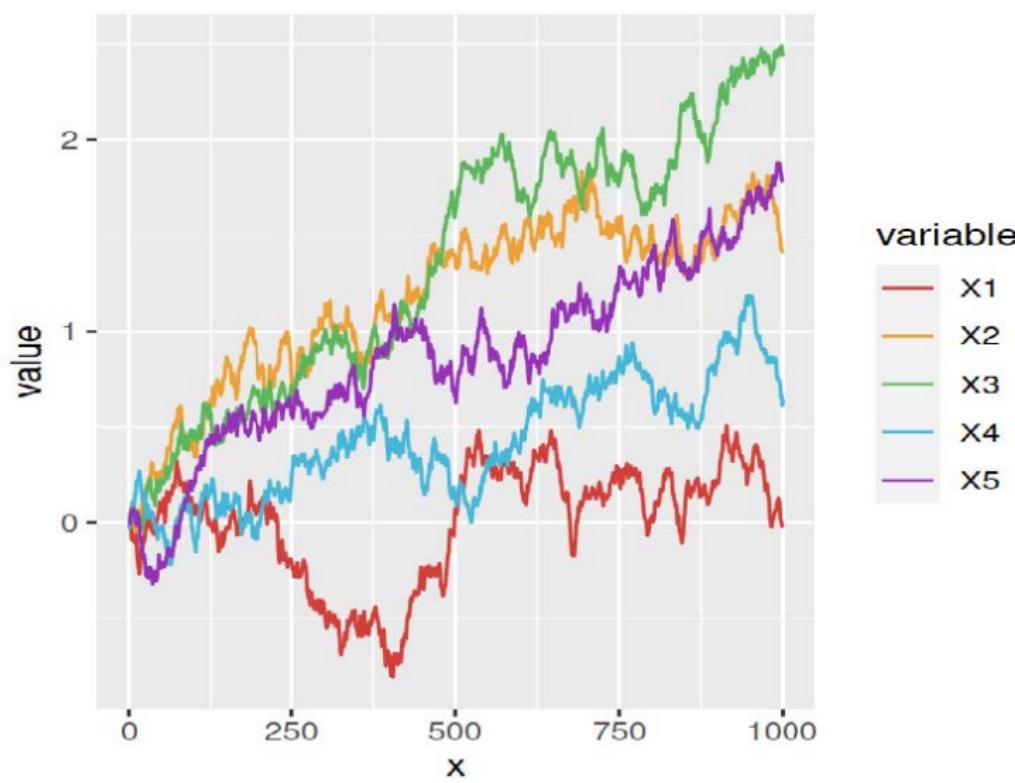
Gráficos : De líneas

Los gráficos de líneas se utilizan para mostrar el valor cuantitativo en un intervalo o intervalo de tiempo continuo. Se usa con mayor frecuencia para mostrar tendencias y relaciones (cuando se agrupan con otras líneas). Los gráficos de línea también ayudan a dar un «panorama general» en un intervalo, para ver cómo se ha desarrollado durante ese período. Los gráficos de líneas se representan dibujando primero los puntos de datos en una cuadrícula cartesiana, y luego conectando una línea entre estos puntos. Típicamente, el eje Y tiene un valor cuantitativo, mientras que el eje X tiene una escala de categoría o secuenciada. Los valores negativos se pueden mostrar debajo del eje X. Los gráficos de líneas se componen de un rango continuo de fechas o números en el eje x, y de un valor numérico asociado en el eje y. El eje x de un gráfico de líneas muestra una variable continua, como el tiempo o la distancia, y dibuja una línea que visualiza el cambio de valores entre cada intervalo de tiempo o distancia consecutivos. Cada intervalo está marcado con un punto correspondiente a un valor numérico medido por el eje y.

Gráficos : De líneas



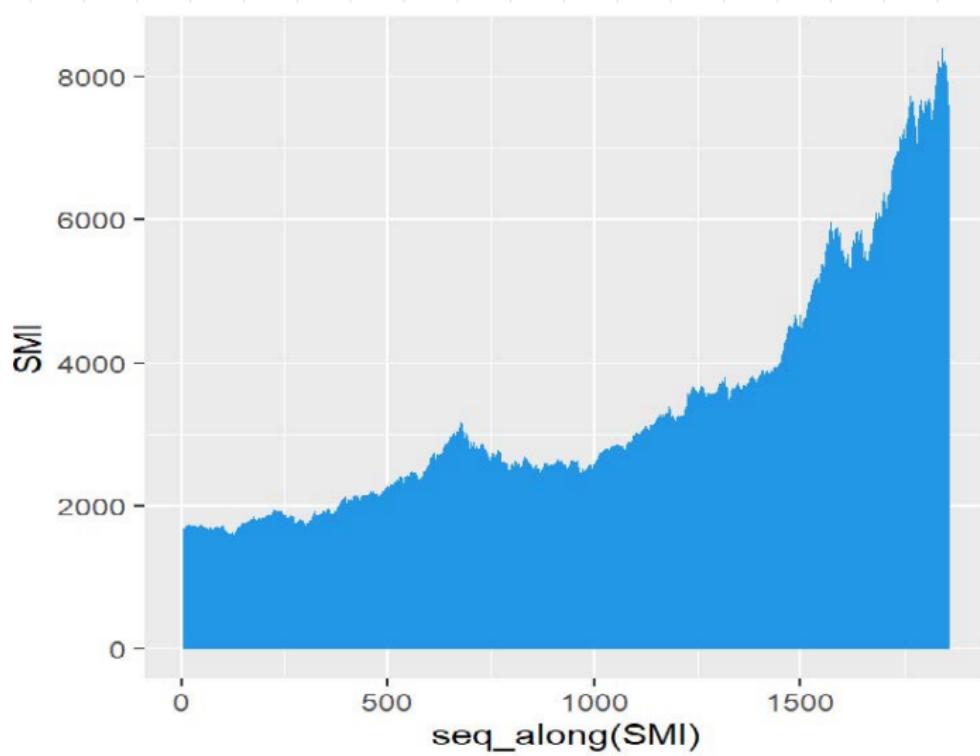
Gráficos : De líneas



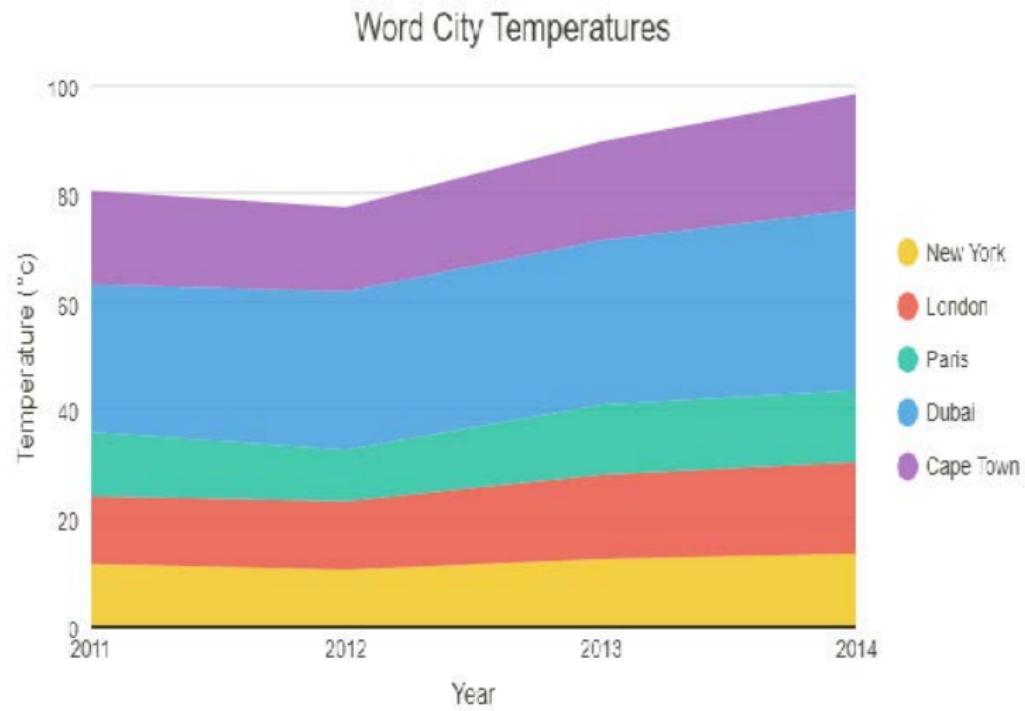
Gráficos : De área

Un gráfico de áreas representa el cambio en una o más cantidades a lo largo del tiempo. Es similar a un gráfico de líneas. Tanto en los gráficos de área como en los gráficos de líneas, los puntos de datos se trazan y luego se conectan mediante segmentos de línea para mostrar el valor de una cantidad en varios momentos diferentes. Sin embargo, los gráficos de áreas son diferentes de los gráficos de líneas porque el área entre el eje XY y la línea se rellena con color o sombreado. Un gráfico de áreas apiladas muestra cuánto contribuye cada parte a la cantidad total. Por ejemplo, el propietario de una cadena de tiendas de comestibles puede querer hacer un gráfico que muestre las ganancias obtenidas por cada una de sus tiendas y las ganancias totales obtenidas por todas las tiendas juntas. Un gráfico de áreas apiladas sería perfecto para representar este tipo de datos. Aunque los gráficos de áreas se utilizan con mayor frecuencia para mostrar tendencias generales en los datos a lo largo del tiempo, también puede hacer un gráfico de áreas apiladas que muestre explícitamente la contribución exacta de cada cantidad al total.

Gráficos : De área



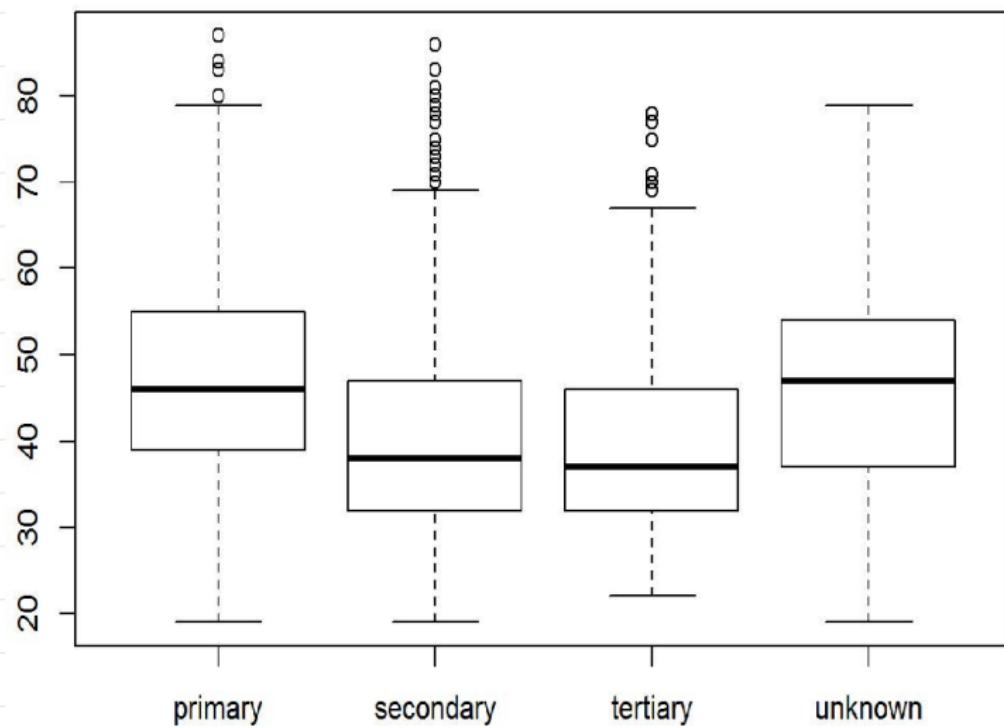
Gráficos : De área



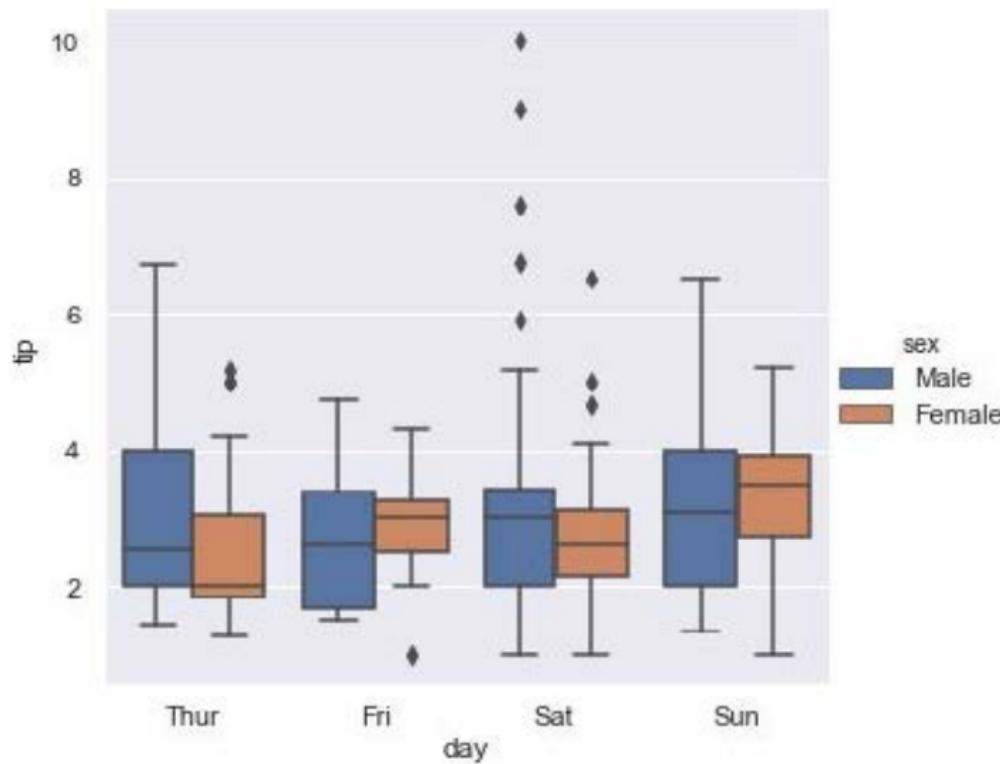
Gráficos : Cajas y bigotes

Los diagramas de Caja-Bigotes (boxplots o box and whiskers) son una presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría. Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente. Una gráfica de este tipo consiste en una caja rectangular, donde los lados más largos muestran el recorrido intercuartílico. Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero(recordemos que el segundo cuartil coincide con la mediana). Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente

Gráficos : Cajas y bigotes



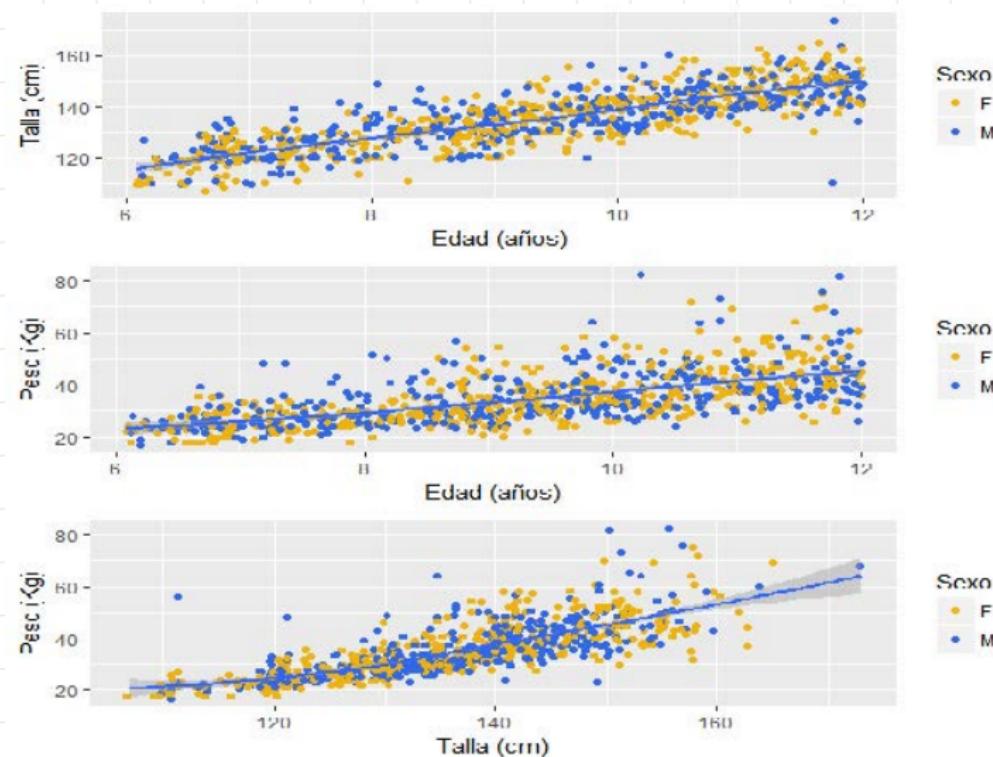
Gráficos : Cajas y bigotes



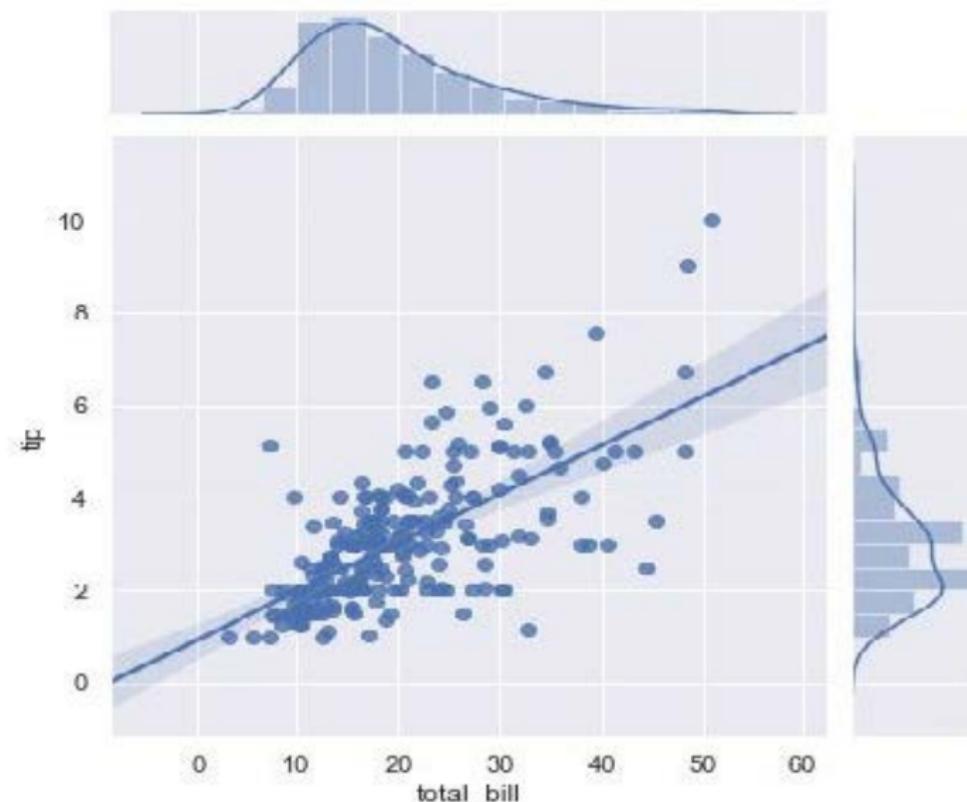
Gráficos : De dispersión

Un diagrama de dispersión o gráfica de dispersión o gráfico de burbujas es un tipo de diagrama matemático que utiliza las coordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos. Se emplea cuando una o varias variables está bajo el control del experimentador. Si existe un parámetro que se incrementa o disminuye de forma sistemática por el experimentador, se le denomina parámetro de control o variable independiente y habitualmente se representa a lo largo del eje horizontal (eje de las abscisas). La variable medida o dependiente usualmente se representa a lo largo del eje vertical (eje de las ordenadas). Si no existe una variable dependiente, cualquier variable se puede representar en cada eje y el diagrama de dispersión mostrará el grado de correlación (no causalidad) entre las dos variables. Un diagrama de dispersión puede sugerir varios tipos de correlaciones entre las variables con un intervalo de confianza determinado. La correlación puede ser positiva (aumento), negativa (descenso), o nula (las variables no están correlacionadas). Se puede dibujar una línea de ajuste (llamada también "línea de tendencia") con el fin de estudiar la correlación.

Gráficos : De dispersión



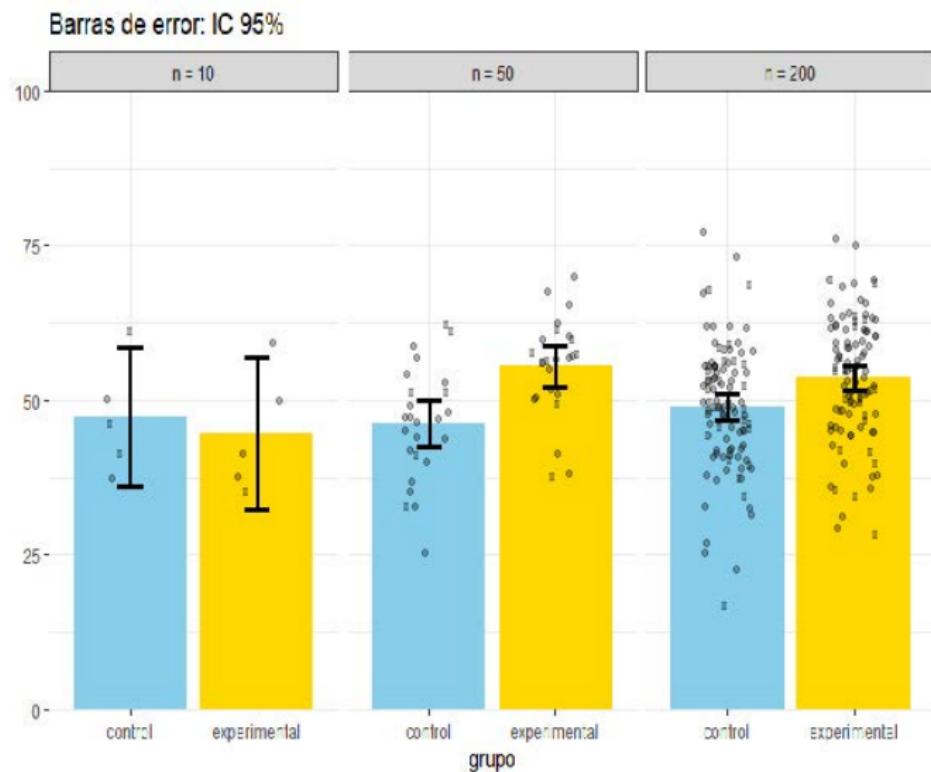
Gráficos : De dispersión



Gráficos : Barra de error

Las barras de error son representaciones gráficas de la variabilidad de los datos, y se usan en gráficos para indicar el error o la incertidumbre en una determinada medida. Dan una idea general de lo precisa que es una medición o, a la inversa, a qué distancia del valor indicado puede estar el valor verdadero (sin errores) del elemento medido. Representan una incertidumbre utilizando una desviación típica, un error estándar o un intervalo de confianza particular (por ejemplo, un intervalo del 95 %). Estas cantidades no expresan necesariamente valores coincidentes, por lo que debe indicarse explícitamente en el gráfico o en el texto de apoyo cuál es el indicador del error utilizado. Se pueden usar para comparar visualmente dos cantidades, e implicitamente, si se cumplen determinadas condiciones, permiten determinar a simple vista si las diferencias son de significación estadística. Las barras de error también pueden sugerir la bondad de ajuste de una función dada, es decir, la exactitud con la que la función describe los datos. Es habitual que los artículos científicos en el campo de las ciencias experimentales incluyan barras de error en todos los gráficos.

Gráficos : Barra de error



Definición de Población

En forma general, en Estadística se denomina población a un conjunto de elementos que consiste de personas, análisis, objetos, etc., en los que se puede observar o medir una o más características de naturaleza cualitativa o cuantitativa. A cada elemento se le denomina unidad elemental o unidad estadística.

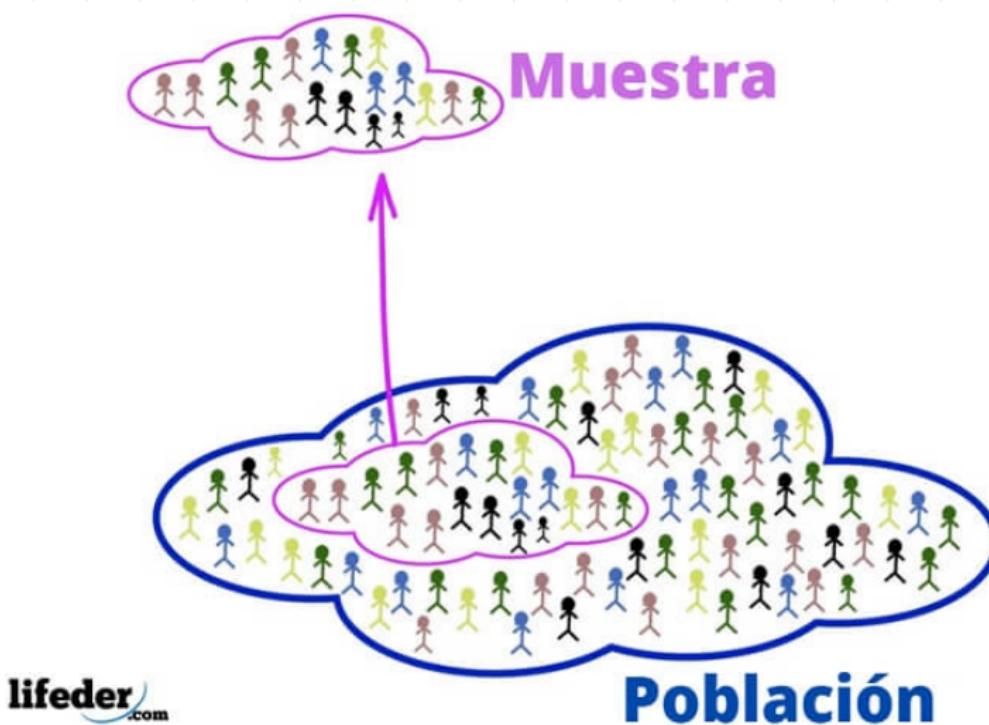
Por ejemplo, los empleados de una empresa en un día laborable, constituyen una población en la que cada empleado (unidad estadístico) tiene muchas características a ser observadas como género, estado civil, lugar de procedencia, grado de instrucción, etc.

Definición de Muestra

Parte de la población seleccionada de acuerdo con un plan o regla para obtener información acerca de la población de la cual proviene. La muestra debe ser seleccionada de manera que sea representativa de la población.

Un ejemplo de muestra se refiere a una selección más pequeña de un grupo más grande, utilizada para representar las características de ese grupo más grande.

Población y muestra



Tipos de muestreo

Tipos de muestreo



Muestreo
probabilístico



Muestreo
aleatorio simple



Muestreo por
conglomerados



Muestreo
sistématico



Muestreo
estratificado



Muestreo no
probabilístico



Muestreo de
bola de nieve



Muestreo por
cuotas



Muestreo intencional
o por conveniencia