

Probabilidad y Estadística

Universidad Tecnológica Izúcar de Matamoros
UTIM

Dr. Alejandro Rodriguez

21 de febrero de 2022

- ① Estadística Descriptiva
- ② Probabilidad
- ③ Estadística Inferencial

Definiciones importantes

Estadística

Es un conjunto de procedimientos que sirven para organizar y resumir datos, hacer inferencias a partir de ellos y transmitir los resultados de manera clara, concisa y significativa.

¹Cada definición será enriquecida más adelante.

Definiciones importantes

Estadística

Es un conjunto de procedimientos que sirven para organizar y resumir datos, hacer inferencias a partir de ellos y transmitir los resultados de manera clara, concisa y significativa.

Estadística Descriptiva

Es un conjunto de procedimientos que sirven para organizar, describir y sintetizar datos, sin que las conclusiones que se extraigan de éstos rebasen su ámbito específico.

¹Cada definición será enriquecida más adelante.

Definiciones importantes

Probabilidad

El término probabilidad se refiere al estudio de azar y la incertidumbre en cualquier situación en la cual varios posibles sucesos pueden ocurrir; la disciplina de la probabilidad proporciona métodos de cuantificar las oportunidades y probabilidades asociadas con varios sucesos planeado o en una investigación científica.

¹Cada definición será enriquecida más adelante.

Definiciones importantes

Probabilidad

El término probabilidad se refiere al estudio de azar y la incertidumbre en cualquier situación en la cual varios posibles sucesos pueden ocurrir; la disciplina de la probabilidad proporciona métodos de cuantificar las oportunidades y probabilidades asociadas con varios sucesos planeado o en una investigación científica.

Estadística Inferencial

Es un conjunto de procedimientos que se emplean para hacer inferencias y generalizaciones respecto a una totalidad, partiendo del estudio de un número limitado de casos tomados de esta última.

¹Cada definición será enriquecida más adelante.

Estadística Descriptiva

Estadística Descriptiva

Estadística Descriptiva

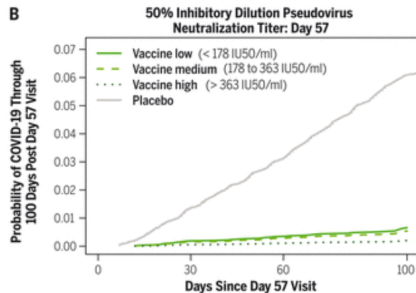
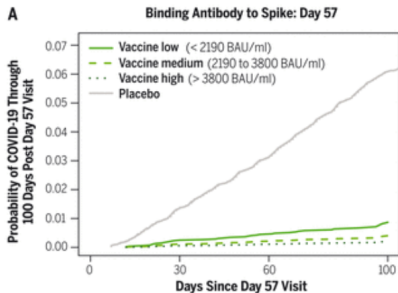
Es un conjunto de procedimientos que sirven para organizar, describir y sintetizar datos, sin que las conclusiones que se extraigan de éstos rebasen su ámbito específico.

Ejemplo A

Table 1. Demographic and Clinical Characteristics of Children in the Phase 2–3 Trial.*

Characteristic	BNT162b2 (N = 1518)†	Placebo (N = 750)†	Total (N = 2268)
Male sex — no. (%)	799 (52.6)	383 (51.1)	1182 (52.1)
Race — no. (%)‡			
White	1204 (79.3)	586 (78.1)	1790 (78.9)
Black	89 (5.9)	58 (7.7)	147 (6.5)
Asian	90 (5.9)	47 (6.3)	137 (6.0)
Multiracial	109 (7.2)	49 (6.5)	158 (7.0)
Other or not reported	26 (1.7)	10 (1.3)	36 (1.6)
Hispanic or Latinx ethnicity — no. (%)‡	319 (21.0)	159 (21.2)	478 (21.1)
Country — no. (%)			
United States	1073 (70.7)	531 (70.8)	1604 (70.7)
Finland	158 (10.4)	81 (10.8)	239 (10.5)
Spain	162 (10.7)	78 (10.4)	240 (10.6)
Poland	125 (8.2)	60 (8.0)	185 (8.2)
Age at vaccination — yr			
Mean ±SD	8.2±1.93	8.1±1.97	8.2±1.94
Median (range)	8.0 (5–11)	8.0 (5–11)	8.0 (5–11)
Obese — no. (%)§	174 (11.5)	92 (12.3)	266 (11.7)
Coexisting conditions¶	312 (20.6)	152 (20.3)	464 (20.5)
Baseline SARS-CoV-2 infection status — no. (%)			
Positive	133 (8.8)	65 (8.7)	198 (8.7)
Negative	1385 (91.2)	685 (91.3)	2070 (91.3)

Ejemplo B

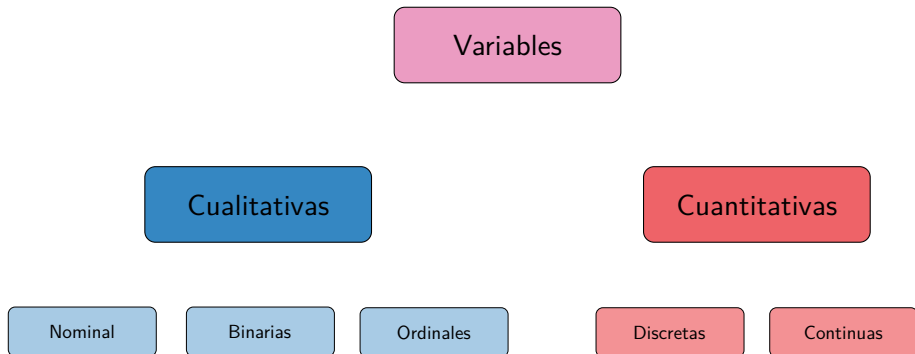


Variable estadística

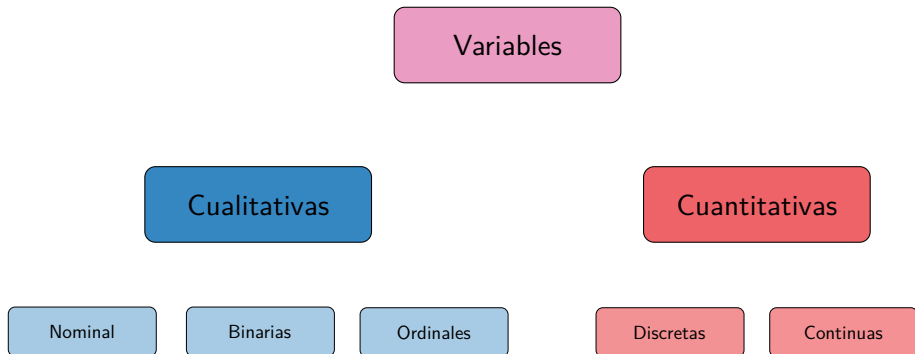
Variable estadística

Variable es cualquier característica cuyo valor puede cambiar de un objeto a otro en la población objeto de estudio. Estos valores a su vez se caracterizan por poder ser medidos.

Variables



Variables



Un estudio de una sola variable se conoce como datos *univariante*. De igual forma si se hace un estudio donde varían dos variables se le llama *bivariante*.

Ejemplo de variables

Id	Nombre	Peso	Escolaridad	Alergia	Localidad
1	Richard Feynman	80.6Kg	PhD	No	NY City
2	Yuri Gagarin	95.3Kg	Coronel	SI	Smolensk
3	Manuel Mercado	70.0Kg	Licenciado	NO	C. de México
C.D	C.N	C.D	C.O	C.B	C.N

Población, Muestra y Elemento

Población, Muestra, Elemento

Población

La **población** o también llamada **universo**, es todo conjunto de personas, cosas u objetos con ciertas características comunes sujetos a un estudio en cuestión.

Población, Muestra, Elemento

Población

La **población** o también llamada **universo**, es todo conjunto de personas, cosas u objetos con ciertas características comunes sujetos a un estudio en cuestión.

Muestra

Definida una población cualquiera, se llama **muestra** a toda porción de elementos sacada de ella.

Población, Muestra, Elemento

Población

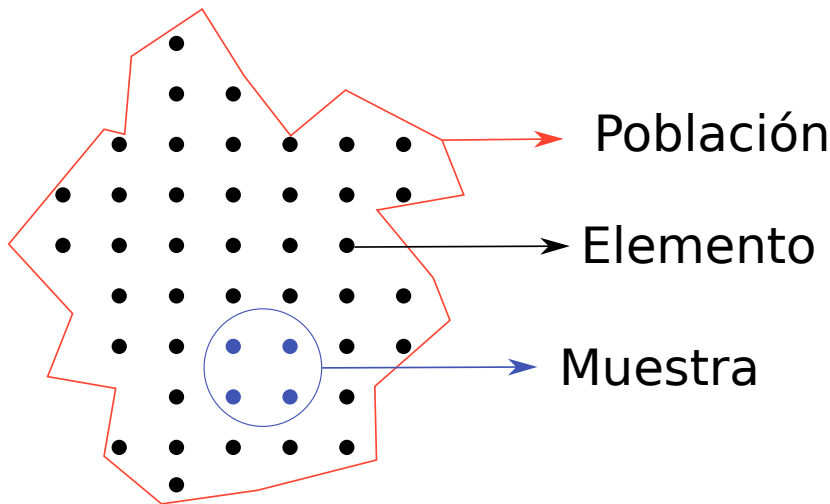
La **población** o también llamada **universo**, es todo conjunto de personas, cosas u objetos con ciertas características comunes sujetos a un estudio en cuestión.

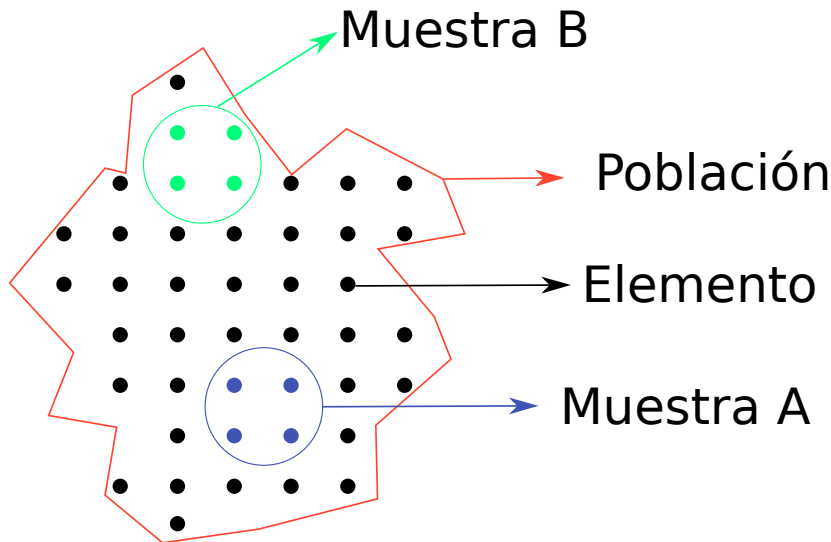
Muestra

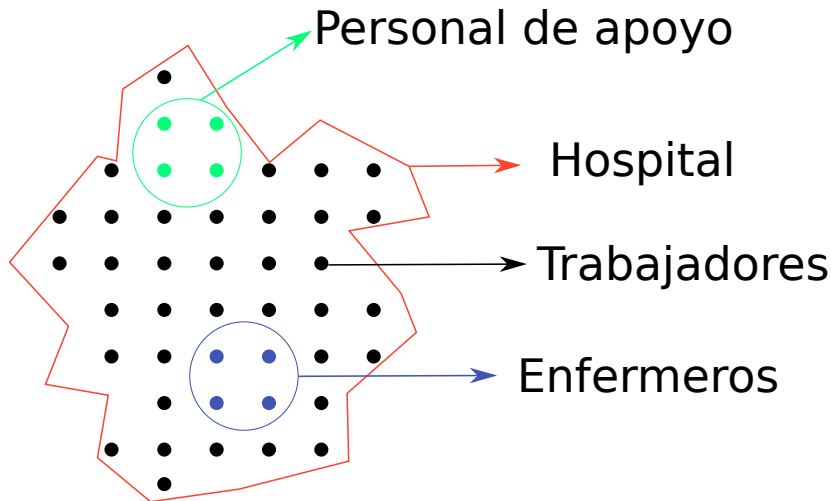
Definida una población cualquiera, se llama **muestra** a toda porción de elementos sacada de ella.

Elemento o Unidad

Cada uno de los componentes de una población recibe el nombre de **elemento** o **unidad** esencial.







Tarea

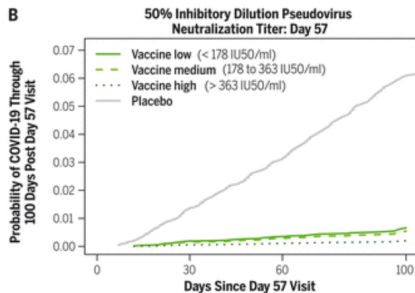
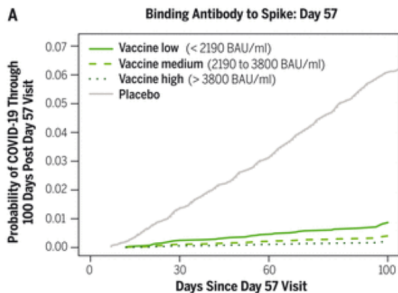
Tarea 1

Table 1. Demographic and Clinical Characteristics of Children in the Phase 2–3 Trial.^a

Characteristic	BNT162b2 (N=1518) [†]	Placebo (N=750) [†]	Total (N=2268)
Male sex — no. (%)	799 (52.6)	383 (51.1)	1182 (52.1)
Race — no. (%) [‡]			
White	1204 (79.3)	586 (78.1)	1790 (78.9)
Black	89 (5.9)	58 (7.7)	147 (6.5)
Asian	90 (5.9)	47 (6.3)	137 (6.0)
Multiracial	109 (7.2)	49 (6.5)	158 (7.0)
Other or not reported	26 (1.7)	10 (1.3)	36 (1.6)
Hispanic or Latinx ethnicity — no. (%) [‡]	319 (21.0)	159 (21.2)	478 (21.1)
Country — no. (%)			
United States	1073 (70.7)	531 (70.8)	1604 (70.7)
Finland	158 (10.4)	81 (10.8)	239 (10.5)
Spain	162 (10.7)	78 (10.4)	240 (10.6)
Poland	125 (8.2)	60 (8.0)	185 (8.2)
Age at vaccination — yr			
Mean \pm SD	8.2 \pm 1.93	8.1 \pm 1.97	8.2 \pm 1.94
Median (range)	8.0 (5–11)	8.0 (5–11)	8.0 (5–11)
Obese — no. (%) [§]	174 (11.5)	92 (12.3)	266 (11.7)
Coexisting conditions [¶]	312 (20.6)	152 (20.3)	464 (20.5)
Baseline SARS-CoV-2 infection status — no. (%)			
Positive	133 (8.8)	65 (8.7)	198 (8.7)
Negative	1385 (91.2)	685 (91.3)	2070 (91.3)

Tarea

Tarea 2



Identifique: Población, Muestra y elemento

Nombre	email
Alexis Cocoyutla Gonzáles	alexcoco223@gmail.com
Gabriela Onofre Rodríguez	gabrielaonofre201@gmail.com
Zuriel Elias Angular Campos	zurielnx@gmail.com
Alejandro Soriano Maya	alejandrosorianomaya@gmail.com
Mary Carmen Espejel Sanchez	marycarmenespejel29@gmail.com
Evelia Arai Benitez Carre'on	eveliaabc66@gmail.com
Daniela Aroche Orta	danielaarochorta@gmail.com
Ximena Pedraza Hern'andez	time.dehernandez2001@gmail.com
Jorge Martinez Lucaro	jorge2313fiva@gmail.com

Censo

Censo

Cuando la información deseada está disponible para todo los objetos de la población se tiene lo que se llama un ***censo***.

Censo

Censo

Cuando la información deseada está disponible para todo los objetos de la población se tiene lo que se llama un ***censo***.

En otra palabras: Cuando podemos saber lo deseado porque tenemos la posibilidad de cuestionar o averiguar con cada individuo lo deseado, estamos haciendo un censo.

Las restricciones de tiempo, dinero y recursos escasos casi siempre hacen que un censo sea impractico o infactible.

¿Qué es el muestreo?

¿Qué es el muestreo?

El muestreo tiene por objetivo seleccionar una muestra de modo que la variable a estudiar guarde la relación existente entre la muestra y la población.

¿Qué es el muestreo?

El muestreo tiene por objetivo seleccionar una muestra de modo que la variable a estudiar guarde la relación existente entre la muestra y la población.

Muestreo

El muestreo o la acción de muestrear, es la toma de una porción proveniente de una población que nos permita describir cierta característica de la misma.

¿Qué es el muestreo?

El muestreo tiene por objetivo seleccionar una muestra de modo que la variable a estudiar guarde la relación existente entre la muestra y la población.

Muestreo

El muestreo o la acción de muestrear, es la toma de una porción proveniente de una población que nos permita describir cierta característica de la misma.

Independientemente del procedimiento de selección de muestra, el conocimiento adquirido es una *estimación*.

Tipos de muestreo

MAPA MENTAL

Muestreo Probabilístico

Método de muestreo probabilístico

Los métodos de muestreo probabilísticos son aquellos que se basan en el principio de equiprobabilidad. Es decir, aquellos en los que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra.

Muestreo Probabilístico. Método aleatorio simple

Método aleatorio simple

Procedimiento: Se asigna un número a cada individuo de la población y a través de algún medio se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

Muestreo Probabilístico. Método aleatorio simple

Método aleatorio simple

Procedimiento: Se asigna un número a cada individuo de la población y a través de algún medio se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

Observaciones:

- Garantiza que todos los individuos que componen la población blanco tengan la misma *oportunidad*.

Muestreo Probabilístico. Método aleatorio simple

Método aleatorio simple

Procedimiento: Se asigna un número a cada individuo de la población y a través de algún medio se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

Observaciones:

- Garantiza que todos los individuos que componen la población blanco tengan la misma *oportunidad*.
- Procedimiento atractivo y *simple*.

Muestreo Probabilístico. Método aleatorio simple

Método aleatorio simple

Procedimiento: Se asigna un número a cada individuo de la población y a través de algún medio se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

Observaciones:

- Garantiza que todos los individuos que componen la población blanco tengan la misma *oportunidad*.
- Procedimiento atractivo y *simple*.
- Poca o *nula utilidad* práctica cuando la población es *muy grade*.

Muestreo Probabilístico. Muestreo aleatorio sistemático

Muestreo aleatorio sistemático

De igual forma se asigna un número a cada individuo de la población, el primer elemento se toma al azar, las siguientes unidades o elementos se toman de forma sistemática a partir de un número que se obtiene de:

$$k = \frac{N}{n}$$

N = tamaño de la población

n = tamaño de la muestra

Muestreo Probabilístico. Muestreo aleatorio sistemático

Muestreo aleatorio sistemático

De igual forma se asigna un número a cada individuo de la población, el primer elemento se toma al azar, las siguientes unidades o elementos se toman de forma sistemática a partir de un número que se obtiene de:

$$k = \frac{N}{n}$$

N = tamaño de la población

n = tamaño de la muestra

Observaciones:

- Se pueden introducir uniformidades que no existen en la población.

Ejemplo: $N = 150$ $n = 45$

Muestreo Probabilístico. Muestreo aleatorio sistemático

Muestreo aleatorio sistemático

De igual forma se asigna un número a cada individuo de la población, el primer elemento se toma al azar, las siguientes unidades o elementos se toman de forma sistemática a partir de un número que se obtiene de:

$$k = \frac{N}{n}$$

N = tamaño de la población

n = tamaño de la muestra

Observaciones:

- Se pueden introducir uniformidades que no existen en la población.

Ejemplo: $N = 150$ $n = 45 \Rightarrow k = \frac{150}{45}$

Muestreo Probabilístico. Muestreo aleatorio sistemático

Muestreo aleatorio sistemático

De igual forma se asigna un número a cada individuo de la población, el primer elemento se toma al azar, las siguientes unidades o elementos se toman de forma sistemática a partir de un número que se obtiene de:

$$k = \frac{N}{n}$$

N = tamaño de la población

n = tamaño de la muestra

Observaciones:

- Se pueden introducir uniformidades que no existen en la población.

Ejemplo: $N = 150$ $n = 45 \Rightarrow k = \frac{150}{45} \Rightarrow K = 3.3$

Muestreo Probabilístico. Muestreo aleatorio estratificado

Cuando las unidades a muestrear no son homogéneas y se dividen naturalmente en grupos que no se traslapan entre y que son homogéneas en su interior.

Muestreo Probabilístico. Muestreo aleatorio estratificado

Cuando las unidades a muestrear no son homogéneas y se dividen naturalmente en grupos que no se traslapan entre y que son homogéneas en su interior.

Muestreo aleatorio estratificado

Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica. Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra.

Muestreo Probabilístico. Muestreo aleatorio estratificado

Cuando las unidades a muestrear no son homogéneas y se dividen naturalmente en grupos que no se traslapan entre y que son homogéneas en su interior.

Muestreo aleatorio estratificado

Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica. Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra.

Observaciones:

- Garantiza que ningún estrato esté *sobre-representado* ni *sub-representado*.

Muestreo Probabilístico. Muestreo aleatorio por conglomerados

Muestreo aleatorio por conglomerados

Este consiste en reunir a los individuos en un grupo que forman un elemento, que tienen a la vez unidades de análisis dentro de ellos, posee la característica de ser diferentes al interior del grupo y homogéneos entre sí.

Muestreo Probabilístico. Muestreo aleatorio por conglomerados

Muestreo aleatorio por conglomerados

Este consiste en reunir a los individuos en un grupo que forman un elemento, que tienen a la vez unidades de análisis dentro de ellos, posee la característica de ser diferentes al interior del grupo y homogéneos entre sí.

Observaciones:

- El muestreo por conglomerados se usa cuando se tiene población muy grande y dispersa.
- Si un conglomerado tiene un peso mayor de unidades puede utilizarse un muestreo proporcional a su tamaño.

Muestreo No Probabilístico

Muestreo No Probabilístico

En general se seleccionan a los sujetos siguiendo determinados criterios procurando, en la medida de lo posible, que la muestra sea representativa. Sin embargo estos criterios son a decisión del investigador. Pueden ser circunstanciales o intencionales.

Intencional: Permite seleccionar casos característicos de una población limitando la muestra sólo a estos casos.

Observaciones:

- La muestra puede ser representativa, pero no permite calcular el error de muestreo ni el nivel de confianza.

Ejercicios .1

Se desea tomar una muestra aleatoria estratificada de las personas mayores de edad de un municipio, cuyos estratos son los siguientes intervalos de edades, en años: de 18 a 30, de 31 a 45, de 46 a 60 y mayores de 60. En el primer intervalo hay 7500 personas, en el segundo hay 8400, en el tercero 5700 y en el cuarto 3000. Calcule el tamaño de la muestra total y su composición, sabiendo que el muestreo se hace de forma proporcional y se han elegido al azar 375 personas del primer estrato.

Ejercicio 2.

En un pueblo habitan 700 hombres adultos, 800 mujeres adultas y 500 menores. De él se quiere seleccionar una muestra de 80 personas, utilizando, para ello, muestreo estratificado y se tomaran de forma proporcional proporcional. ¿Cuál será la composición que debe tener dicha muestra?

Evaluación

Una ganadería tiene 3000 vacas. Se quiere extraer una muestra de 120. Explica cómo se obtiene dicha muestra: a) Mediante muestreo aleatorio simple. b) Mediante muestreo aleatorio sistemático.

Tablas y Gráficas en la estadística

Tablas estadísticas

Es necesario presentar los datos colectados de manera **clara, sintética** y significativa para su mejor y fácil entendimiento. Para ello se recurre a la tabla estadística y el gráfico estadístico.

Las tablas estadísticas constan de tres partes fundamentalmente.

- La **cabeza o encabezamiento**, que ocupa la parte superior de la misma y contiene el título y el nombre de cada campo de modo claro.
- El **cuerpo** en el cual se contienen los datos a exponer.
- El **pie de tabla** el cual esta en la parte inferior. Donde se suelen dejar aclaraciones o notas que faciliten la información contenida en la tabla.

Tablas estadísticas. Ejemplo

Nombre	email
Alexis Cocoyutla González	alexcoco223@gmail.com
Gabriela Onofre Rodríguez	gabrielaonofre301@gmail.com
Zuriel Elias Angular Campos	zurielnx@gmail.com
Alejandro Soriano Maya	alejandrosorianomaya@gmail.com
Mary Carmen Espejel Sanchez	marycarmenespejel29@gmail.com
Evelia Arai Benitez Carreón	eveliaabc66@gmail.com
Daniela Aroche Orta	danielaarochorta@gmail.com
Ximena Pedraza Hernández	xime.dehernandez2001@gmail.com
Jorge Martinez Lucaro ¹	jorge231five@gmail.com

Cuadro: Tabla de ejemplo

¹Esto es un nota.

Tablas estadísticas. Notas adicionales.

Hay algunas reglas que no deben ser tomadas como normas rígidas, para la confección y representación de los datos.

- De ser necesario los valores numéricos deben ser simplificados, de modo que sean legibles.
- En caso de usar unidades de medidas, especificarla en título de la columna. En caso de ser diferentes unidades de medidas dedicarle una columna.
- La tabla debe ser lo más breve y concisa posible, para facilitar su lectura y análisis.
- Por hacer cumplir la regla anterior, no puede haber omisión de indicadores indispensables en el análisis.
- Se debe de dejar para el pie de tabla las aclaraciones y textos largos.

Ejemplo

Table 3.5.1: C_{1s} and N_{1s} core-levels binding energy for the complexes Fe(L)₂[Ni(CN)₄] where L = pyridine, 3F-pyridine, 3Cl-pyridine, 3Br-pyridine and 3I-pyridine.

Ligand	Pyridine	3F-pyridine	3Cl-pyridine	3Br-pyridine	3I-pyridine	Assignment
C 1s	284.1 (1.2)	284.2 (1.0)	284.0 (1.3)	283.9 (1.4)	284.0 (1.4)	†
	284.8 (1.3)	284.8 (1.0)	284.8 (1.3)	284.8 (1.4)	284.8 (1.4)	‡
	285.7 (1.3)	285.6 (1.0)	285.7 (1.3)	285.5 (1.4)	285.5 (1.4)	✕
	-	286.6 (1.9)	-	-	-	♠
						♣
N 1s	399.0 (1.3)	398.9 (1.4)	398.0 (1.4)	398.2 (1.3)	398.1 (1.4)	♣
	399.9 (1.4)	400.0 (1.5)	399.0 (1.3)	399.3 (1.4)	399.1 (1.4)	★

Within brackets, the peak width is indicated. The error in the values of BE is no higher than 0.1 eV.

† C atom in CNligands, ‡ Cring away from N atom, ✕ Cring closest to N atom, ♠ Cring linked to F atom,

♣ N atom in CNligands, ★ N atom in the ring

Gráficos estadístico

Gráficos estadístico

Si uno de los objetivos de la estadística es comunicar los resultados de una investigación de manera clara y significativa, la representación gráfica de esos resultados es un valioso recurso para facilitar este objetivo.

Gráficos estadístico

Si uno de los objetivos de la estadística es comunicar los resultados de una investigación de manera clara y significativa, la representación gráfica de esos resultados es un valioso recurso para facilitar este objetivo.

Gráfico estadístico

Un **gráfico estadístico** es la representación de datos estadísticos por medio de figuras geométricas (puntos, líneas, rectángulos, etc.). cuya dimensión son proporcionales al valor numérico de los datos.

Su fin es permitir, de un solo vistazo, la captación rápida del conjunto de características presentadas y evidenciar su variación.

Gráficos estadístico. Reglas para su confección.

- 1 Cuando se hace la representación gráfica de una sola variable, es costumbre indicar los datos de ésta en el eje horizontal.
- 2 Cuando la variable es de tiempo, las unidades en que se exprese éste se colocaran en el eje horizontal (horas, días, meses, años, etc.)
- 3 Cuando se representen los datos de dos variables (X , Y), se toma como convención representar la variable independiente en el eje horizontal y la variable dependiente en el eje vertical.
- 4 La disposición general de un gráfico debe ser de izquierda a derecha.

Gráficos estadístico. Ejemplo

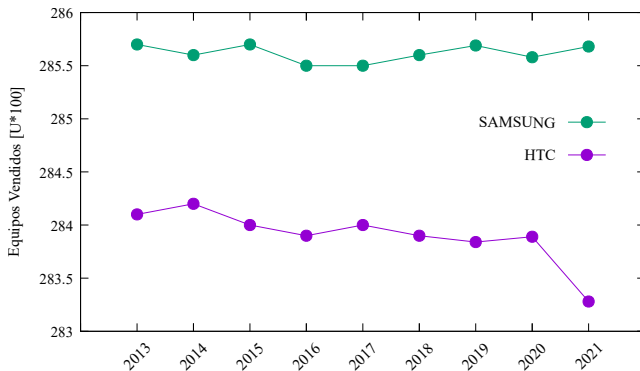


Figura: Equipos celulares vendidos por las compañías SAMSUNG y HTC en el periodo 2013-2021.

Gráficos estadístico. Ejemplo

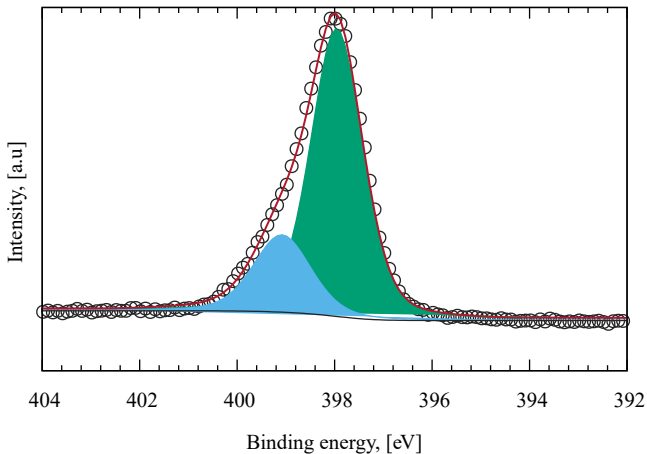


Figura: Espectro XPS de una muestra de piridina.

Gráficos estadístico. Representación de tronco y hoja

Las gráficas de tallos y hojas son una forma rápida de obtener la representación visual informativa de un conjunto de datos.

Es un método exploratorio, que permite visualizar la distribución de los datos de una manera gráfica, pero sin llegar a elaborar un gráfico.

Pasos para construir una gráfica de tallos y hojas

- 1 Seleccione uno o más de los primeros dígitos para los valores de tallo. Los segundos dígitos se convierten en hojas.
- 2 Enumere los posibles valores de tallos en una columna vertical.
- 3 Anote la hoja para cada observación junto al valor de tallo.
- 4 Indique las unidades para tallos y hojas en algún lugar de la gráfica.

Gráficos estadístico. Representación de tronco y hoja

Pasos para construir una gráfica de tallos y hojas

- 1 Seleccione uno o más de los primeros dígitos para los valores de tallo. Los segundos dígitos se convierten en hojas.
- 2 Enumere los posibles valores de tallos en una columna vertical.
- 3 Anote la hoja para cada observación junto al valor de tallo.
- 4 Indique las unidades para tallos y hojas en algún lugar de la gráfica.

Tomemos la siguiente tabla como ejemplo:

5.0	5.5	6.2	6.5	6.6
6.7	6.7	7.1	7.3	7.7
7.7	7.9	8.3	8.5	8.5
8.8	8.8	8.9	9.0	9.1
9.1	9.1	9.2	9.3	9.3

Gráficos estadístico. Representación de tronco y hoja

5|0 5
6| 2 5 6 7 7
7|1 3 7 7 9
8| 3 5 5 8 8 9
9|0 1 1 2 3 3

Gráficos estadístico. Histogramas. Tabla de Frecuencias.

Para las distribuciones de frecuencias la representación gráfica más común es el histograma.

Frecuencia y Frecuencia relativa

Considérense datos compuestos de observaciones de una variable discreta x . La **frecuencia** de cualquier valor x particular es el número de veces que ocurre un valor en el conjunto de datos. La **frecuencia relativa** de un valor es la fracción o proporción de veces que ocurre el valor. El cual se obtiene dividiendo la **frecuencia** de un valor entre el **número de observaciones en el conjunto de datos**.

Gráficos estadístico. Histogramas. Tabla de Frecuencias.

Ejemplo

Supóngase, por ejemplo, que un conjunto de datos se compone de 200 observaciones y x representa el número de cursos que un estudiante está tomando en un cuatrimestre. Si 70 de estos valores x es igual a 3, entonces:

Gráficos estadístico. Histogramas. Tabla de Frecuencias.

Ejemplo

Supóngase, por ejemplo, que un conjunto de datos se compone de 200 observaciones y x representa el número de cursos que un estudiante está tomando en un cuatrimestre. Si 70 de estos valores x es igual a 3, entonces:

frecuencia del valor 3 es 70 veces, o dicho de otro modo, hay 70 estudiantes que tomaron 3 materias.

Gráficos estadístico. Histogramas. Tabla de Frecuencias.

Ejemplo

Supóngase, por ejemplo, que un conjunto de datos se compone de 200 observaciones y x representa el número de cursos que un estudiante está tomando en un cuatrimestre. Si 70 de estos valores x es igual a 3, entonces:

frecuencia del valor 3 es 70 veces, o dicho de otro modo, hay 70 estudiantes que tomaron 3 materias.

$$FR = \frac{70}{200} = 0,35$$

Si se multiplica una frecuencia relativa por 100 se obtiene un 35 %.

Gráficos estadístico. Histogramas. Tabla de Frecuencias.

Ejemplo

Supóngase, por ejemplo, que un conjunto de datos se compone de 200 observaciones y x representa el número de cursos que un estudiante está tomando en un cuatrimestre. Si 70 de estos valores x es igual a 3, entonces:

frecuencia del valor 3 es 70 veces, o dicho de otro modo, hay 70 estudiantes que tomaron 3 materias.

$$FR = \frac{70}{200} = 0,35$$

Si se multiplica una frecuencia relativa por 100 se obtiene un 35 %.

Una **distribución de frecuencia** es una **tabla de las frecuencias** o de las frecuencias relativas, o de ambas.

Gráficos estadístico. Histogramas. Distribución de frecuencias agrupadas

Distribución de frecuencias agrupadas

La distribución de frecuencias agrupadas o tabla con datos agrupados se emplea si las variables toman un número grande de valores o la variable es continua. Se agrupan los valores en intervalos que tengan la misma amplitud denominados clases. A cada clase se le asigna su frecuencia correspondiente.

Gráficos estadístico. Histogramas. Distribución de frecuencias agrupadas

Límites de la clase

Cada clase está delimitada por el límite inferior de la clase y el límite superior de la clase.

Amplitud de la clase

La amplitud de la clase es la diferencia entre el límite superior e inferior de la clase.

Marca de clase

La marca de clase es el punto medio de cada intervalo y es el valor que representa a todo el intervalo para el cálculo de algunos parámetros.

La marca de clase se representa por c_i

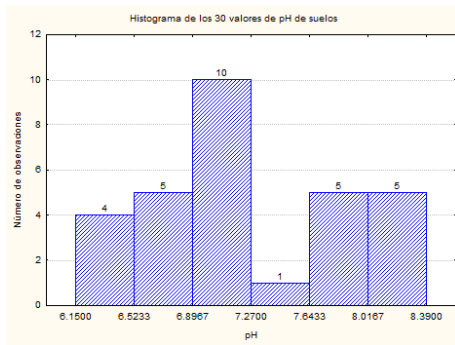
$$c_i = \frac{L_{sup} - L_{inf}}{2}$$

Gráficos estadístico. Histogramas

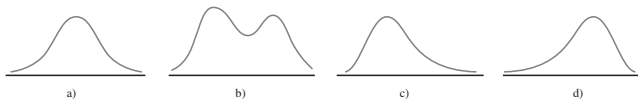
Gráficos estadístico. Histogramas

Construcción de un histograma

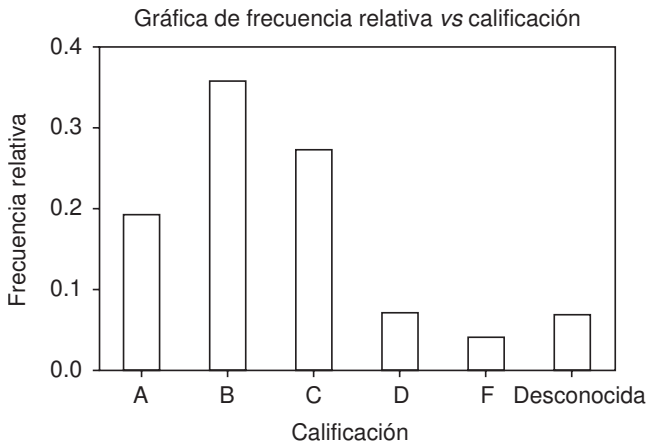
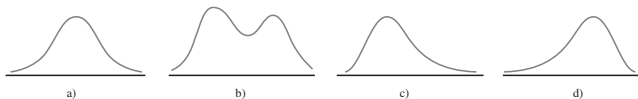
En primer lugar, se determina la frecuencia y la frecuencia relativa de cada valor x . Luego se marcan los valores x posibles en una escala horizontal. Sobre cada valor, se traza un rectángulo cuya altura es la frecuencia relativa o alternatively, la frecuencia de dicho valor.



Gráficos estadístico. Formas de los Histogramas

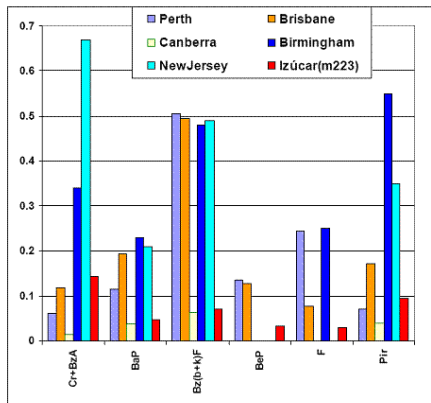


Gráficos estadístico. Formas de los Histogramas



Gráficos estadístico. Gráficos de barras

Un tipo de gráfico muy parecido al histograma es la gráfica de barras. A diferencia del histograma, no es necesario tener una escala horizontal continua. Se pueden colocar varias series, por lo que deben seleccionarse colores y anchos de las barras que nos permitan diferenciarlos.



Gráficos estadístico. Gráficos de barras

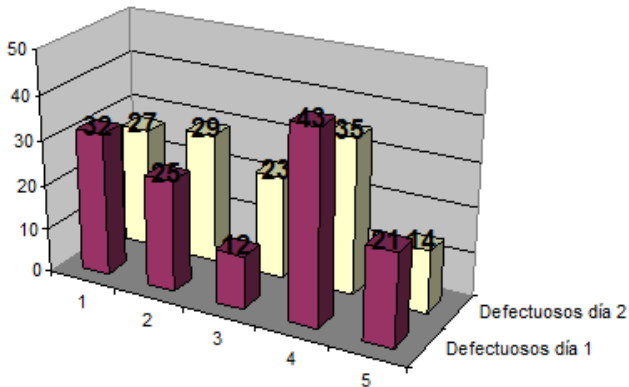


Figura: Gáfico de barras 3D.

Gráficos estadístico. Gráficos de barras

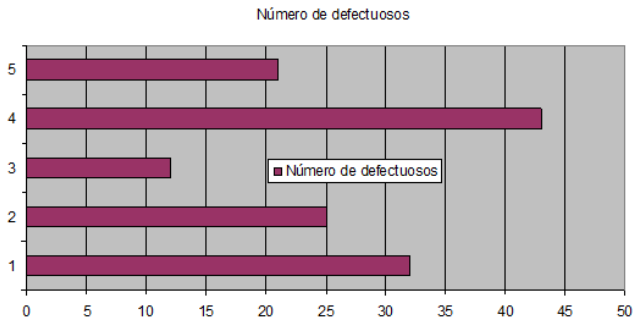
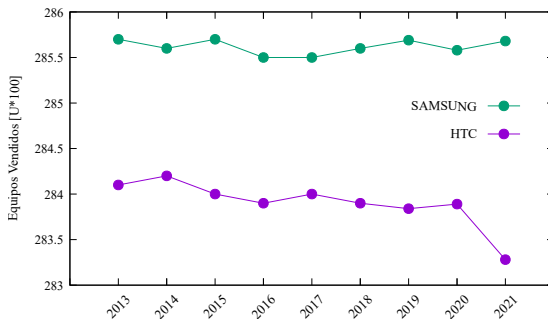


Figura: Gáfico de barras horizontal.

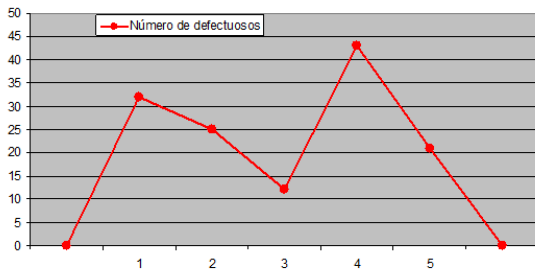
Gráficos estadístico. Gráfica de líneas

Cuando los datos se relacionan entre sí, es decir, existe cierta continuidad entre las observaciones (como por ejemplo el crecimiento poblacional, la evolución del peso de una persona, las variaciones presentadas en la medición realizada en algún experimento con determinada temporalidad) se pueden utilizar las gráficas de líneas.



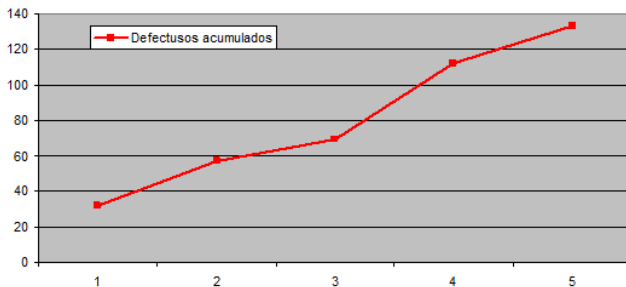
Gráficos estadístico. Polígono de frecuencias

En el polígono de frecuencias se añaden dos clases con frecuencias cero: una antes de la primera clase con datos y otra después de la última. El resultado es que se "sujeta" la línea por ambos extremos al eje horizontal y lo que podría ser una línea separada del eje se convierte, junto con éste, en un polígono.



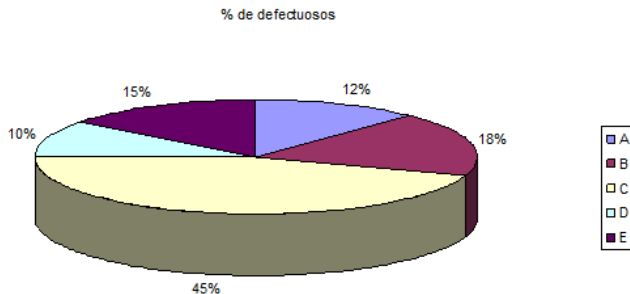
Gráficos estadístico. Gráfico de frecuencias acumuladas u ojiva

Su objetivo, al igual que el histograma y el polígono de frecuencias es representar distribuciones de frecuencias de variables cuantitativas, pero sólo para frecuencias acumuladas. No se utilizan barras en su confección, sino segmentos de recta.



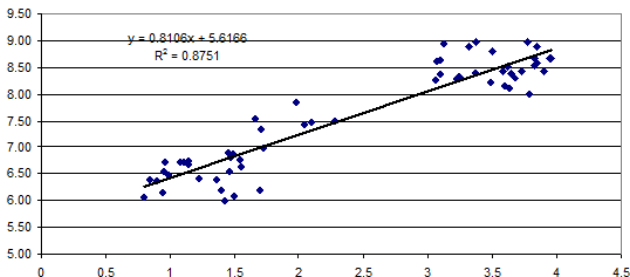
Gráficos estadístico. Gráfica de pastel

Cuando se desea resaltar las proporciones que representan algunos subconjuntos con respecto al total, son útiles las gráficas de pastel.



Gráficos estadístico. Gráficas de dispersión

Se utilizan mucho para evaluar datos experimentales. Se representan los valores de una variable dependiente, en el eje de las ordenadas, contra los valores de las variables independientes que corresponden, en el eje de las ordenadas. Esto nos permite evaluar tendencias, correlaciones, etc.



Medidas de tendencia central, localización y dispersión

Introducción Medidas de tendencia central, localización y dispersión.

Para describir un conjunto de datos se utilizan **medidas de resumen**, las cuales nos permiten **condensar** los datos de tal manera que con dos o tres valores obtenemos una idea sobre los *valores alrededor de los cuales se agrupan los datos*, la *dispersión de ese agrupamiento* y la forma de la *distribución de los datos*.

Introducción Medidas de tendencia central, localización y dispersión.

Para describir un conjunto de datos se utilizan **medidas de resumen**, las cuales nos permiten **condensar** los datos de tal manera que con dos o tres valores obtenemos una idea sobre los *valores alrededor de los cuales se agrupan los datos*, la *dispersión de ese agrupamiento* y la forma de la *distribución de los datos*. Estos valores representativos se denominan **estadísticos** o **estadígrafos**. Su ventaja radica en que facilitan considerablemente el manejo de los datos y la transmisión de la información.

Clases de estadísticos o estadígrafos

- 1 **Los estadísticos de posición o localización.** Indican alrededor de cuál valor se agrupan los datos obtenidos, interesando más en este curso los que caracterizan la tendencia central. La media, la mediana, la moda son los más comunes.

Clases de estadísticos o estadígrafos

- 1 **Los estadísticos de posición o localización.** Indican alrededor de cuál valor se agrupan los datos obtenidos, interesando más en este curso los que caracterizan la tendencia central. La media, la mediana, la moda son los más comunes.
- 2 **Los estadísticos de dispersión.** Indican qué tan esparcidos están los datos obtenidos respecto de la medida o medidas de posición que se adopten.

Clases de estadísticos o estadígrafos

- ❶ **Los estadísticos de posición o localización.** Indican alrededor de cuál valor se agrupan los datos obtenidos, interesando más en este curso los que caracterizan la tendencia central. La media, la mediana, la moda son los más comunes.
- ❷ **Los estadísticos de dispersión.** Indican qué tan esparcidos están los datos obtenidos respecto de la medida o medidas de posición que se adopten.
- ❸ **Los estadísticos de forma.** Estudian la asimetría – simetría y deformación respecto de una distribución modelo denominada distribución normal.

Clases de estadísticos o estadígrafos

- ① **Los estadísticos de posición o localización.** Indican alrededor de cuál valor se agrupan los datos obtenidos, interesando más en este curso los que caracterizan la tendencia central. La media, la mediana, la moda son los más comunes.
- ② **Los estadísticos de dispersión.** Indican qué tan esparcidos están los datos obtenidos respecto de la medida o medidas de posición que se adopten.
- ③ **Los estadísticos de forma.** Estudian la asimetría – simetría y deformación respecto de una distribución modelo denominada distribución normal.
- ④ **Los estadísticos específicos.** Son propios de cada especialidad, por ejemplo el índice de mortalidad infantil en una población, el producto interno bruto, etc.

Estadísticos de posición y localización. Tendencia central

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana
 - ▶ Moda

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana
 - ▶ Moda
- b) Medidas de posición o tendencia no centrales: informan de como se distribuye el resto de los valores de la serie.

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana
 - ▶ Moda
- b) Medidas de posición o tendencia no centrales: informan de como se distribuye el resto de los valores de la serie.
 - ▶ Cuartiles

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana
 - ▶ Moda
- b) Medidas de posición o tendencia no centrales: informan de como se distribuye el resto de los valores de la serie.
 - ▶ Cuartiles
 - ▶ Deciles

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana
 - ▶ Moda
- b) Medidas de posición o tendencia no centrales: informan de como se distribuye el resto de los valores de la serie.
 - ▶ Cuartiles
 - ▶ Deciles
 - ▶ Percentiles

Estadísticos localización o tendencia central

Las medidas de **posición** (**localización**) o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana
 - ▶ Moda
- b) Medidas de posición o tendencia no centrales: informan de como se distribuye el resto de los valores de la serie.
 - ▶ Cuartiles
 - ▶ Deciles
 - ▶ Percentiles

Estadísticos localización o tendencia central

Las medidas de **posición (localización)** o **tendencia** nos facilitan información sobre la serie de datos que estamos analizando. Estas las podemos agrupar en dos tipos:

- a) Medidas de posición o tendencia central: informan sobre los valores medios de la serie de datos.
 - ▶ Medias (aritmética, aritmética ponderada, geométrica, armónica, etc.)
 - ▶ Mediana
 - ▶ Moda
- b) Medidas de posición o tendencia no centrales: informan de como se distribuye el resto de los valores de la serie.
 - ▶ Cuartiles
 - ▶ Deciles
 - ▶ Percentiles

Este último (**b**) también es reverenciado como Estadísticos de Localización.

Estadísticos tendencia central. Media aritmética o promedio (MA)

Media aritmética o promedio (MA)

Es la suma de todos los valores de la variable dividida entre el número total de elementos, N.

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n}{N} = \frac{\sum_{n=i}^n x_i}{N}$$

Estadísticos tendencia central. Media aritmética o promedio (MA)

Media aritmética o promedio (MA)

Es la suma de todos los valores de la variable dividida entre el número total de elementos, N.

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n}{N} = \frac{\sum_{n=i}^n x_i}{N}$$

Media aritmética o promedio (MA). Para el calculo de tablas sin agrupar.

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

donde n_i representa la frecuencia absoluta del valor x_i

Media aritmética o promedio (MA). Ejemplo

Ejemplo

En un almacén de trigo se realizan determinaciones de humedad a varias muestras de harina, obteniendo los resultados de la tabla siguiente. Determine el % de humedad promedio.

$x_i, \%$	n_i
10	3
11	2
12	3
13	2

Empleando la ecuación:

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

Media aritmética o promedio (MA). Ejemplo

Obtenemos:

$x_i, \%$	n_i	$x_i * n_i$
10	3	30
11	2	22
12	3	36
13	2	26
Suma	10	116

Media aritmética o promedio (MA). Ejemplo

Obtenemos:

$x_i, \%$	n_i	$x_i * n_i$
10	3	30
11	2	22
12	3	36
13	2	26
Suma	10	116

Sustituimos:

$$\bar{X} = \frac{\sum x_i n_i}{N} = \frac{116}{10} = 11,6 \%$$

Media aritmética o promedio (MA). Propiedades

Propiedades de la media:

- La suma de las desviaciones de los valores de la variable con respecto a la media aritmética es 0. La media aritmética equilibra esos desvíos, o sea, se puede considerar a la media como el centro de gravedad de los datos.
- Si a todos los valores de la variable se le suma una misma cantidad, la media aritmética queda aumentada en dicha cantidad
- Cuando a la media se la multiplica o divide por una constante, la nueva media queda multiplicada o dividida por dicha constante.

Media aritmética o promedio (MA). Propiedades

Propiedades de la media:

- La suma de las desviaciones de los valores de la variable con respecto a la media aritmética es 0. La media aritmética equilibra esos desvíos, o sea, se puede considerar a la media como el centro de gravedad de los datos.
- Si a todos los valores de la variable se le suma una misma cantidad, la media aritmética queda aumentada en dicha cantidad
- Cuando a la media se la multiplica o divide por una constante, la nueva media queda multiplicada o dividida por dicha constante.

La media aritmética presenta un problema. Los valores extremos, influyen mucho en ella, restándole en ocasiones significación.

Estadísticos tendencia central. Mediana (Me)

Mediana (Me)

La mediana es el valor que deja por debajo y por encima de él el mismo número de datos.

Observaciones:

- No presenta el problema de estar influida por los valores extremos.
- No utiliza en su cálculo toda la información de la serie de datos.

Ejemplo:

9, 1, 34, 8, 5, 10, 13, 17, 6

Estadísticos tendencia central. Mediana (Me)

Mediana (Me)

La mediana es el valor que deja por debajo y por encima de él el mismo número de datos.

Observaciones:

- No presenta el problema de estar influida por los valores extremos.
- No utiliza en su cálculo toda la información de la serie de datos.

Ejemplo:

9, 1, 34, 8, 5, 10, 13, 17, 6

Solución:

9, 1, 34, 8, **5**, 10, 13, 17, 6

Estadísticos tendencia central. Mediana (Me)

Los datos pueden estar agrupados o no.

Estadísticos tendencia central. Mediana (Me)

Los datos pueden estar agrupados o no.

Si los datos están sin agrupar:

Si el número de observaciones es impar, la mediana corresponde al valor central. Si el número de observaciones es par, la mediana corresponde al promedio entre los dos valores centrales.

Ejemplo para observaciones impar:

9, 1, 34, 8, 5, 10, 13, 17, 6

Ejemplo para observaciones par:

9, 1, 34, 8, 10, 13, 17, 6

promedio entre 8 y 10 \Rightarrow 9

Estadísticos tendencia central. Mediana (Me)

Estadísticos tendencia central. Moda (Mo)

Es el valor de la variable de mayor frecuencia. Se le usa en lugar de la media o la mediana, cuando se desea señalar el valor más típico o común del total de observaciones efectuadas. Cuando la distribución de datos presenta más de un máximo relativo, se le llama bimodal. Cuando los datos están sin agrupar la moda se determina por la simple inspección de la lista ordenada:

X: 10 10 11 12 13 13 13 14 15 15 16

Estadísticos tendencia central. Moda (Mo)

Es el valor de la variable de mayor frecuencia. Se le usa en lugar de la media o la mediana, cuando se desea señalar el valor más típico o común del total de observaciones efectuadas. Cuando la distribución de datos presenta más de un máximo relativo, se le llama bimodal. Cuando los datos están sin agrupar la moda se determina por la simple inspección de la lista ordenada:

X: 10 10 11 12 13 13 13 14 15 15 16

la moda es 13

Cuando están agrupados es necesario aplicar fórmulas, por lo que resulta más sencillo utilizar los paquetes estadísticos.

Estadísticos tendencia central. Cuartiles, Deciles y Percentiles

Cuantiles

Representa las fracciones del total, en otras palabras, la son cuartas partes del total. El primer cuartil Q_1 deja el 25 % de los valores por debajo. El segundo cuartil es igual a la mediana y el tercero Q_3 deja el 75 %.

Estadísticos tendencia central. Cuartiles, Deciles y Percentiles

Cuantiles

Representa las fracciones del total, en otras palabras, la son cuartas partes del total. El primer cuartil Q_1 deja el 25 % de los valores por debajo. El segundo cuartil es igual a la mediana y el tercero Q_3 deja el 75 %.

Deciles

Las fracciones son décimas partes del total. El primer decil D_1 deja el 10 % de los valores por debajo y el resto por encima. El quinto decil es la mediana.

Estadísticos tendencia central. Cuartiles, Deciles y Percentiles

Cuantiles

Representa las fracciones del total, en otras palabras, la son cuartas partes del total. El primer cuartil Q_1 deja el 25 % de los valores por debajo. El segundo cuartil es igual a la mediana y el tercero Q_3 deja el 75 %.

Deciles

Las fracciones son décimas partes del total. El primer decil D_1 deja el 10 % de los valores por debajo y el resto por encima. El quinto decil es la mediana.

Percentil

Las fracciones son centésimas partes del total. Así, el percentil veinte (P_{20}), deja el 20 % de los valores por debajo.

Medidas de dispersión

Se denomina dispersión o variabilidad, a la mayor o menor separación de los valores de la muestra, respecto de las medidas de tendencia central que se hayan calculado.

Hay varios estadísticos para ello:

- Rango
- Varianza
- Desviación estándar
- Desviación media

Medidas de dispersión. El Rango

Es la medida más simple y se define como la diferencia entre el valor máximo y el valor mínimo del grupo de datos:

$$R = X_{max} - X_{min}$$

Es sencilla de calcular y de fácil comprensión pero tiene la desventaja de ser la medida más grosera.

Dos grupos de datos con igual recorrido puede tener una dispersión diferente:

Medidas de dispersión. La varianza

La varianza

Se define la varianza como la sumatoria del cuadrado de las desviaciones de cada medición con respecto a la media aritmética dividida entre el número de mediciones.

Definamos la varianza como s^2 o σ^2 .

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

Medidas de dispersión. Desviación media

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

Medidas de dispersión. Desviación estándar

Desviación estándar

Si analizamos la fórmula, las unidades de la varianza están dadas en las unidades de la variable, pero elevadas al cuadrado. Para poder tener entonces una idea de la dispersión en las propias unidades de la variable, se toma la raíz cuadrada de este valor, que es la desviación estándar o típica

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

TO BE CONTINUE