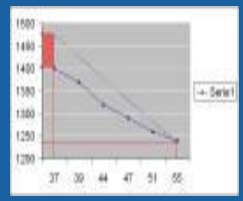
World Training Colombia

Uso de la Estadística en la Preparación de Planes de Seguridad Vial

Bienvenidos

aboadar@gmail.com

Uso de la Estadística en la Preparación de Planes de Seguridad Vial



AGENDA

1. **Definiciones**
2. **Métodos descriptivos**
3. **Accidentalidad**
4. **Calculo de índices**
5. **Métodos multivariados**

"Llegará el día en el que el pensamiento estadístico será una condición tan necesaria para la convivencia eficiente como la capacidad de leer y escribir"

H.G. Wells

aboadar@gmail.com

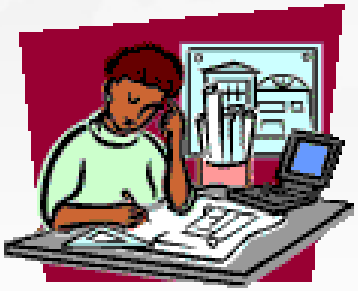
¡Objetivos de esta sesión¡

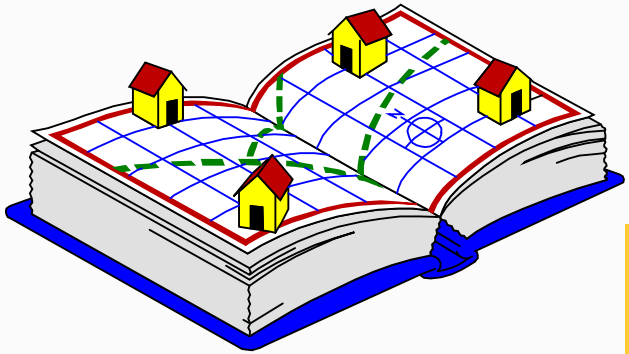
- ✓ Recordar qué significa estadística y su relación con el transporte y los planes de prevención vial.
- ✓ Explicar qué es estadística descriptiva y estadística inferencial y como se relaciona con los planes de prevención vial.
- ✓ Diferenciar entre una variable cualitativa y una variable cuantitativa utilizadas en la accidentalidad en el transporte
- ✓ Identificar variables relacionadas con el flujo vehicular y la accidentalidad
- ✓ Diferenciar entre niveles de medición nominal, ordinal, por intervalo y de razón.
- ✓ Aplicaciones y talleres
- ✓ Métodos descriptivos

aboadar@gmail.com

1. Definiciones

Carecer de datos estadísticos en cuanto a lo que acontece tanto interna como externamente, impide decidir sobre bases racionales, y adoptar las medidas preventivas y correctivas con el suficiente tiempo para evitar daños, en muchos casos irreparables, para cualquier entidad u organización.





***PLAN NACIONAL DE SEGURIDAD VIAL*: Reducir la probabilidad en la ocurrencia de los accidentes, mediante la unión de esfuerzos del sector público y privado, de una manera sistemática y coordinada, que permita ofrecer una protección a los usuarios y controlar las amenazas existentes en el sistema vial"**

aboadar@gmail.com

1. Definiciones

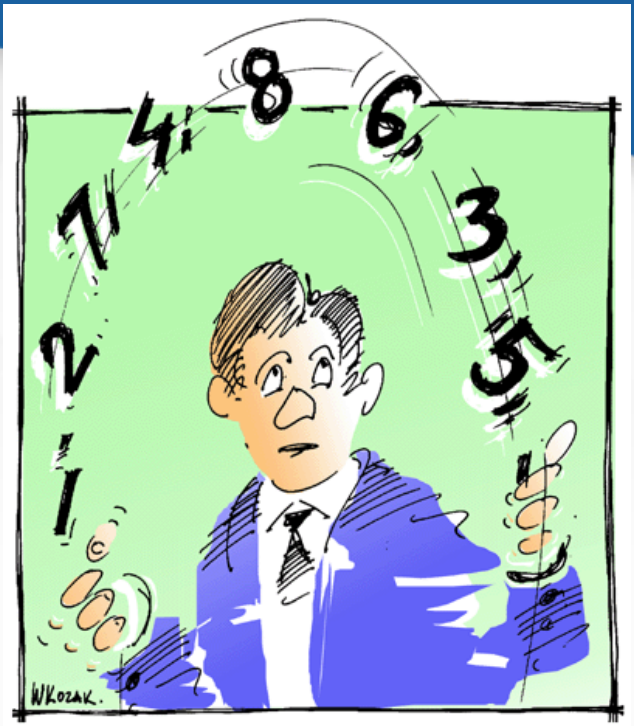
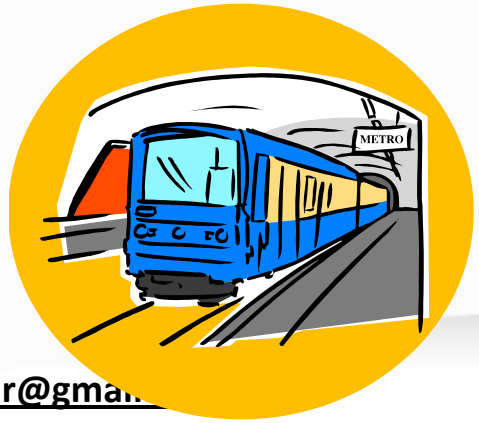


Imagen popular de la estadística:
"Existen medias mentiras, mentiras y estadísticas".

Dos significados:

(1) Colección de datos numéricos (una estadística).

(2) Ciencia: obtener regularidades de fenómenos de masas (la estadística).



Según recientes estadísticas, en los accidentes ferroviarios, el mayor número de víctimas son del último vagón. Si esto es cierto, ¿por qué no lo quitan?

aboadar@gmail

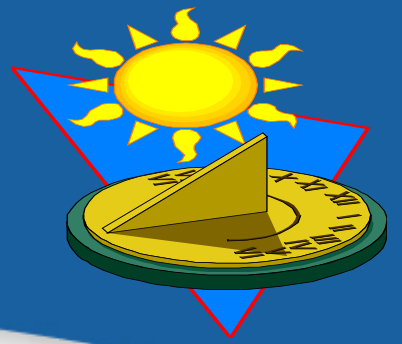
¿Cómo se define estadística?

- *Estadística* es la ciencia que se encarga de recolectar, organizar, presentar, tabular, graficar, procesar, analizar e interpretar datos con el propósito de ayudar a una toma de decisiones más efectiva.

Estudia el comportamiento de los fenómenos de interés en los distintos campos del conocimiento.

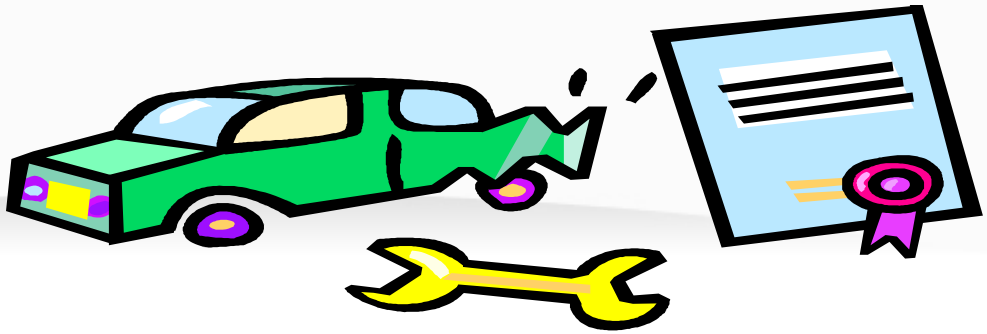
aboadar@gmail.com

Relación estadística y planes de seguridad vial



Objetivos Plan de Seguridad Vial:

- ✓ Disminuir las cifras de mortalidad en las carreteras.
- ✓ Correlacionar el aumento de la movilidad con incremento seguridad vial.
- ✓ Mejorar las condiciones de seguridad en las carreteras
- ✓ Establecer procedimientos de diseño y construcción que garanticen que las nuevas carreteras dispongan de las mejores características de seguridad.
- ✓ Hacer que se respete la señalización, evitando incongruencias y haciéndola más útil.
- ✓ Desarrollar y divulgar investigaciones relacionadas con la seguridad vial, fomentando el intercambio de información con otros organismos.
- ✓ Fomentar la difusión de información sobre siniestralidad, con el objetivo de concientizar a la sociedad.



aboadar@gmail.com

Transformación de los datos en información

Datos

Administración Base de Datos

Recolectar - depurar

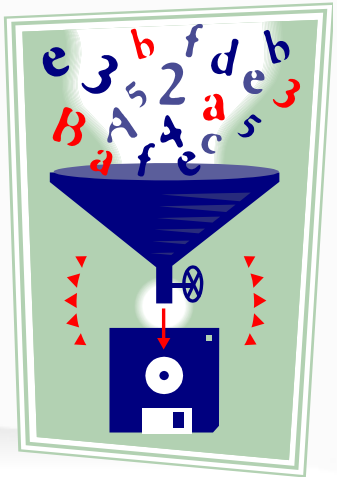
Tabular - graficar

Procesar - analizar

Interpretar

Sistema de Información estadístico

Información



aboadar@gmail.com

Resumen de las definiciones

La Estadística es la Ciencia que se encarga de

Descriptiva
- **Recolectar, ordenar, sistematizar y presentar** datos referentes a un fenómeno que presenta variabilidad o incertidumbre para su estudio metódico, con objeto de

Probabilidad
- **deducir las leyes** que rigen esos fenómenos,

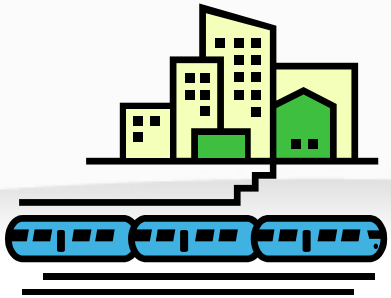
Inferencia
- y poder de esa forma hacer estimaciones sobre los mismos para tomar **decisiones, disminuyendo la incertidumbre** y alcanzar**conclusiones**.

aboadar@gmail.com

¿Por qué la estadística en la preparación de planes de prevención vial

- Las técnicas estadísticas se usan ampliamente en áreas de comercialización, contabilidad, control de calidad, consumidores, deportes, administración de hospitales, educación, política, medicina, etcétera... *¿Y en los planes de prevención vial?*

La necesidad de vías … construir el mayor kilometraje de calles, carreteras en el menor tiempo.... Disminuir las tasas de accidentes…Se requiere tener en cuenta el *aspecto estructural* … y garantizar una mayor duración …. Entonces la parte *operacional* se vuelve importante y su eficiencia depende de tener en el momento con exactitud y precisión … ***D A T O S***



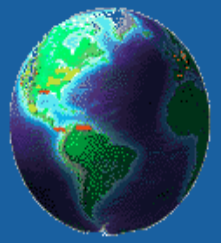
aboadur@gmail.com

Algunas de las muchas aplicaciones, son:

- Determinar las velocidades de los vehículos en una vía
- Evaluar la tasa de crecimiento promedio anual en el parque automotor
- Predecir el número de accidentes semanales en una intersección
- Calcular las tasas de la accidentalidad por kilómetros recorridos
- Estimar el número de kilómetros por pavimentar en una vía y su costo
- Entender el problema del congestionamiento
- Evaluar el número de habitantes por vehículo desde ….
- Describir la accidentalidad de Bogotá y la región…
- Estimar el número heridos en accidentes de tránsito por cada 100 mil habitantes

aboadar@gmail.com

Tipos de estadísticas



- Estadística descriptiva: Métodos para organizar, resumir y presentar datos de manera informativa.
- Ejemplo:

ACCIDENTALIDAD VIAL 2002	
ACCIDENTES	189.933
HERIDOS	42.837
MUERTOS	6.063

aboadar@gmail.com

Tipos de estadísticas

- Estadística inferencial: es una decisión, estimación, predicción o generalización sobre una población, con base en una muestra.

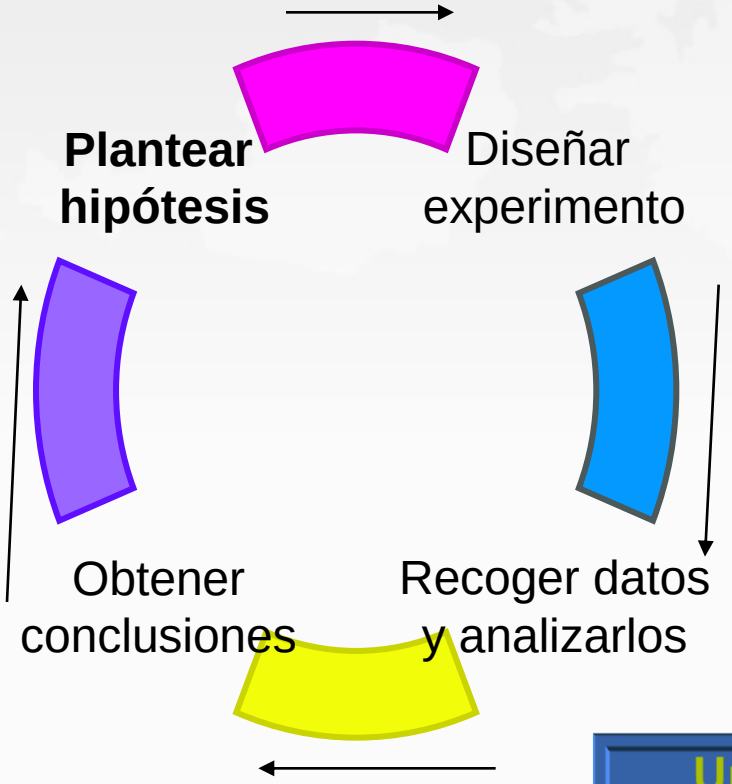
Ejemplo

ESTIMACIÓN COSTOS DE LOS ACCIDENTES DE TRANSITO

ACCIDENTE	COSTO PROMEDIO
Con sólo daños materiales	$ 4.6 Millones
Con heridos	$ 20.8 Millones
Con Muertos	$ 118.5 Millones

ab[illegible]dar[illegible]mail.com

Método científico y estadística

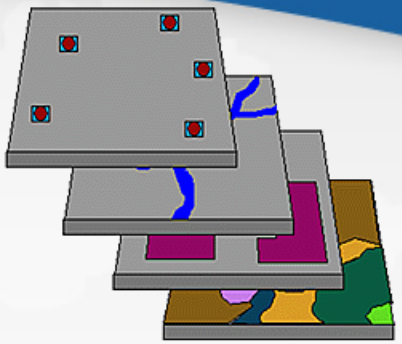


Un viejo refrán dice: No hay ciencia sin estadística

aboadar@gmail.com

El rol de la estadística en los planes de prevención vial

DATOS



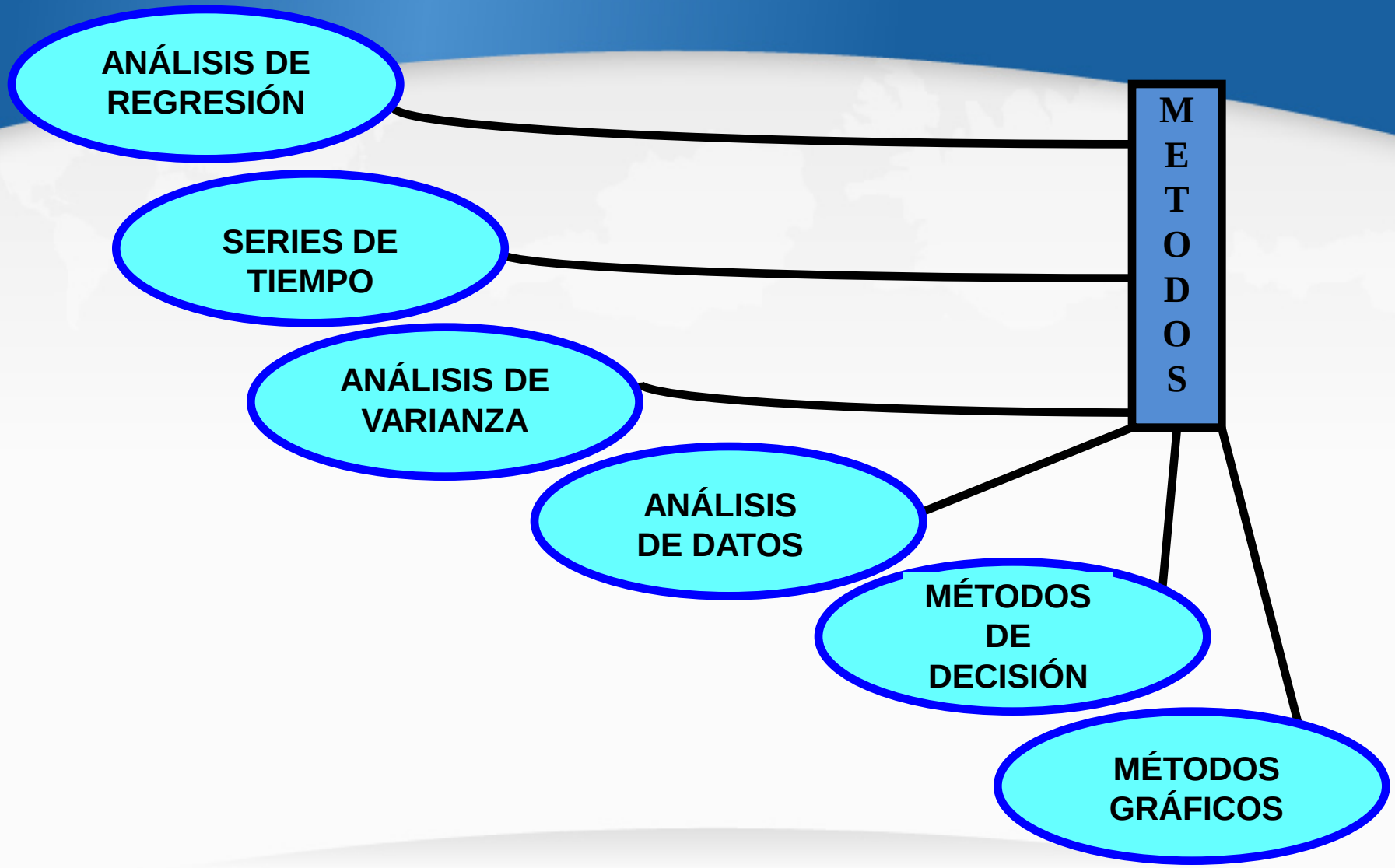
PROCESOS

INFORMACIÓN

TOMA DE DECISIONES

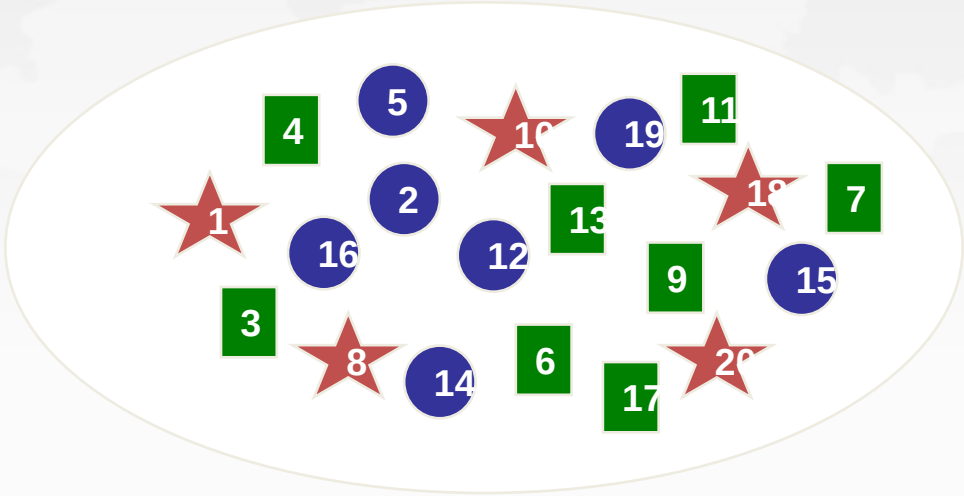
aboadar@gmail.com

ALGUNOS MÉTODOS ESTADÍSTICOS

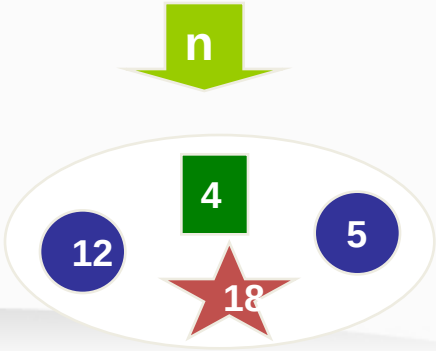


Definiciones básicas

Población es el conjunto de todos los posibles individuos, objetos o medidas de interés.



Una muestra es una porción, o parte, de la población de interés.

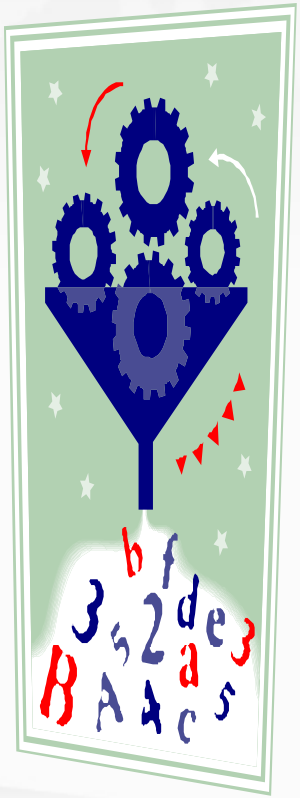


aboadar@gmail.com

Parámetros y estadísticos

Parámetro: Es una cantidad numérica calculada sobre una población.

- Heridos en accidentes de transito en 2009
- La idea es resumir toda la información que hay en la población en unos pocos números (parámetros).



Estadístico: Es una cantidad numérica calculada sobre una muestra

- La altura media de los que estamos en este salón.
  - Somos una muestra (¿representativa?) de la población.
- Si un estadístico se usa para aproximar un parámetro también se le suele llamar estimador.

aboadar@gmail.com

Tipos de variables

- Variable cualitativa o de atributos: la característica o variable que se estudia no es numérica, *son categóricas. observables*.

- ✓ Sexo
- ✓ afiliación religiosa
- ✓ tipo de automóvil que se posee
- ✓ lugar de expedición del pase
- ✓ Modelo de un auto
- ✓ estado de la vía
- ✓ clasificación de un accidente de tráfico
- ✓ causa del accidente (exceso de velocidad, embriaguez…).

aboadar@gmail.com

Tipos de variables

- Variable cuantitativa:

  la variable se puede registrar numéricamente, es medible.

✓ Numero de autos que circulan en Bogotá en las horas pico

✓ Número de heridos en un accidente de transito

✓ Longitud de las vías de Bogotá

✓ Calificación de los usuarios al medio de transporte que más usan

✓ Minutos que faltan para que termine la clase

aboadar@gmail.com

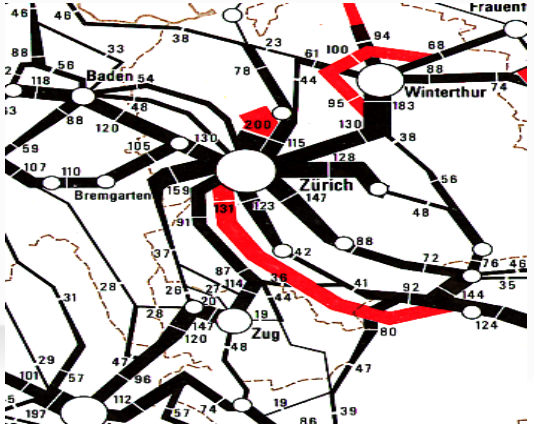
Tipos de variables

- Las variables cuantitativas se pueden clasificar como discretas o continuas.
- Variables discretas: sólo pueden adquirir ciertos valores y casi siempre hay “brechas” entre esos valores.

✓ número de habitaciones en una casa

✓ número de comparendos impuestos en el ultimo día sin carro

✓ número de accidentes



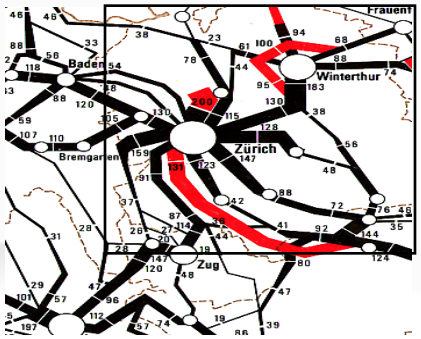
aboadar@gmail.com

Variables relacionadas con el flujo vehicular

- ***Tasa de flujo*:**

Frecuencia a la cual pasan los vehículos por un punto o sección transversal de un carril o calzada, es decir número de vehículos *N*, que pasan durante un intervalo de tiempo específico *T,* inferior a una hora en unidades de minutos o segundos, luego la tasa de flujo

Q = N / T, se expresa en vehículos por hora



aboadar@gmail.com

Variables relacionadas con el flujo vehicular

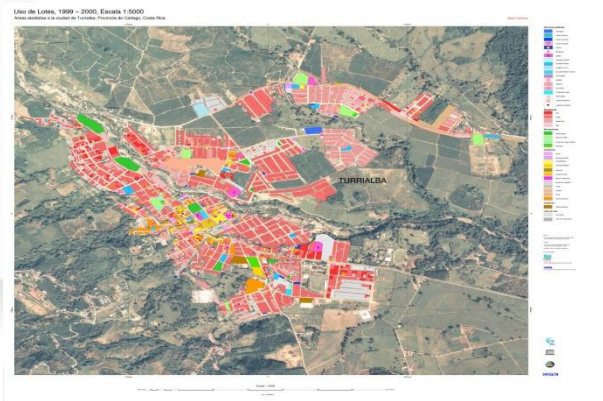
Intervalo simple (h_i)

Es el intervalo de tiempo entre el paso de dos vehículos consecutivos, generalmente expresados en segundos y medido entre puntos homólogos del par de vehículos

- ***Intervalo promedio ($\bar{h}$)***

Es el promedio de todos los intervalos simples h_i existentes entre los diversos vehículos que circulan por una vía

$$\bar{h} = \sum_{i=1}^{N-1} hi \; / \; N - 1$$



aboadar@gmail.com

Variables relacionadas con el flujo vehicular

Intervalo promedio (h)

Es el promedio de todos los intervalos simples h_i existentes entre los diversos vehículos que circulan por una vía

$$\overline{h} = \sum_{i=1}^{N-1} hi \; / \; N - 1$$

donde:

$\overline{h}$ = intervalo promedio (s/veh)

N = número de vehículos (veh)

N – 1 = número de intervalos (veh)

hi = intervalo simple entre el vehículo i y el vehículo i + 1

aboadar@gmail.com

• Variables relacionadas con la velocidad

• Velocidad instantánea o de punto *(instante preciso)*

• Velocidad media temporal *(promedio velocidades de punto)*

• Velocidad media espacial *(promedio velocidades de punto en un tramo)*

• Velocidad de recorrido *(velocidad global o de viaje, resulta de dividir la distancia, desde principio a fin del viaje, entre tiempo total, incluye demoras)*

• Velocidad de marcha o crucero *(resultado de dividir la distancia recorrida entre el tiempo durante el cual el vehículo estuvo en movimiento)*

• Distancia de recorrido

• Tiempo de recorrido

***Velocidad*: relación entre el espacio recorrido y el tiempo que se tarda en recorrerlo. Para un vehículo es la relación de movimiento (km/h) V = d / t**

aboadar@gmail.com

• Variables relacionadas con la densidad

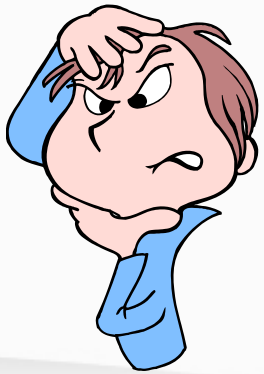
• ***Densidad de concentración (K)***

Número, N, de vehículos que ocupan una longitud especifica, d, de una vía en un momento dado

• ***Espaciamiento simple (Si)***

Es la distancia entre el paso de dos vehículos consecutivos, en metros, y medida entre sus defensas traseras.

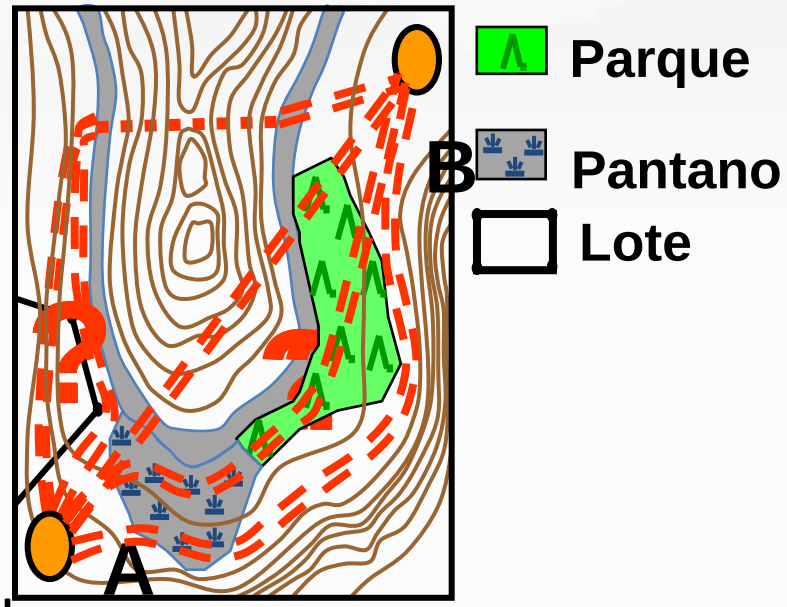
• ***Espaciamiento promedio***



aboadar@gmail.com

Niveles o escalas de medición

- Nivel nominal: los datos sólo se puede clasificar en categorías, no se pueden ordenar.



- ✓ color de los ojos
- ✓ Sexo
- ✓ afiliación religiosa.
- ✓ marca de un automóvil

aboadar@gmail.com

Niveles de medición

- Mutuamente excluyente: un individuo, objeto o artículo, al ser incluido en una categoría, debe excluirse de las demás.
- EJEMPLO: color de un automóvil
- Exhaustivo: cada persona, objeto o artículo debe clasificarse en al menos una categoría.
- EJEMPLO: Categoría del pase.

aboadar@gmail.com

Niveles de medición

- Nivel ordinal: involucra datos que se pueden ordenar, pero no es posible determinar las diferencias entre los valores de los datos o no tienen significado.

✓ Estado de una vía

✓ Clasificación de las vías

✓ Estratificación socio - económica

aboadar@gmail.com

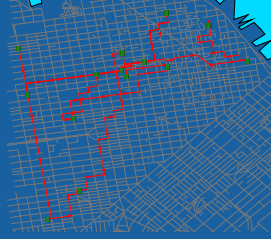
Niveles de medición

- Nivel de intervalo: similar al nivel ordinal, con la propiedad adicional de que se pueden determinar cantidades significativas de las diferencias entre los valores. No existe un punto cero natural.

✓ temperatura en la escala de grados Fahrenheit.

aboadar@gmail.com

Niveles de medición



- Nivel de razón: el nivel de intervalo con un punto cero inicial inherente. Las diferencias y razones son significativas para este nivel de medición.

✓ Ingreso por comparendos en un municipio

✓ Número de habitaciones en el *Hotel* .. *Lucho* .

aboadar@gmail.com

Pensamiento proporcional

- **Proporción**: es un concepto matemático relacionado con fracciones y porcentajes. Parte de la cantidad total o número de observaciones, expresada en forma decimal. (Las proporciones tienen un común denominador de 1; los porcentajes tienen un común denominador de 100)

- **Proporción** = Parte / Todo

- **Proporción** = # en una categoría / # en grupo total

- **Tasa**: frecuencia de ocurrencia de un fenómeno en relación con un número "base" especificado de sujetos de una población (Las tasas tienen un común denominador útil seleccionado en múltiplos de 10)

aboadar@gmail.com

Pensamiento proporcional

Proporción:

Es un cociente en el que el numerador está incluido en el denominador.

Por ejemplo, si en una población de 25.000 habitantes se diagnostican 1.500 pacientes con diabetes, la proporción de diabetes en esa población es:

Proporción de diabetes = (1.500 / 25.000) = 0.06

6 % expresada en porcentaje

El valor de una proporción puede variar así de 0 a 1, y suele expresarse como un porcentaje, que varia entre 1 y 100.-

Pensamiento proporcional

El concepto de tasa es similar al de una proporción, con la diferencia de que las tasas llevan incorporado el concepto de tiempo. Toman todos los casos de un evento (enfermedad o muerte) por una causa, pertenecientes a una población total, en un lugar y período determinado.

Las tasas incorporan el concepto de tiempo y lugar, o sea que numerador y denominador deben estar referidos, al mismo tiempo y lugar.

Las tasas representan la fuerza con que se produce un hecho determinado en una población, y esto es igual a riesgo.

Se pueden hacer pronósticos en base a tasas calculadas en período inmediatamente anteriores.

Toda tasa es una expresión numérica de un riesgo al que estuvo sometida una población, Mide un riesgo de salud (enfermedad o muerte) en términos probabilísticos.-

Las Tasas pueden ser generales, específicas y particulares

Generales:

Toman todos los casos, por ejemplo, de muerte o todos los casos de muerte por una causa, con respecto a la población total de un lugar y período determinado:

Tasa general de Mortalidad =

(Nº total de muertes en un lugar y tiempo determinado/ Población en ese lugar a mitad del período) *1000

Ejemplo: en 2005 se produjeron 81.632 muertes por enfermedades del corazón

Tasas Específicas

Son las tasas que se construyen relacionando el fenómeno a un sector de la población (por edad por sexo etc.)

Tasa Específica de Mortalidad infantil =

(Nº de muertes < de 1 año para un área y tiempo determinado) / (Nacidos Vivos área y tiempo determinado) *1000

Ejemplo: En 2000 se produjeron 17.348 muertes de menores de 1 año. Los nacidos vivos, en promedio, en ese año fueron de 577.463

aboadar@gmail.com

Pensamiento proporcional

- **Razón**: Cociente de dos números o de dos cantidades comparables entre sí
- **Razón** = $\Sigma X_i / \Sigma Y_i$
- **Tasas de crecimiento**: Permiten analizar el comportamiento de una variable en un determinado período de tiempo en relación con el anterior.

$$T_t = (X_t - X_{t-1}) / X_{t-1}$$

Donde t puede referirse a un día, mes, trimestre, año, quinquenio, etc.

- **Datos transversales**: Datos reunidos en el mismo, o aproximadamente en el mismo, punto en el tiempo.
- **Datos de una serie de tiempo**: Datos reunidos en diferentes períodos sucesivos.

aboadar@gmail.com

Taller No. 1

1. **Defina y clasifique las variables que se deben tener en cuenta en los planes de prevención vial**
2. **¿Cuáles son sus escalas de medición?**
3. **¿Qué método usaría para recolectarlas?**
4. **¿Cómo las organizaría en una base de datos?**

aboadar@gmail.com

Accidentalidad

- ***Causa aparente de los accidentes***

En el informe del agente o policía de transito esta la base de la estadística.

Informe perfila la *"causa"* del accidente: Ubicación, frecuencia, número de heridos…

Con esta información se logra determinar las causas reales (vía o calle, vehículo, usuario…)

Al relacionar los saldos en muertos, heridos, o el kilometraje recorrido, se dispondrá de cifras o índices que permitan hacer comparaciones acerca del comportamiento de la accidentalidad.

aboadar@gmail.com

Índices con respecto a la población

• **Índices de Accidentalidad (No. De accidentes), morbilidad (No. De heridos) y mortalidad (No. De muertos), con respecto al número de habitantes en el año de que se trate expresado por cada 100.000 habitantes.**

Índice de accidentalidad ($I_{A/P}$) = (No. De accidentes en el año X 100000)/No. De habitantes

Índice de morbilidad ($I_{morb/P}$) = (No. De heridos en el año X 100000)/No. De habitantes

Índice de mortalidad ($I_{mort/P}$) = (No. De muertos en el año X 100000)/No. De habitantes

Útiles para comparar ciudades, entidades de tránsito, sistemas de carreteras, semejantes en la base económica

aboadar@gmail.com

Índices con respecto al parque vehicular

• **Índices de Accidentalidad (No. De accidentes), morbilidad (No. De heridos) y mortalidad (No. De muertos), con respecto al número de vehículos registrados en el año respectivo, expresado por cada 10.000 vehículos.**

Índice de accidentalidad ($I_{A/P}$) = (No. De accidentes en el año X 10000)/No. De vehículos

Índice de morbilidad ($I_{morb/P}$) = (No. De heridos en el año X 10000)/No. No. De vehículos

Índice de mortalidad ($I_{mort/P}$) = (No. De muertos en el año X 10000)/No. No. De vehículos

Útiles para comparar ciudades, entidades de tránsito, sistemas de carreteras, aunque exista diferente base económica

aboadar@gmail.com

Otros índices de accidentes

Índices de Accidentalidad con respecto al kilometraje de viaje ($I_{A/K}$). Es el número de accidentes por un millón de vehículos-kilómetros de viaje, se expresa como:

Índice de accidentalidad ($I_{A/K}$) = (No. De accidentes en el año X 100000)/VK

Donde VK representa el número de vehículos- kilómetros de viaje al año y es igual a:

$$VK = TDP(365)(L)$$

TPD es el transito promedio diario y L es la longitud del viaje (tramo determinado de una vía). El valor de VK también se puede determinar multiplicando el consumo anual de combustible por el rendimiento promedio.

aboadar@gmail.com

Otros índices de accidentes

Índices de Accidentalidad con respecto al número de vehículos que entran a una intersección ($I_{A/VEI}$). Es el número de accidentes por un millón de vehículos que entran a una intersección, se expresa como:

$I_{A/VEI}$ = (No. De accidentes en el año X 1000000)/V

Donde V representa el número de vehículos que entran a la intersección en un año

$$V = TDP(365)$$

Este índice se utiliza para medir las tasas de accidentes en intersecciones, y así con base en un índice de accidentalidad definido como peligroso, se pueden determinar los puntos críticos de accidentalidad de la ciudad

aboadar@gmail.com

Otros índices de accidentes

Índices de severidad en intersecciones (IS): Este índice tiene en cuenta la gravedad de los accidentes en términos de daños materiales, heridos y muertos, con respecto al número de vehículos que entran a la intersección.

IS = (NAD_E. X 1000000)/(TPD(365))

Donde NAD_E es el número de accidentes por daños materiales, heridos y muertos, equivalentes en daños materiales. Esto es:

NAD_E. = NAD + NAH(F_1) + NAM(F_2)

Donde:

NAD = Número de accidentes con daños materiales solamente

NAH = Número de accidentes con heridos

NAM = Número de accidentes con muertos

F_1 = (Costo de AH) / (Costo de AD) ; F_2 = (Costo de AM) / (Costo de AD)

aboadar@gmail.com

Taller No. 2

(http://www.ccb.org.co/contenido/contenido.aspx?catID=127&conID=6589)

1. **Bogotá en 2009 según el DANE tenia una población estimada de 7.259.597 habitantes y 900.000 vehículos. Durante dicho año la cámara de comercio de Bogotá reporto en un estudio que se presentaron 31.562 accidentes de transito con 9.117 heridos y 528 muertos. Determine e interprete los índices de accidentalidad, morbilidad y mortalidad teniendo como base la población y el parque vehicular.**
2. **Para una intersección de alta accidentalidad, en un año en particular, se reporta la siguiente información: 10.500 vehículos como transito promedio diario servido en toda la intersección. 40 accidentes: 32 con daños materiales, 6 con heridos y 2 con muertos. Se estima que los costos de los accidentes con respecto al de los daños materiales son: 1.5 veces con heridos y 8.0 veces con muertos. Determine el índice de severidad de la intersección.**

aboadar@gmail.com

!Gracias por su atención!

aboadar@gmail.com

Descripción de los datos: Distribuciones de frecuencias y representaciones gráficas

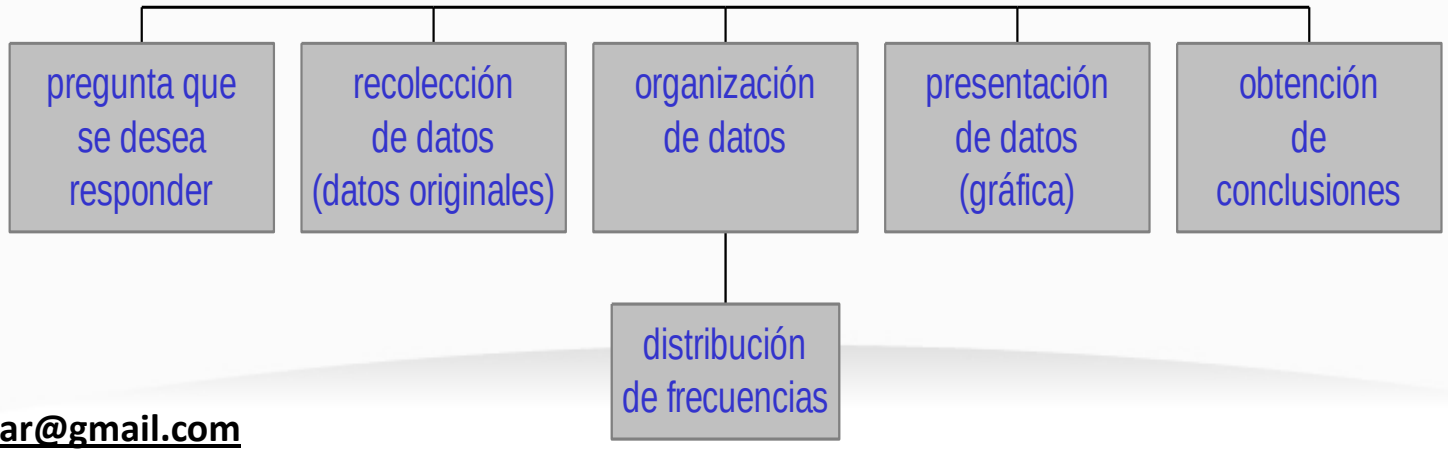
- ✓ Organizar los datos en una distribución de frecuencias.
- ✓ Presentar una distribución de frecuencias en un histograma, un polígono de frecuencias y un polígono de frecuencias acumuladas (ojiva).
- ✓ Desarrollar una representación de tallo y hoja.
- ✓ Presentar datos mediante gráficas de líneas, de barras, bastones, circulares.

aboadar@gmail.com

Distribución de frecuencias

- Distribución de frecuencias: agrupamiento de datos en categorías que muestran el número de observaciones en cada categoría mutuamente excluyente.

Elaboración de una tabla de distribución de frecuencias

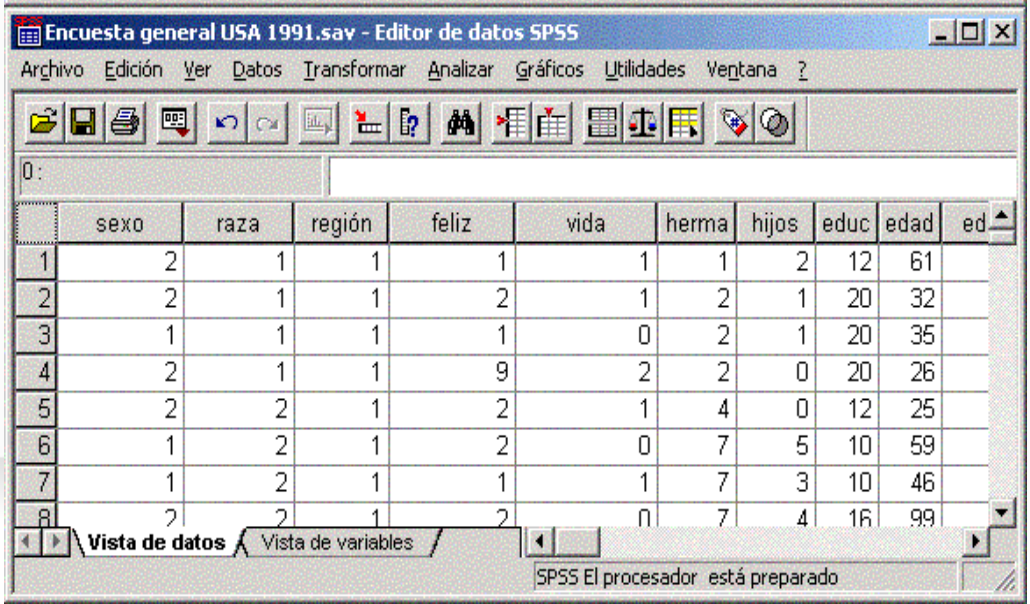


aboadar@gmail.com

Elaboración de tablas de distribución de frecuencias

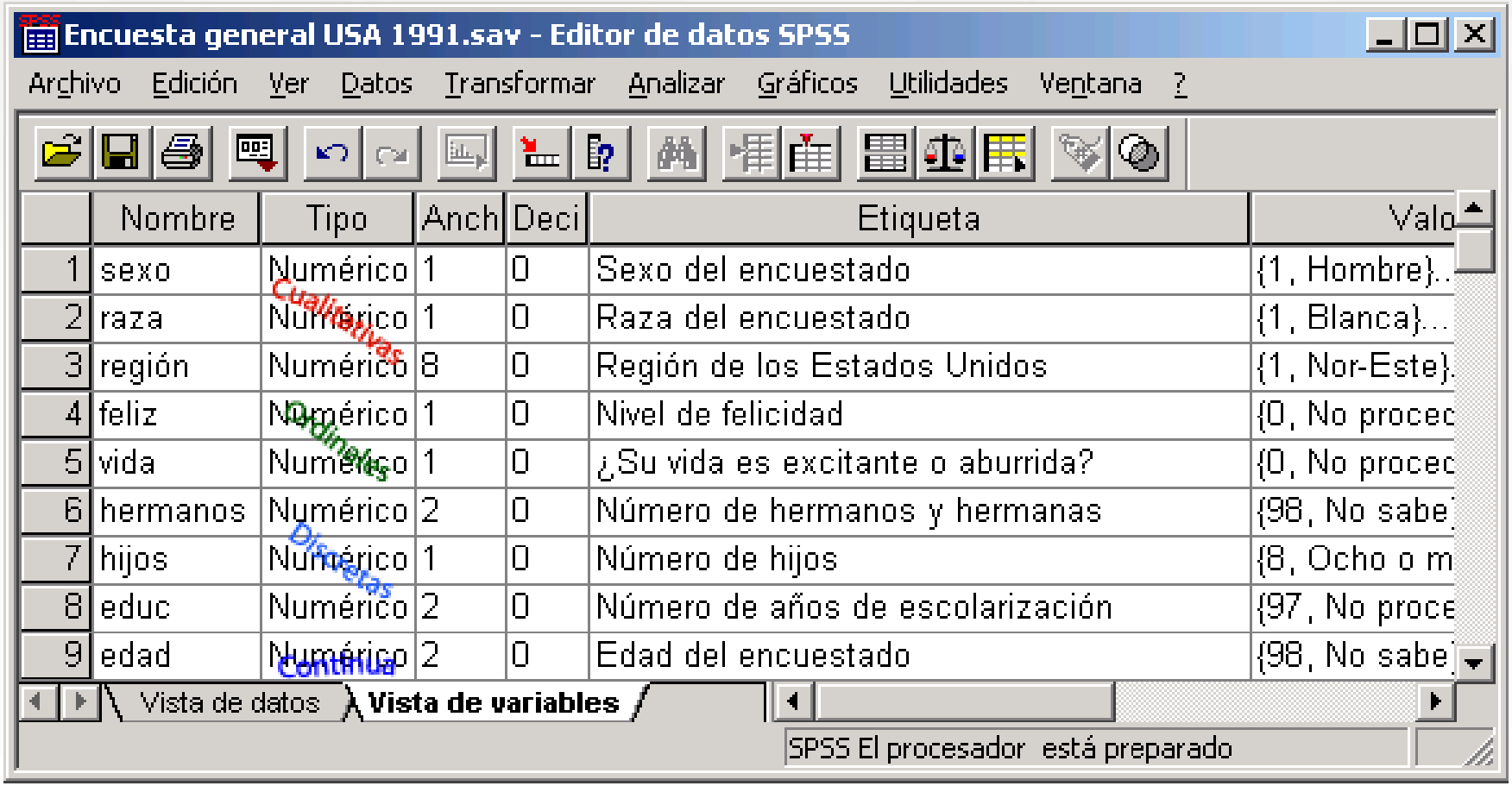
- Es buena idea codificar las variables como números para poder procesarlas con facilidad en un computador.
- Es conveniente asignar "etiquetas" a los valores de las variables para recordar qué significan los códigos numéricos.
  - Genero (Cualitativa: Códigos arbitrarios)
    - 1 = Hombre
    - 2 = Mujer
  - Raza (Cualitativa: Códigos arbitrarios)
    - 1 = Blanca
    - 2 = Negra,…
  - Felicidad Escala Ordinal: Respetar un orden al codificar.
    - 1 = Muy feliz
    - 2 = Bastante feliz
    - 3 = No demasiado feliz
- Se pueden asignar códigos a respuestas especiales como
    - 0 = No sabe
    - 99 = No contesta…
- Estas situaciones deberán ser tenidas en cuentas en el análisis. Datos perdidos ('missing data')

aboadar@gmail.com



Elaboración de tablas de distribución de frecuencias

- Aunque se codifiquen como números, se debe recordar siempre el verdadero tipo de las variables y su significado cuando se vaya a usar programas de cálculo estadístico.
- Atención…**N**o todo está permitido con cualquier tipo de variable.



Elaboración de tablas de distribución de frecuencias

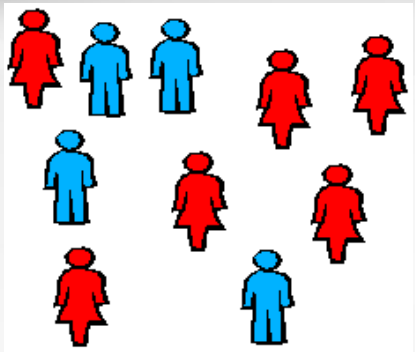
- Los posibles valores de una variable suelen denominarse modalidades.
- Las modalidades pueden agruparse en clases (intervalos)
  - Edades:
    - Menos de 20 años, de 20 a 50 años, más de 50 años
  - Velocidades:
    - Menos de 30 km/h, De 30 a 60 km/h, 60 o más km/h

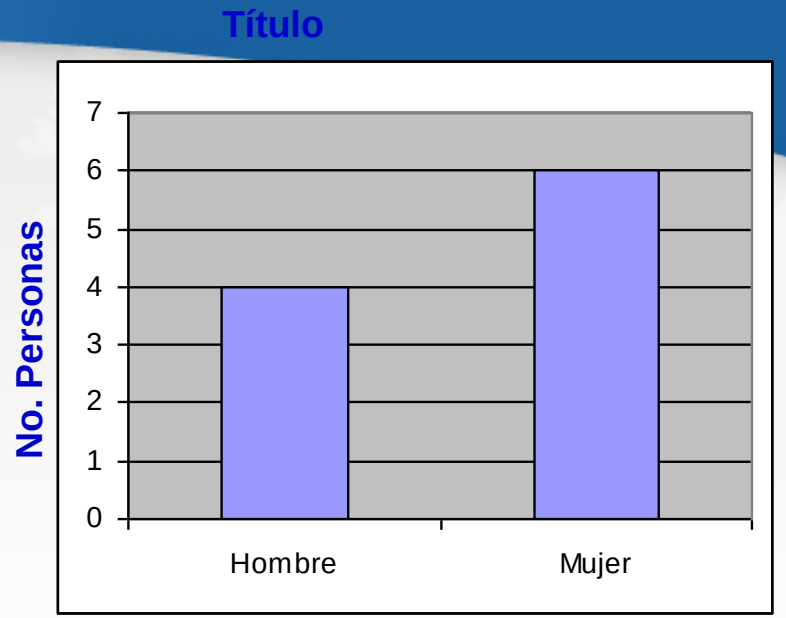
- Las modalidades/clases deben formar un sistema exhaustivo y excluyente
  - Exhaustivo: No se puede olvidar ningún posible valor de la variable
    - **Incorrecto: ¿Cuál es su música preferida: (Reggeton, RocK)**
    - Bien: ¿Cuál es su grupo sanguíneo?
  - Excluyente: Nadie puede presentar dos valores simultáneos de la variable
    - Estudio sobre el ocio
      - **Incorrecto** : De los siguientes, qué le gusta: (deporte, cine)
      - Bien: Le gusta el deporte: (Sí, No)
      - Bien: Le gusta el cine: (Sí, No)
      - **Incorrecto** : Cuántos hijos tiene: (Ninguno, Menos de 5, Más de 2)

aboadar@gmail.com

Presentación ordenada de datos

Género	Frec.
Hombre	4
Mujer	6





Fuente: …….

- Las tablas de frecuencias y las representaciones gráficas son dos maneras ***equivalentes*** de presentar la información.
- Las dos exponen ordenadamente la información de una población y/o muestra**.**

aboadar@gmail.com

Tablas de distribución de frecuencias

- Exponen la información recogida en la muestra, de forma que no se pierda nada de información (o poca).
  - Marca de clase (punto medio): punto que divide a la clase en dos partes iguales. Es el promedio entre los límites superior e inferior de la clase.
  - Intervalo de clase: para una distribución de frecuencias que tiene clases del mismo tamaño, el intervalo de clase se obtiene restando el límite inferior de una clase del límite inferior de la siguiente.
  - Frecuencias absolutas: Contabilizan el número de individuos de cada modalidad
  - Frecuencias relativas (porcentajes): Contabilizan el número de individuos de cada modalidad, pero dividido por el total
  - Frecuencias acumuladas: Sólo tienen sentido para variables ordinales y numéricas
    - Muy útiles para calcular cuantiles (ver más adelante)
      - ¿Qué porcentaje de individuos tiene 3 o menos hijos? Sol: 83,8
      - ¿Entre 4 y 6 hijos? Soluc 1ª: 8,4%+3,6%+1,6%= **13,6%**. Soluc 2ª: 97,3% - 83,8% = **13,5%**

Sexo del encuestado

		Frecuencia	Porcentaje	Porcentaje válido
Válidos	Hombre	636	41,9	41,9
	Mujer	881	58,1	58,1
	Total	1517	100,0	100,0

Nivel de felicidad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Muy feliz	467	30,8	31,1	31,1
	Bastante feliz	872	57,5	58,0	89,0
	No demasiado feliz	165	10,9	11,0	100,0
	Total	1504	99,1	100,0	
Perdidos	No contesta	13	,9		
Total		1517	100,0		

Número de hijos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	419	27,6	27,8	27,8
	1	255	16,8	16,9	44,7
	2	375	24,7	24,9	69,5
	3	215	14,2	14,2	83,8
	4	127	8,4	8,4	92,2
	5	54	3,6	3,6	95,8
	6	24	1,6	1,6	97,3
	7	23	1,5	1,5	98,9
	Ocho o más	17	1,1	1,1	100,0
	Total	1509	99,5	100,0	
Perdidos	No contesta	8	,5		
Total		1517	100,0		

Datos desordenados y ordenados en tablas

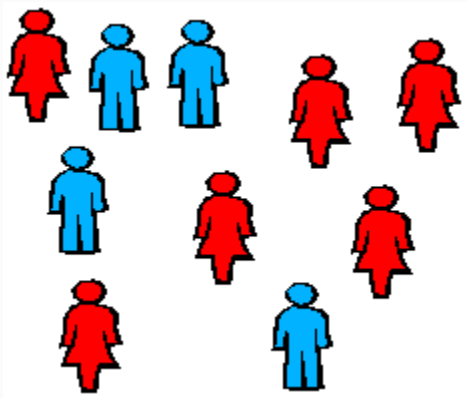
- Variable: Género
  - Modalidades:
    - H = Hombre
    - M = Mujer

Género	Frecuencia	Frecuencia relativa (%)
Hombre	4	4/10=0,4=40%
Mujer	6	6/10=0,6=60%
	10	

- Muestra:

M H H M M H M M M H

  - equivale a
    HHHH MMMMMM



aboadar@gmail.com

Aplicaciones: Tablas de distribución de frecuencias

- ¿Cuántos individuos tienen menos de 2 hijos?
  - Frecuencia absoluta sin hijos
    +
    frecuencia absoluta con 1 hijo
    = 419 + 255
    = 674 individuos

- ¿Qué porcentaje de individuos tiene 6 hijos o menos?
  - 97,3%

- ¿Qué cantidad de hijos es tal que al menos el 50% de la población tiene una cantidad inferior o igual?
  - 2 hijos

Número de hijos

	Frec.	Porcent. (válido)	Porcent. acum.
0	419	27,8	27,8
1	255	16,9	44,7
2	375	24,9	69,5 ≥50%
3	215	14,2	83,8
4	127	8,4	92,2
5	54	3,6	95,8
6	24	1,6	97,3
7	23	1,5	98,9
Ocho+	17	1,1	100,0
Total	1509	100,0	

aboadar@gmail.com

EJEMPLO 1

- Zoila Pérez Sosa desea determinar cuánto estudian los alumnos en cierto curso de Estadística. Selecciona una muestra aleatoria de 30 estudiantes y determina el número de horas por semana que estudia cada uno: 15.0, 23.7, 19.7, 15.4, 18.3, 23.0, 14.2, 20.8, 13.5, 20.7, 17.4, 18.6, 12.9, 20.3, 13.7, 21.4, 18.3, 29.8, 17.1, 18.9, 10.3, 26.1, 15.7, 14.0, 17.8, 33.8, 23.2, 12.9, 27.1, 16.6.
- Organice los datos en una tabla de distribución de frecuencias para datos continuos.

aboadar@gmail.com

EJEMPLO 1 *continuación*

Considere las clases 8-12 y 13-17. Las marcas medias de clase son 10 y 15. El intervalo de marca de clase es 5 (17 – 13 y/o 17-12).

Horas de estudio	Frecuencia: n_j
8-12	1
13-17	12
18-22	10
23-27	5
28-32	1
33-37	1

aboadar@gmail.com

Sugerencias para elaborar una distribución de frecuencias

- Se sugiere que los intervalos de clase usados en la distribución de frecuencias sean iguales.
- Para determinar el intervalo de clase use:

$C_j = R / m$ *Donde: R = Rango; m = Rango/número de marcas de clases*

- Formulas alternas

$m = 1 + 3.322(\log_{10} n)$

$m = \sqrt{n}$

$2^m \geq n$

aboadar@gmail.com

Tabla de Distribución de frecuencias

Y'_{j-1}	Y'_j	n_j	h_j
8	12	1	1/30=.0333
13	17	12	12/30=.400
18	22	10	10/30=.333
23	27	5	5/30=.1667
28	32	1	1/30=.0333
33	37	1	1/30=.0333
TOTAL		30	30/30=1

La frecuencia relativa de una clase se obtiene dividiendo la frecuencia de clase entre la frecuencia total.

aboadar@gmail.com

Representaciones de tallo y hoja

- Representaciones de tallo y hoja: Técnica estadística para representar un conjunto de datos. Cada valor numérico se divide en dos partes: los dígitos principales son el tallo y el dígito siguiente es la hoja.
- *Nota*: una ventaja de la representación de tallo y hoja comparado con la distribución de frecuencias es que no se pierde la identidad de cada observación.

aboadar@gmail.com

EJEMPLO 2

- Emma Madera de Gallo registró las siguientes velocidades km/h, en la vía Bogotá – Villeta: 86, 79, 92, 84, 69, 88, 91, 83, 96, 78, 82, 85. Construya una representación de tallo y hoja para los datos.

tallo	hoja
6	9
7	8 9
8	2 3 4 5 6 8
9	1 2 6

aboadar@gmail.com

Presentación gráfica de una distribución de frecuencias

- Las tres formas de gráficas más usadas son histogramas, polígonos de frecuencia y distribuciones de frecuencias acumuladas (ojiva).
- Histograma: gráfica donde las clases se marcan en el eje horizontal y las frecuencias de clase en el eje vertical. Las frecuencias de clase se representan por las alturas de las barras y éstas se trazan adyacentes entre sí.

aboadar@gmail.com

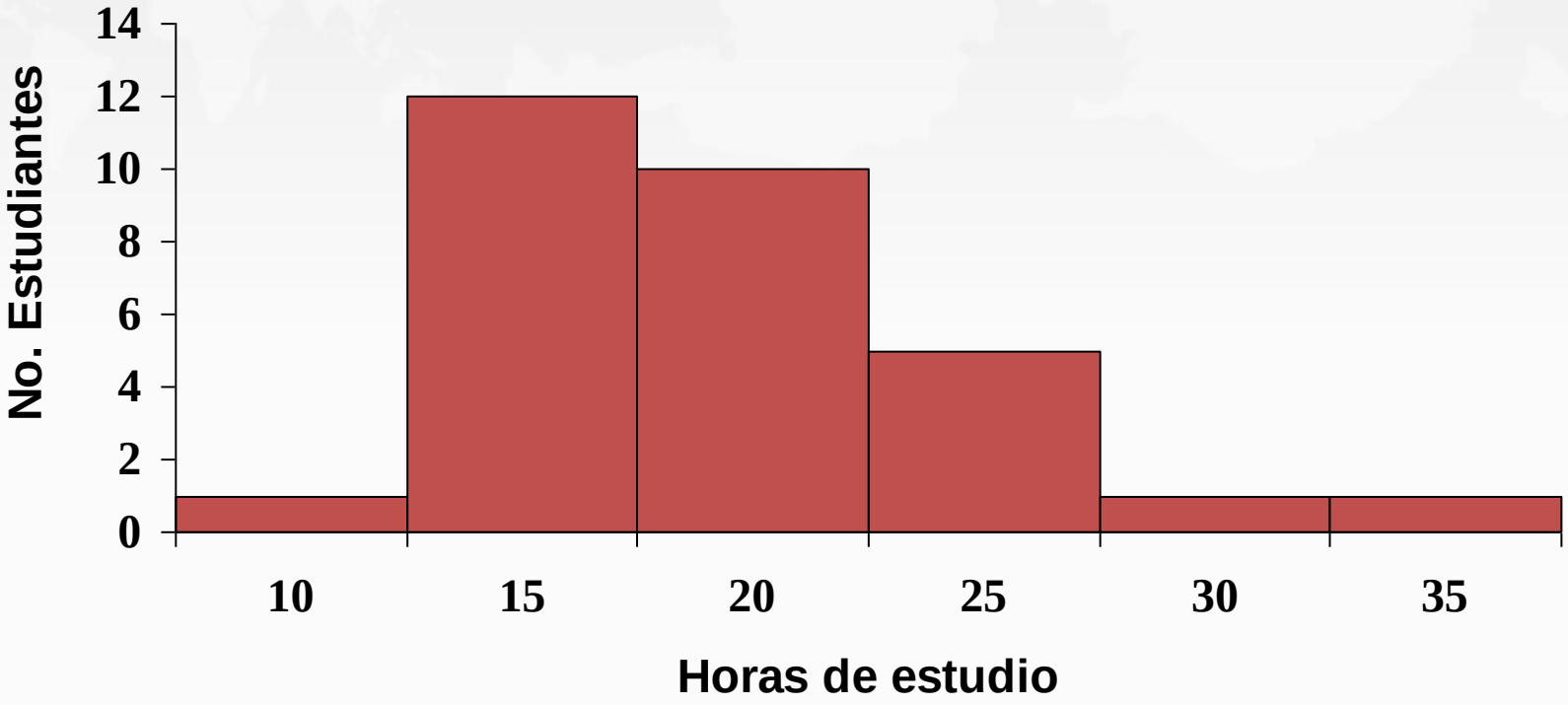
Presentación gráfica de una distribución de frecuencias

- Un polígono de frecuencias consiste en segmentos de línea que conectan los puntos formados por el punto medio de la clase y la frecuencia de clase.
- Una distribución de frecuencias acumulada (ojiva) se usa para determinar cuántos o qué proporción de los valores de los datos es menor o mayor que cierto valor.

aboadar@gmail.com

Histograma para las horas de estudio

Distribución de las horas de estudio de un grupo de estudiantes

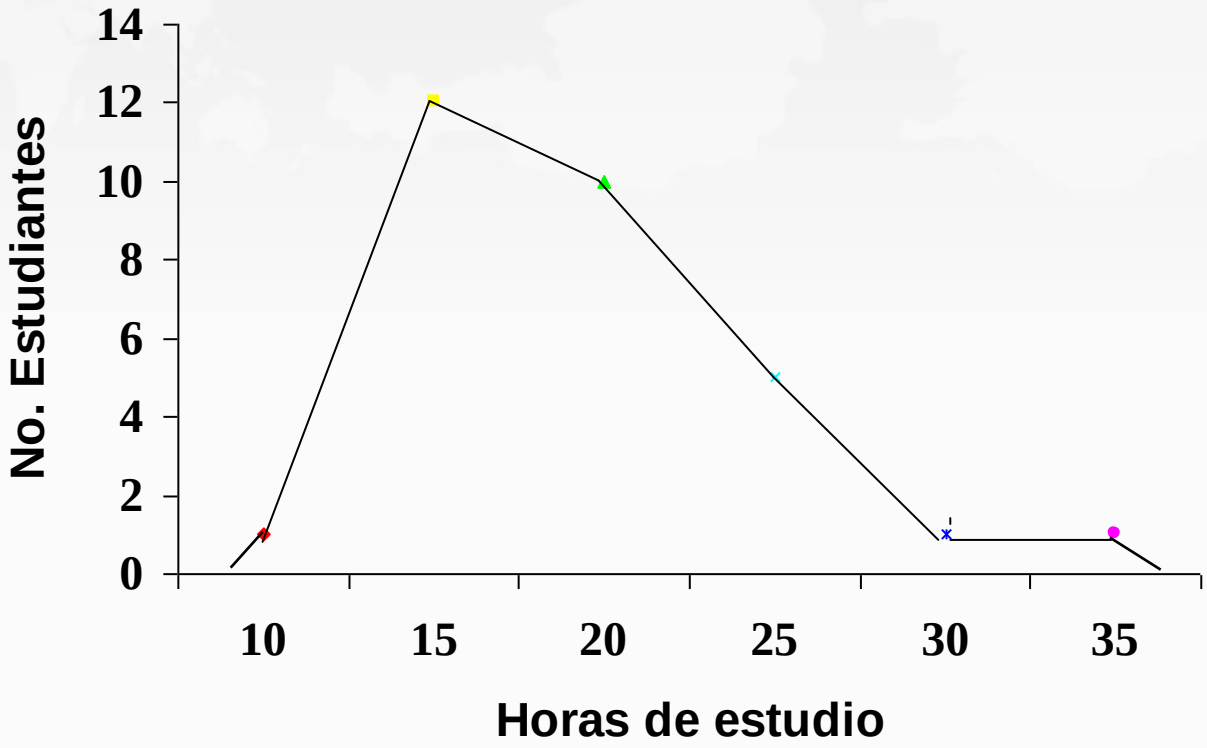


Fuente: Estudiantes Escuela de Ingeniería I-2010

aboadar@gmail.com

Polígono de frecuencias para las horas de estudio

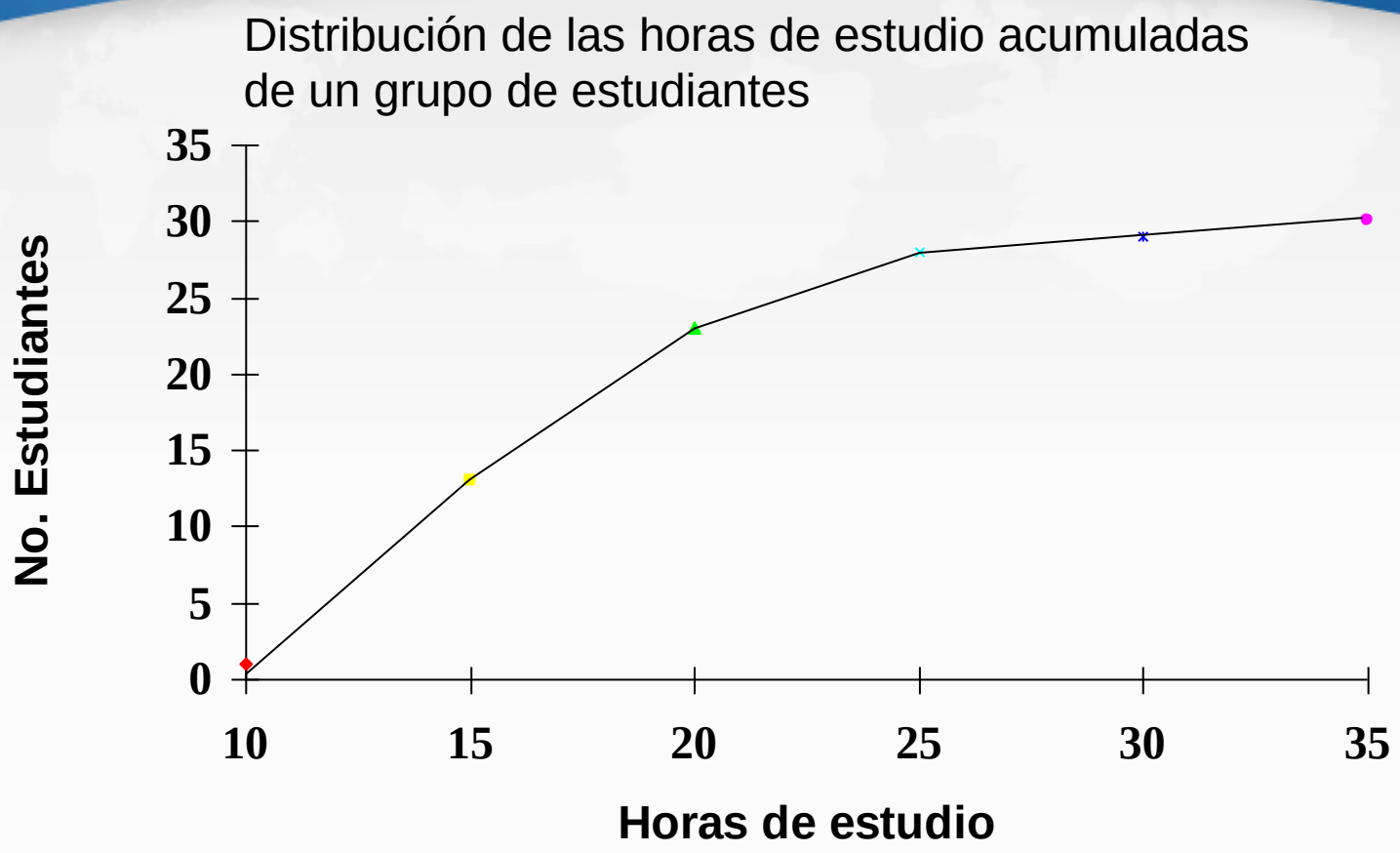
Distribución de las horas de estudio de un grupo de estudiantes



Fuente: Estudiantes Escuela de Ingeniería I-2010

aboadar@gmail.com

Distribución de frecuencias acumuladas menor que para las horas de estudio



Fuente: Estudiantes Escuela de Ingeniería I-2010

aboadar@gmail.com

Gráfica de barras

- Una gráfica de barras se puede usar para describir cualquier nivel de medición (nominal, ordinal, de intervalo o de razón).
- EJEMPLO 3: construya una gráfica de barras para el número de personas desempleadas por cada 100.000 habitantes de ciertas ciudades en 2009.

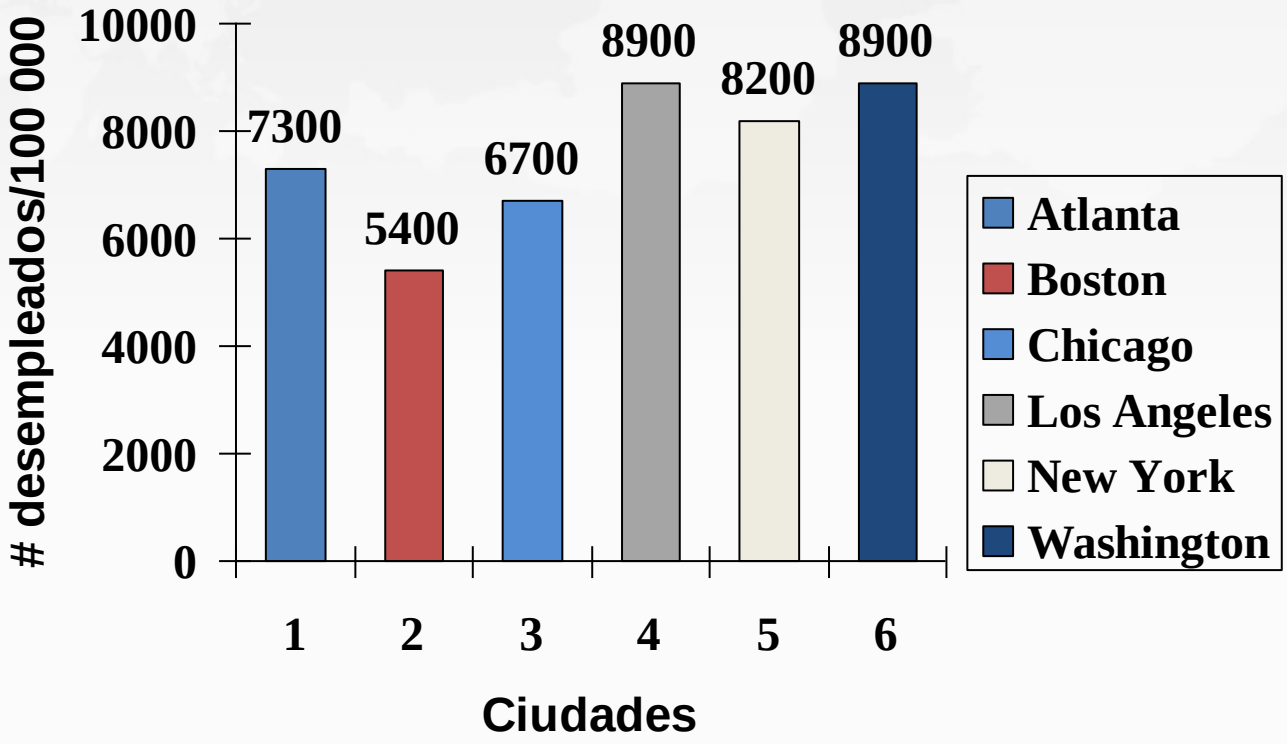
aboadar@gmail.com

EJEMPLO 3 *continuación*

Ciudad	Número de desempleados por 100 000 habitantes
Atlanta, GA	7300
Boston, MA	5400
Chicago, IL	6700
Los Angeles, CA	8900
New York, NY	8200
Washington, D.C.	8900

aboadar@gmail.com

Gráfica de barras para los datos de desempleados



aboadar@gmail.com

Gráfica circular

- Una gráfica circular es en especial útil para desplegar una distribución de frecuencias relativas. Se divide un círculo de manera proporcional a la frecuencia relativa y las rebanadas representan los diferentes grupos.
- EJEMPLO 4: se pidió a una muestra de 200 corredores que indicaran su tipo favorito de zapatos para correr.

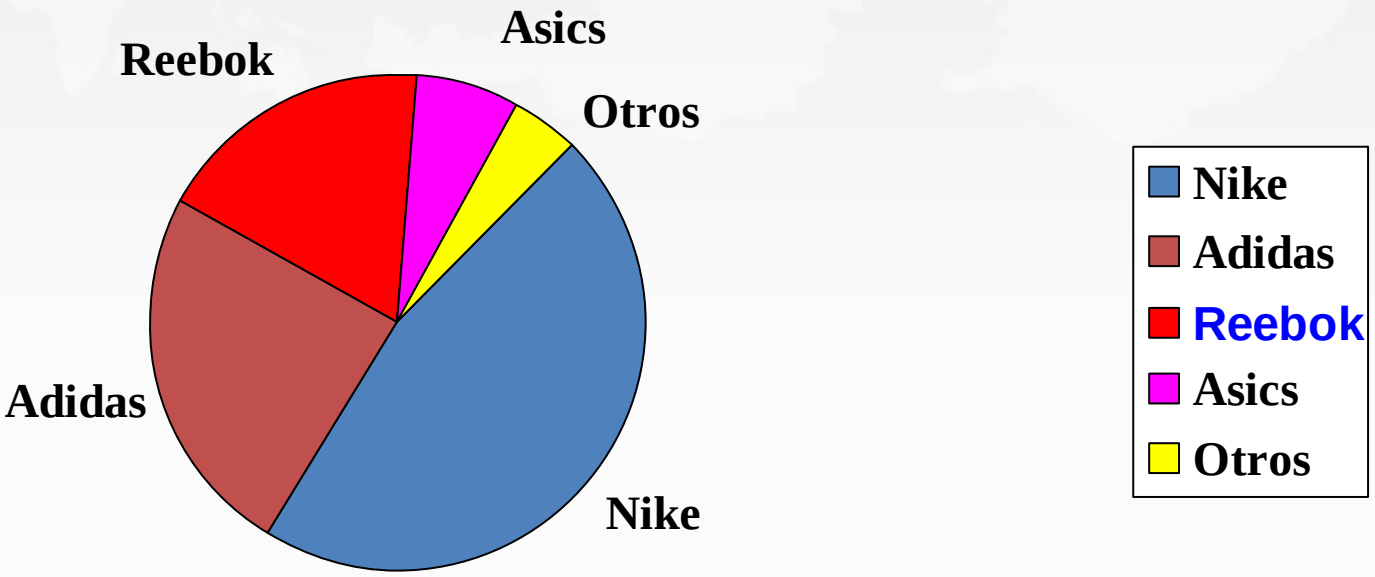
aboadar@gmail.com

EJEMPLO 4 *continuación*

- Dibuje una gráfica circular basada en la siguiente información.

Tipo de zapato	# de corredores
Nike	92
Adidas	49
Reebok	37
Asics	13
Otros	9

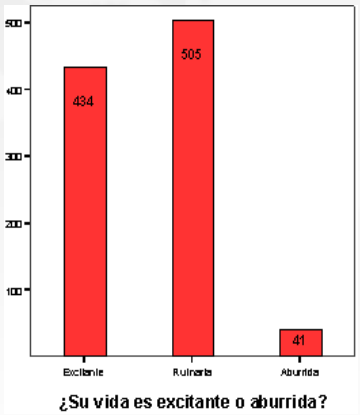
Gráfica circular para tipos de zapatos



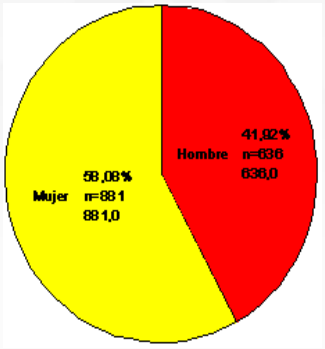
aboadar@gmail.com

Gráficos para variables cualitativas

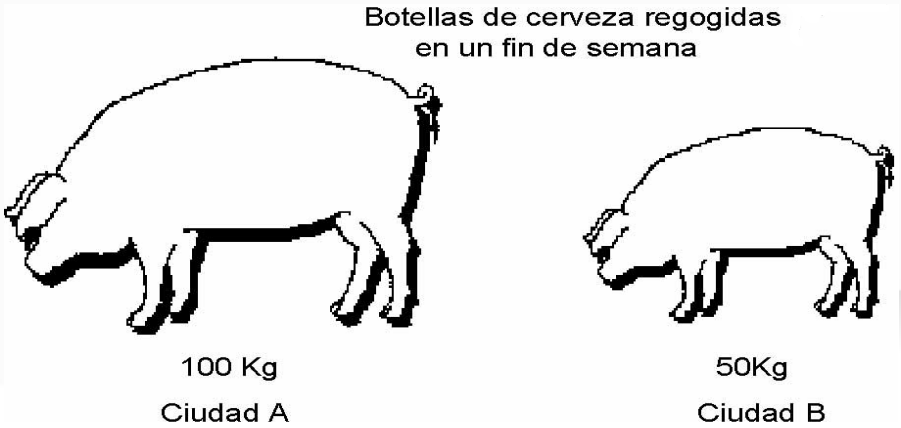
- Diagramas de barras
  - Alturas proporcionales a las frecuencias (abs. o rel.)
  - Se pueden aplicar también a variables discretas

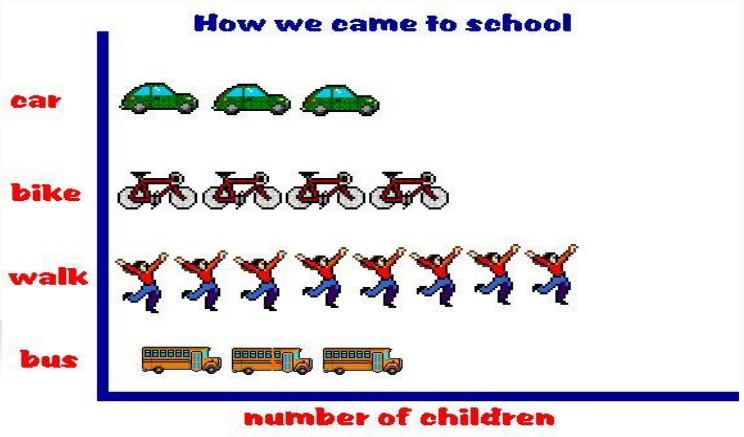


- Diagramas de sectores (tortas, polares)
  - No usarlo con variables ordinales.
  - El área de cada sector es proporcional a su frecuencia (abs. o rel.)



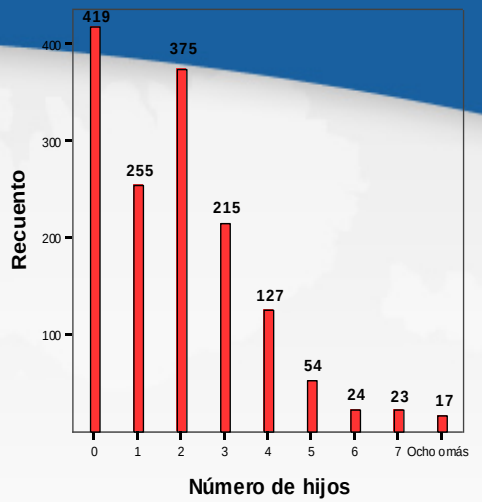
- Pictogramas
  - Fáciles de entender.
  - El área de cada modalidad debe ser proporcional a la frecuencia. ¿De los dos, cuál es incorrecto?.

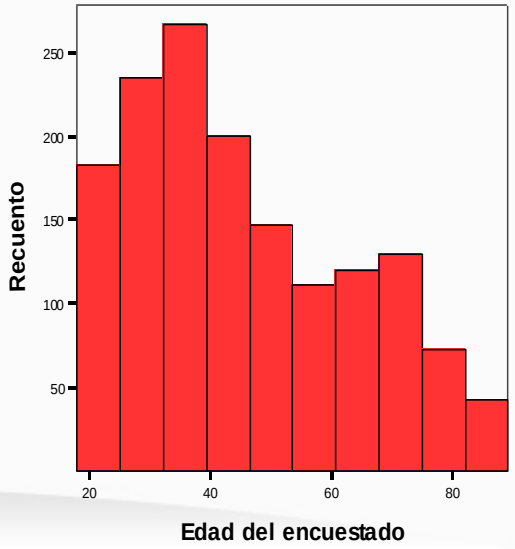




Gráficos diferenciales para variables numéricas

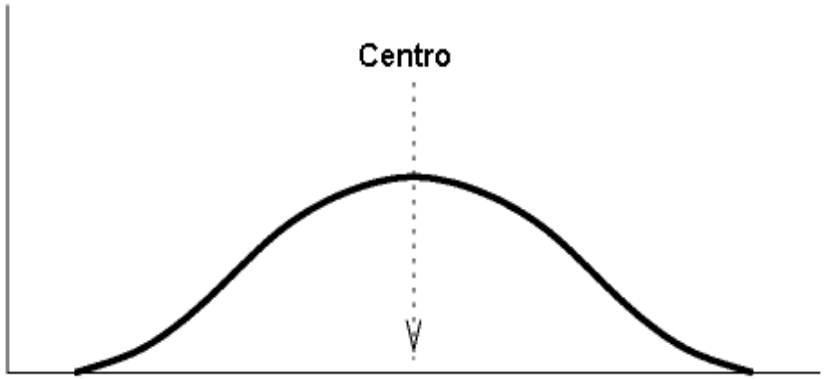
- Son diferentes en función de que las variables sean discretas o continuas. Se elaboran con frecuencias absolutas o relativas.
  - **Diagramas barras para variables discretas**
    - Se deja un hueco entre barras para indicar los valores que no son posibles
  - **Histogramas para variables continuas**
    - El área que hay bajo el histograma entre dos puntos cualesquiera indica la cantidad (porcentaje o frecuencia) de individuos en el intervalo.

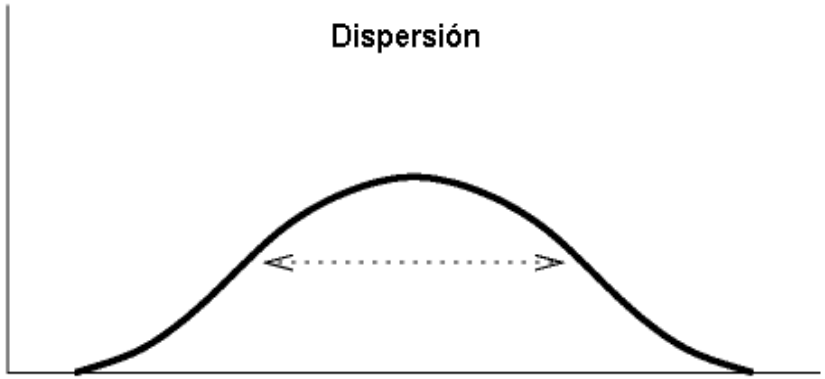


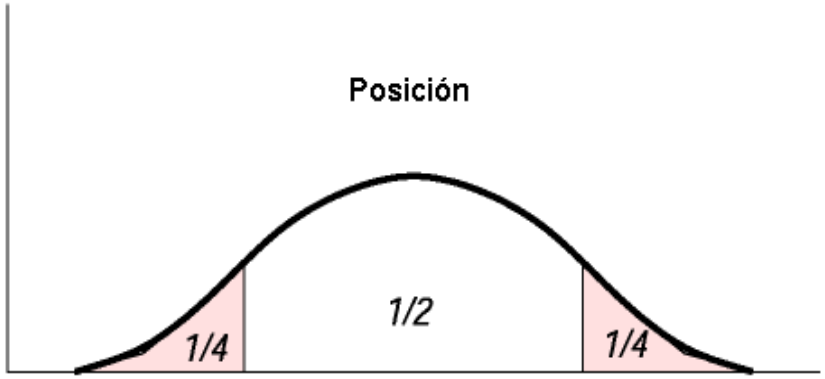


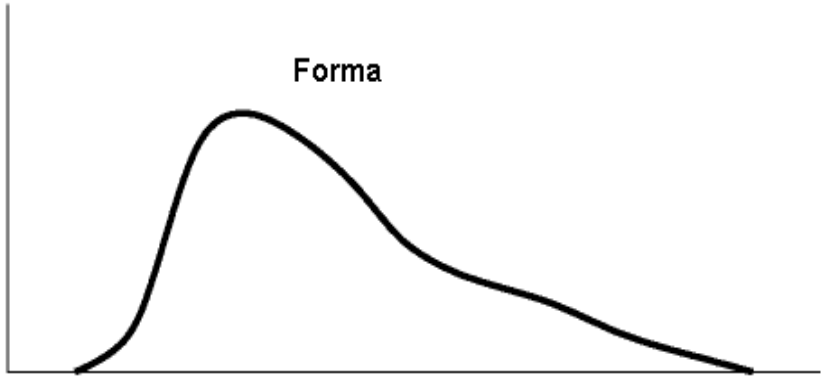
aboadar@gmail.com

Estadísticos en el Análisis Exploratorio de datos AED









aboadar@gmail.com

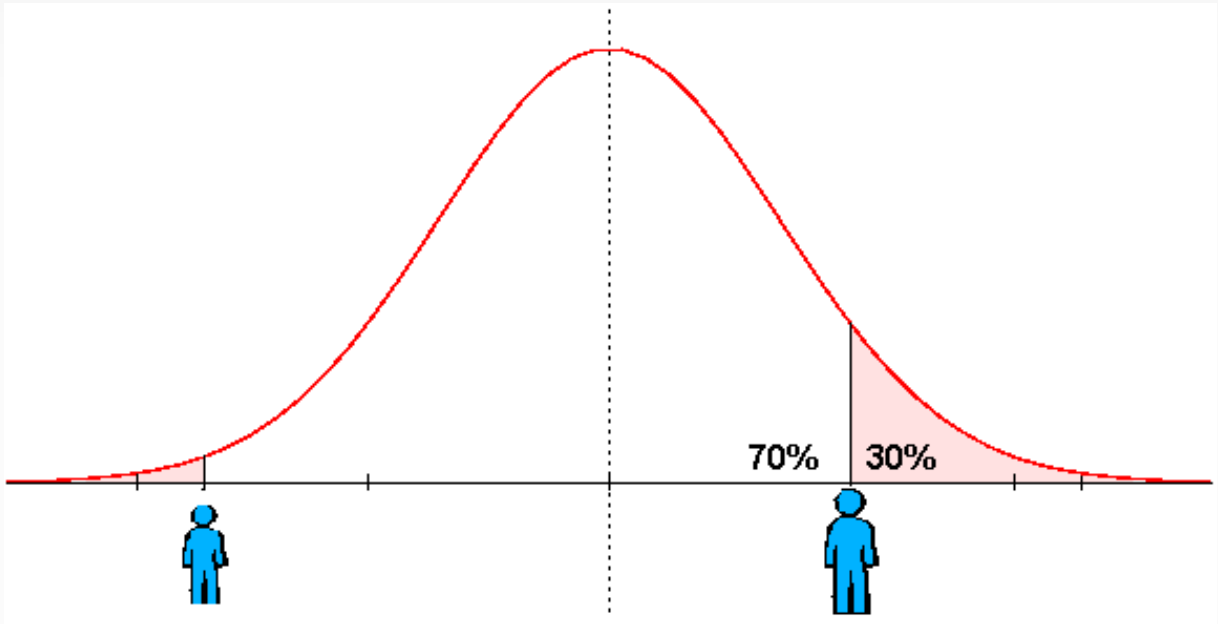
Estadísticos – Análisis Exploratorio de datos

- Posición
  - Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.
    - Cuantiles, percentiles, cuartiles, deciles,...
- Centralización
  - Indican valores con respecto a los que los datos parecen agruparse.
    - Media, mediana, moda, media armónica, media geométrica, etc.
- Dispersión
  - Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización.
    - Desviación típica, coeficiente de variación, rango, varianza
- Forma
  - Asimetría (Skewness)
  - Apuntamiento o curtosis

aboadar@gmail.com

Estadísticos de posición

- Se define el **cuantil** de orden α como un valor de la variable por debajo del cual se encuentra una frecuencia acumulada α.
- Casos particulares de un cuantil son los percentiles, cuartiles, deciles, quintiles,...



aboadar@gmail.com

Estadísticos de posición

- **Percentil** de orden k = cuantil de orden k/100
  - La **mediana** es el percentil 50.
  - El percentil de orden 15 deja por debajo al 15% de las observaciones. Por encima queda el 85%.

- **Cuartiles**: Dividen a la muestra en 4 grupos con frecuencias similares.
  - Primer cuartil = Percentil 25 = Cuantil 0,25.
  - Segundo cuartil = Percentil 50 = Cuantil 0,5 = **mediana.**
  - Tercer cuartil = Percentil 75 = cuantil 0,75.

aboadar@gmail.com

Estadísticos de posición

- ***Ejemplos:*** El 5% de los recién nacidos tiene un peso demasiado bajo. ¿Qué peso se considera "demasiado bajo"?
  - Percentil 5 o cuantil 0,05.
- ¿Qué peso es superado sólo por el 25% de los individuos?
  - Percentil 75.
- El colesterol se distribuye simétricamente en la población. Se considera patológico los valores extremos. El 90% de los individuos son normales. ¿Entre qué valores se encuentran los individuos normales?
  - Entre el percentil 5 y el 95.
- ¿Entre qué valores se encuentran la mitad de los individuos "más normales" de una población?
  - Entre 1º y 3º cuartil (Q_1 y Q_3).

Niveles de Hb en 61 adultos normales

105	110	112	112	118	119	120	120	120
125	126	127	128	130	132	133	134	135
138	138	138	138	141	142	144	145	146
148	148	148	149	150	150	150	151	151
153	153	154	154	154	154	155	156	156
158	160	160	160	163	164	164	165	166
168	168	170	172	172	176	179		

133.5

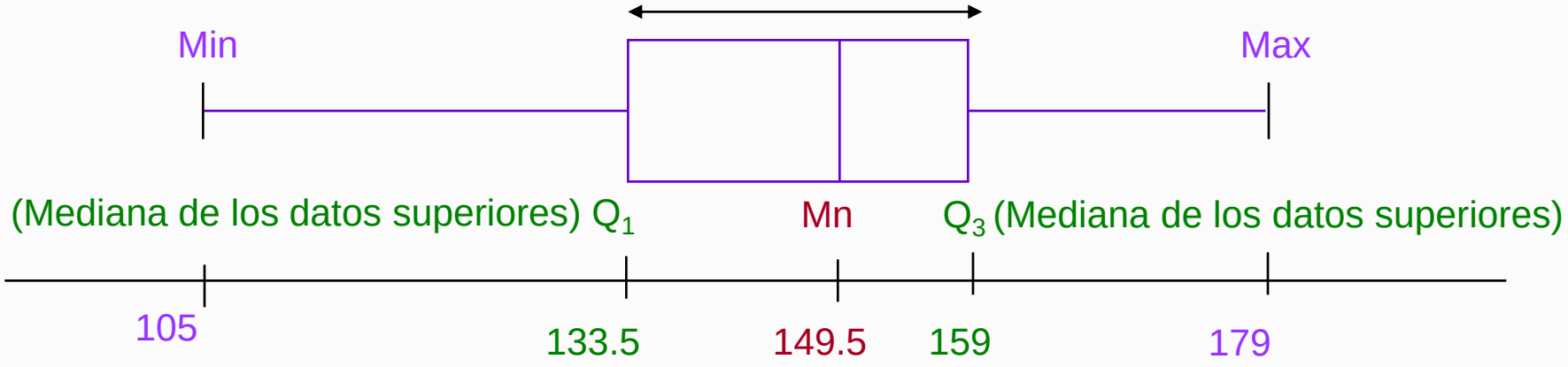
149.5

159

Un resumen de esta serie en 5 valores

Min = 105 ; Max =179; Q_1 = 133.5 ; Q_3 = 159 ; Q_2 = Mn = 149.5

Recorrido intercuartílico: IQR = $Q_3 - Q_1$



("Box-and-Whisker" plot)

aboadar@gmail.com

Medidas de tendencia central - Centralización

- Añaden unos cuantos casos particulares a las medidas de posición. Son medidas que buscan posiciones (valores) con respecto a los que los datos muestran tendencia a agruparse.

- Media (‘mean’) Es la media aritmética (promedio) de los valores de una variable. Suma de los valores dividido por el tamaño muestral.
  - Media de {2, 2, 3, 7} es (2+2+3+7)/4 = 3,5
  - Conveniente cuando los datos se concentran simétricamente con respecto a ese valor.
  - Muy sensible a valores extremos.
  - Centro de gravedad de los datos.

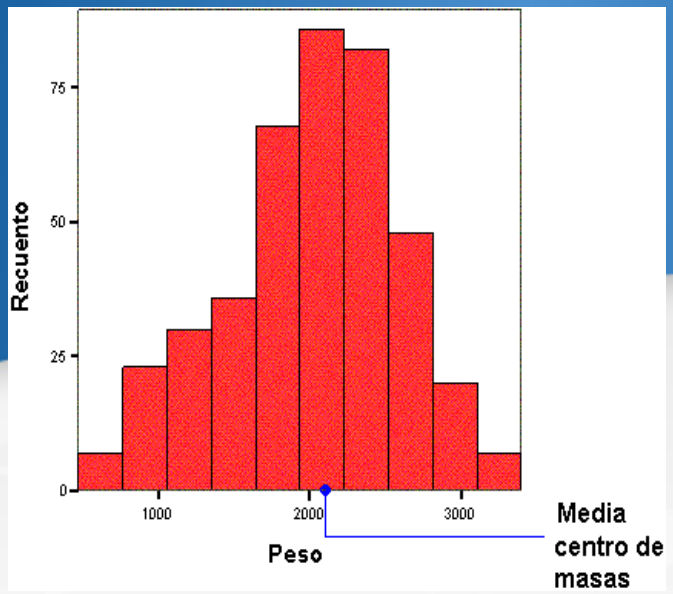
aboadar@gmail.com

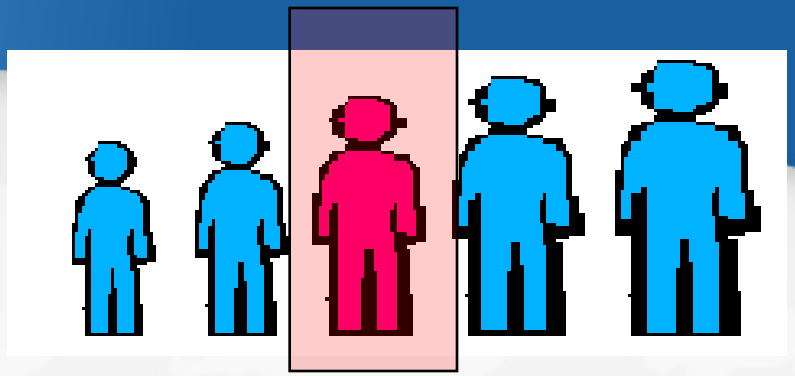
Medidas de tendencia central - Centralización

- Mediana (‘median’) Es un valor que divide a las observaciones en dos grupos con el mismo número de individuos (percentil 50). Si el número de datos es par, se elige la media de los dos datos centrales.
  - Mediana de 1, 2, 4, **5**, 6, 6, 8 es 5
  - Mediana de 1, 2, 4, **5**, **6**, 6, 8, 9 es (5+6)/2 = 5,5
  - Es conveniente cuando los datos son asimétricos. No es sensible a valores extremos.
    - Mediana de 1, 2, 4, **5**, 6, 6, 800 es 5. ¡La media es 117,7!

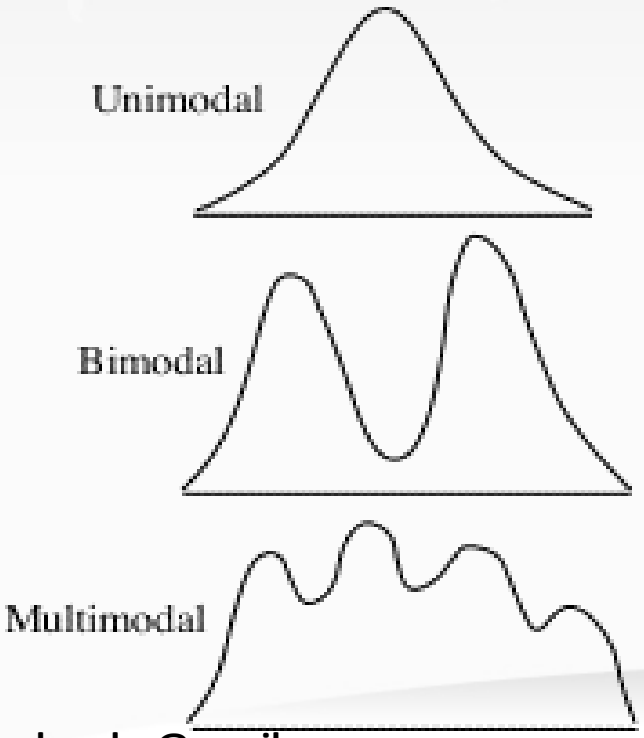
- Moda (‘mode’) Es el/los valor/es donde la distribución de frecuencia alcanza un máximo.

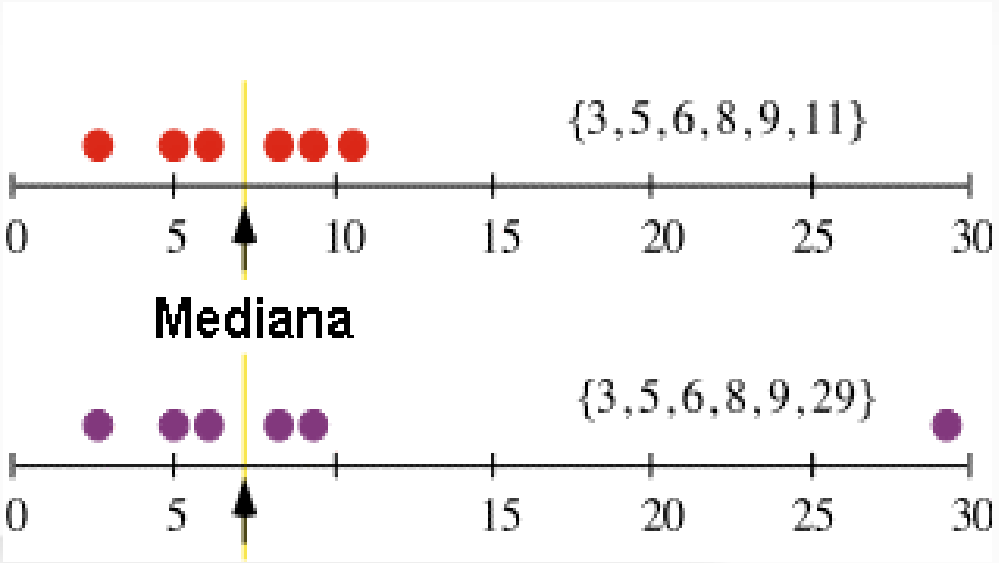
aboadar@gmail.com





Altura mediana

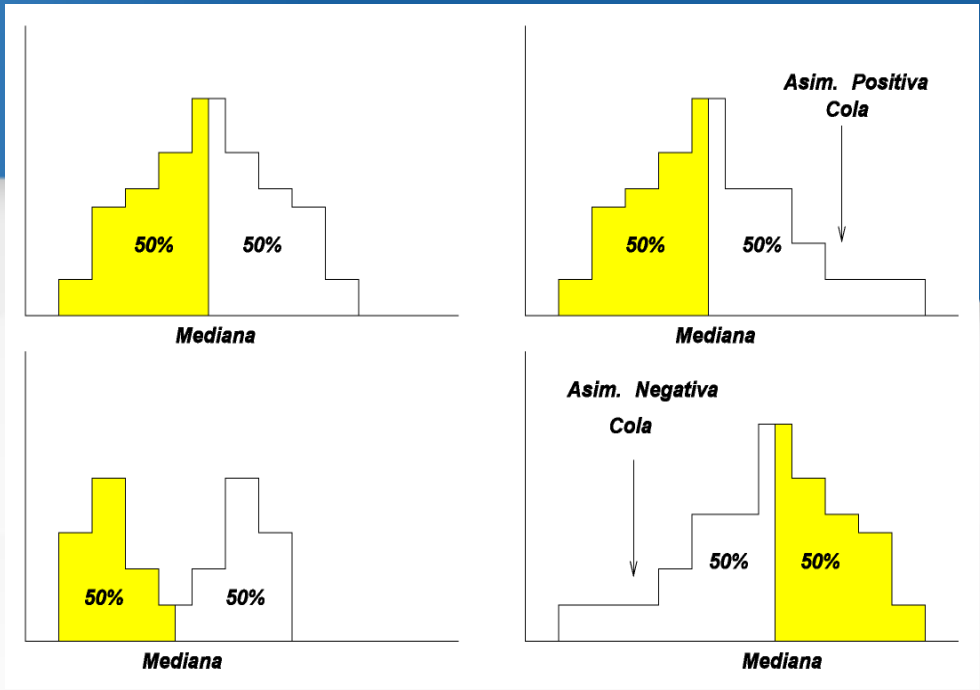


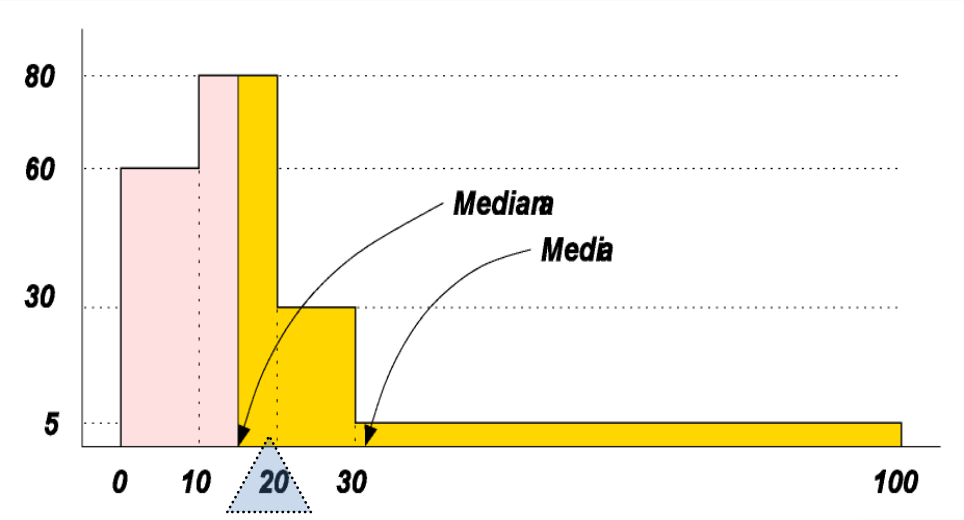


aboadar@gmail.com

Asimetría o sesgo

- Una distribución es simétrica si la mitad izquierda de su distribución es la imagen especular de su mitad derecha.
- En las distribuciones simétricas media y mediana coinciden. Si sólo hay una moda también coincide.
- La asimetría es positiva o negativa en función de a qué lado se encuentra la cola de la distribución.
- La media tiende a desplazarse hacia las valores extremos (colas).
- Las discrepancias entre las medidas de centralización son indicación de asimetría.



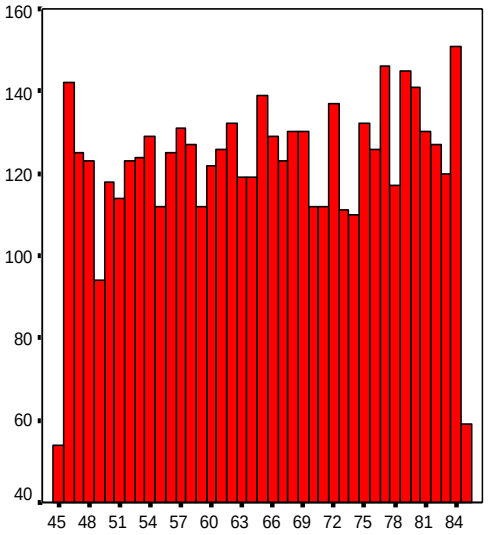


aboadar@gmail.com

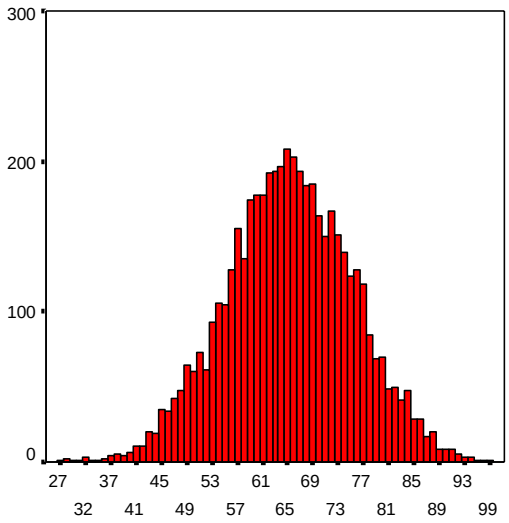
Apuntamiento o curtosis (kurtosis)

- La curtosis nos indica el grado de apuntamiento (aplastamiento) de una distribución con respecto a la distribución normal o gaussiana.
  Es adimensional.
- Platicúrtica: curtosis < 0
- Mesocúrtica: curtosis = 0
- Leptocúrtica: curtosis > 0

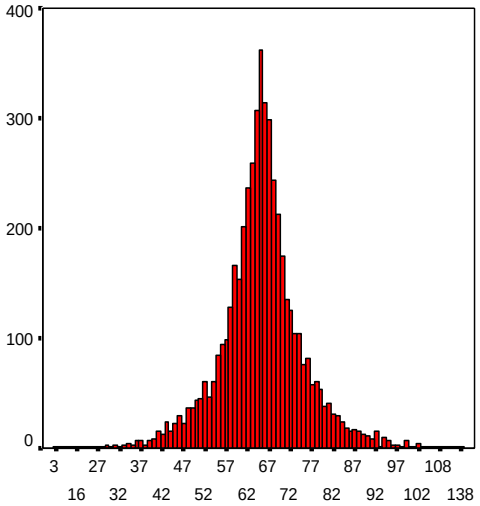
Los gráficos poseen la misma media y desviación típica, pero diferente grado de apuntamiento o curtosis.



Platicúrtica



Mesocúrtica

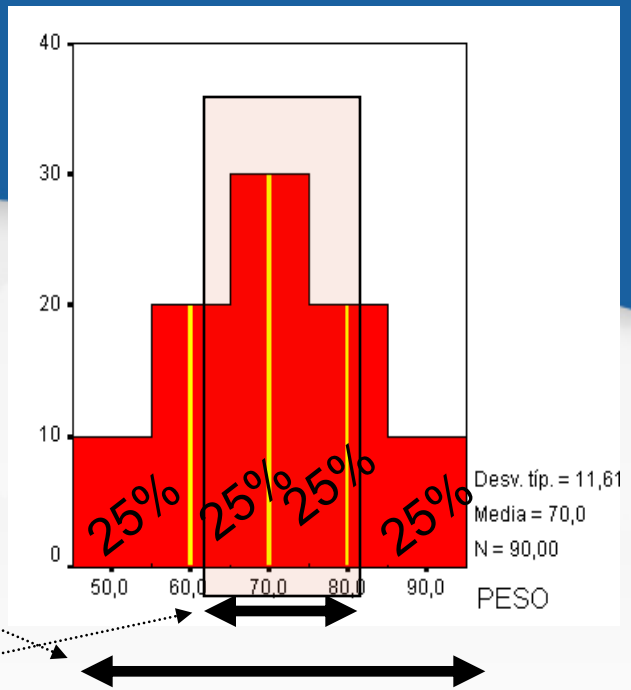


Leptocúrtica

aboadar@gmail.com

Medidas de dispersión

• Miden el grado de dispersión (variabilidad) de los datos, independientemente de su causa.



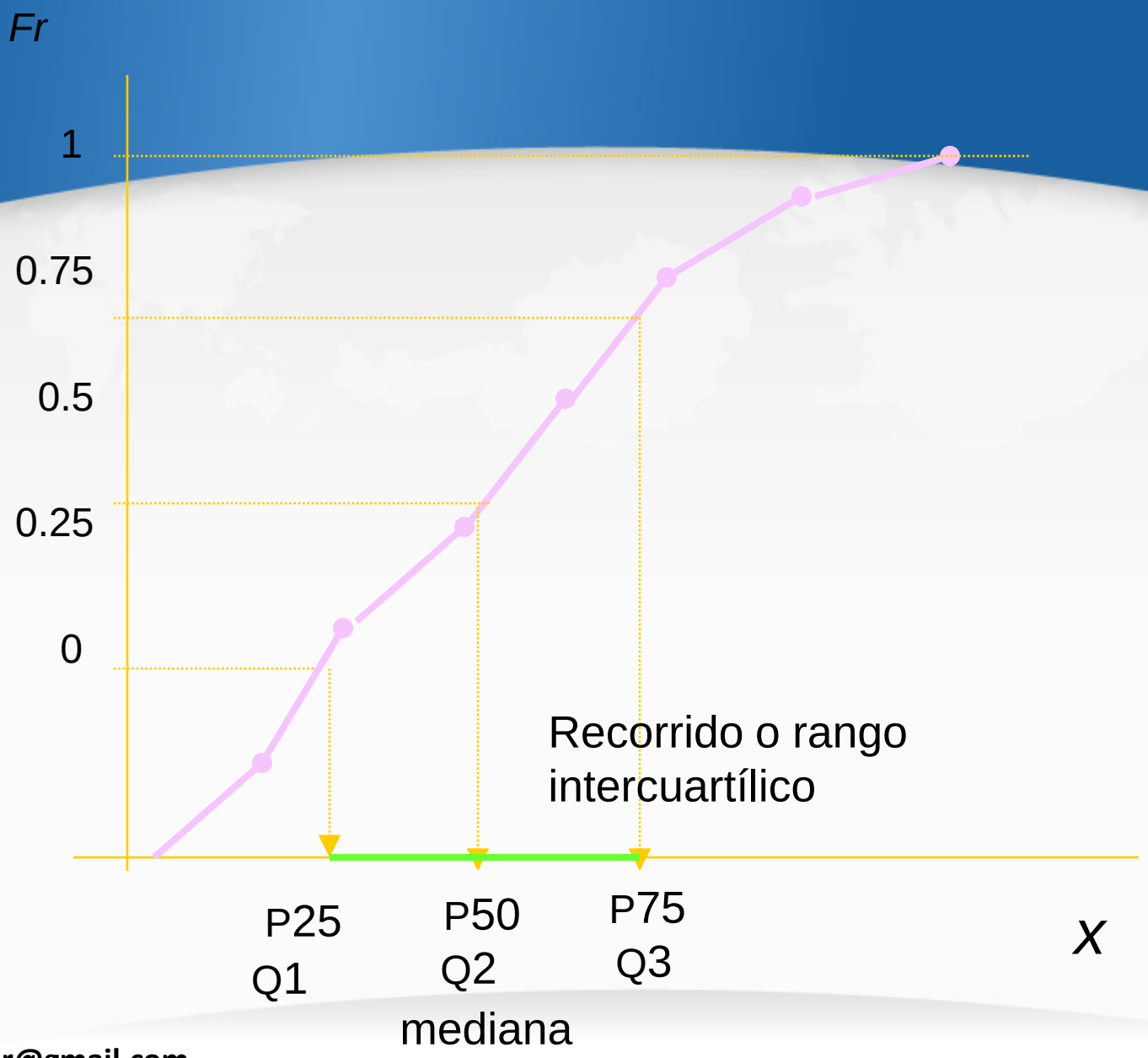
•**Amplitud o Rango** ('range'):
La diferencia entre las observaciones extremas.

- 2,1,4,3,8,4. El rango es 8-1=7
- Es muy sensible a los valores extremos.

•**Rango intercuartílico** ('interquartile range'):

- Es la distancia entre el primer y tercer cuartil.
  - Rango intercuartílico = $P_{75} - P_{25}$
- Parecida al rango, pero eliminando las observaciones más extremas inferiores y superiores.
- No es tan sensible a valores extremos.

aboadar@gmail.com



aboadar@gmail.com

Medidas de dispersión

- Varianza S^2 ('Variance'): Mide el promedio de las desviaciones (al cuadrado) de las observaciones con respecto a la media.

$$S^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2$$

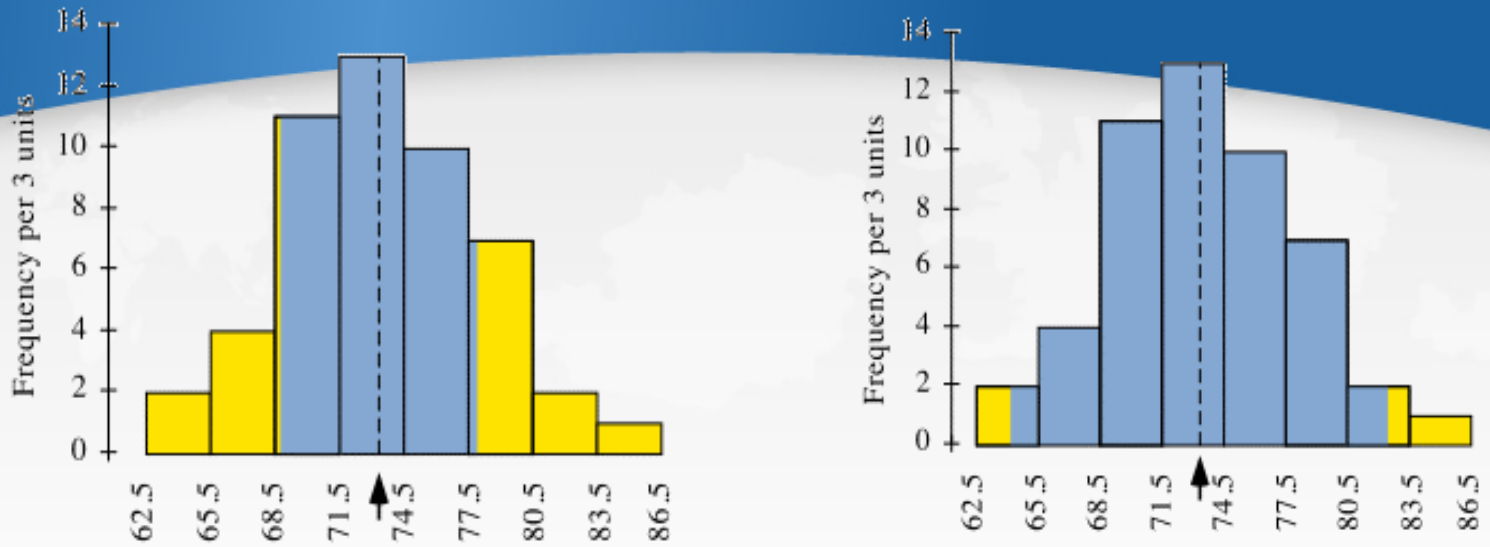
  - Es sensible a valores extremos (alejados de la media).
  - Sus unidades son el cuadrado de las de la variable.

- Desviación típica ('standard deviation')
  Es la raíz cuadrada de la varianza. Tiene las misma dimensionalidad (unidades) que la variable.

$$S = \sqrt{S^2}$$

aboadar@gmail.com

Medidas de dispersión



- Centrados en la media y a una desviación típica de distancia tenemos más de la mitad de las observaciones (izquierda.)

- A dos desviaciones típicas las tenemos a casi todas (derecha.)

aboadar@gmail.com

Medidas de dispersión

- **Coeficiente de variación**
- Es la razón entre la desviación típica y la media.

$$CV = \frac{S}{\overline{x}}$$

  - Mide la desviación típica en forma de "qué tamaño tiene con respecto a la media"
  - También se la denomina variabilidad relativa.
  - Es frecuente mostrarla en porcentajes
    - Si la media es 80 y la desviación típica 20 entonces CV=20/80=0,25=25% (variabilidad relativa)

- Es una cantidad **adimensional**. Interesante para comparar la variabilidad de diferentes variables.
  - Si el peso tiene CV=30% y la altura tiene CV=10%, los individuos presentan más dispersión en peso que en altura.

- No debe usarse cuando la variable presenta valores negativos o donde el valor 0 sea una cantidad fijada arbitrariamente
  - Por ejemplo 0ºC ≠ 0ºF
- Los ingenieros electrónicos hablan de la razón 'señal/ruido' (su inverso).

aboadar@gmail.com

Desigualdad de Chebyshev (1821-1894)

Si un conjunto de datos posee una varianza pequeña no existirán "muchos valores" alejados de la media. Precisemos: sea el intervalo alrededor de la media:

$$\bar{x} - k\sigma < x_i < \bar{x} + k\sigma$$

$$S^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \cdot f_i$$

$$S^2 = \underbrace{\frac{1}{n}\sum_{\substack{i\,dentro\\ del\ entorno}} (x_i - \bar{x})^2 \cdot f_i}_{>0} + \underbrace{\frac{1}{n}\sum_{\substack{i\ fuera\\ del\ entorno}} (x_i - \bar{x})^2 \cdot f_i}_{>0}$$

aboadar@gmail.com

Demostración:

$$S^2 \geq \frac{1}{n} \sum_{\substack{i \text{ fuera} \\ \text{del entorno}}} (x_i - \bar{x})^2 \cdot f_i \geq \frac{1}{n} \sum_{\substack{i \text{ fuera} \\ \text{del entorno}}} k^2 S^2 \cdot f_i =$$

$$= k^2 S^2 \frac{1}{n} \sum_{\substack{i \text{ fuera} \\ \text{del entorno}}} f_i$$

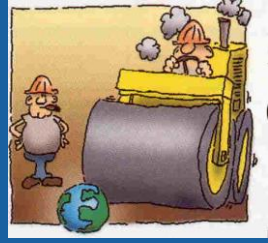
$$\frac{1}{n} \sum_{\substack{i \text{ fuera} \\ \text{del entorno}}} f_i \leq \frac{1}{k^2}$$

La frecuencia relativa de los datos que caen fuera del intervalo de centro media y radio k veces la varianza es igual o menor que

$1/k^2$

aboadar@gmail.com

Buen uso de la estadística



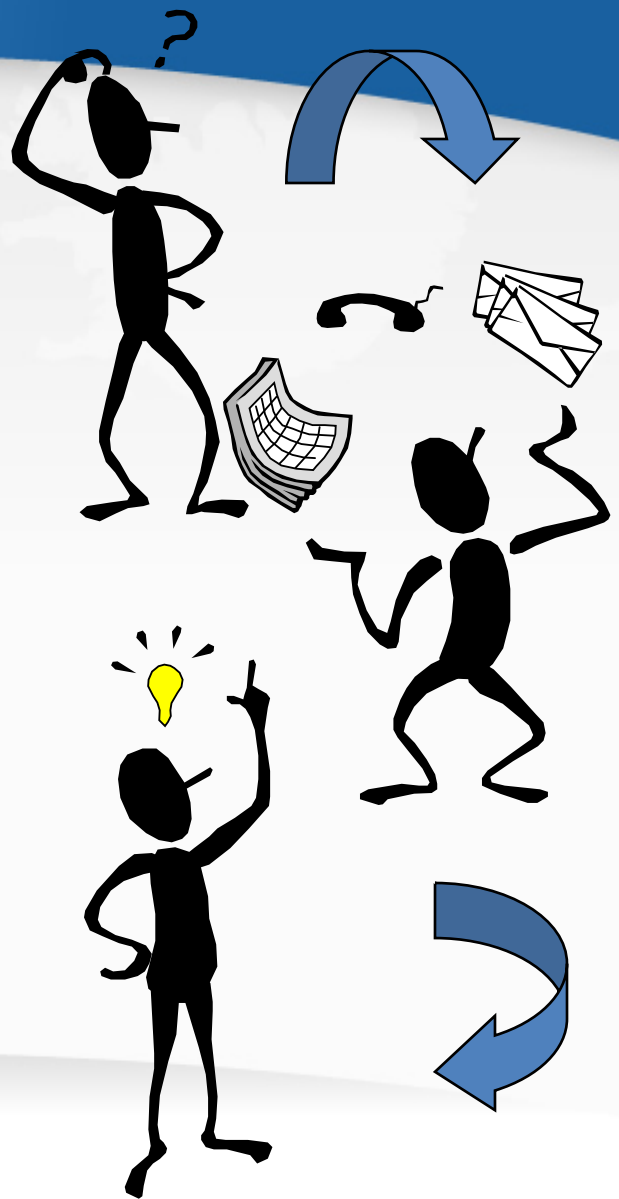
Eliminar el Sistema Gerencial de Cristóbal Colon:

Cuando partió.....
NO sabía para donde iba

Cuando tocó tierra.....
NO sabía donde estaba

Cuando regreso.....
NO sabía lo que había descubierto

Sin embargo era un excelente marino y realizó muchos viajes con éxito



aboadar@gmail.com

World Training Colombia

Gracias por su atención

aboadar@gmail.com