La Estadística y sus aportaciones

Curso Cero de Estadística y Probabilidad

Universidad de Salamanca

Grados en Economía y ADE

Contenidos

¿Cómo aprender Estadística?

¿Cómo aprender Estadística?

Ideas iniciales

1. La Estadística es una ciencia que acompaña una habilidad: el análisis de datos
2. El análisis de datos consiste en aprender a interpretar lo que observas
3. La Estadística es el lenguaje universal para interpretar resultados

El árbol de la Estadística

1. Conceptos base (raíces)
2. Tronco (metodología y software)
3. Hojas y ramas (técnicas de análisis)

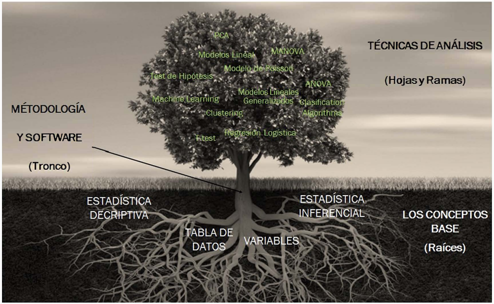


Figure 1: Fuente: Cómo aprender Estadística

¿Cómo aprender Estadística?

Las raíces. Los conceptos base

Parte fundamental centrada en conocer los conceptos relacionados con:

1. **Las tablas de datos**: Cómo se organizan los datos, qué son las variables y sus tipos
2. **La Estadística Descriptiva**: Conocer los conceptos fundamentales de descripción de datos y para que se utilizan
3. **La Estadística Inferencial**: Conocer los conceptos esenciales que nos servirán para aprender y dominar las ramas del árbol

Figure 2: Fuente: Cómo aprender Estadística

¿Cómo aprender Estadística?

- La **Estadística descriptiva** resume y representa los datos. Se utiliza para entender la información escondida en los datos
- La **Estadística matemática o inferencial** pretende ir un paso más allá y aportar conclusiones generales a partir de los datos

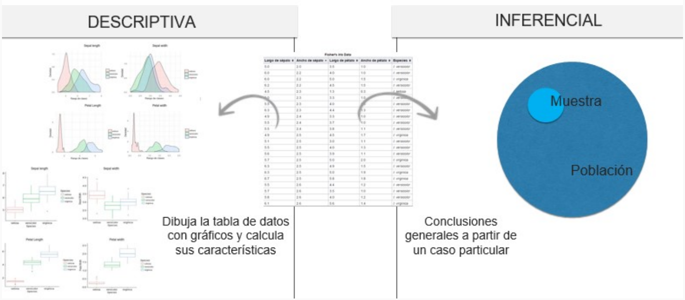


Figure 3: Fuente: Cómo aprender Estadística

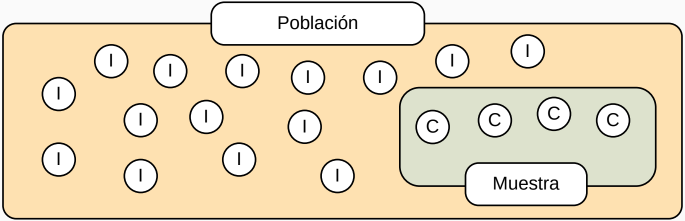
¿De donde salen los números?

Población

Conjunto o colección completa de todos los elementos/objetos/sujetos de interés sobre los que se va a realizar el estudio estadístico

Muestra

Un subconjunto de la población elegida habitualmente para investigar las propiedades de la población subyacente



Conceptos básicos

Parámetro

Una característica específica de una población (fijo)

Estadístico

Una característica específica de una muestra (varía de muestra en muestra)

Variable

Una característica de un individuo

Los números nos dicen cosas

Tipos de datos: Escalas de medida

Los datos se clasifican por niveles de medición o escalas de medida

- **Escala nominal**. Los datos sólo se clasifican

  Ejemplos: Sector de la actividad económica (primario, secundario, terciario), tener o no video en casa, sexo, grupo sanguíneo

- **Escala ordinal**. Los datos se ordenan

  Ejemplo: Nivel de estudios (sin estudios, primaria, secundaria, bachillerato, universitario)

Tipos de datos: Escalas de medida

- **Escala de intervalo**. Diferencia significativa entre valores

  Ejemplo: Temperatura, edad

- **Escala de razón o proporción**. Punto 0 significativo y razón entre valores

  Ejemplo: Longitud, peso, salario, número de pacientes atendidos, número de llamadas de ventas realizadas

Tipos de datos: Escalas de medida

Una misma variable puede ser medida en distintas escalas

Ejemplos:

- **Longitud de una pieza**

  Escala de intervalo: milímetros
  Escala ordinal: corta, normal, larga
  Escala nominal: aceptable, defectuosa

- **Tiempo de entrega a un cliente**
  Escala de intervalo: Días, horas, etc.
  Escala ordinal: Muy poco, poco, normal, etc.
  Escala nominal: dentro del plazo, fuera del plazo

Organización de los datos

Variable estadística

La materia prima para el análisis: **los datos** (variables)

Definición

Son características de interés de una población sobre las que se quiere hacer un estudio estadístico

- Variable aleatoria: Referida a la población
- Variable estadística: Referida a la muestra

Tipos

- Variables categóricas o cualitativas
- Variables cuantítativas o numéricas

Variable estadística: Tipos

Variables categóricas o cualitativas

Categorías exhautivas y mutuamente excluyentes

- **Nominales**: categorías no ordenables (sexo, grupo sanguineo, origen de un producto)
- **Ordinales**: categorías con orden natural (nivel de estudios)

Variables cuantítativas o numéricas

- **Discretas**: conjunto de valores aislados, finito o numerable (número de llamadas, número de hijos)
- **Continuas**: valores en un intervalo de la recta real de forma continua (temperatura, concetración de $C0_2$)

Variable estadística: Tipos

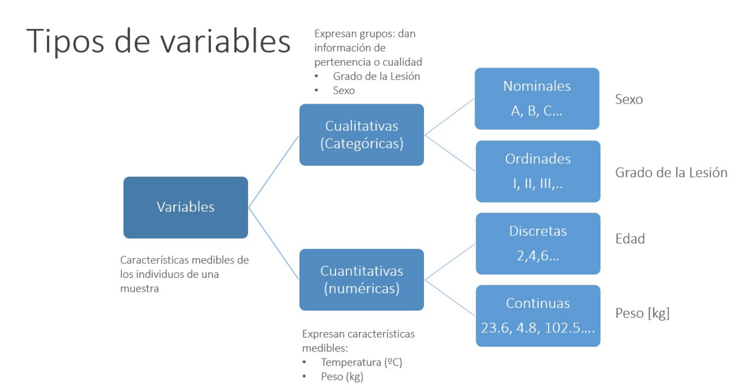


Figure 4: Fuente: Cómo aprender Estadística

Variable estadística

Datos (Variables)

Categóricos (Cualitativos)		Numéricos (Cuantitativos)	
Ordinales	Nominales	Discretos	Continuos
clases ordenables	sin orden natural	entero	no entero
Ejemplo	**Ejemplo**	**Ejemplo**	**Ejemplo**
Talla de ropa:	Grupo sanguíneo:	nº de hijos:	Altura:
G>M>P	A,B,AB,O	0,1,2,...	1.55cm, 1.71cm

Estadística Descriptiva

Variables cuantitativas:

- Tabla de frecuencias
- Valor central (media o mediana)
- Dispersión (desviación estándar, rango intercuartílico)
- Medidas de posición (cuartiles, percentiles)
- Distribución (histograma, boxplot, densidad)
- Correlación y covarianza

Variables cualitativas

- Tablas de contingencias
- Concepto de proporciones
- Distribución
- Diagrama de barras y sectores

Estadística Inferencial

Importante entender

- El concepto de muestra y población
- Qué son las distribuciones de variables
- Qué es y para qué sirve la distribución de densidad de probabilidad y las probabilidades
- El poder de la distribución normal
- Qué son los intervalos de confianza y su gran poder
- Tipos importantes de distribuciones de probabilidad
- El contraste de hipótesis y el p-valor
- La esencia de los modelos predictivos y sus dos super poderes

Los números en Estadística Descriptiva

Curso Cero de Estadística y Probabilidad

Universidad de Salamanca

Grados en Economía y ADE

Contenidos

Frecuencias y distribuciones de frecuencias

Distribuciones de frecuencias

Definición

Resumen de los datos que permite revelar la naturaleza de la información. Patrón de variabilidad mostrado por los datos de una variable. La distribución muestra la frecuencia de cada valor de la variable

¿Por qué usar distribuciones de frecuencias

- Es una forma de resumir los datos
- La distribución condensa los datos primarios de forma más útil
- Permite una interpretación visual y rápida de los datos

Distribución de frecuencias: Variable categórica y discreta

Frecuencia absoluta de la categoría

- **Definición**: Número de individuos o datos de la muestra con la característica o valor A_i
- **Notación**: n_i
- **Propiedades**:
  - $n_i \geq 0, \ i = 1, \ldots, k$
  - $n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i = N$

Frecuencia relativa de la categoría

- **Definición**: Frecuencia absoluta dividida por el número total de observaciones (se puede expresar en porcentaje)
- **Notación**: $f_i = \frac{n_i}{N}, \ i = 1, \ldots, k$
- **Propiedades**:
  - $0 \leq f_i \leq 1, \ i = 1, \ldots, k$
  - $f_1 + f_2 + \ldots + f_k = \sum_{i=1}^{k} f_i = 1$

Distribución de frecuencias: Variable categórica y discreta

Tiene sentido hablar de valores acumulados cuando los valores de la variable se han ordenado de menor a mayor (ordinales y discretos)

Frecuencia absoluta acumulada de la categoría

- **Definición**: Número de individuos que hay hasta la característica o valor A_i (inclusive)
- **Notación**: N_i
- **Propiedades**:
  - $n_1 = N_1$
  - $N_2 = n_1 + n_2 = \sum_{j=1}^{2} n_j = N_1 + n_2$
  - $\cdots$
  - $N_k = n_1 + n_2 + \ldots + n_k = \sum_{j=1}^{k} n_j = N_{k-1} + n_k = N$

Distribución de frecuencias: Variable categórica y discreta

Frecuencia relativa acumulada de la característica o valor A_i

- **Definición**: Frecuencia absoluta acumulada dividida entre el número total de observaciones (se puede expresar en porcentaje)
- **Notación**: $F_i = \frac{N_i}{N}$
- **Propiedades**:
  - $F_1 = \frac{N_1}{N} = f_1$
  - $F_2 = \frac{N_2}{N} = f_1 + f_2 = \sum_{j=1}^{2} f_j = F_1 + f_2$
  - $\ldots$
  - $F_k = \frac{N_k}{N} = f_1 + f_2 + \ldots + f_k = \sum_{j=1}^{k} f_j = F_{k-1} + f_k = 1$

Distribución de frecuencias: Variable categórica y discreta

X	n_i	$f_i = \frac{n_i}{N}$	N_i	F_i
A_1	n_1	$f_1 = \frac{n_1}{N}$	$N_1 = n_1$	$F_1 = f_1$
A_2	n_2	$f_2 = \frac{n_2}{N}$	$N_2 = N_1 + n_2$	$F_2 = F_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_k	n_k	$f_k = \frac{n_k}{N}$	$N_k = \mathbf{N}$	$F_k = \mathbf{1}$
Total	$\mathbf{N}$	$\mathbf{1}$		

Distribución de frecuencias: Variable cuantitativa continua

- Son aquellas que toman, en lugar de categorías, valores numéricos
- Hay que distinguir entre variables **discretas** (si la variable toma un número finito o infinito numerable de observaciones) y **continuas** (infinito no numerable)
- **El primer paso del análisis**: Observar los distintos valores que toma una variable, ordenarlos de menor a mayor y contar el número de veces que aparece cada valor
- **El problema**: Suele tomar un número de valores mucho mayor que las posibles categorías de una variable cualitativa
- **Solución**: Trabajar con intervalos de la variable, **intervalos de clase**

Distribución de frecuencias: Variable cuantitativa continua

Definiciones

- **Soporte de la variable**: (a, b), $a = a_0 < a_1 < \cdots < a_k = b$
- **Intervalo de clase**: Rango utilizado para dividir el conjunto de posibles valores numéricos. Cada clase tiene un extremo superior y uno inferior. Se suele incluir el extremo inferior, pero se excluye el superior (convención típica): $[a_{i-1}, a_i)$
- **Extremos de clase**: a_{i-1}, a_i
- **Marca de clase**: $m_i = \frac{a_{i-1}+a_i}{2}$
- **Amplitud de clase**: $A = a_i - a_{i-1}$

Distribución de frecuencias: Variable cuantitativa

Definiciones

- **Frecuencia absoluta del intervalo de clase**: n_i
- **Frecuencia relativa del intervalo de clase**: $f_i = \frac{n_i}{N}$
- **Frecuencia absoluta acumulada del intervalo de clase**: N_i
- **Frecuencia relativa acumulada del intervalo de clase**: $F_i = \frac{N_i}{N}$

Distribución de frecuencias: Variable cuantitativa

Construcción de una distribución de frecuencias para datos cuantitativos mediante intervalos de clase

1. Elegir un número razonable de intervalos *significativos*. Utilizar la Regla de Sturges:
$$k = 1 + 3.3 \cdot \log N$$
siendo k el número de clases
2. Se recomienda usar clases de igual amplitud
$$A = \frac{a_k - a_0}{k}$$
3. Normalmente se redondea la amplitud del intervalo para obtener los extremos de los intervalos deseados
4. Las clases deben definirse con precisión para poder clasificar sin ambigüedad las observaciones en uno de los intervalos

Distribución de frecuencias: Variable cuantitativa

Int. Clase	m_i	n_i	f_i	N_i	F_i
$[a_0 - a_1)$	m_1	n_1	f_1	N_1	F_1
$[a_1 - a_2)$	m_2	n_2	f_2	N_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[a_{k-1} - a_k)$	m_k	n_k	f_k	$N_k = \mathbf{N}$	$F_k = \mathbf{1}$
Total		$\mathbf{N}$	$\mathbf{1}$		

Representaciones gráficas

Representaciones gráficas

- **Objetivo**: Representar la información numérica contenida en la tabla de frecuencias. Para cada tipo de variable existen representaciones apropiadas que son variaciones de la misma idea

Categórico	Numérico
$\Downarrow$	$\Downarrow$
• diagrama de sectores	• histograma
• diagrama de barras	• polígono de frecuencias
	• diagrama de caja

Representaciones gráficas: Variables categóricas

Diagrama de sectores

- Sobre un círculo se representan sectores asociados a cada valor de la variable
- Los sectores son proporcionales a las frecuencias (absolutas o relativas)
- Se usa preferentemente para variables medidas en escala nominal, pero también para la ordinal
- En el caso ordinal hay que mantener el orden natural de las categorías de la variable
- Es aconsejable que el número de sectores no exceda de 5, ordenados de mayor a menor, iniciando con el más amplio a partir de las 12h (reloj)

Diagrama de sectores

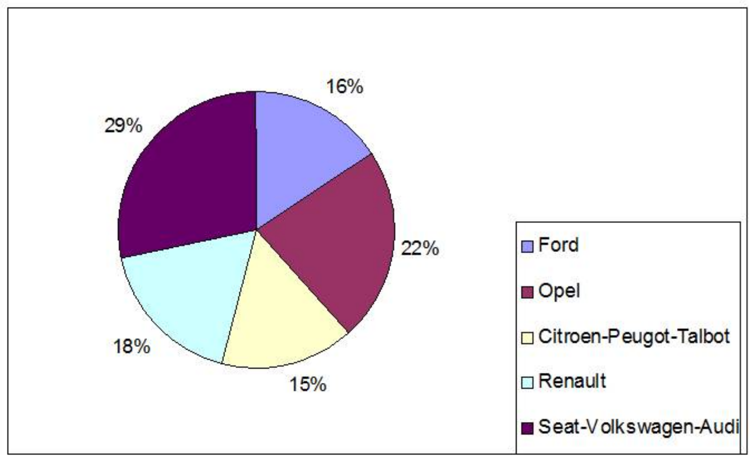
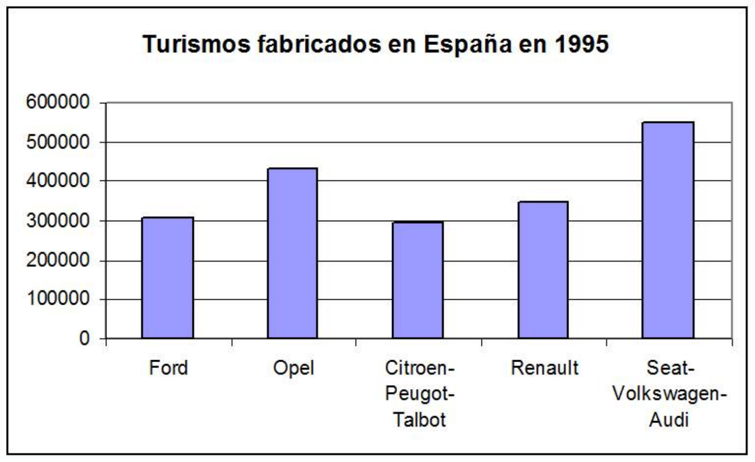


Diagrama de barras

- Sobre un eje simbólico se representan los valores de la variable
- Sobre cada valor se levanta un rectángulo (barra) cuya altura representa la frecuencia
- Se pueden hacer con frecuencias absolutas o relativas Las frecuencias relativas permiten la comparación de muestras diferentes
- Se usan tanto para variables medidas en escala nominal como ordinal. En el caso ordinal hay que mantener el orden natural de las categorías de la variable

Diagrama de barras



Representaciones gráficas: Variables categóricas

Diagrama de Pareto

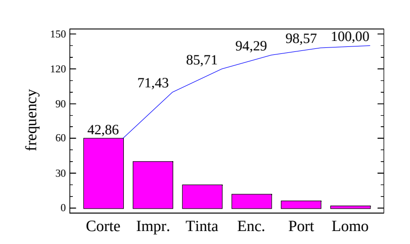
- Diagrama de barras con las categorías de la variable ordenadas de mayor a menor frecuencia
- Sólo para atributos nominales (los ordinales sólo admiten la ordenación natural de las categorías)
- Las clases menos significativas se pueden agrupar como OTROS, que se colocará en último lugar Las frecuencias relativas permiten la comparación de muestras diferentes
- Es un diagrama en barras en el que las categorías se ordenan de mayor a menor frecuencia, dibujando sobre las barras una línea indicativa de la frecuencia acumulada hasta esa categoría

Representaciones gráficas: Variables categóricas

Diagrama de Pareto

EJEMPLO: Defectos en libros en una imprenta

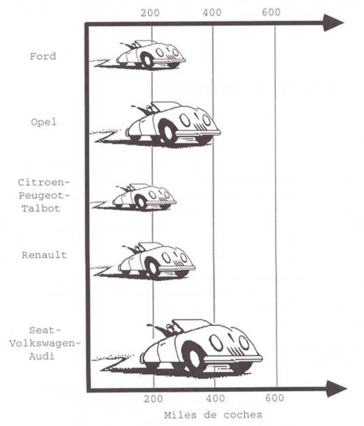
TIPOS	Frecuencia	Frec. Relativa
Corte de las hojas	60	0,43
Mala impresión	40	0,29
Tinta irregular	20	0,14
Encuadernación	12	0,09
Portada	6	0,04
Lomo	2	0,01
TOTAL	140	1,00



Representaciones gráficas: Variables categóricas

Pictograma

- Cada modalidad de la variable está representada por un dibujo alusivo a su significado



Representaciones gráficas: Variables numéricas o cuantitativas

Histograma

- Se levantan rectángulos sobre las clases en que han sido agrupados los valores de la variable cuantitativa
- Los rectángulos tienen que ser adyacentes para reflejar la continuidad de la variable
- La frecuencia se representa a través del área de cada rectángulo
- El área bajo el histograma es 1 si son frecuencias relativas, o N si son absolutas
- Para comparar muestras hay que usar frecuencias relativas
- El aspecto del histograma depende tanto del número de clases como de la posición de éstas

Representaciones gráficas: Variables numéricas o cuantitativas

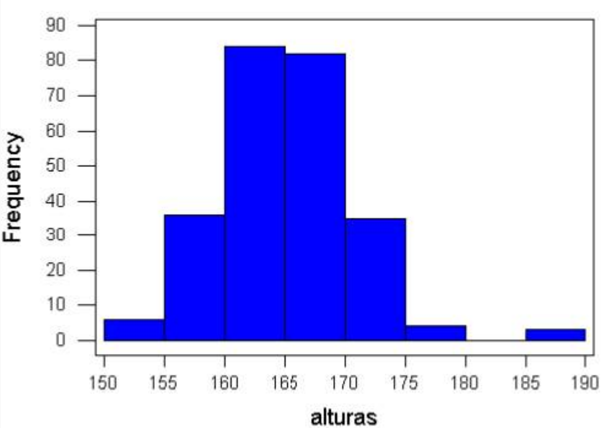
Histograma

- Es conveniente usar clases de igual longitud. De esta manera, la frecuencia también se verá representada por la altura de los rectángulos, lo cual facilita la representación e interpretación
- Para elegir el número de clases y su posición, conviene probar varias posibilidades y elegir la que proporcione una representación más razonable
- El histograma refleja la densidad de aparición de observaciones sobre el soporte de la variable

Histograma

- El histograma pone de manifiesto las características de cada conjunto de datos:
  - Localización
  - Dispersión
  - Simetría o asimetría
  - Unimodalidad o multimodalidad
  - Observaciones atípicas, etc.
- Tomando una muestra representativa, cada vez más grande, y haciendo tender a 0 la amplitud de las clases, aparece en el límite una curva que va a reflejar el modelo de la población
- Es una herramienta importante en el control de calidad

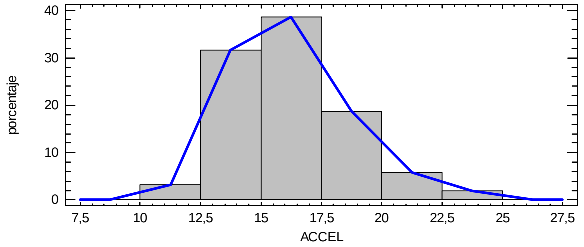
Histograma

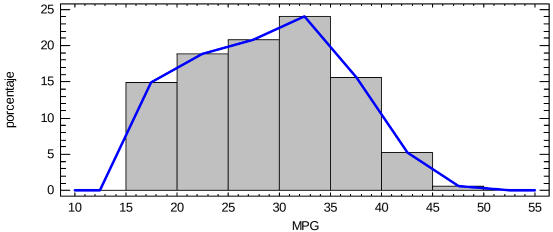

Frequency
90
80
70
60
50
40
30
20
10
0
150
155
160
165
170
175
180
185
190
alturas

Polígono de frecuencias sobre el histograma

- Se traza un polígono uniendo los valores del histograma en las marcas de clase
- Si las clases son de igual amplitud, los extremos del polígono se unen con los puntos correspondientes a las marcas de clase de las que serían la anterior y la posterior a las utilizadas para agrupar los datos
- Si se hace con frecuencias relativas y clases de igual amplitud, el área bajo el polígono es 1
- Es una versión suavizada del histograma, en un intento de aproximar el modelo poblacional que surgiría con muchas observaciones y clases de amplitud tendiendo a 0

Polígono de frecuencias sobre el histograma





Representación de los datos

Curso Cero de Estadística y Probabilidad

Universidad de Salamanca

Grados en Economía y ADE

Contenidos

Descripción numérica de datos

Descripción numérica de variables estadísticas

Interés de los resúmenes numéricos

- Unos pocos números resumen las características fundamentales de la distribución
- Complemento natural de la descripción gráfica
- Facilitan la comparación de muestras con modelos de referencia y la comparación entre muestras.

Tipos de resúmenes numéricos

- **Posición** de la distribución: Media, Mediana, Moda, Cuartiles, Percentiles
- **Dispersión** de la distribución: Rango, Varianza, Desviación típica, Coeficiente de variación
- **Forma** de la distribución: Simetría y Kurtosis

Medidas de posición

Medidas de posición de tendencia central

Media aritmética

- Variable discreta:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum_{i=1}^{k} x_i \cdot n_i = \sum_{i=1}^{k} x_i \cdot f_i$$

- Variable continua (agrupada en k int. clases):

$$\overline{X} \simeq \frac{1}{N}\sum_{i=1}^{k} m_i n_i = \sum_{i=1}^{k} m_i f_i$$

Propiedades de la Media

- Es única
- No tiene porque ser un valor observado de la variable
- En su calculo intervienen todos los datos
- Cambios de origen y escala en los datos $y_i = a + bx_i,\ i = 1, 2, \ldots, n$ conllevan los mismos cambios en la media

$$\overline{Y} = a\overline{X} + b$$

Medidas de posición de tendencia central

Propiedades de la Media

- La media de una suma es la suma de las medias. Sea $(x_1, x_2, \ldots, x_n) \longrightarrow \overline{X}$ e $(y_1, y_2, \ldots, y_n) \longrightarrow \overline{Y}$

$$(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \longrightarrow \overline{X} + \overline{Y}$$

- Si la muestra esta dividida en dos grupos, la media de la muestra es la media ponderada (por los tamaños de los grupos) de las medias
- La suma de las desviaciones con respecto a la media es cero

$$\sum_{i=1}^{N}(x_i - \overline{X}) = 0$$

Medidas de posición de tendencia central

Mediana

Punto que parte la distribución en dos mitades del 50% a cada lado. Observación central en la muestra ordenada

- Si N es impar: Observación central en la muestra ordenada

$$Me = x_{(\frac{n+1}{2})}$$

- Si N es par: Promedio de las dos observaciones centrales en la muestra ordenada

$$Me = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} \in \left(x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}\right)$$

- Datos agrupados: Intervalo $[L_{j-1}, L_j]$

$$Me_X \simeq L_{j-1} + A_i \cdot \left(\frac{\frac{N}{2} - N_{j-1}}{N_j - N_{j-1}} \right)$$

Medidas de posición de tendencia central

Propiedades Mediana

- Es única
- En su calculo no intervienen todos los datos
- Siempre es un valor observado de la variable
- Sea X con media $\overline{X}$ entonces

$$Y = aX + b \Longrightarrow Me_Y = a\,Me_X + b$$

Medidas de posición de tendencia central

Moda

Punto donde se alcanza el máximo de la distribución de frecuencias

Propiedades

- No es única. Hay distribuciones con varias modas locales (bimodales o multimodales)
- Siempre es un valor observado de la variable
- En su calculo no intervienen todos los datos
- Para variables continuas con datos agrupados, hablaremos de intervalo modal
- Sea X con media $\overline{X}$ entonces

$$Y = aX + b \Longrightarrow Mo_Y = a\,Mo_X + b$$

Elección de la medida de tendencia central adecuada

La elección de una medida de posición óptima no tiene una solución universal, depende del criterio de comparación

- **Eficiencia**: Para datos normales, la media es más eficiente que la mediana. La media utiliza todas las observaciones. La mediana sólo las centrales
- **Robustez**: Estabilidad frente a la presencia de observaciones atípicas
  - La Media es poco robusta: una sola observación errónea puede cambiar mucho la media
  - La Mediana es más robusta: se necesitan muchas observaciones erróneas para producir cambios importantes en la mediana

Medidas de posición no centrales

- Permiten dividir la distribución en un variable definida como el numero de segmentos, **cuantiles**, facilitando la ubicación de orden de un sujeto o caso sobre un conjunto de los datos

- Estas medidas requieren que exista un orden en las categorías de la variable, por lo que solo se pueden determinar a partir de la escala ordinal

- Los cuantiles más comunes en el ámbito de la Estadística Aplicada a las Ciencias Sociales son: cuartiles, deciles y percentiles

Medidas de posición no centrales

Cuartiles

Son los valores de la variable que dividen la distribución en cuatro partes iguales, son por tanto tres puntos y se denotan como Q_1, Q_2 y Q_3

Deciles

Los deciles son las nueve marcas que fraccionan la distribución en diez partes iguales, conteniendo cada una de ellas la décima parte de las observaciones y se nombran como $D_1, D_2, \ldots, D_8, D_9$

Percentiles

Los centiles o percentiles segmentan el conjunto de las observaciones en cien partes iguales y su notación se expresa como $C_1, \ldots, C_{20}, \ldots, C_{99}$ ó $P_1, \ldots, P_{20}, \ldots, P_{99}$

Medidas de posición no centrales

Medidas de posición	Cálculo	
	Distribución de datos no agrupados	Distribución de datos agrupados
Cuartiles [Q_m]	Localización	$Q_m = L_i + \frac{mN/4 - N_{ai-1}}{n_m} \times a$
Deciles [D_m]		$D_m = L_i + \frac{mN/10 - N_{ai-1}}{n_m} \times a$
Centiles o Percentiles [C_m] o [P_m]		$P_m = L_i + \frac{mN/100 - N_{ai-1}}{n_m} \times a$

a: Amplitud del intervalo
N_{ai}: Frecuencia absoluta acumulada del intervalo anterior a aquel en que se encuentra el cuantil
m: cuantil que se desea conocer
n_m: frecuencia absoluta total del intervalo cuartil
L_i: límite inferior del intervalo donde se encuentra el cuantil

Medidas de dispersión o variabilidad

Rango

Es la diferencia entre el valor máximo y mínimo

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

- Ignora la manera en en que se distribuyen los datos
- Sensible a observaciones atípicas

Recorrido intercuatílico

Diferencia entre el tercer y el primer cuartil

$$RI = Q_3 - Q_1$$

Medidas de dispersión absolutas: Varianza

Definición

Promedio de las desviaciones cuadráticas en torno a la media

$$S_X^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{X})^2 = \frac{1}{N}\sum_{i=1}^{k} x_i^2 - \overline{X}^2$$

Propiedades

- $S_X^2 \geq 0$
- Mayor varianza $\Rightarrow$ mayor dispersión
- Sea $Y = aX + b \Rightarrow S_Y^2 = a^2\, S_X^2$
- Variable discreta: $S_X^2 = \sum_{i=1}^{k}(x_i - \overline{X})^2 \cdot f_i$
- Variable continua: $S_X^2 = \sum_{i=1}^{k}(m_i - \overline{X})^2 \cdot f_i$

Medidas de dispersión absolutas: Desviación típica y cuasivarianza o varianza corregida

- **Desviación típica**:

$$S_X = \sqrt{\frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{X})^2} = \sqrt{S_X^2}$$

- **Cuasivarianza o varianza corregida**:

$$S_C^2 = \frac{1}{N-1}\sum_{i=1}^{k}(x_i - \overline{X})^2$$

- **Relación varianza y cuasi-varianza**:

$$S_X^2 = \frac{N-1}{N} S_C^2$$

Medidas de dispersión relativas: Coeficiente de variación de Pearson

Definición

$$CV_X = \frac{S_X}{|\,\overline{X}\,|}$$

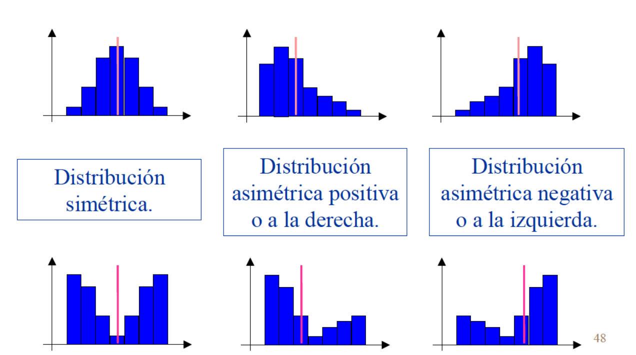
Propiedades

- Relativiza la dispersión en función de la magnitud (escala) de las observaciones
- No tiene unidades. Adimensional
- Facilita la comparación
- A menor $CV_X \Rightarrow$ mayor representatividad de la media $\overline{X}$ y menor dispersión

Medidas de forma

Estudio de la forma: Simetría

Una distribución de frecuencias X es simétrica cuando son iguales las frecuencias correspondientes a valores equidistantes de un valor central considerado como eje



Medidas de asimetría: Coeficiente de asimetría

Definición

- Datos originales:

$$CA = \frac{1}{N} \cdot \frac{\sum_{i=1}^{k} (x_i - \overline{X})^3 \cdot n_i}{S_X^3} = \frac{1}{N} \cdot \sum_{i=1}^{k} \left(\frac{x_i - \overline{X}}{S_X} \right)^3$$

- Datos tipificados:

$$CA = \frac{1}{N} \cdot \sum_{i=1}^{k} z_i^3$$

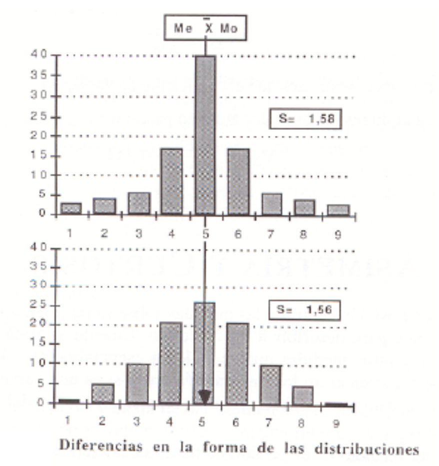
Medidas de asimetría: Coeficiente de asimetría

Propiedades

- Distribución simétrica: $CA \cong 0$
- Distribución asimétrica positiva: $CA > 0$ (cola derecha más pesada)
- Distribución asimétrica negativa: $CA < 0$ (cola izquierda más pesada)

Estudio de la forma: Apuntamiento

Medida de la importancia de las colas de la distribución



Medidas de curtosis o apuntamiento: Coeficiente de apuntamiento o curtosis

Definición

- Datos originales:

$$Cap = \frac{1}{N} \cdot \frac{\sum_{i=1}^{k}(x_i - \overline{X})^4 \cdot n_i}{S_X^4} = \frac{1}{N} \cdot \sum_{i=1}^{k} \left(\frac{x_i - \overline{X}}{S_X} \right)^4$$

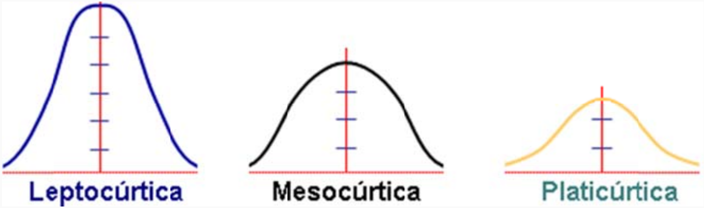
- Datos tipificados

$$Cap = \frac{1}{N} \cdot \sum_{i=1}^{k} z_i^4$$

Medidas de curtosis o apuntamiento: Coeficiente de apuntamiento o curtosis

Propiedades

- Distribución normal (mesocúrtica): $Cap \cong 3$
- Distribución más apuntada (leptocúrtica): $Cap > 3$
- Distribución menos apuntada (platicúrtica): $Cap < 3$



Cuadro resumen Descripción numérica de datos

Tendencia central = RESUMIR	Posición no central = CATEGORIZAR
○ Moda ○ Mediana ○ Media aritmética	○ Cuartiles ○ Deciles ○ Percentiles
Dispersión = **VARIABILIDAD**	**Forma** = **SIMETRÍA y CONCENTRACIÓN**
○ Varianza ○ Desviación típica	○ Asimetría ○ Curtosis

Medidas de concentración o desigualdad

Medidas de concentración o desigualdad

¿Qué miden?

El grado de homogeneidad o desigualdad en el reparto del "todo"

¿Cómo oscila?

- **Concentración máxima**: Un individuo tiene el "todo" y el resto no tiene nada
- **Concentración mínima**: Todos los individuos tienen exactamente la misma cantidad

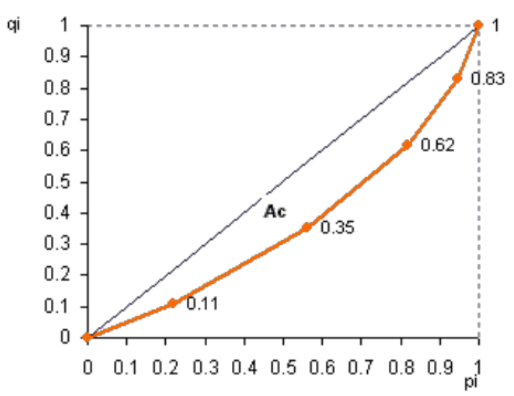
¿Qué utilizamos para medir la concentración?

- **Índice de Gini**: para medir situaciones intermedias
- **Curva de Lorenz**: para visualizar en grado de concentración

Curva de Lorenz

Definición

Es la poligonal que une los pares (p_i, q_i). La representación gráfica estará por debajo de la bisectriz del primer cuadrante



Curva de Lorenz

Cálculo

x_i	n_i	$p_i = F_i$	$x_i n_i$	U_i	$q_i = \frac{U_i}{U_k}$
x_1	n_1	$p_1 = F_1$	$x_1 n_1$	$U_1 = x_1 n_1$	$q_1 = \frac{U_1}{U_k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	$p_k = 1$	$x_k n_k$	$U_k = \sum_{i=1}^{k} x_i n_i$	$q_k = \frac{U_k}{U_k} = 1$

Índice de Gini

Definición

$$I_G = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

Cálculo

x_i	n_i	$p_i = F_i$	$x_i n_i$	U_i	$q_i = \frac{U_i}{U_k}$
x_1	n_1	$p_1 = F_1$	$x_1 n_1$	$U_1 = x_1 n_1$	$q_1 = \frac{U_1}{U_k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	$p_k = 1$	$x_k n_k$	$U_k = \sum_{i=1}^{k} x_i n_i$	$q_k = \frac{U_k}{U_k} = 1$

Índice de Gini

Propiedades

- El índice siempre toma valores entre 0 y 1
- $p_i - q_i \geq 0$
- Si toma el valor cero, implica igualdad entre todos los individuos (mínma concentración)

$$I_G = 0 \Rightarrow x_1 = x_2 = \ldots = x_k$$

- Si toma el valor 1, implica máxima desigualdad, un individuo dispone del todo (máxima concetración)

$$I_G = 1 \Rightarrow q_1 = \ldots = q_{k-1} = 0 \quad q_k = 1$$

Índice de Gini

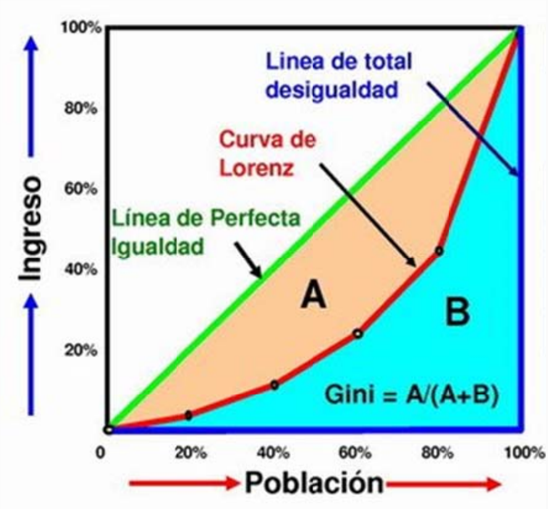


Diagrama de caja

Diagrama de caja

- Resumen rápido de una distribución de frecuencias de una muestra utilizando cinco estadísticos:
  - los cuartiles: Q_1, $Q_2 = Me$ y Q_3
  - las observaciones extremas máximo y mínimo
- Aporta información rápida sobre posición dispersión y forma de la distribución
- Se completa con unos límites:
  - Inferior: $L_I = Q_1 - 1.5 \cdot (Q_3 - Q_1)$
  - Superior: $L_S = Q_3 + 1.5 \cdot (Q_3 - Q_1)$
- Datos fuera de (L_I, L_S) son posibles datos anómalos (outliers), errores de medición o de tecleado, etc.
- Para datos normales, $fr(L_I, L_S) = 99\%$

Diagrama de caja

- **Caja**: Q_1, Me, Q_3 (contiene el 50% de los datos)
- **Patas**: La observación más grande y la más pequeña (L_S, L_I)

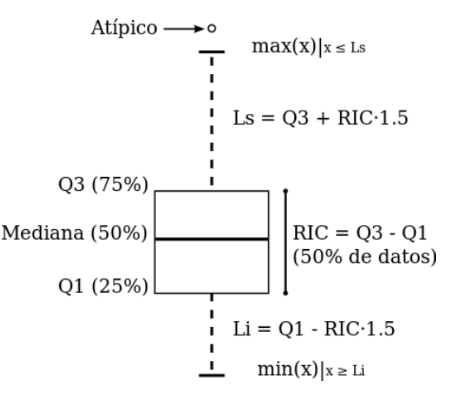


Diagrama de caja

La forma de la caja puede ser:

1. Cuadrada
2. Rectangular-achatada (mayor concentración)
3. Rectangular-expandida (mayor dispersión)

Las distintas morfologías describen si, el 50% de los valores centrales, se distribuyen de forma normal (1) o, si por el contrario sus valores se concentran (2) o, se dispersan (3)

Diagrama de caja

La línea que divide la caja representa la **mediana** de la distribución y es un detector de simetría:

1. Si la línea se sitúa en el centro de la caja $\longrightarrow$ la distribución es **simétrica**
2. Si por el contrario la línea se aproxima al extremo inferior $(Q_1) \longrightarrow$ la distribución es **asimétrica positiva**
3. Si por el contrario la línea se aproxima al extremo superior $(Q_3) \longrightarrow$ la distribución es **asimétrica negativa**

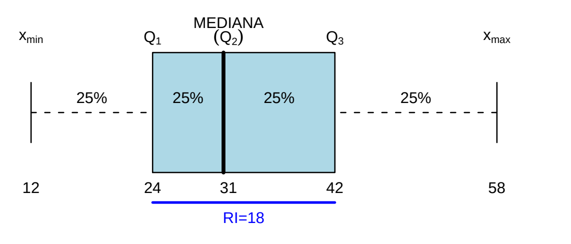


Diagrama de caja

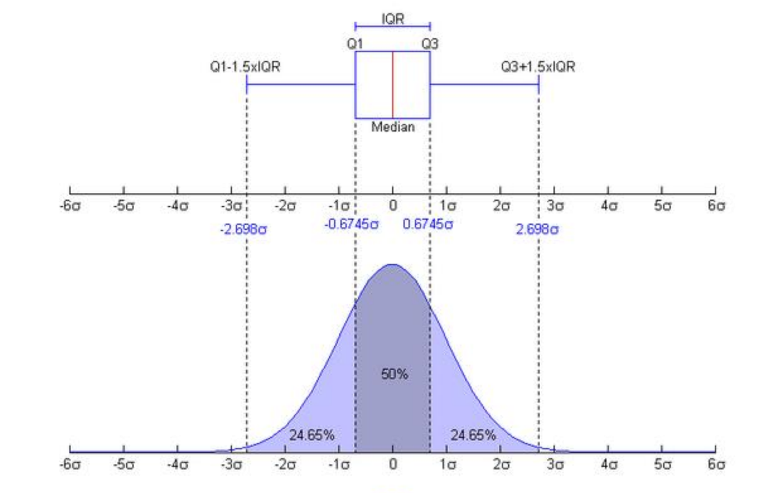
Los bigotes

- De la caja salen dos segmentos (superior e inferior) denominados **bigotes** que señalan el límite para la detección de valores atípicos
- La longitud de los bigotes expresa la **variabilidad** de la distribución en el 25% de los valores bajos (por debajo de Q_1) y altos (por encima de Q_3)

Los valores atípicos

- Los valores que se alejan más de lo esperado del centro de la distribución ($RI = Q_3 - Q_1$) son clasificados como **atípicos** y **extremos**
  - Se considera que un valor es **atípico** cuando excede $1,5$ unidades de longitud de caja su distancia de los percentiles 25 (Q_1) y 75 (Q_3)
  - Se considera que un valor es **extremo** cuando excede 3 unidades de longitud de caja su distancia de los percentiles 25 (Q_1) y 75 (Q_3)

Diagrama de caja y curva Normal



Variables Bidimensionales

Curso Cero de Estadística y Probabilidad

Universidad de Salamanca

Grados en Economía y ADE

Contenidos

Datos bivariantes

Datos bivariantes: Descripción

- En la mayoría de los problemas de interés interviene más de una variable y la información tomada se presenta como un conjunto de datos bivariantes o multivariantes

- Las descripciones univariantes de las variables que intervienen en estos problemas dan información incompleta de los mismos. El mayor interés está en el estudio de la asociación (relaciones) entre las distintas variables

- Descubrir la existencia de cierto tipo de relación entre una variable Y y otra variable X, puede permitir a veces atribuir a esta última parte de la variabilidad de la primera. Existen técnicas descriptivas diferentes para describir la asociación entre variables, según el tipo de variables a estudiar

Datos bivariantes: Descripción

- **Asociación entre variables cualitativas o atributos**. En este caso se utilizan tablas de contingencia o comparación de grupos respecto a una variable cualitativa
- **Asociación entre una variable numérica y una cualitativa**. En este caso se realizan comparaciones entre grupos o subpoblaciones respecto a la variable numérica
- **Asociación entre variables cuantitativas**. Se utiliza el análisis de regresión y de correlación

Datos bivariantes: Tablas de frecuencias

Frecuencia absoluta conjunta

Número de individuos que presentan simultaneamente $X = x_i$ e $Y = y_j$

$X \backslash Y$	y_1	y_2	$\cdots$	y_p	$n_{i.}$
x_1	n_{11}	n_{12}	$\cdots$	n_{1p}	$n_{1.}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2p}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\cdots$	n_{kp}	$n_{k.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.p}$	N

$$\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{p} n_{.j} = N$$

Datos bivariantes: Tablas de frecuencias

Frecuencia relativa conjunta

Proporción de individuos que presentan simultaneamente $X = x_i$ e $Y = y_j$

$X \backslash Y$	y_1	y_2	$\cdots$	y_p	$f_{i.}$
x_1	f_{11}	f_{12}	$\cdots$	f_{1p}	$f_{1.}$
x_2	f_{21}	f_{22}	$\cdots$	f_{2p}	$f_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	f_{k1}	f_{k2}	$\cdots$	f_{kp}	$f_{k.}$
$f_{.j}$	$f_{.1}$	$f_{.2}$	$\cdots$	$f_{.p}$	1

$$f_{ij} = \frac{n_{ij}}{N}$$

$$\sum_{i=1}^{k} \sum_{j=1}^{p} f_{ij} = \sum_{i=1}^{k} f_{i.} = \sum_{j=1}^{p} f_{.j} = 1$$

Datos bivariantes: Distribuciones marginales y condicionadas

A partir de **una variable bidimensional** podemos estudiar **dos tipos de variables unidimensionales**:

- **Distribuciones marginales**
- **Distribuciones condicionadas**

Datos bivariantes: Distribución marginal de X

Frecuencia absoluta marginal

X	$n_{i\cdot}$
x_1	$n_{1\cdot}$
$\vdots$	$\vdots$
x_k	$n_{k\cdot}$
	N

Frecuencia relativa marginal: $f_{i\,\cdot} = \frac{n_{i\,\cdot}}{N}$

X	$f_{i\cdot}$
x_1	$f_{1\cdot}$
$\vdots$	$\vdots$
x_k	$f_{k\cdot}$
	1

Datos bivariantes: Distribución marginal de X

- **Media marginal**:

$$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum_{i=1}^{k} x_i n_{i.}}{N} = \sum_{i=1}^{k} x_i f_{i.}$$

- **Varianza marginal**:

$$S_X^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{X})^2}{N} = \frac{\sum_{i=1}^{k} x_i n_{i.}}{N} - (\overline{X})^2 = \sum_{i=1}^{k} x_i f_{i.} - (\overline{X})^2$$

Datos bivariantes: Distribución marginal de Y

Frecuencia absoluta marginal

Y	$n_{\cdot j}$
y_1	$n_{\cdot 1}$
$\vdots$	$\vdots$
y_p	$n_{\cdot p}$
	N

Frecuencia relativa marginal: $f_{\cdot j} = \frac{n_{\cdot j}}{N}$

Y	$f_{\cdot j}$
y_1	$f_{\cdot 1}$
$\vdots$	$\vdots$
y_p	$f_{\cdot p}$
	1

Datos bivariantes: Distribución marginal de Y

- **Media marginal**:

$$\overline{Y} = \frac{\sum_{j=1}^{N} y_j}{N} = \frac{\sum_{j=1}^{p} y_j n_{.j}}{N} = \sum_{j=1}^{p} y_j f_{.j}$$

- **Varianza marginal**:

$$S_Y^2 = \frac{\sum_{j=1}^{N}(y_j - \overline{Y})^2}{N} = \frac{\sum_{j=1}^{p} y_j n_{.j}}{N} - (\overline{Y})^2 = \sum_{j=1}^{p} y_j f_{.j} - (\overline{Y})^2$$

Distribuciones condicionadas: Distribución de X condicionada por Y

- La tabla se lee por columnas
- Familia de p distribuciones condicionadas de la variable X a cada valor de $Y = y_1, \ldots, y_p \longrightarrow X/Y_{=y_j}$, $j = 1, \ldots, p$

$X \backslash Y$	y_1	y_2	$\cdots$	y_p
x_1	$f_{1/1}$	$f_{1/2}$	$\cdots$	$f_{1/p}$
x_2	$f_{2/1}$	$f_{2/2}$	$\cdots$	$f_{2/p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	$f_{k/1}$	$f_{k/2}$	$\cdots$	$f_{k/p}$
	1	1	1	1

- Distribución de X condicionada por $Y = y_j \Rightarrow X/Y_{=y_j}$
- Frecuencia relativa de $X = x_i$ condicionada por $Y = y_j$

$$f_{i/j} = \frac{n_{ij}}{n_{.j}} = \frac{f_{ij}}{f_{.j}}$$

Distribuciones condicionadas: Distribución de Y condicionada por X

- La tabla se lee por filas
- Familia de k distribuciones condicionadas de la variable Y a cada valor de $X = x_1, \ldots, x_k \longrightarrow Y/X_{=x_i},\ i = 1, \ldots, k$

$X \backslash Y$	y_1	y_2	$\cdots$	y_p	
x_1	$f_{1/1}$	$f_{1/2}$	$\cdots$	$f_{1/p}$	1
x_2	$f_{2/1}$	$f_{2/2}$	$\cdots$	$f_{2/p}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	1
x_k	$f_{k/1}$	$f_{k/2}$	$\cdots$	$f_{k/p}$	1

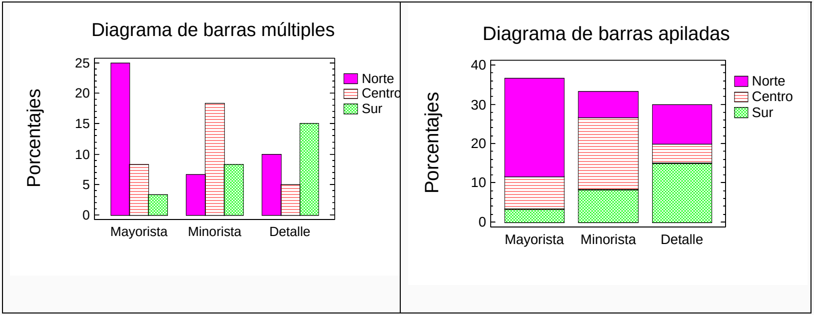
- Distribución de Y condicionada por $X = x_i \Rightarrow Y/X_{=x_i}$
- Frecuencia relativa de $Y = y_j$ condicionada por $X = x_i$

$$f_{j/i} = \frac{n_{ij}}{n_{i\,.}} = \frac{f_{ij}}{f_{i\,.}}$$

Representaciones gráficas

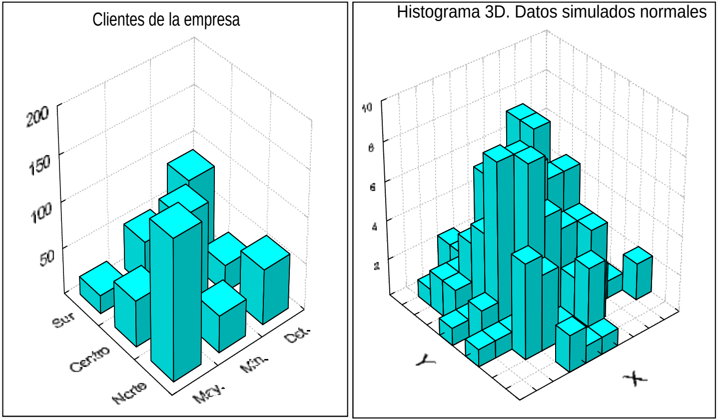
Representaciones gráficas de tablas de doble entrada

- Diagrama de barras agrupadas



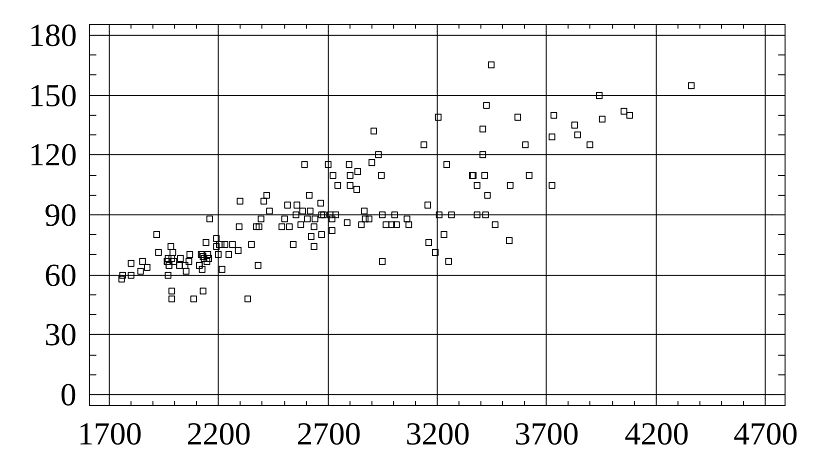
Representaciones gráficas de tablas de doble entrada

- Histograma de frecuencias tridimensional



Representaciones gráficas de datos individualizados

- Diagrama de dispersión o nube de puntos



Relación entre variables

Asociación entre variables numéricas

- En la mayoría de los problemas de interés intervienen más de una variable
- El interés principal es el estudio de las relaciones entre las variables presentes en el problema
- Suelen buscarse relaciones lineales entre las variables:
    - Es el tipo de relación más simple
    - Muchas relaciones no lineales pueden linealizarse a través de transformaciones

Asociación entre variables numéricas: Covarianza

Independencia

Diremos que dos variables son estadísticamente independientes si

$$f_{ij} = f_{i.} f_{.j} \iff n_{ij} = \frac{n_{i.} n_{.j}}{N}$$

¿Cómo medimos la dependencia?

$$\begin{aligned}
S_{XY} = Cov(X, Y) &= \frac{\sum_{i=1}^{N} (x_i - \overline{X})(y_i - \overline{Y})}{N} \\
&= \frac{\sum_{i=1}^{k} \sum_{j=1}^{p} (x_i - \overline{X})(y_i - \overline{Y}) n_{ij}}{N} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{p} (x_i - \overline{X})(y_i - \overline{Y}) f_{ij} \\
&= \frac{\sum_{i=1}^{k} \sum_{j=1}^{p} x_i y_j n_{ij}}{N} - \overline{X}\overline{Y}
\end{aligned}$$

Asociación entre variables numéricas: Covarianza

Propiedades:

- Si $S_{XY} > 0 \longrightarrow$ Existe relación positiva o creciente
- Si $S_{XY} = 0 \longrightarrow$ No existe relación
- Si $S_{XY} < 0 \longrightarrow$ Existe relación inversa o decreciente
- $S_{XY} = S_{YX}$
- $S_{XX} = S_X^2$
- $Cov(a \cdot X, c \cdot Y) = a \cdot\ c \cdot\ Cov(X, Y)$
- $Cov(b + X, Y) = Cov(X, Y)$
- $Cov(b + a \cdot X, d + c \cdot Y) = a \cdot c \cdot Cov(X, Y)$

Asociación entre variables numéricas: Covarianza

Defectos

- **Tiene unidades**: La covarianza se mide en unidades. Sin embargo, el grado de asociación entre dos variables no debería depender de las unidades en que las midamos (cambios de escala lineales). Los cambios de localización no afectan a la covarianza

- **Sólo indica el sentido** de la asociación (a través del signo).

Asociación entre variables numéricas: Coeficiente de correlación

Definición:

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

Interpretación:

- Si $|r_{XY}|$ próximo a 1 $\longrightarrow$ Asociación lineal importante
- Si $|r_{XY}|$ próximo a 0 $\longrightarrow$ Asociación lineal débil

Asociación entre variables numéricas: Coeficiente de correlación

Propiedades

- **Adimensionalidad**: No tiene unidades
- El **sentido** de asociación lineal (creciente o decreciente) lo da el signo, que es el mismo que el de la covarianza
- $-1 \leq r_{XY} \leq 1$
- **Invariancia** frente a cambios de localización u origen y escala (transformaciones lineales) en las variables

$$U = a \cdot X + b$$
$$V = c \cdot Y + d$$
$$r_{UV} = r_{XY}$$

- $r_{XY} = 0 \iff$ Ausencia de relación lineal
- $|r_{XY}| = 1 \iff$ Asociación lineal exacta

Asociación entre variables numéricas: Matriz de varianzas-covarianzas

Definición:

$$S = \begin{pmatrix} S_X^2 & S_{XY} \\ S_{YX} & S_Y^2 \end{pmatrix}$$

Propiedades:

- En la diagonal principal aparecen las varianzas
- En las diagonales secundarias aparecen las covarianzas
- Es simérica
- Es definida positiva

Asociación entre variables numéricas: Matriz de correlaciones

Definición:

$$S = \begin{pmatrix} 1 & r_{XY} \\ r_{YX} & 1 \end{pmatrix}$$

Propiedades:

- Es simérica
- Es definida positiva

Probabilidad

Curso Cero de Estadística y Probabilidad

Universidad de Salamanca

Grados en Economía y ADE

Contenidos

Introducción

Introducción

¿Cuándo se utiliza?

Utilizamos el cálculo de probabilidades cuando necesitamos obtener conclusiones generales de los resultados obtenidos en casos particulares

¿Qué nos suministra el cálculo de probabilidades?

Nos suministra la reglas para el estudio de experimentos aleatorios o de azar, constituyendo la base para la estadística inductiva o inferencial

Experimentos, sucesos y variables aleatorias

Experimentos, sucesos y variables aleatorias

Experimentos deterministas

Son aquellos que realizados de una misma forma y con las mismas condiciones iniciales, ofrecen siempre el mismo resultado

Ejemplo

Lanzar un objeto de cualquier masa (partiendo del reposo)
La velocidad de un objeto al llegar al suelo siempre es

$$v = \sqrt{2gh}$$

Experimentos, sucesos y variables aleatorias

Experimentos aleatorios

Son aquellos que realizados de una misma forma y con las mismas condiciones iniciales, no ofrecen siempre el mismo resultado

Variable aleatoria

Cualquier característica asociada a un experimento aleatorio

Ejemplo

- Juegos del azar cotidianos: Loterías, quinielas, ruletas
- Procesos de producción: Longitudes de piezas, duraciones de máquinas
- Problemas macroeconómicos: Evolución de la tasa de desempleo, del I.P.C., del déficit público
- Problemas microeconómicos: Demanda de un artículo en un establecimiento, número de clientes, volumen de ventas de un comercio
- Problemas sociológicos: Intención de voto, edad, sexo, nivel de estudios, ingresos... dentro de diferentes poblaciones o colectivos

Experimentos, sucesos y variables aleatorias

Espacio muestral

Es el conjunto de todos los posibles resultados de un experimento aleatorio

$$E = \{e_1, \ldots, e_n\}$$

Suceso aleatorio

Cualquier subconjunto del espacio muestral

$$A, B \subset E$$

Tipos de sucesos aleatorios

- **Suceso elemental**: Los formados por un sólo punto del espacio muestral
- **Suceso seguro**: Es otra forma de denominar al propio espacio muestra
- **Suceso imposible**: Aquel que nunca se verifica después del experimento, es decir, el conjunto vacío $\emptyset$

$$\emptyset \subset E \longrightarrow \emptyset$$

- **Suceso contrario o complementario**: Es el suceso que se verifica sí, como resultado del experimento aleatorio, no se verifica A

$$A \subset E \longrightarrow \overline{A} = \{e \in E : e \notin A\}$$

Experimentos, sucesos y variables aleatorias

Ejemplo

Experimento aleatorio: lanzar un dado al aire y observar el resultado

Sucesos:

- Sucesos elementales: $1, 2, 3, 4, 5, 6$
- Espacio muestral: $E = \{1, 2, 3, 4, 5, 6\}$
- Sucesos aleatorios:
  - Suceso imposible: $\emptyset$
  - Suceso seguro: E
  - Suceso complementario: $A = \{2, 4, 6\}$

$$\overline{A} = \{1, 3, 5\}$$

Operaciones con sucesos aleatorios

Operaciones con sucesos aleatorios

Unión

Dados dos sucesos aleatorios $A, B \subset E$

$$A \cup B = \{e \in E; e \in A \text{ ó } e \in B\}$$

Se da al menos uno de los dos sucesos A o B

Ejemplo

Tiramos un dado

$A = \{1, 2, 3\}$ y $B = \{3, 4\}$

$$A \cup B = \{1, 2, 3, 4\}$$

Intersección

Dados dos sucesos aleatorios $A, B \subset E$

$$A \cap B = \{e \in E; e \in A \text{ y } e \in B\}$$

Se dan simultáneamente los dos sucesos A y B

Ejemplo

Tiramos un dado

$A = \{1, 2, 3\}$ y $B = \{3, 4\}$

$$A \cap B = \{3\}$$

Operaciones con sucesos aleatorios

Diferencia

Dados dos sucesos aleatorios $A, B \subset E$

$$A - B = \{e \in E; e \in A \text{ y } e \notin B\} = A \cap \overline{B}$$

Ejemplo

Tiramos un dado

$A = \{1, 2, 3\}$ y $B = \{3, 4\}$

$$A - B = \{1, 2\}$$

$$B - A = \{4\}$$

Operaciones con sucesos aleatorios

Leyes de Morgan

Dados dos sucesos aleatorios $A, B \subset E$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Nota básica

$$B = B \cap E = B \cap (A \cup \overline{A}) = (B \cap A) \cup (B \cap \overline{A})$$

Noción de probabilidad

Noción de Probabilidad

Definición

Una probabilidad sobre un experimento aleatorio es cualquier asignación de números a los sucesos de dicho experimento, satisfaciendo las siguientes condiciones:

- A cualquier suceso A se le asigna un número, $P(A) \geq 0$
- La probabilidad del suceso seguro es $P(E) = 1$
- La probabilidad es aditiva

Noción de Probabilidad

Probabilidad axiomática

- $P(\overline{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- Si $A \subseteq B \Rightarrow P(A) \leq P(B) \Rightarrow P(B - A) = P(B \cap \overline{A}) = P(B) - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Si A y B son incompatibles, $P(A \cup B) = P(A) + P(B)$

Noción de Probabilidad

Probabilidad axiomática

- $P(A \cup B \cup C) =$
  $P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

- Si A, B y C son incompatibles, $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

Probabilidad condicionada

Probabilidad condicionada

Cuando tenemos información parcial sobre el resultado de un experimento, lo razonable es incorporar esa información y reasignar probabilidades

Definición

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Propiedades

$$P(A \cap B) = P(A)\, P(B/A)$$

$$P(A \cap B) = P(B)\, P(A/B)$$

Teoremas fundamentales del cálculo de probabilidades

Teoremas fundamentales del cálculo de probabilidades

Teorema de la probabilidad total

$$P(B) = \sum_{i=1}^{n} P(B/A_i)P(A_i)$$

Teorema de Bayes

$$P(A_k/B) = \frac{P(A_k)P(B/A_k)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$$