

Tema 1

Introducción a la Estadística

Estadística Aplicada al Turismo

Juan Miguel Navarro Ruiz

Grado en Turismo



ÍNDICE

CONTENIDOS

1. Introducción a la estadística
2. Conceptos básicos
3. Variable

Motivación

Imagen 1
*¿Qué hace
única a la
estadística?*

WHAT MAKES STATISTICS UNIQUE IS ITS ABILITY TO *QUANTIFY* UNCERTAINTY, TO MAKE IT *PRECISE*. THIS ALLOWS STATISTICIANS TO MAKE *CATEGORICAL STATEMENTS*, WITH COMPLETE ASSURANCE—ABOUT THEIR LEVEL OF UNCERTAINTY!

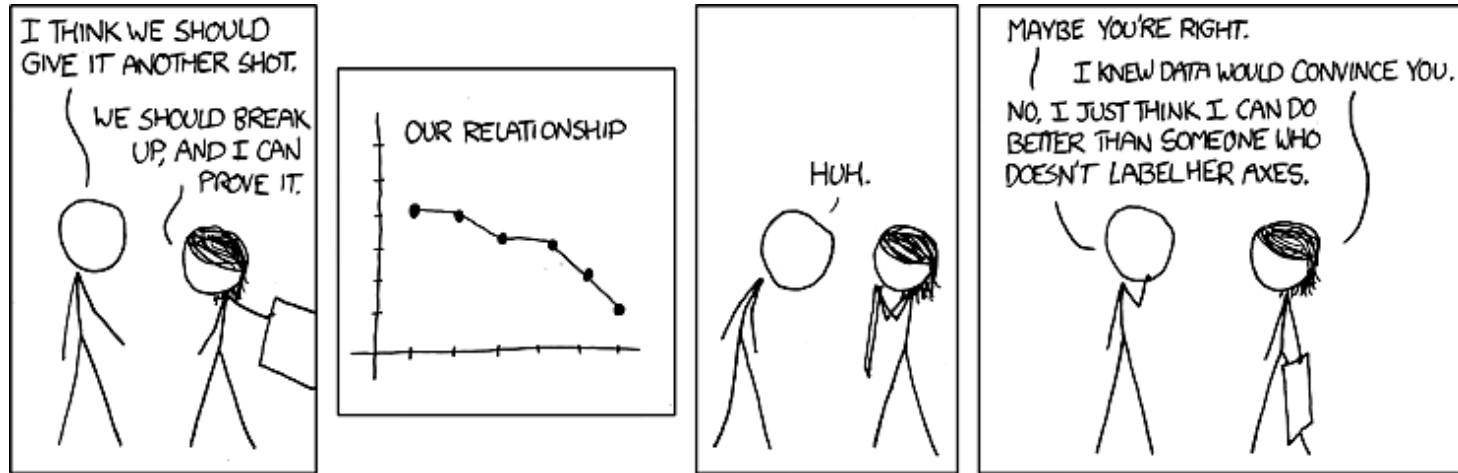


Nota. Adaptado de Título original, por autor, año, Web

Motivación

Imagen 2

El poder de los datos



Nota. Adaptado de Título original, por autor, año, Web

Imagen 3
Gráficos

1. Estadística descriptiva e inferencial

Nota. Adaptado de *Papeles de impresora blancos*, por Lukas, 2017, [Pexels](#).

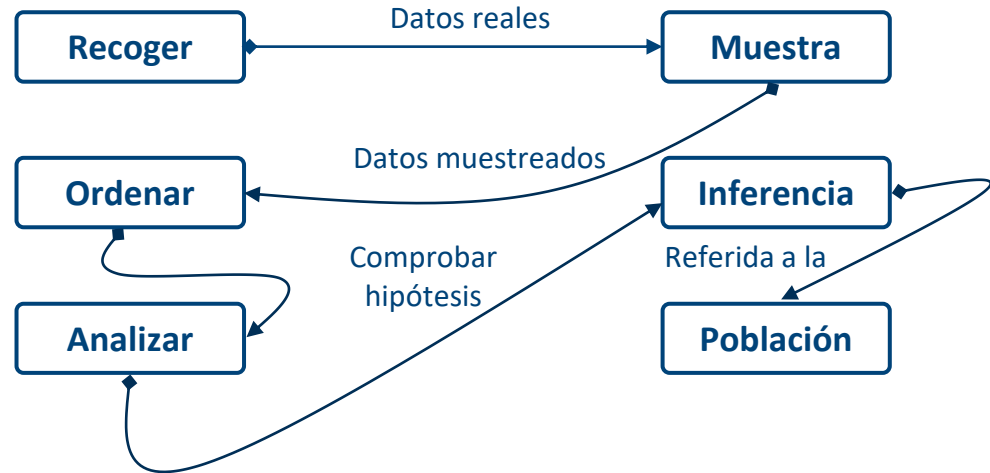
1. Estadística descriptiva e inferencial

- ✓ El análisis estadístico es un conjunto de procedimientos diseñados para **resumir**, **organizar** los datos con el objetivo de extraer **información** y elaborar **conclusiones**.
- ✓ **Todas** las ciencias empíricas hacen uso de la estadística para generar nuevo conocimiento, y en nuestra vida cotidiana continuamente hacemos uso de sus razonamiento.
- ✓ La **estadística** moderna surge de la unión de dos disciplinas independientes:
 1. El **cálculo de probabilidades** que apareció como una aproximación matemáticas a los juegos de azar.
 2. La estadística o **ciencia de estado** dedicada a llevar los registros ordenados (contar, tabular, clasificar, censar, etc.) de los datos del Estado.

1. Estadística descriptiva e inferencial

- ✓ A finales del siglo XIX surge la nueva ciencia que busca extraer conclusiones de la investigación empírica mediante el uso de modelos matemáticos.

Figura 1
Flujo informativo



- ✓ **Flujo de la información:**

Nota. La figura representa un flujo de información en el que se recoge una serie de datos que se ordenan para posteriormente ser analizados con el fin de realizar una inferencia referida a la población.

Fuente: Elaboración propia.

1. Estadística descriptiva e inferencial

1.1. Clasificación

- ✓ **Estadística descriptiva:** procedimientos diseñados para describir la información contenida en un conjunto de datos. **Resume y organiza datos.**
- ✓ **Estadística inferencial:** procedimientos que permiten generalizar (inferir, deducir) la información contenida en ese conjunto particular de datos (muestra) al conjunto de datos (población) a los que representan. **Extrae información y elabora conclusiones.**
- ✓ **Cálculo de probabilidades:** es una parte de la estadística pero debe entenderse más como una herramienta tanto para hacer estadística descriptiva como inferencial.
- ✓ Se habla también de **estadística teórica** (teoriza, propone y desarrolla métodos de análisis) y **estadística aplicada** (aplica a datos reales los métodos propuestos por la estadística teórica) = análisis de datos.

1. Estadística descriptiva e inferencial

1.2. *Utilidad del análisis de datos*

Las ciencias pueden clasificarse en dos tipos:

- ✓ **Formales o deterministas:** se parte de premisas verdaderas y por un proceso de deducción lógica se llega a conclusiones verdaderas.
 - Matemáticas, lógica y ciencias de la computación.
- ✓ **Empíricas:** se intenta ir desde lo que se observa en un conjunto reducido de observaciones hasta la afirmación de que eso es válido para el total. Por un proceso inductivo.
 - Comprenden las ciencias sociales y las ciencias naturales.

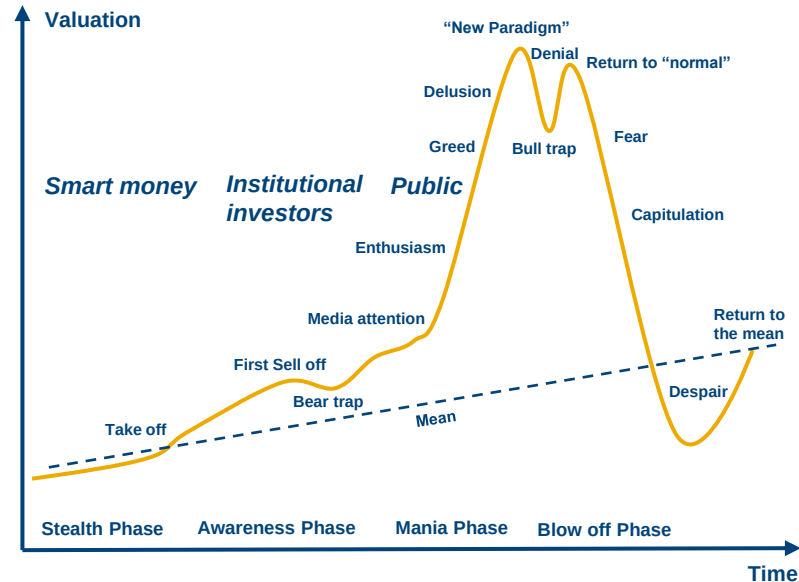
1. Estadística descriptiva e inferencial

1.2.1. Fenómenos de la realidad

- ✓ RIESGO: posibilidad de que suceda o no suceda un contratiempo.

Figura 2
Riesgo

Nota. Esta figura indica cómo se desarrolla el riesgo a lo largo del tiempo.
Fuente: Adaptado de Dr Jean-Paul Rodrigue Dept of Global Studies & Geography Hofstra University



1. Estadística descriptiva e inferencial

1.2.1. *Fenómenos de la realidad*

- ✓ INCERTIDUMBRE: falta de conocimiento seguro y claro de algo.



1. Estadística descriptiva e inferencial

1.2.1. *Fenómenos de la realidad*

- ✓ VARIABILIDAD: ciertas características de seres y procesos presentan variaciones no deterministas.



1. Estadística descriptiva e inferencial

1.2.1. *Fenómenos de la realidad*

- ✓ AZAR: situaciones, ocurrencias y casualidades que no se pueden prever de manera determinista.



1. Estadística descriptiva e inferencial

1.2.1. Fenómenos de la realidad

- ✓ SESGO: en general, alude a una distorsión, bien en la selección de sujetos para un estudio (sesgo muestral) bien en el procesamiento mental de datos (sesgo cognitivo).



1. Estadística descriptiva e inferencial

1.3. Niveles de estudio

1. **Nivel descriptivo:** se intenta obtener conocimiento sobre algo desconocido, identificar problemas de investigación y generar ideas futuras. **¿Cómo son las cosas?**
2. **Nivel relacional:** permite descubrir qué hechos se encuentran relacionados y en qué grado y además poder predecir unos hechos a partir de otros. **¿Cómo se relacionan las cosas?**
3. **Nivel explicativo:** permite establecer relaciones de tipo causal entre los eventos de manera que lo que ocurre con uno o varios de ellos puede ser explicado recurriendo a otro o varios diferentes. **¿Por qué las cosas son como son?**

1. Estadística descriptiva e inferencial

1.4. *Fases de un estudio estadístico*

1. **Planteamiento de hipótesis.** Enunciado del resultado que se quiere probar.
2. **Diseño del estudio.** Diseño del resto del proceso (variables a considerar, procedimiento de muestreo, proceso de análisis e inferencia, etc.)
3. **Muestreo.** Extracción y registro de datos reales (muestra).
4. **Descripción y resumen de los datos.** Análisis descriptivo de los datos.
5. **Inferencia sobre población. Confianza.** Análisis inferencial para validación de la hipótesis de partida.

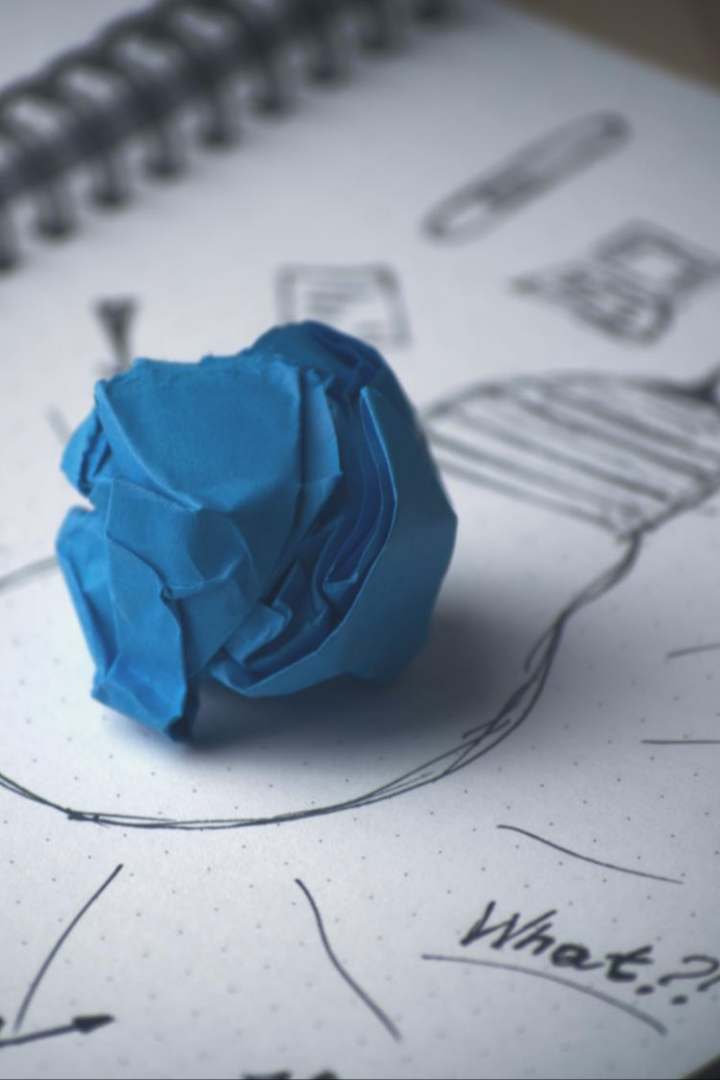


Imagen 4

Boceto de una bombilla

2. Conceptos básicos

Nota. Adaptado de *Arte boceto bombilla*, por Tookapic, 2015, [Pexels](#).

2. Conceptos básicos

- ✓ **Datos.** Elementos de información que se recopilan durante un estudio. Están directamente relacionados con el tipo de variable.
- ✓ **Conjunto de datos.** Totalidad el conjunto de datos de un estudio.
- ✓ **Variable.** Es una característica o un valor numérico que puede ser diferente de un individuo a otro (o de una unidad de análisis a otra).

2. Conceptos básicos

Tabla 1
Título tabla

Company Name / Title										
25 items ☒ Auto-Fit Columns with Data Form										
Form Data Check Sort Filters Fixed Width Columns Hide Rows & Columns Display All Fields Help Purchase / Register										
Name	Member ID	Joined	Membership	Status	Age	Gender	Marital Status	Street	City	State
Arnold Basner	C10010	2001	Gold	Active	28	Male	Married	1070 Massive Ave	Van Nuys	California
Bill Millerton	C10020	2007	Silver	Paused	28	Male	Single	1212 Congo Road	Valencia	California
Bob Franklin	C10030	2004	Bronze	Active	50	Male	Divorced	12112 Duper Road	Valencia	California
Carrie Dater	C10220	2004	Silver	Active	41	Female	Divorced	4447 Willow Drive	Orange County	California
Cathy Deleronsco	C10250	2006	Bronze	Paused	39	Female	Single	2547 Hill Street	Northridge	California
Diane Jenkins	C10180	2003	Gold	Active	38	Female	Married	3247 Zipper Street	Orange County	California
Donna Friendly	C10090	2002	Silver	Active	36	Female	Divorced	1456 Carton Street	Northridge	California
Eric Williams	C10140	2001	Gold	Paused	37	Male	Single	21607 Blender Ave	Tarzana	California
Jill Macy	C10240	2001	Silver	Active	46	Female	Married	666 Bendover Drive	Los Angeles	California
Jim Holloween	C10060	2005	Silver	Active	36	Male	Divorced	13447 Rader Lane	Westlake	California
Karen Hotchick	C10070	2005	Silver	Active	44	Female	Married	13447 Small Lane	Westlake	California
Karen Jacobsen	C10080	2008	Bronze	Active	31	Female	Divorced	1447 Willow Lane Drive	Orange County	California
Kathy Cruster	C10050	2006	Gold	Active	32	Female	Single	13334 Mills Station Road	Los Angeles	California
Larry Write	C10190	2008	Silver	Active	49	Male	Married	3264 Pappy Street	Northridge	California
Lisa Thompson	C10200	2001	Bronze	Active	37	Female	Divorced	333 Happy Street	Northridge	California
Lisa Wantssome	C10040	2001	Silver	Active	36	Female	Married	1321 Stone Creekside Road	Los Angeles	California
Mary Hadalamb	C10100	2004	Gold	Active	28	Female	Single	1607 Blender Ave	Tarzana	California
Mike Stoner	C10230	2002	Gold	Active	43	Male	Single	6321 Bend Drive	Los Angeles	California
Monica Candoit	C10110	2006	Silver	Active	25	Female	Married	166 Over Drive	Los Angeles	California
Paul Jones	C10210	2007	Bronze	Paused	43	Male	Married	33447 Toaste Lane	Westlake	California
Paula Mitchel	C10120	2001	Bronze	Paused	26	Female	Divorced	1687 Stone Road	Los Angeles	California
Peter Notoool	C10130	2001	Silver	Active	48	Male	Divorced	21000 Bloated Ave	Van Nuys	California
Scott Miller	C10170	2002	Silver	Paused	42	Male	Divorced	3212 Smile Road	Valencia	California
Todd Willinters	C10150	2004	Silver	Active	25	Male	Married	2727 Pico Contante Drive	Los Angeles	California
Tom Hasserian	C10160	2004	Bronze	Active	43	Male	Single	27527 Juicer Drive	Los Angeles	California

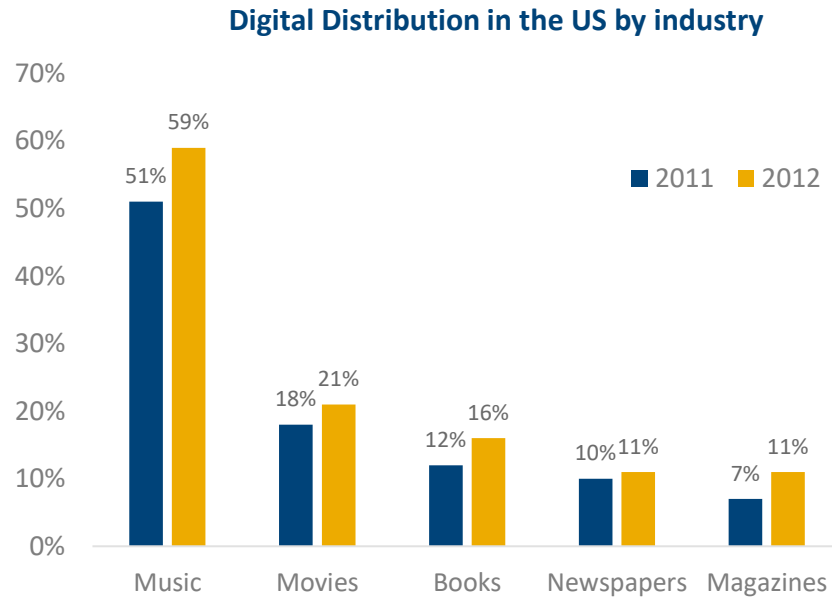
Nota. Esta tabla muestra...

Fuente:

2. Conceptos básicos

Gráfico 1

Digital Distribution in the US by industry



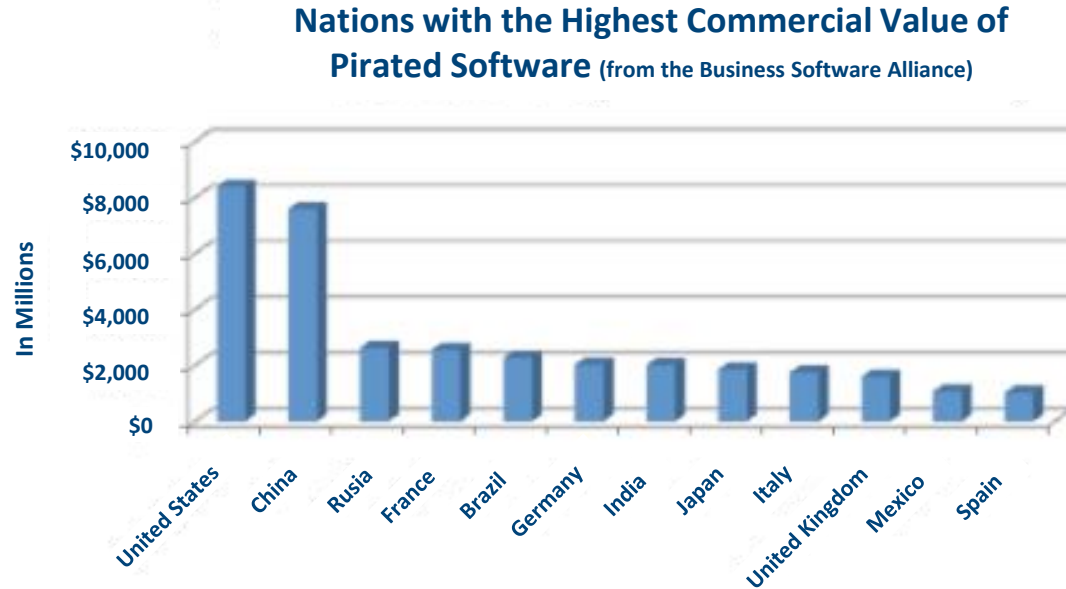
Nota. Este gráfico especifica la distribución digital en la industria estadounidense de la música, de las películas, de los libros, de los periódicos y las revistas.

Fuente: Adaptado de RIAA; RIAA analysis of PWC data.

2. Conceptos básicos

Gráfico 2

Nations with the Highest Commercial Value of Pirated Software



Nota. Este gráfico compara a los países más poderosos dentro del valor del software pirateado, donde Estados Unidos y China obtienen los valores más elevados.

Fuente: Elaboración propia.

2. Conceptos básicos

- ✓ **Población.** Conjunto concreto de unidades de análisis que interesa estudiar. Es muy importante que quede bien definida en el diseño de estudio.
- ✓ **Muestra.** Cada uno de los posibles subconjuntos de la población. Nos interesa que sea representativa y aleatoria.
- ✓ **Estadístico.** Es un número que resume los datos recopilados en una muestra. Un porcentaje, un promedio, la mediana, la desviación típica, etc. son estadísticos si se calculan sobre los datos de la muestra. Se suelen representar con caracteres latinos.
- ✓ **Parámetro.** Es el equivalente a cada estadístico pero en la población. En general son desconocidos y se estiman a partir del correspondiente estadístico muestral. Se representan con letras griegas.

2. Conceptos básicos

2.1. *Tipos de muestreo*

✓ **Algunos ejemplos:**

- Simple.
- Sistemático.
- Estratificado.

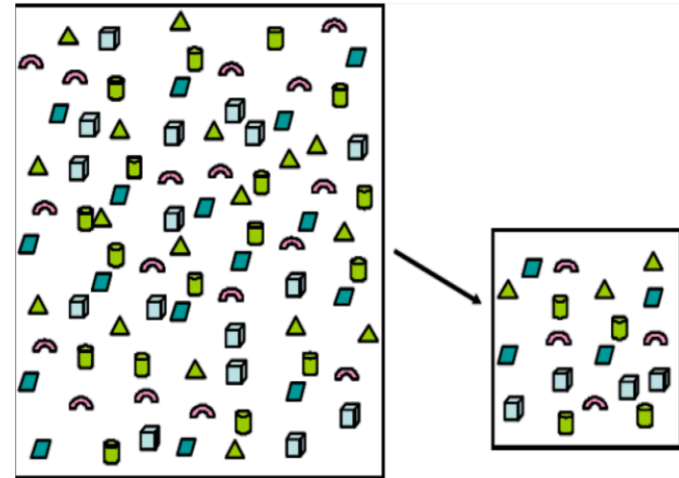
✓ Existen mas tipos.

2. Conceptos básicos

2.1. Tipos de muestreo

- ✓ SIMPLE: Selecciona muestras mediante métodos que aseguran que:
 - Cada posible muestra tiene una igual probabilidad de ser seleccionada y
 - Cada elemento de la población total tiene una oportunidad igual de ser incluido en la muestra.
- ✓ Es el más usado en la práctica.

Figura 3
Muestreo simple



Nota. La figura representa un muestreo simple, donde cada muestra tiene igual posibilidad de ser seleccionada.

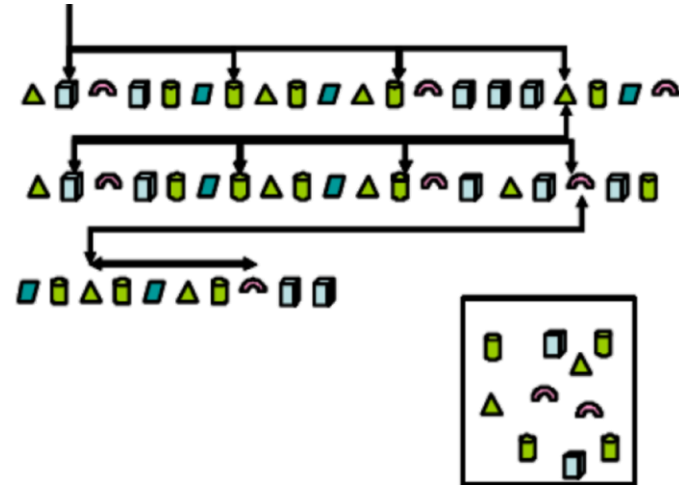
Fuente: -

2. Conceptos básicos

2.1. Tipos de muestreo

- ✓ **SISTEMÁTICO:** selecciona muestras mediante métodos sistemáticos definiendo una regla o reglas.

Figura 4
Muestreo sistemático



Nota. Esta figura simula una selección de muestras mediante métodos sistemáticos.

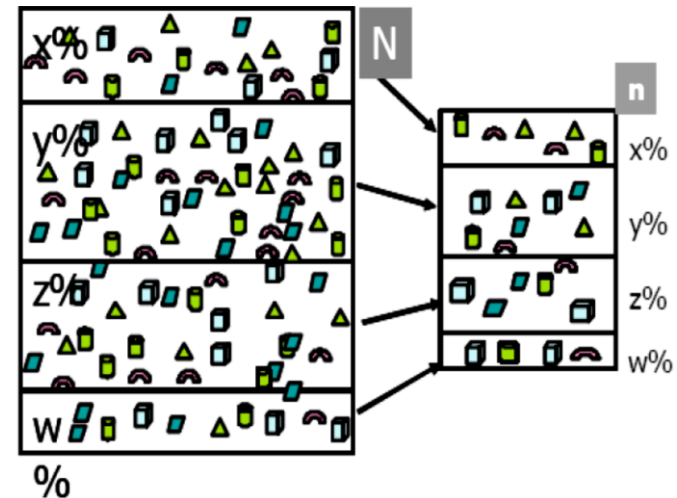
Fuente: -

2. Conceptos básicos

2.1. Tipos de muestreo

- ✓ **ESTRATIFICADO:** Selecciona muestras mediante métodos que agrupan muestras según unos criterios definidos.

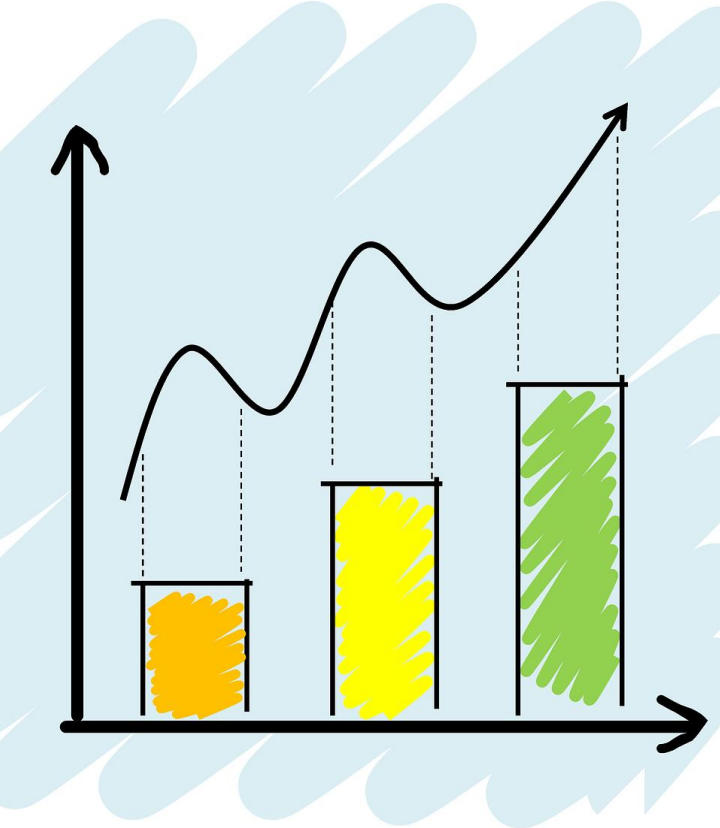
Figura 5
Muestreo estratificado



Nota. Esta figura consiste en la selección de muestras con métodos que las agrupan según criterios definidos.

Fuente: -

Imagen 5
Gráfico de análisis



3. Variable

Nota. Adaptado de *Analytics chart drawing colors*, por Tumisu, 2018, [Pixabay](#).

3. Variable

- ✓ Una variable es una característica o propiedad de cada unidad de análisis de un estudio.
- ✓ Es importante conocer en la fase de diseño de una investigación cómo se va a recoger la información y definir todas y cada una de las variables del estudio.

3. Variable

3.1. Tipos de variable

- ✓ En los estudios en los que se evalúa relación **causal**:

Variable independiente:

Aquella característica que suponemos **causa** u **origen** del fenómeno estudiado

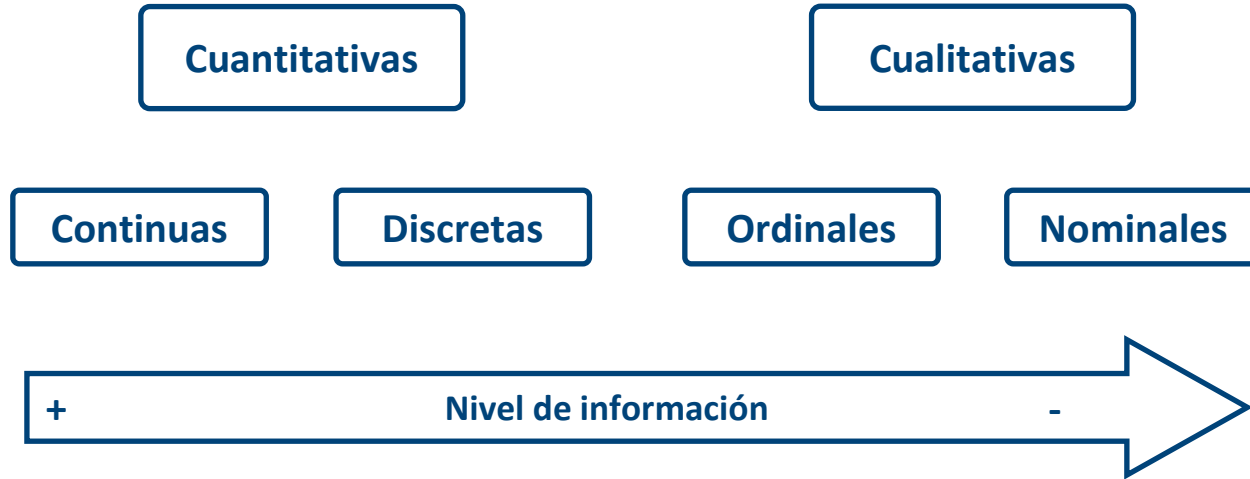
Variable dependiente:

Aquella característica que suponemos **consecuencia** o **efecto** del fenómeno estudiado

3. Variable

3.1. Tipos de variable

✓ Si nos fijamos en el **tipo de datos**:



3. Variable

3.1. *Tipos de variable*

- ✓ Una variable **cuantitativa continua** es una variable de tipo numérico que puede, en teoría, tomar cualquier valor comprendido entre dos números.
- ✓ El hecho de que registremos la variable con más o menos decimales (o sin decimales) no depende de la naturaleza de la variable, depende de la **precisión** con la que queramos registrar la información.
- ✓ **P. ej.:** Peso.

3. Variable

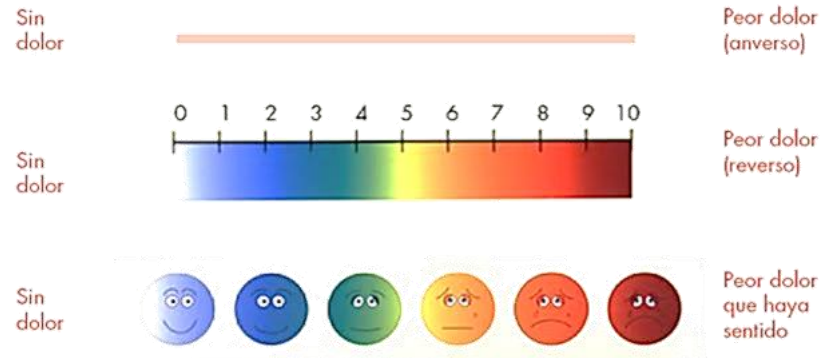
3.1. Tipos de variable

- ✓ Una variable **cuantitativa discreta** es una variable de tipo numérico sólo puede tomar números enteros porque no tiene sentido que tomara decimales.
- ✓ **P.ej.:** El número de productos vendidos. Sólo puede tomar valores enteros. Nadie va al fisioterapeuta 2,38 veces.
- ✓ Una variable **cualitativa ordinal** es una variable no mensurable, no cuantificable con varias categorías en las que cada sujeto puede clasificarse.
- ✓ Las categorías pueden **ordenarse** según algún criterio lógico.
- ✓ **P.ej.:** grande, mediano, pequeño.

3. Variable

3.1. Tipos de variable

- ✓ En ocasiones se las denomina **cuasi-cuantitativas**, en especial a aquellas que se han obtenido a través de puntuaciones en diferentes escalas: EVA, cuestionarios de calidad de vida, de disfuncionalidad, de depresión, etc.
- ✓ Desde el punto de vista operativo se suelen tratar como las cuantitativas. **Cuidado** con las interpretaciones posteriores.



3. Variable

3.1. *Tipos de variable*

- ✓ Una variable **cualitativa nominal** es una variable no mensurable, no cuantificable con varias categorías en las que cada sujeto puede clasificarse.
- ✓ Las categorías **no pueden ordenarse** según algún criterio lógico.

3. Variable

3.1. Tipos de variable

✓ Clasifica las siguientes variables según los criterios anteriores:

- Color de ojos de una muestra de alumnos

Marrón	Verde	Verde	Azul	Marrón	Azul
Marrón	Marrón	Azul	Marrón	Azul	Marrón
Verde	Marrón	Verde	Marrón	Azul	Verde
Marrón	Marrón	Azul	Marrón	Verde	Marrón
Verde	Verde	Marrón	Marrón	Marrón	Marrón
Azul					

- Estatura de los alumnos de la muestra anterior

1,48	1,56	1,56	1,59	1,60	1,61	1,63	1,64	1,64
1,67	1,68	1,68	1,68	1,68	1,69	1,70	1,71	1,72
1,72	1,75	1,76	1,76	1,77	1,77	1,79	1,81	1,81
1,84	1,84	1,88	1,94					

3. Variable

3.1. Tipos de variable

- ✓ En cuanto a la variable del color de ojos resulta obvio que no es de tipo cuantitativo, además no parece lógico establecer ningún tipo de orden en las categorías que componen las posibles respuestas de esta variable (a diferencia, por ejemplo de una variable cuyas posibles respuestas fueran: alto, medio y bajo), por tanto esta variable es de tipo *Categórica nominal*.
- ✓ Respecto a la estatura de los estudiantes, tal y como se nos presenta es una variable de tipo cuantitativo. Además, como la estatura de cualquier persona puede tomar cualquier valor (independientemente de que pueda ser redondeado) esta variable es de tipo *Cuantitativa continua*.

3. Variable

3.1. *Tipos de variable*

Clasifica las siguientes variables según el criterio que acabamos de introducir:

- ✓ Gravedad de un infarto (leve, moderado, fuerte)
- ✓ Número de ataque de asma semanales
- ✓ Sexo
- ✓ Presión arterial
- ✓ Estatura
- ✓ Peso
- ✓ Estado de dolor tras la toma de un fármaco (peor, igual, mejor)
- ✓ Provincia
- ✓ Edad
- ✓ Número de preguntas acertadas en un test
- ✓ Grupo sanguíneo

3. Variable

3.1. Tipos de variable

- ✓ La clasificación correcta de las variables sería la siguiente:

Tabla 2
Variables
cuantitativas
y cualitativas

Cuantitativas	
Continuas	Discretas
Presión arterial Estatura Peso Edad	Número de ataques de asma semanales Número de preguntas acertadas en un test
Cualitativas	
Ordinales	Nominales
Gravedad de un infarto (leve, moderado, fuerte) Estado de dolor tras la toma de un fármaco (Peor, Igual, Mejor)	Sexo Provincia Grupo sanguíneo

Nota. La tabla separa las variables cuantitativas en continuas y discretas, y las cualitativas en ordinales y nominales.
Fuente: Elaboración propia.

3. Variable

3.2. *Intervalos de clase*

- ✓ En ocasiones una variable que se ha registrado como cuantitativa, interesa **transformarla** en una variable cualitativa (en general ordinal).
- ✓ En el proceso se pierde información pero se gana en claridad.
- ✓ Se establecen puntos de corte que permiten clasificar a **todos** los sujetos en **alguna** de las nuevas categorías.
- ✓ Esas categorías reciben el nombre de **intervalos de clase**.

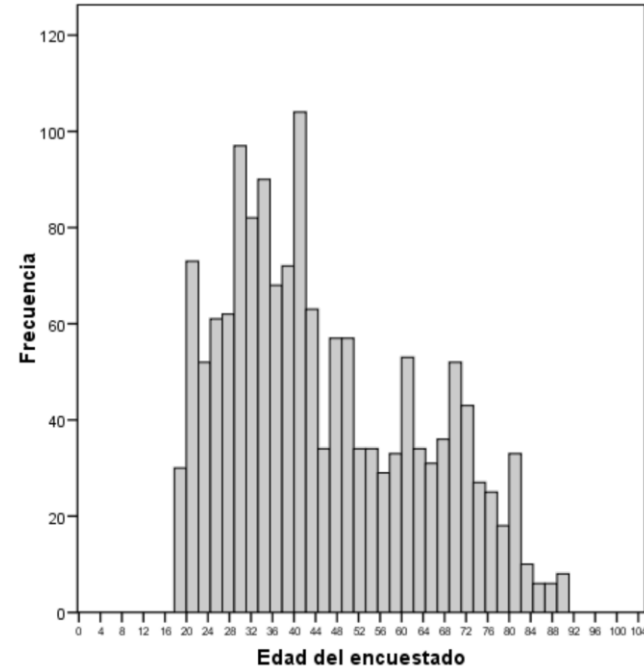
3. Variable

3.2. Intervalos de clase

- ✓ Tenemos la variable **“edad”** que es cuantitativa continua y la queremos transformar en otra variable que llamaremos **“edad categorizada”**.

Gráfico 3

Intervalos de clase



Nota. Este gráfico estudia la variable 'edad' para convertirla en otra por medio de intervalos de clase.

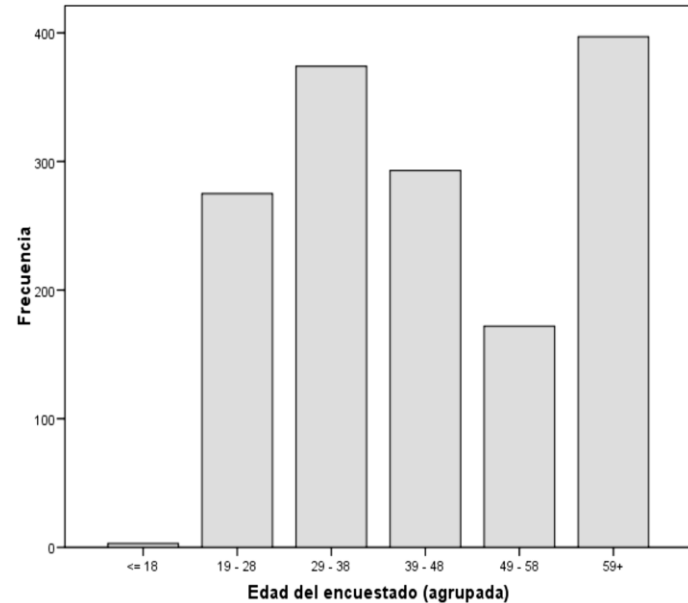
Fuente: -

3. Variable

3.2. Intervalos de clase

- ✓ En esta nueva variable hemos perdido la información particular, pero ahora tenemos clasificados a los sujetos por categorías **ordenadas de edad**.

Gráfico 4
Intervalos de clase



Nota. Este gráfico muestra una nueva clasificación por categorías ordenadas de edad.

Fuente: -

3. Variable

3.3. Etiquetado y codificación

- ✓ Cuando se está preparando la base de datos es necesario definir los nombres de las **variables (etiquetado)** y los posibles códigos de las diferentes **categorías (codificación)** para poder trabajar adecuadamente con los datos .
- ✓ No es complicado y solamente deben tener en cuenta unas **reglas muy sencillas**.
 1. El nombre de la variable no debe contener más de 8-10 caracteres alfanuméricos. Cada nombre debe ser inequívoco y único.
 2. Deben evitarse los espacios. En caso de que sea necesario introducir alguna separación visual, puede hacerse con “punto” o con “guion bajo”
 3. Evitar utilizar caracteres que pueden tener otro significado para los programas de análisis: @, %, &, /, (), #, etc.

3. Variable

3.3. Etiquetado y codificación

4. Debe definirse también el número de decimales que se quieren visualizar. Para trabajar con matrices de datos debe intentarse que sea visualmente cómoda.
5. Cuidado con los redondeos. **NUNCA** se debe redondear para hacer los cálculos, en todo caso se redondea al final. Los programas de análisis nunca redondean.
6. Debe definirse cómo se van a registrar los **valores perdidos**: perdidos por el sistema o perdidos por el usuario.

3. Variable

3.3. *Etiquetado y codificación*

EJEMPLOS ETIQUETADO CORRECTO

ALTURA

GENERO

NUM_HIJOS

INGRESOS

NACIONALIDAD

PESO

NIVEL_ESTUD

EST_CIVIL

EDAD

3. Variable

3.3. *Etiquetado y codificación*

EJEMPLOS ETIQUETADO INCORRECTO

ESTATURA_DE_LA_PERSONA

MUJER O HOMBRE

NUM_HIJ@S

NIVEL ESTUD

FECHA#NACIMIENTO

3. Variable

3.3. Etiquetado y codificación

- ✓ EJEMPLO de codificación.
- ✓ Veremos a continuación un ejemplo de codificación.
 - La variable SEXO puede tomar las categorías Hombre / Mujer que se codificarán como 1 / 2
 - La variable RAZA puede tomar las categorías Blanca / Negra / Otra que se codificarán como 1 / 2 / 3
 - La variable FELIZ puede tomar las categorías Muy feliz / Bastante feliz / No demasiado feliz / No contesta que se codificarán como 1 / 2 / 3 / 9 (9 es el código para valores perdidos).
 - La variable VIDA puede tomar las categorías Excelente / Rutinaria / No procede que se codificarán como 1 / 2 / 0 (0 código para valores perdidos).

3. Variable

3.3. Etiquetado y codificación

Tabla 3
Etiquetado y
codificación

	sexo	raza	región	feliz	vida	hermanos	hijos	edad
1	Mujer	Blanca	Nor-Este	Muy feliz	Excitante	1	2	61
2	Mujer	Blanca	Nor-Este	Bastante feliz	Excitante	2	1	32
3	Hombre	Blanca	Nor-Este	Muy feliz	No procede	2	1	35
4	Mujer	Blanca	Nor-Este	No contesta	Rutinaria	2	0	26
5	Mujer	Negra	Nor-Este	Bastante feliz	Excitante	4	0	25
6	Hombre	Negra	Nor-Este	Bastante feliz	No procede	7	5	59
7	Hombre	Negra	Nor-Este	Muy feliz	Excitante	7	3	46
8	Mujer	Negra	Nor-Este	Bastante feliz	No procede	7	4	No contesta
9	Mujer	Negra	Nor-Este	Bastante feliz	Rutinaria	7	3	57
10	Mujer	Blanca	Nor-Este	Bastante feliz	Excitante	1	2	64
11	Hombre	Blanca	Nor-Este	Bastante feliz	Excitante	6	0	72
12	Mujer	Blanca	Nor-Este	Muy feliz	No procede	2	5	67
13	Hombre	Blanca	Nor-Este	Bastante feliz	No procede	1	0	33



3. Variable

3.3. Etiquetado y codificación

	sexo	raza	región	feliz	vida	hermanos	hijos	edad
14	Hombre	Otra	Nor-Este	Bastante feliz	Rutinaria	2	1	23
15	Mujer	Blanca	Nor-Este	Bastante feliz	Rutinaria	7	1	33
16	Mujer	Blanca	Nor-Este	Muy feliz	Rutinaria	6	2	59
17	Hombre	Blanca	Nor-Este	Bastante feliz	No procede	4	1	60
18	Hombre	Blanca	Nor-Este	Muy feliz	Rutinaria	6	2	77
19	Mujer	Negra	Nor-Este	Bastante feliz	No procede	12	2	52
20	Hombre	Negra	Nor-Este	Muy feliz	Aburrida	5	1	55
21	Mujer	Negra	Nor-Este	Muy feliz	Rutinaria	2	1	37
22	Mujer	Otra	Nor-Este	Bastante feliz	No procede	7	0	45
23	Mujer	Negra	Nor-Este	No demasiado feliz	Rutinaria	4	0	34
24	Mujer	Negra	Nor-Este	No demasiado feliz	Excitante	7	3	35
25	Mujer	Negra	Nor-Este	Bastante feliz	Rutinaria	6	1	35
26	Hombre	Negra	Nor-Este	No demasiado feliz	Excitante	7	0	47

Nota. Esta tabla muestra una forma de etiquetar y codificar el sexo, la raza, la región, la felicidad, la rutina, los hermanos, los hijos y la edad de los encuestados.

Fuente: Elaboración propia.

3. Variable

3.3. Etiquetado y codificación

Tabla 4
Etiquetado y
codificación

	sexo	raza	región	feliz	vida	hermanos	hijos	edad
1	2	1	1	1	1	1	2	61
2	2	1	1	2	1	2	1	32
3	1	1	1	1	0	2	1	35
4	2	1	1	9	2	2	0	26
5	2	2	1	2	1	4	0	25
6	1	2	1	2	0	7	5	59
7	1	2	1	1	1	7	3	46
8	2	2	1	2	0	7	4	99
9	2	2	1	2	2	7	3	57
10	2	1	1	2	1	1	2	64
11	1	1	1	2	1	6	0	72
12	2	1	1	1	0	2	5	67
13	1	1	1	2	0	1	0	33



3. Variable

3.3. Etiquetado y codificación

	sexo	raza	región	feliz	vida	hermanos	hijos	edad
14	1	3	1	2	2	2	1	23
15	2	1	1	2	2	7	1	33
16	2	1	1	1	2	6	2	59
17	1	1	1	2	0	4	1	60
18	1	1	1	1	2	6	2	77
19	2	2	1	2	0	12	2	52
20	1	2	1	1	3	5	1	55
21	2	2	1	1	2	2	1	37
22	2	3	1	2	0	7	0	45
23	2	2	1	3	2	4	0	34
24	2	2	1	3	1	7	3	35
25	2	2	1	2	2	6	1	35
26	1	2	1	3	1	7	0	47

Nota. Esta tabla muestra una forma de etiquetar y codificar el sexo, la raza, la región, la felicidad, la rutina, los hermanos, los hijos y la edad de los encuestados.

Fuente: Elaboración propia.

REFERENCIAS

BIBLIOGRÁFICAS

Feinstein, A.R. (2002). *Principles of medical statistics*. Chapman & Hall/CRC.

Genbeta. (4 de diciembre de 2014). *Once cursos universitarios gratuitos sobre estadística*. <https://www.genbeta.com/web/once-cursos-universitarios-gratuitos-sobre-estadistica>

Hurley, WL., Denegar, CR., Hertel, J., Vazquez JC. (2012). *Métodos de investigación fundamentos de una práctica clínica basada en la evidencia*. Wolter Kluwer.

Pardo Merino, A., Ruiz, M.A., San Martín, R. (2009). *Análisis de datos en ciencias sociales y de la salud I*. Síntesis.

REFERENCIAS

BIBLIOGRÁFICAS

Miriadax (s.f.). *Curso-practico-bioestadística*. <https://miriadax.net/web/curso-practico-bioestadistica-r-3edicion>

Tareasplus para saber más (s.f.). *Estadística Empresarial*.
<https://www.tareasplus.com/Estadistica-Empresarial/Neumann-BS>

Tareasplus para saber más (s.f.). *Probabilidad y Estadística*.
<https://www.tareasplus.com/Curso-Probabilidad-y-Estadistica/Marcela-Gomez>

Juan Miguel Navarro Ruiz

jmnavarro@ucam.edu

UCAM Universidad Católica de Murcia



Tema 2

Análisis Descriptivo Unidimensional

Estadística Aplicada al Turismo

Juan Miguel Navarro Ruiz

Grado en Turismo



ÍNDICE

CONTENIDOS

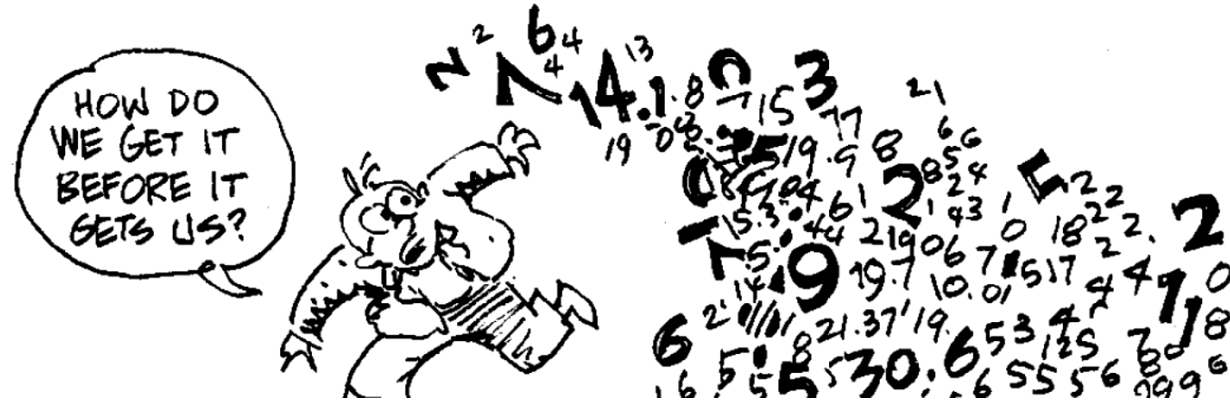
1. Análisis de datos
2. Tablas de frecuencias
3. Representaciones gráficas
4. Medidas de posición
5. Medidas de dispersión
6. Medidas de forma

Motivación

Imagen 1

Descripción de los datos

THIS CHAPTER CONCENTRATES ON DATA DESCRIPTION. HOW CAN WE REPRESENT DATA IN USEFUL WAYS? HOW CAN WE SEE UNDERLYING PATTERNS IN A HEAP OF NAKED NUMBERS? HOW CAN WE SUMMARIZE THE DATA'S BASIC SHAPE?



Nota. Adaptado de *Título original*, por autor, año, Web

Imagen 2

Datos



1. Análisis de datos

Nota. Adaptado de *Analítica datos negocio*, por Tung Nguyen, 2021, [Pixabay](#).

1. Análisis de datos

- ✓ ¿Para qué **organizamos** los datos y los representamos **gráficamente**?

Tener visión en conjunto

Facilitar cálculos posteriores y
decisiones a tomar

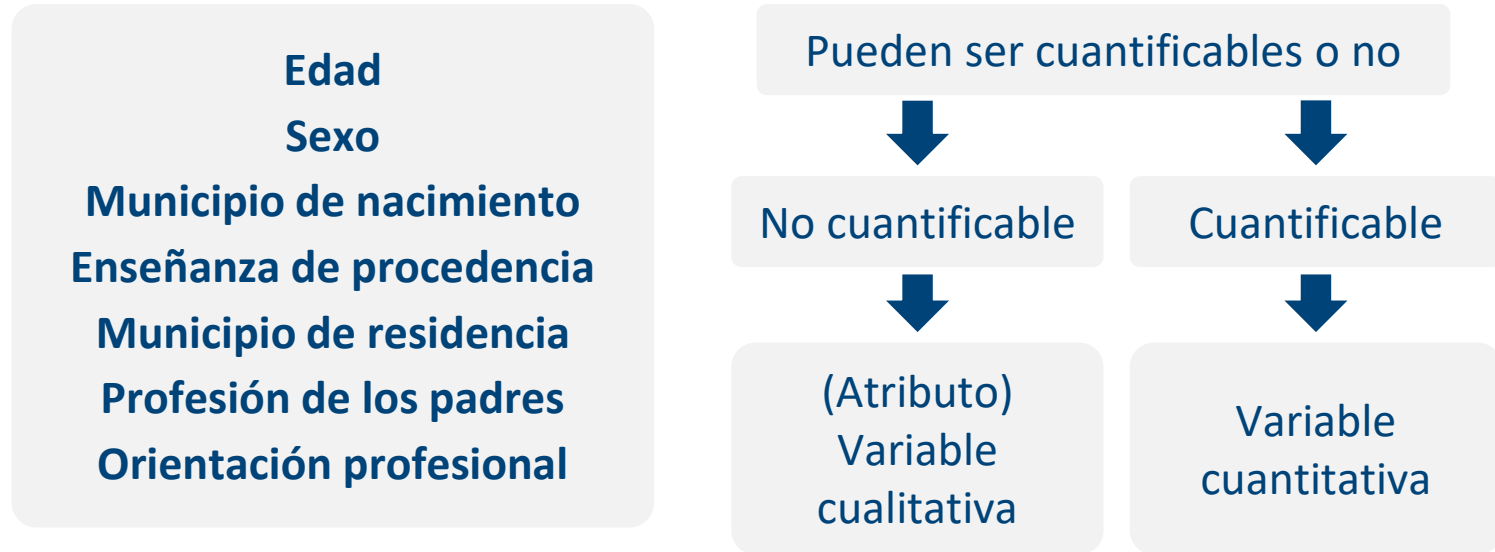
Comunicar y presentar
resultados

1. Análisis de datos

- ✓ En general, tanto si una variable es cualitativa como si es cuantitativa se deberá prestar atención tres características:
 - **El centro:** valor hacia donde tienden a agruparse las medidas.
 - **La dispersión:** grado de variabilidad de las medidas respecto a un valor concreto.
 - **La forma de distribución de los datos:** indicación de la simetría o asimetría y de la tendencia de la distribución.

1. Análisis de datos

- ✓ Es algo que nos interesa conocer de un grupo de individuos. **La indicaremos por X:**



1. Análisis de datos

Variable cualitativa (atributo). Las distintas opciones que puede tomar se denominan **modalidades**.

Variable cuantitativa. Las distintas opciones que puede tomar se denominan **valores**.

X

$X_1, \dots, X_r$

Variable cuantitativa. Las distintas opciones que puede tomar se denominan valores.

1. Análisis de datos

X = Sexo

$x_1 = \text{Varón}, x_2 = \text{Hembra}$ $r=2$

X= nº de asignaturas matriculado

$x_1=1, \dots, x_{12}=12$ $r=12$

X = edad

¿CON QUÉ MEDIDA?

1. Análisis de datos

X = Sexo

¿CON QUÉ MEDIDA?

Años cumplidos

r = finito y pequeño

Variable
aleatoria

Meses cumplidos

r = finito y mediano

Días cumplidos

r = finito y grande

Variable
continua

Horas cumplidas

r = finito y muy grande

Imagen 3
Gráficos



2. Tablas de frecuencias

Nota. Adaptado de *Papeles de impresora blancos*, por Lukas, 2017, [Pexels](#).

2. Tablas de frecuencias

- ✓ Dada una v.e. cada uno de sus valores puede aparecer un determinado número de veces.
- ✓ **Al número de veces que aparece el valor x_i se le llama frecuencia absoluta y se indicará por n_i**
- ✓ **Distribución de frecuencias $\{(x_i, n_i)\}$, $i=1, \dots, r$,**
- ✓ El conjunto formado por todos los valores que puede tomar la variable estadística junto con el número de veces (frecuencia), que aparece cada uno de ellos.

2. Tablas de frecuencias

- ✓ Una **tabla** o **distribución de frecuencias** es una forma particular de ordenar los datos basada en los valores concretos que adopta una **variable categórica** y el número de veces que se repite cada valor.
- ✓ El objetivo principal es organizar la información y **resumirla**.

Tabla 1
Tabla de frecuencias por
opinión sobre una clase

¿Que te ha parecido esta clase?					
Categorías	ni	fi	pi	Ni	Fi
Me ha parecido muy interesante	17	0,23	22,97	17	0,23
Me ha parecido bastante interesante	22	0,30	29,73	39	0,53
Me ha parecido algo interesante	29	0,39	39,19	68	0,92
No me ha parecido interesante	6	0,08	8,11	74	1,00
	74	1	100		

Nota. Esta tabla representa una tabla de frecuencias en la que se recoge la opinión del alumno acerca de la clase recibida.
Fuente: Elaboración propia.

2. Tablas de frecuencias

Concretando:

- ✓ x_i = valor (o cualidad) de la característica observada.
 - Se ordenan de menor a mayor, $i=1, \dots, r$
- ✓ n_i = número de veces que aparece x_i
- ✓ N = número total de observaciones realizadas (individuos) = $\sum n_i, i=1, \dots, r$
- ✓ f_i = frecuencia relativa de la aparición de x_i ; no es más que la proporción de veces que aparece x_i y por tanto es n_i/N
- ✓ p_i = % de veces que aparece x_i , no es más que la $f_i \times 100$

2. Tablas de frecuencias

Tabla 2
*Tabla de frecuencias
 de sexo*

Varón	Hembra	Hembra
Hembra	Varón	Hembra
Hembra	Hembra	Varón
Hembra	Varón	Hembra
Hembra	Varón	Hembra
Varón	Hembra	Varón
Hembra	Hembra	Hembra
Hembra	Varón	Hembra

Nota. Esta tabla indica frecuencias de sexos: hembra y varón.
 Fuente: Elaboración propia.

Característica: SEXO

2. Tablas de frecuencias

Tipo de variable: cualitativa (atributo)

Valores (modalidades)

$X_1 = \text{Varón}; n_1 = 8$

$X_2 = \text{Hembra}; n_2 = 16$

Distribución de frecuencias: $\{(\text{Varón}, 8); (\text{Hembra}, 16)\}$

2. Tablas de frecuencias

Tabla 3

Tabla de frecuencias que calcula el total de cada sexo

x_i	n_i	f_i	p_i
Varón	8	0,33	33,3
Hembra	16	0,67	66,7
Total	24	1	100

Nota. Esta tabla extrae el número total de hembras, varones y, en general, de personas que hay.
Fuente: Elaboración propia.

2. Tablas de frecuencias

Tabla 4
*Tabla de frecuencias de
número de miembros de
familias*

1	4	5
2	3	4
5	3	2
2	6	2
3	2	4
2	2	5
1	4	3
5	2	4

Nota. Esta tabla sintetiza el número de miembros de distintos unidades familiares.
Fuente: Elaboración propia.

Característica: nº miembros de la unidad familiar

2. Tablas de frecuencias

Tipo de variable: cuantitativa discreta

Valores

1, 2, 3, 4, 5, 6

Distribución de frecuencias: $\{(1,2) ; \dots ; (6,1)\}$

Tabla 5
*Tabla de frecuencias que
calcula el número de
miembros por unidad familiar*

x_i	n_i
1	2
2	8
3	4
4	5
5	4
6	1

Nota. Esta tabla extrae los
números más comunes de
miembros totales por
unidades familiares.
Fuente: Elaboración propia.

2. Tablas de frecuencias

Tabla 6
Tabla de
frecuencias
por número
de valores

x_i	n_i	f_i	p_i
1	2	0,083	8,3
2	8	0,333	33,33
3	4	0,167	16,7
4	5	0,208	20,8
5	4	0,167	16,7
6	1	0,042	4,2
7	24	1	100

Nota. Esta tabla de frecuencias se va numerando por valores x_i .
Fuente: Elaboración propia.

Al ser una v.e. cuantitativa cabe hablar de cuántos valores y también de...

2. Tablas de frecuencias

- ✓ $N_i = \text{nº de observaciones} \leq x_i$
- ✓ $N_i = \text{proporción de observaciones} \leq x_i$
- ✓ $P_i = \text{\% de observaciones} \leq x_i$
- ✓ Esto nos lleva a los conceptos de frecuencias acumuladas que se denominan, respectivamente, **frecuencia absoluta acumulada**, **frecuencia relativa acumulada** y **porcentaje acumulado**.

2. Tablas de frecuencias

Tabla 7
Tabla de
frecuencias
desarrollada

x_i	n_i	f_i	p_i	N_i	F_i	P_i
1	2	0,083	8,3	2	0,083	8,3
2	8	0,333	33,33	10	0,417	41,7
3	4	0,167	16,7	14	0,583	58,3
4	5	0,208	20,8	19	0,792	79,2
5	4	0,167	16,7	23	0,958	95,8
6	1	0,042	4,2	24	1	100
7	24	1	100			

Nota. Esta tabla de frecuencias desarrolla los distintos cálculos que puede realizar y recopilar.

Fuente: Elaboración propia.

2. Tablas de frecuencias

- ✓ Su cálculo es inmediato y corresponde a:

$$N_i = \sum_{j=1}^i n_j = n_1 + n_2 + \cdots + n_i$$

$$F_i = \sum_{j=1}^i f_j = f_1 + f_2 + \cdots + f_i$$

$$P_i = \sum_{j=1}^i p_j = p_1 + p_2 + \cdots + p_i$$

2. Tablas de frecuencias

- ✓ La tabla de frecuencias ofrece información sobre 3 aspectos importantes

Tabla 8
Tabla de frecuencias sobre
opinión de una clase

¿Que te ha parecido esta clase?					
Categorías	ni	fi	pi	Ni	Fi
Me ha parecido muy interesante	17	0,23	22,97	17	0,23
Me ha parecido bastante interesante	22	0,30	29,73	39	0,53
Me ha parecido algo interesante	29	0,39	39,19	68	0,92
No me ha parecido interesante	6	0,08	8,11	74	1,00
	74	1	100		

Nota. Esta tabla añade detalles a la tabla anterior en la que alumnos aportaban su juicio sobre la clase recibida.

Fuente: Elaboración propia.

2. Tablas de frecuencias

Tabla 9

Tabla de frecuencias de nivel de estudios

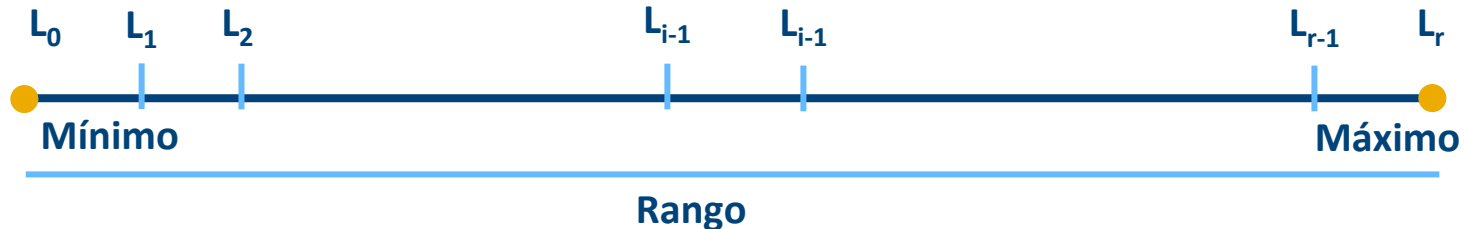
NIVEL DE ESTUDIOS						
	Fr. absolutas (n_i)	Fr. relativas f_i	Fr. porcentuales p_i	Fr. Abs. acumuladas N_i	Fr. Rel. acumuladas F_i	Fr. Porc acumuladas P_i
<i>Sin estudios</i>	5	0,01	1,0	5	0,01	1,0
<i>Primarios</i>	20	0,04	4,0	25	0,05	5,0
<i>Secundarios</i>	270	0,54	54,0	295	0,59	59,0
<i>Medios</i>	115	0,23	23,0	410	0,82	82,0
<i>Superiores</i>	90	0,18	18,0	500	1,00	100,0
Total	500	1,00	100,0			

Nota. Se recogen frecuencias acumuladas, relativas, absolutas y porcentuales del nivel de estudios.

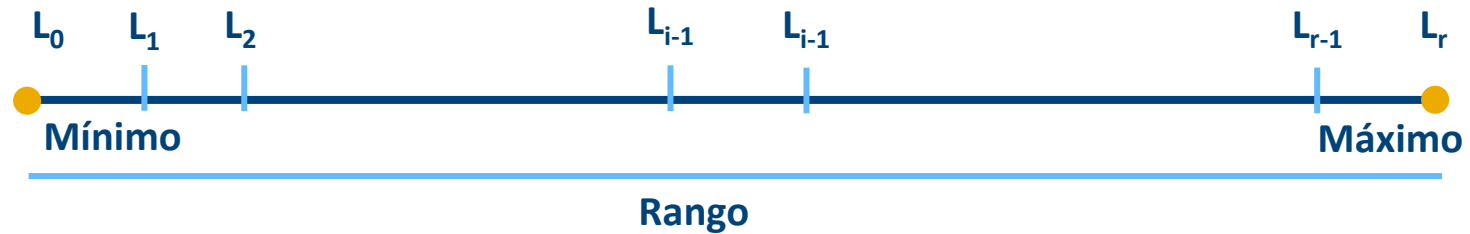
Fuente:...

2. Tablas de frecuencias

- ✓ Tipo de variable: **cuantitativa y continua**
- ✓ Valores: muchos y diferentes
- ✓ Se clasifican en **intervalos** para poder manejarlos.
- ✓ Pero los intervalos han de formar un **sistema exclusivo y excluyente**:
 - Intervalos disjuntos (excluyentes)
 - Que cubran todas las opciones (exclusivo)



2. Tablas de frecuencias



El siguiente paso es “contar” cuantos valores hay en cada intervalo (frecuencia absoluta) y con ello queda definida la **distribución de frecuencias**, pero ahora como: $\{(L_{i-1}, L_i; n_i)\}$

2. Tablas de frecuencias

Tabla 10
Tabla de
frecuencias de
calificaciones

Calificaciones de 120 exámenes							
4,34	7,19	2,85	6,55	3,49	8,93	7,62	3,63
7,53	2,22	1,59	2,74	4,74	3,28	4,16	7,05
9,39	5,59	7,23	6,08	5,63	3,43	4,29	2,33
5,49	3,85	6,76	5,95	6,15	9,73	8,39	3,18
4,96	9,42	4,95	8,01	7,16	3,24	1,98	9,70
9,87	1,48	9,44	7,66	3,88	1,34	6,33	2,52
3,24	1,20	1,49	0,30	9,41	7,05	7,56	9,74
5,59	9,07	9,51	8,53	2,57	4,62	8,07	4,66
0,52	9,68	1,66	5,21	9,64	7,57	6,85	4,06
4,88	3,51	6,64	3,19	1,96	1,23	7,35	2,85
0,98	4,71	4,96	6,29	9,42	1,40	6,65	3,72
0,88	9,15	8,21	5,84	3,33	8,58	3,63	0,13
3,65	8,50	0,78	1,87	9,30	6,30	2,37	9,30
5,37	3,58	1,47	7,05	0,08	7,90	6,11	8,00
7,61	6,52	7,73	7,98	9,43	5,55	5,59	6,87

Nota. Esta tabla contiene las calificaciones obtenidas en 120 exámenes para calcular las frecuencias.

Fuente: Elaboración propia.

2. Tablas de frecuencias

Tabla 11
Intervalos en
la tabla de
frecuencias

$L_{i-1}-L_i$	n_i	f_i	p_i
0-1	7	0,0583	5,83
1-2	12	0,10	10,00
2-3	8	0,0667	6,67
3-4	16	0,133	13,33
4-5	12	0,10	10,00
5-6	10	0,0833	8,33
6-7	13	0,1083	10,83
7-8	16	0,1333	13,33
8-9	9	0,0750	7,50
9-10	17	0,141	14,17
Sumas	120	1	100

Nota. Aquí se estudian los intervalos de las notas más frecuentes de los exámenes, vistas en la tabla anterior.

Fuente: Elaboración propia.

Consideremos los intervalos cerrados por la derecha y abiertos por la izquierda, excepto el primero que es cerrado por ambos lados.

2. Tablas de frecuencias

Tabla 12
Intervalos de
calificaciones de
exámenes

Tabla de frecuencias completa						
$L_{i-1}-L_i$	n_i	f_i	p_i	N_i	F_i	P
0-1	7	0,0583	5,83	7	0,058	5,8
1-2	12	0,10	10,00	19	0,158	15,8
2-3	8	0,0667	6,67	27	0,225	22,5
3-4	16	0,133	13,33	43	0,358	35,8
4-5	12	0,10	10,00	55	0,458	45,8
5-6	10	0,0833	8,33	65	0,542	54,2
6-7	13	0,1083	10,83	78	0,650	65,0
7-8	16	0,1333	13,33	94	0,783	78,3
8-9	9	0,0750	7,50	103	0,858	85,8
9-10	17	0,141	14,17	120	1	100
TOTAL	120	1	100			

Nota. Esta tabla completa la tabla anterior.
Fuente: Elaboración propia.

2. Tablas de frecuencias

La introducción de los intervalos no permite realizar cálculos con los valores de la variable. Se hace preciso “discretizar” la variable continua para poder trabajar con ella. Para eso se introduce el concepto de **marca de clase**, que no es más que sustituir cada intervalo por su centro.

De este modo **sustituiremos**, a efectos de posteriores cálculos y cuando sea preciso, cada intervalo $L_{i-1}-L_i$ por su “representante” X_i

$$x_i = \frac{L_{i-1} + L_i}{2}$$

En cuanto a la diferencia $L_i - L_{i-1} = a_i$ la denominaremos **amplitud** del intervalo i-ésimo

2. Tablas de frecuencias

- ✓ Con lo cual la tabla de frecuencias anterior, quedaría como sigue y en su tratamiento posterior la variable se va a tratar, a efectos de cálculo, como discreta.

Tabla 13
Tabla 12
completa

$L_{i-1}-L_i$	X_i	n_i	f_i	p_i	N_i	F_i	P_i
0-1	0,5	7	0,05	5,8	7	0,05	5,8
1-2	1,5	12	0,10	10,0	19	0,15	15,8
2-3	2,5	8	0,06	6,7	27	0,22	22,5
3-4	3,5	16	0,13	13,3	43	0,35	35,8
4-5	4,5	12	0,10	10,0	55	0,45	45,8
5-6	5,5	10	0,08	8,3	65	0,54	54,2
6-7	6,5	13	0,10	10,8	78	0,65	65,0
7-8	7,5	16	0,13	13,3	94	0,78	78,3
8-9	8,5	9	0,07	7,5	103	0,85	85,8
9-10	9,5	17	0,14	14,2	120	1	100
TOTAL		120	1	100			

Nota. Esta tabla termina de completar las tablas 11 y 12.
Fuente: Elaboración propia.



Imagen 4
Gráficos

3. Representaciones gráficas

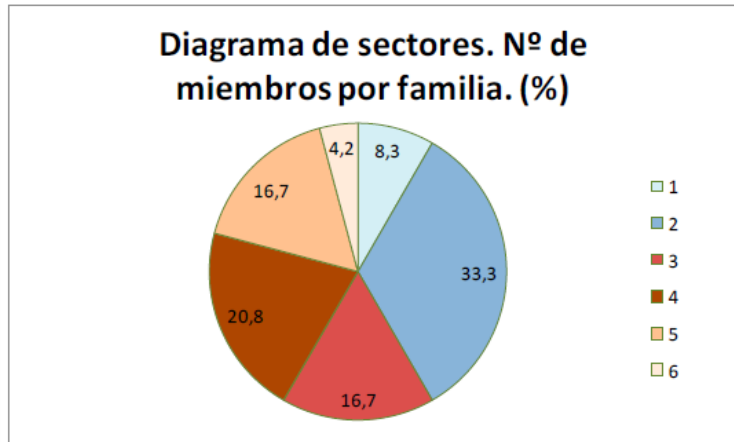
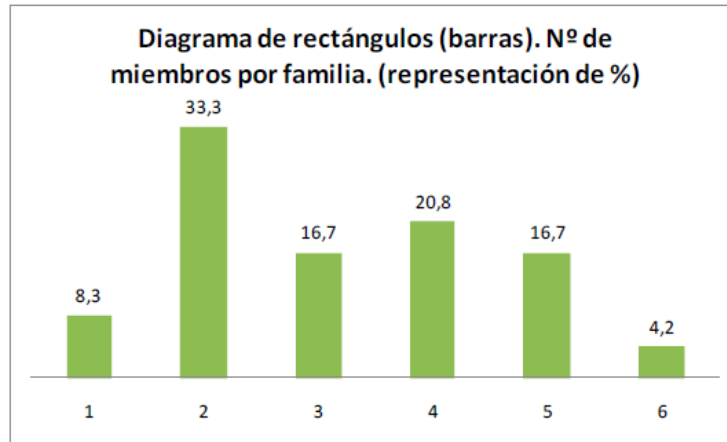
Nota. Adaptado de *Papeles de impresora blancos*, por Lukas, 2017, [Pixabay](#).

3. Representaciones gráficas

- ✓ En el análisis descriptivo **conviene** hacer representaciones **gráficas** de las tablas de frecuencias para hacerse una idea rápida de lo que está ocurriendo.
- ✓ Los gráficos adecuados para representar variables cualitativas son los **gráficos de barras** y los **diagramas de sectores**.
- ✓ Para las frecuencias acumuladas se pueden usar **gráficos de barras** y los **gráficos de líneas**.

3. Representaciones gráficas

✓ DIAGRAMAS DE RECTANGULOS Y/O SECTORES



✓ Válidos para variables cualitativas y cuantitativas discretas.

3. Representaciones gráficas

Correcto diseño - barras

Gráfico 1
Tabaquismo

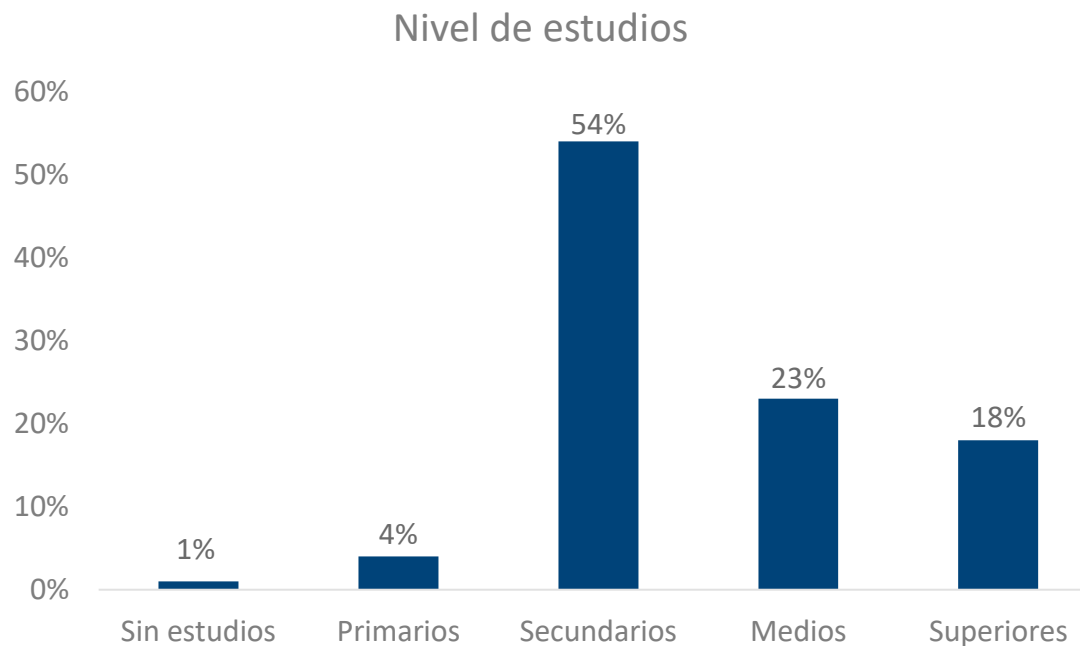


Nota. Este gráfico extrae el porcentaje de fumadores, exfumadores y no fumadores que hay.

Fuente: Elaboración propia.

3. Representaciones gráficas

Gráfico 2
Nivel de estudios

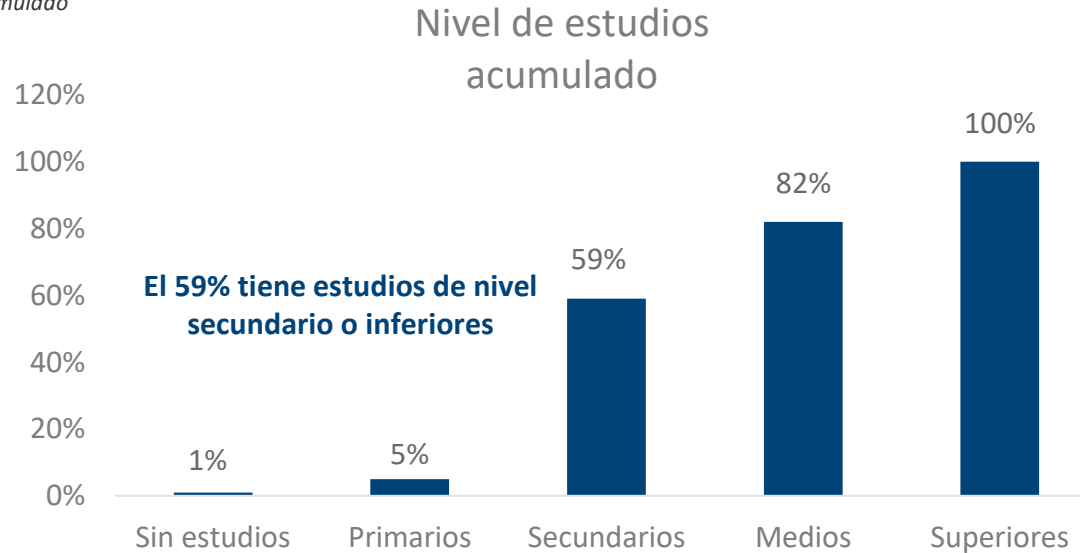


Nota. Este gráfico indica el nivel de estudios que habría por medio de porcentajes: sin estudios, primarios, secundarios, medios y superiores.
Fuente: Elaboración propia.

3. Representaciones gráficas

Barras

Gráfico 3

Nivel de estudios acumulado

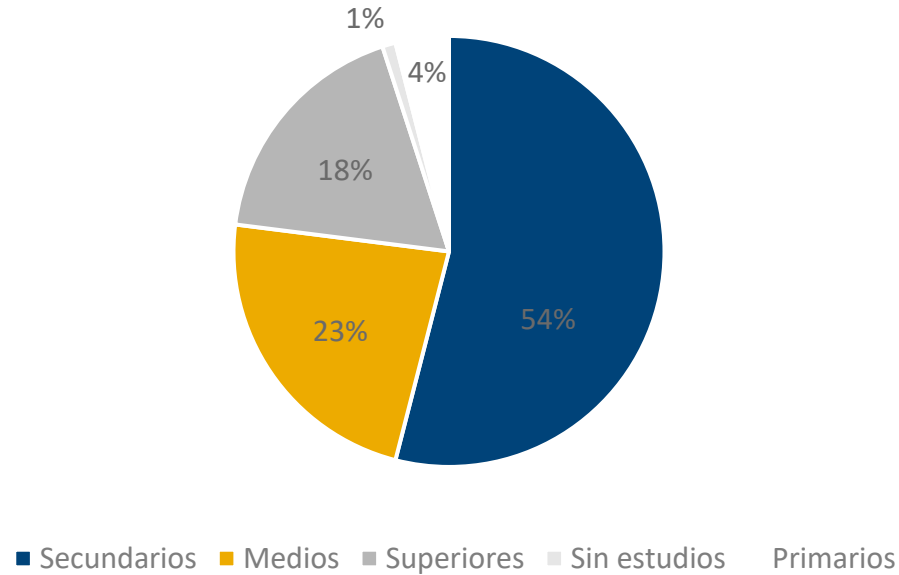
Nota. Este gráfico indica el nivel de estudios acumulados que habría por medio de porcentajes: sin estudios, primarios, secundarios, medios y superiores.

Fuente: Elaboración propia.

3. Representaciones gráficas

Sectores

Gráfico 4
Nivel de estudios por sectores



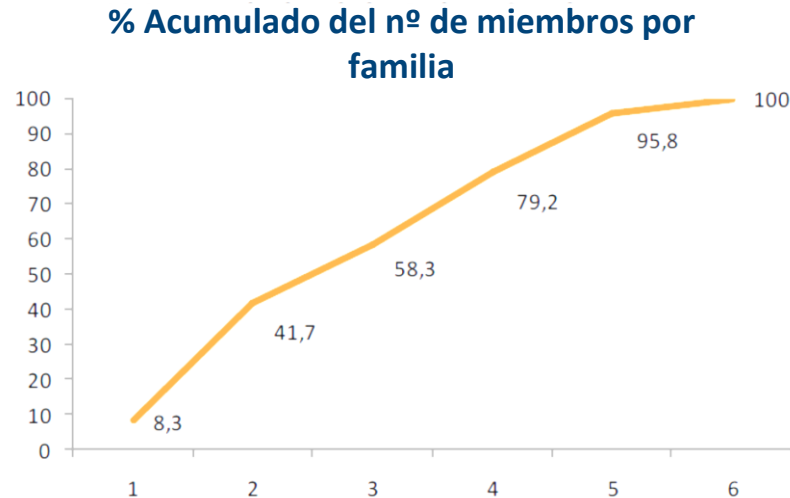
Nota. Este gráfico indica el nivel de estudios que habría por medio de porcentajes y en sectores sin estudios, primarios, secundarios, medios y superiores.
Fuente: Elaboración propia.

3. Representaciones gráficas

Curvas acumuladas. También se puede con barras

- ✓ CURVAS ACUMULATIVAS. Para frecuencias acumuladas.

Gráfico 5
% Acumulado del nº de miembros por familia



Nota. Esta figura muestra el porcentaje acumulado de número de miembros por familia, desde 8,3 a 100.
Fuente: Elaboración propia.

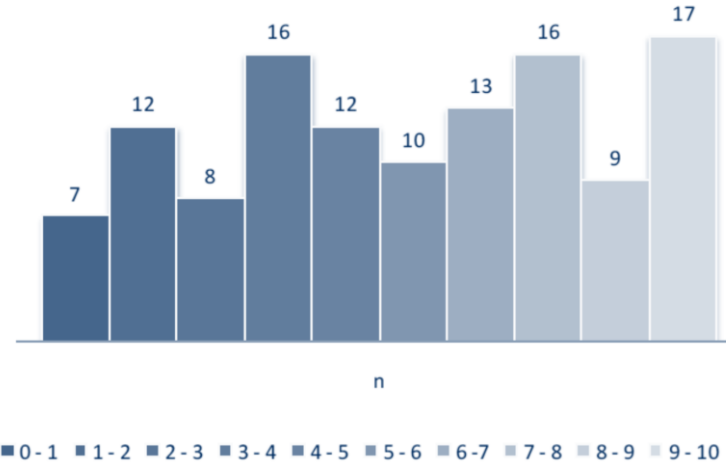
- ✓ Solo tiene sentido para v.e. cuantitativas discretas.

3. Representaciones gráficas

Histograma para continuas

✓ Histograma: calificaciones de examen

Gráfico 6

Calificaciones de examen

Nota. Se recogen las diferentes calificaciones de un examen por medio de un histograma.

Fuente: ...

✓ Casos de variables cuantitativas continuas

3. Representaciones gráficas

Histograma para continuas

- ✓ HISTOGRAMA ACUMULATIVO: Para frecuencias acumuladas.

Gráfico 7

Calificaciones de examen

Nota. Se recogen las diferentes calificaciones de un examen por medio de un histograma acumulativo.

Fuente: ...

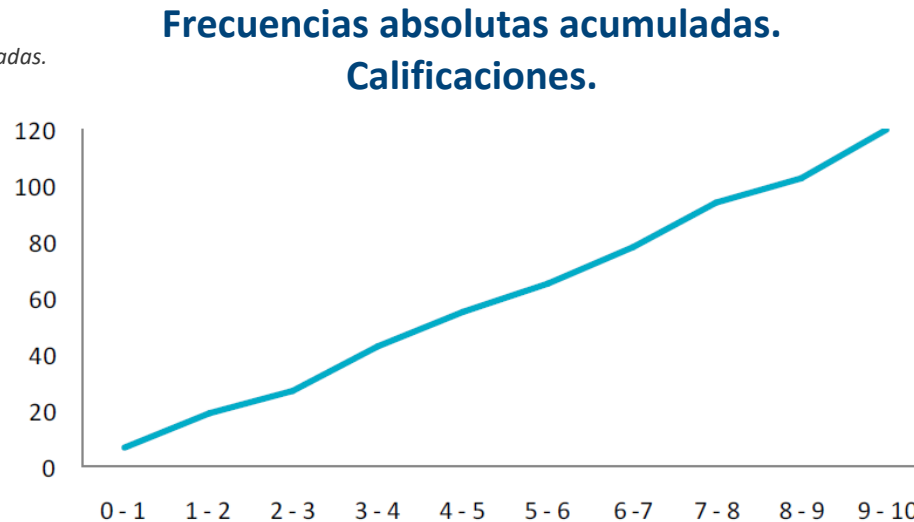


3. Representaciones gráficas

✓ CURVA ACUMULATIVA: Para frecuencias acumuladas.

Gráfico 8
*Frecuencias absolutas acumuladas.
Calificaciones.*

Nota. Se recogen las diferentes calificaciones de un examen por medio de un frecuencias absolutas acumuladas representadas en una curva ascendente..
Fuente: ...



3. Representaciones gráficas

Errores de representación

Gráfico 9
Registro de desempleo

Nota. Se especifica el registro de desempleo por medio de un eje vertical cortado para exagerar las diferencias.

Fuente: Ministerio de Empleo y Seguridad Social, 2014.



Medidas de los datos

- ✓ **OBJETIVO:** Buscar indicadores que representen y sintetizen el conjunto de datos.
- ✓ **Sintetizar:** refundir toda la información en un solo indicador.
- ✓ **Representar:** conocer hasta qué punto ese indicador representa al conjunto de datos.
- ✓ Hacen falta tres clases de indicadores:
 - **De posición,** es decir indicadores que informen y representen al conjunto de datos
 - **De dispersión:** que informen o bien de la representatividad de los anteriores, o bien nos informe del grado de acumulación o dispersión de los datos y por lo tanto de su grado de variabilidad.
 - **De forma:** nos describen la forma que tiene la distribución de los datos.

Medidas de los datos

- ✓ Existen más herramientas para describir el conjunto de valores que toma una variable cuantitativa que cualitativa (aunque siempre se puede categorizar).

Medidas de centralización

Media
Mediana
Moda

Medidas de posición

Cuantiles
Percentiles
Deciles
Cuartiles

Medidas de dispersión

Rango
Varianza
Desviación típica
Coeficiente de variación

Medidas de formas

Asimetría (skewness)
Apuntamiento (kurtosis)

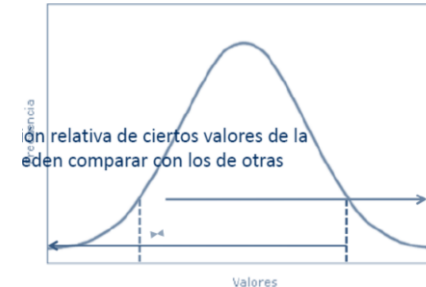
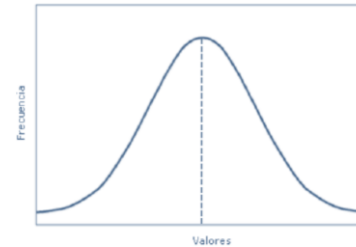


Imagen 6

Escuadra y pizarra



4. Medidas de posición

Nota. Adaptado de *Distancia de los puntos*, por Dids, 2019, [Pexels](#).

4. Medidas de posición

Centralización

- ✓ Medidas de descentralización
 - **Indicadores de hacia qué valores tienden las medidas a agruparse.**

Media

Mediana

Moda

4. Medidas de posición

Centralización

- ✓ **MEDIA ARITMÉTICA:** Es el promedio de todos los datos obtenidos. Al producirse repeticiones de valores (frecuencias), su cálculo se realiza como:

$$\bar{X} = \frac{\sum_1^r x_i n_i}{N} = \sum_1^r x_i f_i$$

- ✓ La frecuencia puede interpretarse como la “importancia” que tiene cada valor, dentro del conjunto de datos.

4. Medidas de posición

Centralización

De manera tradicional, teniendo todos los datos:

✓ Cálculo de la **media**:

- 1º Suma de todos los valores
- 2º Dividir entre el número de valores (tamaño muestral).
- **$(46 + 5 + \dots + 78)/20 = 1308/20 = 65,4$**

Tabla 14
Media de peso

	Peso		Peso
1	46	11	67
2	54	12	67
3	58	13	67
4	58	14	68
5	60	15	72
6	61	16	73
7	62	17	73
8	62	18	76
9	63	19	77
10	66	20	78

Nota. A partir de estas tablas se presentan los diferentes pesos.
Fuente: Elaboración propia.

4. Medidas de posición

Centralización

- ✓ Teniendo la tabla de frecuencias. La ordenación de cálculos para la media es muy simple, basta construir lo siguiente sobre la tabla de frecuencias:

Tabla 15
Cálculo de media

x_i	n_i	$x_i n_i$
1	2	2
2	8	16
3	4	12
4	5	20
5	4	20
6	1	6
TOTAL	24	76
	MEDIA	3,17

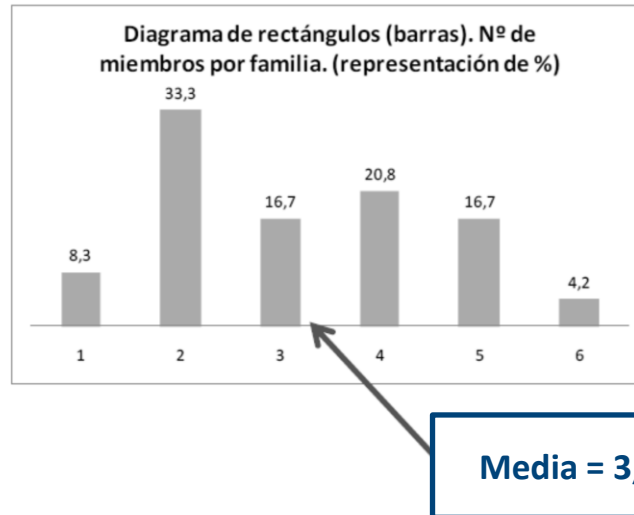
Nota. Esta tabla muestra cómo se calcula fácilmente la media por medio de una tabla de frecuencias.
Fuente: Elaboración propia.

$$\bar{X} = \frac{\sum_{i=1}^N x_i n_i}{N} = \frac{76}{24} = 3,17$$

4. Medidas de posición

Centralización

- ✓ **MEDIA ARITMÉTICA:** al ser un promedio que combina valores con frecuencias, la resultante del cálculo puede interpretarse como el **centro de la distribución de frecuencias**. Y de hecho es el equivalente a su gravedad, hablando en términos físicos.



4. Medidas de posición

Centralización

- ✓ Una cuestión a tener siempre en cuenta es hasta qué punto la media representa adecuadamente el conjunto de datos, además de ser su centro. Dicho de otro modo, para ser un buen indicador debe estar suficientemente cerca de todos los valores de la variable. Esto supone introducir un concepto adicional: la **representatividad de la media aritmética**.
- ✓ Si 5 personas tienen un sueldo de 800 euros y una sexta persona tiene un salario de 5.000 euros mensuales. El salario medio del conjunto es de 1.500 euros. ¿Es un buen indicador de la distribución de los salarios de esa pyme?
- ✓ Si por el contrario las primeras 5 personas mantuvieran su salario de 800 euros y la sexta tuviera de 1.200 la media sería de 866,6. ¿Es mejor representante que en el caso anterior?

4. Medidas de posición

Centralización – Propiedad de la media

- ✓ Una propiedad útil de la media. ¿Qué ocurre si hacemos un cambio de variable?
- ✓ *Por ejemplo si tenemos una distribución de salarios en euros y queremos convertirla a dólares. ¿Qué ocurre con la media? ¿Hay que calcularla de nuevo?*

$$Y = a + bX$$



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

$$\bar{Y} = \frac{\sum_q^r y_i n_i}{N} = \frac{\sum_1^r (a + bx_i) n_i}{N} = \frac{\sum_1^r a n_i}{N} + \frac{\sum_1^r b x_i n_i}{N} = \frac{a \sum_1^r n_i}{N} + \frac{b \sum_1^r x_i n_i}{N} = \frac{aN}{N} + b\bar{X} = a + b\bar{X}$$

4. Medidas de posición

Centralización

- ✓ **MEDIANA:** es aquel valor que es superado por al menos el 50% de las observaciones y que a la vez, supera al menos al 50% de las observaciones. A efectos prácticos es el valor que se corresponde con una frecuencia acumulada del 50% cuando ese valor existe.
- ✓ En el **caso discreto** se puede obtener ordenando todos los valores de menor a mayor, tantas veces como la frecuencia tengan, y designando como mediana el que ocupe el lugar central (o la media de los dos valores centrales, en su caso).
- ✓ En el **caso continuo** hay que calcular en primer lugar el intervalo mediano, que es primer en que el valor de la frecuencia acumulada es al menos del 50%. Dentro de él estará la mediana y su expresión se obtiene mediante un cálculo geométrico.

4. Medidas de posición

Centralización

De manera tradicional con todos los datos:

✓ Cálculo de la **mediana**:

- 1º Ordenación de los datos
- 2º Localización del valor que deja a un lado y el otro mismo numero de casos
- Si hay dos valores centrales se calcula la media de los dos
- **$(66+67)/2=66,5$** ←
- Si hubiera solo 19 valores => Sería el valor 10 => el **66**
- Si hubiera 21 valores => Sería el valor 11 => el **67**

Tabla 16
Cálculo de mediana

	Peso
1	46
2	54
3	58
4	58
5	60
6	61
7	62
8	62
9	63
10	66
11	67

Nota. Esta tabla expone cómo se calcula la mediana del peso en este caso.

Fuente: Elaboración propia.

4. Medidas de posición

Centralización

Tabla 17
Cálculo de
mediana
usando la tabla
de frecuencias
(método
recomendado
en esta
asignatura)

Mediana: $Me=3$						
x_i	n_i	f_i	p_i	N_i	F_i	P
1	2	0,1	8,3	2	0,1	8,3
2	8	0,3	3,33	10	0,4	41,7
3	4	0,2	16,7	14	0,6	58,3
4	5	0,2	20,8	19	0,78	79,2
5	4	0,2	16,7	23	1,0	95,8
6	1	0,0	4,2	24	1	100
TOTAL	24	1	100			

Nota. Esta tabla muestra el resultado de calcular la mediana en la tabla de frecuencias.

Fuente: Elaboración propia.

4. Medidas de posición

Tabla 18
Intervalo
mediano

$L_{i-1}-L_i$	n_i	f_i	p_i
0-1	7	0,058	5,83
1-2	19	0,158	15,8
2-3	27	0,225	22,5
3-4	43	0,358	35,8
4-5	55	0,458	45,8
5-6	65	0,542	54,2
6-7	78	0,65	65,0
7-8	94	0,783	78,3
8-9	103	0,858	85,8
9-10	120	1	100

Nota. Esta tabla señala qué sería el intervalo mediano dentro de una tabla de frecuencias.

Fuente: Elaboración propia.

Intervalo mediano: 5-6

4. Medidas de posición

Centralización

- ✓ **MODA:** es **el valor que más se repite** o que más veces aparece. Puede utilizarse tanto para variables cualitativas como cuantitativas.
- ✓ En el caso de v.e. cualitativas o de cuantitativas discretas es simplemente el valor que presente mayor frecuencia. Esto supone que una distribución de frecuencias no tiene por qué tener una sola moda o valor modal (distribución unimodal), sino que puede ocurrir que tenga varias modas (distribución multimodal).
- ✓ Cuando se trata de v.e. cuantitativas continuas se hablará de intervalos modal, que será aquel que presente mayor frecuencia.

4. Medidas de posición

Centralización

Método tradicional:

- ✓ Cálculo de la **moda**:
 - 1º Localizar la observación que se presenta con más frecuencia
 - **67 (frecuencia absoluta 3, frecuencia relativa 15%)**

Tabla 19
Cálculo de moda

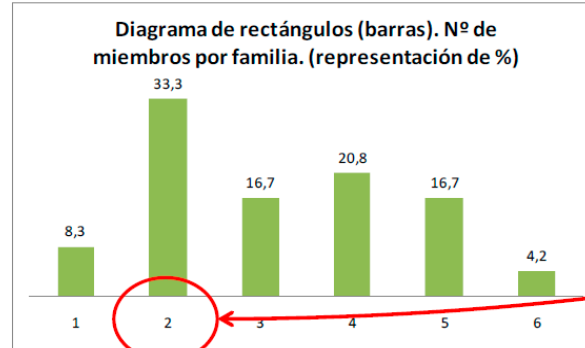
	Peso
1	46
2	54
3	58
4	58
5	60
6	61
7	62
8	62
9	63
10	66
11	67
12	67
13	67

Nota. Esta tabla enseña cómo se calcularía la moda del peso en este caso.

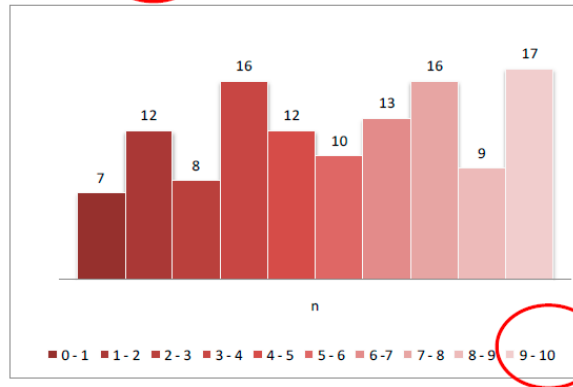
Fuente: Elaboración propia.

4. Medidas de posición

Método recomendado. Gráfica o tabla de frecuencias



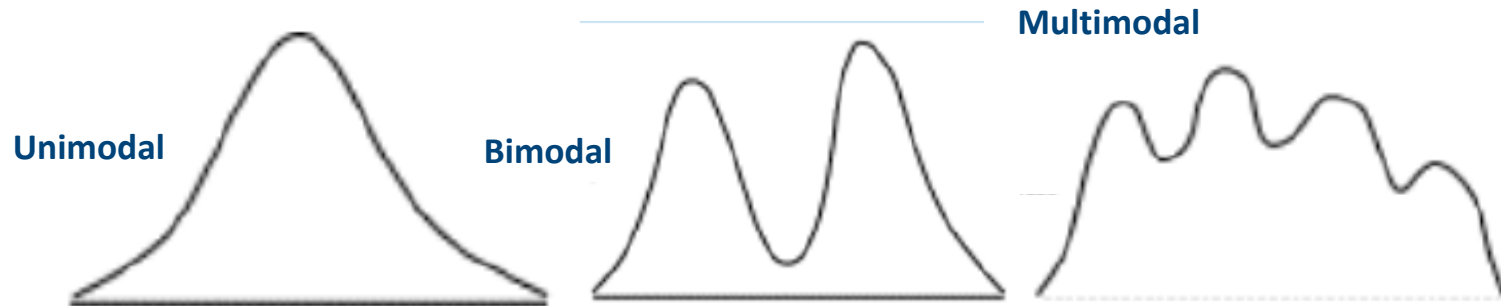
MODA: $M_o=2$



INTERVALO MODAL

4. Medidas de posición

- ✓ El valor que más veces aparece repetido en la distribución (con mayor frecuencia)
 - Pueden existir distribuciones con más de una moda (bimodales, multimodales, etc.)



4. Medidas de posición

Otras posiciones

Cuantiles (de 0 a 1)

Cuantil de orden α es un valor de la variable por debajo del cual se encuentra una frecuencia acumulada de α .

Percentiles (de 0 a 100)

Percentil de orden k es el cuantil de orden $\alpha=k/100$. Tiene el mismo significado que el cuantil, pero se expresa en %.

Deciles (de 0 a 10)

Decil de orden k es el cuantil de orden $\alpha=k/10$. Tiene el mismo significado que el cuantil, pero en escala de 0 a 10.

Cuartiles (de 1 a 3)

Los **cuartiles** dividen a la muestra en cuatro grupos iguales.

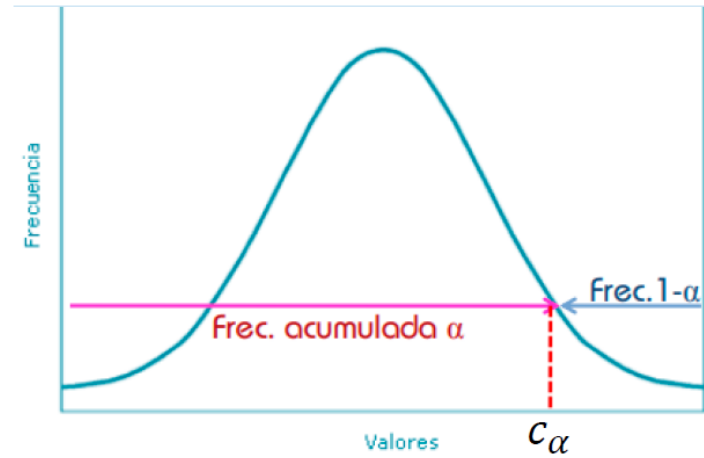
4. Medidas de posición

Otras posiciones

Cuantiles (de 0 a 1)

Cuantil de orden α , c_α es un valor de la variable por debajo del cual se encuentra una frecuencia acumulada de α .

Gráfico 10
Cuantil



Nota. El gráfico representa el cuantil alfa de la frecuencia acumulada alfa.
Fuente: ...

4. Medidas de posición

Otras posiciones

Percentiles (de 0 a 100)

$$\alpha = k/100$$

- ✓ **Percentil** de orden K es el cuantil de orden $\alpha = k/100$. Tiene el mismo significado que el cuantil, pero se expresa en %
- ✓ **Percentil** de orden K es el valor por debajo del cual se encuentra una frecuencia acumulada del k %
- ✓ El **percentil 50** es el valor de la variable que deja a un lado y a otro el mismo número de casos = **MEDIANA**
- ✓ *Para calcularlos es necesario ordenar los datos (de menor a mayor o al revés)*

4. Medidas de posición

✓ ¿Cómo sacar el percentil 90?

1. Hallar el porcentaje acumulado inmediatamente superior a 90% (en este caso es el 90,9%)
2. Mirar a qué valor de la variable corresponde (en este caso 8)

✓ ¿Cómo sacar el percentil 25?

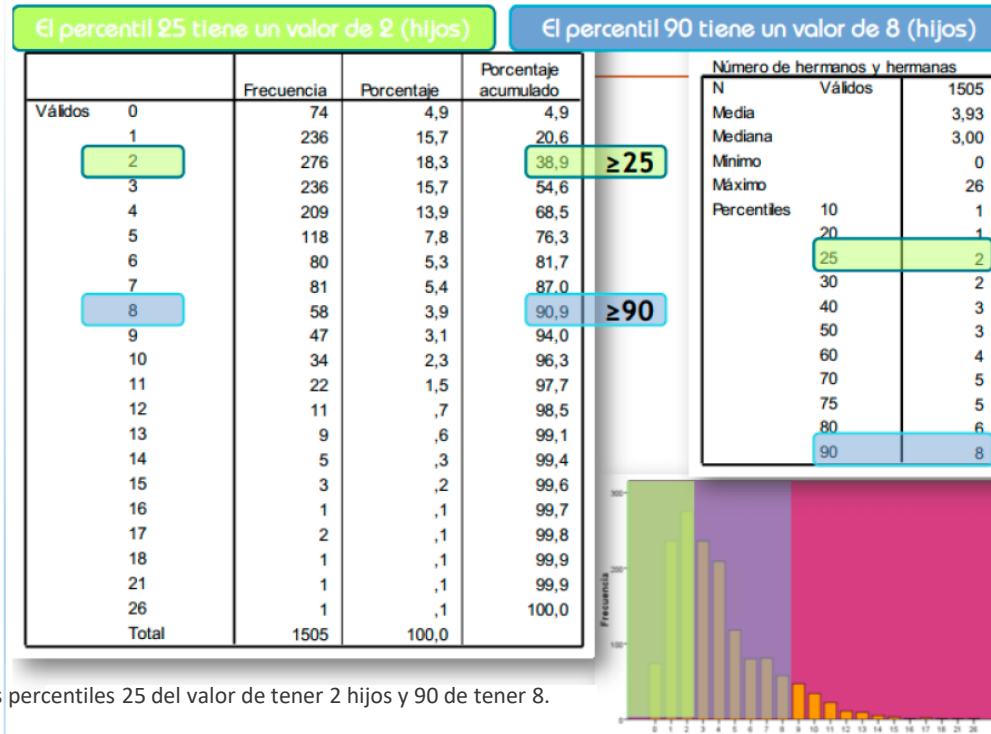
1. Hallar el porcentaje acumulado inmediatamente superior a 25% (en este caso es el 38,9%)
2. Mirar a qué valor de la variable corresponde (en este caso 2)

✓ ¿Cómo sacar la mediana?

1. Porcentaje acumulado inmediatamente superior a 50% => 54,6%
2. Valor correspondiente => 3

4. Medidas de posición

Figura 1
Percentil



Nota. Esta figura extrae los percentiles 25 del valor de tener 2 hijos y 90 de tener 8.

Fuente: ...

4. Medidas de posición

Otras posiciones

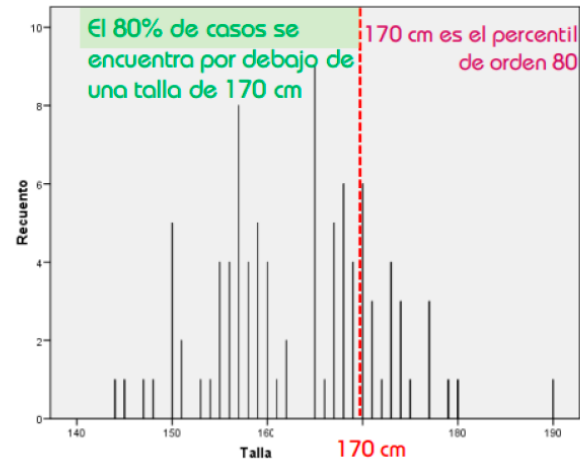
Figura 2
Percentil

	Válidos	Porcentaje acumulado
	144	1,1
	145	2,1
	147	3,2
	148	4,3
	150	9,6
	151	11,7
	153	12,8
	154	13,8
	155	18,1
	156	22,3
	157	30,9
	158	35,1
	159	40,4
	160	44,7
	161	45,7
	162	47,9
	165	57,4
	166	58,5
	167	63,8
	168	70,2
	169	74,6
	170	80,9
	171	84,0
	172	85,1
	173	89,4
	174	92,6
	175	93,6
	177	96,8
	179	97,9
	180	98,9
	190	100,0

Nota. Aquí se calcula el valor o percentil de la variable que queda por debajo del 80% de los casos que miden menos de 170cm.

Fuente: ...

Queremos conocer qué valor de la variable nos deja por debajo de él aprox. Al 80% de los casos



4. Medidas de posición

Otras posiciones

Deciles (de 0 a 10)

$$\alpha = k/10$$

- ✓ **Decil** de orden K es el cuantil de orden $\alpha = k/10$. Tiene el mismo significado que el cuantil, pero se expresa en escala del 0 al 10
- ✓ **Decil** de orden K es el valor por debajo del cual se encuentra una frecuencia acumulada de $(10 \cdot k)\%$
- ✓ El **decil 5** es el valor de la variable que deja a un lado y a otro el mismo número de casos = **MEDIANA**
- ✓ *Para calcularlos es necesario ordenar los datos (de menor a mayor o al revés)*

4. Medidas de posición

Otras posiciones

Cuartiles (de 1 a 3)

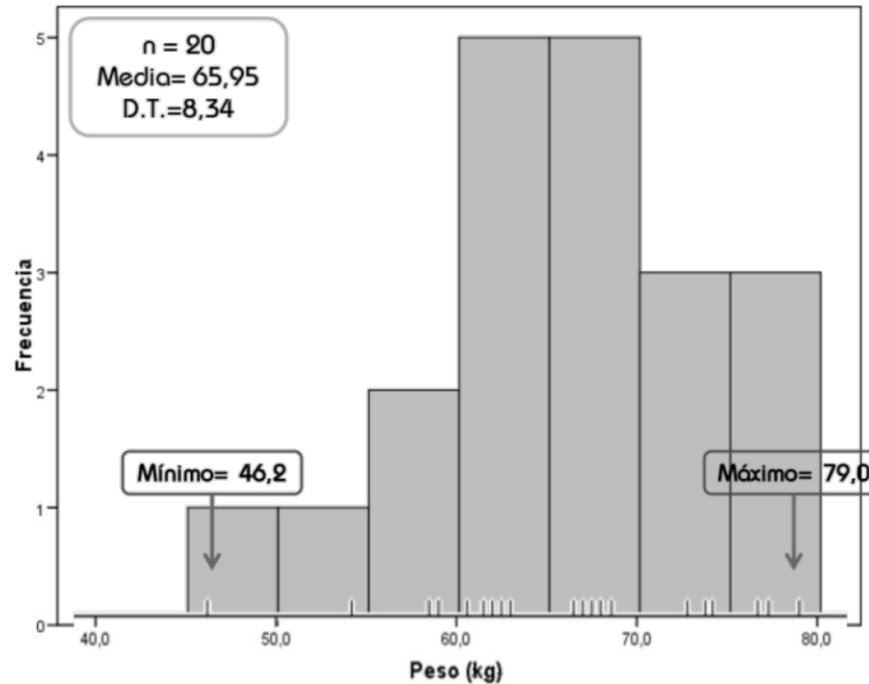
Los **cuartiles** dividen a la muestra en cuatro grupos iguales.

- ✓ Existen tres cuartiles:
 - $Q1 = \text{percentil } 25 = \text{cuantil } 0,25 (= \alpha)$
 - $Q2 = \text{percentil } 50 = \text{cuantil } 0,50 (= \alpha) = \text{decil } 5 = \text{mediana}$
 - $Q3 = \text{percentil } 75 = \text{cuantil } 0,75 (= \alpha)$

4. Medidas de posición

Otras posiciones

Gráfico 11
Titulo figura

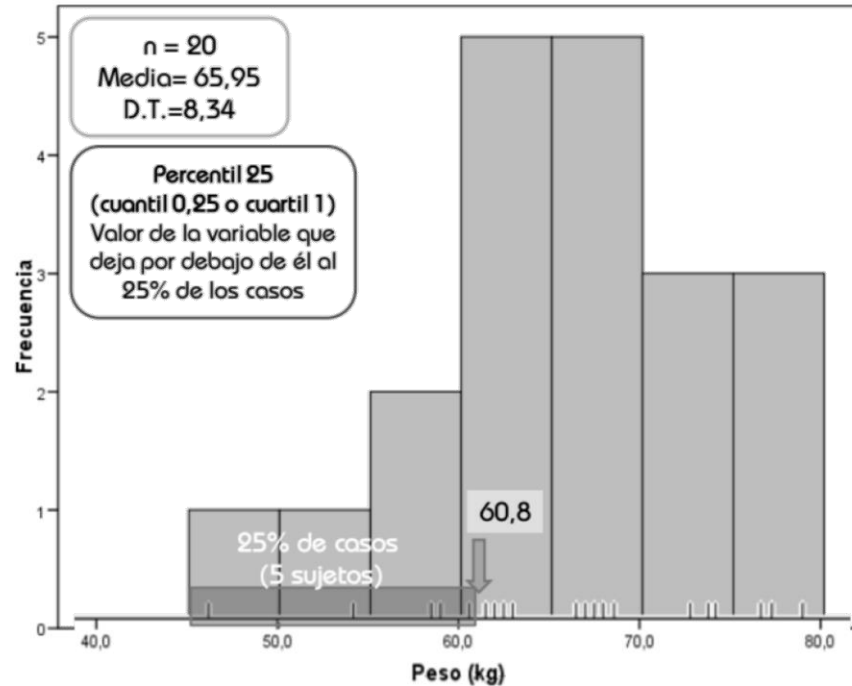


Nota. Esta figura muestra...
Fuente:

4. Medidas de posición

Otras posiciones

Gráfico 12
 Título figura

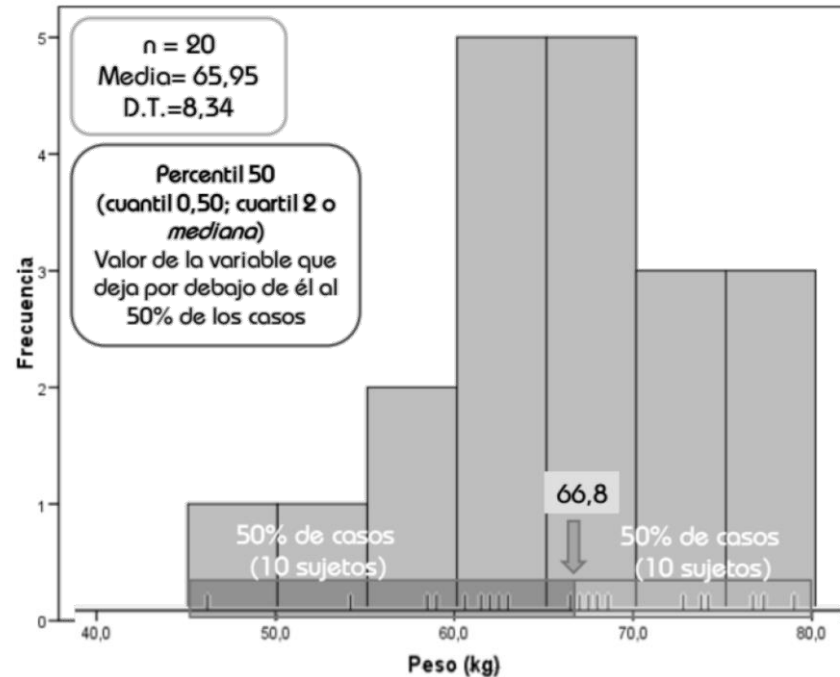


Nota. Esta figura muestra...
 Fuente:

4. Medidas de posición

Otras posiciones

Gráfico 13
 Titulo figura

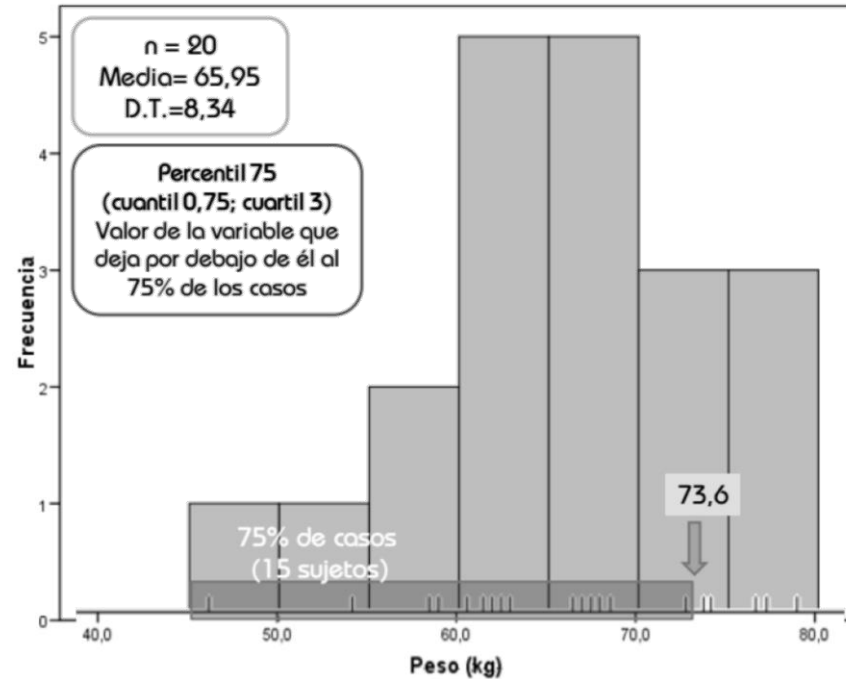


Nota. Esta figura muestra...
 Fuente:

4. Medidas de posición

Otras posiciones

Gráfico 14
Titulo figura

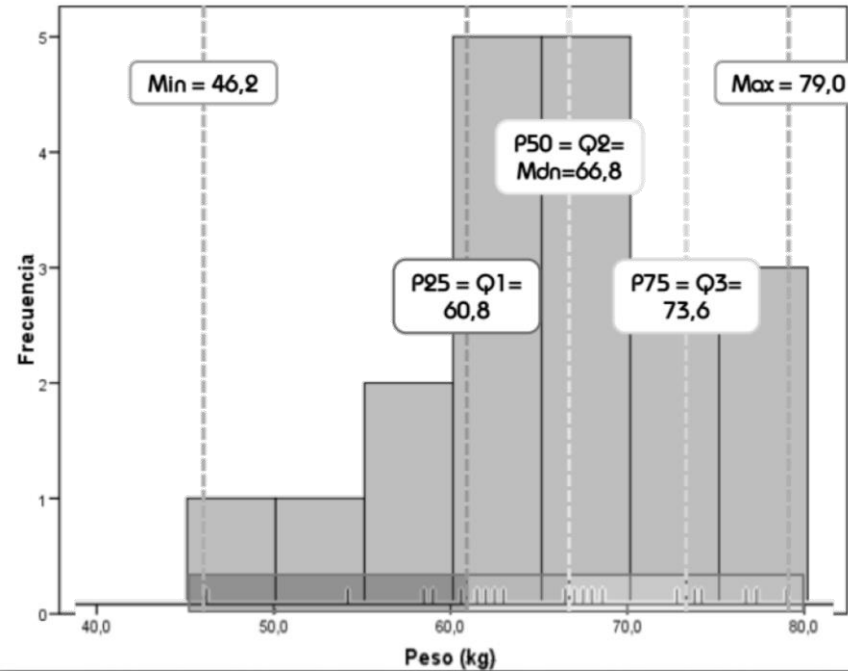


Nota. Esta figura muestra...
Fuente:

4. Medidas de posición

Otras posiciones

Gráfico 15
 Título figura



Nota. Esta figura muestra...
 Fuente:



Imagen 8
Velocímetro

5. Medidas de dispersión

Nota. Adaptado de *Velocímetro vehículo*, por Digital Buggu , 2016, [Pexels](#).

5. Medidas de dispersión

Información de la variabilidad (distancia) que presentan los casos respecto a un valor concreto.

Rango

La distancia (diferencia) entre las dos observaciones más extremas (máximo y mínimo)

Rango intercuartílico

La distancia (diferencia) entre el tercer y el primer cuartil

Varianza (desv. típica)

El promedio de las desviaciones, al cuadrado, de cada una de las observaciones respecto a la media

Coefficiente de variación

Índice que relaciona la dispersión con la media, en %, para comparar el grado de variabilidad de distintas variables.

5. Medidas de dispersión

Rango

Rango intercuartílico

Varianza / D.E.

Coeficiente de variación

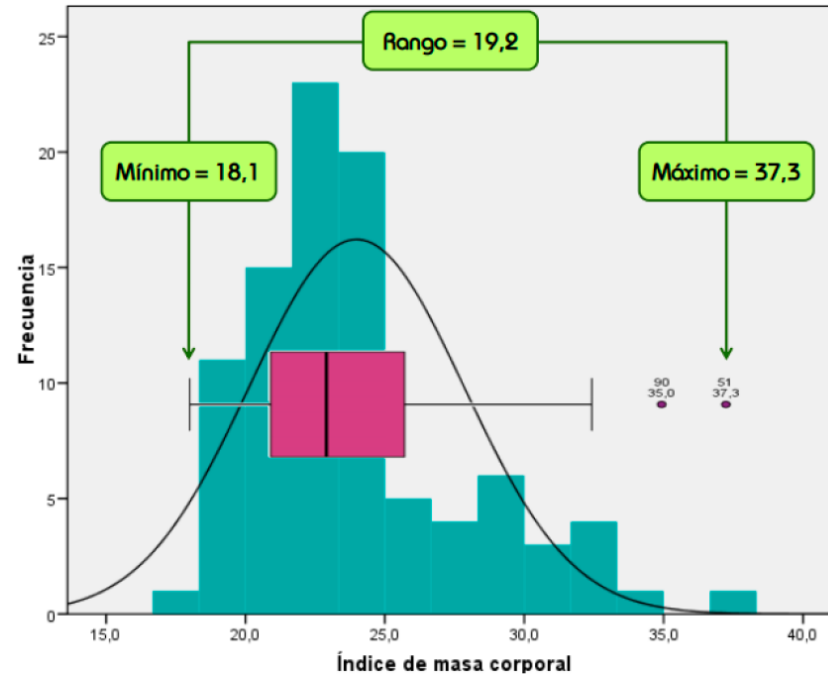
La distancia (diferencia) entre las dos observaciones más extremas (máximo y mínimo).

- ✓ VENTAJAS: es el cálculo sencillo. No está de más calcularla para detectar casos extremos.
- ✗ DESVENTAJAS: solo implica a dos valores de la variable, el mayor y el menor. Es muy sensible a los casos extremos.

5. Medidas de dispersión

Rango

Gráfico 16
Titulo figura



Nota. Esta figura muestra...
Fuente:

5. Medidas de dispersión

Varianza

Rango

Rango intercuartílico

Varianza / D.E.

Coeficiente de variación

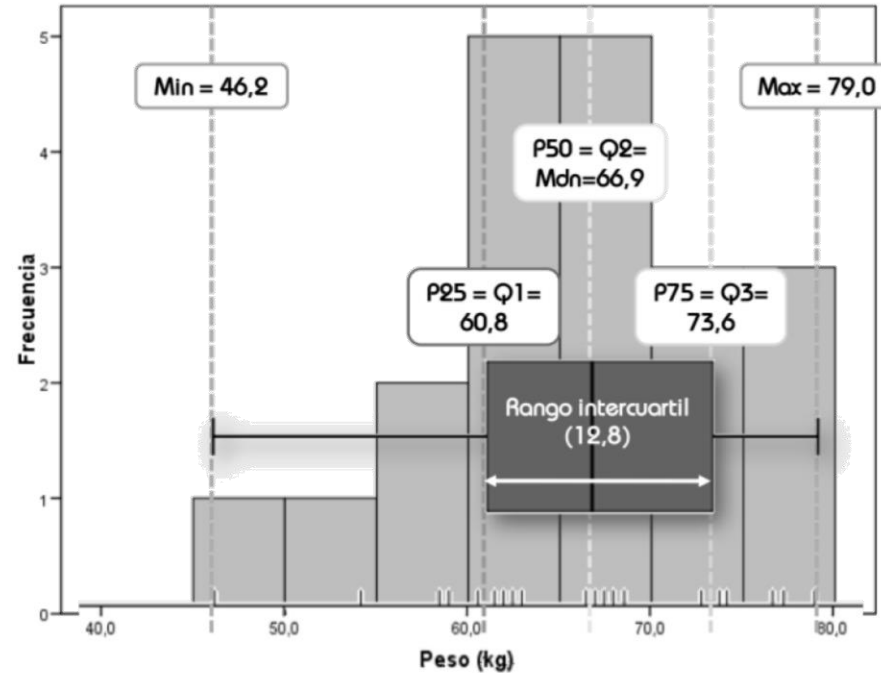
Distancia (diferencia) entre el tercer y el primer cuartil.

- ✓ VENTAJAS: es el cálculo sencillo. Excluye 25% de los casos más grandes y el 25% de los casos más pequeños y por tanto utiliza en sus cálculos el 50% de las observaciones centrales. No es afectada por casos extremos.
- ✗ DESVENTAJAS: solo participan en su cálculo el valor del cuartil 1 (P25) y es el Cuartil 3 (P75%).

5. Medidas de dispersión

Rango y rango intercuartílico

Gráfico 17
 Título figura



Nota. Esta figura muestra...
 Fuente:

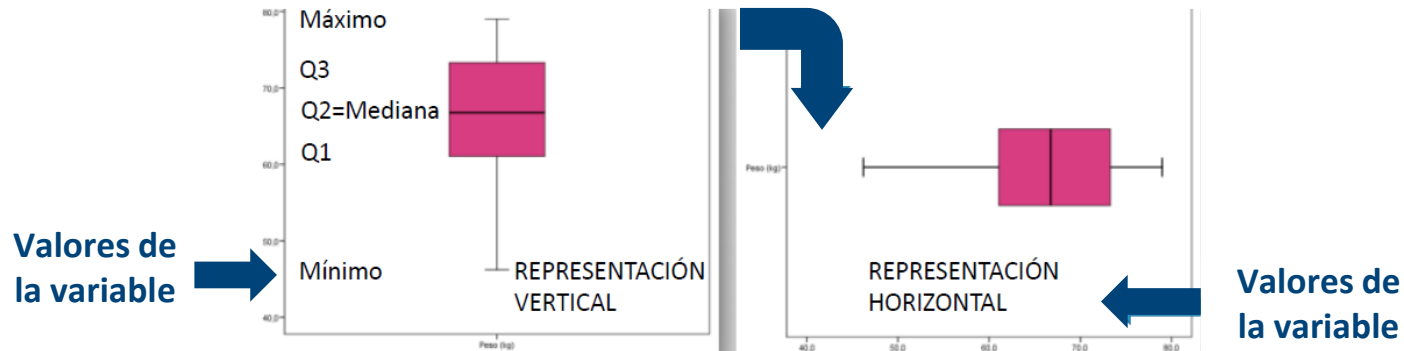
5. Medidas de dispersión

Rango intercuartílico

Los diagramas de Turkey (diagramas de cajas, Box-plot) se construyen con estos 5 valores:

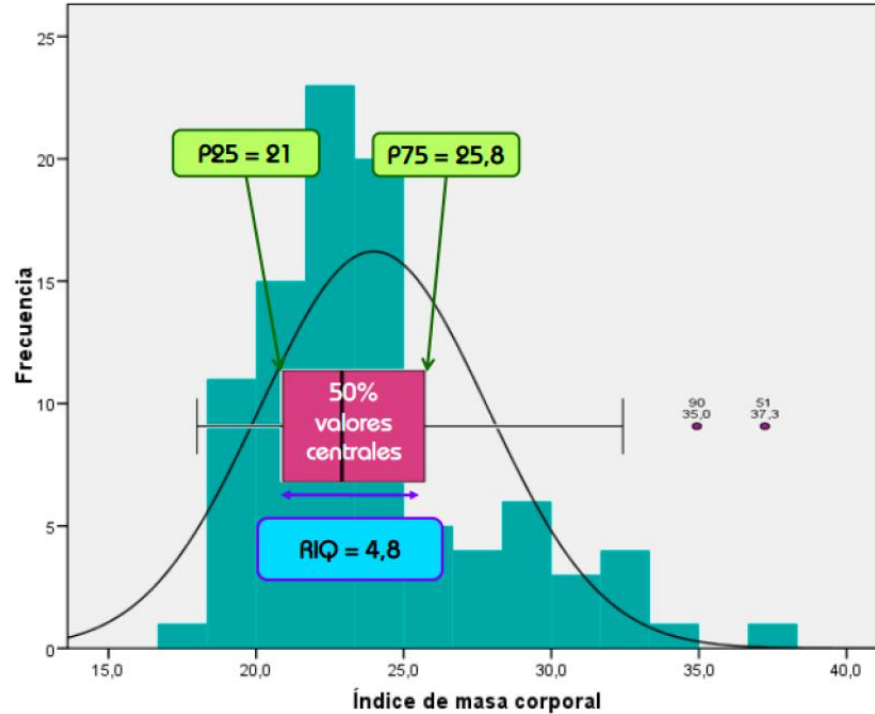
Mínimo, Q1, Mediana, Q3, Máximo

Muy útiles para resumir la distribución



5. Medidas de dispersión

Gráfico 18
Titulo figura



Nota. Esta figura muestra...
Fuente:

5. Medidas de dispersión

Varianza

Rango

Rango intercuartílico

Varianza / D.E.

Coeficiente de variación

Varianza (S^2 , σ^2) es el promedio de las desviaciones, al cuadrado, de cada una de las observaciones respecto de la media.

✓ **VENTAJAS:** participa en los test estadísticos que se aplican si la distribución se parece a una normal (pruebas paramétricas).

✗ **DESVENTAJAS:**

- Sensible a los valores extremos alejados de la media.
- Está en unidades “al cuadrado” por lo que su interpretación se complica (por ello no se suele usar para interpretación).

5. Medidas de dispersión

Varianza

Rango

Rango intercuartílico

Varianza / D.E.

Coeficiente de variación

Desviación típica o **desviación estándar** (S, σ) es la raíz cuadrada de la varianza: $S = \sqrt{S^2}$, $\sigma = \sqrt{\sigma^2}$

- ✓ **VENTAJAS:** se encuentra en las mismas unidades que la variable y se interpreta fácilmente.
- ✗ **DESVENTAJAS:** no es una medida de dispersión adecuada cuando existan valores extremos (los extremos pueden afectarle demasiado).

Se llama también **D.E.** = Desviación Estándar o **D.T.** = Desviación típica.

5. Medidas de dispersión

Cálculo con todos los datos

Edad (años)	21	32	15	59	60	61	64	60	71	80
-------------	----	----	----	----	----	----	----	----	----	----

Media

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{21 + 32 + 15 + \dots + 80}{10} = 52,3 \text{ años}$$

Varianza

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{(21 - 52,3)^2 + (32 - 52,3)^2 + \dots + (80 - 52,3)^2}{10} = 427,61 \text{ (años)}^2$$

**Desviación
típica**

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{(21 - 52,3)^2 + (32 - 52,3)^2 + \dots + (80 - 52,3)^2}{10}} = \sqrt{427,61} = 21,79 \text{ años}$$

5. Medidas de dispersión

- ✓ **VARIANZA S^2 :** expresión general y expresión reducida

$$S^2 = \sum_l^r (X_i - \bar{x})^2 fi = \sum_l^r X_i^2 fi - \bar{x}^2 = \overline{[x_i^2]} - \bar{x}^2$$

- ✓ **Mide la dispersión en torno a la media.** Qué tan cerca o lejos se encuentran, en promedio, todos los valores de su media.
- ✓ Cuanto menor sea su valor más representa la media al conjunto de datos

5. Medidas de dispersión

Varianza

- ✓ Teniendo la tabla de frecuencias. La ordenación de cálculos para la varianza, basta construir lo siguiente sobre la tabla de frecuencias y usar el cálculo previo de la media:

Tabla 15
Cálculo de media

x_i	n_i	$x_i^2 n_i$
1	2	2
2	8	32
3	4	36
4	5	80
5	4	100
6	1	36
TOTAL	24	286
	$[x_i^2]$	11,9

Nota. Esta tabla muestra cómo se calcula fácilmente la media por medio de una tabla de frecuencias.
Fuente: Elaboración propia.

$$[x_i^2] = 286/24 = 11,9$$

$$S^2 = [x_i^2] - \bar{x}^2 = 11,9 - 3,17^2 = 1,85$$

5. Medidas de dispersión

Propiedad de la varianza

✓ Cambio de variable

$$Y = a + bX$$



$$S^2(y) = b^2 S^2(X)$$

$$S^2(y) = \sum_1^r (y_i - \bar{y})^2 f_i = \sum_1^r [(a + bx_i) - (a + b\bar{x})]^2 f_i = \sum_1^r [(bx_i + b\bar{x})]^2 f_i = b^2 \sum_1^r (x_i - \bar{x})^2 f_i = b^2 S^2(X)$$

5. Medidas de dispersión

Coeficiente de Variación

Rango

Rango intercuartílico

Varianza (desv. típica)

Coeficiente de variación

CV = es un índice adimensional (sin unidades en %) que relaciona la dispersión y la media. Sirve para comparar el grado de variabilidad de 2 variables procedentes de una muestra diferente o que están en unidades diferentes.

$$CV = \frac{S}{\bar{x}} * 100$$

5. Medidas de dispersión

Coeficiente de Variación

$$CV = \frac{S}{\bar{x}} * 100$$

PESO: Media = 69,6 Kg; D.T.= 10,44 Kg

$$CV_{\text{peso}} = (10,44/69,6) \times 100 \approx \mathbf{15\%}$$

TAS: Media = 166 mmHg; D.T.=21,3 mmHg

$$CV_{\text{TAS}} = (21,3/166) \times 100 \approx \mathbf{13\%}$$

Cuidado porque no debe utilizarse cuando la variable pueda tomar valores negativos



Imagen 9
Cinta métrica

6. Medidas de forma

Nota. Adaptado de *Cinta métrica-metro largo*, por Holger Langmaier, 2017, [Pixabay](#).

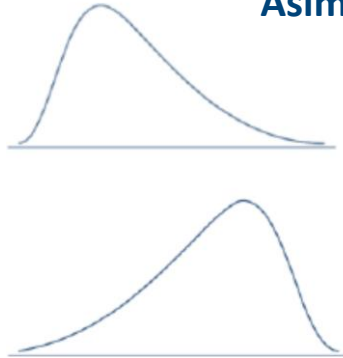
6. Medidas de forma

Medidas de forma

Asimetría (skewness)

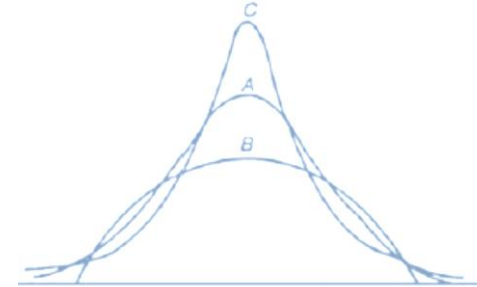
Apuntamiento (curtosis)

Asimetría



Simetría respecto a la normal

Apuntamiento



“Aplastamiento” frente a la normal

- ✓ Informan sobre el parecido de la distribución de los datos respecto a la distribución normal.

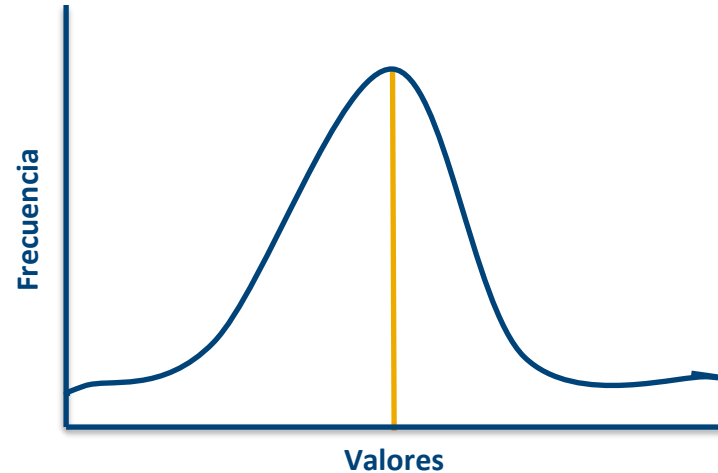
6. Medidas de forma

Apuntamiento o curtosis

- ✓ Asimetría: si la distribución es asimétrica, la distribución es dividida por la media en dos mitades iguales (asimetría = 0).

Asimetría
(skewness)

Apuntamiento
(curtosis)



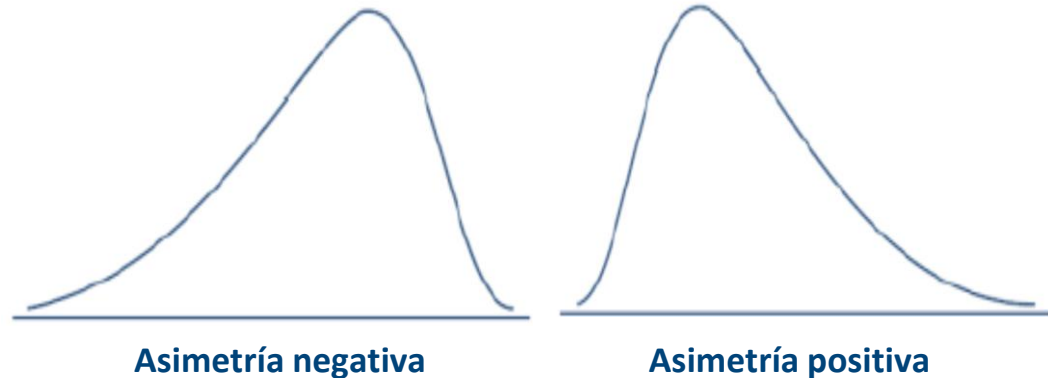
6. Medidas de forma

Apuntamiento o curtosis

Asimetría
(skewness)

Apuntamiento
(curtosis)

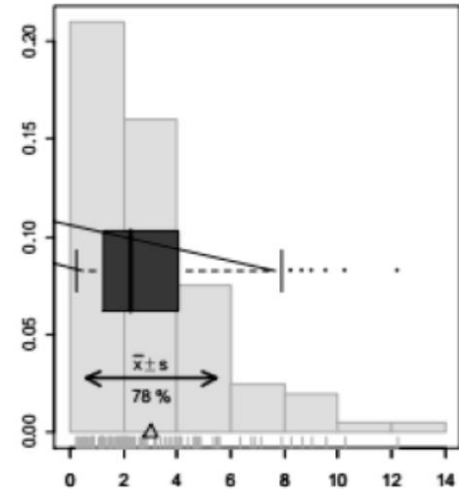
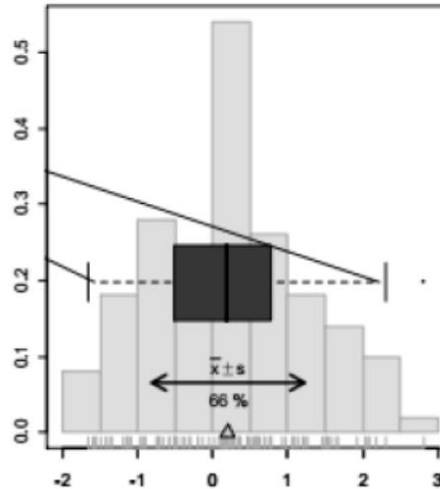
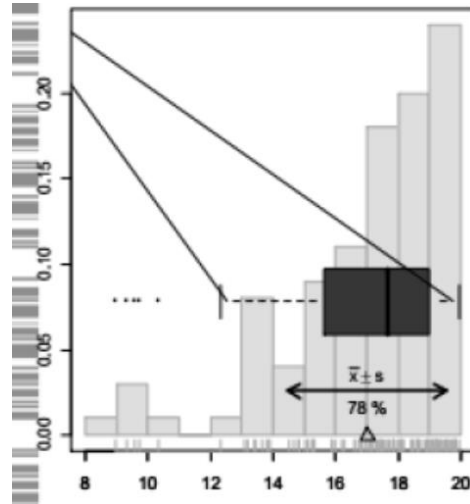
- ✓ Cuando la distribución tiene una cola hacia la derecha presenta una **asimetría positiva**; cuando la cola parece a la izquierda la **asimetría** es **negativa**.



6. Medidas de forma

Asimetría

- ✓ Ejemplos de asimetría negativa (izquierda), aproximadamente nula (centro) y positiva (derecha).



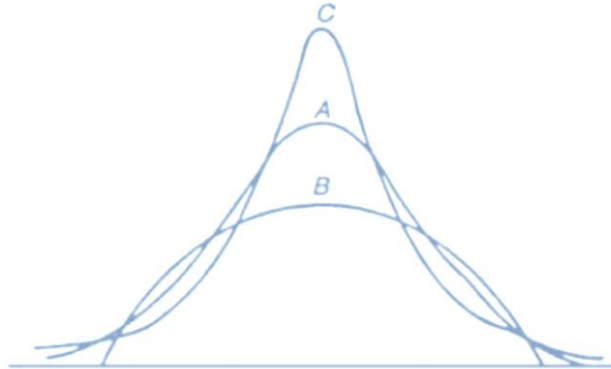
6. Medidas de forma

Apuntamiento o curtosis

- ✓ **Curtosis:** se refiere a la altura de la distribución respecto a la distribución normal.

Asimetría
(skewness)

Apuntamiento
(curtosis)



A = Mesocúrtica

(como la normal)

B = Platicúrtica

(aplanada)

C = leptocúrtica

(apuntada)

6. Medidas de forma

Apuntamiento o curtosis

Asimetría
(skewness)

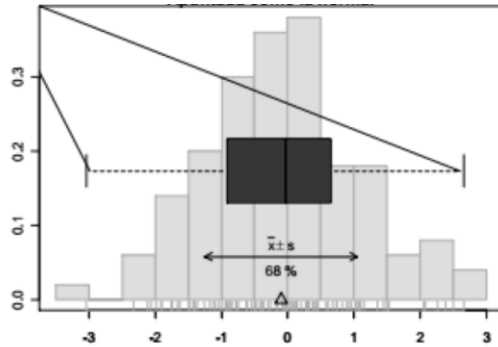
Apuntamiento
(curtosis)

- ✓ **Curtosis:** se refiere a la altura de la distribución.
- ✓ El **coeficiente de aplastamiento Fisher** (curtosis) es un índice que nos indica si la distribución es más o menos apuntada que la distribución normal.
 - Coeficiente = 0 distribución mesocúrtica (como la normal).
 - Coeficiente > 0 distribución leptocúrtica (apuntada).
 - Coeficiente es < 0 , distribución platicúrtica (aplanada).

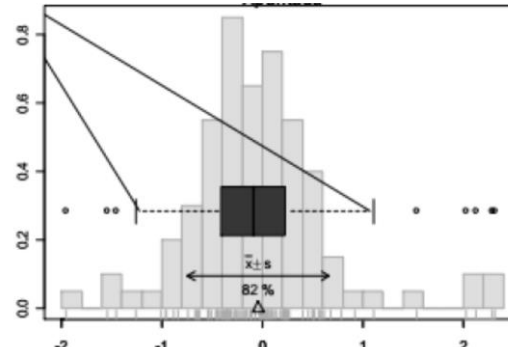
6. Medidas de forma

Apuntamiento o curtosis

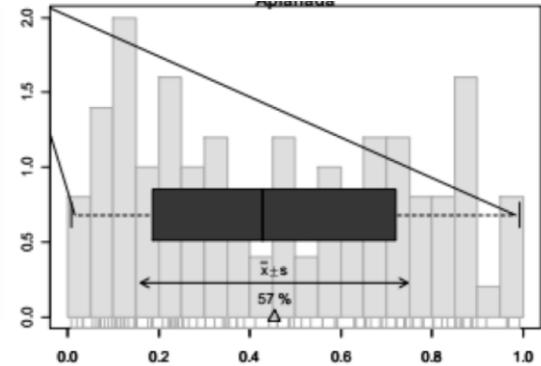
Apuntada como la normal



Apuntada

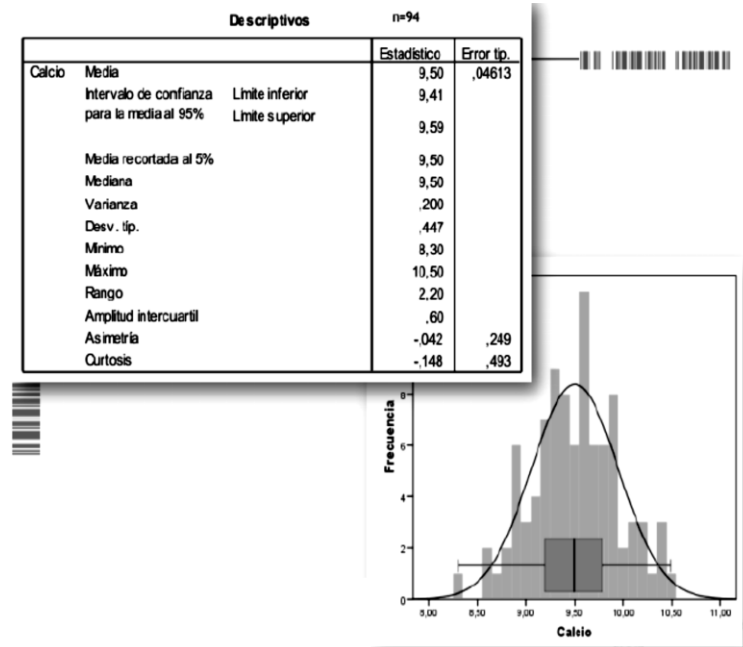


Aplanada



6. Medidas de forma

- ✓ Vamos a ver la simetría y la curtosis con un ejemplo de programa de estadística.
- ✓ Esta variable tiene:
 - Tiene gran simetría porque el valor de asimetría es casi 0.
 - Curtosis prácticamente nula, ligeramente negativa (forma mesocúrtica).



REFERENCIAS

BIBLIOGRÁFICAS

Pardo Merino A, Ruiz MA, San Martín R (2009). *Análisis de datos en ciencias sociales y de la salud I*. Síntesis.

Feinstein AR. (2002). *Principles of medical statistics*. Chapman & Hall/CRC.

Hurley WL, Denegar CR, Hertel J, Vazquez JC. (2012). *Métodos de investigación fundamentos de una práctica clínica basada en la evidencia*. Wolter Kluwer.

Genbeta. (4 de diciembre de 2014). *Once cursos universitarios gratuitos sobre estadística*. <https://www.genbeta.com/web/once-cursos-universitarios-gratuitos-sobre-estadistica>

REFERENCIAS

BIBLIOGRÁFICAS

Tareasplus para saber más (s.f.). *Probabilidad y Estadística*.

<https://www.tareasplus.com/Curso-Probabilidad-y-Estadistica/Marcela-Gomez>

Tareasplus para saber más (s.f.). *Estadística Empresarial*.

<https://www.tareasplus.com/Estadistica-Empresarial/Neumann-BS>

Miriadax (s.f.). *Curso-practico-bioestadística*. <https://miriadax.net/web/curso-practico-bioestadistica-r-3edicion>

Juan Miguel Navarro Ruiz

jmnavarro@ucam.edu

UCAM Universidad Católica de Murcia



UCAM

Tema 3**Análisis Descriptivo Bidimensional**

Estadística Aplicada al Turismo

Juan Miguel Navarro Ruiz

Grado en Turismo

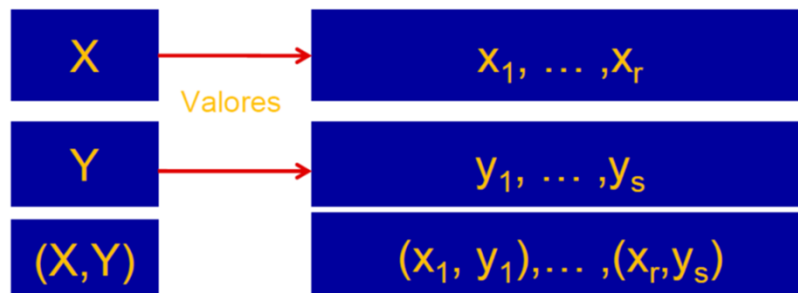
UCAM

ÍNDICE**CONTENIDOS**

1. Variables Bidimensionales
2. Tabulación de datos: distribución de frecuencias bidimensionales
3. Distribuciones marginales y condicionadas
4. Independencia estadística
5. Covarianza y coeficiente de correlación
6. Recta de regresión.

1. Variables Bidimensionales

Como en el caso unidimensional, ahora las variables X e Y pueden ser cualitativas, cuantitativas, discretas o continuas, pudiendo presentarse cualquier combinación de ambas opciones. Se utilizará la notación siguiente



Miguel Navarro – UCAM - 3

2. Distribuciones Bidimensionales

Porcentaje de hombres fumadores entre los que todos los sujetos del estudio

Sexo	Tabaquismo			Total
	Fumadores	Exfumadores	No fumadores	
Hombres	Recuento	18	7	69
	% del total	9,0%	3,5%	34,5%
Mujeres	Recuento	42	6	58
	% del total	21,0%	3,0%	29,0%
Total	Recuento	60	13	127
	% del total	30,0%	6,5%	63,5%

Los **porcentajes del total** indican el porcentaje de cada frecuencia conjunta respecto al número total de casos de la tabla.

Cada porcentaje es la **PROBABILIDAD CONJUNTA** de que sea hombre y fumador

Juan Miguel Navarro – UCAM - 4

2. Distribuciones Bidimensionales

Se pueden utilizar los conceptos de frecuencias acumuladas (absolutas, relativas, porcentajes)

N_{hk} = Número de observaciones para las que $X \leq x_h, Y \leq y_k$

F_{hk} = Proporción de observaciones para las que $X \leq x_h, Y \leq y_k$

P_{hk} = %de observaciones para las que $X \leq x_h, Y \leq y_k$

$$N_{hk} = \sum_{i=1}^h \sum_{j=1}^k n_{ij}$$

$$F_{hk} = \sum_{i=1}^h \sum_{j=1}^k f_{ij}$$

$$P_{hk} = \sum_{i=1}^h \sum_{j=1}^k p_{ij}$$

Juan Miguel Navarro – UCAM - 5

3. Distribuciones Marginales y Condicionadas

Distribuciones marginales

Tabla de contingencia Sexo * Tabaquismo

		Tabaquismo			
		Fumadores	Exfumadores	No fumadores	Total
Sexo	Hombres	18	7	69	94
	Mujeres	42	6	58	106
Total		60	13	127	200

Frecuencias marginales: son los totales de cada fila o de cada columna. Equivalen a las tablas de frecuencias individuales

Marginales para la variable Tabaquismo

Marginales para la variable Sexo

Juan Miguel Navarro – UCAM - 6

3. Distribuciones Marginales y Condicionadas

Distribuciones condicionadas

El concepto de distribución condicionada puede generalizarse planteando cuál es el comportamiento de una variable, por ejemplo la X, cuando la otra Y cumple una cierta condición.

En este caso tendríamos la distribución $(X/Y \in A)$, con lo cual los valores de la variable seguirán siendo los x_i en tanto que las frecuencias hay que construirlas como:

$$f_{i/Y \in A} = \frac{\text{Nº de observaciones con } X=x_i, Y \in A}{\text{Nº de observaciones con } Y \in A}$$

Juan Miguel Navarro – UCAM - 7

3. Distribuciones Marginales y Condicionadas

Distribuciones condicionadas

Porcentaje de fumadores respecto al total de hombres

Sexo		Tabaquismo			Total
		Fumadores	Exfumadores	No fumadores	
Hombres	Recuento	18	7	69	94
	% dentro de Sexo	19,1%	7,4%	73,4%	100,0%
Mujeres	Recuento	42	6	58	106
	% dentro de Sexo	39,6%	5,7%	54,7%	100,0%
Total	Recuento	60	13	127	200
	% dentro de Sexo	30,0%	6,5%	63,5%	100,0%

Los **porcentajes por fila** indican el porcentaje de cada frecuencia conjunta respecto a la frecuencia marginal de su fila. Es una **distribución condicionada**.

Condicionada quiere decir: "La probabilidad de que sea fumador, CUANDO SÉ que es un hombre, es de 19.1%"
La variable tabaquismo **ESTÁ CONDICIONADA** a la variable sexo

Porcentaje de exfumadores respecto al total de mujeres

Juan Miguel Navarro – UCAM - 8

3. Distribuciones Marginales y Condicionadas

Distribuciones condicionadas

Porcentaje de hombres entre los que todos los fumadores

Sexo	Tabaquismo			Total
	Fumadores	Exfumadores	No fumadores	
Hombres	Recuento	18	7	94
	% dentro de Tab.	30,0%	53,8%	54,3%
Mujeres	Recuento	42	6	58
	% dentro de Tab.	70,0%	46,2%	45,7%
Total	Recuento	60	13	200
	% dentro de Tab.	100,0%	100,0%	100,0%

Los **porcentajes por columna** indican el porcentaje de cada frecuencia conjunta respecto a la frecuencia marginal de su columna. Es una **distribución condicionada**.

Ahora es la probabilidad de que sea hombre, CUANDO SÉ que es fumador
La variable sexo **ESTÁ CONDICIONADA** a la variable tabaquismo

Juan Miguel Navarro – UCAM - 9

4. Independencia Estadística

Cualitativas

$$\frac{n_{i1}}{n_{.j}} = \frac{n_{i2}}{n_{.j}} = \dots = \frac{n_{is}}{n_{.j}} = \frac{\sum_{j=1}^s n_{ij}}{\sum_{j=1}^s n_{.j}} = \frac{n_{i.}}{N}$$

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

$$\forall ij$$

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

CONDICIÓN DE INDEPENDENCIA

Juan Miguel Navarro – UCAM - 10

4. Independencia Estadística.

Cualitativas

Sexo		Tabaquismo			Total
		Fumadores	Exfumadores	No fumadores	
Hombres	Frec. observada	18	7	69	94
	Frec. esperada	28,2	6,1	59,7	94
Mujeres	Frec. observada	42	6	58	106
	Frec. esperada	31,8	6,9	67,3	106
Total		60	13	127	200

La frecuencia **observada** para mujeres Y fumadoras es de 42 (la muestra real)

La frecuencia **esperada** para mujeres Y fumadoras sería de 31,8 mujeres en el caso de las variables fuesen independientes => $(106 * 60)/200$

Este número corresponde al número de mujeres (106) por el número de fumadores (60) partido por el número total de sujetos en el estudio (200)

4. Independencia Estadística.

Prueba Chi² de Pearson

Sexo		Tabaquismo			Total
		Fumadores	Exfumadores	No fumadores	
Hombres	Frec. observada	18	7	69	94
	Frec. esperada	28,2	6,1	59,7	94
Mujeres	Frec. observada	42	6	58	106
	Frec. esperada	31,8	6,9	67,3	106
Total		60	13	127	200

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$$

$$\chi^2 = \frac{(18 - 28,2)^2}{28,2} + \frac{(7 - 6,1)^2}{6,1} + \dots + \frac{(58 - 67,3)^2}{67,3} = 9,95$$

Todas las categorías conjuntas

Ejemplo: hombres y exfumadores

Cuanto mayor sea el valor de Chi-cuadrado, mas fuertemente relacionadas están las variables. Pero ¿cómo hacer para tomar la decisión?

4. Independencia Estadística.

Prueba χ^2 de Pearson

GRADO DE LIBERTAD ν

☐ Para calcular el grado de libertad (ν) se realiza:

$$\nu = (Cantidad\ de\ filas - 1)(Cantidad\ de\ columnas - 1)$$

EJEMPLO:

TABLA DE VALORES OBSERVADOS

50	25
40	45

$$\nu = (2-1)(2-1)$$

$$\nu = 1(1) = 1$$

NIVEL DE SIGNIFICANCIA

☐ Es el error que se puede cometer al rechazar la hipótesis nula siendo verdadera.

☐ Por lo general se trabaja con un nivel de significancia de 0.05, que indica que hay una probabilidad del 0.95 de que la hipótesis nula sea verdadera.

4. Independencia Estadística.

Prueba χ^2 de Pearson

Tabla Distribución Chi Cuadrado χ^2

V/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055
2	13,815	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052
3	16,266	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,017
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616
9	27,8767	25,4625	23,5893	21,666	19,0228	16,919	14,6837
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,307	15,9872

ν : grados de libertad

χ^2 Tabla = 5,991 χ^2 Calculado = 9,95

Si χ^2 Calculado > χ^2 Tabla Las variables **SI** están relacionadas.

Si χ^2 Calculado < χ^2 Tabla Las variables **NO** están relacionadas.

4. Independencia Estadística.

Prueba χ^2 de Pearson

- Estadístico Chi cuadrado y coeficiente de contingencia

$$\chi_E^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (\text{Estadístico Chi cuadrado})$$

$$\chi_{(k-1)(m-1), 0.05}^2$$

$$\text{donde } e_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{N}$$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \quad (\text{Coeficiente de contingencia})$$

Juan Miguel Navarro – UCAM - 15

4. Independencia Estadística.

Las variables, que utilizaremos para el ejemplo son:

1. Taller de capacitación.
2. Actitud empresarial.

Las dos variables, tienen dos categorías. Los datos del primer cuadro se llaman "**Valores observados**", pues fueron obtenidos, u observados en el experimento. Los datos del segundo cuadro, se llaman "**Valores esperados**". Estos valores se obtienen multiplicando el total de cada fila por el total de cada columna, y dividiendo entre el total general (200)

	Pre Test	Post Test	Total
Si	4	14	18
No	96	86	182
Total	100	100	200

	Pre Test	Post Test	Total
Si	9	9	18
No	91	91	182
Total	100	100	200

Juan Miguel Navarro – UCAM - 16

4. Independencia Estadística.

Prueba Chi² de Pearson. Ejemplo

	Pre Test	Post Test	Total
Si	4	14	18
No	96	86	182
Total	100	100	200

	Pre Test	Post Test	Total
Si	9	9	18
No	91	91	182
Total	100	100	200

Formula de chi cuadrado:

$$\chi^2 = \frac{\sum (v.o - v.e)^2}{v.e} = \chi^2 = \frac{(4-9)^2}{9} + \frac{(96-91)^2}{91} + \frac{(14-9)^2}{9} + \frac{(86-91)^2}{91} = 6,105$$

Coefficiente de contigencia: **0,17** Asociación casi nula o débil entre las variables.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

El valor de C varía entre 0 y 1.

C = 0, significa que no hay asociación entre las variables.

C > .30, indica una buena asociación entre las variables.

Juan Miguel Navarro – UCAM - 17

4. Independencia Estadística.

Prueba Chi² de Pearson. Ejemplo

	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	
9 d.f.	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	9 d.f.
2	13,816	10,597	9,555	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,468	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,486	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,188	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,257	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15
16	39,252	34,267	32,000	29,633	28,845	28,191	27,136	26,296	23,542	21,793	20,465	19,369	18,418	17,565	16,780	16
17	40,790	35,718	33,409	30,995	30,191	29,523	28,445	27,587	24,769	22,977	21,615	20,489	19,511	18,633	17,824	17
18	42,312	37,156	34,805	32,346	31,526	30,845	29,745	28,669	25,989	24,155	22,760	21,605	20,601	19,699	18,868	18
19	43,820	38,582	36,191	33,687	32,852	32,158	31,037	30,144	27,204	25,329	23,900	22,718	21,689	20,764	19,910	19
20	45,315	39,997	37,566	35,020	34,170	33,462	32,321	31,410	28,412	26,498	25,038	23,828	22,775	21,826	20,951	20

Significancia = 0,01 Chi² Tabla = 6,635 Chi² Calculado = 6,105

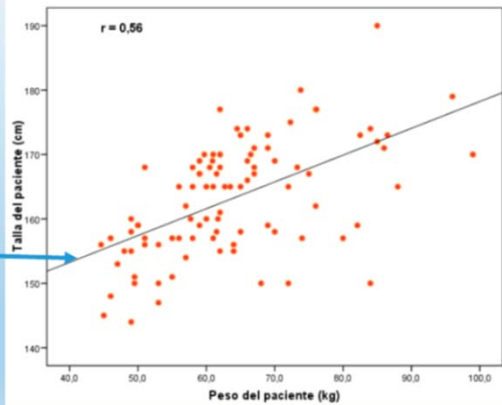
Si Chi² Calculado > Chi² Tabla Las variables SI están relacionadas.

Si Chi² Calculado < Chi² Tabla Las variables NO están relacionadas.

Juan Miguel Navarro – UCAM - 18

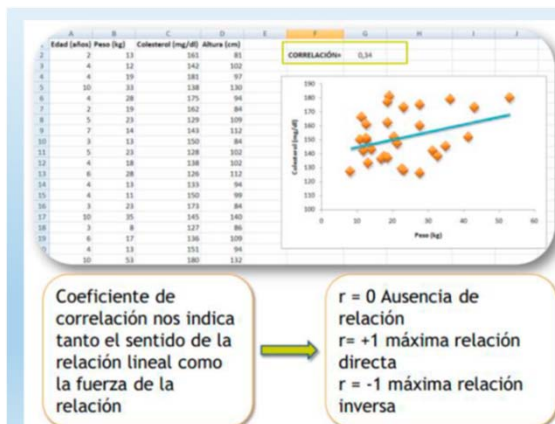
5. Covarianza y coeficiente de correlación

- Entre las variables talla y peso del paciente (cuantitativas) se puede apreciar cierta relación.
- Una técnica permite predecir la recta que “pasa más cerca” de todos los puntos
- Se trata de un “diagrama de dispersión”



Juan Miguel Navarro – UCAM - 19

5. Covarianza y coeficiente de correlación



- Correlación = 0,34
- Están relacionados “positivamente” (a mayor peso mayor colesterol) pero no totalmente (hay puntos por encima y por debajo de la línea) por eso no vale exactamente 1

Juan Miguel Navarro – UCAM - 20

5. Covarianza y coeficiente de correlación

- Interpretación de la covarianza

$$S_{x,y} = \frac{\sum_i \sum_j (x_i - \bar{x})(y_j - \bar{y})n_{ij}}{N} \quad S_{xy} = a_{11} - \bar{x}\bar{y} \quad a_{11} = \frac{1}{N} \sum_i \sum_j x_i y_j n_{ij}$$

Si $S_{xy} > 0$ Relación directa
 Si $S_{xy} < 0$ Relación inversa
 Si $S_{xy} = 0$ No hay relación

CUIDADO PORQUE NO INFORMA DE LA FUERZA DE LA RELACIÓN ¿por qué?

- Habría que **descontar** la influencia que puedan tener las varianzas de X e Y

5. Covarianza y coeficiente de correlación

- El coeficiente de correlación lineal de Pearson (r) mide el grado de relación entre dos variables

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

- Tiene el mismo signo que la covarianza y nos informa sobre el grado de relación lineal entre las dos variables.
- Al ser un coeficiente, es adimensional y podrá tomar valores comprendidos entre +1 (correlación lineal directa máxima) y -1 (correlación lineal inversa máxima). $r = 0$ (o próximo) cuando no exista correlación

Ejemplo 1

X \ Y	1	2	4	6	$n_{i\bullet}$
1	2	0	1	1	4
3	3	1	0	1	5
5	0	1	0	5	6
$n_{\bullet j}$	5	2	1	7	15

$$a_{11} = \frac{1}{N} \sum_i \sum_j x_i y_j n_{ij}$$

$$\bar{x} = \frac{\sum_{i=1}^3 x_i n_{i\bullet}}{N} = \frac{1.4 + 3.5 + 5.6}{15} = \frac{49}{15} = 3,26$$

$$\bar{y} = \frac{\sum_{j=1}^4 y_j n_{\bullet j}}{N} = \frac{1.5 + 2.2 + 4.1 + 6.7}{15} = \frac{55}{15} = 3,6$$

$$a_{11} = \frac{\sum_{i=1}^3 \sum_{j=1}^4 x_i y_j n_{ij}}{N} = \frac{[1.1.2 + 1.2.0 + 1.4.1 + 1.6.1] + [3.1.3 + 3.2.1 + 3.4.0 + 3.6.1] + [5.1.0 + 5.2.1 + 5.4.0 + 5.6.5]}{15} = \frac{205}{15} = 13,66$$

$$S_{xy} = a_{11} - \bar{x} \bar{y} = 13,66 - 3,26 \cdot 3,6 = 1,924$$

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

Ejemplo 2

Las calificaciones obtenidas por un grupo de alumnos en Estadística (E) y Macroeconomía (M):

E	3	4	6	7	5	8	7	3	5	4	8	5	5	8	8	8	5
M	5	5	8	7	7	9	10	4	7	4	10	5	7	9	10	5	7

- Hallar la tabla de frecuencias
- Hallar las distribuciones marginales, media y varianza de las mismas
- Covarianza

Solución:

- La variable E (Estadística) toma seis valores diferentes. La variable M (Macroeconomía) toma siete valores distintos, por lo que para formar la tabla bastará hacer el recuento de las veces que se repite cada par.

E \ M	4	5	6	7	8	9	10	$n_{i\bullet}$
3	1	1						2
4	1	1						2
5		1		4				5
6					1			1
7				1			1	2
8		1				2	2	5
$n_{\bullet j}$	2	4	0	5	1	2	3	17

Ejemplo 2

b)

E_i	$n_{i\bullet}$	$E_i n_{i\bullet}$	$E_i^2 n_{i\bullet}$
3	2	6	18
4	2	8	32
5	5	25	125
6	1	6	36
7	2	14	98
8	5	40	320
	17	99	629

M_j	$n_{\bullet j}$	$M_j n_{\bullet j}$	$M_j^2 n_{\bullet j}$
4	2	8	32
5	4	20	100
6	0	0	0
7	5	35	245
8	1	8	64
9	2	18	162
10	3	30	300
	17	119	903

- Distribución Marginal de Estadística:

$$\bar{E} = a_{10} = \frac{\sum_{i=1}^6 E_i n_{i\bullet}}{N} = \frac{99}{17} = 5,82 \quad a_{20} = \frac{\sum_{i=1}^6 E_i^2 n_{i\bullet}}{N} = \frac{629}{17} = 37 \quad s_E^2 = a_{20} - a_{10}^2 = 37 - 5,82^2 = 3,13$$

- Distribución Marginal de Macroeconomía:

$$\bar{M} = a_{01} = \frac{\sum_{j=1}^7 M_j n_{\bullet j}}{N} = \frac{119}{17} = 7 \quad a_{02} = \frac{\sum_{j=1}^7 M_j^2 n_{\bullet j}}{N} = \frac{903}{17} = 53,11 \quad s_M^2 = a_{02} - a_{01}^2 = 53,11 - 7^2 = 4,11$$

Juan Miguel Navarro – UCAM - 25

Ejemplo 2

c) Para hallar la covarianza: $s_{xy} = a_{11} - a_{10} a_{01}$

$$a_{11} = \frac{\sum_{i=1}^6 \sum_{j=1}^7 E_i M_j n_{ij}}{N} = \frac{3.4.1 + 3.5.1 + 4.4.1 + 4.5.1 + 5.5.1 + 5.7.4 + 6.8.1 + 7.7.1 + 7.10.1 + 8.5.1 + 8.9.2 + 8.10.2}{17}$$

$$a_{11} = \frac{739}{17} = 43,47 \quad s_{xy} = a_{11} - a_{10} a_{01} = 43,47 - 5,82 \cdot 7 = 2,73$$

Juan Miguel Navarro – UCAM - 26

Ejemplo 3

A partir de la siguiente distribución bidimensional $(X_i, Y_j; n_{ij})$, calcular: $\bar{x}, \bar{y}, s_x^2, s_y^2$ y s_{xy} . ¿Son independientes las variables X e Y?

X \ Y	1	2	3
-1	0	1	0
0	1	0	1
1	0	1	0

Solución:

X \ Y	1	2	3	$n_{i\bullet}$
-1	0	1	0	1
0	1	0	1	2
1	0	1	0	1
$n_{\bullet j}$	1	2	1	4

Las variables X e Y son independientes cuando se verifica $\frac{n_{ij}}{N} = \left(\frac{n_{i\bullet}}{N}\right)\left(\frac{n_{\bullet j}}{N}\right) \forall i, j$

No son independientes porque no se verifica la relación: $\frac{0}{4} \neq \frac{2}{4} \cdot \frac{2}{4} \quad \left[\frac{n_{22}}{N} \neq \left(\frac{n_{2\bullet}}{N}\right)\left(\frac{n_{\bullet 2}}{N}\right) \right]$

Juan Miguel Navarro – UCAM - 27

Ejemplo 3

$$a_{11} = \frac{\sum_{i=1}^3 \sum_{j=1}^3 x_i y_j n_{ij}}{N} = \frac{1}{4} [-1 \cdot 2 \cdot 1 + 1 \cdot 2 \cdot 1] = 0$$

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^3 x_i n_{i\bullet}}{N} = \frac{1}{4} [-1 \cdot 1 + 0 \cdot 2 + 1 \cdot 1] = 0 \quad a_{20} = \frac{\sum_{i=1}^3 x_i^2 n_{i\bullet}}{N} = \frac{1}{4} [(-1)^2 \cdot 1 + 0 \cdot 2 + 1^2 \cdot 1] = \frac{2}{4} = 0,5$$

$$s_x^2 = a_{20} - a_{10}^2 = 0,5 - 0 = 0,5 \quad \mapsto \quad s_x = \sqrt{0,5} = 0,7$$

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^3 y_j n_{\bullet j}}{N} = \frac{1}{4} [1 \cdot 1 + 2 \cdot 2 + 3 \cdot 1] = 2 \quad a_{02} = \frac{\sum_{j=1}^3 y_j^2 n_{\bullet j}}{N} = \frac{1}{4} [1^2 \cdot 1 + 2^2 \cdot 2 + 3^2 \cdot 1] = \frac{18}{4} = 4,5$$

$$s_y^2 = a_{02} - a_{01}^2 = 4,5 - 2^2 = 0,5 \quad \mapsto \quad s_y = \sqrt{0,5} = 0,7$$

$$\text{covarianza } s_{xy} = a_{11} - a_{10} \cdot a_{01} = 0 - 0 \cdot 2 = 0$$

Adviértase que la covarianza es cero por la simetría de la distribución.

Si (X, Y) independientes	$\mapsto$	$s_{yx} = 0$
Si $s_{yx} = 0$	$\mapsto$	(X, Y) No independientes

Juan Miguel Navarro – UCAM - 28

6. Recta de Regresión

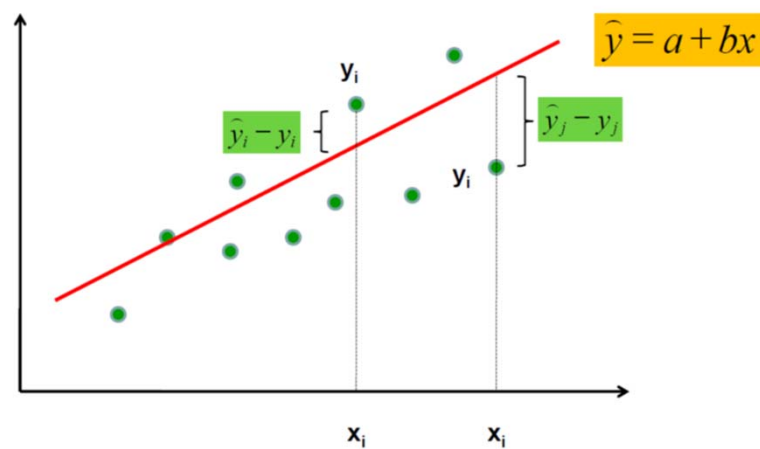
Regresión lineal

- Es una técnica estadística que se utiliza para estudiar la **relación entre variables** cuantitativas y normalmente lleva implícito el objetivo de predecir el comportamiento de una variable (dependiente) a partir de otra u otras variables conocidas (independientes o predictoras).
- Tras el análisis se obtendrá una ecuación denominada **ecuación de regresión**

Juan Miguel Navarro – UCAM - 29

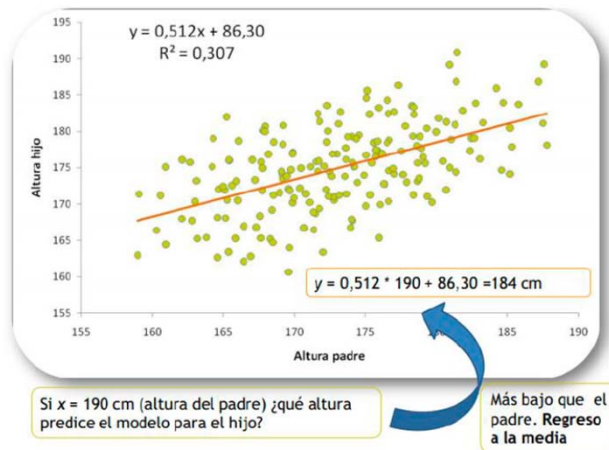
6. Recta de Regresión

Regresión lineal Simple



Juan Miguel Navarro – UCAM - 30

6. Recta de Regresión



Juan Miguel Navarro – UCAM - 31

6. Recta de Regresión

Regresión lineal Simple

Recta de regresión de
Y sobre X

$$\hat{y} = a + bx$$

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

$$b = \frac{S_{xy}}{S_x^2}$$

Recta de regresión de
X sobre Y

$$\hat{x} = a^* + b^*y$$

$$a^* = \bar{x} - b^*\bar{y}$$

$$\frac{1}{b^*} = \frac{S_y^2}{S_{xy}}$$

Pendiente de
ambas rectas

Juan Miguel Navarro – UCAM - 32

6. Recta de Regresión

Coefficiente de correlación lineal

En el caso de que se pueda afirmar que existe algún tipo de relación lineal, se hace preciso introducir un indicador del grado de relación. Dicho de otro modo, de hasta qué punto la recta de regresión representa o se ajusta adecuadamente al conjunto de puntos.

Esto se realiza a partir de la introducción del **coeficiente de correlación lineal** r

$$r = \frac{S_{xy}}{S_x S_y}$$

Iguel Navarro – UCAM - 33

6. Recta de Regresión

Coefficiente de correlación lineal

$$\frac{S_{xy}^2}{S_x^2 S_y^2} = r^2$$

A esta expresión, en el caso de regresión lineal, se la denomina Coeficiente de Determinación R^2 y puede considerarse igualmente como una medida del grado de ajuste de la recta al conjunto de puntos, variando ahora entre 0 y 1

Juan Miguel Navarro – UCAM - 34

6. Recta de Regresión

Ejemplo

Los datos de la producción de trigo en toneladas (X) y el precio del kilo de harina en pesetas (Y) en la década de los 80 en España fueron:

Producción de trigo	30	28	32	25	25	25	22	24	35	40
Precio de la harina	25	30	27	40	42	40	50	45	30	25

Ajusta la recta de regresión por el método de mínimos cuadrados

Resultados

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{10} x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^{10} x_i^2 - n \bar{x}^2} = \frac{9734 - 10 \times 28,6 \times 35,4}{8468 - 10 \times 28,6^2} = -1,3537$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 35,4 + 1,3537 \times 28,6 = 74,116$$

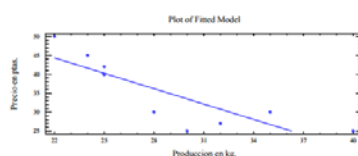
La recta de regresión es:

$$\hat{y} = 74,116 - 1,3537x$$

Juan Miguel Navarro – UCAM - 35

6. Recta de Regresión

Ejemplo



Regression Analysis - Linear model: Y = a + b*X

Dependent variable: Precio en ptas.

Independent variable: Produccion en kg.

Parameter	Estimate	Standard Error	T Statistic	P-Value
$\hat{\beta}_0$ Intercept	74,115	8,73577	8,4841	0,0000
$\hat{\beta}_1$ Slope	-1,3537	0,3002	-4,50924	0,0020

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	528,475	1	528,475	20,33	0,0020
Residual	207,925	8	25,9906		
Total (Corr.)	736,4	9			

Correlation Coefficient = -0,84714

R-squared = 71,7647 percent

Standard Error of Est. = 5,0981

Juan Miguel Navarro – UCAM - 36

6. Recta de Regresión

Las notas de 12 alumnos de una clase en Matemáticas y Física son las siguientes:

Matemáticas	Física
2	1
3	3
4	2
4	4
5	4
6	4
6	6
7	4
7	6
8	7
10	9
10	10

Hallar las rectas de regresión y representárlas.

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
2	1	2	4	1
3	3	9	9	9
4	2	8	16	4
4	4	16	16	16
5	4	20	25	16
6	4	24	36	16
6	6	36	36	36
7	4	28	49	16
7	6	42	49	36
8	7	56	64	49
10	9	90	100	81
10	10	100	100	100
72	60	431	504	380

1ª Hallamos las **medias aritméticas**.

$$\bar{x} = \frac{72}{12} = 6 \quad \bar{y} = \frac{60}{12} = 5$$

2ª Calculamos la **covarianza**.

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

3ª Calculamos las **varianzas**.

$$\sigma_x^2 = \frac{504}{12} - 6^2 = 6 \quad \sigma_y^2 = \frac{380}{12} - 25 = 6.66$$

4ª Recta de regresión de Y sobre X.

$$y - 5 = \frac{5.92}{6} (x - 6) \quad y = 0.987x - 0.922$$

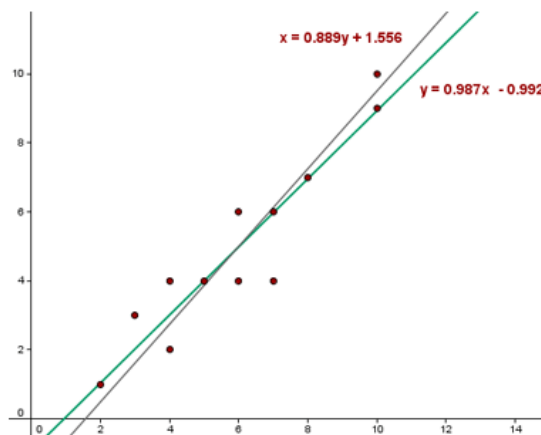
4ª Recta de regresión de X sobre Y.

$$x - 6 = \frac{5.92}{6.66} (y - 5) \quad x = 0.889y + 1.556$$

Juan Miguel Navarro – UCAM - 37

6. Recta de Regresión

Ejemplo



Juan Miguel Navarro – UCAM - 38

Ejemplo 1

Un psicólogo afirma, basándose en los datos obtenidos, que a medida que el niño crece menores son las respuestas inadecuadas que da en el transcurso de una situación experimental:

Edad	Número respuestas inadecuadas
2	11
3	12
4	10
4	13
5	11
5	9
6	10
7	7

Edad	Número respuestas inadecuadas
7	12
9	8
9	7
10	3
11	6
11	5
12	5

- Determinar la validez de las conclusiones del psicólogo
- María, de diez años y medio, participa en el experimento, ¿cuál es el número de respuestas inadecuadas que se puede predecir para ella?

Juan Miguel Navarro – UCAM - 39

Ejemplo 1

Solución:

- La validez de la afirmación se obtendrá en función del coeficiente de correlación: $r = \frac{s_{xy}}{s_x s_y}$

Como no hay pares repetidos se entiende que son 15 pares de la forma (x_i, y_i) que representará x_i : edad e y_i : número respuestas inadecuadas de modo que la frecuencia de cada par es la unidad.

x_i	2	3	4	4	5	5	6	7	7	9	9	10	11	11	12
y_i	11	12	10	13	11	9	10	7	12	8	7	3	6	5	5

$$a_{11} = \frac{\sum_{i=1}^{15} x_i y_i}{N} = \frac{2 \cdot 11 + 3 \cdot 12 + 4 \cdot 10 + \dots + 11 \cdot 5 + 12 \cdot 5}{15} = \frac{789}{15} = 52,6$$

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^{15} x_i}{N} = \frac{2 + 3 + 4 + 4 + 5 + \dots + 11 + 11 + 12}{15} = \frac{105}{15} = 7$$

$$a_{01} = \bar{y} = \frac{\sum_{i=1}^{15} y_i}{N} = \frac{11 + 12 + 10 + 13 + \dots + 6 + 5 + 5}{15} = \frac{129}{15} = 8,6$$

En consecuencia, $s_{xy} = a_{11} - a_{10} a_{01} = 52,6 - 7 \cdot 8,6 = -7,6$

Juan Miguel Navarro – UCAM - 40

Ejemplo 1

Para el cálculo de las desviaciones típicas (s_x, s_y):

$$a_{20} = \frac{\sum_{i=1}^{15} x_i^2}{N} = \frac{2^2 + 3^2 + 4^2 + 4^2 + 5^2 + \dots + 11^2 + 11^2 + 12^2}{15} = \frac{877}{15} = 58,46$$

$$a_{02} = \frac{\sum_{i=1}^{15} y_i^2}{N} = \frac{11^2 + 12^2 + 10^2 + 13^2 + \dots + 6^2 + 5^2 + 5^2}{15} = \frac{1237}{15} = 82,46$$

$$s_x^2 = a_{20} - a_{10}^2 = 58,46 - 7^2 = 9,46 \quad \mapsto \quad s_x = \sqrt{9,46} = 3,07$$

$$s_y^2 = a_{02} - a_{01}^2 = 82,46 - 8,6^2 = 8,5 \quad \mapsto \quad s_y = \sqrt{8,5} = 2,91$$

$$\text{El coeficiente de correlación: } r = \frac{s_{xy}}{s_x s_y} = \frac{-7,6}{3,07 \cdot 2,91} = -0,85 \quad \text{correlación inversa del 85\%}$$

La validez solicitada es del 85% en correlación inversa, es decir, a medida que aumenta la edad del niño (X) disminuye las respuestas inadecuadas (Y).

Juan Miguel Navarro – UCAM - 41

Ejemplo 1

b) Para poder predecir el número de respuestas para cada edad determinada (caso de María) será necesario hallar la ecuación de regresión de Y (nº respuestas inadecuadas) sobre X (edad del niño):

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x}) \quad \text{pendiente de la recta} = \text{coeficiente de regresión: } b_{yx} = \frac{s_{xy}}{s_x^2}$$

Adviértase que la pendiente de la recta o coeficiente de regresión b_{yx} viene determinado por el signo de la covarianza s_{xy}

$$b_{yx} = \frac{s_{xy}}{s_x^2} = \frac{-7,6}{9,46} = -0,80 \quad (\text{recta de regresión decreciente})$$

$$\text{La ecuación de la recta de regresión será: } y - 8,6 = -0,80 (x - 7) \quad \mapsto \quad y = 14,2 - 0,80x$$

En consecuencia, para la edad de María ($x = 10,5$) el número de respuestas inadecuadas que se puede predecir será:

$$y = 14,2 - 0,80 \cdot 10,5 = 5,8 \approx 6 \quad \text{respuestas inadecuadas.}$$

Juan Miguel Navarro – UCAM - 42

UCAM

Juan Miguel Navarro Ruiz

jmnavarro@ucam.edu

UCAM Universidad Católica de Murcia



© UCAM

UCAM

Tema 4

Números Índice

Estadística Aplicada al Turismo

Juan Miguel Navarro Ruiz

Grado en Turismo



UCAM

ÍNDICE CONTENIDOS

1. Introducción
2. Números índices simples
3. Números índices complejos
4. Operaciones con números índices
5. Tasas de variación

1. Introducción

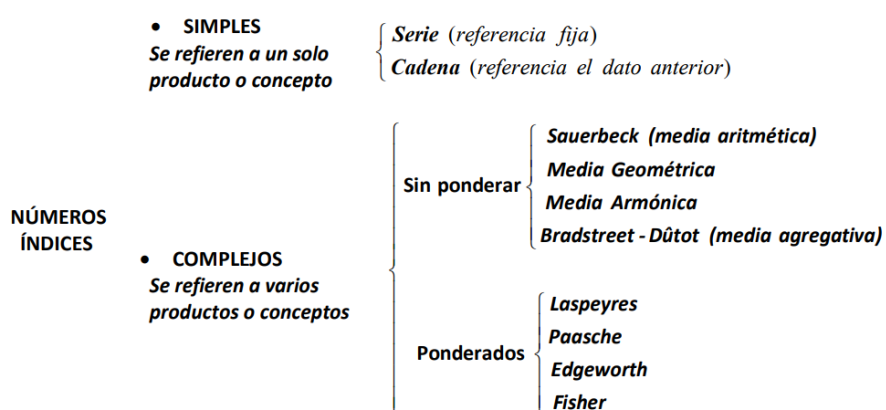
Un **NÚMERO ÍNDICE** es una medida estadística que nos proporciona la variación relativa de una magnitud simple o compleja a lo largo del tiempo o del espacio.

Lo más corriente es que se estudie a lo largo del tiempo.

Período inicial o base sobre el que se va comparando la evolución de la magnitud o variable estadística.

1. Introducción

Una clasificación sencilla de los números índices sería:



2. Números Índices Simples

NÚMEROS ÍNDICES SIMPLES.- Son los índices que proporcionan la variación que ha sufrido una magnitud o concepto entre dos períodos o lugares distintos. Generalmente, esta comparación se realiza con el valor de un período fijo (*periodo base*).

Dependiendo de si la referencia es fija o no, se habla de **índices en serie** (referencia fija) e **índices en cadena** (referencia variable).

2. Números Índices Simples

En serie

NÚMEROS ÍNDICES SIMPLES EN SERIE.- Sean x_t y x_0 dos valores de una variable X , el valor del número índice en serie que corresponde al valor x_t tomando como referencia o base fija x_0 se representa mediante $I_0^t(X)$ y se define:

$$I_0^t(X) = \frac{x_t}{x_0} \cdot 100$$

2. Números Índices Simples

En serie

Estudian la evolución a lo largo del tiempo de una magnitud que tiene un solo componente (no admite desagregación)

Dada una magnitud simple X_t , su número índice en el período t será el siguiente:

$$I_{it} = \frac{X_{it}}{X_{i0}} \cdot 100$$

Donde:

X_{it} : valor de la magnitud en el periodo t

X_{i0} : valor de la magnitud en el periodo base

Juan Miguel Navarro – UCAM - 7

2. Números Índices Simples

En serie

Ejemplo 1.- En la tabla se presenta el número de mujeres (en miles) activas en España desde el tercer trimestre de 2009 hasta el tercer trimestre de 2010. En la última columna se representan los números índices simples en serie con base el tercer trimestre de 2009.

NÚMEROS ÍNDICES SIMPLES EN SERIE

Año	Trimestre	Mujeres activas (miles)	Base (2009-3º)
2009	3	10089,4	100
2009	4	10139,3	$(10139,3 / 10089,4) \cdot 100 = 100,4946$
2010	1	10213,3	$(10213,3 / 10089,4) \cdot 100 = 101,2280$
2010	2	10250,5	$(10250,5 / 10089,4) \cdot 100 = 101,5967$
2010	3	10265,2	$(10265,2 / 10089,4) \cdot 100 = 101,7424$

Los índices reflejan la variación porcentual que experimentan los distintos valores de la variable con respecto al valor que se ha tomado como referencia (3º trimestre de 2009).

Observando la tabla, el número de mujeres activas en España en el tercer trimestre de 2010 es un 1,74% superior al que había en el tercer trimestre del año anterior.

Los índices que se obtienen respecto de una base (periodo de referencia) fija se denominan **índices en serie**.

Juan Miguel Navarro – UCAM - 8

2. Números Índices Simples

En cadena

NÚMEROS ÍNDICES SIMPLES EN CADENA.- Cuando el índice correspondiente a cada dato se calcula tomando como referencia el dato inmediatamente anterior.

Sean x_{t-1} y x_t los valores observados de una variable X en dos instantes consecutivos, el índice en cadena que corresponde al valor x_t se representa mediante IC^t y se define:

$$IC^t = \frac{x_t}{x_{t-1}} \cdot 100$$

Para series de observaciones temporales, estos índices reflejan la variación porcentual que experimenta la variable entre cada dos observaciones consecutivas.

NÚMEROS ÍNDICES SIMPLES EN CADENA

Año	Trimestre	Mujeres activas (miles)	Base(2009-3º)
2009	3	10089,4	-----
2009	4	10139,3	$(10139,3 / 10089,4) \cdot 100 = 100,49$
2010	1	10213,3	$(10213,3 / 10139,3) \cdot 100 = 100,73$
2010	2	10250,5	$(10250,5 / 10213,3) \cdot 100 = 100,36$
2010	3	10265,2	$(10265,2 / 10250,5) \cdot 100 = 100,14$

Juan Miguel Navarro – UCAM - 9

2. Números Índices Simples

En cadena

Los índices en cadena reflejan la variación porcentual entre trimestres del número de mujeres activas en España.

En esta línea, en el segundo trimestre de 2010 el número de mujeres activas fue un 0,36% superior al dato del trimestre anterior.

$$IC_{2010-2}^{2010-3} = \frac{10250,5}{10213,3} \cdot 100 = 100,36$$

Juan Miguel Navarro – UCAM - 10

2. Números Índices Simples

Relación entre índices en serie y en cadena

- Los índices en cadena se pueden obtener a partir de los índices en serie

$$IC^t = \frac{x_t}{x_{t-1}} \cdot 100 = \frac{\frac{x_t}{x_0}}{\frac{x_{t-1}}{x_0}} \cdot 100 = \frac{I_0^t}{I_0^{t-1}} \cdot 100$$

En el ejemplo, $IC_{2010-2}^{2010-3} = \frac{101,5967}{101,2280} \cdot 100 = 100,36$

- Los índices en serie se pueden obtener a partir de los índices en cadena

$$I_0^t(X) = \frac{x_t}{x_0} \cdot 100 = \frac{x_t}{x_{t-1}} \cdot \frac{x_{t-1}}{x_{t-2}} \cdot \frac{x_{t-2}}{x_{t-3}} \cdot \dots \cdot \frac{x_2}{x_1} \cdot \frac{x_1}{x_0} \cdot 100 = \frac{IC^t(X)}{100} \cdot \frac{IC^{t-1}(X)}{100} \cdot \dots \cdot \frac{IC^2(X)}{100} \cdot \frac{IC^1(X)}{100} \cdot 100$$

En el ejemplo, $I_0^{2010-3}(X) = \frac{100,14}{100} \cdot \frac{100,36}{100} \cdot \frac{100,73}{100} \cdot \frac{100,49}{100} \cdot 100 = 101,7424$

Juan Miguel Navarro – UCAM - 11

2. Números Índices Simples

Ejemplo 2.- En la tabla adjunta recoge los índices en cadena trimestrales para el número de parados en los sectores de la construcción y servicios en España

Año	Trimestre	IC Parados construcción	IC Parados servicios
2009	1	-----	-----
2009	2	94,37	101,33
2009	3	88,64	95,84
2009	4	98,79	100,7
2010	1	97,87	106,35
2010	2	87,71	95,91
2010	3	87,4	96,04

$$\frac{96,04}{100} \cdot \frac{95,91}{100} \cdot \frac{106,35}{100} \cdot \frac{100,70}{100} \cdot 100$$

a) Determinar la variación porcentual que experimento el número de parados en el sector servicios durante el tercer trimestre de 2009 al tercer trimestre de 2010.

b) Sabiendo que en el sector de la construcción el número de parados ascendió a 527,6 miles de personas durante el segundo trimestre de 2010. Hallar la serie expresada en miles de trabajadores parados en la construcción.

Juan Miguel Navarro – UCAM - 12

2. Números Índices Simples

Solución: En el apartado (a)

$$\begin{aligned}
 IC_{2009-3}^{2010-3}(\text{parados servicios}) &= \frac{IC^{2010-3}}{100} \cdot \frac{IC^{2010-2}}{100} \cdot \frac{IC^{2010-1}}{100} \cdot \frac{IC^{2009-4}}{100} \cdot 100 = \\
 &= \frac{96,04}{100} \cdot \frac{95,91}{100} \cdot \frac{106,35}{100} \cdot \frac{100,70}{100} \cdot 100 = 98,6468
 \end{aligned}$$

En consecuencia, la variación porcentual que corresponde al periodo comprendido entre el tercer trimestre de 2009 y el tercer trimestre de 2010 es de $[98,6468 - 100 = -1,3532\%]$

Juan Miguel Navarro – UCAM - 13

2. Números Índices Simples

b) En la construcción, primero se obtienen los **índices en serie** con base primer trimestre de 2009:

$$I_{2009-2}^{2009-3}(X) = \frac{IC^{2009-3}}{100} \cdot \frac{IC^{2009-2}}{100} \cdot 100 = \frac{88,64}{100} \cdot \frac{94,37}{100} \cdot 100 = 83,6496$$

$$I_{2009-2}^{2009-4}(X) = \frac{98,79}{100} \cdot \frac{88,64}{100} \cdot \frac{94,37}{100} \cdot 100 = 82,6374$$

Año	Trimestre	IC Parados construcción	I Parados (2009-100)	Miles de parados (construcción)
2009	1	-----	100	$x_{2009-1} = 743,754$
2009	2	94,37	94,3700	$743,754 \times 94,37 = 701,881$
2009	3	88,64	83,6496	$743,754 \times 83,6496 = 622,147$
2009	4	98,79	82,6374	$743,754 \times 82,6374 = 614,619$
2010	1	97,87	80,8772	$743,754 \times 80,8772 = 601,528$
2010	2	87,71	70,9374	527,600
2010	3	87,4	61,9993	$743,754 \times 61,9993 = 461,122$

Utilizando el dato de 527,6 mil parados para el segundo trimestre de 2010 se calcula el dato de paro para el primer trimestre de 2009:

$$I_0^t = \frac{x_t}{x_0} \cdot 100 \rightarrow \frac{527,6}{x_{2009-1}} \cdot 100 = 70,9374 \rightarrow x_{2009-1} = 743,754$$

Juan Miguel Navarro – UCAM - 14

2. Números Índices Simples

ÍNDICES SIMPLES MÁS UTILIZADOS

- **PRECIO RELATIVO:** Relación entre el precio de un bien en el período actual p_{it} y el precio del mismo en el período base p_{i0} : $p_0^t = \frac{p_{it}}{p_{i0}} \cdot 100$
- **CANTIDAD RELATIVA:** Razón entre la cantidad producida o vendida de un bien en sus períodos actual q_{it} y base q_{i0} : $q_0^t = \frac{q_{it}}{q_{i0}} \cdot 100$
- **VALOR RELATIVO:** Valor de un bien en un período cualquiera se define como el producto del precio de ese bien y la cantidad producida (vendida). El valor relativo será la razón entre los valores de ese bien en el período actual ($p_{it} \cdot q_{it}$) y en el período base ($p_{i0} \cdot q_{i0}$):

$$V_0^t = \frac{V_t}{V_0} = \frac{p_{it} \cdot q_{it}}{p_{i0} \cdot q_{i0}} \cdot 100 = \left(\frac{p_{it}}{p_{i0}} \right) \cdot \left(\frac{q_{it}}{q_{i0}} \right) \cdot 100 = p_0^t \cdot q_0^t \cdot 100$$

El valor relativo de un bien es igual al producto de su precio relativo y su cantidad relativa.

Juan Miguel Navarro – UCAM - 15

2. Números Índices Simples

Precio relativo

Estudian la evolución del precio de un único producto.

Dado el precio de un producto p_i :

$$P_{it} = \frac{p_{it}}{p_{i0}} \cdot 100$$

Ejemplo 1. Los precios corrientes de las mesas de escritorio de una determinada marca en el periodo 2007-2012 han sido: 240, 242, 245, 250, 255, 260. Obtenga la serie de números índices simples de la magnitud precio de las mesas tomando como periodo base 2007.

Juan Miguel Navarro – UCAM - 16

2. Números Índices Simples

Precio relativo

t	P _i	$P_i = P_{it}/P_{i0} * 100$
2007	240	100
2008	242	100.83
2009	245	102.08
2010	250	104.17
2011	255	106.25
2012	260	108.33

El precio ha crecido un 0.83% del 2007 al 2008, un 2.08% del 2007 al 2009, un 4.17% del 2007 al 2010,... En el periodo completo, el incremento del nivel de precios ha sido de un 8.33%

Juan Miguel Navarro – UCAM - 17

2. Números Índices Simples

Ejemplo - Se desea conocer la evolución del precio de la barra de pan ente 2005 y 2010 en España. Para ello se dispone de la siguiente información:

Años	Precio barra de pan (céntimos euro)	Índices
		Variación precio barra de pan
2005	25	100
2006	30	$\frac{P_{2006}}{P_{2005}} = \frac{30}{25} \cdot 100 = 120$
2007	32	$\frac{P_{2007}}{P_{2005}} = \frac{32}{25} \cdot 100 = 128$
2008	38	$\frac{P_{2008}}{P_{2005}} = \frac{38}{25} \cdot 100 = 152$
2009	44	$\frac{P_{2009}}{P_{2005}} = \frac{44}{25} \cdot 100 = 176$
2010	48	$\frac{P_{2010}}{P_{2005}} = \frac{48}{25} \cdot 100 = 192$

Calculada la serie de índices de variación, se observa que el precio de la barra de pan en 2007 fue 1,28 veces el de 2005; el de 2010 fue 1,92 veces la de 2005, y así sucesivamente.

Señalar que el índice es una medida adimensional, numerador y denominador vienen dados en las mismas unidades de medida.

Juan Miguel Navarro – UCAM - 18

3. Números Índices Complejos

ÍNDICES COMPLEJOS.- Generalmente el interés no se encuentra en comparar precios, cantidades o valores individuales, sino que se comparan fenómenos del mundo real donde intervienen muchas variables. Como consecuencia, la información suministrada por los índices de diferentes bienes debe de ser resumida en un único índice al que se denomina **índice complejo**.

La construcción de un índice complejo no es una tarea fácil. Para elaborar la evolución del coste de la vida de un país (IPC en España) habría que seleccionar un grupo de bienes que reflejaran dicho coste, teniendo en cuenta la importancia relativa de cada uno de esos bienes, decidiendo finalmente la forma de unificar toda la información para obtener un único índice.

El objetivo es llegar a un número índice sencillo que reúna la mayor cantidad posible de información.

De esta manera, se llega a dos tipos de índices complejos: **índices complejos no ponderados** (cuando prima la sencillez) e **índices complejos ponderados** (cuando se desea que contengan la mayor cantidad de información).

Juan Miguel Navarro – UCAM - 19

3. Números Índices Complejos

ÍNDICES COMPLEJOS DE PRECIOS NO PONDERADOS.- Mediante los **Índices de Precios** se analiza el estudio de magnitudes económicas, que cuantifican la evolución de la magnitud precio de un conjunto de bienes y servicios.

Se tendría la información que proporciona un cuadro análogo al siguiente:

Artículos	1	2		n
Épocas				
0	p_{10}	p_{20}		p_{n0}
1	p_{11}	p_{21}		p_{n1}
2	p_{12}	p_{22}		p_{n2}
⋮	⋮	⋮		⋮
⋮	⋮	⋮		⋮
t	p_{1t}	p_{2t}		p_{nt}

Artículos	1	2	...	n
Índices simples	$\frac{p_{1t}}{p_{10}} 100$	$\frac{p_{2t}}{p_{20}} 100$	...	$\frac{p_{nt}}{p_{n0}} 100$

El objetivo será encontrar una medida estadística que resuma toda la información y permita conocer cuál ha sido la variación experimentada por los precios en el período t respecto al período base.

Para resumir la información obtenida a través de los índices simples, es lógico promediar éstos. De este modo, los índices complejos van a ser medias aritméticas, geométricas, armónicas y agregativas de los índices simples.

Señalar que estos cuatro tipos de índices compuestos sin ponderar se pueden utilizar para estudiar la evolución de cualquier otra variable distinta del precio.

Juan Miguel Navarro – UCAM - 20

3. Números Índices Complejos

ÍNDICE DE SAUERBECK: Considerando los precios relativos $I_i = \frac{p_{it}}{p_{i0}}$, es la media aritmética no

ponderada de los índices simples: $S_p = \frac{1}{n} \cdot \sum_{i=1}^n \frac{p_{it}}{p_{i0}} \cdot 100$

ÍNDICE MEDIA GEOMÉTRICA: $I_0^t = \sqrt[n]{\prod_{i=1}^n \frac{p_{it}}{p_{i0}}} \cdot 100$

3. Números Índices Complejos

ÍNDICE MEDIA ARMÓNICA: $I_0^t = \frac{n}{\sum_{i=1}^n \frac{p_{i0}}{p_{it}}} \cdot 100$

De los tres índices el que se utiliza con mayor frecuencia es el índice de Sauerbeck.

ÍNDICE MEDIA AGREGATIVA SIMPLE O DE BRADSTREET-DÛTOT: Consiste en considerar un índice simple de agregados de magnitudes (precios). Es decir, se calcula la razón de la media aritmética de los precios de n artículos (en el período t como en el período base):

$$B-D_p = \frac{\sum_{i=1}^n p_{it}}{\sum_{i=1}^n p_{i0}} \cdot 100$$

Señalar que los índices analizados tienen la ventaja de ser fáciles de aplicar, pero presentan inconvenientes importantes: << No tienen en cuenta la importancia relativa de cada uno de los diferentes artículos en el conjunto total, puesto que no son ponderados >>

3. Números Índices Complejos

Ejemplo 7.- En la tabla adjunta aparecen distintos artículos y los precios (en céntimos de euros) entre 2008 y 2010. Se pide calcular los índices compuestos.

Artículos	Precios		
	2008	2009	2010
Pan	38	44	48
Huevos	130	150	215
Leche	88	100	110
Pollo	160	190	205

Solución:

Índice de Sauerbeck: $S_p = \frac{1}{n} \sum_{i=1}^n \frac{p_{it}}{p_{i0}} \cdot 100$ (media aritmética simple)

$$S_{p_{2009}} = \frac{1}{4} \sum_{i=1}^4 \frac{p_{it}}{p_{i0}} \cdot 100 = \frac{1}{4} \left[\frac{44}{38} + \frac{150}{130} + \frac{100}{88} + \frac{190}{160} \right] \cdot 100 = 115,89$$

$$S_{p_{2010}} = \frac{1}{4} \sum_{i=1}^4 \frac{p_{it}}{p_{i0}} \cdot 100 = \frac{1}{4} \left[\frac{48}{38} + \frac{215}{130} + \frac{110}{88} + \frac{205}{160} \right] \cdot 100 = 136,21$$

Índice media Geométrica: $I_t^g = \sqrt[n]{\prod_{i=1}^n \frac{p_{it}}{p_{i0}}} \cdot 100$

$$I_{2009}^{2008} = \sqrt[4]{\frac{44}{38} \cdot \frac{150}{130} \cdot \frac{100}{88} \cdot \frac{190}{160}} \cdot 100 = 115,88$$

$$I_{2010}^{2008} = \sqrt[4]{\frac{48}{38} \cdot \frac{215}{130} \cdot \frac{110}{88} \cdot \frac{205}{160}} \cdot 100 = 135,25$$

Índice media Armónica: $I_0^t = \frac{n}{\sum_{i=1}^n \frac{p_{i0}}{p_{it}}} \cdot 100$

$$I_{2009}^{2008} = \frac{4}{\frac{38}{44} + \frac{130}{150} + \frac{88}{100} + \frac{160}{190}} \cdot 100 = 115,86$$

$$I_{2010}^{2008} = \frac{4}{\frac{38}{48} + \frac{130}{215} + \frac{88}{110} + \frac{160}{205}} \cdot 100 = 134,37$$

Índice media agregativa simple o de Bradstreet-Dútót: $B-D_p = \frac{\sum_{i=1}^n p_{it}}{\sum_{i=1}^n p_{i0}} \cdot 100$

$$B-D_{p_{2009}} = \frac{\sum_{i=1}^4 p_{it}}{\sum_{i=1}^4 p_{i0}} \cdot 100 = \frac{44 + 150 + 100 + 190}{38 + 130 + 88 + 160} \cdot 100 = 116,35$$

$$B-D_{p_{2010}} = \frac{\sum_{i=1}^4 p_{it}}{\sum_{i=1}^4 p_{i0}} \cdot 100 = \frac{48 + 215 + 110 + 205}{38 + 130 + 88 + 160} \cdot 100 = 138,94$$

3. Números Índices Complejos

Estudian la evolución a lo largo del tiempo de una magnitud que tiene más de una componente (admite desagregación)

Números Índices Complejos ponderados

Se le asigna un determinado coeficiente de ponderación W_i a cada componente de la magnitud compleja.

Por ejemplo: Índice de Precios al Consumo, Índice de Producción Industrial,...

$$I_t = \frac{\sum_{i=1}^N I_{it} W_i}{\sum_{i=1}^N W_i}$$

3. Números Índices Complejos

INDICES COMPLEJOS DE PRECIOS PONDERADOS.- Una presentación sobre los sistemas de ponderaciones propuestos tradicionalmente:

$p_{i0} \cdot q_{i0} \equiv$ valor de la cantidad consumida del bien i-ésimo en el período base, a precios de período base. *(situación real)*

$p_{i0} \cdot q_{it} \equiv$ valor a precios del período base de la cantidad consumida del bien i-ésimo en el período actual. *(situación con valoración ficticia)*

Los índices complejos ponderados más utilizados son: Laspeyres, Paasche, Edgeworth y Fisher.

Juan Miguel Navarro – UCAM - 25

3. Números Índices Complejos

Ponderaciones utilizadas en el Índice de Precios al Consumo

Grupo	2011	2012
01. Alimentos y bebidas no alcohólicas	18,16	18,26
02. Bebidas alcohólicas y tabaco	2,87	2,89
03. Vestido y calzado	8,59	8,34
04. Vivienda	11,70	12,00
05. Menaje	6,84	6,67
06. Medicina	3,21	3,14
07. Transporte	14,74	15,16
08. Comunicaciones	3,98	3,85
09. Ocio y Cultura	7,64	7,54
10. Enseñanza	1,38	1,42
11. Hoteles, cafés y restaurantes	11,52	11,46
12. Otros bienes y servicios	9,37	9,26
Total	100	100

Miguel Navarro – UCAM - 26

3. Números Índices Complejos

Analizan las variaciones debidas a los cambios en los precios de un conjunto de artículos ponderándolos siempre por las mismas cantidades.

El índice de Laspeyres se define como la media aritmética ponderada de los índices simples de precios. El criterio de ponderación es $p_{i0} \cdot q_{i0}$, con lo cual:

i. LASPEYRES

Ponderación: $W_i = p_{i0} \cdot q_{i0}$

$$P_L = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{i0}} \cdot W_i}{\sum_{i=1}^N W_i} \cdot 100 = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{i0}}{\sum_{i=1}^N p_{i0} \cdot q_{i0}} \cdot 100$$

$$P_L = \frac{\sum_{i=1}^N p_{it} \cdot q_{i0}}{\sum_{i=1}^N p_{i0} \cdot q_{i0}} \cdot 100$$

Ventaja: ponderaciones fijas todos los periodos

Inconveniente: la representatividad disminuye a medida que nos alejamos del periodo base

Juan Miguel Navarro – UCAM - 27

3. Números Índices Complejos

El índice de Laspeyres se cuestiona en ocasiones, ya que parece poco realista suponer que las cantidades compradas o adquiridas en el año de referencia no varían en el tiempo.

Como ejemplo, no parece muy realista la hipótesis de que en años de sequía, y en consecuencia, de subidas importantes de los precios de los productos agrarios, las cantidades demandadas sean iguales.

ii. PAASCHE

Ponderación: $W_i = p_{it} \cdot q_{it}$

$$P_P = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{i0}} \cdot W_i}{\sum_{i=1}^N W_i} \cdot 100 = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{it}}{\sum_{i=1}^N p_{i0} \cdot q_{it}} \cdot 100$$

$$P_P = \frac{\sum_{i=1}^N p_{it} \cdot q_{it}}{\sum_{i=1}^N p_{i0} \cdot q_{it}} \cdot 100$$

Ventaja: actualización de las ponderaciones cada periodo

Inconveniente: se complica el cálculo respecto a ponderaciones fijas

Juan Miguel Navarro – UCAM - 28

3. Números Índices Complejos

iii. EDGEWORTH

Ponderación: $W_i = p_{i0} \cdot q_{i0} + p_{it} \cdot q_{it}$

$$P_E = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{i0}} \cdot W_i}{\sum_{i=1}^N W_i} \cdot 100 = \frac{\sum_{i=1}^N \frac{p_{it}}{p_{i0}} \cdot (p_{i0} \cdot q_{i0} + p_{i0} \cdot q_{it})}{\sum_{i=1}^N (p_{i0} \cdot q_{i0} + p_{i0} \cdot q_{it})} \cdot 100$$

$$P_E = \frac{\sum_{i=1}^N p_{it} \cdot q_{i0} + \sum_{i=1}^N p_{it} \cdot q_{it}}{\sum_{i=1}^N p_{i0} \cdot q_{i0} + \sum_{i=1}^N p_{i0} \cdot q_{it}} \cdot 100$$

Es una medida agregativa ponderada de precios cuyo coeficiente de ponderación es $(q_{i0} + q_{it})$:

Juan Miguel Navarro – UCAM - 29

3. Números Índices Complejos

I. Fisher propuso como número índice de precios la media geométrica de los índices de precios de Laspeyres y Paasche, es decir:

iv. FISHER

$$P_F = \sqrt{P_L \cdot P_P}$$

Juan Miguel Navarro – UCAM - 30

3. Números Índices Complejos

Ejemplo Se dispone también de las cantidades vendidas de cada componente. Calcule los índices de precios complejos ponderados de 2012 tomando como periodo base 2010.

Artículos	2010		2011		2012	
	p_{i10}	q_{i10}	p_{i11}	q_{i11}	p_{i12}	q_{i12}
Silla	90	100	100	105	105	110
Mesa	250	260	255	250	260	270
Estante	49	150	55	125	59	160

Juan Miguel Navarro – UCAM - 31

3. Números Índices Complejos

Coeficientes de ponderación

Artículos	$p_{i12}q_{i10}$	$p_{i10}q_{i10}$	$p_{i12}q_{i12}$	$p_{i10}q_{i12}$
Silla	$105 \cdot 100$	$90 \cdot 100$	$105 \cdot 110$	$90 \cdot 110$
Mesa	$260 \cdot 260$	$250 \cdot 260$	$260 \cdot 270$	$250 \cdot 270$
Estante	$59 \cdot 150$	$49 \cdot 150$	$59 \cdot 160$	$49 \cdot 160$

Juan Miguel Navarro – UCAM - 32

3. Números Índices Complejos

Artículos	$P_{i12}q_{i10}$	$P_{i10}q_{i10}$	$P_{i12}q_{i12}$	$P_{i10}q_{i12}$
Silla	10500	9000	11550	9900
Mesa	67600	65000	70200	67500
Estante	8850	7350	9440	7840
	86950	81350	91190	85240

$$P_L = \frac{86950}{81350} \cdot 100 = 106.88$$

$$P_P = \frac{91190}{85240} \cdot 100 = 106.98$$

$$P_E = \frac{86950 + 91190}{81350 + 85240} \cdot 100 = 106.93 \quad P_F = \sqrt{106.88 \cdot 106.98} = 106.93$$

Los precios del mobiliario de oficina han crecido un 6.88% de 2010 a 2012, según Laspeyres, un 6.98% según Paasche, un 6.93%, según Edgeworth y un 6.93% según Fisher.

Juan Miguel Navarro – UCAM - 33

3. Números Índices Complejos

Ejemplo con ponderados

Ejemplo Supongamos que en el ejemplo 7 disponemos de información adicional sobre la cantidad vendida en cada uno de los períodos, como se detalla en la tabla adjunta. Determinar los índices de Laspeyres, Paasche, Edgeworth y Fisher para 2010, siendo el año base 2008.

Artículos	2008		2009		2010	
	precios	cantidad vendida	precios	cantidad vendida	precios	cantidad vendida
Pan	38	150	44	200	48	240
Huevos	130	400	150	580	215	560
Leche	88	700	100	780	110	925
Pollo	160	400	190	400	205	375

Juan Miguel Navarro – UCAM - 34

3. Números Índices Complejos

Solución:

Artículos	Laspeyres		Paasche		Edgeworth		
	$p_{110} \cdot q_{108}$	$p_{108} \cdot q_{108}$	$p_{110} \cdot q_{110}$	$p_{108} \cdot q_{110}$	$(q_{108} + q_{110})$	$p_{110} \cdot (q_{108} + q_{110})$	$p_{108} \cdot (q_{108} + q_{110})$
Pan	7200	5700	11520	9120	390	18720	14820
Huevos	86000	52000	120400	72800	960	206400	124800
Leche	77000	61600	101750	81400	1625	178750	143000
Pollo	82000	64000	76875	60000	775	158875	124000
	252200	183300	310545	223320		562745	406620

$$\text{Índice de Laspeyres: } L_{p_{2008}}^{2010} = \frac{\sum_{i=1}^4 p_{110} \cdot q_{108}}{\sum_{i=1}^4 p_{108} \cdot q_{108}} \cdot 100 = \frac{252200}{183300} \cdot 100 = 137,59$$

$$F_{p_{2008}}^{2010} = \sqrt{L_{p_{2008}}^{2010} \cdot P_{p_{2008}}^{2010}} = \sqrt{137,59 \cdot 139,06} = 138,32$$

$$\text{Índice de Paasche: } P_{p_{2008}}^{2010} = \frac{\sum_{i=1}^4 p_{110} \cdot q_{110}}{\sum_{i=1}^4 p_{108} \cdot q_{110}} \cdot 100 = \frac{310545}{223320} \cdot 100 = 139,06$$

$$\text{Índice de Edgeworth: } E_{p_{2008}}^{2010} = \frac{\sum_{i=1}^4 p_{110} \cdot (q_{108} + q_{110})}{\sum_{i=1}^4 p_{108} \cdot (q_{108} + q_{110})} \cdot 100 = \frac{562745}{406620} \cdot 100 = 138,40$$

Juan Miguel Navarro – UCAM - 35

3. Números Índices Complejos

PROBLEMAS CON LA UTILIZACIÓN DE NÚMEROS ÍNDICES.- Fundamentalmente son referentes a dos cuestiones:

PONDERACIONES.- En la medida de lo posible, el tipo de ponderación debe reflejar la importancia relativa de cada bien en particular. En los índices expuestos las ponderaciones más apropiadas se basan en cantidades o valores para los índices de precios, y en precios o valores para los índices de cantidad.

En la práctica, cada bien incluido en un índice complejo se suele interpretar como representativo de toda la clase de artículos relacionados y no como bien individual. En este sentido, la ponderación asignada a cada artículo individual refleja la importancia de toda la clase que representa.

PERÍODO BASE.- Es aquél período con respecto al que se efectúan las comparaciones, por lo que para que muchas comparaciones no pierdan significado, se suele elegir como tal un período no alejado excesivamente del período corriente. En esta línea, se hace necesario renovar periódicamente la información relativa al año base.

Juan Miguel Navarro – UCAM - 36

4. Operaciones con Números Índices

Recordar: Relación entre índices en serie y en cadena

- Los índices en cadena se pueden obtener a partir de los índices en serie

$$IC^t = \frac{x_t}{x_{t-1}} \cdot 100 = \frac{\frac{x_t}{x_{t-1}}}{\frac{x_{t-1}}{x_0}} \cdot 100 = \frac{l_0^t}{l_0^{t-1}} \cdot 100$$

$$\text{En el ejemplo, } IC_{2010-2}^{2010-3} = \frac{101,5967}{101,2280} \cdot 100 = 100,36$$

- Los índices en serie se pueden obtener a partir de los índices en cadena

$$l_0^t(X) = \frac{x_t}{x_0} \cdot 100 = \frac{x_t}{x_{t-1}} \cdot \frac{x_{t-1}}{x_{t-2}} \cdot \frac{x_{t-2}}{x_{t-3}} \cdot \dots \cdot \frac{x_2}{x_1} \cdot \frac{x_1}{x_0} \cdot 100 = \frac{IC^1(X)}{100} \cdot \frac{IC^{t-1}(X)}{100} \cdot \dots \cdot \frac{IC^2(X)}{100} \cdot \frac{IC^1(X)}{100} \cdot 100$$

$$\text{En el ejemplo, } l_0^{2010-3}(X) = \frac{100,14}{100} \cdot \frac{100,36}{100} \cdot \frac{100,73}{100} \cdot \frac{100,49}{100} \cdot 100 = 101,7424$$

Juan Miguel Navarro – UCAM - 37

4. Operaciones con Números Índices

Cambio de base

Los **índices en cadena** son enlaces de índices cuya base es siempre el periodo de comparación del índice precedente.

$$I_{ot} = I_{01} \cdot I_{12} \cdot I_{23} \cdot \dots \cdot I_{t-1,t}$$

El **cambio de base** consiste en transformar una serie con un periodo base 0 en una nueva serie con un periodo base t':

$$I_{t't} = I_{0t} / I_{0t'} \cdot 100$$

donde $1/I_{0t'}$ es el coeficiente de transformación de la serie.

Juan Miguel Navarro – UCAM - 38

4. Operaciones con Números Índices

CAMBIOS DE BASE Ó REVISIÓN DE LA BASE EN ÍNDICES SIMPLES.- Al alejarse del período base el índice sufre una pérdida de representatividad, en especial cuando para ponderar magnitudes actuales se utilizan precios relativos referidos al período base. Este problema se resuelve haciendo un **cambio de base** a período más próximo al actual.

Para relacionar series de índices referidos a distintos períodos base se utilizan **enlaces técnicos** entre ambas series.

Período	Índice (período 0)	Índice (período h)
0	I_0^0	I_h^0
1	I_0^1	I_h^1
⋮	⋮	⋮
i	I_0^i	I_h^i
⋮	⋮	⋮
h	I_0^h	I_h^h
⋮	⋮	⋮
t	I_0^t	I_h^t

La nueva serie de índices se obtiene:

$$I_h^i = \frac{I_0^i}{I_0^h} \cdot I_h^h = \frac{I_0^i}{I_0^h}$$

donde I_0^h es el índice que hace de **enlace técnico** entre las dos series.

Juan Miguel Navarro – UCAM - 39

4. Operaciones con Números Índices

Ejemplo Dada la serie adjunta con base año 2000, se desea cambiar la base al año 2005

Años	Precio refresco (euros)	Índices Simples Base 2000	Índices Simples Base 2005
2000	1,2	100	$(100 / 145) \cdot 100 = 68,97$
2001	1,3	$(1,3 / 1,2) \cdot 100 = 108,33$	$(108,33 / 145) \cdot 100 = 74,71$
2002	1,42	$(1,42 / 1,2) \cdot 100 = 118,33$	$(118,33 / 145) \cdot 100 = 81,61$
2003	1,54	$(1,54 / 1,2) \cdot 100 = 128,33$	$(128,33 / 145) \cdot 100 = 88,51$
2004	1,65	$(1,65 / 1,2) \cdot 100 = 137,50$	$(137,5 / 145) \cdot 100 = 94,83$
2005	1,74	145	100
2006	1,86	$(1,86 / 1,2) \cdot 100 = 155$	$(155 / 145) \cdot 100 = 106,90$
2007	1,94	$(1,94 / 1,2) \cdot 100 = 161,67$	$(161,67 / 145) \cdot 100 = 111,49$
2008	2,15	$(2,15 / 1,2) \cdot 100 = 179,17$	$(179,17 / 145) \cdot 100 = 123,56$
2009	2,25	$(2,25 / 1,2) \cdot 100 = 187,50$	$(187,50 / 145) \cdot 100 = 129,31$
2010	2,30	$(2,30 / 1,2) \cdot 100 = 191,67$	$(191,67 / 145) \cdot 100 = 132,18$

El interés del cambio reside en tener los datos más actuales, con la transformación se puede observar como el precio de la botella de refrescos en el año 2010 aumento el 32,18% en relación al año 2005.

Señalar que para realizar un cambio de base en los índices simples basta dividir cada uno de los índices de la base antigua por el valor del índice correspondiente al periodo seleccionado como nueva base y multiplicarlo por 100.

Juan Miguel Navarro – UCAM - 40

4. Operaciones con Números Índices

Como alternativa a la actualización del período base descrito para los **sistemas de base fija**, se viene utilizando con mayor frecuencia los **sistemas de índices de base variable o encadenada** (sistemas que utilizan como base el período inmediatamente anterior).

Observando la tabla anterior, utilizando la **BASE VARIABLE o ENCADENADA**:

Años	Precio refresco (euros)	Índices Simples 2005=100	Índices Simples Base variable o Encadenada	Tasa variación (interanual)
2005	1,74	100	-----	
2006	1,86	106,90	$(106,90 / 100) \cdot 100 = 106,90$	6,90
2007	1,94	111,49	$(111,49 / 106,90) \cdot 100 = 104,30$	4,30
2008	2,15	123,56	$(123,56 / 111,49) \cdot 100 = 110,82$	10,82
2009	2,25	129,31	$(129,31 / 123,56) \cdot 100 = 104,65$	4,65
2010	2,30	132,18	$(132,18 / 129,31) \cdot 100 = 102,22$	2,22

En la última columna, se observa que entre 2006 y 2005 el precio de la botella de refrescos varió un 6,90%, entre 2006 y 2007 un 4,30%, etc. En este ejemplo, de índices de base variable o encadenada, cada índice se calcula respecto a un año distinto.

Destacar que a partir de la serie de base variable (última columna) se puede calcular el índice para base fija de cualquier período. De esta manera, el índice de los refrescos de 2010 con base 2005 sería:

$$I_{\frac{2010}{2005}} = I_{\frac{2006}{2005}} \cdot I_{\frac{2007}{2006}} \cdot I_{\frac{2008}{2007}} \cdot I_{\frac{2009}{2008}} \cdot I_{\frac{2010}{2009}} = \frac{106,9}{100} \cdot \frac{104,30}{100} \cdot \frac{110,82}{100} \cdot \frac{104,65}{100} \cdot \frac{102,22}{100} \cdot 100 = 132,18$$

Juan Miguel Navarro – UCAM - 41

5. Tasas de Variación

Tasa de variación anual: mide el cambio relativo del valor de una magnitud entre dos momentos.

$$TV = \left[\frac{X_t}{X_{t-1}} - 1 \right] \cdot 100$$

Tasa media de variación acumulada: para conocer el cambio relativo medio mantenido a lo largo de un periodo.

$$TMA = \left[\left(\frac{X_t}{X_0} \right)^{1/t} - 1 \right] \cdot 100$$

Juan Miguel Navarro – UCAM - 42

5. Tasas de Variación

Relación entre Tasas e Índices

Sea x_{t_1} el valor de una variable X en el instante o periodo de tiempo t_1 y x_{t_2} el valor de la misma en un instante o periodo posterior t_2 , la **tasa de variación** de X en t_2 con respecto a t_1 se define como:

$$\text{Tasa}_{t_1}^{t_2}(x) = \frac{x_{t_2} - x_{t_1}}{x_{t_1}} \cdot 100$$

$$\text{Adviértase que } \text{Tasa}_{t_1}^{t_2}(x) = \frac{x_{t_2} - x_{t_1}}{x_{t_1}} \cdot 100 = \left(\frac{x_{t_2}}{x_{t_1}} - 1 \right) \cdot 100 = I_{t_1}^{t_2}(x) - 100$$

5. Tasas de Variación

Tasas a partir de índices en cadena

- ♦ La tasa de variación entre dos observaciones consecutivas (x_{t-1} , x_t) de X , se denota por **Tasa^t(x)**, y se calcula a partir del **índice en cadena**:

$$\text{Tasa}^t(x) = \frac{x_t - x_{t-1}}{x_{t-1}} \cdot 100 = IC^t(x) - 100$$

Se suele utilizar la expresión **tasa de variación interanual, intertrimestral o intermensual**, para referirse a la tasa de variación entre observaciones consecutivas correspondientes a años, trimestres o meses, respectivamente.

Cuando se trabaja con series de datos mensuales o trimestrales correspondientes a distintos años, también se utiliza la expresión **tasa de variación interanual correspondiente a un determinado mes (o trimestre)** para referirse a la variación porcentual que experimenta la variable en un determinado mes (o trimestre) del año inmediatamente anterior.

5. Tasas de Variación

Ejemplo En la tabla adjunta se refleja el gasto total en viajes turísticos (en millones de euros) de los residentes en España para el periodo 2001-2004.

Datos	2001	2002	2003	2004
Viajes turísticos	12815,2	12093	12743,7	14568,5
Destino España	9896	9274,6	9828,5	11154,3
Destino extranjero	2919,2	2814,4	2915,2	3414,2

- ¿Cuál fue el incremento porcentual de gasto en viajes turísticos de los residentes en España entre los años 2001-2004?
- Hallar la variación porcentual del gasto en viajes turísticos con destino al extranjero correspondiente a cada año respecto al año 2001.
- Determinar las tasas de variación interanual (%) para el gasto por viajes con destino a España correspondientes al periodo 2001-2004, sabiendo que en el año 2001 respecto al 2000 fue de un 16,2%.

Juan Miguel Navarro – UCAM - 45

5. Tasas de Variación

Solución: En el apartado (a)

$$\text{Tasa}_{2001}^{2004}(x) = \frac{x_{2004} - x_{2001}}{x_{2001}} \cdot 100 = \frac{14568,5 - 12815,2}{12815,2} \cdot 100 = 13,6814$$

$$\text{o bien, } \text{Tasa}_{2001}^{2004}(x) = I_{2001}^{2004}(x) - 100 = \frac{14568,5}{12815,2} - 100 = 13,6814$$

- Se obtiene la serie de índices simples en serie con base 2001 para el gasto en viajes turísticos y después se obtiene la tasa porcentual restando 100 a cada índice.

Datos	2001	2002	2003	2004
Destino extranjero	2919,2	2814,4	2915,2	3414,2
I_{2001}^t	100	96,410	99,863	116,957
$\text{Tasa}_{2001}^t = I_{2001}^t - 100$	0	-3,590	-0,137	16,957

irro – UCAM - 46

5. Tasas de Variación

También a partir de la expresión: $Tasa_{2001}^t(x) = \frac{x_t - x_{2001}}{x_{2001}} \cdot 100$

c) Las tasas de variación interanual se calculan a partir de los índices de cadena:

Datos	2001	2002	2003	2004
Destino España	9896	9274,6	9828,5	11154,3
IC ^t	116,2	93,72	105,97	113,49
Tasa interanual = IC ^t - 100	16,20	-6,28	5,97	13,49

$$I_{2001}^{2002}(X) = \frac{9274,6}{9896} \cdot 100 = 93,72 \quad I_{2002}^{2003}(X) = \frac{9828,5}{9274,6} \cdot 100 = 105,97 \quad I_{2003}^{2004}(X) = \frac{11154,3}{9828,5} \cdot 100 = 113,49$$

Juan Miguel Navarro – UCAM - 47

5. Tasas de Variación

Ejemplo En la tabla adjunta figuran el número de hipotecas inmobiliarias para fincas rústicas entre junio y diciembre de 2004.

Mes	2004 - 06	2004 - 07	2004 - 08	2004 - 09	2004 - 10	2004 - 11	2004 - 12
Hipotecas	4155	3836	3380	4212	4119	3927	3801

- Determinar las tasas intermensuales de variación correspondientes
- Conociendo que el número de hipotecas en septiembre de 2005 ascendió a 4410, obtener la tasa de variación interanual correspondiente al mes de septiembre.

Juan Miguel Navarro – UCAM - 48

5. Tasas de Variación

Solución: En el apartado (a)

Mes	2004 - 06	2004 - 07	2004 - 08	2004 - 09	2004 - 10	2004 - 11	2004 - 12
Hipotecas	4155	3836	3380	4212	4119	3927	3801
IC ^t (mes)	-----	92,32	88,11	124,62	97,79	95,34	96,79
Tasa IC ^t - 100	-----	-7,68	-11,89	24,62	-2,21	-4,66	-3,21

4410

b) La tasa interanual para el mes de septiembre de 2005:

$$\text{Tasa interanual}^{2005-09} = \text{IC}^{2005-09} - 100 = \frac{4410}{4212} \cdot 100 - 100 = 4,7\%$$

Juan Miguel Navarro – UCAM - 49

5. Tasas de Variación

Tasa media de variación

Se denomina **tasa media de variación** de la variable X en el periodo $[t, t+k]$, o **tasa media de crecimiento acumulativo**, a la tasa T_k que permite obtener la observación x_{t+k} en el instante o periodo $t+k$, partiendo de la observación x_t en el instante t , aplicando entre instantes o periodos consecutivos un incremento porcentual constante e igual a T_k .

$$x_{t+1} = x_t + \frac{T_k}{100} \cdot x_t = \left(\frac{100 + T_k}{100} \right) \cdot x_t$$

$$x_{t+2} = x_{t+1} + \frac{T_k}{100} \cdot x_{t+1} = \left(\frac{100 + T_k}{100} \right) \cdot x_{t+1} = \left(\frac{100 + T_k}{100} \right)^2 \cdot x_t$$

$$x_{t+3} = x_{t+2} + \frac{T_k}{100} \cdot x_{t+2} = \left(\frac{100 + T_k}{100} \right) \cdot x_{t+2} = \left(\frac{100 + T_k}{100} \right)^3 \cdot x_t$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

Juan Miguel Navarro – UCAM - 50

5. Tasas de Variación

Tasa media de variación

$\vdots$
 $\vdots$
 $\vdots$
 $\vdots$

$$x_{t+k} = x_{t+k-1} + \frac{T_k}{100} \cdot x_{t+k-1} = \left(\frac{100 + T_k}{100} \right) \cdot x_{t+k-1} = \left(\frac{100 + T_k}{100} \right)^k \cdot x_t$$

$$\text{Siendo } x_{t+k} = \left(\frac{100 + T_k}{100} \right)^k \cdot x_t \rightarrow \frac{x_{t+k}}{x_t} = \left(\frac{100 + T_k}{100} \right)^k \rightarrow \sqrt[k]{\frac{x_{t+k}}{x_t}} = \frac{100 + T_k}{100}$$

$$\text{Por tanto, } T_k = \left(\sqrt[k]{\frac{x_{t+k}}{x_t}} - 1 \right) \cdot 100$$

Juan Miguel Navarro – UCAM - 51

5. Tasas de Variación

Ejemplo En la tabla adjunta figuran el número de hipotecas inmobiliarias para fincas rústicas entre junio y diciembre de 2004.

Mes	2004 - 06	2004 - 07	2004 - 08	2004 - 09	2004 - 10	2004 - 11	2004 - 12
Hipotecas	4155	3836	3380	4212	4119	3927	3801

4410

- a) Determinar la tasa media de variación intermensual de 2004
- b) Conociendo que el número de hipotecas en septiembre de 2005 ascendió a 4410, obtener el crecimiento medio mensual acumulativo para el periodo septiembre 2004 - septiembre 2005

Solución: En el apartado (a)

$$T_6 = \left(\sqrt[6]{\frac{x_{2004-12}}{x_{2004-06}}} - 1 \right) \cdot 100 = \left(\sqrt[6]{\frac{3801}{4155}} - 1 \right) \cdot 100 = \left(\sqrt[6]{0,9148} - 1 \right) \cdot 100 = -1,47\%$$

- b) El crecimiento medio mensual acumulativo entre septiembre 2004-2005 (periodo de 13 meses):

$$T_{12} = \left(\sqrt[12]{\frac{x_{2005-09}}{x_{2004-09}}} - 1 \right) \cdot 100 = \left(\sqrt[12]{\frac{4410}{3380}} - 1 \right) \cdot 100 = \left(\sqrt[12]{1,3047} - 1 \right) \cdot 100 = 2,24\%$$

Juan Miguel Navarro – UCAM - 52

Ejemplo

Ejemplo Con la tabla adjunta de precios de productos agrícolas (arroz, trigo y patatas). Calcular los índices de precios de Sauerbeck y de Bradstreet-Dùtot, así como las tasas de variación intermensuales.

Meses	Precio Arroz	Precio Trigo	Precio Patatas
0	50	30	40
1	60	30	40
2	70	35	45
3	75	40	45
4	80	45	50
5	90	50	50

Solución:

Los índices complejos de Sauerbeck y Bradstreet-Dùtot se obtienen, respectivamente, como media

aritmética simple $\left[\text{índices } I_0^t(X) = \frac{x_t}{x_0} \cdot 100 \right]$ y media agregativa simple $\left[\text{índices } (I_A)_0^t(X) = \frac{(x_A)_t}{(x_A)_0} \cdot 100 \right]$

Juan Miguel Navarro – UCAM - 53

Ejemplo

Meses	Arroz Precio	Trigo Precio	Patatas Precio	Arroz l. simple	Trigo l. simple	Patatas l. simple	Índice Sauerbeck M. aritmética	TOTAL x_A	Bradstreet-Dùtot M. agre x_A
0	50	30	40	100	100	100	100	120	100
1	60	30	40	120	100	100	106,67	130	108,33
2	70	35	45	140	116,667	112,5	123,06	150	125
3	75	40	45	150	133,333	112,5	131,94	160	133,33
4	80	45	50	160	150	125	145	175	145,83
5	90	50	50	180	166,667	125	157,22	190	158,33

El índice de Sauerbeck es la media aritmética de los índices simples: $I_0^t(X) = \frac{x_t}{x_0} \cdot 100$

$$S_{p2} = \frac{120 + 100 + 100}{3} = 106,77 \quad S_{p3} = \frac{140 + 116,667 + 112,5}{3} = 123,06 \quad \dots\dots\dots$$

El índice de Bradstreet-Dùtot es la media agregativa simple: $(I_A)_0^t(X) = \frac{(x_A)_t}{(x_A)_0} \cdot 100$

$$B-D_{p0}^1 = \frac{130}{120} \cdot 100 = 108,33 \quad B-D_{p0}^3 = \frac{160}{120} \cdot 100 = 133,33 \quad \dots\dots\dots$$

Juan Miguel Navarro – UCAM - 54

Ejemplo

Las tasas de variación intermensuales: $Tasa_0^t(x) = \frac{x_t - x_0}{x_0} \cdot 100 = \left(\frac{x_t}{x_0} - 1 \right) \cdot 100 = I_0^t(x) - 100$

Meses	0	1	2	3	4	5
B-D (media agregativa)	100	108,33	125	133,33	145,83	158,33
Tasas variación (intermensuales)	-----	8,33	25	33,33	45,83	58,33

Juan Miguel Navarro – UCAM - 55

Juan Miguel Navarro Ruiz

jmnavarro@ucam.edu

UCAM Universidad Católica de Murcia



Tema 6

Fuentes de Información Estadística del Sector Turismo

Estadística Aplicada al Turismo

Juan Miguel Navarro Ruiz

Grado en Turismo



ÍNDICE

CONTENIDOS

1. Introducción
2. Fuentes de información turística

1. Introducción

En esta tema se ofrecen enlaces a algunos proveedores de información estadística sobre turismo. No se pretende ofrecer un catálogo completo y exhaustivo de los mismos sino una pequeña selección que sirva de puerta de acceso y motive a los estudiantes a buscar sus propias fuentes de información.

1. Introducción

Para facilitar la localización de datos, esta sección se ha estructurado de la siguiente forma:

- Datos de ámbito nacional
- Datos de ámbito internacional
- Datos de ámbito regional

2. Fuentes de información turística

Ámbito Nacional



El [Instituto Nacional de Estadística](#) elabora algunas de las [estadísticas sobre turismo](#) con más tradición del país. Entre ellas se encuentran las Encuestas de ocupación de Alojamientos Turísticos entre las que destaca la Encuesta de ocupación hotelera. Además de esto el INE también elabora unos índices de precios para estas mismas modalidades de alojamientos y la [Estadística de productos en el sector servicios](#) que complementa a la [Encuesta anual de servicios](#). También se encarga de la realización de la [Cuenta satélite del turismo de España](#) con la que se pretende ofrecer una representación sistemática, comparable y completa de la actividad turística española.

2. Fuentes de información turística

Ámbito Nacional



Junto a esto la página del INE proporciona a través de [INEBASE](#) información sobre población, sociedad, cultura, medioambiente, economía, territorio, etc. toda ella de gran utilidad para el estudio del turismo.



2. Fuentes de información turística

Ámbito Nacional

La Subdirección General de Conocimiento y Estudios Turísticos depende del Instituto de Turismo de España y tiene entre sus funciones, el análisis de los mercados emisores de turismo, la elaboración de proyecciones y prospectivas sobre flujos turísticos, pernoctaciones e ingresos por turismo, el análisis de los factores que inciden en la economía turística, la coordinación de la información sobre el sector turístico, y en general la difusión del conocimiento turístico.



2. Fuentes de información turística

Ámbito Nacional

Como productora de información, se encarga de la Encuesta de Ocupación Turística (Ocupatur), de la Encuesta de Hábitos Turísticos (Habitur) y del Índice de tendencia de competitividad turística (ITCT). De igual modo, explota información turística procedente de fuentes externas, como son la evolución de llegadas de pasajeros extranjeros en compañías aéreas de bajo coste y el empleo en el sector turístico, análisis de información de gasto en pasarelas de pago y de cualquier otra que sea de interés para el mejor conocimiento del comportamiento del turismo en España.



2. Fuentes de información turística

Ámbito Nacional

Hasta fechas recientes se encargaba de las tres grandes operaciones estadísticas sobre turismo en España: [Frontur](#) (datos relativos a la entrada en España de visitantes no residentes), [Egatur](#) (datos relativos al gasto que realizan en España los visitantes no residentes en ella) y [Familitur](#) (datos relativos a los viajes que realizan los residentes en España). Todas ellas han pasado a realizarse con ciertas modificaciones por el [INE](#). En cualquier caso todavía puede accederse a la información anterior a este momento a través de informes y tablas de datos y documentos metodológicos.



2. Fuentes de información turística

Ámbito Nacional

Actualmente ofrece un [Boletín trimestral de coyuntura turística](#) (COYUNTUR), que contiene un análisis de los principales indicadores de la actividad turística. Además de lo anterior ofrece [otra información](#) a partir de la explotación de datos proporcionados por otros organismos. Esta información no siempre está actualizada y entre ella merece la pena destacar la siguiente:

[Balanza de Pagos](#): Elaborada mes a mes por el Banco de España ofrece información sobre ingresos y pagos de la rúbrica de turismo y viajes.

[Compañías de Bajo Coste](#): Información elaborada en base mensual, a partir de los datos obtenidos de 3 fuentes de información (Datos de vuelos aportados por AENA, encuestas de la Subdirección General de Conocimiento y Estudios Turísticos asociadas a la operación de EGATUR y los datos de la base de datos de vuelos de OAG).



2. Fuentes de información turística

Ámbito Nacional

Empleo Turístico: Incluye información del mercado de trabajo en la industria turística, elaborada a partir de datos obtenidos del Ministerio de Empleo y Seguridad Social (datos mensuales de afiliación a la Seguridad Social) y del INE (datos trimestrales de la Encuesta de Población Activa, EPA).

Junto a esta información se ofrecen una serie de informes trimestrales de prospectiva, que utilizando el software TRAMO-SEATS, realizan una proyección a futuro del comportamiento de los principales mercados emisores de turismo hacia España. La información está estructurada en forma de fichas.

2. Fuentes de información turística

Ámbito Nacional



El [Banco de España](#) elabora los datos de la balanza de pagos en los que siempre ha tenido especial importancia la rubrica de turismo como un elemento fundamental de la Balanza por Cuenta Corriente. En su página no se encuentran largas series históricas, pero si datos actualizados sobre la evolución mensual de las grandes partidas de la Balanza de Pagos. Los datos se pueden consultar directamente en la sección de [Estadísticas del Sector Exterior](#). (Los datos de 2014 y años anteriores pueden consultarse en [Balanza de Pagos y Posición de Inversión Internacional de España](#)).



2. Fuentes de información turística

Ámbito Nacional

Exceltur es una asociación sin ánimo de lucro formada por algunos de los más relevantes grupos empresariales turísticos españoles que se organiza como un foro de análisis, opinión y diálogo abierto, desde el que valorar el presente y el futuro del turismo español. Su departamento de estudios elabora informes monográficos sobre diversos aspectos del turismo y una serie de estudios periódicos que sirven para estudiar la situación y evolución del sector.

2. Fuentes de información turística

Ámbito Nacional

Los más importantes son los siguientes: [Perspectivas turísticas](#), (Informe trimestral sobre la situación del sector), [Impactur](#) (Impacto económico del turismo en las comunidades autónomas), [Monitur](#) (Un instrumento para valorar la capacidad de cada comunidad autónoma de consolidar un posicionamiento turístico diferencial y sostenible en el tiempo), [Urbantur](#) (Monitor de competitividad turística de los destinos urbanos españoles).

También elabora un [Barómetro de la rentabilidad y empleo de los destinos turísticos españoles](#) y una serie de indicadores de actividad que reflejan la situación del sector turístico español y sus destinos



2. Fuentes de información turística

Ámbito Internacional

La [Organización Mundial del Turismo](#) es una de las instituciones que dedica más esfuerzos a la elaboración y sistematización de datos estadísticos sobre el turismo. Parte de la información que ofrece es de pago pero hay otra que se puede obtener gratuitamente, en algunos casos con cierto retraso. Es muy recomendable navegar por su página web especialmente por la sección de [estadísticas](#), pero si se están buscando datos se puede acudir a la sección [Facts & Figures](#).



2. Fuentes de información turística

Ámbito Internacional

La [Organización de las Naciones Unidas](http://data.un.org/Default.aspx) (ONU) dispone de un banco de datos (UNdata) al que ofrece acceso gratuito a través de la web. (<http://data.un.org/Default.aspx>) Este repositorio incluye todo tipo de datos organizados por temáticas. Además, en la página de inicio del banco de datos se ofrece acceso a los organismos estadísticos de cada país por lo que es fácil ampliar la información directamente de fuentes originales.

El turismo está presente en este banco de datos con información sobre [turismo emisor y turismo receptor en cada país](#). Las series abarcan desde 1995 y se utiliza información de la OMT, del FMI y en ocasiones de organismos nacionales. Esta disparidad y los problemas que tienen los propios datos de la OMT hacen que sea muy recomendable revisar la información complementaria que se ofrece en cada caso.



2. Fuentes de información turística

Ámbito Internacional

La [OCDE](#) (Organisation for Economic Co-operation and Development) es una organización internacional actualmente integrada por 35 países, fundamentalmente europeos. En su portal especializado [OCDE DATA](#) ofrece multitud de [datos estadísticos](#) entre ellos unos cuantos sobre turismo dentro de la sección dedicada a la industria y los servicios. También existe un [Comité de Turismo de la OCDE](#) y periódicamente se hacen circular publicaciones con datos del sector.

2. Fuentes de información turística

Ámbito Internacional



Eurostat es la agencia estadística de la Unión Europea. Trabaja en la homogeneización de los datos para poder estudiarlos en un contexto europeo. Se nutre de los datos proporcionados por los diferentes países pero también elabora sus propias estadísticas. Entre sus áreas de interés se encuentra el turismo del que ofrece datos y documentos metodológicos.

Además de lo anterior la Comisión Europea ha puesto en marcha el Virtual Tourism Observatory con el fin de proporcionar apoyo a los responsables de la política turística y los empresarios del sector en la elaboración de estrategias destinadas a mejorar la competitividad del sector.

2. Fuentes de información turística

Ámbito Regional

Además de los datos de carácter regional proporcionados por los organismos e instituciones estadísticas de ámbito nacional, muchas Comunidades Autónomas han puesto en marcha centros de estadística que ofrecen los datos proporcionados por otros organismos al tiempo que explotan las fuentes de datos que les son propias. Algunos de estos organismos y operaciones estadísticas se ofrecen a continuación.

- **Andalucía:** Instituto de Estadística y Cartografía de Andalucía
- **Andalucía:** Estadísticas de Turismo de Andalucía
- **Aragón:** Instituto Aragonés de Estadística
- **Canarias:** ISTAC Instituto Canario de Estadística
- **Cantabria:** ICANE Instituto Cántabro de Estadística
- **Castilla – La Mancha:** Servicio de estadística de Castilla La Mancha

2. Fuentes de información turística

Àmbito Regional

- **Castilla y León:** Coyuntura Turística de Castilla y León
- **Castilla y León:** Servicio de Información Estadística de Castilla y León
- **Cataluña:** IDESCAT Institut d' Estadística de Catalunya
- **Cataluña:** Observatori del Treball i Model Productiu (Àmbit Turisme)
- **Extremadura:** IEEX Instituto de Estadística de Extremadura
- **Galicia:** Instituto Galego de Estadística
- **Islas Baleares:** IBESTAT Institut d'Estadística de les Illes Balears
- **La Rioja:** Instituto de Estadística de La Rioja
- **Madrid:** Instituto de Estadística de la Comunidad de Madrid
- **Navarra:** Nastat – Instituto de Estadística de Navarra
- **País Vasco:** Eustat – Instituto Vasco de Estadística
- **Valencia:** Institut Valencià d' Estadística
- **Asturias:** SADEI Sociedad Asturiana de Estudios Económicos e Industriales
- **Asturias:** Sistema de información turística de Asturias (SITA)
- **Murcia:** CREM – Centro Regional de Estadística de Murcia

Juan Miguel Navarro Ruiz

jmnavarro@ucam.edu

UCAM Universidad Católica de Murcia

