ANÃ LISIS CUANTITATIVO

DE LA

ACTIVIDAD TURÃ STICA

Ã NDICE:

TEMA 1. ANÃ LISIS DE LA ACTIVIDAD TURÃ STICA

02/10/2008

1.1.LA NECESIDAD Y LA UTILIDAD DEL ANÃ LISIS CUANTITATIVO

- La **actividad turÃ-stica o turismo** es toda actividad que hacen las personas cuando se desplazan fuera de su entorno habitual, para una duraciÃ³n inferior a 12 meses y con la finalidad de no ejercer actividades remuneradas en el lugar de destino.

- **Problemas en la cuantificaciÃ³n**: muchas de las actividades turÃ-sticas sirven a consumidores que pueden ser o no turistas (restaurantes, tiendas diversas, etc.). No todas las personas que viajan son turistas.

- Para intentar solucionar estos problemas, **la OrganizaciÃ³n Mundial del Turismo** (OMT), a partir de los aÃ±os 80, comenzÃ³ a unificar, a nivel mundial, las **definiciones bÃ¡sicas** para la estadÃ-stica de la actividad turÃ-stica. Estas definiciones deben ser homogÃ©neas y utilizadas de igual forma, en todo el mundo. La OMT tambiÃ©n da unas clasificaciones sobre los turistas, que deben seguir todos los paÃ-ses y asÃ-, **poder hacer comparaciones entre los diferentes lugares y tener unos anÃ¡lisis homogÃ©neos.** La OMT recomienda el uso de estas definiciones y clasificaciones a todos los paÃ-ses.

1.2.UNIDADES DE ANÃ LISIS DE LA ACTIVIDAD TURÃ STICA

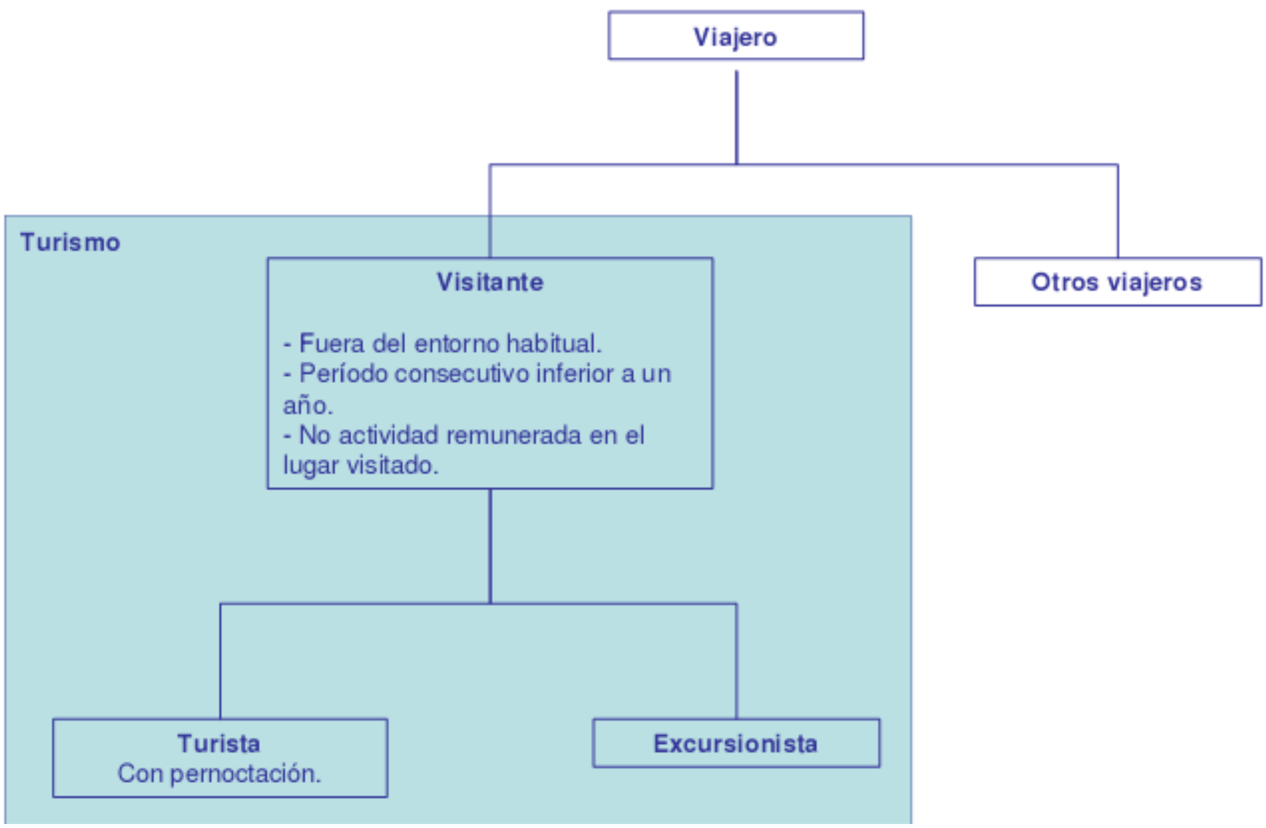
Residente: persona que estÃ¡ durante la mayor parte del aÃ±o en un paÃ-s o lugar.

Viajero: toda persona que se desplaza entre dos o mÃ¡s paÃ-ses diferentes, o entre dos o mÃ¡s lugares dentro de su paÃ-s de residencia, pero fuera de su entorno habitual.

Visitante: Toda persona que se desplaza a un lugar diferente a su entorno habitual, por una duraciÃ³n inferior a 12 meses y cuya finalidad principal en el viaje NO es ejercer una actividad que sea remunerada en el lugar visitado.

Turista: es un visitante (que se desplaza a un lugar diferente a su entorno habitual, por una duraciÃ³n inferior a 12 meses y cuya finalidad principal en el viaje NO es ejercer una actividad que sea remunerada en el lugar visitado) que pernocta como mÃ-nimo una noche en el lugar visitado.

Excursionista: es un visitante (que se desplaza a un lugar diferente a su entorno habitual, por una duraciÃ³n inferior a 12 meses y cuya finalidad principal en el viaje NO es ejercer una actividad que sea remunerada en el lugar visitado) que realiza una visita de dÃ-a y NO pernocta en el lugar visitado.



En **otros viajeros** englobamos:

- Personas que viajan dentro de su entorno habitual (viajeros laborales fronterizos, trabajadores fronterizos, viajeros en vecindad directa de lugar de residencia),
- Personas que cambian de lugar de residencia (Migrantes a largo plazo, y personas que se trasladan a otro lugar dentro de su paÃ-s de residencia, ambos grupos con un propÃ³sito de estancia de mÃ¡s de 12 meses).

- Personas <u>sin lugar fijo de residencia</u> (nÃ³madas, vagabundos, refugiados).
- Personas que <u>viajan a lugares donde perciben remuneraciÃ³n</u> (migrantes a corto plazo con propÃ³sito de estancia igual o inferior a 12 meses, trabajadores estacionales, conferenciantes, artistas de espectÃ¡culos, Au pair,)
- <u>Otros excluidos por convenciÃ³n</u> (pasajeros de trÃ¡nsito, miembros de las fuerzas armadas, representaciÃ³n de consulados, diplomÃ¡ticos, prisioneros)

1.3.LA HETEROGENEIDAD DE LA ACTIVIDAD TURÃ STICA Y LA HOMOGENIZACIÃ“N DE CONCEPTOS.

Viaje: es el que realiza un individuo o un grupo, cada vez que deja su lugar de residencia habitual por un motivo o mÃ¡s, recorre una determinada distancia para visitar uno o mÃ¡s destinos, con uno o mÃ¡s medios de transporte, y vuelve a su residencia. (Un viaje de negocios es un viaje, pero no es turismo).

Viaje de vacaciones: es un viaje en el que el motivo principal es el ocio, e incluye <u>cuatro pernoctaciones como mÃ-nimo</u>.

Turismo interno: es el que realizan los residentes de un paÃ-s que viajan dentro del propio paÃ-s.

Turismo receptor: es el que realizan los no residentes que viajan dentro de un determinado paÃ-s.

Turismo emisor: es el que realizan los residentes de un determinado paÃ-s, que viajan a otro paÃ-s.

[OJO, si se hace un estudio de una zona en concreto, por ejemplo a nivel de las Islas Baleares, el concepto paÃ-s debe entenderse como esa zona en concreto, en este caso, sÃ³lo las Islas Baleares]

TIPOLOGÃ A DE TURISMO

TURISMO INTERIOR = TURISMO INTERNO + TURISMO RECEPTOR

TURISMO INTERIOR = el que realizan los residentes de un paÃ-s que viajan dentro del propio paÃ-s + el que realizan los no residentes que viajan dentro de un determinado paÃ-s.

TURISMO NACIONAL = TURISMO INTERNO + TURISMO EMISOR

TURISMO NACIONAL = el que realizan los residentes de un paÃ-s que viajan dentro del propio paÃ-s + el que realizan los residentes de un determinado paÃ-s, que viajan a otro paÃ-s.

TURISMO INTERNACIONAL = TURISMO RECEPTOR + TURISMO EMISOR

TURISMO INTERNACIONAL = el que realizan los no residentes que viajan dentro de un determinado paÃ-s + el que realizan los residentes de un determinado paÃ-s, que viajan a otro paÃ-s.

La definiciÃ³n de Turismo no acaba de determinar de una manera inequÃ-voca y Ãºnica, los principales componentes y unidades que la integran, por lo que es necesario unificar conceptos:

- Unificar ciertas definiciones (ver material de apoyo 1)

- Contabilizar las actividades y los bienes y servicios turÃ-sticos

Los **bienes y servicios** turÃ-sticos son los que se destinan a satisfacer las necesidades de los <u>visitantes</u> de un determinado paÃ-s o regiÃ³n. Se dividen en:

- **Bienes y servicios primarios o puramente turÃsticos**: son los que dependen totalmente de la actividad turÃstica (hoteles, apartamentos...)

- **Bienes y servicios secundarios o complementarios**: son los que dependen parcialmente de la actividad turÃstica. Este tipo de bienes y servicios son los mÃ¡s difÃciles de clasificar (por ej. un restaurante tendrÃ¡ clientes que sean visitantes y otros que no lo sean, pero dependiendo de donde estÃ© situado es posible que sÃ³lo tenga clientes visitantes o sÃ³lo clientes del lugar).

- **Actividades y servicios intermedios o indirectos**: dependen indirectamente de la actividad turÃstica (por ej. la construcciÃ³n).

1.4.PROBLEMÃ TICA DE LA ACTIVIDAD TURÃ STICA

- Hay dificultad para conseguir datos creÃbles para describir la magnitud y las consecuencias del turismo en la economÃa.

- Hay una gran diversidad en el sector, lo que lleva a plantearse si el turismo es realmente un sector o un conjunto de sectores.

- Hay grandes variaciones del fenÃ³meno turÃstico segÃºn el lugar y la diversidad de tamaÃ±o de esos lugares.

- Hay fragmentaciÃ³n y falta de organizaciÃ³n en la actividad turÃstica; es un sector muy complejo.

- Es difÃcil hacer predicciones a medio y largo plazo debido al gran dinamismo del sector y a la exposiciÃ³n del sector a cambios imprevistos o imprevisibles (guerras, cambio de gustos del cliente, etc). Esta dificultad de predicciÃ³n es caracterÃstica del sector turÃstico, a pesar de que estas predicciones son necesarias y muy importantes para su funcionamiento.

Estos problemas se intentan solucionar o clarificar con la homogeneizaciÃ³n de definiciones y clasificaciones.

OBJETIVO DE LA ASIGNATURA: TENER HERRAMIENTAS VÃ LIDAS PARA TRABAJAR CON SEGURIDAD, ES DECIR, QUE A PARTIR DE UNOS DATOS, SE PUEDA TRABAJAR CON ELLOS Y SACAR CONCLUSIONES VÃ LIDAS.

TEMA 2. LAS VARIABLES QUE MIDEN LA ACTIVIDAD TURÃ STICA

9/10/2008

2.1.VARIABLES E INDICADORES DE LA ACTIVIDAD TURÃ STICA.

VARIABLE: Es el conjunto de observaciones sobre una determinada caracterÃstica (IMP). Por ejemplo, si se hace un estudio sobre la edad de los alumnos que hay en una clase, la caracterÃstica es la edad del alumno y cada una de las edades serÃ¡ una observaciÃ³n, pero el conjunto de edades serÃ¡ la variable. Las variables deben estar bien definidas: en el ejemplo anterior lo correcto serÃa Variable: la edad de los alumnos de la clase de AnÃ¡lisis).

CONSTANTE: es una observaciÃ³n de un hecho concreto. Por ej: la edad de un alumno concreto.

INDICADOR: son variables que sin medir directamente una caracterÃstica, nos aproximan a ella de una manera indirecta, recogiendo informaciÃ³n que se relaciona con la que queremos obtener realmente. Por ejemplo: nÃºmero de visitas a un destino como indicador de la fidelidad del turista a ese destino:

Alemanes		Británicos	
Una	39,74%	Una	23,10%
Dos	21,33%	Dos	18,29%
Tres	12,74%	Tres	16,41%
Cuatro	7,28%	Cuatro	10,44%
Más de cuatro	18,91%	Más de cuatro	31,76%

TIPOLOGÃ A DE LAS VARIABLES: Es importante saber quÃ© tipo de variable estamos analizando en cada momento, ya que segÃºn de quÃ© tipo se trate, se pueden aplicar un tipo de normas y operaciones u otro.

Variables temporales: son las que se observan en varios perÃ-odos de tiempo consecutivos. Por ej: gasto total *per cÃ¡pita* y dÃ-a entre 1989 y 2004.

Variables transversales o atemporales: son las que se observan en un mismo momento del tiempo para varios individuos. Por ej: gasto total *per cÃ¡pita* y dÃ-a por nacionalidades en 2004.

Gasto turístico en las Islas Baleares

Gasto total per cápita y día. 1989-2004	
Año	
89	61,76
90	57,02
91	56,54
92	55,10
93	55,03
94	57,93
95	57,50
96	57,31
97	63,98
98	67,04
99	66,34
00	67,49
01	72,77
02	65,41
03	61,82
04	66,17

Gasto total per cápita y día por nacionalidades. 2004	
NACIONALIDAD	
ALEMANA	58,71
BRITÁNICA	73,34
ESPAÑOLA	70,45
FRANCESA	66,39
BELGA	63,68
SUECA	56,89
NORUEGA	68,01
DANESA	67,44
HOLANDESA	50,42
ESTADOUNIDENSE	54,21
SUIZA	83,97
AUSTRIACA	64,78
ITALIANA	76,84
IRLANDESA	66,69
OTRAS	69,14

Fuente: Encuesta de Gasto Turístico de las Islas Baleares Conselleria d'Economia i Hisenda del Govern de les Illes Balears.

Variables cualitativas (IMP): son las que expresan caracterÃ-sticas o categorÃ-as de una cualidad y NO pueden expresarse numÃ©ricamente:

Variables cualitativas nominales son las que Ãºnicamente ponen nombre a una caracterÃ-stica: sexo, nacionalidad, profesiÃ³n, tipo alojamiento, modo de realizaciÃ³n de la reserva, lugar de vacaciones del aÃ±o anterior, motivo del viaje, intenciÃ³n de retorno, tipo de moneda... Estas variables NO se pueden ordenar.

Variables cualitativas ordinales son las que llevan asociadas un orden en las respuestas: tipo de paquete turÃ-stico seleccionado, categorÃ-a del establecimiento de alojamiento, opiniÃ³n sobre los precios, impresiÃ³n acerca del viaje, comparaciÃ³n de la calidad con anteriores visitas. Estas variables se pueden ordenar de forma creciente o decreciente: por ej estrellas de un hotel: 1, 2, 3, 4, 5) Si hablamos de cuÃ¡ntos hoteles de 2* hay en un lugar, serÃ¡ la frecuencia de una variable cuantitativa, pero si nos referimos sÃ³lo al tipo de hotel que es, serÃ¡ cualitativa.

Variables cuantitativas: son las que se expresan numÃ©ricamente y se puede operar matemÃ¡ticamente con esos datos:

Variable cuantitativa discreta: son las que toman un nÃºmero finito de valores. Son valores enteros y con valor mÃ¡ximo, como por ej. la edad, nÃºmero de estrellas de los hoteles, dÃ-as de estancia, nÃºmero de viajes realizados, personas incluidas en el paquete turÃ-stico, etc.

Variable cuantitativa continua: son las que toman un nÃºmero infinito de valores, por ej: el gasto turÃ-stico. Pueden ser cualquier tipo de valor: decimales, negativos, etc.

Hay variables que pueden ser de un tipo o de otro (cualitativa o cuantitativa) y nos puede interesar analizarlas como variables cuantitativas (pero no al revÃ©s). Salvo excepciones, la variable edad es una variable cuantitativa, pero podemos analizarla como cualitativa si la expresamos en forma de intervalos. O sea, **una variable cuantitativa puede ser tambiÃ©n cualitativa.**

2.2.LA CUANTIFICACIÃ“N DE LA DEMANDA

Las variables que **miden la demanda son aquÃ©llas que miden el consumo de bienes o servicios turÃ-sticos por parte de los visitantes**, ya sean turistas o excursionistas.

Indicadores de la demanda turÃ-stica:

- **En unidades fÃ-sicas**: el flujo turÃ-stico o volumen turÃ-stico (nÂº de viajes, nÂº de pernoctaciones, nÂº de turistas, nÂº de...)

- **En unidades monetarias**: gasto turÃ-stico, es decir, todo gasto de consumo efectuado por un visitante o a cuenta de un visitante, durante y para su desplazamiento y permanencia turÃ-stica en el lugar de destino.

    - ♦ Gasto realizado en la ciudad o paÃ-s de origen
    - ♦ Gasto realizado en el lugar de destino
    - ♦ Gasto total = gasto en origen + gasto en destino

2.3.LA CUANTIFICACIÃ“N DE LA OFERTA

La cuantificaciÃ³n de la oferta es mÃ¡s complicada que la de la demanda, debido a la **gran diversidad de actividades que pueden ser clasificadas como turÃ-sticas**. La actividad mÃ¡s destacada es la **oferta hotelera**, que es una actividad primaria o de primer nivel. La oferta hotelera se mide principalmente por la capacidad de alojamiento turÃ-stico:

- NÂº de establecimientos por categorÃ-as
- NÂº de plazas
- NÂº de habitaciones
- NÂº de meses de apertura a lo largo del aÃ±o

El grado de utilizaciÃ³n de la capacidad productiva hotelera se mide por los siguientes indicadores:

- NÂº de viajeros

- NÂº de pernoctaciones

Estos indicadores permiten calcular la TASA DE OCUPACIÃ“N que es la proporciÃ³n de habitaciones o plazas/cama de un establecimiento de alojamiento colectivo de turismo, ocupadas durante un cierto perÃodo de tiempo:

Total de pernoctaciones

Tasa de ocupaciÃ³n = −−−

Total de plazas X NÂº de noches

Este cociente darÃ¡ un nÃºmero con varios decimales, pero aporta mÃ¡s informaciÃ³n si lo transformamos en un porcentaje, multiplicÃ¡ndolo por 100:

Total de pernoctaciones

Tasa de ocupaciÃ³n en % = −−− X 100

Total de plazas X NÂº de noches

El Total de pernoctaciones es el nÂº de noches que un turista se aloja en el establecimiento y el NÂº de noches es el nÂº de dÃas o noches del perÃodo que se considere (un mes, varios meses, un aÃ±o). Este nÃºmero debe ser exacto, es decir si se trata del mes de febrero se contabilizarÃ¡n los 28 dÃas, si es agosto 31, etc...

Ejemplo: Calcular la tasa de ocupaciÃ³n de un hotel de 250 plazas en el mes de marzo, que ha tenido 770 pernoctaciones:

Total de pernoctaciones 770

Tasa de ocupaciÃ³n = −−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−− X 100 ==> −−−−−−−−−−−−−−− X 100

Total de plazas X nÂº de noches 250 X 31

T.O. = 0,099 X 100 = 9,9 % â‰ˆ 10%

Ejercicio: Calcular las pernoctaciones mensuales (mes estÃ¡ndar de 30 dÃas) de un alojamiento turÃstico que tiene 25 habitaciones individuales, 15 dobles y 30 triples. Ocupa un 73% de su capacidad.

Total de pernoctaciones

Tasa de ocupaciÃ³n = −−− X 100

Total de plazas X nÂº de noches

Total de pernoctaciones

73% = −−− X 100

[25 + (15 X 2) + (30 X 3)] X 30 noches

Total de pernoctaciones 73 X 145 X 30

73% = −−−−−−−−−−−−−−−−−−−−−−−−−−−− X 100 total pernoctaciones = −−−−−−−−−−−−−−−−− = 3175,5

145 X 30 100

total pernoctaciones = 3176 (se redondea)

Si se usa la fÃ³rmula sin porcentajes serÃa:

Total de pernoctaciones Total de pernoctaciones

0,73 = −− ==> 0,73 = −−−−−−−−−−−−−−−−−−−−−−−−−−−−−

[25 + (15 X 2) + (30 X 3)] X 30 noches 145 x 30

0,73 x 145 x 30 = 3175.5 = total de pernoctaciones

Ejercicio: Un hotel presenta una T.O. del 16,76% con un total de 2345 pernoctaciones mensuales y abre 26 dÃas de un mes de 31. Calcular el nÂº de plazas del hotel.

2345 234500

16,76% = −−−−−−−−−−−−−−−−−−−−−−−−−− X 100 16,76 = −−−−−−−−−−−−−−−−−−−−−−−−−−

Total de plazas X 26 Total de plazas X 26

234500

16,76 X 26 X Total de plazas = 234500 Total de plazas = −−−−−−−−−−−−−−−−−−− = 538,1 plazas

16,76 X 26

538,1 plazas â‰ˆ 269 habitaciones dobles

2.4. LAS CLASIFICACIONES UNIFORMES DE LAS ACTIVIDADES

Una clasificaciÃ³n uniforme de actividades **consiste en la clasificaciÃ³n de todas las actividades econÃ³micas de manera exhaustiva, coherente y completa, de forma que no exista ambigÃ¼edad posible**, en el sentido de que cada una ocupe un lugar y sÃ³lo uno en la clasificaciÃ³n. Esto permite:

- Decidir si la producciÃ³n de una empresa con una actividad econÃ³mica determinada pertenece a un sector u otro.
- Afrontar que existen productos y servicios que sin cambiar de naturaleza, parece que pueden estar clasificados en dos sectores diferentes.

A nivel mundial, existe la clasificaciÃ³n de actividades de la **C.I.I.U.** (ClasificaciÃ³n Industrial Internacional Uniforme de actividades, elaborada por la DivisiÃ³n de EstadÃstica de la SecretarÃa de Naciones Unidas). Se puede consultar en la web de la **OMT**, www.world−tourism.org.

A-	Agricultura.
B-	Pesca.
C-	Explotación de minas y canteras.
D-	Manufactura.
E-	Suministro de electricidad, gas y agua.
F-	Construcción.
G-	Ventas al por mayor y al por menor.
H-	Hoteles y restaurantes.
I-	Transporte, almacenamiento y comunicaciones.
J-	Intermediación financiera.
K-	Actividades inmobiliarias y de alquiler.
L-	Administración Pública.
M-	Enseñanza.
N-	Actividades sociales y de salud.
O-	Viviendas.
P-	Otras actividades de servicios colectivos, sociales y personales.
Q-	Organizaciones y entidades extraterritoriales.

Los sectores de la C.I.I.U. mÃ¡s relacionados con el turismo, estÃ¡n desagregados en la **C.I.U.A.T.,** que es otra clasificaciÃ³n mÃ¡s especÃ-fica: Recoge la especificaciÃ³n de quÃ© subsectores se dedican totalmente al turismo (T) y cuÃ¡les se dedican sÃ³lo parcialmente (P):

Cuadro 1.3 Sector H: Hoteles y restaurantes

P/T	Nombre	% Venta al Turismo	Compras al Turismo
T	**Hoteles, campamentos y otros establecimientos de alojamiento comerciales:**		
T	Hoteles y moteles con restaurante	A	M
T	Hoteles y moteles sin restaurante	A	M
T	Hostales y refugios	A	B
T	Estacionamiento para vehículos de caravaning y terrenos de camping	A	B
T	Alojamiento con fines sanitarios	A	B
T	Otras instalaciones de alojamiento, N.C.O.P.	M	B
P	**Restaurantes, bares y cantinas:**		
P	Cafés y otros lugares de consumo	M	M
P	Restaurantes de servicio completo	M	M
P	Restaurantes de comida rápida (fast food) y cafeterías	M	M
P	Servicios y proveedores de comida para colectividades	M	M
P	Quioscos, puestos de refrescos y vendedores de comida	M	M
P	Salas de fiestas (nights clubs) y cenas espectáculo	M	M

Otro sistema muy importante a nivel mundial y compatible con las anteriores clasificaciones, es la ClasificaciÃ³n Industrial General de Actividades EconÃ³micas de la Comunidad Europea (**N.A.C.E.).**

2.5.LAS ESTADÃ STICAS TURÃ STICAS Y SUS FUENTES

Antes de hacer un anÃ¡lisis turÃ-stico, hay que obtener la informaciÃ³n que serÃ¡ objeto del estudio. Es necesario conocer las fuentes para la elaboraciÃ³n de la informaciÃ³n (encuestas y cuestionarios) y las estadÃ-sticas sobre turismo que publican diversos organismos:

A nivel mundial:

OMT

Eurostat

A nivel estatal:

IET

INE

AENA

A nivel autonÃ³mico:

Conselleria de Turisme

Conselleria d'Economia, Hisenda i InnovaciÃ³

Conceptos:

- **PoblaciÃ³n estadÃ-stica**: son todos aquellos elementos que se quieren analizar.
- **Muestra:** es un subconjunto de elementos de la poblaciÃ³n. Tienen que ser representativos de la poblaciÃ³n ya que resulta muy costoso y difÃ-cil conocer todos los elementos de esa poblaciÃ³n.
- **Inferir**: generalizar los resultados de la muestra a toda la poblaciÃ³n.
- **Unidad mostral:** elemento o unidad que se escoge como elemento de estudio. Esta serÃ¡ la unidad que serÃ¡ entrevistada.
- **Encuesta**: entrevista personal, telefÃ³nica o por correo postal, etc...
- **Cuestionario**: conjunto de preguntas abiertas (la respuesta es libre por parte del entrevistado, por ej., Nacionalidad) o cerradas (hay que escoger la respuesta entre unas opciones determinadas).

Las estadÃ-sticas de turismo en EspaÃ±a

El **I.E.T.** (Instituto de Estudios TurÃ-sticos) elabora tres encuestas fundamentales para el conocimiento del sector turÃ-stico en EspaÃ±a: la encuesta sobre Movimientos en Frontera (FRONTUR), la encuesta sobre Movimientos TurÃ-sticos de los EspaÃ±oles (FAMILITUR) y la encuesta sobre Gasto TurÃ-stico (EGATUR).

− **FRONTUR** cuantifica y caracteriza el nÃºmero de visitantes que llegan a EspaÃ±a por las distintas vÃ-as de acceso a las fronteras.

− **FAMILITUR** cuantifica y caracteriza los flujos de viajeros espaÃ±oles entre las distintas Comunidades AutÃ³nomas y hacia el extranjero.

− **EGATUR** cuantifica el gasto realizado en el turismo receptor y el emisor.

El **I.N.E.** (Instituto Nacional de EstadÃ-stica) es un organismo autÃ³nomo adscrito al Ministerio de EconomÃ-a y Hacienda. Las principales estadÃ-sticas turÃ-sticas que realiza son:

- EstadÃ-stica de movimientos de viajeros en establecimientos hoteleros.

- I.P.C., se extraen los I.P.H (Ã-ndice de precios por hoteles y otros establecimientos), I.P.R. (Ã-ndice de precios de restaurantes y cafeterÃ-as) y el Ã ndice de servicios turÃ-sticos. Estos 3 Ã-ndices se combinan en el I.P.T.H. (Ã-ndice de precios de turismo y hotelerÃ-a). No hay que confundir estos datos con el Gasto TurÃ-stico.

La ***Conselleria de Turisme*** y la *Conselleria d'Economia, Hisenda i InnovaciÃ³* del *Govern Balear*, realizan las siguientes encuestas:

- Encuesta sobre el cÃ¡lculo de turistas

- Encuesta del gasto turÃ-stico de las Islas Baleares: determina el gasto turÃ-stico total que se hace en la CAIB durante todo un aÃ±o. El gasto se puede desglosar segÃºn se haga en el lugar de origen o en el de destino. La poblaciÃ³n serÃ¡ aproximadamente de 10 millones de turistas y la muestra serÃ¡ de 5500 turistas. La muestra se estratifica en grupos segÃºn su nacionalidad y segÃºn la isla que visiten. La entrevista se realiza en el momento de la salida en el aeropuerto o en el puerto.

23/10/2008

Una variable puede ir acompaÃ±ada de otros datos que serÃ¡n frecuencias, pesos, ponderaciones.

Variables <u>cualitativas nominales</u>: sexo, tipo de moneda

Variables <u>cualitativas ordinales</u>: estrellas de un hotel

Por ej: se encuesta a 100 personas sobre el tipo de hotel en el que se han alojado:

Estrellas	NÂº de personas	Frecuencia relativa
1*	15	15 %
2*	45	45%
3*	25	25%
4*	15	15%
N=	100	Total =100%

El nÃºmero de personas que ha contestado sobre cada categorÃ-a de hotel serÃ-a la frecuencia.

No hay que confundir las variables con las frecuencias. **Lo primero que hay que hacer <u>siempre</u> es identificar el tipo exacto de variable** con el que vamos a trabajar.

Sexo	Frecuencia
Hombre	7
Mujer	13

Las variables serÃ-an hombre o mujer y las frecuencias 7 y 13.

Una constante es la observaciÃ³n de un hecho concreto.

Necesitamos que la variable vaya acompaÃ±ada de una frecuencia.

Variable cuantitativa temporal: gasto medio del aÃ±o 2000 al 2002

AÃ±o	Gasto	frecuencia
2000	X	1
2001	Y	1
2002	Z	1

En este caso la variable es gasto y sus frecuencias son unitarias, ya que los valores de la variable no se repiten.

Â			
Â		x	
Â	x		
x			
Â			
Â	Â	Â	Â
2000	2001	2002	

Variable cuantitativa no temporal

- Gasto de los alumnos de la clase en los viajes de este verano:

X €

Y €

Z €

W €

X €

Las frecuencias serÃan unitarias en los valores que no se repiten, excepto en X € que sÃ se repite.

- Edad (en intervalos)

Edad	frecuencia
0−10	
11−20	
21−30	

Los intervalos se pueden tratar tambiÃ©n como variable cualitativa.

El nÃºmero de personas que estÃ¡n en cada intervalo serÃa la frecuencia.

TEMA 3. ANÃ LISIS DESCRIPTIVO DE UNA VARIABLE

6/11/2008

HERRAMIENTAS DE ANÃ LISIS UNIVARIANTE DE UNA VARIABLE: PARÃ METROS, ESTADÃ STICOS Y GRÃ FICOS

BibliografÃa: Tema 3 del libro y material de apoyo 2.

AnÃ¡lisis descriptivo o EstadÃstica descriptiva: conjunto de tÃ©cnicas para describir de forma resumida y ordenada el comportamiento de una variable:

− estadÃsticos

− grÃ¡ficos

3.1. ANÃ LISIS DESCRIPTIVO DE UNA VARIABLE CUALITATIVA

- DISTRIBUCIÃ“N DE FRECUENCIAS (numÃ©rica y grÃ¡fica)
- ESTADÃ STICOS DESCRIPTIVOS: moda y mediana

1er PASO. FRECUENCIA O FRECUENCIA ABSOLUTA (ni): nÃºmero de veces que se repite uno de los valores de una variable (Xi). Por ejemplo, una encuesta de edad a un grupo de 100 personas:

Xi	ni
−16	25
17−30	15
31−50	10
51−70	30
+70	20
N =	**100 personas**

25 personas de las 100 encuestadas, han contestado que tienen menos de 16 aÃ±os.

15 personas de las 100 encuestadas, han contestado que tienen entre 17 y 30 aÃ±os.

10 personas de las 100 encuestadas, han contestado que tienen entre 31 y 50 aÃ±os.

30 personas de las 100 encuestadas, han contestado que tienen entre 51 y 70 aÃ±os.

20 personas de las 100 encuestadas, han contestado que tienen mÃ¡s de 70 aÃ±os.

2Âº PASO. FRECUENCIA RELATIVA (fi): es la frecuencia absoluta **(ni)** de cada categorÃa de la variable dividida por el nÃºmero total de observaciones (N).

Xi	ni	fi = (ni/N)
−16	25	0,25
17−30	15	0,15
31−50	10	0,1
51−70	30	0,3
+70	20	0,2
	Total =	**1**

Si tenemos las frecuencias relativas unitariamente (en tantos por 1), la suma de todas ellas debe dar 1. Si lo queremos representarlas en porcentajes, hay que multiplicar cada frecuencia relativa por 100, y la suma de todas ellas darÃ¡ 100. Es recomendable hacerlo en porcentajes y comprobar siempre que la suma de todas es igual a 100.

Xi	ni	fi = (ni/N)	fi %
−16	25	0,25	25%
17−30	15	0,15	15%
31−50	10	0,1	10%
51−70	30	0,3	30%
+70	20	0,2	20%
	Total =	**1**	**100**

DistribuciÃ³n de frecuencias: registro de todos los posibles valores de la variable, junto con sus frecuencias asociadas. (Cuadro de Xi, ni, fi, fi %). La distribuciÃ³n de frecuencias suele ir acompaÃ±ada de representaciones grÃ¡ficas que facilitan y clarifican la lectura de la informaciÃ³n.

DIAGRAMA DE BARRAS: representa cada una de las categorÃ-as de la variable (en el eje X) y su frecuencia relativa o absoluta (en el eje Y) en forma de rectÃ¡ngulos. Siempre hay que indicar quÃ© tipo de frecuencia estamos empleando. Cuando se trata de variables cualitativas, las barras pueden ser de cualquier manera, no hay condiciÃ³n fija, simplemente que quede claro y estÃ©tico. Siguiendo con el ejemplo anterior, quedarÃ-a asÃ-:

valores de la variable

PICTOGRAMA: es un cÃ-rculo donde se **representan las categorÃ-as de la variable**, proporcionalmente a las frecuencias. Siempre se debe empezar a representar desde la lÃ-nea de las 12 en punto y hacia la derecha (si se trata de variables nominales, se empieza igual pero el orden no es tan importante. Se puede hacer calculando los Ã¡ngulos o bien por porcentajes (25, 50, 75%). Los pictogramas nos aportan datos por la forma en que estÃ¡n confeccionados (IMP)

Frecuencia relativa X 360Âº

Ã ngulo del pictograma = −−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−− = Frecuencia relativa X 3,6

100

100 % 25%

75 % 50%

3er PASO. ESTADÃ STICOS DESCRIPTIVOS O DE RESUMEN: son los valores que aportan mÃ¡s informaciÃ³n sobre la variable. Los estadÃ-sticos que mÃ¡s se utilizan son la Moda y la Mediana.

La **MODA** es aquella categorÃ-a de la variable que tiene una mayor frecuencia. Es aquel valor (unimodal) o valores (bimodal o multimodal) de la variable que tienen una mayor frecuencia.

Xi	ni
EspaÃ±ola	40
BritÃ¡nica	10
Alemana	40
otros	10

En este ejemplo sobre nacionalidades, la distribuciÃ³n serÃa bimodal (EspaÃ±ola y Alemana) ya que hay dos valores cuyas frecuencias son las mayores.

Xi	ni
−16	25
17−30	15
31−50	10
51−70	30
+70	20

En este ejemplo de las edades, 30 es la frecuencia mayor y el valor de la variable es entre 51 y 70 aÃ±os. La moda de esta variable es entre 51 y 70 aÃ±os. La variable menor de 16 aÃ±os tiene una frecuencia muy alta, pero por debajo de la moda, pero el resto de variables quedan muy lejos de estos dos valores, por lo que menor de 16 serÃa la casi−moda de esta distribuciÃ³n. **En el pictograma podemos detectar la moda por el Ã¡ngulo mayor**.

La **MEDIANA** es aquel valor de la distribuciÃ³n que ocupa el valor central de la misma. Ordenados los valores de la variable (no de las frecuencias), de menor a mayor, la mediana define aquel punto que deja por debajo de sÃ mismo el 50% de las observaciones. **Para poder calcular la mediana deben cumplirse dos condiciones:**

- Debe tratarse de una variable cualitativa **ORDINAL**
- Hay que calcular las **frecuencias relativas acumuladas** (Fi = sumar cada frecuencia relativa con la siguiente hasta que de 100):

Xi	ni	fi = (ni/N)	fi %	Fi
−16	25	0,25	25%	25%
17−30	15	0,15	15%	40%
31−50	10	0,1	10%	50%
51−70	30	0,3	30%	80%
+70	20	0,2	20%	100%
	Total =	**1**	**100**	

Para saber la mediana, necesitamos el primer valor que contenga el 50%. En este caso, tenemos el valor exacto de 50%, lo que quiere decir que por encima de este valor estÃ¡ el 50% de las observaciones (no de valores) y por debajo, el otro 50%. La mediana de esta variable serÃa entre 31 y 50 aÃ±os. Pero, si en lugar de las frecuencias, pusiÃ©ramos los valores en lÃnea, quedarÃa:

−16, −16, −16,....., 17−30, 17−30, 17−30..., 31−50, 31−50, 31−50..., 51−70, 51−70....+ de 70....

25 veces 15 veces 10 veces 30 veces 20 veces

Si el nÃºmero de observaciones es par, no hay mediana porque ese puesto quedarÃa vacÃo . Es este caso, como hay 100 observaciones, la mediana serÃa tanto el valor 31−50, como el valor 51−50, ya que la observaciÃ³n nÂº 50 serÃa 31−50, y la nÃºmero 51 serÃa 51−70, ya que la mediana se encontrarÃa entre la posiciÃ³n 50 y 51 de las observaciones. Si hubiera 99 observaciones, la mediana serÃa el valor que estuviera en la posiciÃ³n 50. Si cambiamos las frecuencias, queda:

Xi	ni	fi = (ni/N)	fi %	Fi

−16	25	0,25	25%	25%
17−30	15	0,15	15%	40%
31−50	12	0,12	12%	**52%**
51−70	30	0,3	30%	82%
+70	18	0,18	18%	100%
	Total =	**1**	**100**	

En este caso, no habrÃ-a duda de que la mediana es 31−50, porque el valor 52% supera el 50% de las observaciones.

EJEMPLO:

Nacionalidad	ni
Alemana	1500
BritÃ¡nica	1980
EspaÃ±ola	420
Francesa	200
otros	900

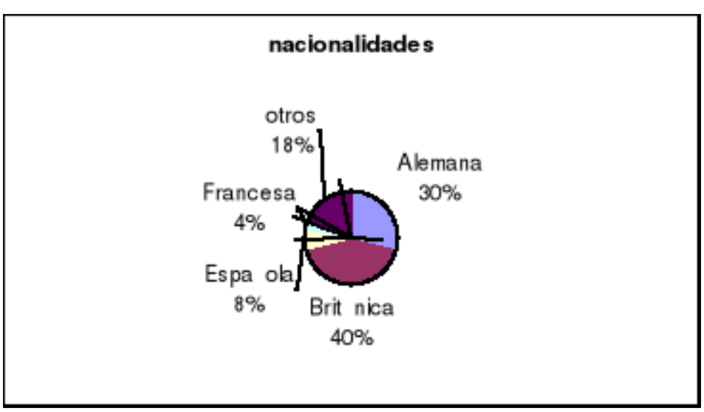
1er paso:

Nacionalidad	ni	fi %	Fi
Alemana	1500	30%	30
BritÃ¡nica	1980	39,6%	69,6
EspaÃ±ola	420	8,4%	78
Francesa	200	4%	82
otros	900	18%	100

N = 5000

2Âº paso: Diagrama de barras:

Pictograma:



<u>Con las frecuencias acumuladas podemos ir seÃ±alando aproximadamente los Ã¡ngulos</u>, suponiendo que el circulo estÃ¡ dividido en 4 partes iguales:

30	Un poco mÃ¡s del 25%
69,6	Menos del 75 %
78	MÃ¡s del 75%
82	
100	

La moda de la variable Nacionalidad es BritÃ¡nica y habrÃa tambiÃ©n una casi−moda que serÃa Alemana. La mediana no se puede calcular porque se trata de una variable cualitativa nominal. <u>No se puede calcular la mediana</u> porque esta variable no se puede ordenar de menor a mayor y ademÃ¡s si cambiÃ¡ramos el orden de las nacionalidades, la mediana cambiarÃa de lugar con cada cambio.

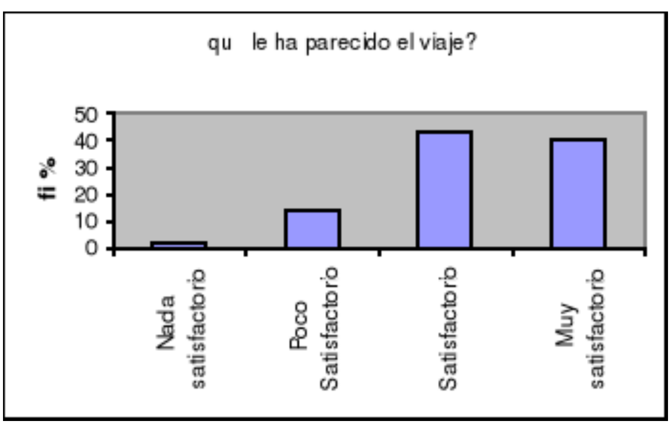
EJEMPLO: (los valores de No sabe/no contesta sÃ³lo se consideran si son muy significativos)

Â¿QuÃ© le ha parecido el viaje?	ni
Nada satisfactorio	6
Poco Satisfactorio	31
Satisfactorio	96
Muy satisfactorio	90

Â¿QuÃ© le ha parecido el viaje?	ni	fi %	Fi
Nada satisfactorio	6	2,69	2,7
Poco Satisfactorio	31	13,90	16,6
Satisfactorio	96	43,049	59,6
Muy satisfactorio	90	40,35	100

N = 223

Diagrama de barras:



Pictograma:

La moda es satisfecho y casi−moda muy satisfecho.

SegÃºn las frecuencias acumuladas, el valor que contiene el 50% es Satisfecho con una Fi del 59,6% (la variable es cualitativa ordinal y estÃ¡ ordenada de menor a mayor). **En el pictograma podemos saber cuÃ¡l es la mediana, por el valor que estÃ¡ en el punto del 50% del cÃrculo.**

EJERCICIO:

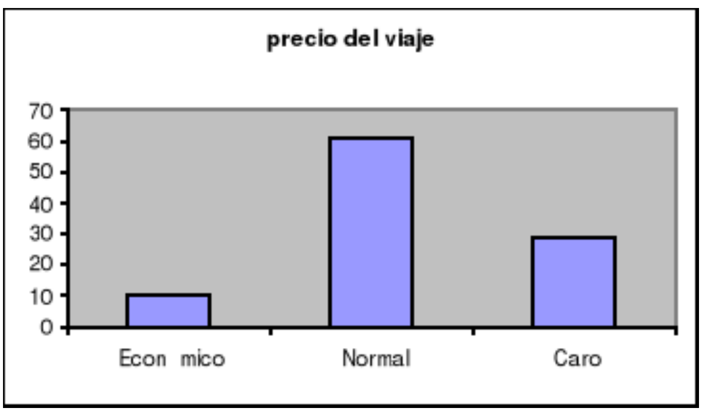
Precio del viaje	ni
EconÃ³mico	480
Normal	2970
Caro	1395

DistribuciÃ³n de frecuencias:

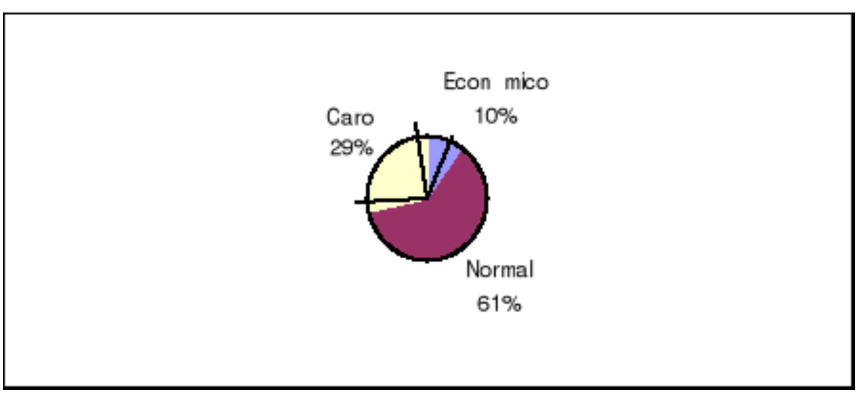
Precio del viaje	ni	fi %	Fi
EconÃ³mico	480	9,9%	9,9
Normal	2970	61,3%	71,2
Caro	1395	28,79%	100

N = 4845

Diagrama de barras:



Pictograma:



Moda: Normal es la variable con mayor frecuencia.

Mediana: Es una variable cualitativa ordinal y ya estÃ¡ ordenada de menor a mayor. SegÃºn las frecuencias acumuladas y el pictograma, el 1er valor que estÃ¡ por encima del 50% es de la variable Normal.

Si se trata de una variable cualitativa ordinal, se ordena dicha variable (no las frecuencias) de menor a mayor y se indican los cuartiles (ver pÃ¡g 28).

13/11/2008

3.2.ANÃ LISIS DESCRIPTIVO DE UNA VARIABLE CUANTITATIVA

La distribuciÃ³n de frecuencias de las variables cuantitativas se calcula igual que con las variables cualitativas. La diferencia es que **con las variables cuantitativas es posible obtener estadÃsticos adicionales a la moda y la mediana**, que facilitan su descripciÃ³n. Para la elaboraciÃ³n de frecuencias es necesario, aunque no siempre, agrupar los valores de la variable en intervalos.

Siempre hay que ordenar la variable de menor a mayor.

1er paso: distribuciÃ³n de frecuencias:

Cuando los valores son unitarios (que su frecuencia sea 1) no es necesario hacer la distribuciÃ³n de frecuencias, y tampoco hay moda (porque todos los valores tienen la misma frecuencia, que serÃa uno). **En el caso de la mediana** de una variable cuantitativa, **si los valores son pares, hay que hacer la media aritmÃ©tica de los valores centrales**, pero si son impares, la mediana serÃ¡ un solo valor.

Ejemplo:

Edad : 20, 21, 23, 23, 25, 27, 28, 29, 29, 30

Moda: 23 y 29 son los valores con mayores frecuencias (distribuciÃ³n bimodal).

Mediana: como el nÃºmero de valores es par (ya estÃ¡n ordenados de menor a mayor) la mediana estarÃa entre 25 y 27, por lo que hay que calcular la media aritmÃ©tica:

25 + 27

−−−−−−−−−− = 26

2

Si el nÃºmero de valores fuera impar, la mediana serÃa un solo valor.

Para tratar variables que toman gran cantidad de valores, el procedimiento usual consiste en agrupar los valores en intervalos. La distribuciÃ³n de frecuencias se realizarÃa tomando como referencia, los distintos intervalos de la variable (marca de clase).

La **marca de clase** es un nÂº que representa a un intervalo. Es el valor medio y central del intervalo y se calcula con la semisuma de los valores del intervalo. Cuando calculamos la marca de clase, Ã©sta serÃ¡ Xi, para poder trabajar con el intervalo.

Edad	Marca de clase (Xi)
[1,19]	10
[20,39]	29,5
[40,59]	49,5
[60,79]	69,5
+ 79	89,5

(1 + 19) / 2 = 10

(20 + 39) / 2 = 29,5

(40 + 59) / 2 = 49,5

(60 + 79) / 2 = 69,5

+ 79 es lo mismo que [80,âˆ]. = 89,5

Para poder saber la marca de clase, en el caso [80,âˆ], observamos cÃ³mo son las de los otros intervalos, quÃ© criterio cumplen. Vemos que las otras marcas de clase son el primer nÂº del intervalo mÃ¡s 9,5. Para el intervalo [80,âˆ] haremos lo mismo, y su marca de clase serÃ¡ 89,5. Los intervalos [...] son cerrados, mientras que + 79 serÃa un intervalo abierto.

2Âº paso: representaciÃ³n grÃ¡fica: HISTOGRAMA

La representaciÃ³n grÃ¡fica de la distribuciÃ³n de frecuencias de una variable <u>continua</u> (son las que toman un nÃºmero infinito de valores) se representa mediante el histograma. El histograma es como un diagrama de barras, pero cada una de las barras es un rectÃ¡ngulo cuya Ã¡rea es la frecuencia relativa, es decir:

ni

Ã rea = Base x Altura = −−−−−−−−

N

Donde la base es la amplitud del intervalo y la altura es:

1 ni

Altura = −−−−−−−−−− x −−−−−−−−

base N

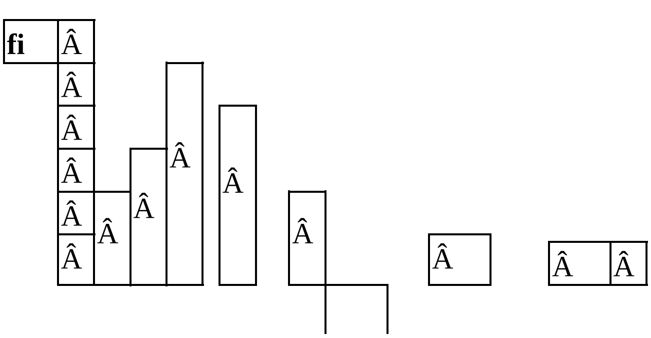
Para que la representaciÃ³n grÃ¡fica sea mÃ¡s sencilla, le damos valor 1 a la base del rectÃ¡ngulo y asÃ conseguimos que el Ã¡rea sea igual a la altura:

ni ni ni

Base x Altura = −−−−−−−− 1 x altura = −−−−−−−− altura = −−−−−−−

N N N

En el histograma no se deja espacio entre las barras, a diferencia del diagrama de barras anterior:



Xi = 1

<u>ESTADÃ STICOS DESCRIPTIVOS: POSICIÃ“N, DISPERSIÃ“N Y FORMA</u>

<u>3.3. ESTADÃ STICOS O MEDIDAS DE POSICIÃ“N (centrales y no centrales)</u>

Ofrecen informaciÃ³n de dÃ³nde se sitÃºan los valores caracterÃsticos de la variable.

MEDIDAS DE POSICIÃ“N CENTRAL: moda, mediana, media aritmÃ©tica, media ponderada y media geomÃ©trica.

Por ej. si los dÃas de vacaciones que han tenido 10 turistas son:

13,14,15,16,17,25,26,26,29,31

La **moda** serÃ¡ 26 (es la que tiene mayor frecuencia) y la **mediana** (valor que se sitÃºa en el punto medio de la distribuciÃ³n) como hay un nÂº par de observaciones, serÃ¡ la media de los valores centrales 17 y 25, es decir, 21 dÃas [(17 + 25) / 2].

Media aritmÃ©tica: suma de todos los valores de la variable, dividida por el total de observaciones. Se distingue de la mediana en que utiliza en su cÃ¡lculo todas las observaciones de la muestra:

âˆ‘ Xi

_ i â†’1

X = −−−−−−−−−−−−−−−−−−−

n

Xi = valores de la variable

n = nÃºmero total de observaciones

Si usamos distribuciones de frecuencias, la fÃ³rmula quedarÃa:

âˆ‘ Xi * ni

_ i â†’1

X = −−−−−−−−−−−−−−−−−−−

n

Xi = valores de la variable

n = nÃºmero total de observaciones = âˆ‘ni = N

ni = frecuencia absoluta

Con la 1Âº fÃ³rmula serÃa:

âˆ‘ Xi

_ i â†’1 13 + 14 + 15 + 16 +17 +25 +26 +26 +29 +31

X = −−−−−−−−−−− = −−

n 10

Con la 2Âª fÃ³rmula:

âˆ‘ Xi * ni

_ i â†’1 13*1 + 14*1 + 15*1 + 16*1 +17*1 +25*1 +(26* 2) +29*1 +31*1

X = −−−−−−−−−−− = −−−

n 10

Supongamos que tenemos las siguientes observaciones:

10,10,10,15,15,20,20,20,20,20,21

La moda serÃ-a 20 y la mediana serÃ-a 20

_

La media aritmÃ©tica (X) serÃ-a:

âˆ‘ Xi

_ i â†’1 10+10+10+15+15+20+20+20+20+20+21

X = −−−−−−−−−−− = −−−

n 11

Si pasamos las observaciones a una distribuciÃ³n de frecuencias, quedarÃ-a (con los valores de la variable ordenados de menor a mayor, no las frecuencias):

Xi	ni
10	3
15	2
20	5
21	1

n = 11

Y calcularÃ-amos la media aritmÃ©tica:

âˆ‘ Xi * ni

_ i â†’1 (10*3) + (15*2) + (20*5) + (21* 1)

X = −−−−−−−−−−− = −−

n 11

Si hay muchos datos, es mejor emplear esta segunda fÃ³rmula.

PROPIEDADES DE LA MEDIA ARITMÃ‰TICA (IMP)

− La suma de las desviaciones de los valores de la variable (diferencias entre cada valor y la media aritmÃ©tica), respecto a la media es igual a cero:

n _

âˆ‘ (Xi − x) = 0

i = 1

Xi	ni	media arit	Xi − media arit
10	3	16,45	−6,45
15	2	16,45	−1,45
20	5	16,45	3,55
21	1	16,45	4,55
		âˆ‘ (Xi − media arit)=	0

- Si a los valores de la variable se les suma una constante, la media de los valores transformados se incrementa en esa cantidad:

Xi	Xi + 5
10	15
15	20
20	25
21	26

âˆ‘ Xi+5 * ni

_ i â†’1 (15*3) + (20*2) + (25*5) + (26* 1)

Xi + 5 = −−−−−−−−−−−−−− = −−− = 21,45

n 10

_ _

Xi + 5 = Xi + 5 = 16,45 + 5 = 21,45

- Si los valores se multiplican por una constante, la media de los valores transformados serÃ¡ la media original multiplicada por esa constante:

Xi	Xi * 2
10	20
15	30
20	40
21	42

âˆ‘ Xi+2 * ni

_ i â†’1 (20*3) + (30*2) + (40*5) + (42* 1)

Xi*2 = ––––––––––––––– = –– = 32,9

n 10

_ _

Xi*2 = Xi * 2 = 16,45 * 2 = 32,9

Media ponderada: los valores promediados son ponderados mediante un peso determinado. Si tenemos p valores distintos de la variable Xi y a cada uno de ellos se le da un peso wi, la media ponderada se define como:

$$\bar{x}_w = \frac{\sum_{i=1}^{p} w_i X_i}{\sum_{i=1}^{p} w_i}$$

Las observaciones de una variable tambiÃ©n podrÃan ponderarse por sus frecuencias relativas: wi = ni/n, es decir, por la importancia relativa de cada valor en la distribuciÃ³n

En ocasiones, la ponderaciÃ³n tiene que ver con el peso en la poblaciÃ³n de las distintas observaciones. Supongamos, por ejemplo, que se dispone del gasto turÃstico diario de tres personas que representan a los turistas que han visitado tres municipios distintos de una regiÃ³n. El gasto de cada uno de ellos es de 200, 200 y 100 euros. Si se pidiese calcular el gasto turÃstico medio diario en la regiÃ³n, una opciÃ³n directa serÃa la siguiente:

En ese perÃodo, el nÃºmero de turistas que ha visitado cada municipio ha sido 10.000, 20.000 y 1.000.000 turistas, respectivamente. ResultarÃa lÃ³gico utilizar una media ponderada, donde el gasto de cada turista de la muestra serÃa asignado a todos los turistas del correspondiente municipio:

Dependiendo del peso que empleemos, la media aritmÃ©tica saldrÃ¡ distinta.

Media geomÃ©trica: se utiliza para calcular los promedios de valores que son porcentajes, tasas, tipos de interÃ©s, nÃºmeros Ãndices..., es decir, valores que representan variaciones acumulativas, de un perÃodo anterior. Cada valor se eleva a su frecuencia relativa y se hace raÃz cuadrada n.

$$g=\sqrt[n]{X_1^{n_1}\ X_2^{n_2}\dots X_p^{n_p}}$$

Por ej: 15%, 18%, 20%

g =

Los valores los pondrÃ-amos en tantos por 1 y harÃ-amos la raÃ-z 3, porque hay 3 valores.

MEDIDAS DE POSICIÃ“N (NO CENTRALES): CUARTILES, DECILES, Y PERCENTILES

Los **cuartiles** (primer, segundo y tercer cuartil) **son los valores que** (con la variable ordenada de menor a mayor) **dejan por debajo de su posiciÃ³n el 25%, 50% y 75% de las frecuencias acumuladas,** respectivamente. Los cuartiles dividen la variable en cuatro grupos con igual nÃºmero de observaciones (el 25% de valores mÃ¡s bajos, el 25% siguiente,).

Los **deciles** dividen la muestra en porcentajes del 10, 20, ..., hasta el 90 por ciento.

Los **percentiles** amplÃ-an esta idea para definirse sobre porcentajes del 1, 2, 3, , hasta el 99 por ciento.

Xi	fi	Fi %			
10	29	29	1er cuartil	1y 2Âº decil	1−28 percentil
20	15	44	Â	3 y 4Âº decil	29−43 percentil
30	19	63	2Âº cuartil	5 y 6Âº decil	44−62 percentil
40	22	85	3er cuartil	7 y 8Âº decil	63−84 percentil
50	15	100	Â	9Âº decil	85−99 percentil

Los deciles y percentiles se emplean poco. Sirven cuando hay muchÃ-simos valores.

EQUIVALENCIAS IMPORTANTES:(examen)

Mediana = Segundo cuartil = 50 percentil

Primer cuartil = 25 percentil

Tercer cuartil = 75 percentil

Hay 3 cuartiles

Hay 9 deciles

Hay 99 percentiles

3.4. MEDIDAS O ESTADÃ STICOS DE DISPERSIÃ“N

Estas medidas son Ãºtiles cuando la media no es representativa de lo que ocurre en los datos y necesitamos saber cuÃ¡n cerca o lejos de la media se sitÃºan esos datos, es decir, su dispersiÃ³n. Las medidas de posiciÃ³n ofrecen informaciÃ³n de dÃ³nde se sitÃºan los valores caracterÃ-sticos de la variable, pero las medidas de posiciÃ³n deben complementarse con las medidas de dispersiÃ³n de la variable: **Rango, rango intercuartÃ-lico, varianza, desviaciÃ³n tÃ-pica o estÃ¡ndar, y coeficiente de variaciÃ³n.**

Rango: El rango de una variable es la diferencia entre el mayor y el menor valor de las observaciones (valores de la variable, no de las frecuencias):

Rango = X MAX −X MIN

Como medida de dispersiÃ³n el rango es muy sensible a los valores extremos. Si comparamos dos variables, la que tenga un **rango mayor**, serÃ¡ la **mÃ¡s dispersa** y la que tenga rango menor, tendrÃ¡ menos dispersiÃ³n o mÃ¡s concentraciÃ³n.

Rango intercuartÃ-lico: es la diferencia entre el tercer y primer cuartil de la variable. En ese intervalo se incluyen, por tanto, el 50% de las observaciones situadas en la zona central de la distribuciÃ³n:

Rango IQ = Q3 − Q1

Â	50 % central = RIQ	Â
	100 % de las observaciones	

Por lo tanto, nos da informaciÃ³n sobre **cuÃ¡n concentrados o dispersos estÃ¡n los valores centrales.(IMP)**

Varianza: La varianza es el promedio de las desviaciones de la variable con respecto a la media al cuadrado, es decir, que indica cuÃ¡n concentrados o dispersos son los valores respecto a la media aritmÃ©tica:

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

Si utilizamos frecuencias absolutas, hay que multiplicar la fÃ³rmula por ni:

n _

âˆ‘ (Xi− x)2 * ni

i = 1

Sx2 = −−−−−−−−−−−−−−−−−

N

La varianza es la medida de dispersiÃ³n mÃ¡s importante para indicar la distancia de las observaciones respecto a la media aritmÃ©tica. **A mayor varianza, mayor dispersiÃ³n**. Su valor **depende de la unidad de medida de la variable**. La varianza de una sola distribuciÃ³n no nos aporta mucha informaciÃ³n por sÃ-sola, pero **es muy Ãºtil para compararla con otras distribuciones.**

Otra forma de calcular la varianza es:

$$s_x^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{x}^2$$

N = n = âˆ‘ni

<u>PROPIEDADES DE LA VARIANZA:(examen)</u>

- La varianza es siempre positiva (al ser el cuadrado de un nÃºmero, aunque Ã©ste fuera negativo, el resultado serÃ-a positivo)
- Es sensible a los valores extremos porque es una medida de dispersiÃ³n.
- Si a los valores de la variable se les suma una constante c, la varianza de los valores modificados no se modifica, es decir, que seguirÃ-a siendo la misma.

Sc +x2 = Sx2

- Si a los valores de la variable se les multiplica por una constante c, la varianza de los valores modificados queda multiplicada por el cuadrado de la constante:

Sc *x2 = c2 * Sx2

- Otra forma de calcular la varianza, si empleamos frecuencias absolutas es:

n

âˆ‘ Xi2* ni

i = 1 _

Sx2 = −−−−−−−−−−−−− − x 2

N

Xi	ni	Xi*ni	Xi2	Xi2*ni
Â	Â	Â	Â	Â
Â	Â	Â	Â	Â

DesviaciÃ³n tÃ-pica o EstÃ¡ndar: es la raÃ-z cuadrada de la varianza y tiene la misma unidad de medida que la variable.

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

Como medidas de dispersiÃ³n, la desviaciÃ³n estÃ¡ndar y la varianza muestran su utilidad, especialmente, cuando se emplean de manera comparativa.

Coeficiente de variaciÃ³n (MUY IMP.): es la relaciÃ³n entre la desviaciÃ³n estÃ¡ndar (medida de dispersiÃ³n) y la media aritmÃ©tica de la variable (medida de posiciÃ³n). Facilita la discusiÃ³n de la importancia de la dispersiÃ³n de una variable. **Hay que expresarlo en tantos por ciento** (para hacerlo se multiplica por 100).

$$CV = \frac{s_x}{\bar{x}}$$

sx

CV en % = _ * 100

X

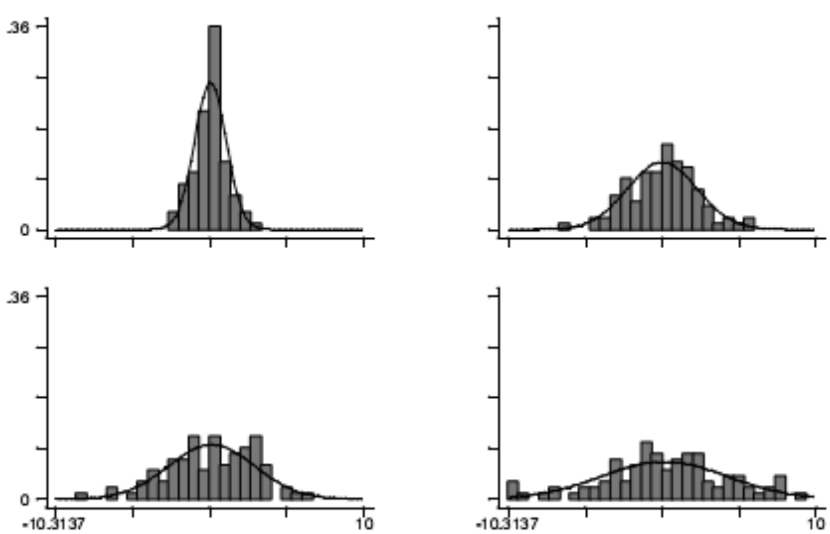
Resulta de especial interÃ©s en dos circunstancias:

- Cuando se quieren comparar distribuciones de variables con diferentes unidades de medida.
- Cuando se comparan distribuciones que, aun teniendo igual unidad de medida, toman valores muy distintos.

El resultado del coeficiente de variaciÃ³n en tantos por ciento, **debe compararse con la media aritmÃ©tica para determinar si el % resultante es grande o pequeÃ±o**. Este coeficiente <u>no depende de la unidad de medida de la variable</u>, por eso sirve para comparar distribuciones y unidades de medidas diferentes. **Cuanto mayor sea el C.V., mayor serÃ¡ la dispersiÃ³n.**

Si por ej., el C.V. da 0,02, lo multiplicamos por 100 para ponerlo en %, lo que nos darÃa 2%. Esto significa que el C.V. es de un 2% respecto a la media aritmÃ©tica, lo que serÃa un porcentaje bastante bajo.

Como medidas de dispersiÃ³n, la desviaciÃ³n estÃ¡ndar y la varianza son Ãºtiles para comparar diferentes distribuciones:



Ejemplo 3.19. Con los siguientes datos: {9, 23, 25, 28, 32, 47, 50} calcularemos las diversas medidas de dispersión:

$$Rango = X_{MAX} - X_{MIN} = 50 - 9 = 41$$

$$Rango_{IQ} = Q_3 - Q_1 = 47 - 23 = 24$$

Para calcular la varianza emplearemos los datos del cuadro 3.17, y nos basaremos en su expresión.

i	X_i	X_i^2
1	9	81
2	23	529
3	25	625
4	28	784
5	32	1024
6	47	2209
7	50	2500

$\sum_{i=1}^{7} X_i = 214$	$\sum_{i=1}^{7} X_i^2 = 7752$
$\bar{x} = \frac{\sum_{i=1}^{7} X_i}{7} = 30{,}571$	$\frac{\sum_{i=1}^{7} X_i^2}{n} = 1107{,}43$
$s_x^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{x}^2 = 172{,}84$	

La desviación estándar será, por tanto:

$$s_x = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n}} = \sqrt{172{,}84} = 13{,}147$$

Y el coeficiente de variación:

$$CV = \frac{s_x}{\bar{x}} = \frac{13{,}147}{30{,}571} = 0{,}43$$

3.5. SIMETRÃ A DE UNA DISTRIBUCIÃ“N (histogramas)EXAMEN

SegÃºn su forma, las distribuciones pueden dividirse en simÃ©tricas, asimÃ©tricas por la derecha y asimÃ©tricas por la izquierda. **Las distribuciones simÃ©tricas tiene la misma forma a izquierda y derecha de la mediana. Las distribuciones unimodales asimÃ©tricas por la derecha tiene mÃ¡s valores a la derecha del intervalo modal que a su izquierda**. Una distribuciÃ³n con la forma contraria es asimÃ©trica por la izquierda.

Aunque existen estadÃsticos que describen la simetrÃa (o asimetrÃa) de la distribuciÃ³n, los valores de la media, mediana y moda de la variable facilitan tambiÃ©n esta informaciÃ³n. En una distribuciÃ³n simÃ©trica la media, la mediana y la moda tienden a coincidir, mientras que en las distribuciones asimÃ©tricas se dispersan:

MEDIDAS DE ASIMETRÃ A Y MEDIDAS DE APUNTAMIENTO O CURTOSIS

Existen medidas de forma que proporcionan informaciÃ³n numÃ©rica sobre dos caracterÃsticas de la distribuciÃ³n: su simetrÃa y su apuntamiento o curtosis. **El apuntamiento o curtosis se refiere a la importancia de la concentraciÃ³n de las observaciones en la zona central de la variable**, mostrÃ¡ndose con ello mÃ¡s o menos apuntada.

Las distribuciones simÃ©tricas tiene la misma forma a izquierda y derecha de la mediana. Cualquier medida que recoja alteraciones de esta situaciÃ³n proporcionarÃ¡ una **cuantificaciÃ³n de la asimetrÃa de la distribuciÃ³n**. Una primera propuesta podrÃa ser la siguiente:

n _

âˆ‘ (Xi− x)

i = 1

−−−−−−−−−−−−−−−−−

N

en tanto que define un promedio de las desviaciones de la variable con respecto al valor medio. En el caso de que se dieran muchas observaciones por encima de la media, se esperarÃa un valor positivo, mientras que una mayor proporciÃ³n de valores por debajo de la media proporcionarÃa un valor negativo. El problema es que si se cumple una de las propiedades de la media aritmÃ©tica, que es

− La suma de las desviaciones de los valores de la variable (diferencias entre cada valor y la media aritmÃ©tica), respecto a la media es igual a cero:

n _

âˆ‘ (Xi − x) = 0

i = 1

Una modificaciÃ³n alternativa es tomar las desviaciones de la variable respecto a la media, pero elevadas a alguna potencia. Al elevar las desviaciones al cuadrado se obtiene la expresiÃ³n de la varianza, una medida de dispersiÃ³n de los valores respecto a la media, pero que no ayuda a seÃ±alar la posiciÃ³n de las observaciones a derecha o izquierda de la medida de posiciÃ³n central. El cubo de las desviaciones, al respetar el signo de la diferencia de la operaciÃ³n (Xi − media arit.), permite promediar tanto la importancia de la desviaciÃ³n como su direcciÃ³n. Se utilizarÃa por tanto, un estadÃstico conocido como el **momento de orden tres con respecto a la media (m3):**

n _

âˆ‘ (Xi− x)3

i = 1

m3 = −−−−−−−−−−−−−−−−−

N

(Esta fÃ³rmula habrÃa que multiplicarla por ni si empleamos frecuencias absolutas).

Xi	ni	Xi*ni	Xi2	Xi2* ni	(Xi − media)3	(Xi − media)3* ni

- Si m3 es igual a cero, es simÃ©trica.
- Si m3 es > cero, es asimÃ©trica por la derecha
- Si m3 es < cero, es asimÃ©trica por la izquierda

MEDIDA DE ASIMETRÃ A DE FISHER

Con la siguiente fÃ³rmula se evita que la medida varÃe si se produce un cambio de escala, y se normaliza mediante el cubo de la desviaciÃ³n estÃ¡ndar:

m3

g1 = −−−−−−−

Sx3

- Si g1 es igual a cero, es simÃ©trica.
- Si g1 es > cero, es asimÃ©trica por la derecha
- Si g1 es < cero, es asimÃ©trica por la izquierda

MEDIDA DE ASIMETRÃ A DE PEARSON

Relaciona la media aritmÃ©tica y la moda respecto a la desviaciÃ³n estÃ¡ndar:

_

X − moda

AS = −−−−−−−−−−−

SX

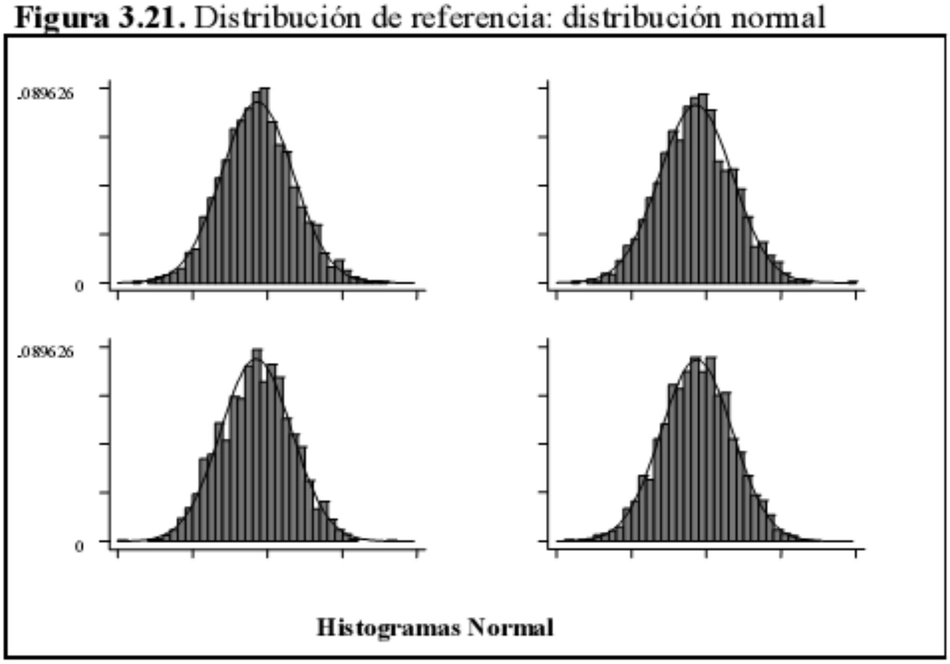
- Si AS es igual a cero, es simÃ©trica.
- Si AS es > cero, es asimÃ©trica por la derecha

Si AS es < cero, es asimÃ©trica por la izquierda

MEDIDAS DE APUNTAMIENTO O CURTOSIS

Se aplican a distribuciones <u>simÃ©tricas y unimodales</u>, pretenden medir hasta quÃ© punto las observaciones de la variable se acumulan en la parte central de la distribuciÃ³n.

Figura 3.21. Distribución de referencia: distribución normal



$$Curtosis = \frac{\sum_{i=1}^{n}(X_i - \bar{x})^4 / n}{s_x^4} = \frac{m_4}{s_x^4}$$

En las distribuciones mesocÃºrticas, la curtosis es igual a 3. (DistribuciÃ³n normal)

Coeficiente de curtosis: permite la comparaciÃ³n directa con la distribuciÃ³n normal:

$$g_2 = \frac{m_4}{s_x^4} - 3$$

$g_2 = 0$: distribución mesocúrtica (normal).

$g_2 < 0$: distribución platicúrtica (“achatada”).

$g_2 > 0$: distribución leptocúrtica (“apuntada”).

RESUMEN:

Variables cualitativas nominales: Moda. Variables cualitativas ordinales: Moda y mediana. Variables cuantitativas temporales y atemporales: TODO

Si un valor se repite, significa que hay frecuencia. Si no se repite, su frecuencia es 1. Si no hay muchas frecuencias, no vale la pena calcular los deciles ni los percentiles.

Las frecuencias acumuladas sÃ³lo sirven para la mediana.

Si la moda tiene un valor prÃ³ximo a los demÃ¡s valores, hay que indicarlo. TambiÃ©n hay que indicar cuÃ¡n diferente es la moda de los demÃ¡s valores.

La varianza por sÃ- sola no aporta mucha informaciÃ³n. El coeficiente de variaciÃ³n indica el % de variabilidad.

De cada fÃ³rmula, hay que sacar una conclusiÃ³n sobre la informaciÃ³n que aporta.

TEMA 4. ANÃ LISIS CONJUNTO DE DOS VARIABLES CUALITATIVAS

11/12/2008

El anÃ¡lisis conjunto de dos variables consiste en averiguar la relaciÃ³n que existe entre esas variables. Cruzar los datos de dos variables aporta mucha mÃ¡s informaciÃ³n que una sola variable.

4.1. OBTENCIÃ“N DE INFORMACIÃ“N A TRAVÃ‰S DE ENCUESTAS

DespuÃ©s de realizar una encuesta, si queremos averiguar si hay relaciÃ³n de dependencia o no entre dos variables, es decir, intentar establecer cuÃ¡l es la variable que influye en la otra, lo primero que hay que saber es si esas variables son cualitativas o si son cuantitativas:

- Si son CUALITATIVAS, se estudia la ASOCIACIÃ“N entre ellas.

- Si son CUANTITATIVAS, se estudia la CORRELACIÃ“N o RELACIÃ“N LINEAL que hay entre ellas.

- En el caso de que una variable sea cualitativa y la otra sea cuantitativa, se tratarÃ¡n ambas variables como cualitativas. Si la variable es cuantitativa, pero se puede tratar como cualitativa ordinal (definiÃ©ndola en intervalos de valores) se estudiarÃa tambiÃ©n la asociaciÃ³n. Hay variables cuantitativas que se pueden estudiar como cualitativas, pero no ocurre lo mismo al revÃ©s.

4.2. RELACIÃ“N ENTRE DOS VARIABLES CUALITATIVAS.

1Âº. CONSTRUCCIÃ“N DE LA TABLA DE CONTINGENCIA.

(Se puede hacer lo mismo con las variables cuantitativas)

Para describir la relaciÃ³n entre dos variables cualitativas, se emplea la **tabla de frecuencias conjunta o tabla de contingencia.** Se considera que cada una de las variables puede tomar una serie de valores mutuamente excluyentes. Como en el caso de la estadÃstica univariante, la primera manera de resumir la informaciÃ³n es el puro cÃ³mputo. Empleando una tabla de doble entrada se muestran todas las posibles combinaciones de las categorÃas de las dos variables, anotando en cada una de las celdas resultantes el nÃºmero de casos que pertenecen a las dos categorÃas. En las tablas de contingencia **se muestran las frecuencias (absolutas o relativas) en las que ocurren las categorÃas de filas y columnas.**

Ejemplo. Con el objetivo de conocer el nivel de satisfacciÃ³n de los turistas que han pasado sus vacaciones en una regiÃ³n, se ha realizado una encuesta en la que se pregunta al turista si estÃ¡ satisfecho de sus vacaciones. Las posibles respuestas a la pregunta son Mucho, Bastante, Poco y Nada. Considerando que la satisfacciÃ³n alcanzada puede estar en funciÃ³n de la zona concreta donde se ha realizado la estancia, se ha cruzado la pregunta sobre satisfacciÃ³n con una variable que identifica cuatro zonas de la regiÃ³n.

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**
MUCHO	**395**	**593**	**358**	**200**
BASTANTE	**100**	**182**	**120**	**94**
POCO	**167**	**233**	**102**	**90**
NADA	**212**	**249**	**118**	**115**

Esta tabla se interpretarÃa por ej., que 395 turistas que se alojaron en la zona 1 les ha gustado mucho el viaje o bien que 395 turistas a los que les ha gustado mucho el viaje, se alojaron en la zona 1.

Las variables se pueden colocar indistintamente en la vertical o en la horizontal de la tabla. O sea, se podrÃan colocar las zonas en la vertical y el grado de satisfacciÃ³n en la horizontal.

2Âº. CALCULAR LAS DISTRIBUCIONES MARGINALES: Son las frecuencias (absolutas y relativas) de cada una de las variables. Para ello es suficiente con sumar todas las celdas correspondientes a cada una de las filas o de las columnas.

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**	**dist. Marginal satisfacciÃ³n**
MUCHO	**395**	**593**	**358**	**200**	**1546**
BASTANTE	**100**	**182**	**120**	**94**	**496**
POCO	**167**	**233**	**102**	**90**	**592**
NADA	**212**	**249**	**118**	**115**	**694**
dist. Marginal Zonas	**874**	**1257**	**698**	**499**	**3328**

En la fila distribuciÃ³n marginal Zonas estÃ¡ la suma de los totales de cada zona, es decir, la suma de cada columna. En la columna distribuciÃ³n marginal SatisfacciÃ³n del viaje, estÃ¡ la suma de los totales de cada grado, es decir, la suma de cada fila.

En las distribuciones marginales, **podemos saber cuÃ¡l es la moda de cada variable**. En el caso de las zonas, la moda es la Zona 2 y en el caso de la satisfacciÃ³n la moda es Mucho.

El dato 3328 es el **total de frecuencias,** es decir N o n.

3Âº. TABLA DE DISTRIBUCIÃ“N DE FRECUENCIAS RELATIVAS:

Consiste en calcular las frecuencias relativas de cada una de las celdas, respecto al total de observaciones(n). Estos datos se ponen en una tabla nueva. Las frecuencias relativas se calculan ni/n si se calculan en tantos por uno, y se multiplica por 100 para ponerlo en %.

Por ej:

La tabla de distribuciones marginales es:

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**	**dist. Marginal satisf.**
MUCHO	**395**	**593**	**358**	**200**	**1546**
BASTANTE	**100**	**182**	**120**	**94**	**496**
POCO	**167**	**233**	**102**	**90**	**592**
NADA	**212**	**249**	**118**	**115**	**694**
dist. Marginal zonas	**874**	**1257**	**698**	**499**	**3328**

El cÃ¡lculo para la primera casilla serÃ-a:

395

−−−−−− = 0,1187 * 100 = 11,87

3328

La **tabla de distribuciÃ³n de frecuencias relativas** serÃ-a:

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**	**TOTAL**
MUCHO	**11,87**	**17,82**	**10,76**	**6,01**	**46,45**
BASTANTE	**3**	**5,47**	**3,61**	**2,82**	**14,90**
POCO	**5,02**	**7**	**3,06**	**2,70**	**17,79**
NADA	**6,37**	**7,48**	**3,55**	**3,46**	**20,85**
TOTAL	**26,26**	**37,77**	**20,97**	**14,99**	**100**

Las frecuencias relativas se suman por filas y por columnas. **Para comprobar si estÃ¡ bien calculado, la suma de la fila Total y la suma de la columna Total, debe ser igual a 100**.

Una vez se ha completado la tabla, se comentan los datos, como cuÃ¡l es el porcentaje mayor y el menor, etc. En este caso, el 17,82% de turistas, que se alojaron en la zona 2, les gustÃ³ mucho el viaje y el 2,7%, que estaban en la zona 4, les gustÃ³ poco el viaje. Pero a un 46,45% de los turistas, les gustÃ³ mucho el viaje (sin distinguir zona) y un 37,77% de los turistas se alojÃ³ en la zona 2.

4Âº. TABLA DE PERFILES FILA (PORCENTAJES FILA):

Recogen las frecuencias relativas de cada una de las celdas con respecto al total de las filas. Estos datos se ponen en una tabla nueva. Ahora se dividen los valores de la tabla original, entre el total de la fila de la tabla original. Los perfiles fila se calculan en tantos por uno, y se multiplica por 100 para ponerlo en %.

La tabla de distribuciones marginales es:

	ZONA 1	ZONA 2	ZONA 3	ZONA 4	dist. Marginal Satisfacción.
MUCHO	**395**	**593**	**358**	**200**	**1546**
BASTANTE	**100**	**182**	**120**	**94**	**496**
POCO	**167**	**233**	**102**	**90**	**592**
NADA	**212**	**249**	**118**	**115**	**694**
dist. Marginal Zonas	**874**	**1257**	**698**	**499**	**3328**

El cÃ¡lculo para la primera casilla serÃa:

395

−−−−−− = 0,2555 * 100 = 25,55

1546

Los perfiles fila se suman por filas. **Para comprobar si estÃ¡ bien calculado, la suma de cada una de las filas debe ser igual a 100**. Si vemos una tabla con una columna en la que todos los datos son 100, nos indica que es una tabla de perfiles fila.

La **tabla de perfiles fila** quedarÃa:

	ZONA 1	ZONA 2	ZONA 3	ZONA 4	TOTAL
MUCHO	**25,55**	**38,36**	**23,16**	**12,94**	**100**
BASTANTE	**20,16**	**36,69**	**24,19**	**18,95**	**100**
POCO	**28,21**	**39,36**	**17,23**	**15,20**	**100**
NADA	**30,55**	**35,88**	**17,00**	**16,57**	**100**

Hay que tener cuidado cuando se interpretan las tablas de perfilas fila. En este caso, de los datos de la fila Nada se interpretarÃan que del total de turistas que no les ha gustado nada el viaje, el 17% se ha alojado en la zona 3.

Cada dato de la fila, es un parcial del 100% de la fila.

5Âº. TABLA DE PERFILES COLUMNA (PORCENTAJES COLUMNA):

Recogen las frecuencias relativas de cada una de las celdas con respecto al total de las columnas. Estos datos se ponen en una tabla nueva. Ahora se dividen los valores de la tabla original, entre el total dela columna de la tabla original. Los perfiles columna se calculan en tantos por uno, y se multiplica por 100 para ponerlo en %.

La tabla de distribuciones marginales es:

	ZONA 1	ZONA 2	ZONA 3	ZONA 4	dist. Marginal SatisfacciÃ³n.
MUCHO	**395**	**593**	**358**	**200**	**1546**
BASTANTE	**100**	**182**	**120**	**94**	**496**
POCO	**167**	**233**	**102**	**90**	**592**
NADA	**212**	**249**	**118**	**115**	**694**
dist. Marginal Zonas	**874**	**1257**	**698**	**499**	**3328**

El cÃ¡lculo para la primera casilla serÃa:

395

−−−−−− = 0,4519 * 100 = 45,19

874

Los perfiles columna se suman por columnas. **Para comprobar si estÃ¡ bien calculado, la suma de cada una de las columnas debe ser igual a 100**. Si vemos una tabla con una fila en la que todos los datos son 100, nos indica que es una tabla de perfiles columna.

La **tabla de perfiles columna** quedarÃa:

	ZONA 1	ZONA 2	ZONA 3	ZONA 4
MUCHO	**45,19**	**47,18**	**51,29**	**40,08**
BASTANTE	**11,44**	**14,48**	**17,19**	**18,84**
POCO	**19,11**	**18,54**	**14,61**	**18,04**
NADA	**24,26**	**19,81**	**16,91**	**23,05**
TOTAL	**100**	**100**	**100**	**100**

En este caso, los datos de la columna zona 2 se interpretarÃan como que de todos los turistas alojados en la zona 2, el 14,48% ha dicho que les ha gustado bastante el viaje.

Cada dato de la columna, es un parcial del 100% de la columna.

EJERCICIO: Se ha pedido a 219 usuarios de un servicio si han quedado satisfechos o no, teniendo en cuenta si ya lo habÃan utilizado antes. Hacer la descripciÃ³n de datos de la siguiente tabla:

	REPETIDORES	NO REPETIDORES
SATISFECHOS	86	43
NO SATISFECHOS	36	54

219 = n o N

Distribuciones marginales segÃºn repeticiÃ³n y satisfacciÃ³n:

REPETIDORES	NO REPETIDORES	DIST. MARGINAL SATIS.

SATISFECHOS	86	43	129
NO SATISFECHOS	36	54	90
DISTRIBUCIÃ“N MARGINAL REP.	122	97	219

La moda es repetidores y en cuanto a la satisfacciÃ³n, la moda es satisfechos.

Frecuencias relativas: Todas las casillas se dividen entre N (219):

	REPETIDORES	NO REPETIDORES	TOTAL
SATISFECHOS	39,27	19,63	58,90
NO SATISFECHOS	16,44	24,66	41,10
TOTAL	55,71	44,29	100

El 39,27% de los encuestados eran repetidores y estÃ¡n satisfechos y el 16,44 tambiÃ©n son repetidores y no estÃ¡n satisfechos.

Perfiles fila: ahora se divide por los nÂº marginales:

	REPETIDORES	NO REPETIDORES	TOTAL
SATISFECHOS	66,67	33,33	100,00
NO SATISFECHOS	40,00	60,00	100,00

Del total de satisfechos, un 66,67% son repetidores. Del total de los no satisfechos, el 60% no son repetidores.

Perfiles Columna:

	REPETIDORES	NO REPETIDORES
SATISFECHOS	70,49	44,33
NO SATISFECHOS	29,51	55,67
TOTAL	100,00	100,00

Del total de repetidores, un 70,49% estÃ¡n satisfechos y del total de no repetidores un 55,67% no estÃ¡n satisfechos.

4.3. ESTADÃ STICOS DE ASOCIACIÃ“N. INDEPENDENCIA Y ASOCIACIÃ“N

18/12/2008

EstadÃ-sticos de asociaciÃ³n para variables cualitativas nominales u ordinales:

chi−cuadrado

C de contingencia

lambda

EstadÃ-sticos de asociaciÃ³n para variables ordinales:

Gamma

CHI− CUADRADO Y C DE CONTINGENCIA

Una medida sintÃ©tica del grado de asociaciÃ³n se calcula a partir de la comparaciÃ³n entre los valores observados y los valores que uno esperarÃ-a encontrar en el caso de inexistencia de asociaciÃ³n.

Las **frecuencias esperadas** son el producto de las frecuencias marginales dividido entre el total de observaciones N:

ni x nj

eij = −−−−−−−−−−−−

n

Ahora en cada casilla tendremos las dos frecuencias: la absoluta observada y la esperada:

Tabla de distribuciones marginales:

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**	**dist. Marginal satisf.**
MUCHO	**395**	**593**	**358**	**200**	**1546**
BASTANTE	**100**	**182**	**120**	**94**	**496**
POCO	**167**	**233**	**102**	**90**	**592**
NADA	**212**	**249**	**118**	**115**	**694**
dist. Marginal zonas	**874**	**1257**	**698**	**499**	**3328**

Los cÃ¡lculos de las frecuencias esperadas serÃ-an, entre otros:

1546 x 874 1546 x 1257

−−−−−−−−−−−−− = 406,1 −−−−−−−−−−−−−−− = 583,93

- 3328

496 x 874 496 x 1257

−−−−−−−−−−−−− = 130,26 −−−−−−−−−−−−−−− = 184,37

- 3328

Tabla de frecuencias esperadas:

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**	**dist. Marginal satisfacciÃ³n**
MUCHO	**406,01**	**583,93**	**324,25**	**231,81**	**1546**
BASTANTE	**130,26**	**187,34**	**104,03**	**74,37**	**496**

POCO	**155,47**	**223,6**	**124,16**	**88,764**	**592**
NADA	**182,26**	**262,13**	**145,56**	**104,06**	**694**
dist. Marginal zonas	**874**	**1257**	**698**	**499**	**3328**

CHI−CUADRADO

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}}$$

Se calcula a partir de las frecuencias absolutas (nij) observadas y de las esperadas (eij). Este estadÃ-stico suma para todas las celdas la diferencia, elevada al cuadrado, entre la frecuencia observada y la esperada. Si no existe ningÃºn grado de asociaciÃ³n entre las variables cualitativas, los valores esperados serÃ-an iguales a los valores observados, con lo que el valor del estadÃ-sticos en este caso serÃ-a igual a cero.

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**	**dist. Marginal satisf.**
MUCHO	**395**	**593**	**358**	**200**	**1546**
BASTANTE	**100**	**182**	**120**	**94**	**496**
POCO	**167**	**233**	**102**	**90**	**592**
NADA	**212**	**249**	**118**	**115**	**694**
dist. Marginal zonas	**874**	**1257**	**698**	**499**	**3328**

Para cada casilla calculamos:

(395− 406,01)2 (593− 583,93)2

−−−−−−−−−−−−−−−−−−−− = 0,2986 −−−−−−−−−−−−−−−−−−−− = 0,1409

406,01 583,93

Tabla de *chi− cuadrado*:

	ZONA 1	**ZONA 2**	**ZONA 3**	**ZONA 4**
MUCHO	**0,2986**	**0,1409**	**3,5127**	**4,3644**
BASTANTE	**7,0294**	**0,1523**	**2,452**	**5,1812**
POCO	**0,8549**	**0,3951**	**3,9562**	**0,0172**
NADA	**4,8533**	**0,6574**	**5,2169**	**1,1505**

La suma de todas estas cantidades proporciona el valor del estadÃ-stico *chi−cuadrado*, que en este caso es igual a 40,233. Dado que este valor no estÃ¡ cercano a cero, puede afirmarse que existe algÃºn tipo de asociaciÃ³n entre las dos variables. En este caso, se puede afirmar que el nivel de satisfacciÃ³n que el turista declara, tiene algÃºn tipo de relaciÃ³n con la zona de estancia de sus vacaciones.

El problema de este estadÃ-stico es que aunque tiene un lÃ-mite inferior (cero), **no proporciona un lÃ-mite superior que permita referirnos al grado de asociaciÃ³n existente.** Muchos estadÃ-sticos tiene una mayor utilidad si definen lÃ-mites (tanto superior como inferior) a los que tomar como referencia. En este caso, **el lÃmite inferior es cero e indica la ausencia total de dependencia (o asociaciÃ³n) entre las dos variables**, pero desconocemos el lÃmite superior que puede tomar y, por tanto, no podrÃamos afirmar si la asociaciÃ³n detectada es fuerte o dÃ©bil. Es decir que **sÃ³lo nos indica si existe asociaciÃ³n o no entre las dos variables**.

Ï‡2 â‰ˆ 0 ==> no hay asociaciÃ³n

Ï‡2 â‰ 0 ==> hay asociaciÃ³n

C DE CONTINGENCIA

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

El estadÃstico C de Contingencia es una alternativa al estadÃstico Chi− cuadrado. El coeficiente C toma su valor mÃnimo cuando Ï‡2 es cero, es decir, si entre las dos variables no existe asociaciÃ³n. El valor mÃ¡ximo del coeficiente, que se da cuando existe una asociaciÃ³n completa entre las variables, depende del nÃºmero de modalidades de las variables. El mÃ¡ximo que puede alcanzar el coeficiente es uno (correspondiente a dos variables con infinitos valores).

El coeficiente C de contingencia toma valores entre 0 y 1. Valores de C cercanos a 0 indican un grado de asociaciÃ³n pequeÃ±o. Valores cercanos a uno serÃan sÃntoma de asociaciÃ³n entre las variables.

0 â‰¤ C â‰¤ 1

El lÃmite mÃ¡ximo del coeficiente C es:

$$\sqrt{1-\frac{1}{min(I,\ J)}}$$

El mÃnimo (I, J) significa que hay que coger el nÃºmero que sea menor del total de columnas o filas. Si hay 3 columnas y 2 filas, escogemos el 2. **El lÃmite mÃ¡ximo de C sirve para poder comparar el valor de C con su valor mÃ¡ximo.**

Por ej: 0 â‰¤ C â‰¤ 0,739 y el coeficiente C tiene un valor de 0,534 calcularemos el porcentaje de C respecto a su valor mÃ¡ximo:

0,534

−−−−−− x 100 = 72,26% de C

0,739

Al poder comparar el valor de C entre 0 y 100 podremos decidir si el tipo de asociaciÃ³n es muy dÃ©bil, dÃ©bil, fuerte o muy fuerte.

EJEMPLO: Se ha pedido a 219 usuarios de un servicio si han quedado satisfechos o no, teniendo en cuenta si ya lo habÃan utilizado antes. Hacer la descripciÃ³n de datos de la siguiente tabla:

	REPETIDORES	NO REPETIDORES
SATISFECHOS	86	43
NO SATISFECHOS	36	54

219 = n o N

Los valores esperados serÃ¡n:

	REPETIDORES	NO REPETIDORES	DIST. MARGINAL SATIS.
SATISFECHOS	71,8630137	57,1369863	129
NO SATISFECHOS	50,1369863	39,8630137	90
DISTRIBUCIÃ“N MARGINAL REP.	122	97	219

La tabla de chi−cuadrado serÃ¡:

	REPETIDORES	NO REPETIDORES
SATISFECHOS	2,781	3,498
NO SATISFECHOS	3,986	5,014

La suma de estos valores, es decir, chi−cuadrado es 15,279.

Y el estadÃstico C de contingencia es:

15,279

−−−−−−−−−−−−−−− = 0,255

219 + 15,279

Su posible valor mÃ¡ximo es:

1

1 − −−−−−−−− = 0,707

2

El % de C serÃa:

0,255

−−−−−−− x 100 = 36%

0,707

Es decir, que el estadÃstico alcanza un 36% de su mÃ¡ximo, indicando que hay una asociaciÃ³n moderada entre la satisfacciÃ³n respecto al servicio y el haber sido usuario anterior.

TEMA 5. ANÃ LISIS CONJUNTO DE DOS VARIABLES CUANTITATIVAS

15/01/2009

5.1. <u>RELACIÃ“N ENTRE DOS VARIABLES CUANTITATIVAS. REPRESENTACIÃ“N GRÃ FICA.</u>

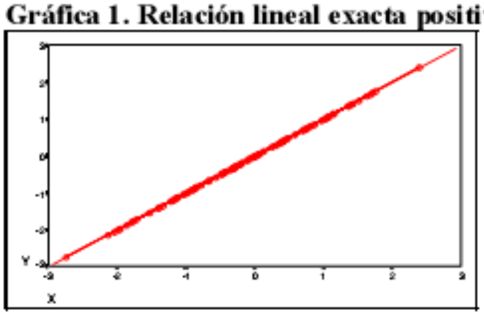
En el caso de las variables cuantitativas, estudiamos su relaciÃ³n, no su asociaciÃ³n.

RelaciÃ³n lineal: relaciÃ³n entre dos variables que puede representarse aproximadamente como una lÃ-nea recta.

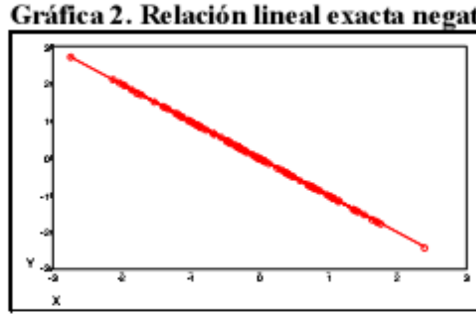
La asociaciÃ³n no implica causalidad

Dos tipos de asociaciÃ³n lineal: positiva y negativa.

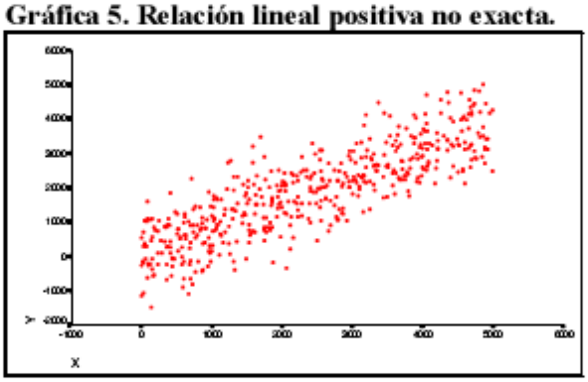
Gráfica 1. Relación lineal exacta positiva.



Gráfica 2. Relación lineal exacta negativa.



Gráfica 5. Relación lineal positiva no exacta.

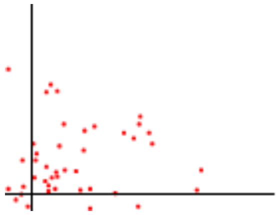


Estas variables siempre serÃ¡n unitarias, por lo que no habrÃ¡ frecuencias. Tendremos parejas de valores (X,Y) que se pueden representar en un eje de coordenadas.

Si **los puntos estÃ¡n exactamente sobre la recta**, diremos que la **relaciÃ³n lineal** es **exacta** y la ecuaciÃ³n de la recta serÃ-a:

Yi = a + b Xi

Si **los puntos estÃ¡n dispersos**, en forma de nube, se dice que **no existe relaciÃ³n lineal**:



Pueden existir otros tipos de relaciones entre las variables: parabÃ³licas, etc...

5.2.<u>MEDIDAS O ESTADÃ STICOS DE RELACIÃ“N LINEAL: COVARIANZA Y CORRELACIÃ“N</u>

COVARIANZA: mide la dispersiÃ³n entre las dos variables estudiadas, es decir, que **mide la dispersiÃ³n entre ambas variables.**

$$s_{XY}=\frac{\sum_{i=1}^{n}\left(X_i-\bar{x}\right)\left(Y_i-\bar{y}\right)}{n}$$

Con esta fÃ³rmula, necesitaremos calcular la media de cada variable y hacer la siguientes columnas:

X	Y	– Xi − x	– Yi −y	– – (Xi − x)* (Yi −y)
x1	y1			
x2	y2			
xn	yn			

âˆ‘=

Hay otra fÃ³rmula que es mÃ¡s fÃ¡cil y rÃ¡pida, y que ayuda a evitar errores para calcular la covarianza, que es:

$$=\frac{\sum_{i=1}^{n}X_iY_i}{n}-\overline{X}\,\overline{Y}$$

Para aplicar esta fÃ³rmula, nos basta calcular la media de cada variable y los productos de las variables X e Y.

X	Y	Xi *Yi
x1	y1	
x2	y2	
xn	yn	

n âˆ‘=

PROPIEDADES DE LA COVARIANZA:

- No tiene limite inferior ni superior (puede ser positiva o negativa)
- Depende de las unidades de medida de las variables, por lo que no se puede comparar rÃ¡pidamente con una cifra estÃ¡ndar que permita hablar de mucha o poca relaciÃ³n.
- El signo de la covarianza es importante porque determina la pendiente de la recta y nos indica el tipo de relaciÃ³n:

Covarianza positiva (sxy > 0) ==>AsociaciÃ³n lineal positiva.

Covarianza negativa (sxy < 0) ==>AsociaciÃ³n lineal negativa.

Covarianza nula (sxy = 0) ==>AsociaciÃ³n lineal inexistente.

Lo que no nos indica es cÃ³mo es la recta, es decir, no indica cuÃ¡n dispersos estÃ¡n los puntos respecto a la recta (IMP).

Otras Propiedades de la covarianza:

- Si se suma a la variable X una constante b y a la variable Y una constante c, la covarianza entre las dos nuevas variables transformadas serÃ¡ igual a la covarianza original.
- Si se multiplica la variable X por una constante b y la variable Y por una constante c, la covarianza entre las dos nuevas variables transformadas serÃ¡ igual a la covarianza original multiplicada por las constantes bc.
- La covarianza de una variable y una constante es cero.

COEFICIENTE DE CORRELACIÃ“N DE PEARSON (COEFICIENTE DE CORRELACIÃ“N LINEAL SIMPLE)

Es el cociente entre la covarianza y el producto de las desviaciones estÃ¡ndar de las 2 variables.

$$r_{XY} = \frac{s_{XY}}{s_X \, s_Y}$$

OJO, **en esta fÃ³rmula empleamos desviaciones tÃ-picas**, no varianzas. La fÃ³rmula desarrollada serÃ-a:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{\sum_{i=1}^{n} (X_i - \overline{x})(Y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{y})^2}} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\overline{x}^2} \sqrt{\sum_{i=1}^{n} Y_i^2 - n\overline{y}^2}}$$

O sea, que la fÃ³rmula mÃ¡s sencilla es (hay que tener en cuenta que n es el nÃºmero de observaciones):

$$\frac{\sum_{i=1}^{n} X_i Y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} Y_i^2 - n\bar{y}^2}}$$

X	Y	– x2	– y2	Xi *Yi	– n * x2	– n * y2	Xi 2
x1	y1						x12
x2	y2						x22
xn	yn						xn2

âˆ‘Xi2=

PROPIEDADES DEL COEFICIENTE DE CORRELACIÃ“N (examen):

- TendrÃ¡ el mismo signo que la covarianza
- **Su valor estÃ¡ entre −1 y 1**
- No depende de las unidades de medida de las variables

AsociaciÃ³n lineal positiva ==> Sxy> 0 ==> rxy>0

AsociaciÃ³n lineal negativa ==> Sxy< 0 ==> rxy <0

Ausencia de asociaciÃ³n lineal ==> Sxy= 0 ==> rxy =0

− El valor del coeficiente de correlaciÃ³n entre dos variables no se modifica si una (o ambas) variables se multiplica por una constante.

− El coeficiente de correlaciÃ³n toma valores en el intervalo −1 y 1. Los valores mÃ¡ximo y mÃ-nimo se alcanzan cuando se da una relaciÃ³n lineal exacta entre las dos variables, de tipo positivo o de tipo negativo, respectivamente. (Cuanto mÃ¡s cerca estÃ© de cero, peor).

− Valores del coeficiente prÃ³ximos a 1 indican la existencia de una asociaciÃ³n positiva fuerte entre las variables; valores cercanos a −1 indican la existencia de una asociaciÃ³n negativa fuerte entre las variables; **valores cercanos a cero seÃ±alan la ausencia de una asociaciÃ³n lineal**.

El coeficiente de correlaciÃ³n toma valores entre −1 y 1.

rxy = 1 AsociaciÃ³n lineal exacta de tipo positivo.

rxy = −1 AsociaciÃ³n lineal exacta de tipo negativo.

rxy = 0 Ausencia de asociaciÃ³n lineal.

Por ejemplo:

Si rxy = 0,95 ==> asociaciÃ³n lineal positiva fuerte

Si rxy = 0,6 ==> asociaciÃ³n lineal positiva moderada

Si rxy = 0,2 ==> asociaciÃ³n lineal positiva leve

Ejemplo:

i	X_i	Y_i	X_iY_i	X_i^2	Y_i^2
1	12	14,55	174,6	144	211,70
2	10	12,85	128,5	100	165,12
3	11	13,3	146,3	121	176,89
4	13	13,53	175,89	169	183,06
5	15	18,18	272,7	225	330,51
6	14	18,94	265,16	196	358,72
7	12	16,11	193,32	144	259,53
8	11	13,82	152,02	121	190,99
9	19	23,53	447,07	361	553,66
10	20	23,02	460,4	400	529,92
Suma	137	167,83	2415,96	1981	2960,12
Media	13,7	16,783			

$S_{XY} = 11{,}67$

$$S_X = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2} = \sqrt{\frac{1981}{10} - 13{,}7^2} = 3{,}23$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n} Y_i^2}{n} - \overline{Y}^2} = \sqrt{\frac{2960{,}12}{10} - 16{,}783^2} = 3{,}79$$

$$\boxed{r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{11{,}67}{3{,}23 \cdot 3{,}79} = 0{,}95}$$

Por tanto, existe ***asociación positiva muy fuerte*** entre ambas variables.

COEFICIENTE DE CORRELACIÃ“N DE RANGOS DE SPEARMAN (sÃ³lo para variables cuantitativas continuas)

Los rangos de cada variable son los valores que se les da, **teniendo en cuenta el orden que tendrÃ-an los valores de las variables, si las ordenÃ¡ramos de menor a mayor**. Con ello se obtiene una medida de grado de relaciÃ³n en los posicionamientos de las observaciones de las dos variables. Tiene las mismas caracterÃ-sticas que el coeficiente de Pearson.

Si por ejemplo, tenemos estas dos variables:

X	Y
19	73
55	110
110	9
3	230
220	150

n = 5

Y ordenamos de menor a mayor cada una de ellas:

X	Y	Rgx	Rgy
19	73	2	2
55	110	3	3
110	9	4	1
3	230	1	5
220	150	5	4

n = 5

En el caso de que se repitan valores, no se les asignarÃ-an rangos consecutivos, sino que harÃ-amos el promedio de los rangos que les corresponderÃ-an y les asignarÃ-amos el mismo rango a todos los valores repetidos y los rangos promediados ya no se asignarÃ-an a ningÃºn otro valor:

X	Rgx
7	3
9	5
4	2
9	5
3	1
9	5
10	7
12	8

El valor 9 se repite 3 veces y ocuparÃ-a los rangos 4, 5 y 6, por lo tanto el nuevo rango se calcularÃ-a:

4 + 5 + 6 15

−−−−−−−−−−−− = −−−−−−− = 5

- 3

La fÃ³rmula para calcular el coeficiente de correlaciÃ³n de los rangos de las variables o correlaciÃ³n de Spearman es:

n

6 * âˆ‘ di2

i = 1

RSpearman = Ï = 1 − −−−−−−−−−−−−−−

n3 − n

Donde n es el nÃºmero de observaciones y di es la diferencia entre los rangos:

di = Rango(Xi) − Rango (Yi)

SegÃºn el ejemplo anterior, calcularÃ-amos:

X	Y	Rgx	Rgy	di	di2
19	73	2	2	0	0
55	110	3	3	0	0
110	9	4	1	3	9
3	230	1	5	−4	16
220	150	5	4	1	1

n = 5 âˆ‘= 26

6 * 26 156

Ï = 1 − −−−−−−−−−− = 1 − −−−−−−− = −0,3

53 − 5 120

Para estas variables, el coeficiente de correlaciÃ³n de Pearson sale − 0,153, pero hay que tener en cuenta que ambos coeficientes no tienen por quÃ© tener valores parecidos.

EJEMPLO: Calcular la covarianza segÃºn este cuadro de datos:

X	Y	– Xi − x	– Yi −y	– – (Xi − x)* (Yi −y)
x1	y1	−2	−6	12
x2	y2	−1	−3	3
x3	y3	−4	−2	8
x4	y4	0	−5	0
x5	y5	2	−4	−8
x6	y6	3	6	18
x7	y7	2	4	8

âˆ‘= 41

41

$$s_{XY} = \frac{\sum_{i=1}^{n} (X_i - \bar{x})(Y_i - \bar{y})}{n}$$

sxy = −−−−−−− = 5,857

7

La covarianza es positiva y por lo tanto, podrÃa existir relaciÃ³n lineal positiva entre estas dos variables. **Para saber si hay asociaciÃ³n lineal y de quÃ© tipo, tendrÃamos que calcular la correlaciÃ³n de Pearson**.

EJEMPLO: si la covarianza entre dos variables es 0,558, la varianza de X es 0,5 y la de Y es 0,75, calcular el coeficiente de correlaciÃ³n de Pearson.

sXY 0,558

rXY = −−−−−−−−−− = −−−−−−−−−−−−− = 0,907

sX * sY 0,5 * 0,75

OJO, como la varianza es sx2 hay que hacer la raÃ-z cuadrada, para calcular las desviaciones estÃ¡ndar.

El comentario sobre el resultado serÃ-a que hay una fuerte correlaciÃ³n lineal positiva entre las variables X e Y.

(ver otros ejemplo del libro, tema 5)

5.3. RELACIÃ“N DE CAUSALIDAD O DE DEPENDENCIA

La asociaciÃ³n entre dos variables no implica causalidad o dependencia entre ellas. Proponer una relaciÃ³n de dependencia (suponer que una es dependiente de la otra) supone aÃ±adir una hipÃ³tesis adicional. Normalmente llamamos Y a la variable dependiente y calculamos sus valores en funciÃ³n de X, que es la variable independiente. No hay que confundir asociaciÃ³n con dependencia.

Cuando la variable X toma valores, ocasiona variaciones en la variable Y, por lo tanto, Y depende de X. Se estudian los valores de Y con los diferentes valores de X con la funciÃ³n de la recta. **Ajustar una recta** (**examen**)a una nube de puntos supone determinar los valores de a y b. Esta recta se llama **recta de regresiÃ³n simple**.

Si creemos que la variable X es causa de la variable Y, escribimos la relaciÃ³n como:

Yi = a+bX, i = 1, 2, , n

Y: variable endÃ³gena, variable dependiente o variable a explicar.

X: variable exÃ³gena, variable independiente o variable explicativa.

La recta Yi = a + bX se conoce como **recta de regresiÃ³n simple**.

a: es el valor que tomarÃ-a la variable Y cuando el valor de X es cero. Se conoce como **tÃ©rmino independiente de la regresiÃ³n.**

b: mide el impacto que una variaciÃ³n en una unidad de la variable X tiene sobre la variable Y. Se conoce como la **pendiente de la recta**.

Pendiente positiva Pendiente negativa

b

b

x1 x2 x2 x1

Las observaciones se pueden situar mÃ¡s o menos sobre la recta, aunque difÃ-cilmente lo harÃ¡n de manera perfecta. Cuanto mÃ¡s cerca estÃ©n las observaciones de la recta, mejor, pero puede que sÃ³lo haya algunos puntos sobre la recta, o incluso que no haya ninguno. La distancia del punto a la recta se llama **error de**

ajuste. El punto serÃ-a el valor observado y el punto sobre la recta serÃ-a el valor ajustado.

El **ajuste de una recta** puede expresarse como:

Yi = a + bXi + ei

Error de ajuste: diferencia entre los valores observados y los valores ajustados:

ei = Yi − a − bXi

En este marco de causalidad resulta necesario disponer no sÃ³lo de una medida de grado de asociaciÃ³n (el coeficiente de correlaciÃ³n lineal), sino ademÃ¡s de los valores numÃ©ricos de *a* y *b*. El valor de *a* es el valor que tomarÃ-a la variable Y cuando el valor de X es cero. El valor de *b* mide el impacto que una variaciÃ³n en una unidad de la variable X tiene sobre la variable Y y es la pendiente de la recta:

âˆ†Yi

b = −−−−−−−−−

âˆ†Xi

5.4. <u>OBTENCIÃ“N DE *a* Y *b* POR MÃ NIMOS CUADRADOS ORDINARIOS (M.C.O)</u>

El mÃ-nimo error global de un ajuste se obtiene utilizando los valores de *a* y *b* que minimizan la siguiente expresiÃ³n:

n n

MÃ-nimoâˆ‘ei 2 = âˆ‘(Yi − a − bXi)2

i = 1 i = 1

Los valores de *a* y *b* deben elegirse para que se alcance el mÃ-nimo de la suma de cuadrados de los errores. Este principio se denomina de **mÃ-nimos cuadrados ordinarios (mco).** Las fÃ³rmulas para *a* y *b* que se obtienen como resultados son las siguientes:

_ _

a = y − bx

n _ _

âˆ‘(Xi − x) * (Yi − y)

i = 1

b = −−−−−−−−−−−−−−−−−−−−−−−−−−

n _

âˆ‘(Xi − x) 2

i = 1

- Es conveniente calcular primero la *b*
- Es importante decidir cuÃ¡l va a ser la variable independiente (X)

Una fÃ³rmula mÃ¡s sencilla serÃa (examen):

n _ _

â^‘(Xi * Yi) −(n* x * y)

i = 1 sXY

b = −−−−−−−−−−−−−−−−−−−−−−−−−−−−−−− = −−−−−−−−−−

n _ sX2

â^‘ Xi2 − (n* x2)

i = 1

Con esta fÃ³rmula, las columnas que debemos calcular son:

X	Y	Xi * Yi	Xi2

EJEMPLO:

n=10

X	Y
17	−21
20	−20
23	−23
24	−24
24	−30
25	−31
25	−25
26	−18
28	−27
30	−29

Calculamos las medias de las variables y despuÃ©s x*y y x2:

_ 242 242 _ −248 −248

x = −−−−−−−−−−− = −−−−−−−−−− = 24,2 y = −−−−−−−−−−− = −−−−−−−−−− = −24,8

n 10 n 10

	X	Y	X*Y	X2
	17	−21	−357	289
	20	−20	−400	400
	23	−23	−529	529
	24	−24	−576	576
	24	−30	−720	576
	25	−31	−775	625
	25	−25	−625	625
	26	−18	−468	676
	28	−27	−756	784
	30	−29	−870	900
sumatorios	242	−248	−6076	5980

Ahora, ya se puede calcular *a* y *b* de la recta:

n _ _

âˆ‘(Xi * Yi) −(n* x * y)

i = 1 (−6076) − (10*24,2*(−24,8))

b = −−−−−−−−−−−−−−−−−−−−−−−−−−−−−− = −−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−− = −0,6 (pendiente neg.)

n _ 5980 − (10* (24,2)2)

âˆ‘ Xi2 − (n* x2)

i = 1

_ _

a = y − bx ==> a = (−24,8) − ((−0,6)*24,2) ==> **a = −10,2**

El ajuste de la recta serÃa:

Yi = a + bXi Y = −10,2 + (−0,6)X Y = −10,2 −0,6X

La pendiente es negativa y un incremento de X disminuirÃa en 0,6 unidades el valor de Y. Cuando X aumenta, Y disminuye y cuando X disminuye, Y aumenta.

5.5. COEFICIENTE DE DETERMINACIÃ“N: BONDAD DEL AJUSTE LINEAL

Este estadÃstico mide hasta quÃ© punto la variable exÃ³gena X de la regresiÃ³n explica o determina la evoluciÃ³n de la variable endÃ³gena Y.

Entre el coeficiente de correlaciÃ³n lineal (coeficiente de correlaciÃ³n de Pearson) rxy y el coeficiente de determinaciÃ³n R2, se da la siguiente relaciÃ³n:

$$r_{XY} = \frac{s_{XY}}{s_X \, s_Y}$$

$$\frac{\sum_{i=1}^{n} X_i Y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} Y_i^2 - n\bar{y}^2}}$$

El coeficiente de determinaciÃ³n toma valores entre 0 y 1. Valores cercanos a cero indican que el ajuste es insuficiente, valores cercanos a uno se interpretarÃ¡n como un buen ajuste del modelo.

$$\boxed{0 \leq R^2 \leq 1}$$

El coeficiente de determinaciÃ³n mide la bondad del ajuste de una recta a una nube de puntos. El coeficiente de correlaciÃ³n lineal (coeficiente de correlaciÃ³n Pearson) mide el grado de asociaciÃ³n lineal entre dos variables.

Un coeficiente de correlaciÃ³n lineal igual a 1 o −1 indica que los puntos estÃ¡n situados sobre una recta (con pendiente positiva o negativa respectivamente) y esta situaciÃ³n se reflejarÃa en el coeficiente de determinaciÃ³n con un valor igual a 1 (ajuste perfecto de la recta). Si el coeficiente lineal o el coeficiente de determinaciÃ³n son iguales a cero, no existirÃ¡ una relaciÃ³n lineal entre las dos variables.

- Si no existe relaciÃ³n lineal entre X e Y: r XY = 0 y R2 = 0

(valores de 0 o cercanos a 0)

- Si existe una relaciÃ³n lineal positiva exacta entre X e Y: r XY = 1 y R2 = 1

(valores de 1 o cercanos a 1)

- Si existe una relaciÃ³n lineal negativa exacta entre X e Y: r XY = −1 y R2 = 1

(valores de 1 o cercanos a 1)

Por ejemplo, si los valores son de 0,5 dirÃamos que la relaciÃ³n lineal es moderada.

HACER EJEMPLOS DEL LIBRO. TEMA 6.

EJEMPLO: A una muestra de ciudadanos se les ha solicitado su opiniÃ³n sobre la actual situaciÃ³n turÃstica en su regiÃ³n (entre 1 y 5: 1, mala; 5 excelente) y el efecto que creen que tendrÃ¡ un impuesto turÃstico (entre 1 y 5: 1, la situaciÃ³n empeorarÃ¡ mucho; 5 la situaciÃ³n mejorarÃ¡ mucho). En este caso, el efecto del impuesto en la situaciÃ³n turÃstica es la variable dependiente.

n	âˆ‘xi	âˆ‘yi	âˆ‘xi* yi	âˆ‘xi2	âˆ‘yi2
20	57	49	131	181	127

_ 57 57 _ 49 49

x = −−−−−−−−−−− = −−−−−−−−−− = 2,85 y = −−−−−−−−−−− = −−−−−−−−−− = 2,45

n 20 n 20

Ahora, ya se puede calcular a y b de la recta:

n _ _

âˆ‘(Xi *Yi) −(n* x * y)

i = 1 131− (20* 2,85* 2,45) −8,65

b = −−−−−−−−−−−−−−−−−−−−−−−−−−−− = −−−−−−−−−−−−−−−−−−−−−−−−−−−− = −−−−−− = −0,46 (pendiente neg.)

n _ 1812 − (20* 2,852) 18,55

âˆ‘ Xi2 − (n* x2)

i = 1

_ _

a = y − bx ==> a = 2,45 − ((−0,46)*2,85) ==> **a = 3,778**

La recta de regresiÃ³n serÃ-a:

Yi = a + bXi Y = 3,778 + (−0,46)X Y = 3,778 −0,46X

Cuanto mÃ¡s valoramos la situaciÃ³n turÃ-stica (X) menos se valora el efecto del impuesto (Y). La recta tiene pendiente negativa, por lo tanto, cuanto mayor es X, menor es Y.

Ahora calculamos rxy:

$$\frac{\sum_{i=1}^{n} X_i Y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} Y_i^2 - n\bar{y}^2}}$$

131 − (20*2,85* 2,45) −8,65

rxy = −− = −−−−−−−−−−−−− = −0,76

4,30 * 2,63

181− (20*2,852) * 127− (20*2,452)

R2 = rxy2 R2 = −0,762 R2 = 0,5776 â‰ˆ 0,58 ==> 0,58 * 100 = 58%

El resultado de R2 indica que la recta es entre moderada y buena. La percepciÃ³n de la situaciÃ³n turÃstica explica bastante bien la valoraciÃ³n del efecto del impuesto, observÃ¡ndose que cuanto mejor se considera la situaciÃ³n actual, peor valoraciÃ³n se hace del impuesto.

EJEMPLOS IMPORTANTES:

La demanda turÃstica (gasto turÃstico, nÃºmero de turistas, etc) **es siempre una variable dependiente de las siguientes variables independientes**:

- **Con pendiente negativa**: los precios del paquete turÃstico y Precios relativos paÃs destino/ paÃs emisor. Cuando estos precios suben, el gasto turÃstico y nÃºmero de turistas bajan.
- **Con pendiente positiva**: tipo de cambio de moneda, Precios relativos de los paÃses competidores (comparando precios turÃsticos), Precios de los paquetes de los competidores y Renta del paÃs emisor. Cuando estas variables suben, el gasto turÃstico y el nÃºmero de turistas suben. Cuando estas variables bajan, el gasto turÃstico y el nÃºmero de turistas tambiÃ©n bajan.

Demanda i = a + b* Renta i

Demanda i = a + b* Precio del Paquete i

Â¡Â¡Â¡OJO PARA LAS FÃ“RMULAS!!!: NO ES LO MISMO âˆ'Xi2 QUE (âˆ' Xi)2

Lo que aparece en las fÃ³rmulas es âˆ'Xi2 ES:

Xi	**Xi2**
X1	**X12**
X2	**X22**
Xn	**Xn2**

âˆ'Xi2=

El fallo es hacer (âˆ' Xi)2:

Xi
X1
X2
Xn

âˆ'(Xi) = ...y despuÃ©s al cuadrado

26/02/2009

TEMA 6. LA EVOLUCIÃ“N TEMPORAL DE UNA VARIABLE CUANTITATIVA

Cuando se estudia la evoluciÃ³n temporal de una variable cuantitativa, se estudian a la vez dos variables. Una de las variables es el tiempo, que siempre es cuantitativo y la otra es cualquier variable cuantitativa que cambia a lo largo del tiempo.

6.1. LA PERSPECTIVA TEMPORAL EN EL ANÃ LISIS DE UNA VARIABLE.

Lo que interesa es **observar los valores** de una determinada variable **a travÃ©s del tiempo, para estudiar su evoluciÃ³n**. Es decir, que lo que se quiere es conocer quÃ© ha sucedido en el pasado, quÃ© sucede en el presente y **hacer predicciones de lo que puede suceder en el futuro**, siendo Ã©sta Ãºltima la opciÃ³n mÃ¡s interesante.

Se trata de estudiar los valores de la variable en diferentes momentos del tiempo.

VARIABLE TEMPORAL: tambiÃ©n llamadas series temporales, cronolÃ³gicas o histÃ³ricas. En este caso, estudiamos variables temporales cuantitativas.

TIEMPO: elemento fundamental en una serie temporal. Es una variable discreta que siempre toma valores consecutivos y variando de uno en uno (de dÃa en dÃa, de mes en mes, de trimestre en trimestre, de aÃ±o en aÃ±o...)

Por ejemplo, si tenemos la siguiente serie mensual:

	ENERO
	FEBRERO
	MARZO
	ABRIL
2001	MAYO
	JUNIO
	JULIO
	AGOSTO
	SEPTIEMBRE
	OCTUBRE
	NOVIEMBRE
	DICIEMBRE
	ENERO
	FEBRERO
	MARZO
	ABRIL
	MAYO
2002	JUNIO
	JULIO
	AGOSTO
	SEPTIEMBRE
	OCTUBRE
	NOVIEMBRE
	DICIEMBRE

La variable Tiempo: enero, febrero, marzo... no nos sirve para trabajar con ella, por lo tanto, se le asignan nÃºmeros consecutivos, que serÃ¡n los valores de la variable tiempo (t):

ENERO	1
FEBRERO	2
MARZO	3

	ABRIL	4
2001	MAYO	5
	JUNIO	6
	JULIO	7
	AGOSTO	8
	SEPTIEMBRE	9
	OCTUBRE	10
	NOVIEMBRE	11
	DICIEMBRE	12
	ENERO	13
	FEBRERO	14
	MARZO	15
	ABRIL	16
	MAYO	17
2002	JUNIO	18
	JULIO	19
	AGOSTO	20
	SEPTIEMBRE	21
	OCTUBRE	22
	NOVIEMBRE	23
	DICIEMBRE	24

Si se trata de una serie anual:

1900	1
1901	2
1902	3
1903	4
1904	5
1905	6
1906	7
1907	8
1908	9

Y se harÃ-a lo mismo si tuviÃ©ramos otras medidas de tiempo, como horas, semanas, etc. En la tabla, aÃ±adirÃ-amos otra columna para los valores de la variable Xi, por ejemplo temperaturas.

Siempre se tienen en cuenta dos variables, pero una de ellas Yi, se estudia asociada a la otra (tiempo). La variable dependiente serÃ¡ Yi, y la independiente serÃ¡ el tiempo.

Los grÃ¡ficos de las variables temporales aportan mucha informaciÃ³n. En este estudio, la variable Xi se pone en el eje vertical y la variable tiempo (t) se pone en el eje horizontal:

6.2. COMPONENTES SISTEMÃ TICAS DE UNA SERIE TEMPORAL

- Tendencia

- Estacionalidad
- Ciclo

No todas las series temporales siguen el mismo patrÃ³n de comportamiento, pero se pueden sistematizar algunas caracterÃ-sticas de su comportamiento.

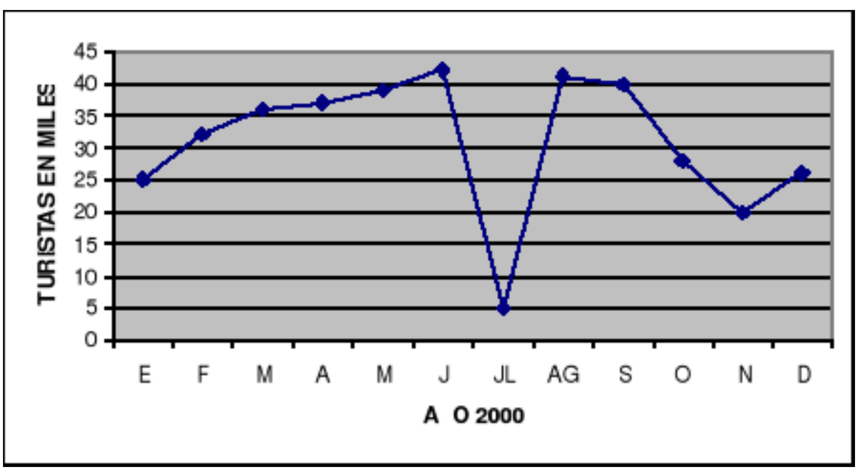
Tendencia: es la evoluciÃ³n GLOBAL de una serie temporal dentro del perÃ-odo que se observa. Es la evoluciÃ³n de la serie en el largo plazo. En turismo, muchas series tienen un comportamiento tendencia claramente creciente. La tendencia puede ser creciente, decreciente o constante (estancada). En el grÃ¡fico anterior, vemos que crece en los meses de verano y despuÃ©s decrece. Cuando se estudia la tendencia, hay que responder a dos preguntas: Â¿hay tendencia?, y si la hay, Â¿de quÃ© tipo es la tendencia?.

Estacionalidad o Componente estacional: representa el conjunto de pautas que ocurren de modo muy similar, <u>en los mismos perÃ-odos</u> (**meses, trimestres**, etc.), <u>repitiÃ©ndose sistemÃ¡ticamente cada aÃ±o</u>. (OJO: no tiene sentido hablar de estacionalidad si son datos anuales). Para poder estudiar la estacionalidad necesitaremos los datos de dos aÃ±os, como mÃ-nimo. En general, el comportamiento estacional de las series temporales se debe al clima, las costumbres vacacionales, sociales y religiosas y a los efectos del calendario. Cuando se estudia la estacionalidad, hay que responder a dos preguntas: Â¿hay estacionalidad? y si la hay, Â¿de quÃ© tipo es la estacionalidad?.

Ciclo o componente cÃ-clica: representa las oscilaciones que <u>se repiten cada cierto nÃºmero de aÃ±os</u>. SÃ³lo se busca ciclo en las **series anuales**. Siempre consideraremos que si existe ciclo, se encuentra incluido en la tendencia. Cuando se estudia el ciclo, hay que responder si hay ciclo o no lo hay. (OJO: no tiene sentido hablar de ciclo si NO son datos anuales).

6.3. <u>COMPONENTE NO SISTEMÃ TICA DE UNA SERIE TEMPORAL</u>

Irregularidad o componente irregular: incluye las variaciones aleatorias de la serie. Esta componente refleja los movimientos de muy corto plazo. Es decir, todo lo que se produce entre observaciones consecutivas, pero que no se reproduce de forma similar en los mismos perÃ-odos en aÃ±os diferentes:



En este grÃ¡fico, vemos una irregularidad en el mes de julio, es un valor que se dispara respecto a los demÃ¡s.

EJEMPLO: Tenemos el siguiente grÃ¡fico que representa las ventas de un restaurante a lo largo de cada trimestre entre los aÃ±os 2000 y 2002 (examen):

- La tendencia es creciente. (hay que mirarlo en general, sube durante los tres primeros trimestres y baja en el cuarto durante los 3 aÃ±os).

- Estacionalidad: el tercer trimestre de los tres aÃ±os tiene ventas muy superiores.
- Ciclos: sÃ³lo hay informaciÃ³n de tres aÃ±os y no basta para establecer si hay ciclo. Lo ideal es tener la informaciÃ³n de 6 o 7 aÃ±os.
- Irregularidad: no se observan irregularidades en este grÃ¡fico.

6.4. ANÃ LISIS DE LA TENDENCIA Y DE ESTACIONALIDAD

El hecho de que en una misma serie temporal haya varias componentes (tendencia, estacionalidad e irregularidad), hace que el anÃ¡lisis de cada una de estas componentes por separado, resulte muy complicado.

Por lo tanto, **es conveniente aislar cada una de estas componentes.** En el Ã¡mbito turÃ-stico es muy frecuente que las series combinen estacionalidad, tendencia e irregularidad, aunque algunas series pueden no tener estacionalidad.

Debido a eso, se distingue entre series con tendencia y estacionalidad y series sÃ³lo con tendencia. Tendencia y estacionalidad son las componentes mÃ¡s importantes.

SERIES TEMPORALES CON TENDENCIA E IRREGULARIDAD (SIN ESTACIONALIDAD)

En el caso de series que sÃ³lo tienen tendencia e irregularidad, **aislar la tendencia significa eliminar la irregularidad.**

Si una serie no tiene estacionalidad y por tanto, muestra un crecimiento o un decrecimiento en el tiempo, se trata de determinar la relaciÃ³n existente entre la propia serie y el tiempo, es decir, **ajustar la tendencia a una recta**. La recta de regresiÃ³n serÃ¡:

Vt = a + b*t

Es decir, la recta de las ventas en funciÃ³n del tiempo serÃ¡ igual a la variable *a* mÃ¡s la variable *b* multiplicada por el tiempo. **Siempre hay que calcular la bondad del ajuste lineal** (R2). La *b* indica la pendiente de la recta y tambiÃ©n el tipo de tendencia:

b positiva ==> pendiente positiva ==> tendencia creciente

b negativa ==> pendiente negativa ==> tendencia decreciente

Es importante tener claro cuÃ¡l es la variable dependiente (por ej, las ventas de un restaurante) y cuÃ¡l es la variable independiente (t). Normalmente, la variable independiente se llama X, pero en este caso, la llamaremos t de tiempo y en los grÃ¡ficos se representa en el eje horizontal. La variable dependiente suele llamarse Y, pero en este caso la llamaremos en funciÃ³n del dato que nos den, por ej V de ventas del restaurante y en los grÃ¡ficos se representarÃ¡ en el eje vertical.

SERIES TEMPORALES CON ESTACIONALIDAD Y CON TENDENCIA CONSTANTE O CON UNA LIGERA TENDENCIA LINEAL

En este caso, **aislar la tendencia significa eliminar la irregularidad y la estacionalidad.**

El **coeficiente estacional** es la media aritmÃ©tica (diaria, semanal, mensual, trimestral..) menos la media aritmÃ©tica global**:**

coeficiente estacional = media aritmÃ©tica − media aritmÃ©tica global

EJEMPLO: Tenemos los siguientes datos de ventas de 4 aÃ±os (necesitamos informaciÃ³n de dos aÃ±os como mÃ-nimo):

	trimestre	ventas	
1997	I	20	
		II	107
		III	136
		IV	45
	1998	I	23
		II	109
		III	144
		IV	51
	1999	I	22
		II	115
		III	145
		IV	53
	2000	I	19
		II	115
		III	143
		IV	52

Confeccionamos una nueva tabla para poder hacer las operaciones. En esta tabla, ya se observa la estacionalidad y el tercer trimestre es el mÃ¡s alto en los cuatro aÃ±os:

	1997	1998	1999	2000
I	20	23	22	19
II	107	109	115	115
III	136	144	145	143
IV	45	51	53	52

—

1Âº Calculamos la **media aritmÃ©tica trimestral Vt** (porque los datos son trimestrales):

	1997	**1998**	**1999**	**2000**	**media Vt**
I	20	23	22	19	21
II	107	109	115	115	111,5
III	136	144	145	143	142
IV	45	51	53	52	50,25

—

2Âº Calculamos la **media global**. Se puede hacer sumando la **Vt** y se hace la media o bien, se suman todos lo valores y se divide entre n:

21 + 111,5 + 142 + 50, 25

———————————————————————————— = 81,1875 es la media global

4

3Âº Calculamos el **coeficiente estacional**, es decir cada media trimestral menos la global (por este orden, ya que el signo del resultado es importante):

	1997	1998	1999	2000	media Vt	coef. Estacional
I	20	23	22	19	21	−60,1875
II	107	109	115	115	111,5	30,3125
III	136	144	145	143	142	60,8125
IV	45	51	53	52	50,25	−30,9375
				media global	**81,1875**	

La temporada mÃ¡s baja en ventas es la del coeficiente negativo mÃ¡s pequeÃ±o (−60,1875).

La temporada mÃ¡s alta en ventas es la del coeficiente positivo mÃ¡s grande (60,8125).

4Âº **Desestacionalizar**: eliminar la estacionalidad de la serie. A cada valor (en este caso, trimestral) hay que restarle el coeficiente estacional correspondiente:

- A las ventas del primer trimestre, restarle el coeficiente estacional del primer trimestre.
- A las ventas del segundo trimestre, restarle el coeficiente estacional del segundo trimestre.
- A las ventas del tercer trimestre, restarle el coeficiente estacional del tercer trimestre.
- A las ventas del cuarto trimestre, restarle el coeficiente estacional del cuarto trimestre.

trimestre	ventas	Desestacionalizar
I	20	80,1875
II	107	76,6875
III	136	75,1875
IV	45	75,9375
I	23	83,1875
II	109	78,6875
III	144	83,1875
IV	51	81,9375
I	22	82,1875
II	115	84,6875
III	145	84,1875
IV	53	83,9375
I	19	79,1875
II	115	84,6875
III	143	82,1875
IV	52	82,9375

6.5. TASAS DE VARIACIÃ“N O DE CRECIMIENTO

Las tasas de variaciÃ³n implican siempre una proporciÃ³n y sirven para establecer comparaciones temporales; esto implica la comparaciÃ³n de los valores de una variable en diferentes momentos del tiempo. **Las tasas**

siempre son cocientes o proporciones y <u>deben expresarse en %.</u>

Ejemplos:

Plazas ocupadas

Tasa de ocupaciÃ³n hotelera = −−−−−−−−−−−−−−−−−−−−−−−−−−−− X 100

Plazas ofertadas

NÂº de ocupados

Tasa de paro = −−−−−−−−−−−−−−−−−−−−−−−−−−−− X 100

PoblaciÃ³n activa

Tasa de variaciÃ³n de un perÃodo respecto al inmediatamente anterior: La tasa de variaciÃ³n porcentual o relativa expresa, en tÃ©rminos relativos, <u>el incremento (o decremento) de un valor respecto a su valor previo</u> y se calcula:

$$\boxed{T_1 = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \cdot 100}$$

Por ejemplo, la tasa de variaciÃ³n del IV trimestre respecto al III trimestre de 1997 de la tabla anterior, serÃa:

Ventas IV trim − Ventas III trimestre

TIV = −−− X 100

Ventas III trimestre

45 − 136

TIV = −−−−−−−−−−−−−−−−− X 100 = −66,91%

136

El resultado indica un decrecimiento de casi el 67%

Tasa de variaciÃ³n de un perÃodo respecto al mismo periodo del aÃ±o anterior: tasa interanual: expresa la variaciÃ³n de una valor de la variable respecto al valor del aÃ±o anterior.

La tasa interanual **para una serie mensual** es:

$$\boxed{T_{12} = \frac{Y_t - Y_{t-12}}{Y_{t-12}} \cdot 100}$$

La tasa interanual **para una serie trimestral** es:

$$\boxed{T_4 = \frac{Y_t - Y_{t-4}}{Y_{t-4}} \cdot 100}$$

Por ejemplo, la variaciÃ³n interanual del primer trimestre de 1999, respecto al del aÃ±o anterior de los datos del ejemplo anterior, serÃ-a:

Ventas I trim1999 − Ventas I trim1998

TI, 1999 = −− X 100

Ventas I trim1998

22 − 23

TI, 1999 = −−−−−−−−−−−−−−− X 100 = − 4,5%

23

El resultado indica un decrecimiento interanual del 4,5%.

6.6. NÃšMEROS Ã NDICE

(A PARTIR DE LAS TASAS DE VARIACIÃ“N)

Para facilitar la comparaciÃ³n de una variable en diferentes momentos del tiempo, hace falta analizar la evoluciÃ³n de cada variable en referencia a un mismo momento dado.

Un **nÃºmero Ã-ndice** representa la variaciÃ³n (creciente o decreciente) que experimenta en cada perÃ-odo una variable con referencia a un instante temporal, que se considera como referencia, llamado **perÃ-odo base** del Ã-ndice.

Un **Ã-ndice simple** es el cociente entre las dos magnitudes que se quieren comparar.

De forma general, si Yt es una serie temporal cualquiera, el Ã-ndice de la variable en el perÃ-odo t con perÃ-odo base 0, puede expresarse de la siguiente forma:

$$\boxed{I_{t,0} = \frac{\text{Valor de Y en el período } t}{\text{Valor de Y en el período base}} \cdot 100 = \frac{Y_t}{Y_0} \cdot 100}$$

Un **Ã-ndice de precios** calculado para el perÃ-odo base toma el valor de 100 (examen):

Y0

I 0,0 = −−−−−−−−−−−− X 100 = 100

Y0

Por lo tanto, 100 es el nÃºmero de referencia. Si calculamos un Ã-ndice de precios y el resultado es 85, significa que ha habido una disminuciÃ³n del 15% (85 − 100 = −15). Si el resultado el 110, significa que ha aumentado un 10% (110 − 100 = 10).

Los **Ã-ndices compuestos** sintetizan, en un solo Ã-ndice, la evoluciÃ³n conjunta de varias magnitudes. Se calculan a base de integrar en un solo conjunto de Ã-ndices simples, consiguiendo un indicador agregado a partir de Ã-ndices individuales.

Los **Ã-ndices de precios compuestos** agregan en un solo Ã-ndice los precios de k productos (examen):

$$\boxed{\frac{1}{k}\sum_{i=1}^{k}\frac{\text{Precio en año t del producto i}}{\text{Precio en año base del producto i}}\cdot 100}$$

Ã ndices Compuestos Ponderados. Se da una importancia diferente a los precios de distintos productos. Se calculan las medias ponderadas.

Ejemplo**. Ã ndices de Precios al Consumo**. Los Ã-ndices de precios al consumo **son Ã-ndices de precios compuestos y ponderados** que se calculan muy frecuentemente a travÃ©s de estas dos alternativas de cÃ¡lculo (examen):

Ã ndice de Laspeyres (**L**): El Ã ndice de Laspeyres usa como ponderaciÃ³n el consumo de los productos en el aÃ±o base (**qi0**).

$$\boxed{L_{t,0}=\frac{\sum_{i=1}^{k}P_{it}\cdot q_{i0}}{\sum_{i=1}^{k}P_{i0}\cdot q_{i0}}\cdot 100}$$

Ã ndice de Paasche (**P**): El Ã ndice de Paasche utiliza los consumos del aÃ±o para el que se calcula el Ã-ndice (**qit**).

$$\boxed{P_{t,0}=\frac{\sum_{i=1}^{k}P_{it}\cdot q_{it}}{\sum_{i=1}^{k}P_{i0}\cdot q_{it}}\cdot 100}$$

Pit y Pi0 son los precios y qit y qi0 las cantidades consumidas de cada producto en el aÃ±o actual y en el aÃ±o base, respectivamente. El Ã ndice de Laspeyres usa como ponderaciÃ³n el consumo de los productos en el aÃ±o base. El Ã ndice de Paasche utiliza los consumos del aÃ±o para el que se calcula el Ã-ndice.

Ã ndice de Fischer: es la media geomÃ©trica de los Ã-ndices anteriores.

Ft,0 = L t,0 X P t,0

Los Ãndices de Fisher y Paasche son muy costosos de calcular. **En EspaÃ±a, se utiliza en Ãndice de Laspeyres para calcular el IPC.**

Ã NDICE DE PRECIOS AL CONSUMO (ESPAÃ‘A)

El INE (Instituto Nacional del EstadÃstica) elabora mensualmente el IPC de EspaÃ±a. Con este indicador se quiere medir la variaciÃ³n de los precios de los bienes y servicios que forman la cesta de la compra de la poblaciÃ³n residente en viviendas familiares espaÃ±olas. Esta cesta de la compra es el conjunto de artÃculos que se consumen de forma mayoritaria en una familia durante todo un aÃ±o. El aÃ±o base actual es el 2001 y el anterior fue 1992.

Los productos consumidos se han clasificado en 12 grupos integrados por un total de 471 artÃculos. Las cantidades de cada uno de esos artÃculos consumidas por las familias, proporcionan sus ponderaciones.

De estos 12 grandes grupos, el subgrupo Servicios TurÃsticos estÃ¡ incluido en el grupo 9 Ocio y cultura. TambiÃ©n hay que destacar el grupo 11 Hoteles, cafÃ©s y restaurantes, que incluirÃan alojamiento y parte de la oferta complementaria. Con los precios de los productos de cada grupo y subgrupo, el INE elabora el Ã ndice de Precio de Turismo (IPT). Hay que tener en cuenta que el IPC se refiere a los precios que pagan las familias residentes en EspaÃ±a y no los turistas. Hay productos que sÃ³lo son consumidos por turistas, como los souvenirs, pero no hay estudios sobre esa variaciÃ³n de precios.

Instituto Nacional de Estadística. (National Statistics Institute) - Windows I...

http://www.ine.es/inebase2/datos.jsp

IPC Ponderaciones. Base 2001

Base 2001. Series desde enero de 2002

Indices nacionales: general y de grupos COICOP

Unidades: Tanto por mil

	2002	2003	2004	2005	2006
General	1.000,000	1.000,000	1.000,000	1.000,000	1.000,000
Alimentos y bebidas no alcohólicas	218,630	219,309	226,033	226,033	222,825
Bebidas alcohólicas y tabaco	32,170	31,822	31,710	31,710	30,696
Vestido y calzado	99,280	98,993	97,287	97,287	92,507
Vivienda	110,260	106,841	106,908	106,908	107,094
Menaje	63,571	64,101	64,131	64,131	61,660
Medicina	28,062	27,529	26,831	26,831	27,152
Transporte	155,760	153,233	144,028	144,028	149,103
Comunicaciones	25,729	27,345	29,942	29,942	32,762
Ocio y cultura	67,263	68,338	67,650	67,650	67,779
Enseñanza	17,444	16,747	16,739	16,739	16,755
Hoteles, cafés y restaurantes	112,708	111,810	112,340	112,340	114,480
Otros bienes y servicios	69,124	73,933	76,401	76,401	77,187

1) Las ponderaciones en vigor durante el año 2005 son las mismas del año 2004

Fuente: Instituto Nacional de Estadística

Internet 100%

IPC Ponderaciones.

Base 2006.

Ponderaciones IPC Nacional general y de grupos COICOP

Unidades:Tanto por mil

	2007	2008
General	1.000,000	1.000,000
Alimentos y bebidas no alcohólicas	220,556	202,796
Bebidas alcohólicas y tabaco	28,229	26,678
Vestido y calzado	90,280	88,054
Vivienda	103,607	102,582
Menaje	61,520	66,697
Medicina	28,259	30,416
Transporte	148,879	151,963
Comunicaciones	35,845	36,801
Ocio y cultura	71,089	74,957
Enseñanza	16,027	14,684
Hoteles, cafés y restaurantes	115,477	118,699
Otros bienes y servicios	80,230	85,673

Grupo 11. Hoteles, cafés y restaurantes

Indice de Precios de Consumo

Base 2006. Series desde enero de 2002

Indices nacionales: general y de grupos COICOP

Unidades:Base 2006=100

	Indice	Variación mensual	Variación anual
Hoteles, cafés y restaurantes			
2008M01	106,946	0,7	4,8
2008M02	107,575	0,6	4,8
2008M03	108,628	1,0	5,2
2008M04	109,200	0,5	4,5
2008M05	109,448	0,2	4,8
2008M06	109,970	0,5	4,8
2008M07	111,077	1,0	4,8
2008M08	111,974	0,8	4,9
2008M09	110,941	-0,9	4,8
2008M10	110,744	-0,2	4,7
2008M11	110,465	-0,3	4,5

EXAMEN DE ANÁLISIS CUANTITATIVO DE LA ACTIVIDAD TURÍSTICA
Febrero 2009

M4

Los criterios de evaluación de este test son los siguientes: las preguntas bien contestadas suponen un punto, las incorrectas restan 1/3.

1 El número de turistas llegados a un destino en agosto de 2008 fue de 1.327.543. Esto es:
- a) Una frecuencia
- b) Una observación
- c) Una varianza
- d) Una variable

2 En el siguiente cuadro se recogen los valores de la media y la mediana de tres variables.

Variable	Media	Mediana
1	-0,05	-0,05
2	9,01	0,71
3	-10,33	-1,59

Puede afirmarse que:
- a) La primera y segunda variable son simétricas.
- b) La primera variable es simétrica, la segunda asimétrica por la izquierda y la tercera asimétrica por la derecha.
- c) La primera variable es simétrica, la segunda asimétrica por la derecha y la tercera asimétrica por la izquierda.
- d) Las tres variables son simétricas.

3 En un estudio sobre el gasto turístico por persona y día realizado con 4000 turistas, se ha determinado que el cuartil de menos gasto es 66,07 euros. Esto significa que:
- a) Que el turista que gasta menos, gasta 66,07 euros
- b) Que hay 3000 turistas que gastan menos de 66,07 euros
- c) Que hay 2000 turistas que gastan menos de 66,07 euros
- d) Que hay 1000 turistas que gastan menos de 66,07 euros

4 En una determinada distribución, sabemos que su moda deja el 75% de observaciones a su derecha. Entonces la distribución es:
- a) Simétrica
- b) Asimétrica
- c) Normal
- d) Bimodal

5 Contamos con la siguiente información acerca de la satisfacción expresada por los clientes de un hotel sobre el trato recibido:

	n_i	f_i
Muy satisfecho/a	15	0'3
Satisfecho/a	11	0,22
Ni satisfecho/a, ni insatisfecho/a	9	0,18
Insatisfecho/a	7	0,14
Muy insatisfecho/a	8	0,16
	50	1

Según estos datos:
- a) 7 de los 54 individuos de la muestra, dicen estar insatisfechos con el trato recibido en el hotel.
- b) Las opiniones favorables son del 45%.
- c) Las opiniones desfavorables son del 45%.
- d) La moda de esta distribución es "muy satisfecho/a".

6 Para comparar la dispersión entre los precios de una cadena situados en Francia y en España, la mejor medida de dispersión será:
- a) La desviación estándar.
- b) La varianza.
- c) El coeficiente de variación.
- d) El rango intercuartílico.

7 En una encuesta turística se encuentra la siguiente pregunta: *Además de lo que ha pagado en su ciudad, ¿Cuánto ha gastado en nuestras islas?* Las respuestas de 200 encuestados a esta pregunta se recogen en:
- a) Una variable temporal
- b) Una variable cualitativa
- c) Una constante
- d) Ninguna de las anteriores

8 El volumen de pernoctaciones mensual de cierto alojamiento turístico que posee 25 habitaciones individuales, 15 habitaciones dobles y 30 habitaciones de 3 plazas, si se sabe que ocupa el 73% de su capacidad será de:
- a) 145 pernoctaciones
- b) 105,85 pernoctaciones
- c) 3175,5 pernoctaciones
- d) 4350 pernoctaciones

9 Se conoce el número de pasajeros que han volado en la compañía *Air world* cada trimestre del año 2008. Las frecuencias relativas acumuladas han sido para el primer, segundo, tercer y cuarto trimestre respectivamente: 37,2%, 76,5%, 94,5% y 100%. Las frecuencias relativas de cada trimestre son por tanto:
- a) 37,2 %, 40,1%, 19,3%, 3,4%
- b) 37,2%, 39,3%, 18,1%, 5,4%
- c) 37,2%, 38,4%, 15,9%, 8,5%
- d) No se pueden calcular.

10 Este es el *ranking* de las Cadenas Hoteleras en España por el número de plazas.

Cadena	Nº de plazas (2008)
Grupo Sol-Meliá	56007
Riu Hoteles	21261
Hoteles Husa	12278
Hoteles Barceló	11357
N.H. Hoteles	10722
Paradores de España	10394
Iberostar	9837
Tryp Hoteles	9490
Med Playa Hotels	8743
Fiesta Hoteles	8634

Se puede afirmar que los cuartiles de esta distribución son:
- a) 56007, 12278 y 10394
- b) 56007, 12278 y 9837
- c) Grupo Sol-Melia, Hoteles Husa y Paradores
- d) Grupo Sol-Melia, Hoteles Husa y Iberostar

11 Calcule la covarianza entre Yt y Xt con t=1,...,7, sabiendo que sus desviaciones con respecto a la media son -6, -3, -2, 5, -4, 6, 4 para la variable Yt y -2, -1, -4, 0, 2, 3, 2 para la variable Xt. El resultado es:
- a) 6,857
- b) 5,857
- c) -6,857
- d) 0,536

12 Se ha detectado una relación de causalidad entre dos variables cuantitativas de las que se posee 7 observaciones, tienen respectivamente medias de $\overline{X} = 256$ y $\overline{Y} = 26$. Se realiza un ajuste lineal y se sabe que la pendiente es 2,5. ¿Qué valor se obtendrá del término constante del ajuste?
a) -256
b) –614
c) 191
d) 282

13 Se quiere hacer la media ponderada de los precios de Diciembre de 2008 del trayecto Barcelona – Palma que realizan las compañías AIR EU, IRIA y SPAIR. La compañía AIR EU ha tenido 54.000 pasajeros, la compañía IRIA 64.000 y la compañía SPAIR 35.000. ¿Cuales serán las respectivas ponderaciones (en %) que se utilizarán para la media?
a) 34,77, 22,88 y 40,11
b) 0, 3477, 0,2288 y 0,4011
c) 0,3529, 0,4183 y 0,2012
d) 35,29, 41,83 y 22,88

14 Se han estudiado las respuestas de 283 clientes a un cuestionario. De las respuestas a dos de las cuestiones se ha obtenido esta tabla conjunta de frecuencias relativas:

	Bueno	Normal	Malo
si	15,90%	16,25%	4,24%
No	4,95%	18,73%	15,55%
NS/NC	8,83%	3,89%	11,66%

Esta tabla es:
a) de perfiles fila.
b) de perfiles columna.
c) de frecuencias sobre el total de observaciones.
d) de frecuencias marginales.

15 En la siguiente tabla se presentan una serie de cálculos realizados a partir de los valores de la serie trimestral Y_t:

	Y_t	MM(4)	MMC(4)	Y_t-MMC(4)
1T	20			
2T	30	32,5		
3T	70	32,5	32,5	37,5
4T	10	35	33,75	-23,75
1T	20	32,5	33,75	-13,75
2T	40	32,5	32,5	7,5
3T	60	35	33,75	26,25
4T	10	35	35	-25
1T	30	37,5	36,25	-6,25
2T	40	37,5	37,5	2,5
3T	70			
4T	10			

Los factores estacionales correspondientes al primer, segundo, tercer y cuarto trimestre son:

a) 23,33, 36,66, 66,66 y 10
b) -13,75, 7,5, 26,25 y -25
c) -10,625, 4,375, 31,255 y -25,005
d) 31,88, -24,38, -10 y 5

16 Se ha estudiado la relación lineal entre el número de excursiones a Cabrera vendidas durante siete años consecutivos (Ventas) y los precios de la excursión (expresados en euros) en estos siete años (Precios). Un resultado del estudio es que si aumentan los precios en 1 euro, entonces el número de excursiones vendidas disminuye cada año en 235. ¿De qué recta ajustada se ha extraído este resultado?
a) Precios= 143.234 - 235 · Ventas
b) Ventas = 143.234 - 235 · Precios
c) Precios= 235 – 1,43 · Ventas
d) Ventas = 235 – 1,43 · Precios

17 Un Indice de precios de Laspeyres para el año 2008 en base 2000 combina:
a) Precios de 2008 con cantidades de 2000
b) Precios de 2008 con cantidades de 2008
c) Precios de 2008 con cantidades de 2007
d) Precios de 2000 con cantidades de 2008

18 Si en el mes de Diciembre de 2008 se calcula que el IPC (base Enero 2001) es 129,8 y que la tasa de crecimiento de IPC de este mes con respecto al mismo mes del año anterior es 3,09 % debemos interpretar que :
a) Los precios de consumo de los bienes y servicios de Diciembre de 2008 son un 129,8% más elevados que los de Enero de 2001 y han crecido un 3,09% con respecto a los de Diciembre de 2007.
b) Los precios de consumo de los bienes y servicios de Diciembre de 2008 son un 29,8% más elevados que los de Enero de 2001 y han crecido un 3,09% con respecto a los de Diciembre de 2007.
c) Los precios de consumo de los bienes y servicios de Diciembre de 2008 son un 29,8% más elevados que los de Enero de 2001 y han crecido un 103,09% con respecto a los de Diciembre de 2007.
d) Los precios de consumo de los bienes y servicios de Diciembre de 2008 son un 129,8% más elevados que los de Enero de 2001 y han crecido un 103,09% con respecto a los de Diciembre de 2007.

AnÃ¡lisis cuantitativo de la Actividad turÃstica 2

55

77

Frecuencia relativa X (360/100) = x grados

O lo que es lo mismo:

Frecuencia relativa X 3,6 = x grados

30 x 3,6 = 108Âº

39,6 x 3,6 =142,56Âº

8,4 x 3,6 = 30,24Âº

4 x 3,6 = 14,4Âº

18 x 3,6 = 64,8 Âº

25%

75 % 50%

	Â			
2	x			
	Â			
1	Â	x	x	x
	Â	Â	Â	Â
	X	Y	Z	W