

ESTADÍSTICA I

Cuestiones tipo test con R-Commander

Victoria Alea Riera
Isabel Maqueda de Anta
Carmen Muñoz Vaquer
Elisabeth Torrelles Puig
Núria Viladomiu Canela

Departamento de Econometría, Estadística
y Economía Española

ESTADÍSTICA I

Cuestiones tipo test con R-Commander

Victoria Alea Riera
Isabel Maqueda de Anta
Carmen Muñoz Vaquer
Elisabeth Torrelles Puig
Núria Viladomiu Canela

Departamento de Econometría, Estadística
y Economía Española

Índice

PRÓLOGO	7
1. ESTADÍSTICA Y TIPO DE DATOS	9
1. Instalación de R-Commander	10
2. Ejercicios del capítulo	11
2. DISTRIBUCIÓN DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA	15
1. Tablas de frecuencias	17
2. Diagrama de barras e histograma	20
3. Diagrama de frecuencias acumuladas	24
4. Diagrama de tallo y hojas	27
5. Ejercicios del capítulo	30
3. MEDIDAS DE POSICIÓN	37
1. Media aritmética	38
2. Media, mediana y moda	41
3. Mediana y cuantiles	43
4. Ejercicios del capítulo	47
4. MEDIDAS DE DISPERSIÓN Y FORMA	53
1. Varianza, desviación estándar y coeficiente de variación	55
2. Transformaciones lineales	58
3. Medidas de forma y <i>box plot</i>	60
4. Ejercicios del capítulo	63
5. DISTRIBUCIÓN DE FRECUENCIAS BIDIMENSIONALES	73
1. Distribuciones bidimensionales	74
2. Independencia	77
3. Asociación entre variables cualitativas	79
4. Ejercicios del capítulo	82
6. ASOCIACIÓN ENTRE VARIABLES CUANTITATIVAS	87
1. Covarianza y coeficiente de correlación lineal	88
2. Regresión lineal	91
3. Diagrama de dispersión	94
4. Ejercicios del capítulo	97

7. INTRODUCCIÓN A LA TEORÍA DE LA PROBABILIDAD	105
1. Probabilidad. Axiomas y propiedades	106
2. Probabilidad condicionada. Teorema de la intersección	108
3. Independencia de sucesos	110
4. Teorema de la probabilidad total y teorema de Bayes	112
5. Ejercicios del capítulo	114
8. VARIABLE ALEATORIA	119
1. Variable discreta	121
2. Variable continua	123
3. Esperanza matemática y varianza	126
4. Ejercicios del capítulo	129
9. DISTRIBUCIONES BINOMIAL Y NORMAL	135
1. Ejercicios del capítulo	136
SOLUCIONARIO	141
Soluciones de los ejercicios por epígrafes	141
Soluciones de los ejercicios del capítulo	142
BIBLIOGRAFÍA	143

Prólogo

Este libro ofrece una extensa variedad de cuestiones tipo test de la materia de Estadística I de estudios de grado orientados a las Ciencias Económicas y Sociales.

Las cuestiones han sido diseñadas con la finalidad de:

- Ayudar al lector en el estudio y en la comprensión de los principales conceptos de la materia.
- Poner en práctica su capacidad para resolver problemas aplicando los conocimientos teóricos adquiridos.
- Familiarizar al lector con la utilización del programa R-Commander y la interpretación de los resultados obtenidos.
- Proporcionar un instrumento de autoevaluación.

Las cuestiones se agrupan en ocho capítulos, siguiendo el programa de la asignatura Estadística I. Cada capítulo comienza con una breve introducción de los conceptos principales y un cuadro resumen de fórmulas de la lección con el objetivo de que el lector las tenga presentes a la hora de resolver las cuestiones planteadas. Éstas presentan cuatro respuestas alternativas de las que sólo una es correcta.

Debido a que en la actualidad la incorporación de los programas de análisis de datos resulta imprescindible en el estudio de la estadística, se recomienda al lector que utilice el programa R-Commander, para lo cual se le facilita una breve guía de instalación. Algunas cuestiones deben resolverse con la ayuda de dicho programa y otras presentan resultados ya obtenidos con el programa con el objetivo de interpretarlos. Estas cuestiones están señaladas con el signo (*Rcmdr*). El lector puede encontrar las bases de datos en formato R necesarias para la realización de algunas cuestiones en el apartado «Base de dades» de la dirección <http://www.publicacions.ub.edu/liberweb/estadisticaR-Commander/>.

En cada capítulo, las cuestiones se han agrupado en dos bloques. El primer bloque sigue el orden de los principales epígrafes del capítulo y el segundo recopila ejercicios de todos los epígrafes del capítulo.

El anexo contiene la lista de las soluciones de los ejercicios propuestos, de manera que el usuario pueda realizar su autoevaluación a medida que progresa en el estudio.

Este libro no pretende ser un manual básico de la asignatura, pero sí un útil instrumento de apoyo en el estudio de la misma.

Barcelona, marzo de 2011

ESTADÍSTICA Y TIPO DE DATOS

La *Estadística* es la ciencia que nos proporciona los métodos necesarios para el estudio empírico de los fenómenos aleatorios.

Los conceptos *población*, que es el colectivo cuya descripción es el objetivo final del análisis estadístico, y *muestra*, que es un subconjunto representativo de la población, permiten determinar la información empírica que constituye la masa de *datos* que se analizarán con los métodos estadísticos.

Hay que distinguir entre *estadística descriptiva* e *inferencia estadística*. El análisis descriptivo puede realizarse con datos muestrales o con censos poblacionales, mientras que la inferencia trabaja con la información que proporcionan los datos muestrales, para llegar a conclusiones acerca de la población.

La naturaleza de las características observadas permite clasificar los distintos tipos de datos:

- Si la característica observada es cuantificable se genera una *variable* y cuando la característica es cualitativa se genera un *atributo*.
- Si los datos están referidos al mismo instante de tiempo se denominan de *corte transversal*, pero si se observa el comportamiento de la característica en el tiempo se tiene una *serie temporal*.
- Si se analiza una sola variable o atributo los datos son *unidimensionales*. Si se analizan dos o más características conjuntamente los datos son *bidimensionales* o *multidimensionales*.

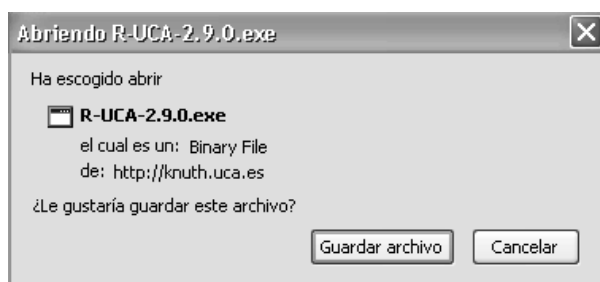
En este capítulo se incluyen las instrucciones de instalación del programa R-Commander. Se recomienda al lector que instale el programa y se familiarice con los procedimientos básicos ya que algunos ejercicios propuestos en este libro se podrán resolver con la ayuda de este programa.

TIPO DE DATOS		
• Según la naturaleza del fenómeno	Cualitativos	Nominales
	Atributos	Ordinales
	Cuantitativos	Discretas
	Variables	Continuas
• Según la escala de medida	Nominal	
	Ordinal	
	De intervalo	
	De razón	
• Según el momento de observación	Corte transversal	
	Series temporales	

1. Instalación de R-Commander

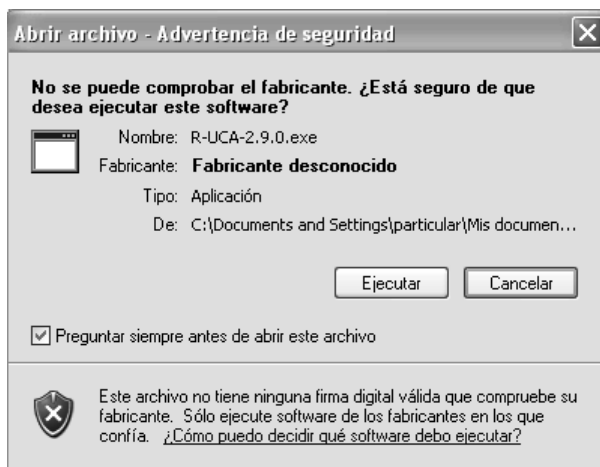
Para instalar la versión de R y R-Commander adaptada por la Universidad de Cádiz es preciso:

- 1) Descargar el programa que se encuentra en la página: <http://knuth.uca.es>.
- 2) Elegir la opción PAQUETE RUCA.
- 3) Descargar la última versión desde el servidor 2. Aparecerá el siguiente cuadro de diálogo. Si se navega con Internet Explorer el cuadro de diálogo es ligeramente distinto y ofrece la posibilidad de ejecutar sin guardar.



Almacenar el fichero R-UCA-xxx.exe en el subdirectorio de las bajadas de internet del ordenador clicando el botón GUARDAR ARCHIVO.

- 4) Ejecutar el programa de instalación clicando sobre R-UCA-xxx.exe. Aparecerá el siguiente cuadro de diálogo:



Activar la opción EJECUTAR.

Finalizada con éxito la instalación en la barra de Programas aparecerá el acceso a R.

2. Ejercicios del capítulo

1. El estudio de los procedimientos de organización, resumen y representación de los datos estadísticos forma parte de la rama de la Estadística llamada:
 - a) Estadística descriptiva.
 - b) Muestreo.
 - c) Estadística inferencial.
 - d) Informática básica.
2. ¿Cuál de los siguientes fenómenos *no* es aleatorio y, por lo tanto, no es objeto de análisis estadístico?
 - a) Proporción de piezas defectuosas que se obtiene en un determinado proceso de producción.
 - b) Cuota de amortización mensual que se debe pagar para amortizar un préstamo a interés fijo.
 - c) Nota mínima a exigir en los próximos cursos para acceder a una determinada facultad.
 - d) Cuota de mercado que alcanzará una determinada marca.
3. El nivel de calidad y servicios de los establecimientos hoteleros se mide mediante la siguiente escala ordinal (de menor a mayor categoría): *, **, ***, ****, *****. Si se recodifica la variable en una escala de 1 a 5 se está midiendo ahora en una escala:
 - a) Ordinal.
 - b) De intervalo.
 - c) De razón.
 - d) Nominal.
4. En el contexto de un estudio encaminado a determinar si los alumnos están de acuerdo con los nuevos planes de estudio, una muestra representativa de este colectivo es:
 - a) Los representantes sindicales.
 - b) Los alumnos de primer curso.
 - c) Un grupo de alumnos becados.
 - d) Ninguna de las anteriores.
5. Los datos procedentes de un censo de población constituyen:
 - a) Una serie temporal.
 - b) Una muestra de corte transversal.
 - c) Información sobre toda la población.
 - d) Una serie temporal de corte transversal.
6. En una encuesta el género del entrevistado se codifica con los valores 0 para los hombres y 1 para las mujeres. Entonces, la característica $X = \text{«género»}$ es:
 - a) Cuantitativa discreta.
 - b) Cualitativa dicotómica.
 - c) Cualitativa ordinal.
 - d) Cuantitativa continua.
7. Indique cuál de las siguientes variables es de naturaleza discreta:
 - a) Peso.
 - b) Tiempo de espera.

- c) Rentabilidad anual de un activo financiero.
 - d) Número de electrodomésticos por unidad familiar.
8. Cuando se observan las temperaturas máximas diarias en una localidad durante n días y los resultados se miden en grados centígrados, la escala de medida utilizada es:
- a) Ordinal.
 - b) Nominal.
 - c) De intervalo.
 - d) De razón.
9. Aquella escala de medida que sólo permite ordenar los datos es:
- a) Nominal.
 - b) Ordinal.
 - c) Escala de intervalo.
 - d) Ninguna de las anteriores.
10. El número de personas que componen una familia es una variable:
- a) Cuantitativa discreta.
 - b) Cualitativa nominal.
 - c) Cualitativa ordinal.
 - d) Cuantitativa continua.
11. Los datos resultantes de observar si un producto es o no defectuoso son:
- a) Cuantitativos discretos.
 - b) Cualitativos nominales.
 - c) Cualitativos ordinales.
 - d) Cuantitativos continuos.
12. Si en una muestra hay elementos para los que no se dispone de información sobre alguna característica se dice que hay:
- a) Datos missing.
 - b) Outliers.
 - c) Datos sobrantes.
 - d) Elementos extremos.
13. Se observa el peso de 300 maletas facturadas en un aeropuerto. La variable X = «peso (en kg)» se mide en una escala:
- a) Ordinal.
 - b) Nominal.
 - c) De intervalo.
 - d) De razón.
14. En el contexto de una encuesta sobre el medio de transporte en una localidad en la que circulan 5 líneas de autobús se ha preguntado a los usuarios qué «línea de autobús» utilizan con más frecuencia, siendo las posibles respuestas, 1, 2, 3, 4 o 5. Los datos resultantes se miden en una escala:
- a) Nominal.
 - b) Ordinal.
 - c) De intervalo.
 - d) De razón.

15. El número de viajeros según el país de procedencia es:

Países	Viajeros	%
Alemania	250.414	46,07%
Austria	12.767	2,35%
Bélgica	37.430	6,89%
Dinamarca	21.552	3,96%
Finlandia	18.732	3,45%
Francia	171.629	31,57%
Otros	31.042	4,71%
Total	543.566	100,00%

La naturaleza de la variable recogida en esta tabla es:

- a) Continua.
 - b) Discreta.
 - c) Ordinal.
 - d) Nominal.
16. En el contexto de un estudio sobre el metro de Barcelona se ha observado la variable X = «número mensual de viajeros» en el periodo de 2009 a 2010. Los datos resultantes son:
- a) Categóricos.
 - b) Corte transversal.
 - c) Serie temporal.
 - d) Dicotómicos.
17. En un instituto se ha observado las asistencias a clase en función de la franja horaria a fin de determinar las franjas con más ausencias. Los valores se han tabulado de acuerdo con las franjas horarias siguientes:

8-9	9-10	10-11	11-12	12-13	13-14
-----	------	-------	-------	-------	-------

La variable «franja horaria» es:

- a) Continua.
 - b) Discreta.
 - c) Ordinal.
 - d) Nominal.
18. Con cuál de los siguientes datos no realizaría inferencia estadística:
- a) Con una muestra de 1.000 familias sobre las que se observa el gasto semanal en alimentación.
 - b) Con una muestra de 600 exámenes de contabilidad sobre los que se observa la nota.
 - c) Con el censo de la población del año 2000 en el que se observa la población activa por provincias.
 - d) Con una muestra de 50 empresas sobre las que se observa el sector al que pertenecen.
19. Una base de datos contiene información acerca de 30 características socioeconómicas observadas en una muestra de 3.000 familias. En 213 de estas familias no se dispone de

información de la variable $X = \text{«número de habitaciones de la vivienda habitual»}$. Para esta variable X el número de casos válidos es:

- a) 3.000
- b) 2.787
- c) 213
- d) 30

20. De una muestra de n electores censados en un municipio en el que se presentan 4 candidatos a la alcaldía, el 27 % tiene decidido votar al candidato A. Si se quiere estimar el porcentaje del total de electores que tiene decidido votar al candidato A debe utilizarse:

- a) Un procedimiento para el análisis de series temporales.
- b) Alguna técnica descriptiva de resumen de datos bidimensional.
- c) Algún método de inferencia estadística.
- d) Una base de datos sin valores missing.

Capítulo 2

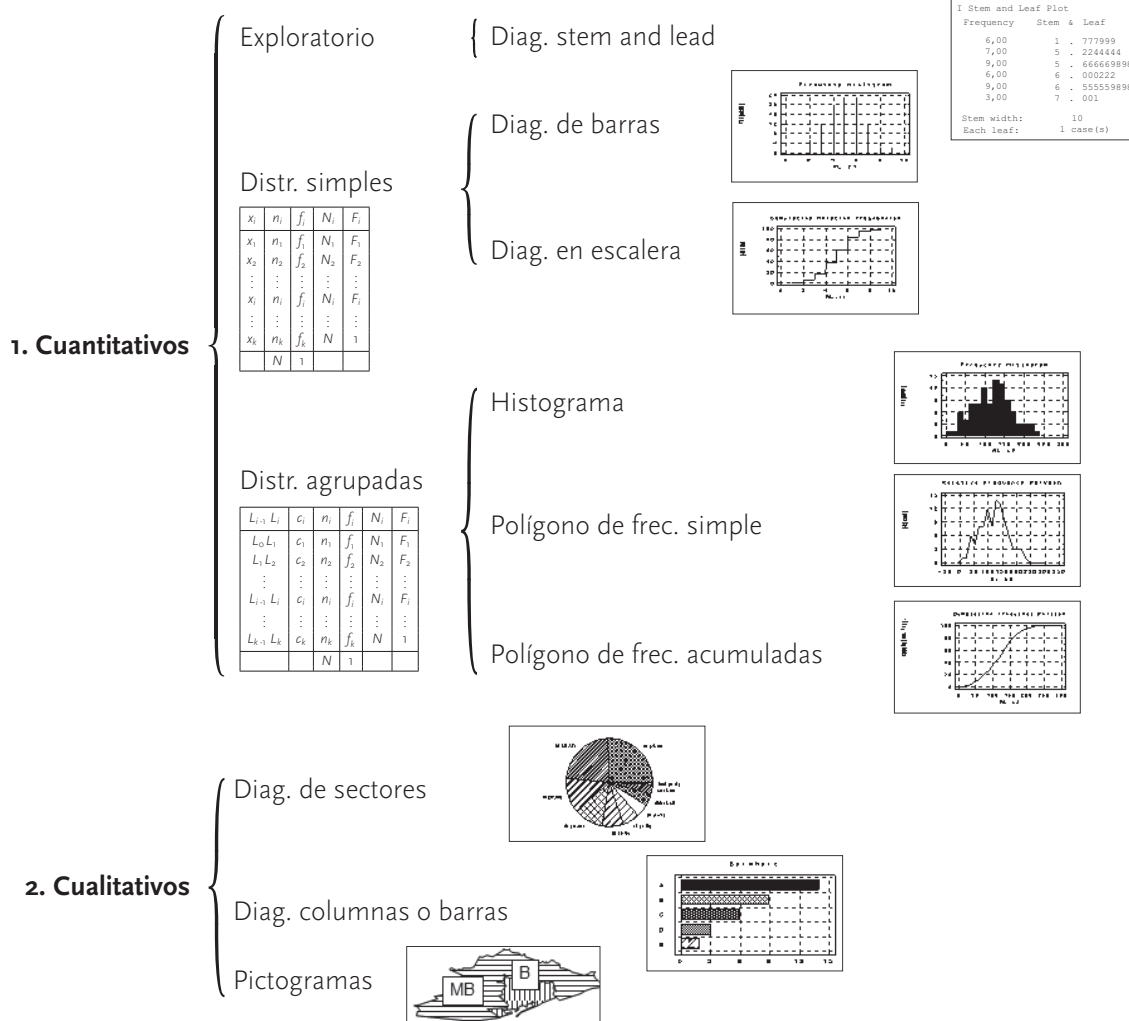
DISTRIBUCIÓN DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA

Una vez obtenidos los datos es preciso presentarlos de forma ordenada, procurando conservar al máximo la información que contienen. Con este fin se introduce el concepto de *distribución de frecuencias* como técnica para la tabulación o representación numérica de variables estadísticas unidimensionales de corte transversal.

A continuación se exponen algunas técnicas para la representación gráfica de los datos, en particular el *diagrama de barras*, el *histograma* y el *polígono de frecuencias*, por el interés que presentan como base empírica para la especificación de un modelo teórico.

Se hace referencia, asimismo, a otros tipos de gráficos como el *diagrama de tallo y hojas* (*stem and leaf*), y en capítulos posteriores, el *diagrama de caja* (*box plot*) y el *diagrama de puntos*, que permiten explorar la distribución de frecuencias de los datos e identificar las medidas de síntesis más adecuadas para su descripción.

TABULACIÓN Y REPRESENTACIÓN GRÁFICA



- Frecuencia absoluta, n_i
- Tamaño de la muestra, $n = \sum_{i=1}^K n_i$
- Frecuencia relativa, $f_i = \frac{n_i}{n}$
- Frecuencia absoluta acumulada hasta el valor x_k , $N_k = \sum_{i=1}^k n_i$
- Frecuencia relativa acumulada hasta el valor x_k , $F_k = \sum_{i=1}^k f_i = \frac{\sum_{i=1}^k n_i}{n}$

1. Tablas de frecuencias

- Una variable X toma únicamente 4 valores distintos: x_1, x_2, x_3, x_4 (en orden creciente). En una muestra de tamaño 500 se observa que: el 45 % de las observaciones toma el valor x_2 ; la proporción de observaciones en las que el valor de X es como máximo x_3 es 0,9 y 150 observaciones toman el valor x_1 . Es cierto que:
 - La frecuencia relativa de x_1 es 0,70.
 - La frecuencia absoluta de x_2 es 375.
 - La frecuencia absoluta de x_4 es 50.
 - La frecuencia relativa de x_3 es 0,20.
- La siguiente distribución de frecuencias relativas acumuladas de la variable X = «número de contratos conseguidos en el mes de enero» corresponde a la observación de la actividad de 150 teleoperadores de una compañía de telefonía móvil:

X_i	58	60	62	65	68	70	71
F_i	0,06	0,2	0,4	0,64	0,8	0,92	1

Indique cuál de las siguientes afirmaciones es falsa:

- El 8 % de los teleoperadores ha conseguido exactamente 71 contratos.
 - 90 teleoperadores han conseguido como mínimo 62 y como máximo 68 contratos.
 - El 60 % de los teleoperadores ha conseguido más de 62 contratos.
 - 40 teleoperadores han conseguido como máximo 62 contratos.
- En una encuesta realizada a 100 familias se ha obtenido la siguiente distribución de frecuencias absolutas acumuladas de X = «número de miembros que componen la unidad familiar»:

X_i	1	2	3	4	5	6	7
N_i	15	32	57	74	87	96	100

Una familia con más de 4 miembros estará entre:

- Las 13 familias más numerosas.
 - Las 32 familias menos numerosas.
 - Las 26 familias más numerosas.
 - Las 57 familias menos numerosas.
- La suma de frecuencias absolutas es igual a:
 - 1.
 - El número de intervalos de clase.
 - El tamaño muestral.
 - El tamaño poblacional.
 - Cuando se afirma que «el 40 % de las exportaciones se envía a destinos que se encuentran a más de 5.000 km», este 40 % es:
 - Una frecuencia absoluta de la distribución de la variable «valor de las exportaciones».
 - Una frecuencia relativa de la distribución de la variable «distancia del destino de las exportaciones».

- c) Un intervalo de clase de la distribución de la variable «distancia del destino de las exportaciones».
- d) Una frecuencia relativa de la distribución de la variable «valor de las exportaciones».
6. Se han realizado 60 llamadas a un servicio de información telefónico a fin de analizar «el tiempo de espera (en min)». Se considerará que el servicio no es operativo cuando más del 30 % de las llamadas tenga un tiempo de espera superior a 3 minutos. Si la distribución de frecuencias relativas del tiempo de espera es:

x_i	f_i
0-1	0,05
1-2	0,20
2-3	0,30
3-4	0,25
4-5	0,15
5-6	0,05

el estudio concluirá que:

- a) El servicio no es operativo.
- b) El servicio es operativo.
- c) No se puede llegar a ninguna conclusión.
- d) Ninguna de las anteriores.
7. Dada la siguiente distribución de frecuencias de la variable X = «número de unidades por caja»:

X_i	100	200	300	400
n_i	4	2	2	2

Indique cuál de las siguientes afirmaciones es falsa:

- a) Hay 6 cajas con 200 o menos unidades.
- b) El 80 % de las cajas contiene como máximo 300 unidades.
- c) El 20 % de las unidades está contenido en 200 cajas.
- d) Estas cajas contienen como máximo 400 unidades.
8. La distribución de frecuencias de la variable X = «importe IBI (en €)» observada en una muestra de pisos de una localidad es:

```
> .Table # counts for Y
      50-150 150-250 250-350 350-450 450-550
      44      24      40      25      17

> 100* .Table/sum(.Table) # percentages for Y
      50-150 150-250 250-350 350-450 450-550
29.333333 16.000000 26.666667 16.666667 11.333333
```

Indique cuál de las siguientes afirmaciones es cierta:

- a) El tamaño de la muestra es 100.
- b) El 59,3 % de los pisos paga como mínimo 450 €.