

PROBLEMES D'ESTADÍSTICA DESCRIPTIVA

Manuela Alcañiz
Montserrat Guillén
Ernest Pons
Héctor Rufino
(eds.)

Departament d'Econometria, Estadística i Economia Espanyola

PROBLEMES D'ESTADÍSTICA DESCRIPTIVA

Manuela Alcañiz
Montserrat Guillén
Ernest Pons
Héctor Rufino

Departament d'Econometria, Estadística
i Economia Espanyola

Índex

INTRODUCCIÓ	7
--------------------------	---

EXERCICIS

Tema 1. Introducció: conceptes bàsics	11
Tema 2. Representació gràfica i tabulació de dades	13
Tema 3. Mesures de síntesi per a dades quantitatives univariants (I)	17
Tema 4. Mesures de síntesi per a dades quantitatives univariants (II)	23
Tema 5. Dades multidimensionals	27
Tema 6. Regressió.....	35
Tema 7. Nombres índexs	41

SOLUCIONS

Solucions als exercicis del tema 1	47
Solucions als exercicis del tema 2	49
Solucions als exercicis del tema 3	51
Solucions als exercicis del tema 4.....	53
Solucions als exercicis del tema 5	55
Solucions als exercicis del tema 6.....	59
Solucions als exercicis del tema 7	63

FORMULARIS

Formulari del tema 2	67
Formulari del tema 3	68
Formulari del tema 4.....	69
Formulari del tema 5	70
Formulari del tema 6.....	72
Formulari del tema 7	73

Introducció

A continuació es presenten un seguit de materials d'estadística descriptiva. El seu objectiu fonamental és servir de suport per a les classes de problemes de l'assignatura del mateix nom que es desenvolupa al grau en Estadística impartit per la Universitat de Barcelona i la Universitat Politècnica de Catalunya. Tanmateix també poden ser utilitzats en altres assignatures que tinguin una part dedicada a aquesta matèria.

Amb més o menys intensitat, es tracten tots els temes que es desenvolupen a les sessions teòriques de l'assignatura d'Estadística Descriptiva, i es proposen nombrosos exercicis per poder treballar amb els conceptes i els procediments explicats.

Per la seva situació en el Pla d'estudis vigent, els problemes no suposen cap coneixement de probabilitat, ja que aquesta s'estudia en l'assignatura d'Introducció a la Probabilitat. Malgrat això, per les característiques dels estudiants que cursen l'assignatura, es considera natural que els problemes continguin raonaments sobre mesures d'estadística descriptiva formulats de manera general, i no només particularment per a un conjunt de dades, com sol ser habitual en la major part dels llibres d'estadística descriptiva.

Amb vista a la utilització del material per a la docència, es recomana seguir la numeració dels problemes proposats. En els casos en què es proporcionin dades, serà interessant dur a terme la resolució mitjançant els programes informàtics que s'estudien en l'assignatura d'Introducció a la Informàtica del mateix ensenyament.

Per tal de facilitar a l'alumne la resolució dels problemes, s'inclouen les solucions de molts dels exercicis, principalment d'aquells de caràcter numèric. També s'inclou un formulari com a compendi d'alguns dels principals conceptes i fórmules que es fan servir a l'assignatura.

Barcelona, febrer de 2014

EXERCICIS

Tema 1

INTRODUCCIÓ: CONCEPTES BÀSICS

1. Entre els estudiants d'una classe se seleccionen, per a fer-los una enquesta, únicament els que no resideixen a Barcelona. Raoneu si es tracta d'una mostra. En cas que ho sigui, creieu que és aleatòria?
2. Poseu un exemple de cadascun dels conceptes següents:
 - a) Variable qualitativa
 - b) Variable quantitativa discreta
 - c) Variable quantitativa contínua
 - d) Categories o modalitats d'una variable qualitativa
 - e) Dada quantitativa
 - f) Dada qualitativa
 - g) Població
 - h) Element de la població
 - i) Subpoblació
 - j) Mostra
 - k) Cens
3. Indiqueu, en cadascun dels casos següents, quines variables són qualitatives i quines són quantitatives (discretes o contínues).
 - a) Temperatura d'un forn
 - b) Trucades telefòniques rebudes per minut en una centraleta
 - c) Temps d'espera en una cua d'un supermercat
 - d) Classe social d'un entrevistat
 - e) Grau d'acceptació d'un producte de cosmètica
 - f) Categoria d'un hotel
 - g) Nombre d'unitats defectuoses en un lot de productes
 - h) Grup sanguini
 - i) Nivell cultural de l'audiència d'una cadena de TV
 - j) Marca del vehicle d'automoció
 - k) Quantitat de pluja enregistrada el mes anterior en un observatori
 - l) Temps d'una reacció química
4. Raoneu quina escala de mesura es farà servir per a cadascuna de les variables següents:
 - a) Temperatura ambient en °C
 - b) Nivell sociocultural d'una persona
 - c) Colors disponibles per a un model d'automòbil
 - d) Pes d'una persona
 - e) Acceptació d'un cert producte
 - f) Esport preferit
 - g) Densitat de població

5. A continuació es presenten un conjunt de casos. Identifiqueu, en cadascun d'ells, quina és la població de referència, quins són els individus i quines les variables mesurades. Per a cadascuna d'aquestes variables comenteu també de quin tipus és (qualitativa, quantitativa, discreta, contínua) i quin nivell de mesura es farà servir:

Cas 1. El cap de vendes d'una empresa dedicada a la venda d'ordinadors està preocupat pel descens de vendes que s'està produint durant l'any 2012 i vol estudiar el perfil de la gent que encara no té ordinador personal a casa. Amb aquest objectiu realitza una enquesta a 1.400 persones de tot Catalunya, seleccionades de manera aleatòria i obté informació sobre si tenen o no ordinador a casa, la seva edat, el seu estat civil i els seus ingressos.

Cas 2. Una important empresa constructora està realitzant un estudi per comparar l'evolució del preu dels pisos a les ciutats de Barcelona i Madrid. Per això ha recollit el preu, la superfície (m^2) i el nombre d'habitacions de tots els pisos que ha construït i venut en aquestes ciutats durant els anys 2010, 2011 i 2012.

Cas 3. Volem conèixer el nombre de germans que tenen els estudiants del grau en Estadística i per disposar d'informació recollim les dades sobre el nombre de germans dels estudiants que hi ha matriculats a l'assignatura d'Estadística Descriptiva.

Cas 4. Un aficionat al bàsquet s'ha dedicat a recollir informació sobre el punts i rebots que aconsegueix el seu jugador preferit cada partit que es juga durant tota la lliga.

6. Indiqueu, en cadascun dels casos següents, si es tracta de sèries temporals, de dades de tall transversal o de dades de panel:

- a) L'atur registrat des del primer trimestre de 2007 als països de la Unió Europea.
- b) L'atur registrat en el segon trimestre de 2010 als països de la Unió Europea.
- c) Els ingressos mitjans i l'edat mitjana dels pilots d'una companyia d'aviació en els últims 10 anys.
- d) Els visitants que ha rebut Barcelona per turisme en els darrers 5 anys.
- e) El pes i l'alçada d'un conjunt de nens des que neixen fins als 10 anys.
- f) El nombre de germans que tenen els habitants d'una població.

Tema 2

REPRESENTACIÓ GRÀFICA I TABULACIÓ DE DADES

7. La taula que es mostra tot seguit conté els ensenyaments amb més estudiants matriculats a la Universitat de Barcelona durant el curs 2009-2010.

Branca	Nombre d'estudiants
Administració i Direcció d'Empreses	3.645
Dret	2.711
Economia	2.086
Belles Arts	1.920
Psicologia	1.504
Altres	15.510
Total UB	27.376

Font: INE

- Quin tipus de variable és la que recull la taula? Quin és el seu nivell de mesura?
- Feu una representació gràfica que mostri la distribució anterior.
- Calculeu les freqüències relatives i les freqüències acumulades, tant absolutes com relatives, només en el cas que els resultats que s'obtinguin siguin interpretables.
- Quin percentatge representen els cinc ensenyaments amb més estudiants sobre el total de la Universitat de Barcelona?

8. Indiqueu de quina manera tabularíeu o faríeu gràfics de les dades següents referides a 1.000 candidats a un cert lloc de treball:

- Estudis realitzats.
- Lloc de residència.
- Edat.
- Si han fet el servei militar o no.
- Idiomes que parlen.
- Sou en l'anterior feina.
- Estat civil.
- Valoració del lloc de treball ofert (de l'1 al 9).
- Puntuació en un test (de l'1 al 100).
- Coneixements informàtics (sí o no).

9. A continuació es mostra una taula que recull la població de les províncies catalanes al 1996 i al 2011:

Població	1996	2011
Barcelona	4.628.277	5.529.099
Girona	530.631	756.810
Lleida	356.456	442.308
Tarragona	574.676	811.401
Total	6.090.040	7.539.618

Font: INE

- a) Quins mètodes gràfics es poden utilitzar per comparar la població de 1996 i la de 2011?
Comenteu els resultats.
- b) És possible presentar l'anterior informació en termes relatius? Com?

10. Les dades següents corresponen al nombre de viatges mensuals que 212 persones enquestades realitzen per les rondes de Barcelona. Es tracta de conductors que fan servir el vehicle particular com a mitjà de desplaçament al treball:

Nombre de viatges al mes	Nombre de persones
0-9	90
10-19	28
20-29	43
30-39	17
40-49	16
50-120	18
Total	212

- a) De quin tipus és la variable analitzada?
- b) Calculeu el nombre de persones que realitzen 20 viatges al mes o més.
- c) Calculeu el percentatge de persones que realitzen entre 20 i 39 viatges mensuals.
- d) Quin percentatge de persones realitzen 40 o més viatges mensuals?

11. Preneu les dades següents de 30 persones corresponents a la variable quantitativa pes. Construïu una taula de freqüències absolutes, relatives, absolutes acumulades i relatives acumulades (podeu agrupar les variables en intervals).

62,	74,	86,	53,	51,	71,	68,	67,	69,	70,	58,	59,	73,	74,	78,
89,	52,	74,	71,	64,	63,	81,	60,	75,	68,	77,	62,	70,	73,	81.

Construïu també un diagrama de troncs i fulles. Comenteu el resultat.

12. Completeu la taula següent a partir de la informació disponible:

x_i	n_i	N_i	f_i	F_i
[0-100)	16			
[100-200)		40	0,12	
[200-350)	42			
[350-500)	56			
[500-700)				
[700-1000)	24			1,00
Total				

13. A la taula següent s'han recollit dades sobre el nombre de passatgers (en milers de persones), en anys diferents, als aeroports comercials de Catalunya. Representeu gràficament aquestes dades de la forma més adient.

Nombre de passatgers	Barcelona	Girona	Reus
2006	29.835,7	3.589,2	1.370,9
2007	32.705,4	4.827,4	1.294,0
2008	30.169,7	5.489,9	1.268,0
2009	27.294,2	5.264,7	1.689,3
2010	29.172,2	4.840,6	1.403,0
2011	34.332,5	2.991,9	1.348,3

Font: Idescat

14. Una empresa en expansió necessita contractar 100 nous venedors. Per tal de dur a terme la selecció dels aspirants s'han realitzat unes proves, a partir de les quals s'ha obtingut la taula següent, que recull les qualificacions dels candidats:

Qualificació	Nombre de candidats
[0, 5)	40
[5, 6)	60
[6, 9)	75
[9, 10)	25

- Calculeu les freqüències relatives, i les freqüències acumulades, tant absolutes com relatives i representeu gràficament la distribució.
- Calculeu la nota mínima que exigirà l'empresa per seleccionar els candidats i el percentatge de candidats que no arriba al 5.
- Quina és la qualificació més comuna entre els candidats?

15. A partir de la informació següent, representeu dos diagrames de *box-plot* (o diagrama de caixa):

- Min = 0, $Q_1 = 0,5$, Me = 1,2, $Q_3 = 2,5$, Max = 4
- Min = 0, $Q_1 = 1,5$, Me = 2,8, $Q_3 = 3,5$, Max = 4

Aquests diagrames corresponen, respectivament, a les variables *consum setmanal de cervesa* (a) i *consum setmanal de llet* (b), tots dos en litres. Les dades corresponen a 200 persones i els seus histogrames són campaniformes. En vista dels gràfics que heu fet, responeu raonadament les preguntes següents:

- Quantes persones consumeixen entre 0,5 i 2,5 litres setmanals de cervesa? I menys de 2,8 litres de llet?
- De mitjana, es consumeix més llet o més cervesa?

Tema 3

MESURES DE SÍNTESI PER A DADES QUANTITATIVES UNIVARIANTS (I)

16. Donades les dades següents: $-4, -1, 2, 4$, és la mitjana geomètrica una mesura representativa de la seva tendència central? Per què?

17. Quina mesura de síntesi faríeu servir per calcular la vida mínima que ha de tenir una de les bombetes d'un cert lot, de manera que pugui ser classificada entre la meitat que més duren?

18. A continuació teniu les dades sobre la superfície de certs habitatges d'una mateixa ciutat, en metres quadrats. Calculeu la mitjana aritmètica, la mediana, la moda i el rang mitjà de les dades i raoneu si totes aquestes mesures han de ser semblants, i, en tot cas, per què ha de ser així:

89,	73,	91,	112,	117,	108,	53,	95,	72,	88,
109,	132,	120,	110,	112,	98,	95,	132,	114,	87.

19. Seleccioneu quin tipus de mitjana calcularíeu en cadascuna de les situacions següents:

- Un inversor té els seus estalvis repartits en tres dipòsits amb 2.000, 5.000 i 10.000 euros, respectivament. Si el primer li rendeix un 5% anual, el segon un 4% anual i el tercer un 2% anual, quin és el tipus d'interès mitjà que rep?
- Per avaluar la velocitat mitjana que s'assoleix amb una nova línia ferroviària s'han fet un total de 5 proves en tot el trajecte (55 km) en què s'han registrat velocitats uniformes de 86 km/h, 87,5 km/h, 89 km/h, 84 km/h i 86 km/h. Calculeu la velocitat mitjana.
- El rendiment d'una màquina depèn del nombre d'operaris que hi treballen. Un dia ha fabricat 3 peces per hora, un altre dia 4 peces per hora, un tercer dia 5 peces per hora i un quart dia 6 peces per hora. Quin és el seu rendiment mitjà?
- L'Ajuntament de Barcelona vol mesurar el que es triga a creuar un determinat semàfor a partir de les dades referides a 4 persones: 15 segons, 13 segons, 12 segons i 8 segons. Calculeu la mitjana del temps necessari per creuar el semàfor.
- Un taronger dóna 100 taronges el primer any, 180 el segon, 234 el tercer i 262 el quart. Quin ha estat l'increment mitjà anual en el nombre de taronges que ha donat?

20. Un inversor en Borsa ha comprat accions d'una mateixa empresa en cinc moments del temps pagant preus de 18, 20, 15, 24 i 25 euros per acció, respectivament. Calculeu la mitjana del preu de les accions que ha comprat en cadascun dels supòsits següents:

- Els cinc cops va comprar el mateix nombre d'accions.
- Els cinc cops va invertir el mateix capital.

21. Una persona que tenia 60.000 € estalviats, els va repartir a finals de l'any 2001 en tres fons d'inversió. Després d'un any, el primer fons (on va posar 10.000 €) li ha donat un rendiment del 10%, el segon fons (on va posar 20.000 €) li ha donat un rendiment del 4% i el tercer fons li ha donat un rendiment de l'1%. Quin és el rendiment mitjà (en tant per cent) que ha rebut aquesta persona pels seus estalvis?

22. Suposa que has tingut la sort de guanyar un premi de loteria i t'ha tocat la quantitat d'un milió d'euros. Per això, reps trucades de quatre grans bancs que t'ofereixen que posis els diners en un dels seus dipòsits a 5 anys. Quina de les quatre ofertes triaries?

Tipus d'interès dipòsits bancaris			
	Banc 1	Banc 2	Banc 3
Any 1	10%	2%	18%
Any 2	10%	6%	14%
Any 3	10%	10%	10%
Any 4	10%	14%	6%
Any 5	10%	18%	2%

23. Es calcula que en un país l'any 2006 hi havia aproximadament 15 milions de telèfons mòbils. Si s'espera que l'any 2013 n'hi hagi 25 milions, quin és el creixement anual mitjà que s'espera?

24. Una empresa està fent un procés de selecció entre dos candidats per a un lloc de treball. Les valoracions que han obtingut a les entrevistes i proves que se'ls han fet són les següents:

Aspecte	Candidat A	Candidat B
Experiència	8	7
Coneixements	6	7
Psicotècnic	4	5

Si l'empresa dóna una importància del 60% a l'experiència, del 25% als coneixements i del 15% a l'habilitat psicotècnica, quin dels dos candidats escollirà?

25. S'ha fet un estudi sobre l'edat dels vianants atropellats en un cert municipi durant el bienni 2010-2011. La taula següent mostra alguns dels resultats obtinguts, en funció del sexe de la persona:

Edat	Home	Dona
Mitjana	40,55	46,96
Mediana	37,50	53,00
1r quartil	13,62	19,50
3r quartil	67,07	70,82
Mínim	1	2
Màxim	95	87
Nombre de casos	194	167

- Seleccioneu la informació rellevant i construïu un diagrama de caixa múltiple (un per a cada sexe). Esbrineu si existeixen *outliers* o observacions extremes.
- Comenteu les diferències que s'observen entre els sexes.
- Pel que fa als infants, quin dels dos sexes presenta més perillositat?
- En vista dels diagrames de caixa, podríeu deduir aproximadament la forma dels histogrames de freqüències? Són campaniformes?
- Calculeu, si és possible, la mitjana, la mediana, el primer quartil, el tercer quartil, el mínim i el màxim de l'edat del total de vianants atropellats.

26. Suposem que tenim n observacions $x_1, x_2, \dots, x_n$ d'una variable quantitativa amb mitjana aritmètica $\bar{X}$. Suposem que afegim una observació $x_{n+1} = k \cdot \bar{X}$, on k és un escalar més gran que 1. És possible saber com variarà la mitjana aritmètica quan afegim aquesta dada a les anteriors? Busqueu com serà aquesta variació i després trobeu un contraexemple per mostrar que la mediana es comporta diferent.

27. Tenint en compte la informació següent sobre la distribució d'una certa variable, creieu possible que la mitjana aritmètica d'aquesta variable sigui igual a 19? I que la desviació típica sigui igual a 15? Raoneu les vostres respostes.

Min = 5	$Q_1 = 10$	Me = 20	$Q_3 = 30$	Max = 35
---------	------------	---------	------------	----------

28. Calculeu la mitjana i la variància de cadascun dels conjunts de dades següents. Què en podeu concloure?

a)	10,	20,	30,	40,	50,	60
b)	10,	10,	10,	60,	60,	60
c)	-130,	20,	30,	40,	50,	200

29. Poseu exemples de dades que compleixin les característiques següents:

- Sis observacions de manera que la mitjana aritmètica sigui igual a 10.
- Cinc observacions de manera que la variància sigui igual a 0.
- Quatre observacions de manera que la variància sigui igual a 5.
- Tres observacions de manera que la mitjana aritmètica sigui 10 i la variància sigui 5.

30. Donades les n dades que es presenten a continuació:

0,	1,	0,	1,	0,	1,	...
----	----	----	----	----	----	-----

Calculeu quant valen en general $\bar{X}$, Me i S^2 , distingint si n és parell o senar. Què passa quan el nombre de dades n tendeix a infinit?

31. Una regió es divideix en tres zones, cadascuna amb el mateix nombre d'habitants. A la primera zona hi ha un habitant per telèfon; a la segona, dos habitants per telèfon, i a la tercera, tres habitants per telèfon. Quina és la mitjana d'habitants per telèfon de la regió sencera? Quina mesura de tendència central heu calculat?

32. Les dades següents corresponen als països africans amb costa al Mediterrani:

País	Població (milions hab.)	Esperança de vida	Morts per cada mil habitants
Marroc	32,3	72	6
Algèria	36,0	73	5
Tunísia	10,7	75	6
Líbia	6,4	75	4
Egipte	82,6	73	5
Total			

Font: 2011 Population Reference Bureau

Completeu la darrera filera de la taula, tot indicant quin tipus de mesura feu servir.

33. Demostreu les afirmacions següents:

- a) La mitjana aritmètica és igual a la mitjana d'ordre 1.
- b) La mitjana harmònica és igual a la mitjana d'ordre -1.

Nota: La mitjana geomètrica és igual a la mitjana d'ordre 0, però no en farem la demostració.

34. Suposem que tenim n dades de dues variables X i Y quantitatives que s'obtenen de la manera següent:

$$X_i = \begin{cases} 0, & \text{si } i \text{ és senar} \\ 1, & \text{si } i \text{ és parell} \end{cases} \quad Y_i = \begin{cases} 0, & \text{si } i \text{ és senar} \\ 1, & \text{si } i \text{ és parell} \end{cases}$$

Considerem ara unes altres dades z_i i t_i , per $i = 1, \dots, n$, que s'obtenen com $z_i = x_i + y_i$ i $t_i = x_i \cdot y_i$. Calculeu quant valen la mitjana i la variància de les variables Z i T .

35. Considerem els n números següents: $k+1, k+2, \dots, k+n$. Calculeu-ne la mitjana aritmètica i la variància, en funció de k i n .

36. Donades les empreses multinacionals següents, es coneixen dades sobre les quantitats de vendes efectuades el 2011 en milions de dòlars. Feu un comentari descriptiu d'aquestes xifres, tot indicant quines mesures quantitatives podríeu utilitzar. A més, feu servir una mesura de localització per a indicar el posicionament de Chevron respecte a les altres.

Empresa	Vendes
Wal-Mart	421.849
Exxon	354.674
Chevron	196.337
Conoco Phillips	184.966
Fannie Mae	153.825
General Electric	151.628
Berkshire Hathaway	136.185
General Motors	135.592
Bank of America Corp	134.194
Ford Motor	128.954

37. La mitjana aritmètica d'una distribució campaniforme és 30 i la seva variància és 16. Construïu un interval que contingui almenys el 95% de les observacions.

38. Sigui una distribució (x_i, n_i) $i = 1, \dots, k$ amb les característiques següents:

$$\bar{X} = k_0, \quad Mo = k_1, \quad S^2 = k_2, \quad n = N$$

Determineu aquestes mateixes mesures per a la distribució $(c_1 \cdot X_i + c_2, c_3 \cdot n_i)$ $i = 1, \dots, k$ on k_2 és una constant positiva, c_3 és un enter estrictament positiu i c_1, c_2, k_0 i k_1 són constants arbitràries. Raoneu la resposta.