



UNIVERSITAT DE
BARCELONA

Facultat d'Economia
i Empresa

TEMA 1. Introducció a l'estadística

Material de suport docent de les sessions teòriques elaborat a partir de la bibliografia referenciada al Pla docent de l'assignatura

Estadística econòmica i empresarial I

Grau d'Economia

Facultat d'Economia i Empresa

Universitat de Barcelona

Autora: Elisabet Motellón

Professora Associada

Departament d'Econometria, Estadística i Economia Aplicada

Objectiu del Tema 1:

“Aquest bloc temàtic té com a objectiu introduir l'estudiant en la disciplina estadística tot introduint el **concepte i contingut de l'estadística** i les seves diferents branques. També s'hi descriu quin **tipus d'informació** és susceptible de ser analitzada mitjançant tècniques estadístiques i es presenten els diferents **tipus de dades** i les maneres d'obtenir-les.

Els **conceptes clau** que integren el bloc temàtic són els següents: estadística, fenomen aleatori, fenomen determinista, estadística descriptiva, teoria de la probabilitat, inferència estadística, població, mostra, variable, dades (qualitatives, quantitatives, contínues, discretes, unidimensionals, bidimensionals, multidimensionals, de tall transversal i de sèrie temporal), errors de mesura, escales de mesura, fonts estadístiques, mostreig, cens i enquesta.”

Pla docent de l'assignatura

ÍNDEX

1.1. Objecte de l'estadística

- Ciència empírica
- Fenòmens aleatoris versus deterministes
- Estadística descriptiva, teoria de la probabilitat i inferència estadística
- Població versus Mostra

1.2. El nivell observacional d'una variable

- Dades, atributs i variables
- Erros de mesura

1.3. Tipologia de dades

- Qualitatives i quantitatives (discretes i contínues)
- Unidimensionals, bidimensionals i multidimensionals
- Tall transversal (*cross-section*) i sèries temporals
- Escales de mesura: nominal, ordinal, d'interval i de proporcions

1.4. Obtenció de dades

- Fonts estadístiques: cens i enquesta
- Mostreig
- Secret estadístic

1.1. Objecte de l'estadística

- Ciència empírica
- Fenòmens aleatoris versus deterministes
- Estadística descriptiva, teoria de la probabilitat i inferència estadística
- Població versus Mostra

1.1. Objecte de l'Estadística

Què és l'estadística?

Diferents accepcions:

- Informació numèrica (recopilació i publicació)
- Ciència encarregada d'utilitzar (*recopilació, organització, resum, presentació i anàlisi*) dades per tal d'ajudar a prendre les decisions d'una manera més efectiva.

1.1. Objecte de l'Estadística

Què és l'estadística?

“La estadística és una ciència que comprèn la recopilació, tabulació, anàlisi i interpretació de les dades quantitatives i qualitatives”

D. Kennedy-Neville

“La Estadística és un camp d'assaig en el qual es recullen i analitzen les dades amb el propòsit de treure conclusions (...) proporciona instruments per a la presa de decisions quan prevalen condicions d'incertesa”

William Guenther

“La Estadística és la ciència que s'ocupa de la recopilació, tabulació, anàlisi, interpretació i presentació de dades quantitatives”

D. D. H. Besterfield

1.1. Objecte de l'Estadística

Què és l'estadística?

Concepte acadèmic-científic:

Ciència q utilitza els números per a l'estudi de les lleis que depenen de l'atzar, tractant de descobrir mitjançant el raonament inductiu la causa al que obeeix el model particular analitzat

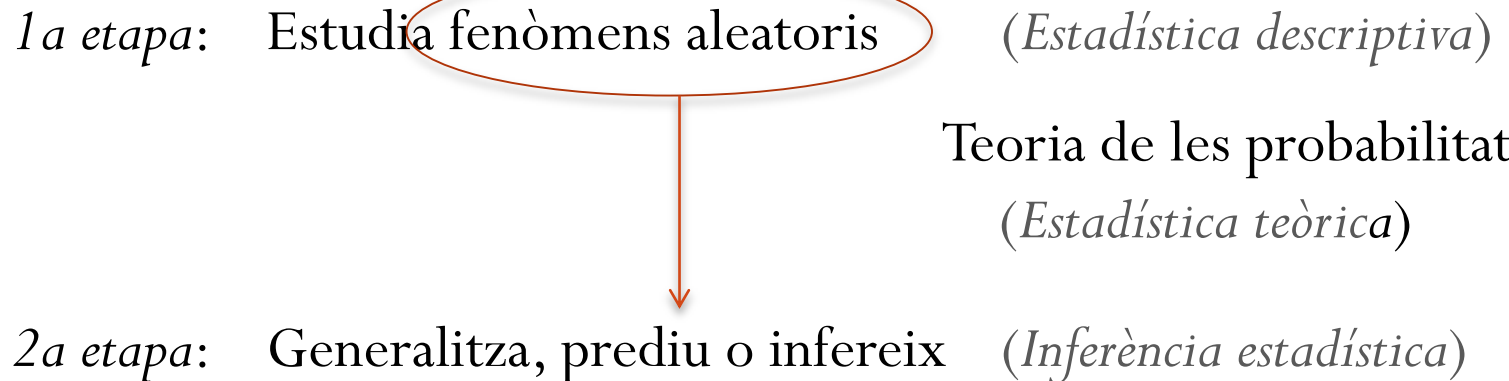
Fenòmens aleatoris: realitzats en = condicions, efectes \neq (deguts a l'atzar)

1.1. Objecte de l'Estadística

Què és l'estadística?

Concepte acadèmic-científic:

Ciència q utilitza els números per a l'estudi de les lleis que depenen de l'atzar, tractant de descobrir mitjançant el raonament inductiu la causa al que obeeix el model particular analitzat




1.1. Objecte de l'Estadística

Fenomen aleatori i Mètode Científic

Fenòmens aleatoris, aquells que realitzats en = condicions tenen efectes \neq degut a l'atzar.

Dualitat en la interpretació de la *Llei de l'atzar*:

- 1) Realitzats en = condicions no es té garantia d'obtenir idèntic resultat
- 2) Resultats \neq pq les condicions no són idèntiques
(*Atzar: incapacitat de repetir experiències idèntiques*)

Mètode científic  Fenòmens aleatoris passen a deterministes

Aprofundeix en els coneixement
de les causes q els expliquen

1.1. Objecte de l'Estadística

Característiques bàsiques

1. Ciència empírica (no formal) social

Ciències classificades en:

- C. Formals: Estudi de fenòmens abstractes que estan en la ment humana (lògica, matemàtiques...)
- C. Empíriques: 2 tipus
 - C. Naturals: estudi fenòmens que es donen en la naturalesa (química i física)
 - C. Socials: estudi de fenòmens relacionats amb l'activitat dels éssers humans en societat (economia, estadística, dret...)

1.1. Objecte de l'Estadística

Característiques bàsiques

1. Ciència empírica (no formal) social
2. Objectiu: característiques, fets o fenòmens reals aleatoris.
Però no tots els fenòmens reals son objecte de l'estadística

Fenòmens reals:

- Deterministes o causals: Coneixem a priori els resultats q es produiran pq coneixem les causes que ho condueixen (Llei i comportament q segueixen)
- Aleatoris o incerts: Caracteritzats per tres elements
 1. Resultat impredecible a priori (depenen de l'atzar)
 2. Compleixen regularitat estadística (pauta de comportament)
 3. Repetits ∞ vegades sota condicions =, o aprox. =

Q factors incontrolable: no predir amb certesa (ni repetir-se en = condicions)

1.1. Objecte de l'Estadística

Característiques bàsiques

1. **Ciència empírica** (no formal) **social**
2. Objectiu: característiques, fets o **fenòmens reals aleatoris**.
Però no tots els fenòmens reals son objecte de l'estadística

Gràcies als Mètodes estadístics, podem prendre **decisions** quan aquestes s'adapten en un **ambient d'incertesa** (fenòmens amb resultat impredecible a priori) sempre que aquesta incertesa pugui ser mesurada en termes de **probabilitat** (compleixen regularitat estadística)

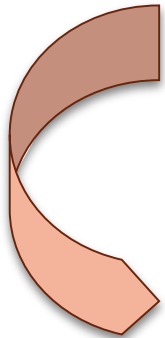
1.1. Objecte de l'Estadística

Anàlisi estadístic

ETAPES

1. Plantejament del problema
2. Recollida d'informació (obtenció dades)
3. Descripció (ordenació i presentació dades) **Estadística descriptiva**
4. Inferència estadística
 - Anàlisi o construcció del model
 - Estimació i contrast d'hipòtesis
 - Validació
 - Aplicació del model o predicció**Estadística inferencial**

Estadística teòrica
(1^a probabilitat)



(anàlisi detallat de les etapes)

1.1. Objecte de l'Estadística

Anàlisi estadístic. Etapa 1 i 2

1. Plantejament del problema

- Definir objecte d'estudi (fenomen)
- Decidir: Població o mostra

2. Recollida d'informació (obtenció dades)

- Importants conseqüències
 - Mala recollida ⓘ \Rightarrow biaix inicial \Rightarrow conclusions errònies
- \neq mètodes d'obtenció:
 - Directe o Indirecte
 - Exhaustiu (cens) o Parcial (mostra, enquesta)

1.1. Objecte de l'Estadística

Anàlisi estadístic. Etapa 3 i 4

3. **Descripció.** Destacar peculiaritats (anàlisi exploratòria):

- Depurar anomalies (errors)
- Construir gràfics
- Obtenir mesures que resumeixin ⓘ (posició, dispersió, concentració, forma, etc.)
 - Mesures paràmetres (context població) ⓘ fi anàlisi estadístic
 - Mesures estadístics (context mostra) ⓘ 4a etapa

4. **Inferència estadística.** Respondre al problema plantejat a partir dels estadístics

- Contrastos d'hipòtesis
- Prediccions

1.1. Objecte de l'Estadística

Població vs Mostra

- **Població (col·lectiu)**

Conjunt total d'individus (elements) que tenen en comú alguna característica q pretenem estudiar. Totalitat d'individus sobre els q desitgem ⓘ (persones, àrees geogràfiques, empreses, objectes, etc.)

- **Mostra**

Subconjunt representatiu d'individus (elements) extrets de la població del com estudiarem alguna característica o variable.

- Extracció aleatòria

- Grandària mostral

} Disseny de la mostra
(conclusions correctes o errònies)

1.1. Objecte de l'Estadística

Població vs Mostra

- 2 tipus de població
 1. Finita: Núm. finit d'individus → controlable, fàcil maneig i accés.
Normalment es treballa directament amb la població
Empreses instal·lades en un polígon industrial
CCAA o províncies d'un Estat en estudi d'economia regional
 2. Infinita: ∞ individus o finit però no controlable.
Es treballa amb mostres
Habitants d'un país
Població activa d'una regió
Llançament de dau, moneda...

1.2. El nivell observacional d'una variable

- Dades, atributs i variables
- Errors de mesura

1.2. Nivell observacional d'una variable

- En f(observacions) i la seva cobertura

1. Exhaustiva: Enumeració completa d'una població (cens)

Cens de població

2. Parcial: Mostreig, estudi de la mostra. Observacions parcials de part de la població per inferir d'aquestes les característiques poblacionals

Enquestes com l'EPA

3 motius per treballar amb mostres:

1. No hi ha accés a \forall individus de la població
2. Pq en cas de tenir accés, \uparrow costos en termes de temps
3. Possibles errors per tractar volum elevat de dades

1.2. Nivell observacional d'una variable

- En f(fonts) de les observacions: mètodes d'obtenció de dades
 1. Indirecta: Fonts estadístiques ja existents. Investigador utilitza dades observades per un tercer per deduir fenòmens
EPA, dades IPC, Enquesta de Costos Laborals, Enquestes d'AATT...
 2. Directa: Dades estadístiques originals per contacte directe de l'investigador amb el fenomen. Eficax i precisa però no sempre es pot aplicar per l'elevat cost
Anotació contingut documents per auditoria, infraestructures d'un municipi

Enquesta: mètode més utilitzat per recopilar ⓘ de persones.
Important: disseny del qüestionari i treball de camp

1.2. Nivell observacional d'una variable

- = població pot utilitzar-se per analitzar \neq problemes

Població activa: atur, dotació HK, temporalitat, tipus d'ocupació, sectors productius, etc.

- Població \Rightarrow mostra (s'amplia objecte d'estudi)

Mostra \Rightarrow població (es redueix l'objecte d'estudi)

1.2. Nivell observacional d'una variable

- **Variable**: Qualsevol característica que desitgem estudiar dels individus d'una població o mostra. Atribut o variable categòrica: característiques qualitatives.

Representades algebraicament amb una lletra majúscula, p. e. X

- **Dades**: observacions d'una variable (característica)

Representades amb lletres minúscules, p. e. x_1, x_2, x_3

$X = \text{Salari}$

algunes observacions que serien $x_1 = 950, x_2 = 830, x_3 = 880$

$Y = \text{nivell d'Estudi}$

i algunes observacions són $y_1 = \text{Primari o inferior}, y_2 = \text{Secundari}, y_3 = \text{Universitari}$

1.2. Nivell observacional d'una variable

Objectiu: “exploració de les dades extretes d'una població (mostra), sintetitzar ⓘ i validar-la per fer-la extensible a la població”

- Treball amb mostres: Dades observades no són perfectes (errors)

$$\begin{array}{ccc} \text{Dades} & = & \text{Veritat} & + & \text{Fortuïtat} \\ \text{(allò que tenim)} & & \text{(allò que volem)} & & \text{(Imperfeccions i} \\ & & & & \text{errors q destorben)} \end{array}$$

- Mostres grans tendeixen a reproduir les característiques poblacionals PERÒ \neq mostres llancen resultats \neq

	Tipus de contracte		
	Indefinit	Temporal	
Probabilitat:	0.75	0.25	% mostrals \neq entre si, i respecte als poblacionals però “connectats entre si” i amb la població per llei de probabilitat q regeix la població
Mostra 1:	0.78	0.22	
Mostra 2:	0.74	0.26	

1.3. Tipologia de dades

- Qualitatives i quantitatives (discretes i contínues)
- Unidimensionals, bidimensionals i multidimensionals
- Tall transversal (cross-section) i sèries temporals
- Escales de mesura: nominal, ordinal, d'interval i de proporcions

1.3. Tipologia de dades

- En f (moment del temps) en què són observades
 1. Sèrie temporal: Conjunt de dades registrades al llarg d'un període de temps. Fan referència a + d'un període.
 - Taxa d'atur trimestral dels últims 20 anys
 - W mensual dels darrers 10 anys
 - PIB des de 1980 a 2019
 2. Tall transversal (*cross-section*): dades reunides de diferents individus en un mateix període de temps
 - Anys d'educació dels aturats (Ilt 2018 EPA)
 - W anual dels assalariats de la construcció (2018)
 - Característiques d'una empresa en el període t
 3. Dades de Panell: Sèrie temporal + Tall transversal

1.3. Tipologia de dades

- En f(núm. característiques o variables) observades
 1. Unidimensionals: S'observa 1 característica del individu
Sector d'activitat dels ocupats a Espanya
Productivitat de les empreses catalanes
 2. Bidimensionals: S'observen 2 variables del individu
Sector d'activitat i ocupació dels ocupats a Espanya
Productivitat i despesa en innovació de les empreses catalanes
 3. Multidimensionals: S'observen + de 2 variables del individu
Sector d'activitat. Ocupació, HK i W dels ocupats a Espanya
Productivitat, despesa en innovació i grandària de les empreses catalanes

1.3. Tipologia de dades

- En f(naturalesa) de la variable observada

1. Dades quantitatives: Observació de variable quantitativa. Són susceptibles de ser mesurades i quantificades (expressades directament amb números)

- Contínues: Pot prendre qualsevol valor numèric
Edat, salari, productivitat, temperatura, alçada, etc.
- Discretes: Valors sencers o fraccions específiques
Núm. membres de la llar, grandària empresarial, etc.



Discretalitzar

2. Dades qualitatives: Observació de variable qualitativa (atribut). Solament poden classificar-se com a posseïdores o no de certa qualitat o propietat (no quantificar, no directament amb números)

Escales de mesura: ordinal i nominal

1.3. Tipologia de dades

Escales de mesura

Mesurament: procés d'assignació de numerals a objectes (categories) d'acord amb certes regles

Característiques del procés

- Vàlid: mesurar el q es pretén mesurar
- Fiable: = procediment \neq vegades sota idèntiques condicions proporciona mateixos resultats
- Exhaustiu: Tenir suficients categories en les quals es poden classificar les dades considerades
- Categories mútuament excloents: Només ha de ser possible classificar cada cas en una sola categoria
- Precís: Major número de distincions possibles

1.3. Tipologia de dades

Escales de mesura

Dades classificades en f(escala de mesura)

1. Mesurament nominal: Dades només és classifiquen
 - Ocupació, sector d'activitat, etc.
2. Mesurament ordinal: Les dades s'ordenen
 - Estatus social, nivell d'ocupació, etc.
3. Mesurament d'interval: Diferència significativa entre valors
 - Temperatura, salari, etc.
4. Mesurament de proporcions: Punt 0 significatiu i raó entre valors
 - Número patents, AATT, etc.

(anàlisi detallat de les escales)

1.3. Tipologia de dades

Escales de mesura

1. Mesurament nominal

- Nivell inferior de mesurament pq les dades només és classifiquen
- No s'assumeix cap supòsit sobre els relacions entre els valor d'una variable (cada valor defineix una categoria diferent i és només un codi)

Sector d'activitat, ocupació, lloc de residència, nacionalitat, etc.

Exemple mesurament nominal: Sector d'activitat de 20 treballadors/es

Automoció	Alimentació	Comerç	Comerç
Automoció	Comerç	Sanitat	Automoció
Química	Construcció	Comerç	Alimentació
Construcció	Construcció	Construcció	Educació
Construcció	Automoció	Educació	Construcció

Associem valors numèrics a cada sector:

Automoció:	1
Comerç:	2
Química:	3
Construcció:	4
Sanitat:	5
Educació:	6
Alimentació:	7

→ Núm. indica sector segons codi preestablert

Substituïm el nom del sector d'activitat pel seu codi:

1	7	2	2
1	2	5	1
3	4	2	7
4	4	4	6
4	1	6	4

1.3. Tipologia de dades

Escales de mesura

2. Mesurament ordinal

- Nivell de mesurament superior al nominal pq és possible ordenar categories d'acord amb algun criteri (“major que”, “menor que”, “més preferit”, etc.)
- Distribuir individus en categories ordenades (jerarquia)
- Ordenació única propietat matemàtica aplicable
- Indiferent el número assignat a la categoria si manté la jerarquia
Estatus social, nivell educatiu, etc.

Exemple mesurament ordinal: Nivell educatiu de 20 individus

Primari	Primari	Primari	Universitari
Primari	Secundari	Universitari	Secundari
Universitari	Primari	Primari	Primari
Secundari	Secundari	Secundari	Universitari
Primari	Primari	Universitari	Secundari

Associem valors numèrics a cada nivell d'estudis:

Primari:	1	→ Núm. indica jerarquia
Secundari:	2	
Universitari:	3	

Substituïm el nom del nivell educatiu pel seu codi:

1	1	1	3
1	2	2	2
3	1	1	1
2	2	2	3
1	1	3	2

1.3. Tipologia de dades

Escales de mesura

Dades numèriques també poden ser tractats qualitativament si ho demanda el problema

Edat (Mesurament d'interval) \Rightarrow Edat de treballar

Important per aplicar mètodes estadístics adequats

**Software estadístic: operar matemàticament amb dades qualitatives?!
obtenir estadístics numèrics de dades nominals?!**

1.4. Obtenció de dades

- Fonts estadístiques: cens i enquesta
- Mostreig
- Secret estadístic

1.4. Obtenció de dades

Fonts estadístiques

2 formes d'obtenir ⓘ d'una població:

1. **Cens:** Relació exhaustiva de tots els elements d'una població (quan existeixi i sigui possible).

Motius per evitar el seu ús:

- Econòmics (costós monetàriament i en temps)
- Individus sotmesos a proves destructives
- Població infinita o molt gran (incontrolable)

Solució:
} mostres (representatives)

2. **Enquesta:** Permet obtenir informació d'un col·lectiu investigant exclusivament una part (mostra) que representa a tota la població. Investigació realitzada per mostreig.

→ ≠ procediments estandarditzats

1.4. Obtenció de dades

Mostreig

Eina de la recerca científica que determina quina mostra s'ha d'examinar per inferir el comportament poblacional

↳ Diferents **tipus de mostreigs**

Problema: **Error de mostreig**

1.4. Obtenció de dades

Secret estadístic

“Prohibició de difondre estadístiques o dades en què no es conserva l'anonimat de cada unitat individual a la qual es refereix la ⓘ” (INE).

Mètodes per garantir privacitat: en origen o en destinació

- Origen: difondre les dades sense els elements més identificadors

Anonimització de microdades, agrupació d'unitats, etc.

- Destinació: limita l'accés als usuaris finals (limitacions s'especifiquen en les llicències d'ús)

Ús individual de dades o prohibició de difondre si no s'ha signat llicència.
Control de publicacions realitzades. Prohibir o limitar la còpia, etc.

Motiu: Confiança del públic. Enquestes han de ser fiables (contestar sincerament si sospita que respostes no són confidencials?)

Com a RECORDATORI final ...

- Aquest material s'utilitza com a guia per realitzar les sessions presencials a l'aula. NO són els apunts de l'assignatura.
- Reviseu amb atenció el document “Conceptes bàsics 1” penjat al campus, en ell trobareu un major detall de la informació referida al tema 1 publicada al Pla docent (contingut, objectius, conceptes claus, bibliografia, etc.)
- No oblideu ampliar els coneixements. Trobareu la bibliografia recomanada a la biblioteca de la Facultat, conjuntament amb un ampli ventall de manuals d'estadística també interessants.



UNIVERSITAT DE
BARCELONA

Facultat d'Economia
i Empresa