

# Treballs de recerca@

## CÓM ES FA UNA ENQUESTA?

A partir de:

Cochran, W.G.: Técnicas de muestreo. Ed. CECSA. 1992.

Apunts de mostratge estadístic i recollida de dades. Diplomatura d'estadística. Facultat de matemàtiques i Estadística. UPC. Tomàs Aluja Banet

### INDEX

- **INTRODUCCIÓ**
  - **TÈCNiques DE MOSTRATGE PROBABILÍSTIC**
    - **MOSTRATGE ALEATORI SIMPLE**
    - **MOSTRATGE PER PROPORCIIONS I PERCENTATGES**
    - **MOSTRATGE ALEATORI ESTRATIFICAT**
    - **MOSTRATGE SISTEMÀTIC**
    - **MOSTRATGE PER CONGLOMERATS**
    - **DETERMINACIÓ DEL NOMBRE DE MOSTRES**
      - **NO PROBABILÍSTICS**
  - **TIPUS FONTS D'ERROR EN LES ENQUESTES**
  - **TIPUS FONTS D'ERROR EN LES ENQUESTES**
- 

### INTRODUCCIÓ

Un estudi estadístic tracta de la recollida, síntesi (o anàlisi) i interpretació de fets numèrics, que anomenem dades, per tal de valorar la informació continguda en aquestes.

#### Classificacions dels estudis estadístics

- Segons els tipus de dades
  - observacionals: l'investigador no intervé en la seva producció
  - experimentals: l'investigador controla la producció de dades
- Segons els tipus d'anàlisi
  - probabilista: fa intervenir la component aleatòria que contenen les dades i permet fer ús de tècniques estadístiques de modelització (per regressions)
  - no probabilista: no utilitza el component aleatori de les dades, permet utilitzar eines de tipus descriptiu o exploratori
- Segons la seva interpretació
  - descriptiu: descriu les dades presents a l'estudi
  - explicatiu: valora la influència de les variables

#### Definició d'objectius

- Formular el problema de forma precisa per fixar el propòsit central de l'estudi i apareixen una sèrie de propòsits secundaris
- Definir la població i els individus objecte de l'estudi

#### Fons d'informació

- Secundàries: son dades que ja existeixen perquè les han recollit organismes com l'IDESCAT o l'INE
- Primàries: Cal fer la recollida de dades. Es poden obtenir a partir de:
  - entrevista: cal disseny qüestionari o enquesta
  - experimentació: cal disseny experimental
  - observació: cal definir què i com s'observa

#### Tipus de recollida de dades

- Disseny transversal: es prenen les dades en un determinat moment, és el més habitual
- Disseny factorial: pensats per comparar dades de diferents situacions
- Estudis "abans" - "després": serveix per comparar l'evolució de l'opinió davant un determinat fet. Ex: què s'opina inicialment d'un producte comercial i com ha canviat l'opinió sobre aquest després de veure un vídeo informatiu.
- Estudis longitudinals: s'agafa una mostra i es va seguint en el temps. Ex: A quina hora van a dormir les criatures i se segueix estudiant aquest hàbit al llarg del anys.  
Recollida de dades segons enquestes o qüestionaris .

## Recollida de dades segons enquestes o qüestionaris

### Alguns usos de les enquestes per mostratge

Donen informació de gran interès sobre la planificació nacional en àrees de gran importància. També s'utilitzen en estudis més especialitzats, com per exemple, les vacances anuals, les causes de divorci, el consum domèstic d'aigua...

Les enquestes s'utilitzen per a conèixer les opinions, desitjos, creences, condicions de vida de la població que és vol estudiar. Consisteixen en formular una sèrie de preguntes a un subconjunt representatiu de la població, a una mostra representativa

Els nostres coneixements, actituds i accions estan basats en gran part sobre mostres. Aquest fet és tan real en la vida quotidiana com en la investigació científica. D'aquesta manera l'opinió que tenim d'alguna cosa, o fet, es basa en el contacte mantingut amb aquesta unes poques vegades.

**Població o "target"** : Tots els elements del grup que es vol estudiar.

**Unitat de mostratge**: Nombre d'elements de la població no solapats que s'estudiaran.

**Unitat d'anàlisi**: Individu del que es vol obtenir informació.

**Mostra**: Conjunt dels elements d'anàlisi.

### Avantatges del mètode de mostratge

1. Menor cost al obtenir les dades d'una petita fracció.
2. Al ser una petita fracció, les dades es recullen amb més rapidesa.
3. El mostratge és més exacte a causa del reduït volum de treball, que es pot realitzar amb més personal qualificat i més intensivament.

### Etales principals en una enquesta per mostratge

1. Exposició clara dels objectius.
2. Definir la població de la qual es tria la mostra.
3. Comprovar que totes les dades són pertinents a l'enquesta i que no s'omet cap dada essencial.
4. Definir els mètodes de mesura (qüestionari autoadministrat, entrevistes...) i els mitjans pels quals es realitzen (telèfon, correu, visita personalitzada...).
5. Establir un marc. Subdivisió de la població en **unitats de mostratge** (família, barri...).
6. Seleccionar la mida de la mostra. Depèn de la grandària de la població i de la variabilitat (p.e. si tothom pensés igual tant sols cal fer una enquesta, si tothom pensés diferent la mostra hauria de ser tota la població).
7. Establir un grau de precisió ja que les dades de l'enquesta provoquen incertesa al mesurar només una part de la població.
8. Enquesta pilot (opcionalment).
9. Organització del treball de camp.
10. Editar els qüestionaris per a corregir errors i determinar els tipus de càlculs que es duran a terme. Aquests càlculs ens permetran realitzar les estimacions.
11. Aconseguir el màxim d'informació de la població per a obtenir unes estimacions més exactes.

### mostratge probabilístic i no probabilístic

En els mostratges probabilístics tots els elements de la població tenen una determinada probabilitat de ser escollits.

Tots els mostratges probabilístics permeten extreure conclusions sobre la població de la que provenen. Mitjanes, desviacions, intervals de confiança, percentatges d'error...

En els mostratges no probabilístics no es pot determinar la probabilitat que cada element de la població té de pertànyer a la mostra. No permeten fer especulacions sobre la població de la que provenen, si no és comparant-les amb enquestes del mateix tipus fetes de manera probabilística.

**Població o "target"** : Tots els elements del grup que es vol estudiar.

**Unitat de mostratge**: Nombre d'elements de la població no sobreposats que s'estudiaran.

**Unitat d'anàlisi**: Individu del que es vol obtenir informació.

**Mostra**: Conjunt dels elements d'anàlisi.

## TÈCNiques DE MOSTRATGE. PROBABILÍSTICS.

### MOSTRATGE ALEATORI SIMPLE

Mètode de selecció de  $n$  unitats d'un conjunt  $N$  de manera que cadascuna de les mostres tinguin la mateixa oportunitat de ser escollides.

### MOSTRATGE PER PROPORCIIONS I PERCENTATGES

En determinades ocasions volem estimar el número total, la proporció, o el percentatge d'unitats en la població, que posseeix alguna característica o atribut, o que s'engloben dins d'alguna classe definida.

La classificació es pot incloure a l'enquesta (sí, no) o pot introduir-se després (preguntar l'edat i al final determinar si està jubilat/da o no)

### MOSTRATGE ALEATORI ESTRATIFICAT

En el mostatge estratificat, la població de  $N$  unitats es divideix en subpoblacions anomenades estrats que volem que estiguin representats a la mostra. Passos:

1. Determinar els estrats (p.e. un barri, un districte)
2. Nombre d'elements de la mostra en cada estrat
3. Es pren una mostra aleatòria simple en cada estrat.

### MOSTRATGE SISTEMÀTIC

Suposem que les  $N$  unitats de la població es numeren de 1 a  $N$  en un cert ordre. Per a escollir una mostra de  $n$  unitats, prenem una unitat al atzar entre les  $k$  primeres, amb  $k$  entre 1 i  $n/n$ , i a continuació prenem les següents a intervals de  $k$ . La selecció de la primera unitat determina tota la mostra, que es denomina mostra de totes les  $k$ -èsimes unitats.

### MOSTRATGE PER CONGLOMERATS

S'utilitza quan es té una idea clara del total de la població, o quan seria massa costós generar-ne la llista.

Les unitats conglomerats (municipis, ciutats...) contenen números diferents d'elements o subunitats (regions geogràfiques, escoles, habitatges, persones). Un cop determinats els conglomerats i els individus a seleccionar dins de cadascun es pot aplicar qualsevol dels mètodes anteriors.

### DETERMINACIÓ DEL NOMBRE DE MOSTRES

El pas previ a iniciar una enquesta per mostatge és enunciar la hipòtesi de partida. En general aquesta hipòtesi s'expressa en termes dels límits d'error esperats o tolerables. Aquesta decisió l'han de prendre les persones que utilitzaran els resultats de l'enquesta, basant-se en experiències anteriors.

Generalment es mesura més d'un atribut o característica en una enquesta per mostatge: en ocasions, el nombre d'atributs és gran. Si s'estipula un grau de precisió per a cada atribut, els càlculs ens conduiran a una sèrie de valors conflictius de  $n$ ; un per a cada atribut. Pel tant s'ha de trobar un mètode per reconciliar aquests valors.

En ocasions és clar que  $n$  s'ha de reduir dràsticament, i llavors és necessari prendre una decisió difícil, que és la de procedir amb una mostra molt més petita, la qual cosa redueix la precisió, o bé, abandonar els esforços fins a comptar amb majors recursos.

En planejar una enquesta per mostatge, sempre s'arriba a una etapa on cal prendre una decisió respecte a la mida de la mostra. Aquesta decisió és important. Una mostra massa gran implica un malbaratament de recursos i una de molt petita disminueix la utilitat dels resultats. La decisió no sempre pot prendre's satisfactòriament; sovint no disposem de la suficient informació per saber si la mida de la mostra seleccionada és l'òptima. La teoria del mostatge proporciona un marc dins del qual es pot pensar intel·ligentment respecte a aquest problema.

A continuació veurem un exemple hipotètic en el qual es ressalten els passos necessaris per arribar a una solució. Un antropòleg està preparant un estudi sobre els habitants de certa illa. Entre altres coses, desitja estimar el percentatge d'habitants que pertanyen al grup sanguini O. Li han assegurat la cooperació necessària per extreure una mostra aleatòria simple. Què tan gran ha de ser la mostra?

Aquest assumpte no es pot discutir sense abans contestar a una altra pregunta. Amb quin grau de precisió desitja conèixer l'antropòleg el percentatge de persones del grup sanguini O? A la pregunta, el nostre amic contesta que estarà satisfet si el percentatge no conté un error major de més o menys el 5%, és a dir que si la mostra indica que el 43% de la població és del grup sanguini O, el percentatge per a tota l'illa es trobarà amb certesa entre el 38 i el 48%.

Per evitar malentesos, és aconsellable aclarir l'antropòleg que no li podem assegurar una exactitud dins d'un 5%, llevat que s'analitzi el tipus sanguini de tots els habitants. Per molt gran que prengui  $n$ , existeix la possibilitat d'una mostra desafortunada que presenti un error major al 5% desitjat. L'antropòleg contesta fredament que ho sap, que accepta el risc d'una possibilitat en vint d'obtenir una mostra poc afortunada i que tot el que demana és un valor de  $n$  i no una classe d'estadística.

Ara estem en posició d'estimar en termes generals el valor de  $n$ . Per simplificar les coses s'ignora la **cpf** (correcció per població finita) i el percentatge de la mostra  $p$  se suposa normalment distribuït.

En termes tècnics,  $p$  ha de trobar-se en l'interval  $(P \pm 5)$ , excepte per a un cas en 20 (veieu que  $1/20=0.05$  un 5%). Ja que  $p$  se suposa normalment distribuïda al voltant de  $P$ , es trobarà en l'interval  $(P \pm 2sp)$  ( $sp$ =desviació standard), llevat d'una possibilitat entre 20. A més (atenció !!  $Q = 1 - P$ ),

$$sp = \text{SQRT}(PQ/n) \quad \text{SQRT} = \text{arrel quadrada}$$

De manera que podem escriure (creieu-vos-ho o llegiu el capítol 3, apartat 3.2 i, al capítol 4 del llibre):

$$n = 4PQ/25$$

En aquest moment apareix una dificultat comuna a tots els problemes per a l'estimació de la mida de la mostra. S'ha obtingut una fórmula per a  $n$ , però  $n$  depèn d'una propietat de la població subjecta al mostratge. En aquest cas, la propietat és la quantitat  $P$  que desitjaríem mesurar. Per tant, preguntem a l'antropòleg si ens pot donar una idea del valor que s'espera de  $P$ . Ell contesta que amb base en dades obtingudes prèviament amb altres grups ètnics, i d'acord a les seves especulacions sobre la història racial de l'illa, li sorprendria que  $P$  fos fora de l'interval del 30 al 60%.

Aquesta informació serà suficient per proporcionar una resposta útil. Per a qualsevol valor de  $P$  entre 30 i 60, el producte  $PQ$  és entre 2100 i un màxim de 2500 amb  $P=50$ . El valor corresponent de  $n$  és entre 336 i 400. Per no córrer riscos en pren 400 com a estimació inicial de  $n$ .

Ara poden reexaminar-se les hipòtesis fetes en aquesta anàlisi. Amb  $n=400$  i  $p$  entre 30 i 60, la distribució de  $p$  hauria d'estar a prop de la normal. Si es requereix o no la **cpf**, depenen del nombre d'habitants de l'illa. Si la població en supera 8000, la fracció de mostratge és menor al 5% i no es necessita ajust per a la **cpf**.

## No probabilístics

Hi ha diferents mètodes:

1. La mostra és una part de la població fàcilment accessible (p.e. es mira la qualitat del carbó de la part superior d'una vagoneta)
2. La mostra es selecciona a l'atzar.
3. Es fa una visió superficial de tota la població i s'agafa una mostra que sembla que siguin típiques (que s'acostin a la mitjana de la població).
4. S'agafa una mostra de voluntaris.

No permeten dir res de la població d'on provenen les mostres.



---

## FONTS D'ERROR EN LES ENQUESTES

1. Error mostral pel fet de triar una determinada mostra i no una altra
2. Error no mostral
  - o coverage: Produït pel fet de no incloure tota la població
  - o Error en mesurar algunes de les unitats en la mostra triada. Pot ocórrer per omissió o fracàs en la localització d'alguns individus o per la seva renúncia a contestar.
  - o Error degut a la forma en que es redacten les preguntes. És molt important dissenyar be els qüestionaris
  - o Errors de mesura en una unitat. El mitjà de mesura pot ser imprecís al rebre respostes imprecises per part d'algun individu.
  - o Errors introduïts en l'edició, codificació i tabulació dels resultats.



---

## DISSENY DE QÜESTIONARIS

### Tipus de preguntes

- Socioeconòmiques: son preguntes obligades tipus sexe, edat, població, districte, estudis, professió...
- Factuals: son preguntes molt precises, indiquen fets. Ex. tens mp3? m2 de la teva habitació ...
- Actitud o opinió: son difícils de mesurar. Val la pena posar més d'una pregunta idèntica formulada de diferent manera

### Tipus de resposta

- Obertes: son més difícils de valorar. L'enquestat pot respondre qualsevol cosa
- Tancades: És dona una, o varies, opcions. La resposta ja està definida, no es pot posar qualsevol cosa

### Escala de mesura de les variables

- Nominal: s'identifica una sola resposta de totes les possibles. Permet calcular freqüències, modes, taules...
- Ordinal: cal ordenar una determinada llista. Ex. ordenar de més conegut a menys una sèrie de marques
- Interval: es posa una nota entre 0 i 10. Permet calcular mitjanes, desviacions, correlacions...