

La mesura de la informació

Thierry Lafouge
Yves-François Le Coadic
Christine Michel

PID_00189701

Índex

1. El nombre	5
2. La mesura	7
3. Sistema d'informació i de mesura	9
3.1. Per què cal mesurar?	10
3.2. Què cal mesurar?	10
3.2.1. Les magnituds mesurables	10
3.2.2. Les escales de mesura	12
3.3. Com cal mesurar?	13
3.3.1. Els mètodes de mesura	14
3.3.2. Els instruments de mesura	14
3.3.3. Els errors de mesura	14
3.4. Com s'ha de calcular?	15
3.4.1. Del càlcul manual al càlcul informàtic	15
3.4.2. Calculadores i microordinadors	17
3.4.3. Els programes estadístics i matemàtics	19
4. Mesures i lleis	21

1. El nombre

“Només es coneix bé un problema quan és possible expressar-lo amb nombres.”

Lord Kelvin

Com a objecte principal (juntament amb les magnituds i les xifres) dels primers matemàtics, el nombre sorgeix de l'experiència quotidiana; podem explicar un llibre, dos llibres, tres llibres, etc. I són sovint les necessitats de càlcul les que es troben en l'origen de l'extensió de la noció de *nombre*. Actualment, els objectes matemàtics, com els vectors, les gràfiques, etc., són més abstractes i no tenen cap interpretació tan directa.

És el nombre present en l'experiència quotidiana d'un centre de documentació, d'una biblioteca, d'un museu, d'un centre servidor o d'un mitjà d'informació? El conte dolent que expliquem a continuació potser us portarà a pensar que no.

El conte de les mil i una biblioteques

“Quan el gran visir va decidir crear la Biblioteca del Califat transformant i modernitzant la Biblioteca del Visirat, es van examinar diverses qüestions arquitectòniques, biblioteconòmiques i financeres. La sorpresa del Gran Geòmetra encarregat de dissenyar la nova biblioteca va ser constatar que no hi havia dades sobre l'ús o la consulta de les obres de l'antiga biblioteca del visirat.

Com es pot dissenyar una biblioteca adaptada a la demanda si no se sap quina és aquesta demanda, es planyia...” (Tirèsies).

No obstant això, els nombres estan molt presents. El recompte de tot el que entra o surt d'una biblioteca, d'un centre de documentació, d'un museu, d'un centre servidor o d'un mitjà de comunicació, és una mesura molt important de l'activitat d'aquests organismes (mesures de gestió o de biblioteconomia, “d'economia” del centre, del museu o del mitjà de comunicació).

Exemples

1) Com a operació important en alguns d'aquests organismes, la classificació també va introduir un element alfanumèric, que és la signatura. Ja en l'ancestral biblioteca es trobaven pesos i mesures temporals i mètrics que no tenien, en general, cap valor informatiu. El format dels llibres encara continua essent la base per a alguns plans de classificació. Els dos formats més coneguts són el format *in quarto* (*in-4°*), en el qual el full de paper es doblega dues vegades i forma quatre quaderns de vuit pàgines, i el format *in octavo* (*in-8°*), en el qual el full d'impressió es doblega quatre vegades en vuit quaderns de setze pàgines. Aquest format apareixia en la signatura. La catalogació dels llibres antics en la biblioteca del CNAM es presenta de la manera següent:

- **signatura = 4° A 21** - Nouveau dictionnaire universel des arts et des sciences - Thomas DYGHE - Avinyó, 1756.
- **signatura = 8° C 4** - Traité de physique expérimentale et mathématique - J. B. BIOT - París, 1816.

2) En el centre de documentació i informació (CDI) d'un col·legi es pren nota de la circulació dels llibres al llarg de l'any; aquesta circulació va dels trenta-cinc als dos-cents deu préstecs per dia, però ningú no els ha analitzat.

3) En un diari regional es poden mesurar les superfícies de redacció i estudiar-ne el repartiment, però això és una cosa que no es fa habitualment.

Així, de la biblioteca de Babel...

"L'univers (que altres anomenen *la Biblioteca*) es compon d'un nombre indefinit, i tal vegada infinit, de galeries hexagonals [...]. La distribució de les galeries és invariable. Vint lleixes, amb cinc lleixes llargues per costat, cobreixen tots els costats menys dos; l'alçada, que és la dels pisos, excedeix amb prou feines la d'un bibliotecari normal [...].

[...] cada lleixa tanca trenta-dos llibres de format uniforme; cada llibre és de quatre-centes deu pàgines; cada pàgina, de quaranta línies; cada línia, d'unes vuitanta lletres de color negre. També hi ha lletres en el dors de cada llibre" (J. L. Borges).

Atès que l'home és la mesura de totes les coses, aquí es trobarà una il·lustració de la primera etapa del desenvolupament de les nocions metrològiques, la de l'antropomorfisme. El cos humà (i els seus membres) serveix com a unitat de mesura principal.

[...] a la secció d'arxius del centre de préstecs de la Biblioteca del Congrés de Washington, es pot constatar aquesta presència constant dels nombres. No obstant això, encara que hi ha molts nombres en els centres de documentació, les biblioteques, els museus, els centres servidors i els mitjans de comunicació, es constata que ni els documentalistes, ni els bibliotecaris, ni els conservadors, ni els informàtics, ni els periodistes els han utilitzat sovint per a analitzar-los.

% sol·lic.	% bytes	Bytes enviats	Sol·licituds	Secció d'arxiu
0,05	0,06	2.139.768	289	/exhibits/lawlib/...
0,04	0,10	3.934.779	233	/exhibits/russian/
0,18	0,74	27.764.173	1.057	/exhibits/scrolls/...
0,05	0,14	5.287.058	323	/exhibits/treasures/
0,01	0,00	118.931	72	/exhibits/treasures/trupscale/
0,03	0,21	7.870.117	202	/exhibits/us.capitol/
0,09	0,37	140.029.931	505	/exhibits/vatican/
0,013	0,61	34.300	82	/exhibits/wcf/

Cal desconfiar del poder de fascinació que exerceixen els nombres i les xifres simples en les organitzacions. La complexitat d'aquestes organitzacions és molt gran, per la qual cosa reduir els judicis a unes quantes xifres simples respon a unes expectatives molt potents. El llenguatge secret de l'estadística, tan atractiu en una societat que viu molt de fets i de xifres, es pot utilitzar de manera sensacionalista, per a inflar els resultats o per a simplificar fins a l'extrem.

Podeu consultar el primer exemple que apareix en el mòdul 2, "Estadística de la informació", apartat 2.3.2, que mostra les dificultats que podem trobar en interpretar resultats numèrics.

2. La mesura

El nombre és present en els sistemes d'informació, per la qual cosa és possible mesurar certes característiques d'objectes, de fets, de documents o d'informacions presents en aquests sistemes. És a dir, és possible determinar el valor de certes magnituds comparant-les amb una altra magnitud de la mateixa espècie presa com a unitat. Perquè una magnitud es pugui mesurar, cal poder definir la igualtat i la suma de dues magnituds de l'espècie que es pren en consideració.

El mètode de mesura aplicable a qualsevol magnitud G (additiva o no) consisteix a avaluar la intensitat i d'un fenomen fàcilment determinable, invariablement vinculada al valor de la magnitud G mitjançant una relació del tipus:

$$G = f(i)$$

Perquè la mesura ofereixi una seguretat total, la relació que representa la funció f ha de ser unívoca. Perquè la mesura sigui precisa, cal que el fenomen sigui apreciable. A més, el fenomen ha de ser fidel, és a dir, s'ha de poder reproduir en el temps. El paràmetre "temps" té un paper cada vegada més important dins dels nous sistemes d'informació. En aquest sentit, es pot parlar d'una de les dificultats de l'aplicació d'eines d'avaluació realment eficaces per a fer mesuraments en el Web, un mitjà en perpètua evolució; aquesta dificultat està vinculada, en part, al factor temporal.

Així, doncs, veiem que en les ciències de la informació apareix la noció de la mesura. Aquesta noció, que actualment té un paper bàsic en les ciències físiques, s'ha imposat progressivament en les primeres. Tampoc no ens hem de sorprendre que aquesta noció pugui trobar certes dificultats per a instal·lar-se en les ciències socials en general, i en les ciències de la informació en particular.

Actualment, en una societat marcada pels avenços científics i tècnics, la mesura sembla regir fins a les decisions més mínimes. Es reflecteix de múltiples maneres en les condicions de l'existència humana.

“Des de la seguretat viària fins a la purificació de l'aigua, des de la comptabilitat i les assegurances a la física de partícules o els pigments de color, des de les institucions escolars fins als processos industrials, les tècniques de mesura remetent indefectiblement a totes les modalitats de la nostra relació amb el món. Tant si la mesura és directa com estadística, la seva dimensió d'objectivitat, sense la qual no seria res, remet a les normes que alguns col·lectius humans (laboratori, escola, comunitat d'investigadors, revista, biblioteca, centre de documentació, museu, hospital, administració) elaboren per distingir millor la realitat de fenòmens que la nostra percepció immediata només aconsegueix modelitzar quan la successió dels fets observables respon a cicles regulars. I fins i tot en aquest cas, sabem que no sempre els diferents observadors es detenen en els mateixos aspectes i que no sempre divideixen la realitat de manera idèntica.

Només aconseguim mesurar quan ens allunyem de les percepcions singulars en benefici dels procediments regulats que es van haver de determinar en el seu moment i que generalment condicionen la nostra percepció de la realitat. Així, doncs, és important comprendre les bases de les tècniques de mesura actualment en vigor, les pràctiques de mesura i els seus límits en el camp de les ciències contemporànies i en l'anàlisi dels comportaments socials, i preguntar-se sobre la manera com afecten el nostre comportament quotidià.”

(*La mesure*, actes de les 7^e Entretiens de la Villette, GSI-GNDP, 1996).

Les mesures en el sector de la informació es van començar a utilitzar en el seu objecte principal, que és el llibre. Es feien recomptes dels llibres i d'altres suports escrits, i a les xifres obtingudes s'aplicaven algunes tècniques matemàtiques simples, bàsicament estadístiques. La bibliometria va néixer, doncs, amb l'objectiu de mesurar les activitats de gestió dels llibres i de les col·leccions de llibres i documents.

Tot seguit, amb el creixement de la ciència i de la tecnologia, calia donar resposta a les necessitats de l'anàlisi del desenvolupament de les activitats científiques i tècniques. Així, doncs, a la Unió Soviètica entre les dues guerres, van aparèixer la “ciència de la ciència” (*nauchno i nauchno*) i la cienciometria (i també la tecnometria), una mesura de les activitats de recerca científica i tecnològica per mitjà de les entrades o *inputs* (mà d'obra, pressupostos) i de les sortides o *outputs* (dispositius, productes, publicacions); aquí, la ciència i la tecnologia s'entenien en el sentit més ampli, des de les ciències de la matèria fins a les ciències humanes, passant per les ciències de la vida i les ciències socials. Si les mesures principals de les “entrades”, dels recursos, consistien a mesurar les despeses i la mà d'obra, les mesures de les sortides només podien ser els productes principals de les seves activitats, és a dir, les publicacions, com els articles per a activitats científiques, és a dir, la informació científica, o bé les patents que constituïen els productes indicadors d'activitats tecnològiques, és a dir, la informació tecnològica. Els articles i les patents són, en efecte, recursos bàsics per a l'avaluació de laboratoris o de països.

A continuació es va passar de l'article a la informació que contenia; va ser llavors quan va aparèixer, en la dècada dels vuitanta, la infometria, que és la mesura de les activitats d'informació, i en particular de les activitats d'informació científica i tècnica. És, abans de res, una quantificació de les informacions bibliogràfiques que té com a objectiu l'anàlisi (i per això l'ús continuat del terme *bibliometria*). La literatura científica i tècnica es presta, en efecte, a una quantificació, en el sentit que tots els articles d'una revista, tots els informes, llibres, notes, etc. contenen elements bibliogràfics paratextuals, com els títols dels articles, les patents, els noms dels autors, les paraules clau o les referències bibliogràfiques. Aquests elements, correctament disposats, constitueixen un conjunt de variables textuals, personals i temporals sobre les quals es poden fer mesures i anàlisis quantitatives i qualitatives dels temes d'investigació, de les relacions entre les disciplines i de la informació científica i tècnica. La combinació d'aquests elements d'informació ha permès elaborar indicadors quantitius i qualitius de les activitats de construcció, de comunicació i d'ús de les informacions científiques i tècniques.

3. Sistema d'informació i de mesura

L'associació d'aquests dos termes és nova. És formular la hipòtesi (que nosaltres hem fet des de fa alguns anys) que la informació pot ser l'objecte d'una ciència i que es pot estudiar de manera científica i experimental. No hi ha cap ciència ni cap tecnologia sense mesures i, cada vegada menys, sense unes mesures precises.

Però la mesura en si mateixa encara no forma part de l'univers quotidià del sistema d'informació o del centre de documentació, ni encara menys de la biblioteca, del museu, del diari o de la cadena de televisió. Per no haver volgut identificar magnituds mesurables i, correlativament, per no haver posat a punt instruments de mesura, però també per falta de rigor en caracteritzar els elements dels sistemes d'informació, actualment no hi ha cap metrologia de la informació. La responsabilitat també recau sobre la falta de preocupació per la demanda; la producció preval sobre l'ús. En els centres de documentació, per exemple, s'ha passat d'un sistema amb pocs usuaris i que emprava professionals que creien que coneixien les seves necessitats d'informació i el fons disponible, a un sistema que ha vist multiplicar el nombre d'usuaris, el nombre de títols i l'especialització cada vegada més gran de les obres. Actualment, l'enginyeria del document electrònic, que ha permès automatitzar moltes tasques, brindarà al gestor la possibilitat d'adoptar una gestió "orientada a l'usuari".

Exemples de magnituds mesurables

Fenòmens, qualitats o propietats que apareixen en diferents intensitats.

Recentment, l'arribada dels ordinadors ha posat de manifest la possibilitat de mesurar; no estem parlant solament dels recomptes i de les tècniques quantitatives de gestió que se'n deriven i que constitueixen l'essencial del que es coneix, en el camp de la documentació i de la biblioteconomia, com a *bibliometria*, i en el camp dels mitjans de comunicació com a *mediametria*, sinó també de les mesures de la informació que arriba a aquests organismes i s'hi emmagatzema o hi transita, la qual cosa es coneix com a *infometria*. Biblioteques, museus, centres servidors, mitjans de comunicació, nodes de xarxes i tots els sistemes d'informació, tant electrònics com tradicionals, han de poder utilitzar la mesura i poder-se valer dels mètodes infomètrics. A més, davant el desenvolupament de les xarxes electròniques i els nous processos de recerca i tractament de la informació que aquestes indueixen, és necessari, més enllà dels enfocaments infomètrics tradicionals, pensar en mètriques noves dirigides a capturar aquests nous fenòmens. Així, doncs, actualment es tendeix a definir, en les ciències de la informació, un cos de magnituds i d'unitats de mesura que, d'una manera molt apropiada, els proporcionarà el sistema de mesures que encara els fa falta.

3.1. Per què cal mesurar?

A més de les raons científiques i tècniques esmentades més amunt, hi ha diverses raons socials, econòmiques i polítiques que justifiquen el fet de recórrer a la mesura.

Com a raons **socials** i **econòmiques** podem esmentar:

- 1) La mesura dels avantatges i els inconvenients per als usuaris en el moment de substituir els mitjans d'informació tradicionals per un sistema més modern.
- 2) La mesura dels nous productes de treball en grup a distància que cal avaluar.
- 3) La mesura de l'audiència dels diaris i de les emissions de televisió i de ràdio.
- 4) La mesura del cost de les consultes en un banc d'informacions, etc.

Quant a les raons **polítiques**, l'elaboració d'indicadors xifrats respon a les necessitats de l'anàlisi del desenvolupament a llarg termini i a la voluntat dels organismes de millorar el coneixement per a il·luminar les seves accions. Han de permetre modificar aquesta realitat. Amb l'ajuda de variables es vol descriure, comparar, explicar i preveure. Per exemple, per respondre a aquesta preocupació, l'economia ha desenvolupat una branca d'estudis estadístics i matemàtics coneguda com a *econometria*.

3.2. Què cal mesurar?

3.2.1. Les magnituds mesurables

En el camp de la informació, les **magnituds mesurables** (és a dir, susceptibles d'augmentar o de disminuir) són el senyal, el text, el temps, la població i la utilitat, com es mostra en la figura 1.

Figura 1. Les magnituds mesurables en la informació

Magnitud	Símbol	Unitat	Notació
El senyal	S	Bit	bit
El text	T	Paraula	m
El temps	t	Segon	s
La població	G	Individu	g
La utilitat	U	Ús	u

Les unitats de mesura corresponents són:

- El **bit** és un logaritme en base 2 del nombre d'estats possibles del dispositiu, i els seus múltiples (kilobit, megabit, gigabit, terabit); en l'electrònica, el bit es defineix per les seves característiques físiques de durada i intensitat del senyal.
- La **paraula** (m) és una forma lingüística mínima independent dotada de sentit, i els seus múltiples (kiloparaula, megaparaula).
- El **segon** (s), els seus múltiples (minut, hora, dia, setmana, mes, any) i els seus submúltiples (mil·lisegon, microsegon, etc.).
- L'**individu** (g) és autor, lector, visitant, oient, espectador.
- L'**ús** (u) és el nombre de préstecs, de citacions, de consultes, de compres, de lectures, d'escoltes, etc. d'un objecte d'informació.

Consulteu el subapartat 3.2 del mòdul 3, "Matemàtica de la informació".

Aquestes unitats es consideren les fonamentals del sistema d'unitats de mesura de la informació, per la qual cosa ens permetran determinar altres unitats derivades com, per exemple, el cabal de senyals:

$\frac{S}{t}$ El nombre de bits per unitat de temps (bits per segon).

També es poden construir diverses mesures definides a partir d'informes relatius al consum i a la producció d'informació:

$\frac{T}{G}$ El nombre de paraules produïdes (consumides) per individu.

$\frac{U}{G}$ El nombre d'usos per individu.

$\frac{U}{t}$ El nombre d'usos per unitat de temps, és a dir, la freqüència d'ús de la informació, també coneguda com a *usabilitat*.

També és necessari que busquem establir, entre aquestes diverses magnituds conegudes en la matemàtica com a variables (la notació de les quals és x , y , z , etc.), unes relacions de naturalesa quantitativa i que es puguin expressar sota la forma d'una funció. Aquestes relacions es traduiran en una correspondència entre aquestes variables:

$$y = f(x)$$

La variable y es considera dependent (ja s'explicarà més endavant) de la variable independent x (explicativa) si la dada del valor de x fixa el valor de y .

Consulteu el subapartat 2.1 del mòdul 2, "Estadística de la informació".

Consulteu l'apartat 2 del mòdul 3, "Matemàtica de la informació".

3.2.2. Les escales de mesures

Les escales de mesura són tres:

1) L'escala nominal (designa una paraula o un nom).

Els objectes analitzats es classifiquen en categories no ordenades segons el tipus. Per exemple, l'esquema analític de qualsevol ciència és nominal, és a dir, classificatori; la gent o els fets es classifiquen per tipus. No hi ha cap categoria superior a una altra.

Exemple

En el camp de la documentació, la classificació decimal universal (CDU) és un exemple d'escala nominal de mesura; aquesta escala distingeix els llibres segons la seva pertinença a una classe o a una altra.

Classe tecnologia (ciències aplicades) = 600
Fabricació de productes per a usos concrets = 680
Classe filosofia = 100
Psicologia = 150
Consciència = 150.1

Les altres escales nominals utilitzades per a fer mesures de públics, d'audiències, etc. són la professió, el diploma, la nacionalitat, etc. Com a primera fase de qualsevol mesura, aquestes classificacions són molt útils per a trobar factors comuns: hi ha més homes que dones entre els usuaris o els no usuaris, hi ha més estudiants de primer cicle que de segon cicle, etc. Aquest tipus de mesura també subministra xifres sobre les quals es poden fer alguns càlculs dins d'una categoria concreta.

No cal considerar aquestes classes ordenades o contínues, ja que són discretes i qualitatives.

2) L'escala ordinal (que marca l'ordre, el rang, el grau, la qualitat o la posició dins d'una sèrie: 1r., 2n., 3r., etc.).

Els individus, els objectes o els fets es classifiquen o se situen dins de categories ordenades (inferior, superior, igual). Així, en una enquesta, les respostes a una pregunta que ofereix les quatre opcions *molt satisfactori*, *satisfactori*, *poc satisfactori* i *gens satisfactori* defineixen una escala ordinal i permeten obtenir l'ordre de preferència dels usuaris davant els serveis del centre de documentació.

Només s'obtenen posicions relatives, i no valors absoluts. No hi ha cap "distància" coneguda entre les opcions primera i segona, ni tampoc entre la segona i la tercera. Així, doncs, no cal considerar mai aquesta escala com si fos una escala d'interval similar a les escales de temperatura (graus Celsius o Fahrenheit). Però sovint trobem una primera opció molt clara, unes opcions segona i tercera bastant iguals, i finalment una classificació aproximada de les altres opcions en funció del que les distingeix. Per manipular les distribucions de rang s'han desenvolupat eines estadístiques específiques.

Consulteu els subapartats 2.4 i 3.3 del mòdul 2, "Estadística de la informació".

Alguns programes d'estadística manipulen les variables ordenades com variables cardinals sense respectar el que s'ha explicat més amunt.

3) L'escala cardinal (designa una quantitat amb unitat o sense).

No solament categoritzem i ordenem, sinó que també mesurem; transformem una característica física en un nombre. És a dir, definim, en una escala de mesura, una unitat de mesura amb un punt zero que indica l'absència total del que es mesura.

Exemples

- El nombre de visitants d'una exposició.
- El nombre diari de consultes d'un banc d'informacions.
- El nombre de citacions que ha tingut un article.
- El nombre de telespectadors d'una emissió televisiva.
- El nombre mitjà de kilobytes per pàgines web consultades.
- La durada d'una sessió de consulta en línia.

La mesura cardinal necessita l'aplicació d'una escala mètrica al fenomen que s'estudia. Aquesta escala es caracteritza per una unitat de mesura que s'usarà per a comparar els fenòmens. Les unitats són fixes i estan estandarditzades, la qual cosa permet unes anàlisis més rigoroses.

Exemple

Les mesures de cabal de les línies de transmissió dels senyals elèctrics a les xarxes s'expressen en bits per segon (unitat de mesura) sobre una escala que va de 0 (un valor que indica l'absència de transmissió de senyal) fins a diversos milers de milions de bits per segon. Per a expressar aquests cabals tan alts s'usen múltiples (o submúltiples) de la unitat: kbit/s, Mbit/s, Gbit/s, Tbit/s. Així, el 1990, les línies de transmissió de les xarxes següents garantien cabals diferents:

- Cabal de la xarxa telefònica commutada (XTC) = 2.400 bit/s.
- Cabal de la xarxa digital de serveis integrats (XDSI) = 64 kbit/s.
- Cabal de la xarxa de transmissió per paquets (TRANSPAC) = 2 Mbit/s.

Així, doncs, és possible ordenar aquestes xarxes segons els cabals que permeten:

$$XTC < XDSI < TRANSPAC$$

En les ciències de la informació, el recompte permet obtenir nombrosos valors numèrics. Aquests valors no tenen unitat, però es poden analitzar com a variables cardinals.

Els tractaments matemàtics que es podran fer dependran de la naturalesa nominal, ordinal o cardinal de les variables.

Consulteu l'apartat 1 del mòdul 2, "Estadística de la informació".



3.3. Com cal mesurar?

Mesurar una magnitud és comparar-la amb una altra magnitud del mateix tipus, que és la que s'adopta com a unitat.

3.3.1. Els mètodes de mesura

Hi ha nombroses mesures directes. Però també hi ha les mesures indirectes, que impliquen fer un càlcul per mitjà d'una fórmula.

Exemple

Mesura del temps de reacció per a seleccionar una operació en un menú en una pantalla tàctil:

$$t = c + k \cdot \ln_2 b$$

en què

- t és el temps en segons.
- b és el nombre d'alternatives.
- c i k són constants característiques de l'usuari.
- \ln_2 és el logaritme en base 2.

Consulteu el subapartat 2.2 del mòdul 3, "Matemàtica de la informació".



3.3.2. Els instruments de mesura

En general, per a mesurar una magnitud no es compara directament amb la unitat, sinó que s'usa algun aparell de mesura que permeti comparar els efectes de dues magnituds del mateix tipus. En qualsevol cas, aquests aparells han de reunir tres qualitats:

- La fidelitat: les indicacions no varien amb el temps.
- La precisió: les mesures de les magnituds calibrades donen uns resultats concordants.
- La sensibilitat: l'aparell mostra les més mínimes variacions de la magnitud que s'ha de mesurar.

Exemples

Calculador del trànsit en línia, comptadors de visitants, cronòmetre, sonda, etc.

3.3.3. Els errors de mesura

Atenció als errors de mesura

Moltes vegades els valors que s'obtenen són aproximats, a causa dels errors de mesura. Les mesures sense errors no existeixen. Aquests errors es deuen a la imprecisió de les eines o bé a l'acció mateixa de l'observador.

1) Errors provocats per les eines bibliogràfiques clàssiques

Són els errors de mesura en les citacions (tant en el nom de la revista, del llibre, del títol, de l'autor o els autors, del volum, de la data d'edició, etc.).

Exemple

Es van analitzar 131 articles de 5 revistes de ciències de la informació. Es van comptar 1.094 referències. Tot seguit es van comparar amb les citacions publicades en l'índex de citacions de ciències socials (*social science citation index*, SSCI). En 193 referències hi havia 223 errors, la qual cosa dona una taxa d'errors del 18%.

2) Errors provocats per les eines informàtiques

Són, per exemple, els errors en la mesura de la quantitat de visitants (usuaris) dels llocs web. Les visites freqüents dels robots encarregats d'actualitzar els índexs de cerca contribueixen a augmentar aproximadament un 10% les consultes dels llocs web.

3) Errors provocats pels mètodes d'observació

Els sociòlegs coneixen des de fa molt de temps els problemes vinculats a la recollida d'informació entre els usuaris per a avaluar-ne la satisfacció, i han desenvolupat nombroses tècniques i competències per recollir aquesta informació minimitzant els biaixos. Aquest no és, ni de molt, el cas dels mètodes de mesura de l'audiència dels llocs web. Segons si l'audiència la mesura una empresa o una altra, usant cadascuna el seu mètode propi, les diferències de 800.000 a 1.000.000 de visites no són res excepcional. Però també és veritat que prevalen els imperatius publicitaris...

El mostreig usat en les enquestes també comporta errors. La mitjana de la mostra és, en general, lleugerament diferent de la mitjana de la població en conjunt.

Consulteu l'annex 2, "Taules estadístiques".

3.4. Com s'ha de calcular?

Una vegada fetes les mesures, l'important és efectuar càlculs estadístics i matemàtics a partir dels resultats d'aquestes mesures.

3.4.1. Del càlcul manual al càlcul informàtic

Els valors numèrics recollits, i també els càlculs fets sobre aquests valors, a mà o amb alguna calculadora o un microordinador, es presenten en fulls de càlcul de la manera següent.

Exemple

Préstecs de llibres a la biblioteca i edat dels lectors.

Nombre de llibres prestats	1	3	5	1	0	1	1	2	0	4	...
Edat (anys)	18	46	21	15	28	45	19	14	55	15	...

1) Anàlisi manual

Recordem els dos mètodes manuals usats per a analitzar els valors obtinguts a partir de qüestionaris. Per a analitzar les respostes estructurades i tancades es poden usar:

- Un **qüestionari buit**: s'expliquen les respostes a cadascuna de les preguntes i s'introdueixen els totals en el qüestionari en les caselles que cal marcar:

Exemple

Quants llibres heu demanat en préstec aquesta setmana?

Cap		14 IIII IIII IIII	(recompte amb traços)
1	27	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	(recompte amb traços i cercles)
2	51		
> 2	8		
...			

Això només dona la distribució dels valors d'una variable, però no permet cap comparació amb una altra pregunta.

- Un **full de recompte**: permet fer aquesta comparació, però només es pot utilitzar si la quantitat de preguntes és reduïda.

Exemple

Nombre de llibres prestats

Edat	Cap	1	2	> 2	Total
Menys de 21 anys	7	10	28	3	48
21-45 anys	2	9	8	2	21
Edat	Cap	1	2	> 2	Total
46-60 anys	3	5	12	1	21
Més de 60 anys	2	3	3	2	10
Total	14	27	51	8	100

2) Anàlisi informàtica

Avui dia disposem de calculadores electròniques (numèriques i gràfiques) programables amb un rendiment espectacular, tant per la quantitat d'operacions que poden fer com per la rapidesa i la precisió amb la qual donen els resultats. A més, també hi ha una oferta molt completa de programes informàtics d'anàlisi matemàtica i d'anàlisi estadística que es poden usar en la gran majoria dels microordinadors comercials.

Exemple

En el camp del càlcul estadístic, els programes informàtics d'anàlisi estadística permeten accedir a nombrosos procediments: estadística bàsica (mitjana, desviació estàndard,

Consulteu el subapartat 3.4.2 d'aquest mòdul.

quartil, proves, distribucions simples i distribucions creuades), càlcul de probabilitats (distribucions binomials, de Poisson i normal, generació de nombres aleatoris), correlacions, anàlisis de variància, regressions simple i múltiple, tipologies, anàlisis de sèries cronològiques (modelització, descomposició estacional, previsió, segmentació), tractaments lèxics, anàlisis de contingut, anàlisis multivariades (classificació jeràrquica, anàlisi factorial, anàlisi en components principals), mostreig (aleatori, comparació de mostres), mètodes no paramètrics, etc.

En el camp del càlcul matemàtic, els programes informàtics fan càlculs numèrics i càlculs simbòlics, dibuixen gràfiques i corbes en 2D i 3D que representen valors i funcions (hiperbòliques, logarítmiques, exponencials, etc.), ajusten corbes, permeten la resolució d'equacions, la simulació, el càlcul de límits, la interpolació, etc.

El cas del processament d'enquestes: abans de l'aparició dels programes informàtics d'enquestes dedicats a l'elaboració i al processament de qüestionaris, l'ús de l'ordinador exigia temps i esforç per a convertir les respostes a les preguntes en formes adaptades. Les respostes orals o numèriques a les preguntes s'havien de representar mitjançant nombres.

Exemple

- Suposem una pregunta tancada P1 en la qual cal seleccionar una de les dues respostes possibles: *sí* i *no*.

Atribuïm el valor 1 a la resposta *sí* i el valor 2 a la resposta *no*. Si no es dona cap resposta, el valor serà 9.

Considerem una pregunta P2 vinculada a P1. P2 no s'aplica als que han respost *no* a P1, i hem creat un valor 8 que indica que, encara que no hi hagi cap resposta, no per això es tracta d'una falta de resposta.

- L'anàlisi de les preguntes obertes exigeix la creació de variables noves resultants de l'inventari dels temes abordats per les persones interrogades. Quan un tema apareix esmentat en la resposta, s'atribueixen a aquestes variables els valors 0 o 1.

Una vegada s'han fet tots aquests tractaments, és possible aplicar als nombres obtinguts un programa d'anàlisi estadística.

3.4.2. Calculadores i microordinadors

L'estadística i la matemàtica són, sens dubte, els primers camps d'aplicació de la informàtica, que va arrelar en el llindar entre la mecanografia i les matemàtiques. No oblidem que Frederick Bull era al capdavant d'una companyia de material de mecanografia especialitzada en màquines de calcular, i que buscava per a la seva clientela de comptables unes màquines cada vegada més eficaçes. Va ser així com van aparèixer les primeres calculadores electròniques automàtiques, primer analògiques i després digitals.

Exemple

Les màquines analògiques constitueixen models dels sistemes físics que permeten estudiar. Les variables matemàtiques es representen amb magnituds físiques com, per exemple, tensions elèctriques o desplaçaments angulars proporcionals. Les operacions matemàtiques –suma, multiplicació, divisió, integració, diferenciació, etc.– es fan amb òrgans elèctrics, electrònics o electromecànics altament especialitzats.

Les màquines actuals, que són totes digitals, tenen tres funcions molt diferents:

- Una funció d'organització o de programa, que es materialitza en la sèrie o la seqüència de les instruccions que especifiquen la successió d'operacions que s'han de fer sobre els nombres recollits.
- Una funció d'emmagatzematge d'aquests nombres, coneguda també com a *funció de memòria*, que es materialitza en l'escriptura dels nombres en les caselles corresponents del full de càlcul.
- Una funció de càlcul, que es materialitza en la màquina ancestral amb teclat.

1) La calculadora

La primera calculadora va aparèixer el 1972. Utilitzant circuits integrats, aprofitava els components bàsics dels ordinadors miniaturitzant-los. El seu cost elevat en aquells moments es veia compensat per la facilitat d'ús: entrades i sortides amb el sistema decimal, encara que els càlculs interns es feien en el sistema binari, unes dimensions reduïdes, fàcils de transportar i un manteniment que es reduïa al subministrament d'alimentació. En qüestió de vint-i-cinc anys, el seu preu s'ha dividit per cent i altres avenços de la tècnica l'han millorat. El 1975 va aparèixer la primera calculadora programable amb targeta magnètica, després amb impressora... El que el 1972 era un objecte altament sofisticat, avui dia s'ha convertit en un objecte d'allò més habitual, que fins i tot s'ofereix com a clauer.

Exemple

Configuració d'una calculadora per al gran públic:

- 128 kb de RAM, 1 Mb de ROM, 570 kb de memòria de programació (Turbo Pascal, assemblador, etc.).
- Nombroses funcions matemàtiques (logaritmes, integrals, derivades, càlcul matricial, etc.) i estadístiques (mitjanes, variàncies, lleis normals, binomials, de Poisson, proves, regressions, etc.), solucionador d'equacions numèriques, resolució d'equacions diferencials, geometria interactiva.

2) El microordinador

Des dels seus orígens, els ordinadors s'han considerat unes eines de càlcul (*computer*). El desenvolupament de les calculadores electròniques, durant la Segona Guerra Mundial, va confirmar aquesta orientació, ja que el seu objectiu continuava essent resoldre càlculs cada vegada més importants. Durant els anys cinquanta va aparèixer una altra manera d'abordar el càlcul, iniciada pels cibernetistes Norbert Wiener i William Ross Ashby. Les calculadores es van començar a veure no tant com a màquines per a manipular nombres,

sinó com a màquines per a manipular símbols. Va ser llavors quan va néixer l'ordinador.

Des dels anys seixanta i setanta van aparèixer diversos programes informàtics estadístics i matemàtics d'alt nivell basats en els sistemes d'IBM. Quan van aparèixer els microordinadors, desgraciadament no es va produir la transferència d'aquests programes a aquestes noves màquines, encara massa modestes. L'aparició, al començament dels anys vuitanta, dels microordinadors més eficients, com el Macintosh i el PC d'IBM, va consolidar el desenvolupament d'aquests nous mètodes.

Exemple

- Microordinador: processador de 600-800 MHz, 64-128 Mb de RAM, 10-15 Gb disponibles en el disc dur, pantalla de 17 polzades, targeta gràfica, lector de CD-ROM de 48X o lector de DVD.
- Resolució numèrica o simbòlica de problemes matemàtics habituals, nombroses funcions de càlcul predefinides (funcions habituals, càlculs vectorials i matricials), estadístiques generals (distribució, regressió, interpolació i previsió), resolució d'equacions, gràfiques en 2D i 3D, etc.

3.4.3. Els programes estadístics i matemàtics

1) Els programes estadístics

- **Programes generals:** per a justificar el terme de *programa estadístic*, un programa informàtic ha d'oferir diversos tractaments habituals, com l'anàlisi de la variància, la regressió múltiple o fins i tot l'anàlisi factorial.
- **Programes infomètrics i cartogràfics:** integren mètodes estadístics i s'usen habitualment en les ciències de la informació. Normalment inclouen mòduls cartogràfics.
- **Programes lexicomètrics:** programes estadístics especialitzats en el tractament de les cadenes de caràcters.

2) Els programes matemàtics

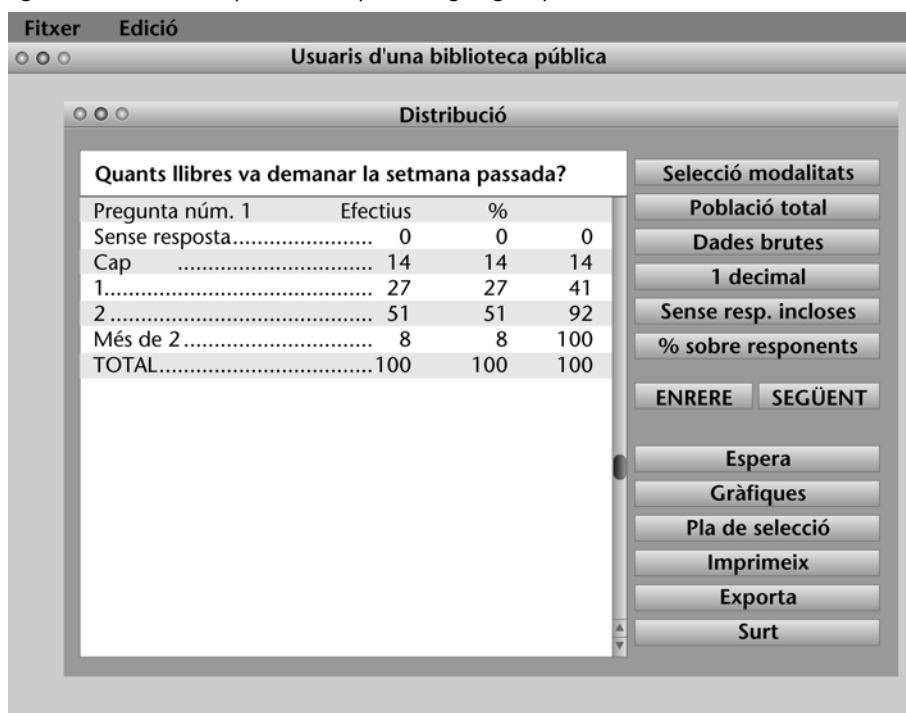
- **Programes generals:** permeten la majoria de les anàlisis matemàtiques i el dibuix de les gràfiques corresponents.
- **Programes de càlcul numèric.**
- **Programes de càlcul simbòlic.**

Observació

Hi ha una oferta molt important de programes dissenyats de manera especial per a fer enquestes i sondejos. En general, aquests programes permeten definir el format del qüestionari, el manual de manteniment, etc., i també l'edició en paper o en format HTML en el cas dels sondejos per Internet. Una vegada s'han emplenat els qüestionaris, l'examen es fa dins d'aquest marc preestablert. Els mòduls estadístics que s'inclouen en aquests programes permeten fer tractaments diversos i detallats.

Abans d'aquests tractaments, les *distribucions simples*, permet obtenir els efectius dels responents per a cada modalitat de resposta a una pregunta (estadística d'una variable). En la figura següent es mostra una distribució simple feta amb el programa Modalisa®. Aquestes distribucions acostumen a anar acompanyades de representacions gràfiques o, fins i tot, les substitueixen.

Figura 2. Distribució simple: efectius, percentatges, gràfiques



Pel que fa a les distribucions simples, podeu consultar el subapartat 2.3 del mòdul 2, "Estadística de la informació".

4. Mesures i lleis

La mesura és el primer pas cap a la matematització de la realitat, que és la base del coneixement científic. Per exemple, la física és la matematització de la naturalesa. La física és singular i s'estructura entorn de les matemàtiques o, més aviat, a partir del projecte de construir una imatge matemàtica de la realitat. En canvi, és perfectament possible imaginar una biologia, una economia i fins i tot una química sense matemàtiques.

Es diu que una ciència es matematitza quan els conceptes de les seves lleis es vinculen mitjançant relacions formals. No hem d'oblidar que una llei estableix unes relacions mesurables, universals i necessàries entre els fets que permeten la previsió.

En física, en la llei de Newton, els conceptes de força, de massa i d'acceleració es vinculen mitjançant una relació formal de tipus vectorial que es coneix com la *llei de la gravetat*.

$$\vec{F} = M\vec{\Gamma}$$

És ben cert que ja hem superat el marc d'interpretació del segle XIX i avui dia es posa en qüestió l'epistemologia positivista de la ciència, que en aquella època es basava en la hipòtesi que el món real tenia una estructura perfectament definida, regida per lleis i comprensible per la raó humana. Pràcticament tots els científics estan d'acord que el moviment dels planetes no es deu a la força de la gravetat, sinó que és la llei de la gravetat que converteix en intel·ligibles els moviments dels astres.

Les ciències de la informació també s'han matematitzat en les seves lleis principals; així, la llei de Lotka, que explica les diferències de productivitat dels investigadors científics dins d'un camp determinat, ens ensenya que la relació autors-articles és inversament proporcional al quadrat del nombre d'articles publicats:

$$G_i = \frac{k}{U_i^2}$$

En què

G_i = nombre d'autors que han publicat U_i articles

U_i = nombre d'articles

k = constant

Certament, igual que moltes altres regularitats estadístiques d'aquest tipus, la llei de Lotka només és vàlida en general i d'una manera aproximada. No obstant això, aquesta llei, que forma part d'un conjunt més ampli conegut com a *lleis de la informació*, sembla tenir una solidesa destacable malgrat les diferents imprecisions degudes a les mesures i a les condicions experimentals.

Consulteu el subapartat 3.4 del mòdul 2, "Estadística de la informació", i el subapartat 2.2 del mòdul 3, "Matemàtica de la informació".

Però llavors, quines matemàtiques cal escollir, tenint en compte que la matemàtica és el conjunt de les ciències que tenen com a objectiu la quantitat i l'ordre? Aquestes ciències són, per exemple, l'àlgebra, l'anàlisi, l'estadística, les probabilitats, les lògiques o les eines de càlcul diferencial, integral i vectorial.

La selecció d'una matemàtica o una altra és determinada pel tipus de fenòmens que actuen dins del camp de coneixements que s'estudia. Les ciències de la informació s'ocupen de les activitats socials que mobilitzen nombroses persones que produeixen, comuniquen i utilitzen moltes informacions, és a dir, fets socials en gran quantitat que generaran una gran quantitat de nombres. Aquestes grans quantitats i la complexitat dels fenòmens informatius fan que en sigui impossible l'estudi exhaustiu. Així, doncs, cal trobar les matemàtiques del "resum". L'estadística serà la primera d'aquestes matemàtiques. I a ella dedicarem el mòdul següent.