

BUTLL. G C A, (1) 1: 9-20 1981

MANEJAMENT ESTADÍSTIC DE LES DADES D'ANELLAMENT

Josep Vilalta (a)

ABSTRACT

Statistical management on Ringing Data

This article is the first in a series planned to give an eye to the statistical terminology in use when working on biometrical data of ringed birds.

It starts with definitions of sampling and sample, types of data and its classification, minimal sampling unit, and individual observation. Afterwards, the different variable types are classified, studying each and every type with different examples.

Other subjects such as population, accuracy, exactness, real limits (implicit) are also explained.

Finally, it studies how to perform a frequency distribution and what are the advantages of its biological interpretation. It deals with a qualitative and quantitative frequency distribution. Different graphical essays are presented on both models as well as the set up of a frequency polygon.

Our basical aim when trying to clarify all these concepts is to get rid of all types of conceptual and linguistic barbarisms liable to mixtify our ringing reports.

(a) TALLER DE NATURA, Comissió d'Ornitologia, Badalona (Barcelona).

Generalitats

En aquest article, així com en d'altres, volem donar un repàs a la terminologia estadística —a fi d'ajudar a no caure en barbarismes lingüístics o conceptuals en publicar els nostres Reports— i als càlculs utilitzats més sovint a l'hora de presentar les dades biomètriques dels ocells anellats. (1)

No obstant, seguirem un criteri ampli en explicar els diferents conceptes sense deixar, però, l'àmbit de l'ornitologia.

Avui més que en cap altre temps tota activitat científica exigeix un coneixement i un domini del càlcul matemàtic. En investigar certs fenòmens o en presentar unes conclusions, el nivell de fiabilitat sovint ve donat per la nostra destresa en emprar els diferents estadístics i per la grandària de la mostra estudiada.

Dins la nostra tasca en operar amb dades biomètriques tenim la doble responsabilitat de manejar escrupolosament els models matemàtics i de donar als resultats una interpretació biològica coherent.

El fet de valorar les freqüències o la dispersió de les mostres que estudiem té una gran transcendència en investigar els efectes selectius del medi sobre un tipus de mesures.

Bé davant un dimorfisme sexual —on els caràcters biomètrics seran cabdals per la identificació—, com davant un entrecreuament de diferents comunitats, o la detecció d'una subespècie, l'exactitud dels resultats amb aquestes operacions força sensibles, com he dit abans, descansarà sobre un mostratge ampli i una correcta aplicació dels estadístics.

Les Dades

- 1 Definicions
- 2 Variables
- 3 Distribució de freqüències
- 4 Assaig gràfic

1 Definicions

"Quan volem obtenir informació d'una població o col·lectiu molt gran, recorrem al mostratge, això consisteix en l'extracció de mostres de l'esmentada població, la qual cosa es fa generalment a l'atzar.

Noteu que l'acció d'extreure mostres és el mostratge i no el "mostreig", i que el verb "mostrejar" és inexistent en català; en la nostra llengua hem de dir, doncs, extreurem (o prendrem, obtindrem...) mostres en comptes de "mostrejarem".

De les mostres obtingudes per un mostratge obtindrem les dades que ens caldran per tal de conèixer allò que volem saber del col·lectiu o població objecte; si el mostratge ha estat ben fet, les dades reflectiran de manera no esbiaixada —sense biaix— (mai s'ha de dir "inesgada") aquella característica de la població que volem conèixer.

Les dades poden ésser el resultat de comptatges, mesures o amidaments (aquest darrer mot és restringit a mesures de longitud), fets sobre l'individu

(1) Les dades emprades en aquest article pertanyen al Centre d'Estudi de la Migració dels Ocells de Tiana (Barcelona).

(2) JORDÍ LLEONART "Els barbarismes en l'estadística"
Full Lexicogràfic núm. 7, Comissió Coordinadora Lexicogràfica de Ciències

o els individus que componen cada mostra.

Si de cada individu hem pres una sola mesura, go és, si disposem de moltes mesures d'una sola variable aleatòria, hom pot fer en primer lloc un ordenament de dades, Taula 6 (A) i (B), i estudiar la seva distribució de freqüències, fig. 3. Després caldrà fer estimacions (no s'ha de dir estimes, mot reservat a la situació d'un vaixell en la mar) per mitjà de diferents estadístics descriptius: la mitjana (no "mitja"), la mediana, la moda, la desviació estàndard i la variància (no la "variança)". (2)

Tot aquest procés el veurem detingudament més endavant, ara cal aprofundir en allò que esdevé quan treballem amb una mostra.

Si prenem un conjunt de mesures, aquestes es basen en observacions individuals, o sigui, en dades o observacions agafades d'una mínima unitat de la mostra, (m u m). Sovint, però no necessàriament, aquestes m u m, són també individus en el sentit biològic ordinari.

Si pesem una mostra de Tallarols de casquet (*Sylvia atricapilla*), el pes de cadascun dels Tallarols és una observació individual, els pesos del total de Tallarols representen la mostra d'observacions (*n*), que es defineix com un conjunt d'observacions individuals seleccionades per un procediment específic, fig. 1.

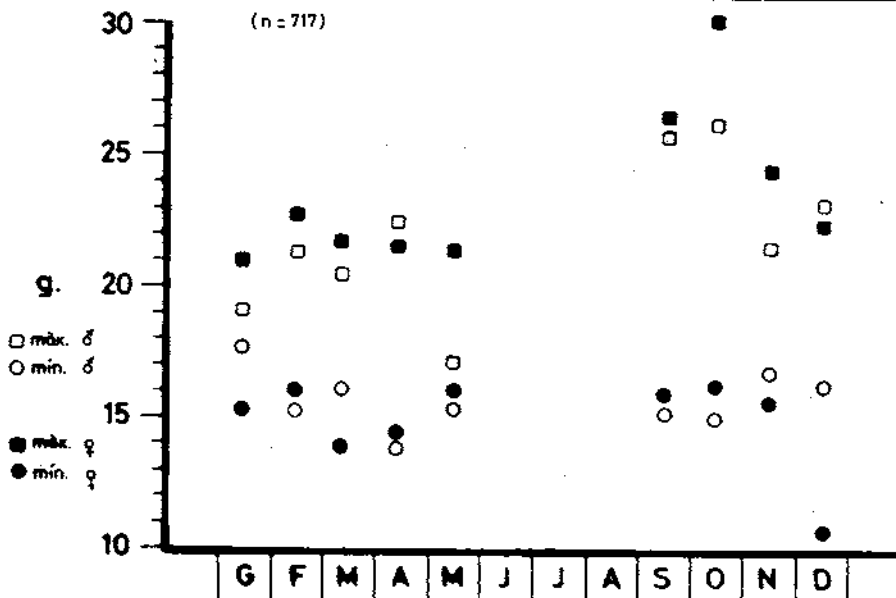
En el darrer exemple, una observació individual està basada en un individu en sentit biològic, un Tallarol de casquet, però, si haguéssim estudiat el pes d'un sol Tallarol a través d'un període de temps, la mostra d'observacions individuals estaria constituïda pels pesos de l'ocell enregistrats en diferents moments successius.

Distribució mensual dels pesos
màx. i mín. segons el sexe.

Fig.1

Sylvia atricapilla

(n = 717)



Els conceptes observació individual i mostra d'observacions, tal com els hem utilitzat abans, sols ens defineixen l'estructura però no la naturalesa de les dades en estudi.

El caràcter o variable és la propietat real mesurada per les observacions individuals.

A cada mínima unitat de la mostra poden mesurar-se més d'un caràcter. Així quan estudiem els ocells anellats, a més de mesurar el pes i d'observar qualsevol tret morfològic -que ajudarà a identificar l'espècie, el seu sexe i l'edat-, segons el tipus d'espècie prenem els següents amidaments:

<u>amidaments utilitzats</u> <u>més sovint</u>	longitud total (llargada)
	" alar
	" amplària
	" cua
	" bec
	" tars
	" ungle

TAULA 1

En l'exemple dels Tallarols de casquet, la mostra és bivariada, ja que tenim 352 observacions de ♂ i 365 observacions de ♀, mostra total d'observacions: 717.

Quan parlem de població, ens referim a tots els individus d'una espècie en concret (o tal vegada d'una fase del cicle vital o d'un sexe).

Definim la població com: la totalitat d'observacions individuals sobre les quals es fan inferències, i que existeixen a qualsevol lloc del món, o bé, dins una àrea de mostratge clarament especificada, limitada en l'espai i el temps.

A partir d'una mostra de 717 Tallarols de casquet treiem conclusions sobre el pes de tota la població d'aquesta espècie de Tallarol; com sempre, la grandària de la mostra serà un factor decisiu que ens ajudarà a una correcta aproximació vers la realitat.

Una població pot referir-se a variables d'un conjunt concret d'objectes, per exemple, les longituds alars de tots els Tallarols de casquet del món, o bé, als resultats d'experiments, tals com la freqüència en que bateguen les ales d'un Gafarró (*Serinus serinus*).

En el primer cas la població és generalment finita, encara que a la pràctica seria impossible mesurar les longituds alars de tots els tallarols. En canvi, en el segon cas, trobem que el experiment pot repetir-se infinites vegades, mentre tinguem humor per a fer-ho, clar està.

La exactitud i la precisió s'utilitzen com a sinònims en el llenguatge col·loquial, cal doncs definir aquests conceptes més rigorosament.

Exactitud: és l'aproximació de la mesura i el càlcul d'una dada al seu veritable valor.

Precisió: és l'aproximació de les mesures repetides d'una mateixa quantitat.

La darrera xifra d'una mesura implica la precisió, go és, els límits entre els quals creiem que es troba el veritable valor. Així una mesura de 60,5 mm. implica que la veritable longitud oscil·la entre 60,45 i 60,55. Entre aquests límits implícits no sabem on està exactament la longitud real.

Suposant que no hi hagi biaix, -"sesgo" en castellà- un número és tant o més exacte, a mesura que som capaços d'escriure'l amb més xifres significatives.

	<u>longitud alar</u>	<u>límits implícits</u>
<u>Sylvia Hortensis</u>	79 mm.	78,5 - 79,5 mm.
	77,5 mm.	77,45 - 77,55 mm.
	76,85 mm.	76,845 - 76,855 mm.

TAULA 2

Amb quantes xifres significatives hem de prendre les mesures ?

Si ordenem la mostra per ordre de magnitud, de l'individu més petit al més gran, una regla fàcil de recordar és que el número de classes consecutives (així anomenem a les xifres ordenades d'una mostra), disposades en un interval de 0,1, de menor a major, ha d'estar entre 30 i 300. Així davant una mostra de Mallerenga cuallarga (*Aegithalos caudatus*), si le dada més petita correspon a 6 g. i la més gran a 10 g., solament hi trobem 4 classes consecutives entre la mesura menor i la major. Per tant haurem d'aproximar els valors amb una xifra decimal més.

Ara, amb uns valors extrems de 6,5 i 10,1, si que hi podem comptar 36 classes consecutives (Taula 6, B), aquesta serà la xifra adequada.

La raó d'aquesta regla és que un error d'una unitat a la darrera xifra significativa de 6 g., constituiria un error inadmissible del 25 %, però, un error d'una unitat a la darrera xifra significativa de 6,5 g., és menor del 2,5 %.

2. Variables

Hem dit que un caràcter o variable és la propietat real mesurada per les observacions individuals. També la podem definir com una propietat respecte a la qual els individus d'una mostra difereixen d'alguna manera verificable.

Per exemple: longitud alar. pes, situació del niu, grandària del territori.

TIPUS DE VARIABLES

- | | |
|-------------------------------------|-----------------|
| | 1. contínues |
| a. Variables mesurables | 2. discontinúes |
| b. Variables classificades en rangs | |
| c. Atributs | 1. proporció |
| d. Variables derivades | 2. percentatge |
| | 3. index, taxa |

TAULA 3

a. Variables mesurables: són totes aquelles en què els diferents valors poden expressar-se d'una forma numèricament ordenada.

a.1. Variables contínues: en teoria poden prendre un número infinit de valors entre dos punts concrets. Per exemple: longituds, àrees, volums, pesos, angles, temperatures, períodes de temps, percentatges, velocitats...

a.2. Variables discontinues (també anomenades discretes o merístiques): solament poden tenir valors numèrics fixes. Així, el número de paràsits que té el plomatge d'un ocell, o bé, el número d'ous d'una niuada, serà 3-5-7, però mai 3,5-5,4-o 7,3.

Exemples: mitja anual d'espècies observades, número total d'ocells anellats.

b. Variables classificades en rangs: són variables que no podem mesurar, però sí podem ordenar-les segons la seva magnitud. Exemple: si estem realitzant estudis sobre el cicle de cria de diferents espècies, podem registrar l'ordre de descloa d'una sèrie d'ous, sense especificar el moment exacte en que ho va fer cadascun d'ells. En expressar una variable com una sèrie de rangs: 1, 2, 3, 4, 5..., no implica que la diferència en magnitud entre els rangs 1 i 2, sigui idèntica ni proporcional, a la diferència entre 2 i 3.

c. Atributs: són variables que al no poder mesurar-les s'han d'expressar qualitativament.

Sempre representen propietats, com ara, negre i blanc, mort o viu, mascle i femella, tipus d'espècie. Quan els atributs es combinen amb freqüències poden manejar-se estadísticament. D'aquest procés en diem fer una «enumeració de dades», i així es presenten en taules apropiades per l'anàlisi estadístic.

Exemple: TAULA 4

<u>ESPÈCIE</u>	<u>(r) d'anellament TIANA (1975/79)</u>
Tallarol de casquet (<u>Sylvia atricapilla</u>)	866
Tallarol gros (<u>Sylvia borin</u>)	330
Tallarol capnegre (<u>Sylvia melanocephala</u>)	103
Tallareta vulgar (<u>Sylvia communis</u>)	41
Tallarol de garriga (<u>Sylvia cantillans</u>)	23
Tallarol trencanates (<u>Sylvia conspicillata</u>)	4
Tallareta cuallarga (<u>Sylvia Undata</u>)	4
Tallarol enmascarat (<u>Sylvia hortensis</u>)	3

d. Variables derivades: quan fem un treball biomètric, la majoria de les variables s'estudien mitjançant observacions registrades com a mesures directes, amb comptatges d'un material biològic, o bé, lectures que són producte de diferents tipus d'instruments. Però, també tenim unes altres variables, que podem anomenar variables derivades, generalment estan basades en dos o més variables mesurades independentment, i expressen llurs relacions d'una manera determinada.

Així tenim: proporcions, percentatges, l'índex i la taxa.

d.1. Proporció: una proporció pot expressar amb una sola xifra, o bé amb un cocient, la relació entre dues variables. 42:28, podria representar la proporció de mascles davant el número de famelles.

Altres relacions són: ous:nius, joves:nius, adults:joves, polls vius:polls morts, espècies diferents registrades:espècies totals observades.

Tots aquests exemples impliquen proporcions basades en comptatges, també podem fer proporcions amb variables contínues, així, 72:180 serà la relació entre llargada i amplària de l'Aguila cuabarrada.

d.2. Percentatge: els percentatges són un tipus de proporció de gran importància pel nostre treball, per exemple, en observar el percentatge mig de Controls -autorecuperacions- anuals entre tres tipus de Tallarols: el T. de casquet, el T. gros i el T. capnegre, els resultats són ben demostratius del caràcter dispers i migratori del primer i el segon, i de la localització i sedentarisme del tercer.

TAULA 5

AULA 5		1975		1976		1977		1978		1979		Total		% mig
		A	C	A	C	A	C	A	C	A	C	A	C	C
Tallarol de casquet (<i>Sylvia atricapilla</i>)		124	222	1	203	1	94	17	223	29	866	48	5,5	
Tallarol gros (<i>Sylvia borin</i>)		17	111	-	183	1	9	-	10	-	330	1	0,30	
Tallarol capnegre (<i>Sylvia melanocephala</i>)		14	17	4	49	12	7	10	16	12	103	38	37	

Codi:- A: ocells anellats

C: Controls -ocells autorecuperats-

d.3. Index: és una proporció que expressada en sentit ampli podria ser la mitjana de dues mesures.

$$1/2 (\text{long. A} + \text{long. B})$$

Quan estudiem les poblacions parlem de l'índex de població, l'índex de supervivència, l'índex de mortalitat.

3 Distribució de Freqüències

1. D.F. quantitatives

Tenim dos tipus:

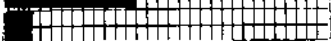
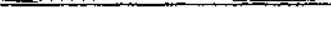
2. D.F. qualitatives

Les que s'utilitzen més sovint són les quantitatives, tot i això, en direm quatre ratlles de les qualitatives.

D.F. qualitatives: açí les diferents classes, en lloc de ser mesures preses d'una variable contínua, són diferents atributs que nosaltres podem ordenar de forma arbitrària.

Exemple: quan fem un estudi ecològic de les poblacions, sovint ens interessa obtenir la freqüència en què les diferents espècies han estat trobades dins l'àrea on hem fet el mostreig.

Fig 2

Itinerari V-4	ESPECIES	Freqüències	(f)
	Aegithalos caudatus		25
	Parus major		20
	Parus caeruleus		15
	Parus cristatus		12
	Parus ater		2

D. F. quantitatives: observem la Taula 6 (A), tenim una mostra de 50 pesos de Mallarenga cuallarga (Aegithalos caudatus). Les diferents lectures estan col·locades sense un ordre previ, ço és, tal com varen ésser obtingudes.

A la Taula 6 (B), tenim les mateixes dades, però, ordenades segons la seva magnitud. A un costat tenim dues columnes, la primera -signes-, anem marcant el número de vegades que cada lectura es repeteix dins la mostra; a la segona -f-, hi anotem la xifra total de repeticions, que serà la freqüència.

Observareu que en ordenar i agrupar el conjunt de mesures originals, la mostra d'observacions, hem aconseguit resumir les 50 lectures o observacions en 14. Així mateix, establint un interval de classe de 0,1 unitat, podem comptar unes 36 classes consecutives entre la lectura menor i la major, i unes 37 marques de classe.

Ara ja podem veure quines són les observacions que presenten freqüències més altes (7 - 7,4 - 7,7).

També advertim que hi ha certes classes (6,8 - 8,2 - 8,7...), que no estan representades per cap observació. Això és degut a què disposem d'una mostra massa petita per a realitzar una distribució de freqüències en 37 classes. Així doncs, la gràfica tindrà un aspecte bastant allargassat i dispers.

Per obtenir una distribució d'aspecte més uniforme, caldrà condensar les observacions en menys classes. Aquest procés conegut com -agrupament de classes-, el podeu observar en la Taula 6 (C) (D).

De fet quan agrupem les observacions en unes classes de rang més ampli, estem prolongant el procés que varem iniciar en obtenir la mesura original.

Com ja hem explicat en parlar de precisió, quan mesurem un pes i anotem 8,3 g., volem expressar amb això, que el veritable valor està entre 8,25 i 8,35 g., però, què som incapaços de mesurar fins a la segona xifra decimal.

Si registrem inicialment 8,3 g., estem fent una estimació que cau dins aquest rang. Si l'estimació excedeix el valor de 8,35 g., per exemple, hauríem d'aplicar la marca superior immediata, 8,4 g. Per tant, totes les mesures entre 8,25 i 8,35 g., són agrupades en realitat, dins la classe identificada per la - marca de classe - 8,3.

El nostre -interval de classe- inicial, en la Taula 6 (A), es de 0,1 unitats. Per tal de veure més clar aquest procés, agruparem les observacions un parell de vegades.

Taula 6

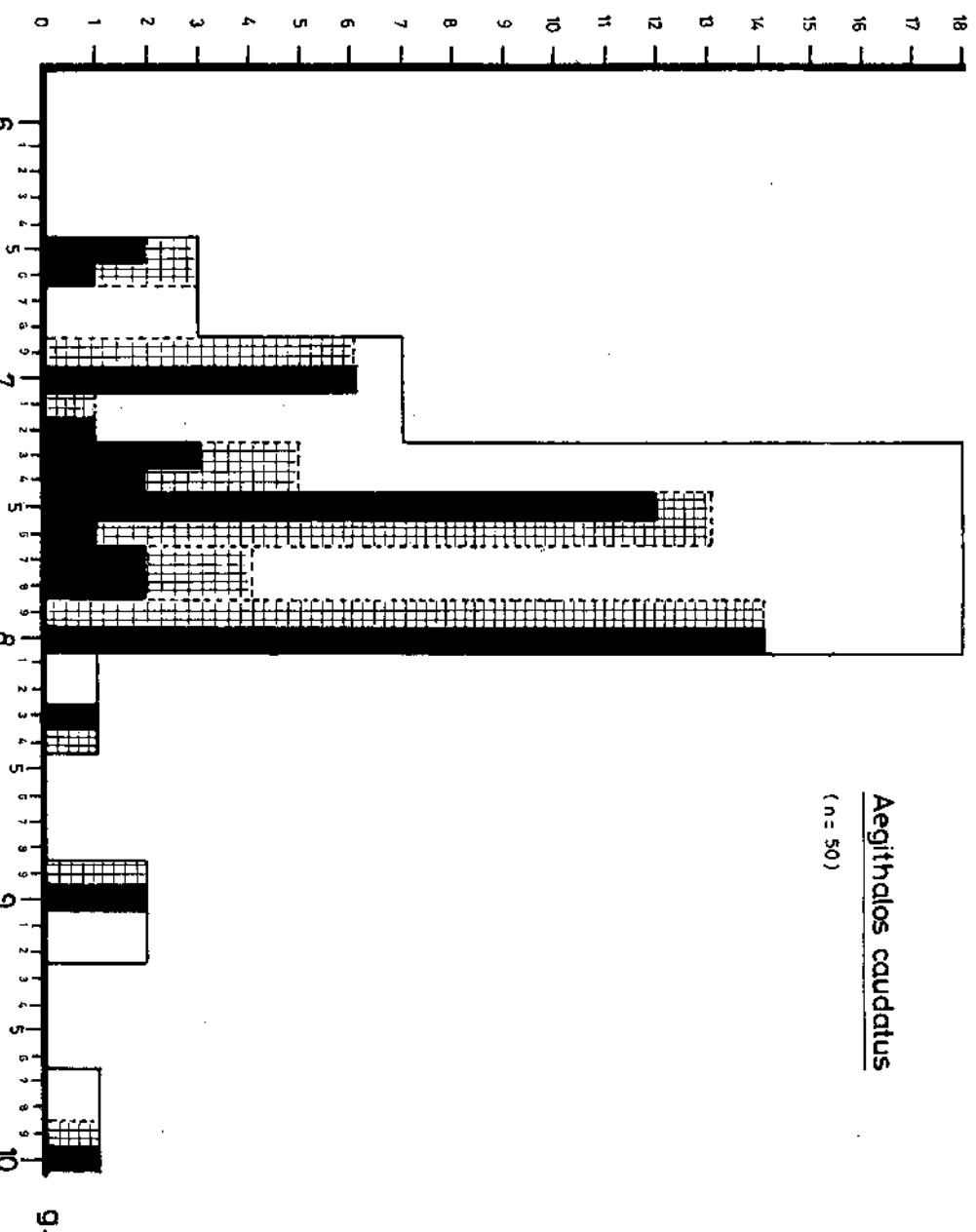
Mallereña Cuallarga (*Aegithalos Caudatus*)
mostra de 50 pesos expressats en grams

(A)

8	8	9	8	8	8	8	7,4	9	6,6	7	7,4	8,3	6,5
7,5	8	7,5	7,5	8	8	7,3	7,5	8	7,5	8	10	7,8	8
7,5	6,5	8	7,5	7,7	7	7,5	7	7,5	8	7	7,6	7,2	7,5
7,8	7,5	7	7,7	7,3	7,3	7	7,5						

(B) Distribució de Freqüències original	(C) Agrupament en 19 classes d'interval 0,2	(D) Agrupament en 9 classes d'interval 0,4
ordre límits implícits Y signes f	límits implícits marca signes f	límits implícits marca signes f
6,5 6,45 - 6,55 6,5 II 2	6,45-6,65 6,55 III 3	6,45-6,85 6,65 III 3
6,6 6,55 - 6,65 6,6 I 1	6,65-6,85 6,75	

Fig. 3



En el primer cas, l'interval de classe s'ha duplicat en amplitut, ço és, s'ha fet de 0,2 unitats. Ara els límits de classe implícits seran de 8,25 - 8,45 , 8,45 - 8,65 ...

Com veieu, en fer intervals de classe més amplis, no fem res més que estendre el rang dintre del qual se situen les mesures d'una classe.

Fet aquest primer agrupament, la nostra tasca ara és trobar les marques de classe representatives.

Quan hem fet la primera distribució de freqüències això ha estat força senzill, ja que les mesures originals s'han utilitzat com marques de classe, ara caldrà trobar el punt mig dels nous intervals de classe, per exemple, el de 6,45 - 6,65 ,serà 6,55. Aquesta nova marca observarem que té una xifra decimal més que les mesures originals, el que no vol dir que s'hagi conseguit una major precisió.

Sempre que escollim un interval de classe on la darrera xifra significativa sigui parell (0,2 - 0,4), la marca de classe portarà una xifra decimal més que les mesures originals.

Una vegada trobats correctament els límits implícits i la marca de classe per a la primera observació, els altres poden enregistrar-se a sota sense cap tipus de càlcul especial. Senzillament es va sumant l'interval de classe a cadascun dels valors.

Cal remarcar, però, que quan més amplis són els intervals de classe més es condensen les dades, i també es fan menys precises. Si ens fixem en la distribució de freqüències dels pesos de Mallarenga cuallarga (*Aegithalos caudatus*), Fig. 3, trobarem que l'estructura inicial es força dispersa, però, es va simplificant per mitjà de l'agrupament.

Quan agrupem la distribució de freqüències en 19 classes amb un interval de classe de 0,2 unitats, aquesta es fa notablement bimodal, ço és, té dos pics de freqüències. Si bé quan fem un agrupament en 9 classes, amb un interval de 0,4 unitats, es perd una part de precisió i la gràfica es presenta amb un sol pic de freqüència.

El número de classes que podem establir quan fem una d.f. dependrà en gran part de la grandària de la mostra. Una mostra menor de 50 dades no hauria de tenir més de 10 classes, ja que en altre cas aquesta inflació de classes provocaria un número massa petit de freqüències per classe. En canvi, mostres d'un miler de dades poden agrupar-se profitosament en més de 20 classes. Per terme mig, i segons la mostra, el número de classes oscil·larà entre 10 i 20.

Quan hauríem d'agrupar les dades en distribucions de freqüències ?

Si disposem d'una calculadora i tenim una mostra de més de 100 observacions, val la pena posar les dades en una d.f. abans de realitzar els càlculs estadístics, ja que la forma que ens donarà la gràfica té força interès per a nosaltres, en representar certes poblacions dimòrfiques. També és del nostre interès mesurar l'asimetria (skewness), o bé, l'apuntament (kurtosis), de tot això ja en parlarem en un altre ocasió.

4 Assaig gràfic

La presentació gràfica de les d.f. es fa mitjançant els diagrames de freqüència.

Tenim dos tipus:

Per una d.f. qualitativa utilitzem el diagrama de barres, Fig. 2. Advertireu una característica important en aquest diagrama, les barres no es toquen, la qual cosa ens indica que la variable no és contínua.

Les d.f. quantitatives es fan per mitjà de variables contínues, com per exemple el pes de la Mallerenga cuallarga, i es representen gràficament amb un histograma, Fig. 3, en el qual l'amplitud de cada barra al llarg de l'abscisa representa un interval de classe de la d.f., i observeu també, que les barres es toquen per a mostrar que els límits implícits de les classes son adjacents. El punt mig de la barra correspon a la marca de classe.

La Fig. 4, ens ensenya un altre tipus de forma gràfica, el polígon de freqüències, per a representar la distribució de freqüències d'una variable contínua.

Deixem per més endavant, l'estudi de les diferents formes que podem obtenir amb aquests diagrames de freqüències i la seva interpretació biològica.

Fig. 4

