

INTRODUCCIÓN

- Estadística descriptiva - análise de datos
romanos → censos de poboación
- Cálculo de probabilidades - xogos de azar

Únen-se no século XIX formando a Estadística.

"BÁSICO"

Usamos a estadística para resolver os seguintes problemas:

- Descrición de datos
- Análise de mostras (resultados electorais, pezas dun coche)
- Construción de hipóteses
- Relacións entre variábeis
- Predicción (mellores cando o tempo é menor)

→Tipos de modelos estadísticos

Dinámicos

Estáticos

- Modelos extrapolativos (depende de sí mesmo)
- Modelos explicativos (imos empregar outros, depende de outras variábeis)

Modelo extrapolativo dinámico (póden-se combinar)

ESTADÍSTICA DESCRIPTIVA:

[Tema II Descrición Estadística dunha Variábel]

TEMA IIA DISTRIBUCIÓN DE FRECUENCIAS

- IIA.1.- INTRODUCCIÓN.
- IIA.2.- DISTRIBUCIÓN DE FRECUENCIAS.
- IIA.3.- REPRESENTACIÓN GRÁFICAS.

IIA.1.- Introducción.

Poboación conxunto de obxectos, persoas ou sucesos obxecto de estudo.

Individuo cada un dos elementos da poboación.

Mostra subconxunto da poboación que a representa.

Variábel unha característica da poboación que podemos estudar.

Variábel estatística

é aquela cuio valor se descoñece a priori.

Modalidade cada un dos valores que pode tomar determinada característica.

As variábeis son cualitativas (atributos) ou cuantivas.

[1] As variábeis cualitativas caracterízan-se porque os seus valores non son numéricos; p.ex. "estado civil" é unha variábel cualitativa (os seus valores son "solteiro/a", "casado/a", "viúvo/a", "divorciado/a").

[2] As variábeis cuantitativas toman valores numéricos:

[2.a] Variábeis cuantitativas discretas: toman valores dentro dun número numerábel de valores. P.ex. a idade dunha persoa en número enteiro de anos (valores 1, 2, 3, ...). P.ex. o número de caras que se obtéñen ao lanzar unha moeda.

[2.b] Variábeis cuantitativas contínuas: en principio existe un número infinito de valores posíbeis. P.ex. o peso dunha persoa, sempre se poderá atopar unha persoa cuio peso está entre o de outras dúas persoas.

Na práctica unha variábel considérase contínua cando toma moitos valores distintos. En realidade as variábeis contínuas acaban sendo *de algún xeito* discretas no sentido de que os instrumentos de medición teñen unha precisión limitada e tamén porque de cote nos interesa traballar con unha precisión de só unha determinada cantidade de cifras decimais.

O importante é que o tratamento que reciben as variábeis contínuas e o que reciben as discretas é totalmente diferente.

As variábeis tamén se distinguen entre unidimensionais e multidimensionais, segundo se fan referencia a unha soa característica da poboación ou se estudan máis dunha característica simultaneamente; p.ex. estudar conxuntamente altura e peso dunha poboación.

Variábeis	{	<u>unidimensionais</u>	cando fai referencia a unha única característica da poboación
		<u>multidimensionais</u>	cando se estudan varias características á vez P.ex. peso e estatura á vez

IIA.2.- DISTRIBUCIÓN DE FRECUENCIAS.

Sexa X una variábel que toma os valores x_i $i=1,2, \dots, m$ (a variábel X toma en cada caso un destes m valores)

Sexa n o número total de valores que toma a variábel na mostra, aínda que se repitan valores cóntan-se igual.

Agrupamos os datos nunha "táboa de frecuencias".

"FRECUENCIA ABSOLUTA" n_i

n_i do valor x_i
é o número de veces que aparece o valor x_i na mostra

$\sum_{i=1}^m n_i = n$

"FRECUENCIA RELATIVA" f_i

é o cociente entre frecuencia absoluta e o número total de valores na mostra
 é o tanto por un de ocurrencias do valor x_i na mostra

$$f_i = \frac{n_i}{n}$$

$$\sum_{i=1}^m f_i = 1$$

"FRECUENCIA ABSOLUTA ACUMULADA"

é o número de veces que se repite o valor x_i e todos os anteriores (menores) a el

$$N_i = \sum_{j=1}^i n_j = n_1 + n_2 + \dots + n_i$$

$$N_m = n$$

"FRECUENCIA RELATIVA ACUMULADA"

$$F_i = \sum_{i=1}^i f_i$$

Exemplo: Realizar a distribución de frecuencias da variábel pts-salário/día dunha compañía que tén un empregado na categoría 10.000 pts-salário/día, outro en 25.000, outro en 30.000; dous empregados en 18.000 e en 20.000 e tres en 15.000.

x_i	n_i	f_i	N_i	F_i
10.000	1	1/10	1	1/10
15.000	3	3/10	4	4/10
18.000	2	2/10	6	6/10
20.000	2	2/10	8	8/10
25.000	1	1/10	9	9/10
30.000	1	1/10	10=n	10/10=1
	<u> </u>	<u> </u>		
	n=10	1		

Nos resultados dos cocientes tóman-se duas ou tres cifras decimais. É importante verificar que unha vez feitos os redondeos se segue cumprindo que $\sum f_i = 1$

Cando temos muitos valores distintos non tén sentido asignar un x_i a cada un deles, senón que o que facemos é agrupar en intervalos que miden o mesmo (abarcán o mesmo rango de números "reais"). Non existen "saltos" entre os intervalos e dado un valor pertence só a un intervalo (i.e. os intervalos son disxuntos).

Como se toman os intervalos?

Tóman-se entre 5 e 20 do seguinte xeito:

$(L_0, L_1]$ x_1 =valor medio do intervalo (é o representante)
 $(L_1, L_2]$ x_2 =valor medio do intervalo (é o representante)

O representante de cada clase chámase marca de clase (m_1, m_2, \dots), deben de ser números sinxelos de manexar. As marcas de clase dos intervalos extremos calcúlan-se mediante a meia aritmética dos valores (e non dos extremos) do intervalo extremo. A todos os efectos de cálculo, os valores numéricos valerán o que a marca da clase dentro da que están.

Como dicimos, tóman-se entre 5 e 20 intervalos; cando o tamaño mostral non é moi grande sóen-se tomar 2^a clases.

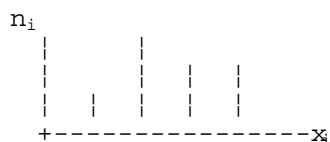
IIA.3.- REPRESENTACIONES GRÁFICAS.

IIA.3.1.- Variábeis cuantitativas: hai representacións non agrupadas e agrupadas en intervalos ou clases.

IIA.3.1.a.- Non agrupadas:

[1] *Diagrama de barras*

eixos cartesianos modalidade (OX) x frecuencia (OY)



* os f_i son proporcionais aos seus respectivos n_i

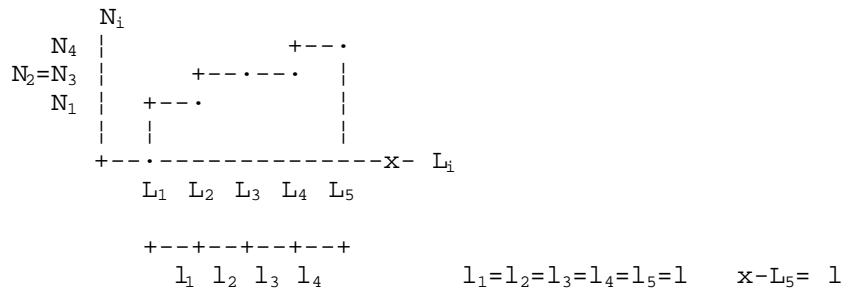
[2] *Polígono de frecuencias*

consiste en unir (x_i, n_i) con (x_{i+1}, n_{i+1})



* os f_i son proporcionais aos seus respectivos n_i

[3] *Polígono de frecuencias acumuladas* unindo $(L_1, 0)$ con (L_2, N_1)
 (L_2, N_1) con (L_3, N_2)
 (L_3, N_2) con (L_4, N_3)
 (L_4, N_3) con (L_5, N_4)
 (L_5, N_4) con x

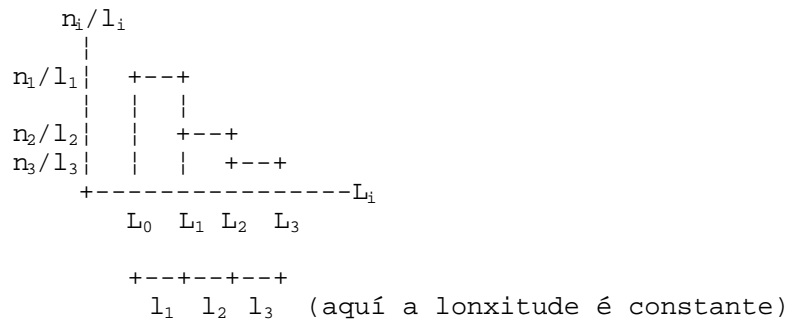


* tamén se lle chama diagrama en escaleira

IIA.3.1.b.- Agrupadas:

[4] *Histograma*

a área de cada rectángulo indica a frecuencia non acumulada



IIA.3.2.- Variábeis cualitativas.

—diagrama de sectores: unha circunferencia dividida en áreas mediante radios. O ángulo de cada sector é proporcional á súa frecuencia.

—pictograma: representación de frecuencias mediante debuxos. O tamaño dos iconos/debuxos é proporcional á frecuencia.

—cartograma: son mapas nos que mediante distintos cores, e áreas raiadas se representan zonas con determinadas características.