

10 Distribucións estatísticas

ÍNDICE DE CONTIDOS

1. VARIABLES ESTATÍSTICAS.....	2
1.1. Táboas de distribución de frecuencias.....	2
1.2. Representacións gráficas.....	3
2. MEDIDAS DE CENTRALIZACIÓN DUNHA VARIABLE ESTATÍSTICA.....	6
2.1. Media aritmética.....	6
2.2. Mediana.....	7
2.3. Moda.....	8
2.4. Cuartís.....	10
3. MEDIDAS DE DISPERSIÓN.....	11
3.1. Percorrido ou rango.....	11
3.2. Desviación media.....	11
3.3. Varianza.....	11
3.4. Desviación típica.....	12
4. COEFICIENTE DE VARIACIÓN.....	14

Na Estatística distinguimos dúas partes perfectamente diferenciadas.

A primeira delas dedícase a recoller datos, ordenalos, simplificalos, clasificalos e obter a partir deles un conxunto de valores que os identifican, e que ademais permiten facer comparacións con outros conxuntos de datos e estudar relacións entre eles; esta parte coñécese como **Estatística Descritiva**.

A outra parte da Estatística denomínase Inferencia estatística ou **Estatística Inferencial** e ten por obxecto facer afirmacións sobre unha poboación a partir dos datos recollidos dunha mostra desa poboación. En segundo curso de Bacharelato estudaremos os conceptos básicos da Inferencia estatística.

1. Variables estatísticas

A estatística estuda unha característica ou carácter dun conxunto de individuos chamado **poboación**. Cando a poboación é moi grande recórrese a un subconxunto denominado **mostra**.

O carácter en estudo pode adoptar diversas modalidades. A todas as modalidades que pode adoptar un determinado carácter chamámoslle **variable estatística**.

A variable estatística pode ser cuantitativa ou cualitativa. Por exemplo, na poboación dos alumnos dun instituto: idade, estatura, cualificacións, número de irmáns, etc., son caracteres cuantitativos; é dicir, son caracteres contables ou medibles. Mentres que o lugar de nacemento, o nome do seu equipo de fútbol favorito, etc., son caracteres cualitativos.

Os números que serven para contar ou medir un carácter cuantitativo, e que poden variar con cada individuo, constitúen o que se chama unha **variable estatística cuantitativa**.

As variables estatísticas cuantitativas poden ser continuas ou discretas. As variables son continuas se os valores que poden tomar son os números reais dun intervalo, mentres que as variables discretas só toman valores enteiros positivos.

1.1. Táboas de distribución de frecuencias

As variables estatísticas, cuantitativas ou cualitativas, xunto coas frecuencias de cada modalidade constitúen unha táboa de distribución de frecuencias ou, simplemente, unha **distribución de frecuencias**. Estas táboas teñen, ademais dos valores das variables e as súas frecuencias, outras columnas: a frecuencia relativa, a frecuencia acumulada, etc. Vexamos nos exemplos como se constrúen.

Exemplos

1. Nunha enquisa realizada a 20 familias, preguntóuselles polo número de fillos e anotáronse os seguintes datos:

2, 0, 1, 3, 1, 1, 2, 0, 1, 0, 2, 3, 5, 2, 1, 0, 2, 1, 1, 4.

Constrúe unha táboa de distribución de frecuencias absolutas e relativas. Poderías dicir qué porcentaxe de familias ten máis de dous fillos? E que porcentaxe de familias ten menos de tres fillos?

Solución: Chamando x_i aos valores da variable estatística, f_i ás súas frecuencias absolutas, h_i ás frecuencias relativas, onde n é o número de datos, F_i as frecuencias absolutas acumuladas e H_i as frecuencias relativas acumuladas podemos confeccionar a táboa seguinte:

x_i	f_i	h_i	F_i	H_i
0	4	0,20	4	0,20
1	7	0,35	11	0,55
2	5	0,25	16	0,80
3	2	0,10	18	0,90
4	1	0,05	19	0,95
5	1	0,05	20	1

As frecuencias relativas expresan porcentaxes, a suma das frecuencias relativas de máis de dous fillos son: $0,10 + 0,05 + 0,05 = 0,20$, o 20% das familias deste estudo ten máis de dous fillos.

A porcentaxe de familias con menos de tres fillos dánolo H_2 , $H_2 = 0,80 = 80\%$.

2. Tómase o pulso a un grupo de 30 persoas, obténdose os datos seguintes: 72, 66, 81, 74, 57, 58, 74, 62, 73, 65, 78, 75, 84, 72, 69, 76, 65, 79, 76, 68, 82, 71, 77, 72, 56, 62, 83, 63, 70, 73. Construír unha táboa de distribución de frecuencias.

Solución: Trátase dunha variable discreta aínda que con moitos valores. Cando os valores da variable son moi numerosos, agrupámoslos en clases ou intervalos.

Cantas clases? Para responder a esta pregunta pódese utilizar o sentido común e establecer que non excedan a decena, e dende logo non menos que 5. Existe unha fórmula chamada fórmula de Sturges que aconsella que sexan:

$$N^{\circ} \text{ de clases} = 1 + 3,32 \cdot \log n,$$

sendo **n** o número de datos; sempre que sexa posible, interésanos que as clases sexan todas da mesma amplitude.

No noso caso, é unha variable discreta cuxo valor menor é 56 e o maior 84, tomamos como rango dos valores de 55 a 84, e como clases: 55 - 59, 60 - 64, 65 - 69, 70 - 74, 75 - 79, 80 - 84. E construímos a táboa seguinte:

Clases	marca de clase x_i	f_i	F_i
55 - 59	57	3	3
60 - 64	62	3	6
65 - 69	67	5	11
70 - 74	72	9	20
75 - 79	77	6	26
80 - 84	82	4	30
		$n=30 \quad \Sigma f_i=30$	

A **marca de clase**, x_i , úsase como un valor representativo de todos os da clase, e corresponde ao valor central da clase, é dicir, a semisuma dos extremos.

1.2. Representacións gráficas

O obxecto das representacións gráficas é mostrar dun modo ordenado e agradable a información numérica para facilitar o seu estudo e interpretación. A partir das táboas de distribución de frecuencias constrúense representacións gráficas para este propósito. Vexamos algunhas moi coñecidas.

O diagrama de rectángulos

É unha gráfica que se emprega para variables cualitativas e cuantitativas discretas. Está formada por tantos rectángulos como valores ou modalidades da variable e todos con igual base. A altura dos rectángulos é igual á frecuencia absoluta ou relativa de cada valor. No caso do exemplo 2 anterior o diagrama de rectángulos sería:

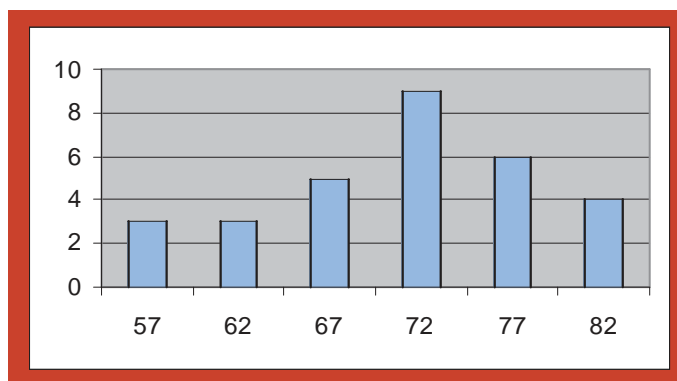


Diagrama de sectores

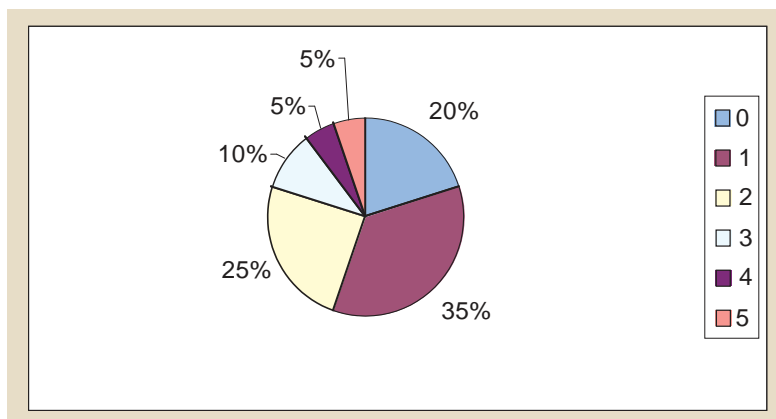
Emprégase tamén con variables cualitativas e cuantitativas discretas. Consiste en dividir un círculo en sectores proporcionais ás frecuencias. Para debuxar un sector necesitamos coñecer o ángulo central, e que se calcula pola proporción:

$$\frac{\text{ángulo do sector}}{360^\circ} = \frac{\text{frecuencia}}{n^\circ \text{ datos}}$$

No caso do exemplo 1 anterior correspondente ao número de fillos de 20 familias, os ángulos centrais son os seguintes:

$$\begin{aligned} \frac{\text{ángulo } 0}{360} &= \frac{4}{20} \Rightarrow \text{ángulo } 0 = 360 \cdot 4/20 = 72^\circ; & \frac{\text{ángulo } 1}{360} &= \frac{7}{20} \Rightarrow \text{ángulo } 1 = 360 \cdot 7/20 = 126^\circ \\ \frac{\text{ángulo } 3}{360} &= \frac{5}{20} \Rightarrow \text{ángulo } 2 = 360 \cdot 5/20 = 90^\circ; & \text{ángulo } 3 &= 36^\circ; \text{ángulo } 4 = 18^\circ; \text{ángulo } 5 = 18^\circ \end{aligned}$$

Cun transportador de ángulos, sobre un círculo, non é difícil debuxar un diagrama de sectores como este:



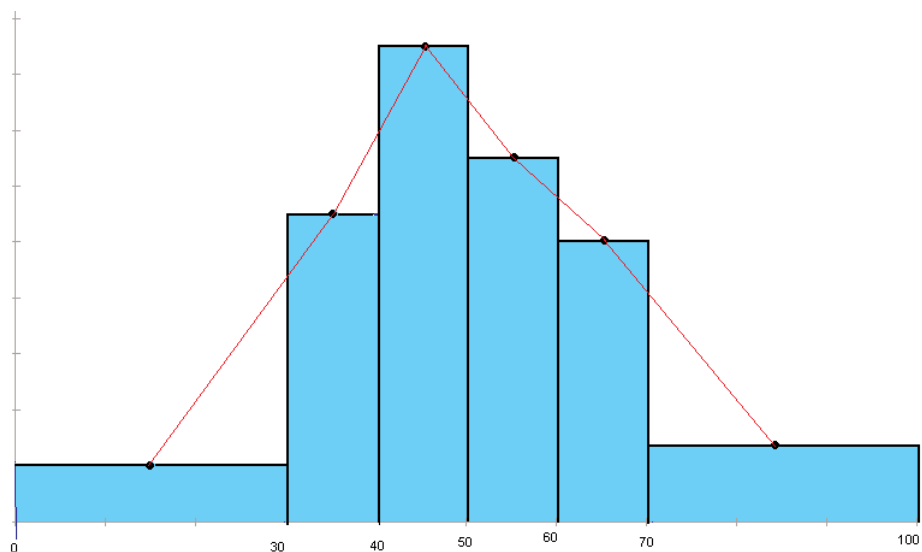
Histograma

Cando a variable é cuantitativa e os datos están agrupados en intervalos, constrúese un diagrama de rectángulos especial, chamado histograma, no que as áreas destes, e non as alturas, indican as frecuencias. Ao ser a área a frecuencia, nos histogramas cúmprese que: **Frecuencia = amplitude da clase · altura**

Por exemplo, se tivésemos unha distribución como:

Clases	f_i	h_i
0-30	6	0,10
30-40	11	0,17
40-50	17	0,26
50-60	13	0,20
60-70	10	0,15
70-100	8	0,12
	65	1

que corresponde ás puntuacións obtidas por 65 persoas nun test de 100 preguntas, entón o histograma sería:



Ao histograma engadímoslle o **polígono de frecuencias**, que aparece se unimos por trazos rectos os puntos medios das bases superiores dos rectángulos do histograma. Cando o número de datos crece indefinidamente e ao mesmo tempo facemos os intervalos cada vez máis pequenos, o polígono de frecuencias tende a unha curva suave e continua que, se se trata de frecuencias relativas, é coñecida co nome de función densidade, e da que oiremos falar na Unidade 12.

2. Medidas de centralización dunha variable estatística

Ás veces é necesario dispoñer dun valor numérico que represente a diversidade de valores dunha distribución de frecuencias dunha variable estatística. Aos valores numéricos que cumpren esta función chámase-lles parámetros centrais ou medidas de centralización dunha distribución, e como sabemos son: **a media, a mediana e a moda**.

2.1. Media aritmética

A **media aritmética** dun conxunto de números é o cociente que resulta de dividir a suma de todos os números polo total destes. Representase por \bar{x} .

Se os valores dunha variable estatística non teñen frecuencias, e son:

$x_1, x_2, x_3, \dots, x_n$ é dicir N valores, a media aritmética calcúlase pola fórmula,

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N} = \frac{\sum_{i=1}^n x_i}{N}$$

Se os datos veñen con frecuencias e os N datos distribúense en k valores da variable, ou das marcas de clase cando están agrupados, e estes: $x_1, x_2, x_3, \dots, x_k$ e f_1, f_2, \dots, f_k , as frecuencias respectivas, onde $f_1 + f_2 + \dots + f_k = N$, entón a media aritmética calcúlase coa fórmula:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + x_3 \cdot f_3 + \dots + x_k \cdot f_k}{N} = \frac{\sum_{i=1}^k f_i x_i}{N}$$

A media aritmética é a medida ou parámetro de centralización máis empregado. Con todo, está influída polos valores extremos dunha distribución ou por algún valor extravagante, nestes caso perde algo de significado. Reflicte, non obstante, todas as alteracións que sufran os datos:

- Se se suma a mesma cantidade a todos os valores da variable, a media resulta aumentada nesa cantidade.
- Se se multiplican todos os valores da variable polo mesmo número, a media resulta multiplicada por ese número.

Existe ademais outra media dun conxunto de n números, que se chama **media ponderada**, e calcúlase cando os datos non teñen todos o mesmo peso. A media ponderada obtense sumando todos os produtos de cada valor polo seu peso dividindo o resultado pola suma dos pesos. A fórmula é:

$$\bar{x} = \frac{x_1 \cdot p_1 + x_2 p_2 + x_3 p_3 \dots + x_n p_k}{p_1 + p_2 + p_3 \dots + p_k} = \frac{\sum_{i=1}^k p_i x_i}{\sum_{i=1}^k p_i}$$

2.2. Mediana

Mediana é o valor da variable estatística, supoñendo que os datos estean ordenados, que ocupa a posición central; é dicir, deixa á súa esquerda o mesmo número de datos que á súa dereita. Simbolízase por *Me*.

Nunha distribución sinxela de variable discreta, a mediana corresponde ao valor central se o número de datos é impar, pero se o número de datos é par, a mediana é a media dos dous valores centrais.

En: 0, 0, 1, 1, 1, 1, 2, 2, 2 => **Me = 1.**

En: 1, 2, 3, 4, 5, 6, 7, 8, => **Me = (4+5)/2 = 4,5**

No caso dunha distribución con moitos datos, pero sen agrupar, a mediana é o primeiro valor cuxa frecuencia absoluta acumulada está por enriba da metade dos datos.

Por exemplo, na distribución

n	x_i	0	1	2	3	4	=27, $n/2 = 13,5$ e frecuencias
as	f_i	5	6	9	4	3	

absolutas acumuladas son: 5, 11, 20, 24, 27, a primeira frecuencia acumulada por enriba do metade dos datos é 20, e corresponde a $x_3=2$, entón **Me = 2.**

Non obstante, na distribución

x_i	0	1	2	3	4
f_i	5	7	6	4	2

$n=24$, $n/2 = 12$ e a frecuencia acumulada correspondente ao valor 1 é tamén 12, neste caso tomamos como mediana a media entre os dous valores as frecuencias acumuladas dos cales son igual e maior que $n/2$,

$$Me = (1+2)/2 = 1,5$$

No caso dunha variable continua, os datos da cal se atopan agrupados en intervalos, chámase **clase mediana** a aquela cuxa frecuencia absoluta acumulada supera a metade de datos.

Como valor aproximado da mediana pode tomarse a marca de clase da clase mediana.

A principal vantaxe da mediana como medida de centralización é que non está influída polos valores extremos, nin por datos extravagantes, aínda que ten serios inconvenientes: non é doado o seu emprego en operacións alxébricas e non ten en conta o valor de todos os datos.

2.3. Moda

Coñécese como **Moda** dunha variable estatística ao valor que ten maior frecuencia absoluta. Simbolízase por **Mo**.

Cando a variable é discreta a moda obtense buscando o valor da variable que ten maior frecuencia. Ás veces, a moda non é única, é dicir, a distribución pode ter dúas, tres ou máis modas, en cuxos casos recibe o nome de bimodal, trimodal, etc.

Se os datos se atopan agrupados en intervalos, a clase de maior frecuencia chámase **clase modal**. É frecuente tomar como moda a marca de clase da clase modal.

A moda ten as mesmas vantaxes e inconvenientes que a mediana: non está influída polos valores extremos, nin por datos extravagantes, pero non ten en conta o valor de todos os datos, nin se pode operar alxeбраicamente con ela.

Exemplo: O tempo en minutos que tardan 87 empregados dunha empresa en chegar dende a súa casa ao traballo móstrase na táboa seguinte:

Minutos	(0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35]
empregado	6	7	18	22	17	11	6

- Calcula a duración media dos traxectos de casa ao traballo dos 87 empregados.
- Calcula a mediana correspondente á variable duración do traxecto casa-traballo.
- Calcula a moda correspondente á variable duración do traxecto casa-traballo.

Solución: a) Construimos unha táboa para calcular a duración media

$$\bar{x} = \frac{1557,5}{87} = 17,9$$

Polo tanto, a duración media é **17,9 minutos**.

Minutos	Marca de clase, x_i	n_i	$x_i n_i$
(0, 5)	2,5	6	15
[5, 10)	7,5	7	52,5
[10, 15)	12,5	18	225
[15, 20)	17,5	22	385
[20, 25)	22,5	17	382,5
[25, 30)	27,5	11	302,5
[30, 35]	32,5	6	195
	Totais	87	1557,5

Cunha calculadora científica é aínda máis doada, coas teclas **MODE** · poñemos a calculadora en modo estatístico, na pantalla aparece SD (estándar deviation), e xa están activas as teclas escritas en azul.

Borramos, antes que nada os datos da memoria, por se os houbese, coa tecla SHIFT SAC e introducimos os valores e as súas frecuencias

2.5 × 6 **DATA** 7.5 × 7 **DATA** 12.5 × 18 **DATA** 17.5 × 22 **DATA** 22.5 × 17 **DATA**

27.5 × 11 **DATA** 32.5 × 6 **DATA** Unha vez introducidas os datos as teclas SHIFT e \bar{x} dan a media.

b) Como $n/2 = 43,5$ a clase mediana é [15, 20), a primeira cuxa frecuencia acumulada supera á metade dos datos, a marca de clase correspondente, 17,5, será a mediana.

c) A maior frecuencia é 22 e a clase modal será tamén [15, 20). Un primeiro valor para a moda sería a marca de clase 17,5.

2.4. Cuartís

Se ordenamos os datos de menor a maior, chámanse **cuartís** aos valores da variable que dividen ao conxunto de datos en catro partes iguais. O primeiro simbolízase por Q_1 e deixa á súa esquerda o 25% dos datos e á súa dereita o 75%, o segundo cuartil coincide coa mediana, $Q_2 = Me$, e o terceiro cuartil simbolízase por Q_3 e deixa á súa esquerda o 75% dos datos e á súa dereita o 25%.

O cálculo dos cuartís primeiro e terceiro é similar ao da mediana e obtéñense polas fórmulas:

$$Q_1 = e_1 + \frac{\frac{n}{4} - F_{Q_1-1}}{f_{Q_1}} \cdot c \quad \text{e} \quad Q_3 = e_3 + \frac{\frac{3n}{4} - F_{Q_3-1}}{f_{Q_3}} \cdot c$$

onde:

e_1 ou e_3 é o extremo inferior da clase á que pertence o cuartil;

c é a amplitude da clase á que pertence o cuartil;

n é o número total de datos;

f_{Q_1} o f_{Q_3} é a frecuencia absoluta da clase á que pertence o cuartil;

F_{Q_1-1} ou F_{Q_3-1} é a frecuencia absoluta acumulada inmediatamente anterior a a da clase á que pertence o cuartil.

Como no caso da mediana, a clase á que pertence Q_1 é aquela cuxa frecuencia absoluta acumulada supera a cuarta parte dos datos, e a clase á que pertence Q_3 é aquela cuxa frecuencia absoluta acumulada supera as tres cuartas partes dos datos.

Exemplo:

Un test de 100 preguntas, realizado a 90 alumnos, deu as puntuacións que figuran na táboa adxunta.

x_i	f_i
0-39	2
40-49	3
50-59	9
60-69	17
70-79	32
80-89	19
90-100	8

Achar Q_1 , Q_2 ou M_e e Q_3 .

Solución: A cuarta parte, a metade e as tres cuartas partes dos datos son: 22,5, 45 y 67,5; polo tanto, Q_1 pertence á clase 60 -69, a clase mediana é 70 -79 e Q_3 pertence a 80 -89.

Empregamos a fórmula que permite calcular os cuartís,

$$Q_1 = 60 + \frac{\frac{90}{4} - 14}{17} \cdot 10 = 65 \quad Q_3 = 80 + \frac{\frac{3 \cdot 90}{4} - 63}{19} \cdot 10 = 82,36$$

Entre 65 e 82,36 están o 50% das puntuacións centrais rexistradas, a diferenza $Q_3 - Q_1$ denomínase o percorrido intercuartílico $82,36 - 65 = 17,64$

Aos cuartís, e a outras medidas semellantes chamadas decís e percentís, tamén se coñecen como medidas de posición porque permiten situar un dato con respecto ao resto da distribución.

3. Medidas de dispersión

Pode acontecer que dúas distribucións distintas teñan os parámetros centrais moi parecidos ou iguais e non obstante as distribucións ser completamente diferentes. Polo tanto, necesítanse outras medidas que nos indiquen o grao de variación dos datos respecto á media. Estas novas medidas, que chamamos parámetros ou medidas de dispersión, informan do grao de dispersión ou de concentración dos datos en relación á media. Vexamos cales son.

3.1. Percorrido ou rango

O **percorrido ou rango** dunha variable estatística é a diferenza entre o maior e o menor valor dos datos. Simbolízase R , é dicir,

$$R = x_{max} - x_{min}$$

3.2. Desviación media

Defínese a **desviación media** dunha variable estatística á suma dos produtos dos valores absolutos das desviacións dos valores ou marcas de clase respecto da media aritmética multiplicados polas súas frecuencias e divididos polo total dos datos. Simbolízase por DM e vén dada por:

$$DM = \frac{\sum_{i=1}^k f_i |x_i - \bar{x}|}{N}$$

3.3. Varianza

A **varianza** é a media aritmética dos cadrados das desviacións de todos os valores da variable estatística ou marcas de clase respecto da media, e simbolízase por s^2 .

Se os valores da variable son $x_1, x_2, x_3, \dots, x_k$ e non se repite ningún, entón

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_k - \bar{x})^2}{N} = \frac{\sum_{i=1}^k (x_i - \bar{x})^2}{N}$$

A varianza, s^2 , tamén se pode expresar por

$$s^2 = \frac{\sum_{i=1}^k (x_i)^2 \cdot f_i}{N} - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

Dicese como nun xogo de palabras que a varianza é a media dos cadrados menos o cadrado da media

Se os valores (ou as marcas de clase) veñen con frecuencias, a varianza calcúlase coas fórmulas que figuran a continuación que, como vimos anteriormente no caso sinxelo, son equivalentes:

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{N}$$

A varianza é sempre positiva, pero ten un pequeno inconveniente non se expresa nas mesmas unidades que os datos; é dicir, se por exemplo os datos son en cm a varianza sería cm^2 .

3.4. Desviación típica

Para emendar o pequeno defecto da varianza defínese a **desviación típica** dunha variable estatística como a raíz cadrada positiva da varianza. E esta si ten as mesmas unidades que os datos. Representámola pola letra s e vén expresada por:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{N}}$$

A desviación típica é o parámetro de dispersión máis utilizado. E ten algunhas propiedades interesantes:

Se sumamos unha cantidade constante a todos os valores da variable, a desviación típica non varía.

Se multiplicamos todos os valores da variable pola mesma cantidade, a desviación típica queda multiplicada por esa cantidade.

Exemplos

1. Rexistráronse as distancias en km, dende a súa casa ao traballo, de 40 empregados dunha empresa situada nun polígono industrial, e resultaron os seguintes datos: 3,

5, 10, 15, 20, 25, 3, 6, 12, 18, 23, 28, 4, 8, 14, 15, 22, 25, 8, 19, 10, 20, 9, 12, 16, 11, 15, 13, 18, 23, 10, 17, 22, 12, 15, 25, 14, 18, 20, 22.

Atopar a media, a varianza e a desviación típica

Solución: Agrupamos os datos en clases, de amplitude 5 km, que inclúan o extremo esquerdo do intervalo, pero non o dereito. Construimos cos datos subministrados a táboa seguinte:

Clases en km	Marcas de clase x_i	f_i	$x_i \cdot f_i$
[0, 5)	2,5	3	7,5
[5, 10)	7,5	5	37,5
[10, 15)	12,5	10	125,5
[15, 20)	17,5	10	175,5
[20, 25)	22,5	8	180
[25, 30)	27,5	4	110
		$\Sigma f_i = 40$	$\Sigma x_i \cdot f_i = 630$

A media das distancias será: $\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum f_i} = \frac{630}{40} = 15,75$ quilómetros

A varianza é a media dos cadrados das desviacións, ou separacións, de cada un dos valores da variable respecto á media aritmética.

A desviación típica dunha distribución estatística é a raíz cadrada da varianza.

$$s^2 = \frac{\sum_{i=1}^k (x_i)^2 \cdot f_i}{N} - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

Completamos o cadro anterior coa columna $x_i \cdot f_i$. Logo multiplicando a primeira columna da táboa pola terceira obtense a columna encabezada por x

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
2,5	3	7,5	18,75
7,5	5	37,5	281,25
12,5	10	125,5	1562,5
17,5	10	175,5	3062,5
22,5	8	180	4050
27,5	4	110	3025
		$\Sigma f_i = 40$	$\Sigma x_i^2 \cdot f_i = 12000$

$$\text{A varianza } s^2 = \frac{1200}{40} - 15,75^2 = 51,9375$$

$$\text{A desviación típica } s = \sqrt{51,9375} = 7,2067$$

2. Ao pesar 32 profesores dun Instituto obtivéronse os seguintes datos en kg:
48,5, 52, 58, 60, 65, 70,5, 77, 83, 89, 49, 53, 59, 61,5, 68, 74, 78,5, 55, 62, 67,
73, 63, 66, 71,5, 62, 65, 72, 64, 66, 71, 63,5, 69, 68,5.

Calcular a media, a varianza e a desviación típica.

Solución. - O maior valor é 89 e o menor 48,5. Agrupámoslos en clases ou intervalos semipechados, de 5 kg, comezando en 45 e rematando en 90. Facemos un cadro coas clases, as marcas de clase e a frecuencia.

Clases	x_i	f_i
[45, 50)	47,5	2
[50, 55)	52,5	3
[55, 60)	57,5	2
[60, 65)	62,5	7
[65, 70)	67,5	8
[70, 75)	72,5	6
[75, 80)	77,5	2
[80, 85)	82,5	1
[85, 90)	87,5	1
		$f_i=32$

Este exemplo ímolo resolver con axuda dunha calculadora.

1º) Establecemos o modo estatístico. Pulsamos

. MODE . . Aparece en pantalla as letras SD

2º) Borramos o contido da memoria estatística

3º) Introducimos os valores da táboa: SHIFT SAC.

47.5 × 2 **DATA** 52.5 × 3 **DATA** 57.5 × 2 **DATA** 62.5 × 7 **DATA** 67.5 × 8 **DATA**

72.5 × 6 **DATA** 77.5 × 2 **DATA** 82.5 **DATA** 87.5 **DATA**

4º) Con SHIFT x obtemos a media, $\bar{x} = 65,78125$

Con SHIFT σ obtemos a desviación típica, $s = 9,2372...$

Con SHIFT σ SHIFT x^2 obtemos a varianza, $s^2 = 85,3271....$

Nas calculadoras aparece a tecla σ_{n-1} que corresponde a un parámetro cuxo cadrado chámase cuasivarianza e ten importancia no estudo de estimacións estatísticas, pero que non é noso caso. Noutros modelos e marcas de calculadoras debe consultarse o manual de uso, aínda que todas teñen instrucións similares.

4. Coeficiente de variación

Ao comparar a dispersión de dúas distribucións, pode acontecer que teñan a mesma desviación típica e que as dispersións sexan totalmente diferentes. Non é o mesmo unha desviación típica de 2 kg nunha poboación de polos que nunha de vacas. Polo que imos definir un parámetro que nos permite medir a desviación típica, tomando a media como unidade, e consiste en calcular o cociente $\frac{s}{\bar{x}}$

A este cociente chámase **coeficiente de variación** e adoita expresarse en tantos por cen, polo tanto haberá que multiplicalo por 100, e simbolizámolo por V_p ,

$$V_p = \frac{s}{\bar{x}} \cdot 100$$

O coeficiente de variación mide, en tantos por cen, a desviación dunha dispersión con respecto á media. Canto máis pequeno sexa o coeficiente de variación, os datos estarán máis concentrados arredor da media, e esta resultará máis significativa.

Exemplo

Temos dúas marcas de lámpadas, A e B, eliximos dúas mostras de 10 lámpadas de cada marca e anotamos a duración, en horas, de cada unha delas, resultando:

A	260	285	250	350	290	280	245	265	230	235
B	250	310	270	290	325	300	280	285	240	315

Calcula o coeficiente de variación de cada mostra.

Solución:

Empregando a calculadora obtemos que nas lámpadas da marca A, $\bar{x}_A = 269$ e $s_A = 33,3$, mentres que nas lámpadas da marca B, $\bar{x}_B = 286,5$ y $s_B = 26,17$.

Os coeficientes de variación serán: $V_A = \frac{33,3}{269} \times 100 = 12,37\%$ para a marca A, e $V_B = \frac{26,17}{286,5} \times 100 = 9,13\%$ para a marca B.

A marca B ten unha duración media maior e ademais a dispersión é menor con respecto á media.