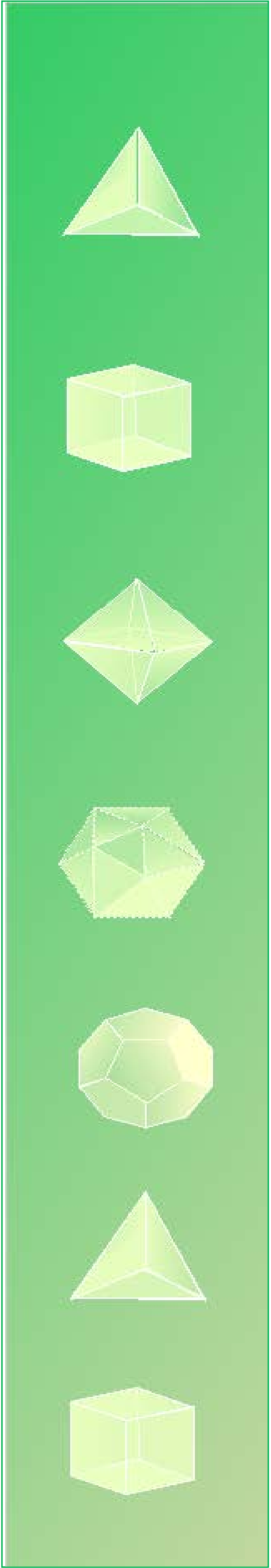


Matemáticas orientadas ás ensinanzas aplicadas

4ºA da ESO

Capítulo 7: Estatística. Azar e probabilidade

www.apuntesmareaverde.org.es

Autores: María Molero e Andrés García Mirantes

Revisoras: Raquel Caro e Nieves Zuasti

Tradutora: Mª Teresa Seara Domínguez

Revisora da tradución ao galego: Fernanda Ramos Rodríguez

Ilustracións: Banco de Imaxes de INTEF

Índice

Resumo

A estatística ocúpase de interpretar gran número de datos. O Instituto Nacional de Estatística recolle estudos de todo tipo sobre a poboación española. Entra en Internet escribindo INE e terás un montón de información ao teu alcance sobre: a) Entorno físico e medio ambiente; b) Demografía e poboación; c) Sociedade; d) Economía…

Nun estudo estatístico conflúen distintas partes da estatística, a Teoría de Mostras que indica a forma de seleccionar unha mostra para que sexa representativa da poboación, a estatística descritiva que utiliza táboas, gráficos e parámetros estatísticos como a media e a desviación típica para describir os datos, e a inferencia estatística que utiliza a Teoría de Probabilidades para obter conclusións.

Como saberás, en tempo de Xesucristo, xa o emperador Augusto fixo censos para coñecer a poboación do Imperio Romano.

A Teoría da probabilidade tivo os seus inicios moi ligados aos xogos de azar, e é sorprendente que con ese inicio teña resultado de tanta utilidade na Ciencia. Preguntábanse que é máis probable ao tirar dous dados, que a suma das súas caras superiores sexa 9 ou sexa 10. Analizando xogos como este foi avanzando a Ciencia.

1. ESTATÍSTICA

1.1. MOSTRAS. ESTUDOS ESTATÍSTICOS

Se queremos facer un estudo estatístico temos que:

a) Recoller os datos.

b) Describir eses datos con táboas e gráficas, cálculo de parámetros estatísticos….

c) Extraer conclusións.

Para recoller os datos e determinar os valores da variable pódese utilizar toda a poboación, todo o universo sobre o que se realiza o estudo, ou facer unha mostra. En moitas ocasións non é conveniente recoller valores de toda a poboación, porque é complicado ou demasiado custoso, ou mesmo porque é imposible como no caso dun control de calidade no que se destrúa o obxecto a analizar. A parte da estatística que se ocupa de cómo seleccionar adecuadamente as mostras denomínase Teoría de Mostras.

Poboación ou universo é todo o conxunto de individuos sobre o que se realiza o estudo.

Unha **mostra** é un subconxunto representativo desa poboación.

Cada un dos elementos da poboación é un **individuo**.

As características da poboación que se estudan denomínanse **variables estatísticas**, que se clasifican en **cuantitativas** e **cualitativas** segundo os valores que tomen sexan ou non numéricos. As variables cuantitativas que toman valores illados denomínanse **variables discretas** e as que poden tomar calquera valor dun intervalo da recta real, **variables continuas**.

A parte da Estatística que ordena, analiza e representa un conxunto de datos para describir as súas características denomínase **Estatística Descritiva**.

Para extraer conclusións utilízanse as probabilidades e a parte da Estatística que se ocupa diso é a **Inferencia Estatística**.

Exemplos:

- Se queremos coñecer as preferencias en deportes do alumnado de 4º, é posible preguntar a toda a poboación (alumnado de 4º), aínda que é adecuado elixir unha mostra representativa, seleccionando algúns estudantes.
- Neste estudo sobre preferencias deportivas, a variable utilizada é cualitativa.
- Para coñecer a intención de voto perante unhas eleccións europeas, municipais, autonómicas… utilízanse mostras, pois preguntar a toda a poboación sería moi custoso (e iso xa se fai nas eleccións). A variable neste caso tamén é cualitativa.
- Para estudar o que máis preocupa a unha poboación: paro, terrorismo, corrupción… tamén se utilizan mostras. Neste caso sería moi custoso preguntar a toda a poboación, aínda que sería factible. A variable neste caso tamén é cualitativa.
- Pero se unha fábrica quere coñecer as horas de vida útil dunha lámpada, unha neveira, un camión… non pode poñer a funcionar toda a poboación (todas as lámpadas ou neveiras ou camións…) ata que se avaríen pois queda sen produción. Neste caso é imprescindible seleccionar unha mostra. A variable neste caso é cuantitativa e o tempo toma calquera valor, é unha variable cuantitativa continua.
- Se preguntamos polo número de irmáns é unha variable cuantitativa discreta.
- En *control de calidade* fanse estudos estatísticos e tómanse mostras.

Actividades propostas

1. Queremos realizar un estudo estatístico sobre o tempo dedicado ao estudo polo alumnado de ESO de Madrid. Para iso selecciónanse adecuadamente 100 alumnos. Indica cal é a poboación, cal a mostra, que tamaño ten a mostra e quen sería un individuo.

2. Queres pasar unha enquisa para coñecer, o mesmo que no problema anterior, o tempo dedicado ao estudo, neste caso o dos compañeiros e compañeiras do teu centro escolar. Pasaríaslla só ás mozas? Só aos mozos? Preguntarías aos mellores da clase? Aos de peores notas? Indica o criterio que seguirías para seleccionar a mostra á que preguntar.

1.2. Variable discreta. Táboas e gráficos

Táboas

Ao facer un estudo estatístico ou realizar un experimento aleatorio a información obtida resúmese nunha táboa ou distribución de frecuencias.

Exemplo:

Posibles resultados	Frecuencia absoluta
Gústalles	28
Non lles gusta	12
Total	40

- Preguntamos a 40 estudantes de 4º se lles gusta, ou non, o fútbol. Na táboa da marxe reflectimos os resultados.

É unha táboa de frecuencias absolutas.

Ao dividir a frecuencia absoluta entre o número total temos a frecuencia relativa, así a frecuencia relativa dos que lles gusta o fútbol é 28/40 = 0.7, e a dos que non lles gusta o fútbol é 12/40 = 3/10 = 0.3.

A **frecuencia absoluta** é o número de veces que se obtivo ese resultado.
A **frecuencia relativa** obtense dividindo a frecuencia absoluta entre o número total de datos.

A suma das frecuencias relativas é sempre igual a 1.

Multiplicando por 100 obtéñense as porcentaxes.

Posibles resultados	Frecuencias relativas	Porcentaxe
Gústalles	0.7	70
Non lles gusta	0.3	30
Suma total	1	100

Actividade resolta

- *Obtivéronse os datos sobre o número de visitas que se fixeron aos Textos Marea Verde de Matemáticas nos meses indicados e reflectíronse nunha táboa. Fai unha táboa de frecuencias absolutas, relativas e porcentaxes, de frecuencias acumuladas absolutas e de frecuencias relativas acumuladas.*

Marea verde	Frecuencias absolutas	Frecuencias relativas	Porcentaxes	Frecuencias acumuladas absolutas	Frecuencias acumuladas relativas
Setembro	1 834	0.51	51	1 834	0.52
Outubro	956	0.26	26	2 790	0.77
Novembro	432	0.12	12	3 222	0.89
Decembro	389	0.11	11	3 611	1
TOTAL	3 611	1	100		

Observa que as **frecuencias acumuladas** se obteñen sumando a frecuencia anterior e indica, neste exemplo, o número de visitas ata ese momento.

Resultados	Frecuencias absolutas
1	17
2	12
3	17
4	15
5	21
6	14

Actividades propostas

3. Copia no teu caderno e completa a seguinte táboa de frecuencias absolutas dos valores obtidos ao tirar un dado, coas frecuencias relativas e porcentaxes, e con frecuencias acumuladas absolutas e frecuencias relativas acumuladas.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Gráficos estatísticos

As representacións gráficas axudan a comprender o significado dos datos.

Dada unha táboa de frecuencias (absolutas, relativas, porcentaxes, acumuladas absolutas ou acumuladas relativas) para representar un **diagrama de rectángulos ou de barras** trázase para cada valor da variable un rectángulo ou barra da altura proporcional á frecuencia que se estea representando.

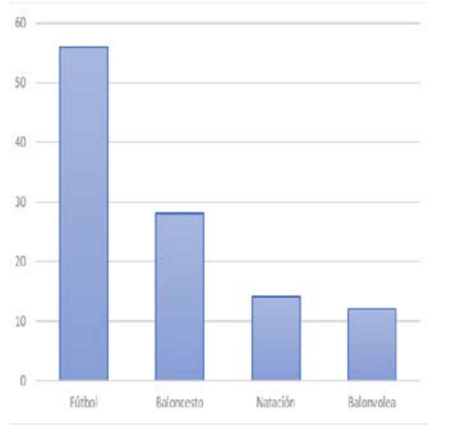
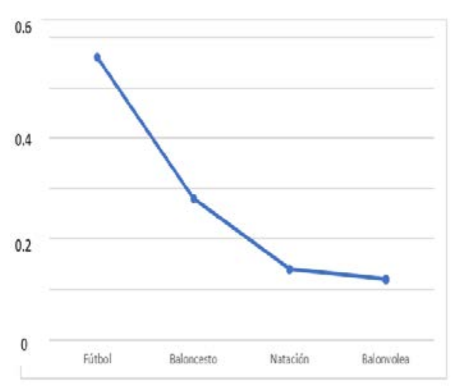
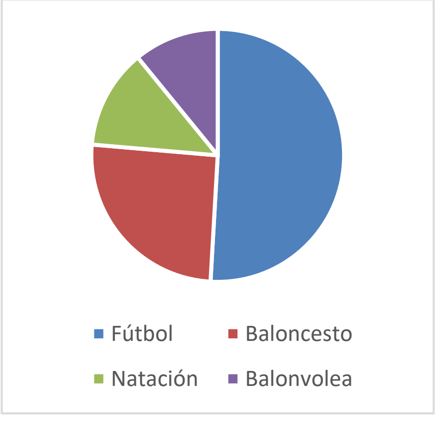
Se se unen os puntos medios dos extremos superiores das barras temos un **polígono de frecuencias ou diagrama de liñas**.

Nun **diagrama de sectores** debúxase un círculo que se divide en sectores de amplitudes proporcionais ás frecuencias.

Actividade resolta

- *Temos un estudo estatístico sobre as preferencias deportivas do alumnado de 4º dun determinado centro escolar. Represéntaos nun diagrama de barras de frecuencias absolutas, nun polígono de frecuencias relativas e nun diagrama de sectores.*

Deportes	Frecuencia Absoluta
Fútbol	56
Baloncesto	28
Natación	14
Balonvolea	12

Diagrama de barras de frecuencias absolutas	Polígono de frecuencias relativas ou diagrama de liñas	Diagrama de sectores
		

Actividades propostas

4. Coa táboa de valores do exercicio anterior, debuxa no teu caderno o diagrama de frecuencias relativas, o polígono de frecuencias absolutas acumuladas e o diagrama de sectores.

5. Fai un estudo estatístico preguntando aos teus compañeiros e compañeiras de clase sobre o número de libros que len ao mes. Confecciona unha táboa e represéntaa nun diagrama de rectángulos, un polígono de frecuencias e un diagrama de sectores.

6. Selecciona unha mostra entre os teus compañeiros e compañeiras e realiza un estudo estatístico sobre o deporte que máis lle gusta a cada un. Fai a representación que sexa máis sinxela de interpretar.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Utiliza o ordenador

- As follas de cálculo son unha ferramenta moi útil para traballar a estatística. Suman, multiplican, e debuxan os gráficos con gran facilidade. Para a Actividade resolta anterior, copiamos a táboa cos datos na folla de cálculo a partir da casa A1. Calculamos a suma total na casa B6, simplemente apertando a tecla: $\sum$, ou ben escribindo =SUMA(B2:B5) que significa que queremos sumar o que hai desde a casa B2 á B5.

Para calcular as frecuencias relativas escribimos en C1: Frecuencia relativa e, en C2, escribimos o signo igual (co que lle estamos dicindo á folla que imos calcular algo), facemos clic na casa B2, escribimos: /, e facemos clic en B6: =B2/B6, sáenos *0,50909…* A casa B2 va ir variando cando calculemos C3, C4…, pero queremos que a casa B6 quede fixa. Para dicir iso, poñemos o símbolo $: =B2/$B$6. E agora arrastramos ata a casa C5 (se arrastramos antes de poñer o $ sáenos un erro, pois está dividindo por cero ao ir modificando a casa). Temos as frecuencias relativas calculadas.

Para debuxar os gráficos só temos que seleccionar as filas e columnas que nos interesen e no menú de "Inserir" seleccionar o tipo de gráfico desexado: Columna, Liña, Circular…

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

1.3. Parámetros de centralización e dispersión

Parámetros de centralización

Xa sabes que os parámetros de centralización nos dan información sobre o "*centro*" dun conxunto de datos. Estudamos a media aritmética, a moda e a mediana.

Actividade resolta

- *Neves obtivo en Matemáticas as seguintes notas: 8, 4, 6, 10 e 10. Calcula a súa media, a súa moda e a súa mediana.*

A súa nota media calcúlase sumando todas as notas: 8 + 4 + 6 + 10 + 10 = 38, e dividindo a suma entre o número total de notas que é 5: 38/5 = 7.6.

A moda é 10 pois é o valor máis frecuente.

Unha forma de calcular a mediana é ordenar os valores de menor a maior, e se o número de datos é impar, o valor central é a mediana. Se o número de datos é par, a mediana é a media dos dous datos centrais.

No noso caso: $4 \leq 6 \leq \mathbf{8} \leq 10 \leq 10$, polo que a mediana é 8.

Para calcular a **media** (**m**) de $x_1, x_2, \ldots, x_n$, súmanse todos e divídese polo número total de datos (n).

$$\textbf{Media} = m = (x_1 + x_2 + \ldots + x_n)/n$$

Que é o que está de moda? O que máis se leva.

A **moda** (***mo***) dunha distribución de frecuencias é o valor máis frecuente.

A **mediana** (***me***) é o valor central que deixa por debaixo o mesmo número de valores da variable que por enriba.

Utiliza o ordenador

- Para calcular a media, a mediana e a moda coa folla de cálculo, copiamos na casa B2, B3… os datos: 8, 4, 6, 10 e 10. Escribimos na casa A7, Media, e para calcular a media escribimos un signo igual en B7. Buscamos, despregando as posibles funcións, a función TERMO MEDIO, e escribimos

  =TERMO MEDIO(B2:B6),

que significa que calcule a media dos valores que hai nas casas desde B2 ata B6.

Do mesmo modo calculamos a mediana buscando nas funcións ou escribindo =MEDIANA(B2:B6) e a moda buscando nas funcións ou escribindo =MODA(B2,B6).

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Actividades propostas

7. Dadas as temperaturas nunha cidade a unha hora determinada o día 1 de cada mes tense a seguinte táboa:

	Xaneiro	Febreiro	Marzo	Abril	Maio	Xuño	Xullo	Agosto	Setembro	Outubro	Novembro	Decembro
Temperatura	–2	5	8	9	11	13	27	33	21	14	9	4

a) Calcula a temperatura media, a moda e a mediana.

b) Utiliza o ordenador para comprobar o resultado.

8. Calcula a media, a mediana e a moda das distribucións seguintes:

a) 2, 3, 4, 5, 7, 9, 9, 1 000 b) 2, 3, 4, 5, 7, 9, 9, 10 c) 0, 0, 4, 5, 7, 9, 9, 100, 200

Utiliza o ordenador para comprobar os resultados.

Observa en cada caso como inflúen os valores extremos. Inflúen na moda? E na mediana? E na media?

Actividade resolta

- *Nunha clase de 40 alumnos as cualificacións foron:*

x_i	0	1	2	3	4	5	6	7	8	9	10	**Suma**
f_i	1	2	0	1	2	8	7	6	6	4	3	**40**

A cada nota chamámola x_i e á frecuencia absoluta desa nota: f_i. Isto significa que houbo un cero, dos uns, ningún 2… e 3 deces.

Para calcular a media aritmética engadimos á táboa unha fila cos produtos $\boldsymbol{x_i \cdot f_i}$ e sumamos esa fila:

$x_i \cdot f_i$	0	2	0	3	8	40	42	42	48	36	30	**251**

Ao ser 40 o número total de estudantes a media é: **Media** = *m* = 251 / 40 = 6.275.

A **moda** é a nota máis frecuente, que é *mo* = 5 pois é a de maior frecuencia.

Para calcular a **mediana** engadimos unha nova fila, a das frecuencias acumuladas:

x_i	0	1	2	3	4	5	6	7	8	9	10
Frecuencias acumuladas	1	3	3	4	6	14	21	27	33	37	40

A metade dos datos é 40/2 = 20, e como 14 < 20 < 21, a mediana é 6.

Se a variable toma os valores $x_1, x_2, …, x_n$, cunha frecuencia absoluta $f_1, f_2, …, f_n$, para calcular a **media** multiplícase cada valor pola súa frecuencia absoluta, súmanse estes produtos e divídese por n o total de valores da variable:

$$\boldsymbol{m} = \textbf{Media} = (x_1 \cdot f_1 + x_2 \cdot f_2 + … + x_n \cdot f_n) / (f_1 + f_2 + … + f_n)$$

A **moda** é a frecuencia máis alta.

Pode ocorrer que unha distribución de frecuencias teña máis dunha moda. Por exemplo, a distribución:

x_i	1	2	3	4	5	6
f_i	10	9	10	8	7	10

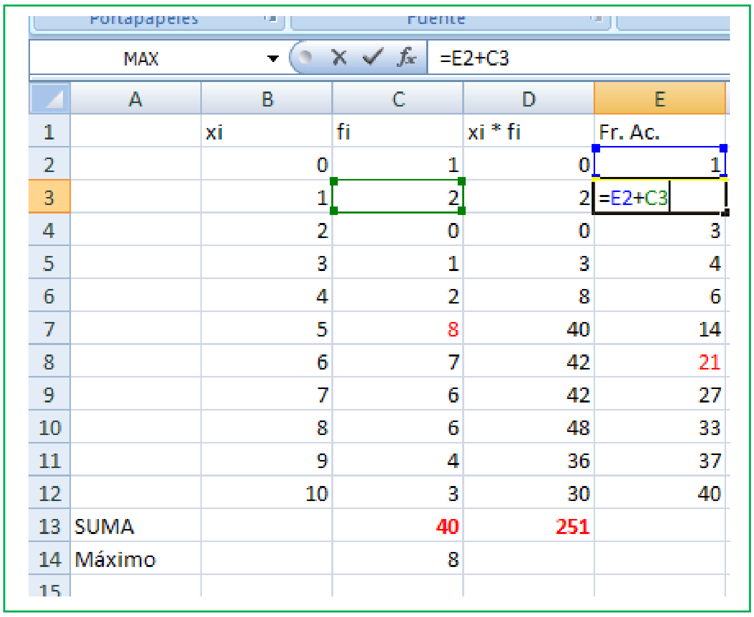
ten 3 modas, 1, 3 e 6, xa que o valor máis alto da frecuencia absoluta é 10 nos tres casos.

A moda permite clasificar os conxuntos de datos en *unimodais*, *bimodais* ou *plurimodais*, segundo o número de modas que teñan.

Para obter a **mediana** calcúlanse as frecuencias acumuladas e búscase o valor da variable que ocupa o lugar central: $n/2$.

Utiliza o ordenador

- Copiamos os datos da actividade resolta nunha folla de cálculo, escribindo xi na casa B1, fi na C1. En B2 escribimos 0, e en B3, 1. Seleccionamos estas dúas casas e arrastramos ata a casa B12. Copiamos as frecuencias na columna C. En A13 escribimos SUMA. Calculamos a suma das frecuencias coa tecla: $\sum$ e obtense 40 na casa C13. Na columna D1 escribimos $x_i \cdot f_i$. En D2 escribimos = e facemos clic en B2, escribimos * e facemos clic en C2 (=B2*C2). Seleccionamos D2 e arrastramos ata D12. Calculamos a suma (251) e dividimos o valor da casa D12 entre o da casa C12.



	A	B	C	D	E
1		xi	fi	xi * fi	Fr. Ac.
2		0	1	0	1
3		1	2	2	=E2+C3
4		2	0	0	3
5		3	1	3	4
6		4	2	8	6
7		5	8	40	14
8		6	7	42	21
9		7	6	42	27
10		8	6	48	33
11		9	4	36	37
12		10	3	30	40
13	SUMA		40	251	
14	Máximo		8		

Podemos calcular o valor máximo das frecuencias, que neste caso se ve a ollo, pero se houbese moitos máis valores, moitas máis filas, pódese utilizar a función MAX.
Para calcular as frecuencias acumuladas utilizamos a columna E. En E2 escribimos =C2. En E3 escribimos =E2+C3. Por que? E seleccionando E3, arrastramos ata E12.

Actividades propostas

9. Lanzouse un dado 100 veces e confeccionouse a seguinte táboa de frecuencias absolutas:

x_i	1	2	3	4	5	6
f_i	18	16	14	16	16	20

a) Calcula a media, a moda e a mediana.
b) Utiliza o ordenador para comprobar os resultados.

10. Lanzamos 2 dados e sumamos os valores obtidos. Repetimos o experimento 1000 veces e obteremos a seguinte táboa de frecuencias absolutas.

x_i	2	3	4	5	6	7	8	9	10	11	12
f_i	24	65	73	81	158	204	148	79	68	59	41

a) Calcula a media, a mediana e a moda.
b) Utiliza o ordenador para comprobar os resultados.
c) Repite ti os lanzamentos, agora só dez veces, e calcula de novo a media, a mediana e a moda.

11. Utiliza o ordenador para calcular a media, a mediana e a moda da seguinte táboa de frecuencias absolutas, que indica o número de fillos que teñen 200 familias entrevistadas:

x_i	0	1	2	3	4	5	6	7	8	9	10
f_i	14	65	73	27	9	6	2	1	0	2	1

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Parámetros de dispersión

Dannos unha medida do "*dispersos*" que están os datos.

A primeira medida dánola o **percorrido** ou o valor máximo menos o valor mínimo.

As máis utilizadas son a **varianza** e a **desviación típica** (ou desviación estándar) que mide a distancia dos datos respecto da media.

Xa sabes que a mediana indica o valor da variable que ocupa o lugar central. Denomínase **primeiro cuartil (Q1)** ao valor da variable que deixa menores ou iguais ca el á cuarta parte dos datos (ou un 25 %), sendo, polo tanto, as tres cuartas partes maiores ou iguais ca el. A mediana é o segundo cuartil, que deixa por debaixo a metade dos datos ou un 50 %. O **terceiro cuartil (Q3)** é o valor da variable que deixa menores ou iguais ca el as tres cuartas partes dos datos ou un 75 % (e maiores ou iguais a cuarta parte). Chámase **rango intercuartílico ou percorrido intercuartílico** á distancia entre o terceiro e o primeiro cuartil (**Q3 – Q1**). Polo que diximos, nese intervalo están a metade dos datos.

Actividade resolta

Seguimos coa mesma actividade anterior.

- Neves tivo en Matemáticas as seguintes notas: 8, 4, 6, 10 e 10. Calcula o seu percorrido, a varianza, a desviación típica, os cuartís e o percorrido intercuartílico.

A maior cualificación foi un 10 e a menor un 4, logo o **percorrido** é 10 – 4 = 6.

Percorrido = Máximo – Mínimo.

A media xa calculamos e é 7.6. Queremos analizar como as observacións se separan da media. Se a cada valor lle restamos a media, uns saen positivos e outros negativos, e se sumamos todos, compénsanse, polo que sae 0. É posible superar esa dificultade calculando esas diferenzas en valor absoluto ou elevándoas ao cadrado. Se as elevamos o cadrado, sumamos todo e dividimos polo número total de valores da variable menos 1, obteremos a varianza.

Divídese por n – 1 para mellorar as propiedades do estatístico: Varianza.

Se despois calculamos a raíz cadrada, obtense a desviación típica. Estamos avaliando a distancia dos valores da variable á media.

	x_i	x_i – media	(x_i – media)^2
1	8	0.4	0.16
2	4	–3.6	12.96
3	6	–1.6	2.56
4	10	2.4	5.76
5	10	2.4	5.76
Media = 7.6			Suma = 27.2

Se dividimos 27.2 entre 5 (*n*), obtense 5.44 que é a varianza.

Calculamos a raíz cadrada: 2.33 que é a desviación típica.

Varianza = ((x_1 – media)^2 + (x_2 – media)^2 + ... + (x_n – media)^2)/n = $\dfrac{\sum_{i=1}^{n}(x_i-m)^2}{n}$

S = Desviación típica = $\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-m)^2}{n}}$

$$\sum_{i=1}^{n}(x_i-m)^2 =$$
$$\sum_{i=1}^{n}(x_i^2-2x_i\cdot m+m^2)=$$
$$\sum_{i=1}^{n}x_i^2-2m\sum_{i=1}^{n}x_i+n\cdot m^2=$$
$$\sum_{i=1}^{n}x_i^2-2m(n\cdot m)+n\cdot m^2=$$
$$\sum_{i=1}^{n}x_i^2-n\cdot m^2$$

Pódese demostrar, facendo operacións, unha fórmula máis cómoda para calcular a varianza e a desviación típica:

Varianza = $\dfrac{\sum_{i=1}^{n}x_i^2}{n}-m^2$ **S** = $\sqrt{\dfrac{\sum_{i=1}^{n}x_i^2}{n}-m^2}$

	x_i	x_i^2
	8	64
	4	16
	6	36
	10	100
	10	100
m = 7.6	Suma = 38	Suma = 316

Varianza = (316/5) – (7.6)^2 = 63.2 – 57.76 = 5.44.

A desviación típica é a raíz cadrada da varianza, é dicir, s = 2.33.

Para calcular os cuartís debemos ordenar os datos; 4 ≤ 6 ≤ 8 ≤ 10 ≤ 10.

1	2	3	4	5
4	**6**	8	**10**	10

O primeiro cuartil deixa por debaixo a cuarta parte ou o 25 % dos datos. Hai 5 datos e 5/4 = 1.25, como 1 < 1.25 < 2, o primeiro cuartil é 6. Q1 = 6. O terceiro cuartil deixa por debaixo as tres cuartas partes ou o 75 % dos datos: 3(5/4) = 3.75. Como 3 < 3.75 < 4, entón Q3 = 10.

Percorrido intercuartílico = Q3 – Q1.

No exemplo o percorrido intercuartílico = Q3 – Q1 = 10 – 6 = 4.

Utiliza o ordenador

Igual que calculamos a media, a mediana e a moda, a folla de cálculo pódese utilizar para obter:

- O percorrido calculando MAX – MIN.
- A varianza utilizando VARP.
- A desviación típica usando DESVESTP.
- Os cuartís (CUARTIL), sendo o cuartil 0 o mínimo; o cuartil 1, Q1; o cuartil 2, a mediana; o cuartil 3, Q3; e o cuartil 4, o máximo.

Actividades propostas

12. Dadas as temperaturas nunha cidade dun exercicio anterior:

Meses	Xaneiro	Febreiro	Marzo	Abril	Maio	Xuño	Xullo	Agosto	Setembro	Outubro	Novembro	Decembro
Temperatura	–2	5	8	9	11	13	27	33	21	14	9	4

a. Calcula o percorrido, a varianza, a desviación típica, os cuartís e o percorrido intercuartílico.

b. Utiliza o ordenador para comprobar os resultados.

13. Calcula o percorrido, a varianza, a desviación típica, os cuartís e o percorrido intercuartílico das distribucións seguintes:

a) 2, 3, 4, 5, 7, 9, 9, 1 000 b) 2, 3, 4, 5, 7, 9, 9, 10 c) 0, 0, 4, 5, 7, 9, 9, 100, 200

Utiliza o ordenador para comprobar os resultados.

1.4. Diagrama de caixa

O **diagrama de caixa ou** de **bigotes** é unha representación gráfica na que se utilizan cinco medidas estatísticas: o valor mínimo, o valor máximo, a mediana, o primeiro cuartil e o terceiro cuartil… intentando visualizar todo o conxunto de datos.

O máis atraente do gráfico é a «caixa». Fórmase un rectángulo (ou caixa) cuxos lados son os cuartís (Q_1 e Q_3) e onde se sinala, no centro, a mediana (*Me*). De maneira que o cadrado/rectángulo contén o 50 por cento dos valores centrais.

Engádense dous brazos (ou *bigotes*) onde se sinalan o valor máximo (*Máx*) e o valor mínimo (*Mín*).

Pódense calcular, ademais, uns límites superior e inferior. O inferior, L_i, é Q_1– 1.5 polo percorrido intercuartílico, e o superior, Ls, é Q_3 + 1.5 polo percorrido intercuartílico.

Exemplo

- Neves tivo en Matemáticas as seguintes notas: 8, 4, 6, 10 e 10. Calcula o seu percorrido, a varianza, a desviación típica, os cuartís e o percorrido intercuartílico.

Ordenamos os datos: $4 \leq 6 \leq 8 \leq 10 \leq 10$, e calculamos que:

Mediana = *Me* = 8.

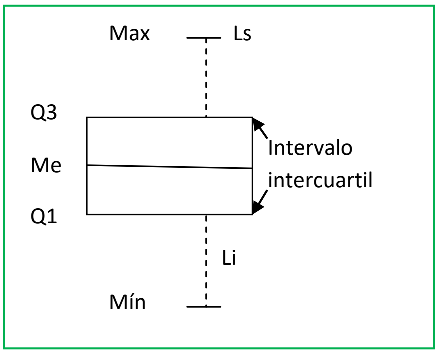
$Q_1 = 6$. $Q_3 = 10$.

Percorrido intercuartílico = 10 – 6 = 4.

Os bigotes indican:

Máx = 10. Mín = 4.

Ls = Q3 + 4*1.5 = 16. Li = Q1 – 4*1.5 = 0.



Neste exemplo o máximo é igual a 10, que é menor que o posible extremo superior, igual a 16. O mínimo é 4, maior que o extremo inferior, logo non hai valores *atípicos* que sexan maiores que o límite superior ou menores que o límite inferior. Os extremos dos bigotes no noso exemplo son 10 e 4. O diagrama de caixa é o da figura da marxe.

Actividades propostas

14. Nunha excursión de montaña participan 25 persoas coas seguintes idades:

9	9	10	11	12	20	36	37	38	40	42	43	43	44
	45	47	48	50	52	53	55	58	61	63	65		

a) Facer unha táboa de frecuencias clasificando as idades en 6 intervalos que comezan en 7.5 e terminan en 67.5. Calcular, a partir dela, os parámetros $\bar{x}$ e σ.

b) Calcular $\bar{x}$ e σ introducindo os 25 datos na calculadora, é dicir, sen agrupalos en intervalos.

c) Prescindindo dos 5 primeiros nenos, obteremos un colectivo de 20 persoas. Calcular de novo os seus parámetros $\bar{x}$ e σ, e comparar cos obtidos no grupo inicial.

d) Calcular os parámetros de posición Q_1, Q_3 e Me, da distribución orixinal, e construír o diagrama de caixa ou de bigotes correspondente.

1.5. Variable continua: intervalos e marcas de clase. Histogramas

Recorda que as variables poden ser cualitativas, se non son numéricas, ou cuantitativas, que á súa vez poden ser discretas ou continuas.

Por exemplo: Se se fai un estudo estatístico sobre a poboación de estudantes, pódese preguntar sobre a profesión dos seus pais e nais, que é unha variable cualitativa, sobre o número de irmáns, que é unha variable cuantitativa discreta (ninguén ten 3,7 irmáns), ou sobre a idade, a estatura, a cualificación media… que son variables cuantitativas continuas.

Coas variables cuantitativas continuas ten sentido agrupar os valores en intervalos.

Ao valor central do intervalo chámaselle **marca de clase**.

A representación gráfica máis adecuada é o **histograma** que é un diagrama de rectángulos no que a área de cada rectángulo é proporcional á frecuencia. Ten a vantaxe de que desa forma a frecuencia de cada suceso vén representada pola área.

Actividade resolta

- Realiza un estudo estatístico sabendo que a táboa de frecuencias absolutas, con intervalos, dos pesos de 40 estudantes dun centro escolar, é:

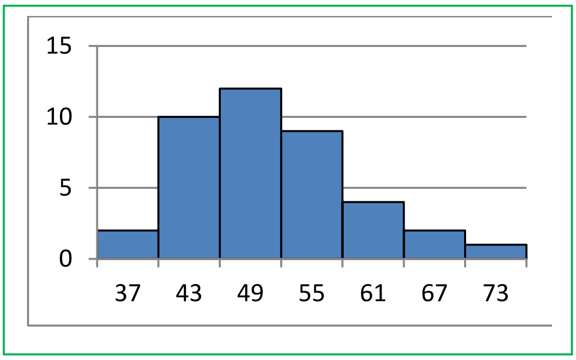
Peso	[34, 40)	[40, 46)	[46, 52)	[52, 58)	[58, 64)	[64, 70)	[70, 76)
Estudantes	2	10	12	9	4	2	1

A táboa dinos que hai 2 estudantes cuxo peso é maior ou igual a 34 e é menor que 40.

Calculamos as marcas de clase buscando o punto medio de cada intervalo: (40 – 34)/2 = 3 e 34+3 =37. Todos os intervalos neste exemplo teñen unha lonxitude de 6. Reescribimos a táboa coas marcas de clase e as frecuencias absolutas:

x_i	37	43	49	55	61	67	73
f_i	2	10	12	9	4	2	1

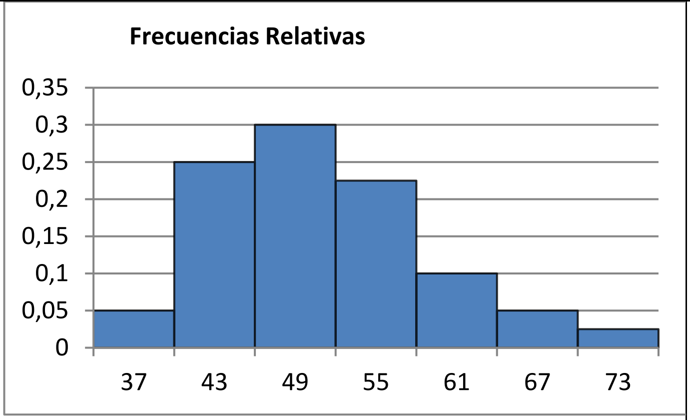
Neste caso o histograma das frecuencias absolutas é moi sinxelo pois todos os intervalos teñen igual lonxitude. Se non fose así, habería que calcular con coidado as alturas dos rectángulos para que as áreas fosen proporcionais ás frecuencias.

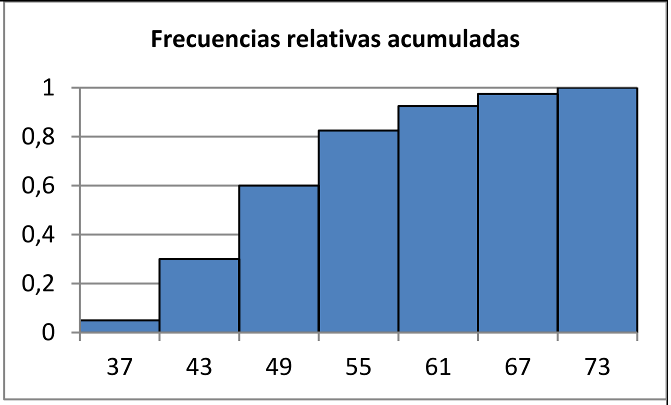


Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Imos representar tamén o histograma das frecuencias relativas e das frecuencias relativas acumuladas:

x_i	37	43	49	55	61	67	73
Frecuencias relativas	0.05	0.25	0.3	0.225	0.1	0.05	0.025
Frecuencias relativas acumuladas	0.05	0.3	0.6	0.825	0.925	0.975	1





Cálculo da media e a desviación típica

Procedemos da forma que xa coñecemos, calculando o produto das marcas de clase polas frecuencias:

x_i	37	43	49	55	61	67	73	Suma
f_i	2	10	12	9	4	2	1	40
$x_i \cdot f_i$	74	430	588	495	244	134	73	2 038

A **media** é igual a 2 038/40 = 50.95

Para calcular a **desviación típica** restamos a cada marca de clase, a media, elevamos ao cadrado e multiplicamos pola frecuencia relativa:

x_i	37	43	49	55	61	67	73	Suma
f_i	2	10	12	9	4	2	1	40
$x_i - m$	−13.95	−7.95	−1.95	4.05	10.05	16.05	22.05	
$(x_i - m)^2$	194.60	63.2025	3.8025	16.4025	101.0025	257.6025	486.2025	1122.8175
f_i	2	10	12	9	4	2	1	40
$(x_i - m)^2 \cdot f_i$	389.20	632.025	45.63	147.62	404.01	515.205	486.2025	2 619.9

A suma das diferenzas da media ao cadrado polas frecuencias relativas é 2 619.9. Agora dividimos entre *n* que no noso caso é 40, e obtense 65.5 que é a varianza. Calculamos a raíz cadrada. A desviación típica é 8.09.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Actividade resolta

- Utilizamos a outra fórmula: *Varianza* = $\frac{\sum_{i=1}^{n}\left(x_i^2 \cdot f_i\right)}{n} - m^2$

x_i	37	43	49	55	61	67	73	**Suma**
f_i	2	10	12	9	4	2	1	40
x_i^2	1 369	1 849	2 401	3 025	3 721	4 489	5 329	22 183
$x_i^2 \cdot f_i$	2 738	18 490	28 812	27 225	14 884	8 978	5 329	106 456

Varianza = (106 456/40) – (50.95)^2 = 2 661.4 – 2 595.9 = 65.5 e *desviación típica* = *s* = 8.09.

- *Vexamos outro exemplo de cálculo da media e da desviación típica utilizando a outra fórmula:*

$$Varianza = \frac{\sum_{i=1}^{n}\left(x_i^2 \cdot f_i\right)}{n} - m^2$$

x_i	64	65	66	67	68	69	70	71	72	73	74	75	**Suma**
f_i	1	0	2	5	9	22	16	12	8	3	1	1	80
$x_i \cdot f_i$	64	0	132	335	612	1 518	1 120	852	576	219	74	75	5 577
x_i^2	4 096	4 225	4 356	4 489	4 624	4 761	4 900	5 041	5 184	5 329	5 476	5 625	
$x_i^2 \cdot f_i$	4 096	0	8 712	22 445	41 616	104 742	78 400	60 492	41 472	15 987	5 476	5 625	389 063

n = 80.

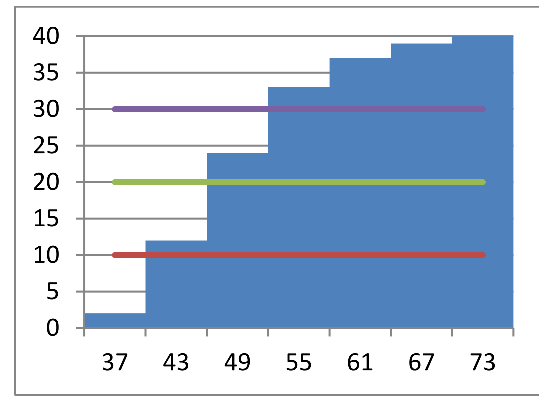
A **media** é igual a *m* = 5 577/80 = 69.7.

A **varianza** é igual a $\frac{\sum_{i=1}^{n}\left(x_i^2 \cdot f_i\right)}{n} - m^2 = \frac{389\,063}{80} - 69.7^2$ = 4 863.2875 – 4 858.09 = 5.1975

A *desviación típica* é igual á raíz cadrada da varianza, *s* = 2.28.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Cálculo da mediana e dos cuartís



- Representamos o histograma de frecuencias absolutas acumuladas e cortamos polas liñas $n/2$ para a mediana, $n/4$ para o primeiro cuartil, e $3n/4$ para o segundo. No noso caso por 20, 10 e 30.

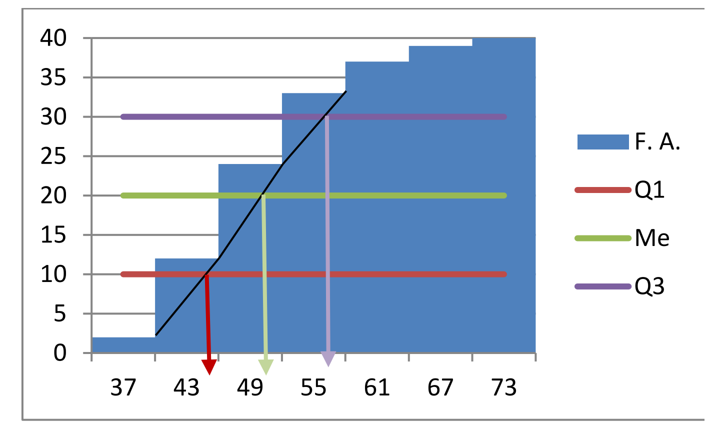
Observamos, vendo onde as rectas horizontais $y = 20$, $y = 10$ e $y = 30$ cortan ao histograma, que a mediana está no intervalo [46, 52) cuxa marca de clase é 49, o primeiro cuartil no intervalo [40, 46) cuxa marca de clase é 43, e o terceiro cuartil en [52, 58) cuxa marca de clase é 55.

x_i	[34, 40)	[40, 46)	[46, 52)	[52, 58)	[58, 64)	[64, 70)	[70, 76)
f_i	2	10	12	9	4	2	1
F_i	2	12	24	33	37	39	40

Podemos axustalo máis facendo unha interpolación lineal, é dicir, aproximando cunha recta.

Para a mediana trazamos a recta que pasa polos puntos (46, 12) e (52, 24) ($y = 2x - 80$) e calculamos onde corta á recta $y = 20$. Corta en $x = 50$. Polo tanto a mediana é $Me = 50$.

O terceiro cuartil está no intervalo [52, 58). Calculamos a ecuación da recta que pasa polos puntos (52, 24) e (58, 33), que é $y = (3/2)x - 54$. Calculamos onde corta o y = 30, que é en $x = 56$. Polo tanto Q3 = 56.



O primeiro cuartil está no intervalo [40, 46). A recta que pasa polos puntos:

(40, 2) e (46, 12)

ten por ecuación $y = (5/3)x - 64.6666$, que corta o y = 10 en $x = 44.79999...$. Q1 = 44.8.

Utiliza o ordenador

- Para debuxar histogramas co ordenador utilizando unha folla de cálculo encontrámonos coa dificultade de que este debuxa os rectángulos separados. Debuxa un diagrama de rectángulos. Para amañalo no caso de que a lonxitude de todos os intervalos sexa a mesma, debes sinalar un dos rectángulos, entrar en "*dar formato á serie de datos*" e, en "*Opcións de serie*", seleccionar en "*Ancho do intervalo*" un ancho do 0 %, é dicir, "*sen intervalo*". Se as lonxitudes son distintas débese calcular previamente as alturas dos rectángulos.

Actividades propostas

15. Utiliza o ordenador para debuxar o histograma da actividade 11.

16. Coñécense as cantidades de residuos sólidos recollidos en m^3/semana durante 12 semanas dunha urbanización: 23, 27, 30, 34, 38, 21, 30, 33, 36, 39, 32, 24. Escribe no teu caderno unha táboa de frecuencias absolutas con catro intervalos: [20, 25), [25, 30), [30, 35) e [35, 40). Calcula as marcas de clase. Debuxa o histograma de frecuencias absolutas. Calcula a media e a desviación típica. Calcula graficamente a mediana e os cuartís.

17. Fai un estudo estatístico preguntando aos teus compañeiros e compañeiras de clase sobre o número de libros que len ao mes. Confecciona unha táboa e represéntaa nun diagrama de rectángulos, un polígono de frecuencias e un diagrama de sectores.

2. DATOS BIDIMENSIONAIS

2.1. Ideas xerais

Posiblemente, a aplicación máis importante da Estatística non sexa o estudo dunha variable illada senón a análise das relacións entre variables. Se temos dúas medidas que se dan xuntas, é lóxico querer saber en que medida unha inflúe na outra. Vexamos algúns exemplos.

Exemplos:

- Nunha tenda de camisas, queremos saber cantas venderemos (por termo medio) en función do prezo.
- Se sabemos a altura do pai dun neno, cal será a altura do fillo?
- Se a un grupo de alumnas lle damos unha paga e medimos as súas cualificacións. As alumnas que reciben máis diñeiro sacan mellores notas? Canto máis? Este mesmo estudo pode facerse cos traballadores dunha empresa. Se se lles paga máis aumenta a produción?
- Son máis intelixentes os homes que as mulleres? Ou viceversa?

Pode parecerche que algún destes casos é elemental. É obvio que os pais altos teñen fillos altos e que se baixo o prezo, vendo máis. Pero o importante é canto. Se eu teño unha tenda, o que quero é gañar diñeiro. E por suposto que se poño as camisas a 0 € vou vender moito… pero non gañarei nada. O que quero é unha estimación de canto vendo a cada prezo para poder saber o prezo que me interesa poñer.

2.2. Variables bidimensionais. Frecuencias conxuntas

Unha **variable bidimensional** son dúas variables que se miden conxuntamente. Se X e Y son as variables, a variable bidimensional é (X, Y).

Exemplos:

- O prezo ao que poñemos as camisas (X) e o prezo anterior (Y).
- A altura dun pai (X) e a altura do fillo (Y)
- A cor do pelo (X) e a cor dos ollos (Y).
- O sexo dunha persoa (X) e o seu coeficiente de intelixencia (Y).

Date conta que as variables bidimensionais poden ser cualitativas ou cuantitativas e mesmo cada unha dun tipo. Así mesmo poderíamos ter os datos agrupados, e entón o que habería sería parellas de intervalos.

A representación de forma de táboa de frecuencias é exactamente igual que no caso unidimensional coa excepción de que agora temos parellas. Imos primeiro cun exemplo e logo introduciremos os conceptos.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Exemplo:

- Temos unha mostra de 8 persoas e miramos a súa cor de ollos e de pelo. Hai 4 morenos de ollos marróns, 1 moreno de ollos verdes, dos louros de ollos azuis e un louro de ollos verdes.

Aínda non definimos as frecuencias pero cremos que o podes entender igual. A táboa é:

Individuo	Frecuencias absolutas	Frecuencias relativas
(Moreno, marróns)	4	0.5 = 4/8
(Moreno, verdes)	1	0.125 = 1/8
(Louro, azuis)	2	0.25 = 2/8
(Louro, verdes)	1	0.125 = 1/10
TOTAL	8	1

Como podes ver, para que dous elementos sexan iguais, deben ser iguais as dúas compoñentes. A variable *X* é a cor do pelo e a variable Y a cor dos ollos. Tense *X* = {“Moreno”,”Louro”} e Y={“Marróns”, “Verdes”, “Azuis”}. Non ten por que haber o mesmo número de valores en cada variable.

As definicións son as mesmas.

A **frecuencia absoluta** é o número de veces que se obtivo esa parella de resultados (dúas parellas son iguais se os seus dous compoñentes son iguais).

A **frecuencia relativa** é a frecuencia absoluta dividida entre o número total de datos.

Táboa de frecuencias conxunta

Ás veces, en vez de amosar os datos en pares, póñense nunha **táboa de dobre entrada ou táboa de continxencia**. Chámase así porque o *X* está en vertical e o Y en horizontal. Nos cruces póñense as frecuencias, xa sexan absolutas ou relativas. Se se poñen as absolutas dise **táboa de dobre entrada de frecuencias absolutas e** se se poñen as relativas pois **táboa de dobre entrada de frecuencias relativas**.

A táboa anterior, con (x_i, y_i), non ten un nome especial universalmente aceptado. Podemos chamala **táboa de frecuencias de pares.**

Exemplo:

- Temos a mesma mostra de antes: 4 morenos de ollos marróns, 1 moreno de ollos verdes, dos louros de ollos azuis e un louro de ollos verdes. Imos colocalos en táboas de dobre entrada de frecuencias absolutas e logo relativas.

Limitámonos a poñer na primeira columna os dous valores que temos do *X*, que é a cor do pelo ("Moreno" e "Louro") e na primeira fila os do *Y*, que é a cor dos ollos ("Marróns", "Verdes" e "Azuis").

Y / *X*	Marróns	Verdes	Azuis
Moreno	4	1	0
Louro	0	1	2

Observa que nesta táboa poden aparecer ceros, que representan que non hai ninguén con esa parella de características.

Se dividimos as frecuencias absolutas polo número total de datos (que neste caso é 8) obteremos a táboa de dobre entrada de frecuencias relativas.

Y / *X*	Marróns	Verdes	Azuis
Moreno	0.5 = 4/8	0.125 = 1/8	0
Louro	0	0.125 = 1/8	0.25 = 2/8

Actividades propostas

18. Coa táboa de valores do exemplo, constrúe a táboa de frecuencias absolutas e relativas da variable *X* ("Cor de pelo") e a variable Y ("Cor de ollos") por separado, como variables unidimensionais.

19. Completa a seguinte táboa e exprésaa en forma de táboa de dobre entrada, primeiro con frecuencias relativas e logo con frecuencias absolutas.

(x_i, y_i)	Frecuencia absoluta	Frecuencia relativa
(0, 1)	12	
(1, 2)	14	
(2, 3)	14	

20. Completa a seguinte táboa de frecuencias conxunta e exprésaa en frecuencias de pares (x_i, y_i), tanto con frecuencias relativas como absolutas.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

2.3. Diagrama de dispersión e recta de regresión

Un **diagrama de dispersión**, tamén chamado **nube de puntos** pola súa aparencia, é un gráfico que se obtén representando cada parella como un punto do plano cartesiano. Úsase principalmente con variables cuantitativas e datos sen agrupar (se estivesen agrupados tomariamos as marcas de clase).

É moi simple de debuxar. Basta con poñer un punto en cada parella. Ás veces, se hai valores repetidos, póñense os puntos máis gordos pero tamén é común poñelos todos igual.

Exemplo:

- Temos unha tenda e queremos estudar as vendas dunha camisa en función do prezo. Para iso, probamos cada semana cun prezo distinto e calculamos as vendas medias. Obteremos así unha táboa como a que segue

Prezo	11	11.5	12	12.5	13	13.5	14	14.5	15	15.5	16
Vendas (medias)	18.2	17.2	16.1	15.3	14.6	13.5	12.5	11.4	10.1	9.1	8.1

Se copiamos os datos a unha folla de cálculo e lle damos a debuxar un diagrama de dispersión, obteremos algo como o seguinte:

que é o típico gráfico que pode verse para facer un estudo de resultados en calquera empresa.

A recta de regresión

O problema coa nube de puntos é que simplemente describe o que pasa. Isto certamente é importante en si mesmo pero o que é realmente interesante é PREDICIR que pasará. No exemplo anterior, os nosos datos chegan a prezos de 16 €. Que pasaría se subimos o prezo a 17 €? Ou se o baixamos a 9 €? E cos prezos intermedios, como 12.25 €?

Como hai infinitos prezos, non imos poder ter en conta infinitos prezos. O interesante é ter un modelo matemático que nos diga, para un prezo dado, cal é o valor esperado das vendas. Ou, en xeral, para un valor de *X* cal é o valor esperado de *Y*.

O máis fácil é facer unha recta que se aproxime. Pódese debuxar practicamente a man, pero hai unha fórmula matemática que a calcula. Esa fórmula é complicada e está fora do alcance deste curso pero si imos ensinarche como facela con ordenador.

Antes de nada, imos mostrarche no exemplo anterior a liña de tendencia.

Observa que non pasa por todos os puntos, senón que uns quedan arriba e outros abaixo. De feito é imposible que unha recta pase por todos e, no mundo real, o axuste nunca é exacto. A recta pasa polo medio dos puntos.

Utiliza o ordenador

- O seguinte son datos da altura dun pai e da do seu fillo con 15 anos de idade. As alturas están en metros.

Pai	1.7	2	1.6	1.7	1.65	1.9	1.9	1.81
Fillo	1.75	1.9	1.7	1.8	1.6	1.88	2	1.95

O primeiro, imos facer o diagrama de dispersión. Copiamos os datos nunha folla de cálculo. Ímolos poñer en vertical para que se vexa mellor, pero podería facerse exactamente igual en horizontal.

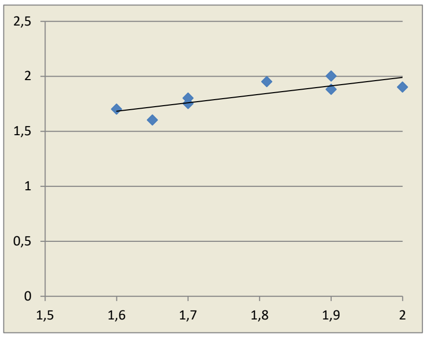
Despois, sinalamos as dúas series e dámoslle a *Inserir gráfico de dispersión*.

Automaticamente aparece o diagrama de dispersión (nube de puntos). Se xogas un pouco coas opcións podes modificar o título, o formato, a escala dos eixes…

Aínda máis, a recta de regresión é moi fácil de debuxar. Basta con que selecciones o gráfico e lle deas a *análise* e a *liña de tendencia*. Escollendo unha tendencia lineal, xa tes a recta de regresión.

Ao final, se o fixeches ben, o debuxo debe ser máis ou menos algo similar a isto:

E fíxate, a recta ten todos os valores posibles. Para ver que valor correspondería a unha altura do pai de 1.75 m, buscámolo na recta.



Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

2.4. Interpretación da recta. Introdución á correlación

Unha vez debuxamos a recta de regresión, podemos ver como é a relación entre as dúas variables. En esencia o tipo de relación vén dada pola pendente da recta.

1. Se a recta de regresión ten pendente positiva (máis informalmente, "se vai cara arriba") dise que a relación entre as variables é **positiva**.
2. Se a recta de regresión ten pendente cero (máis informalmente, "se queda horizontal") dise que a relación entre as variables é **nula ou** que non **hai relación lineal**.
3. Se a recta de regresión ten pendente negativa (máis informalmente "se vai cara abaixo") dise que a relación entre as variables é **negativa**

A cuestión é, pois, sinxela. Basta debuxar a recta e ver cara a onde vai. Pero tamén nos interesa ver se os puntos están preto da recta ou lonxe. Noutras palabras, mirar se a recta *axusta ben* ou *axusta mal*.

Para calcular isto, obtense o que se chama **coeficiente de correlación**. Defínese como:

$$\rho = r = \frac{\dfrac{\sum_{i=1}^{N}(x_i \cdot y_i)}{N} - \overline{x}\cdot\overline{y}}{s_x \cdot s_y}$$

Xa ves, moi complicado! Pero, como antes, basta con usar Excel ou calquera folla de cálculo. A orde en Excel é COEF.DE.CORREL(serie1;serie2).

O coeficiente de correlación mide se a relación é positiva, negativa ou nula. E TAMÉN nos di se o axuste é bo. Imos ver nun cadro os detalles.

O **coeficiente de correlación, ρ,** mide a relación entre as dúas variables. É un número entre –1 e 1 (pode ser exactamente –1 ou exactamente 1).

Se o coeficiente de correlación é exactamente 1 a relación é **perfecta positiva**. A recta vai cara arriba e TODOS os puntos están sobre ela.

Se o coeficiente de correlación está no intervalo (0, 1) a relación é **positiva**. A recta vai cara arriba pero non pasa por todos os puntos.

Se o coeficiente de correlación é exactamente 0, a relación é **nula** (non hai relación lineal).

Se o coeficiente de correlación está en (–1, 0) a relación é **negativa**. A recta vai cara abaixo pero non pasa por todos os puntos.

Resumo

ρ = 1 → correlación perfecta positiva

ρ = –1 → correlación perfecta negativa

ρ = 0 → correlación nula

ρ ∈ (0, 1) → correlación positiva

ρ ∈ (–1, 0) → correlación negativa

Se o coeficiente de correlación é exactamente –1 a relación é **perfecta negativa**. A recta vai cara abaixo e TODOS os puntos están sobre ela.

Isto é o que é obxectivo. Nalgunhas ocasións, fálase de correlación positiva forte (se está achegada a 1) ou positiva débil (se está entre 0 e 1 pero próxima a 0) e o mesmo negativa. Pero claro, iso depende da interpretación de cada un. Así, unha correlación de 0.96 é positiva forte e unha de –0.02 é negativa débil. Pero e 0.55? Pois depende do que consideres. O que si é obxectivo é se é perfecta ou nula, positiva ou negativa.

Utiliza o ordenador

- Cos datos da actividade anterior, imos calcular o coeficiente de correlación.

O único que hai que facer é poñer, na casa correspondente =COEF.DE.CORR. No noso Exemplo é a casa D2.

Automaticamente dános a escoller dúas matrices e escollemos primeiro a do *X* e despois a do *Y*.

Dános o coeficiente de correlación que, neste caso, resulta ser 0.8057828. É unha relación positiva forte como xa imaxinabamos pola nube de puntos e a recta de regresión.

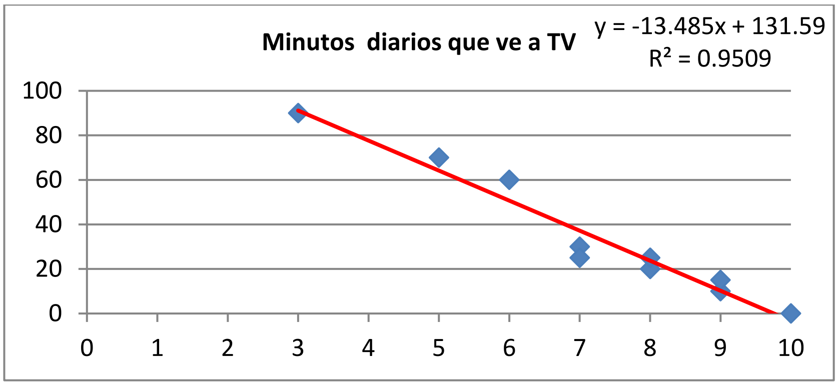
Utiliza o ordenador

- Preguntamos a 10 alumnos de 4º ESO polas súas cualificacións en Matemáticas, polo número de minutos diarios que ven a televisión, polo número de horas semanais que dedican ao estudo, e pola súa estatura en centímetros. Os datos recóllense na táboa adxunta. Queremos debuxar as nubes de puntos que os relacionan coas cualificacións de Matemáticas, o coeficiente de correlación e a recta de regresión.

Cualificacións de Matemáticas	10	3	7	8	5	9	9	8	6	7
Minutos diarios que ve a TV	0	90	30	20	70	10	15	25	60	25
Horas semanais de estudo	15	2	9	12	7	14	13	11	7	8
Estatura (en cm)	177	168	157	159	163	179	180	175	169	170

Para facelo, entramos en Excel, e copiamos os datos. Seleccionamos a primeira e a segunda fila, logo a primeira e a terceira e por último a primeira fila e a cuarta.

Coa primeira e segunda filas seleccionadas, imos a *Inserir*, *Dispersión* e eliximos a *nube de puntos*. Podemos conseguir que o eixe de abscisas vaia de 0 a 10 en "*Dar formato ao eixe*". Pinchamos sobre un punto da nube e eliximos "*Agregar Liña de tendencia*". Para que debuxe o ordenador a recta de regresión a liña de tendencia debe ser *Lineal*. Na pantalla que aparece marcamos a casa que di: "*Presentar ecuación no gráfico*" e a casa que di "*Presentar o valor de R cadrado no gráfico*".



Observa, a recta de regresión, en cor vermella, é decrecente e a súa ecuación é aproximadamente:

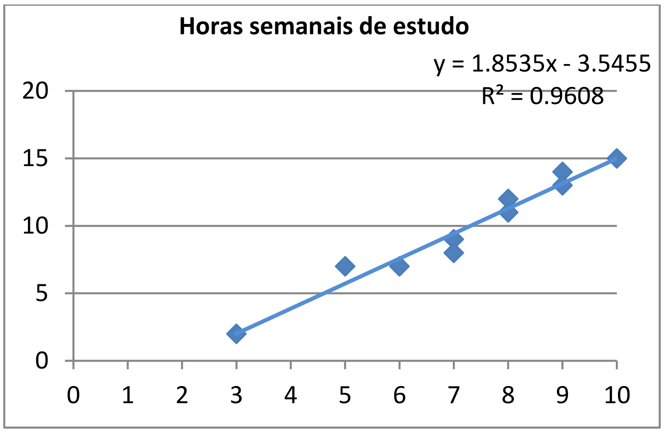
$$y = -13.5\,x + 132.$$

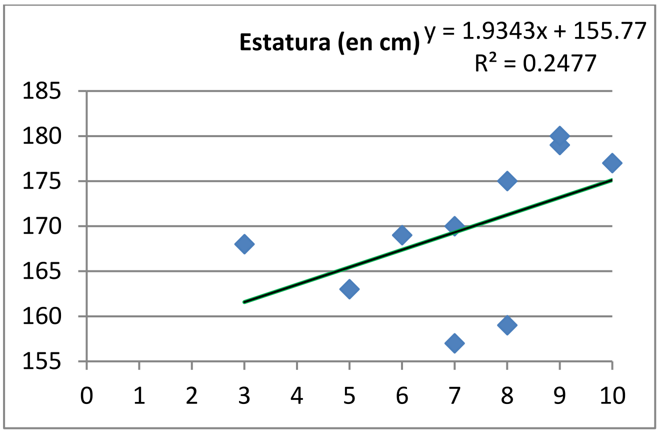
O cadrado do coeficiente de correlación é $\rho^2 = 0.95$. A correlación é negativa e alta:

$$\rho = \sqrt{0.95} = -0.975$$

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Facemos o mesmo coa primeira e terceira filas e coa primeira e cuarta filas. Obteremos os gráficos:





Observa que en ambos os casos a pendente da recta de regresión é positiva pero, no primeiro, o coeficiente de correlación, positivo, é próximo a 1, $\rho = \sqrt{0.96} = 0.98$. A correlación é alta e positiva. No segundo $\rho = \sqrt{0.25} = 0.5$

Actividades resoltas

- O propietario dunha instalación mixta solar-eólica está realizando un estudo do volume de enerxía que é capaz de producir a instalación. Para iso, mide a enerxía ao longo dun total de *N*=16 días que considera suficientemente representativos. A enerxía (en kWh) producida nestes días polas instalacións solar e eólica encóntrase recollida na seguinte táboa:

Xeración solar (x_i)	13.1	10.5	4.1	14.8	19.5	11.9	18	8.6	5.7	15.9	11.2	6.8	14.2	8.2	2.6	9.7
Xeración eólica (y_i)	8.5	14.3	24.7	4	2.3	6.4	3.6	9.2	13.5	1.4	7.6	12.8	10.3	16.5	21.4	10.9

Imos realizar unha actividade resolta completa utilizando as fórmulas da media, a desviación típica e da correlación para que poida servirche de modelo se necesitas algunha vez calculalas sen axuda do ordenador.

Imos denotar á xeración solar como variable ***X*** e a xeración eólica como variable ***Y***. Engadíoslle novas filas á nosa táboa, os cadrados de *x*, de *y* e os produtos de ambas as dúas:

Xeración solar (x_i)	13.1	10.5	4.1	14.8	19.5	11.9	18	8.6	5.7	15.9	11.2	6.8	14.2	8.2	2.6	9.7
Xeración eólica (y_i)	8.5	14.3	24.7	4	2.3	6.4	3.6	9.2	13.5	1.4	7.6	12.8	10.3	16.5	21.4	10.9
x_i^2	171.6	110.3	16.81	219.0	380.3	141.6	324	73.96	32.49	252.8	125.4	46.24	201.6	67.24	6.76	94.09
y_i^2	72.25	204.5	610.1	16	5.29	40.96	12.96	84.64	182.3	1.96	57.76	163.8	106.1	272.3	457.9	118.8
$x_i \cdot y_i$	111.4	150.2	101.3	59.2	44.85	76.16	64.8	79.12	76.95	22.26	85.12	87.04	146.2	135.3	55.64	105.7

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Cálculo das medias

Sumando a primeira fila e dividindo por N = 16, obteremos a media da xeración Solar en Kwh.

Recorda $\bar{x}=\frac{\sum_{i=1}^{N} x_i}{N}$:

$$\bar{x}=\frac{\sum_{i=1}^{N} x_i}{N}=\frac{13.1+10.5+4.1+14.8+19.5+11.9+18+8.6+5.7+15.9+11.2+6.8+14.2+8.2+2.6+9.7}{16}=10.925\text{Kwh}$$

Sumando a segunda fila e dividindo por N = 16 obteremos a media da xeración Eólica en Kwh:

$$\bar{y}=\frac{\sum_{i=1}^{N} y_i}{N}=\frac{8.5+14.3+24.7+4+2.3+6.4+3.6+9.2+13.5+1.4+7.6+12.8+10.3+16.5+21.4+10.9}{16}=10.463\text{Kwh}$$

As medias son:

$$\bar{x}=10.925\text{Kwh e } \bar{y}=10.463\text{ Kwh},$$

Moi parecidas.

Cálculo das desviacións típicas

Na terceira fila calculamos os cadrados dos valores da primeira variable e utilizámolos para calcular a varianza:

Recorda ${s_x}^2=\frac{\sum_{i=1}^{N} {x_i}^2}{N}-\bar{x}^2$:

$${s_x}^2=\frac{\sum_{i=1}^{N} {x_i}^2}{N}-\bar{x}^2=$$

$$\frac{13.1^2+10.5^2+4.1^2+14.8^2+19.5^2+11.9^2+18^2+8.6^2+5.7^2+15.9^2+11.2^2+6.8^2+14.2^2+8.2^2+2.6^2+9.7^2}{16}-10.9^2$$

$$=\frac{141.5}{16}-10.9^2=22.16$$

Na cuarta fila os cadrados dos valores da segunda variable e calculamos a súa varianza:

$${s_y}^2=\frac{\sum_{i=1}^{N} {y_i}^2}{N}-\bar{y}^2=$$

$$\frac{8.5^2+14.3^2+24.7^2+4^2+2.3^2+6.4^2+3.6^2+9.2^2+13.5^2+1.4^2+7.6^2+12.8^2+10.3^2+16.5^2+21.4^2+10.9^2}{16}-10.5^2$$

$$=\frac{150.48}{16}-10.5^2=41.01$$

A desviación típica é a raíz cadrada da varianza, polo tanto:

$$s_x=\sqrt{22.16}=4.71 \text{ e}$$

$$s_y=\sqrt{41.01}=6.4$$

Cálculo do coeficiente de correlación

Para calcular o coeficiente de correlación calculamos na quinta fila os produtos da variable *x* pola variable *y*. Así, 13.1 · 8.5 = 111.4.

Queremos calcular o termo: $\frac{\sum_{i=1}^{N}(x_i \cdot y_i)}{N}$.

Ao sumar esa fila obteremos 1 401.2, que dividimos entre 16, restámoslle o produto das medias e dividimos polo produto das desviacións típicas:

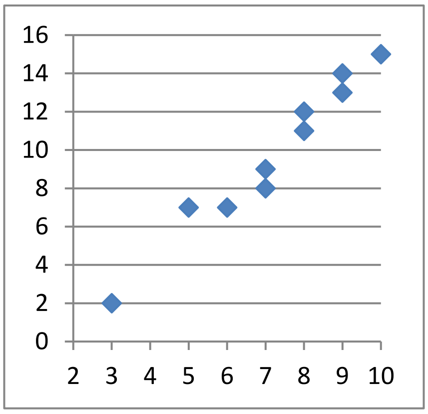
$$\rho = \frac{\frac{\sum_{i=1}^{N}(x_i \cdot y_i)}{N} - \bar{x} \cdot \bar{y}}{s_x \cdot s_y} = \frac{\frac{1\,401.2}{16} - (10.9 \cdot 10.5)}{4.71 \cdot 6.4} = \frac{-26.728}{4.71 \cdot 6.4} = -0.887$$

Este coeficiente de correlación negativo e próximo a –1 indícanos que a relación entre as dúas variables é negativa e bastante importante.

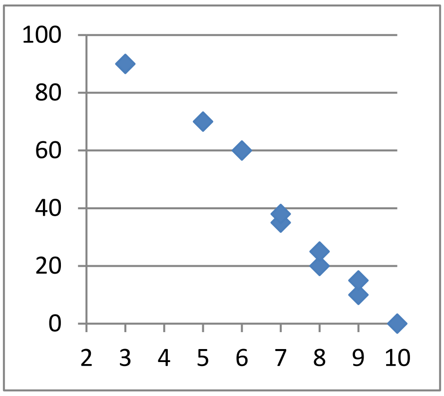
Actividades propostas

21. María calculou os coeficientes de correlación das tres nubes de puntos adxuntas e obtivo: –0.05, 0.98 e –0.99, pero agora non recorda cal é de cada unha. Podes axudala a decidir que coeficiente corresponde con cada nube?

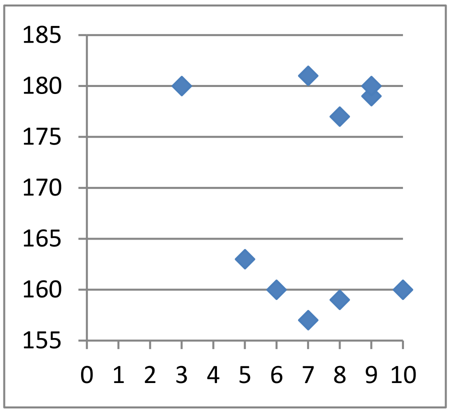
A



B



C



22. Fai unha enquisa entre os teus compañeiros de clase. Con ela vas realizar un traballo de investigación e presentar un informe. Elixe con coidado as preguntas. Vas preguntar a cada un dos teus compañeiros seleccionados, a mostra, dúas preguntas, como por exemplo o que mide a súa man e a súa nota en lingua, pero a ti poden interesarche outras cuestións moi distintas.

a. O primeiro que vas facer é tabular as respostas e confeccionar dúas táboas de frecuencias absolutas. Logo completa esas mesmas táboas coas frecuencias relativas e as frecuencias acumuladas. Fai representacións gráficas desas frecuencias: de barras, de liñas, de sectores.

b. Calcula as medias, modas e medianas así como o percorrido, a desviación típica, os cuartís, o intervalo intercuartílico… Representa os datos nunha táboa de dobre entrada e debuxa a nube de puntos. Calcula o coeficiente de correlación. Presenta un informe deste traballo.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

3. AZAR E PROBABILIDADE

3.1. Experimento aleatorio e suceso

Un **fenómeno ou experimento aleatorio** é aquel que, mantendo as mesmas condicións na experiencia, non se pode predicir o resultado.

- Son experimentos aleatorios:

a) Lanzar unha moeda e anotar se sae cara ou cruz.

b) Lanzar un dado e anotar o número da cara superior.

c) Lanzar dous dados ou dúas moedas.

d) Se nunha urna hai bólas brancas e vermellas, sacar unha ao azar e anotar a cor.

e) Sacar unha carta dunha baralla.

f) Sacar, sen substitución, dúas cartas da baralla.

g) Abrir un libro e anotar a páxina pola que se abriu.

Porén, calcular o custe dunha mercadoría, sabendo o peso e o prezo por kg, non é un experimento aleatorio. Tampouco o é calcular o custe do recibo da luz sabendo o gasto.

- Non son experimentos aleatorios

a) Saír á rúa sen paraugas cando chove e ver se te mollas.

b) O prezo de medio quilo de roscas se as roscas custan 3 € o quilo.

c) Soltar un obxecto e ver se cae.

Actividades propostas

23. Indica se son, ou non, fenómenos aleatorios:

a) A superficie das comunidades autónomas españolas.

b) Anotar o sexo do próximo bebé nacido nunha clínica determinada.

c) A área dun cadrado do que se coñece o lado.

d) Tirar tres dados e anotar a suma dos valores obtidos.

e) Saber se o próximo ano é bisesto.

Ao realizar un experimento aleatorio existen varios posibles resultados ou **sucesos posibles**.

Ao realizar un experimento aleatorio sempre se obterá un dos **posibles resultados**.

Chámase **suceso elemental** a cada un dos posibles resultados dun experimento aleatorio.

O conxunto dos posibles resultados dun experimento aleatorio denomínase **espazo dunha mostra**.

Un **suceso** é un subconxunto do conxunto de posibles resultados, é dicir, do espazo dunha mostra.

Actividade resolta

- ***Por exemplo*** os posibles resultados ao tirar unha moeda son que saia *cara* ou *cruz*. O conxunto de sucesos elementais é {cara, cruz}.
- *O conxunto de posibles resultados dos experimentos aleatorios seguintes:*

a) Extraer unha bóla dunha bolsa con 9 bólas brancas e 7 negras é {*branca*, *negra*}.

b) Sacar unha carta dunha baralla española é {AO, 2O, 3O, …CO, RO, AC, … RC, AB, … RB, AE, …RE}.

c) Tirar dúas moedas é: {(*cara*, *cara*), (*cara*, *cruz*), (*cruz*, *cara*), (*cruz*, *cruz*)}.

- *Ao lanzar un dado, o* conxunto de posibles resultados é {1, 2, 3, 4, 5, 6}, o suceso obter par é {2, 4, 6}, o suceso obter impar é {1, 3, 5}, o suceso obter múltiplo de 3 é {3, 6}, sacar un número menor que 3 é {1, 2}.
- *Ao lanzar dúas moedas o* conxunto de posibles resultados é {(C, C), (C, +), (+, C), (+, +)}. O suceso *sacar cero caras* é {(+, +)}, *sacar unha cara* é {(C, +), (+, C)} e *sacar dúas caras* {(C, C)}.

Actividades propostas

24. Escribe o conxunto de posibles resultados do experimento aleatorio: "*Escribir en cinco tarxetas cada unha das vogais e sacar unha ao azar*".

25. Escribe o conxunto de posibles resultados do experimento aleatorio: "*Tirar unha chincheta e anotar se cae de punta ou non*".

26. Inventa dous sucesos do experimento aleatorio: *Tirar dúas moedas*.

27. No xogo de lotaría, indica dous sucesos respecto á cifra das unidades do primeiro premio.

28. Escribe tres sucesos aleatorios do experimento aleatorio sacar unha carta dunha baralla española.

3.2. Frecuencia e Probabilidade

Non imos definir "Probabilidade" pois existen varias definicións posibles. Existe unha axiomática debida a *Kolmogorov* relativamente recente (1930) pero antes xa fora usado este concepto, por exemplo, por *Fermat* e *Pascal* no século XVII que escribiron cartas reflexionando sobre o que ocorría nos xogos de azar. Cando non comprendían como asignar unha determinada Probabilidade, xogaban moitas veces ao xogo que fose e vían a que valor se aproximaban as frecuencias relativas. Así, a **Probabilidade dun suceso** podería definirse como o **límite ao que tenden as frecuencias relativas** dese suceso cando o número de experimentos é moi alto. Polo tanto,

Para calcular probabilidades úsanse dúas técnicas, unha experimental, analizando as frecuencias relativas de que ocorra o suceso, e a outra por simetría, cando se sabe que os sucesos elementais son **equiprobables,** é dicir, que **todos eles teñen a mesma Probabilidade**, entón **divídese o número de casos favorables polo número de casos posibles.**

Isto último, cando se pode usar, simplifica a forma de asignar probabilidades e coñécese como **Regra de *Laplace*** que di que: *"Se os sucesos elementais son equiprobables, a Probabilidade dun suceso é o número de casos favorables dividido polo número de casos posibles"*.

Actividade resolta

- A Probabilidade de que saia cara ao tirar unha moeda é 1/2, pois só hai dous casos posibles {*cara*, *cruz*}, un único caso favorable, *cara*, e supoñemos que a moeda non está trucada. Se sospeitaramos que a moeda estivese trucada para asignar esa Probabilidade habería que tirar a moeda un montón de veces para observar cara a que valor se achega a frecuencia relativa de obter cara.
- A Probabilidade de sacar un 5 ao tirar un dado é 1/6 pois hai seis casos posibles {1, 2, 3, 4, 5, 6}, un único caso favorable, 5, e supoñemos que o dado non está trucado, logo todos eles son equiprobables.
- A Probabilidade de que ao cruzar a rúa te pille un coche non é 1/2, aínda que só hai dous casos posibles, que te pille o coche e que non te pille, pois xa te tería pillado un montón de veces. Para calcular esa Probabilidade recóllense datos de peóns atropelados e calcúlase utilizando as frecuencias relativas.
- A Probabilidade de sacar unha bóla vermella dunha bolsa con 7 bólas vermellas e 3 brancas é 7/10.
- A Probabilidade de que un bebé sexa nena é aproximadamente 0.5, pero ao facer o estudo coas frecuencias relativas viuse que é 0.49.
- Se consideramos unha baralla española de 40 cartas e eliximos unha carta, algúns dos sucesos que poden ocorrer son "*sacar un ouro*", ou "*sacar un as*", ou "*sacar o cabalo de copas*"… Como de antemán non sabemos o que vai ocorrer dicimos que estes sucesos son *aleatorios* ou de *azar*. Antes de sacar ningunha carta todas elas son igualmente factibles e como pode saír unha calquera das 40 cartas dicimos que a Probabilidade de, por exemplo, *sacar o cabalo de copas* é 1/40, a de *sacar un ouro* é 10/40, e a dun *as* é 4/40.
- Cal é a Probabilidade de sacar o rei de copas? E de sacar un rei? E unha copa?

A Probabilidade de sacar o *as de copas* é 1/40. Pero o suceso *sacar un as* cúmprese se sae o as de ouros, ou de copas, ou de bastos ou de espadas. É dicir, non é un suceso simple, está formado, neste caso por 4 sucesos elementais, logo a súa Probabilidade é 4/40 = 1/10. O mesmo lle ocorre a *sacar unha copa*. É un suceso composto e como hai 10 copas a súa Probabilidade é 10/40 = 1/4.

Actividades propostas

29. Calcula a Probabilidade de que ao sacar unha carta da baralla sexa unha espada.

30. Para saber a Probabilidade de que un recén nado sexa zurdo, basearíaste no estudo das frecuencias relativas ou asignaríala por simetría?

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

3.3. Asignación de probabilidades

Suceso contrario

Actividades resoltas

- *Cal é a Probabilidade de sacar un as na baralla de 40 cartas? E a de non sacar un as? E a de sacar unha copa? E a de non sacar unha copa?*

O suceso non *sacar un as* é o suceso **contrario** ao de *sacar un as*. Cartas que non son ases hai 36, logo a probabilidade de non sacar as é 36/40 = 9/10. Observa que se obtén que p(as) + p(non as) = 1/10 + 9/10 = 10/10 = 1.

A probabilidade de *sacar copa* é 10/40, e hai 30 cartas que non son copas, logo a probabilidade de non *sacar copa* é 30/40, e 10/40 + 30/40 = 1.

Se designamos por p(X) á probabilidade dun suceso X e por p(non X) á probabilidade do seu **suceso contrario** resulta que:

p(X) + p(non X) = 1.

A probabilidade dun suceso máis a probabilidade do seu suceso contrario é igual a 1.

Actividades propostas

31. Cal é a probabilidade de non sacar un 5 ao tirar un dado? E de non sacar un múltiplo de 3? E de non sacar un número menor que 2?

32. Ao tirar unha moeda dúas veces, cal é a probabilidade de non sacar ningunha cara? E de sacar polo menos unha cara? Observa que sacar polo menos unha cara é o suceso contrario de non sacar ningunha cara.

Sucesos dependentes e independentes

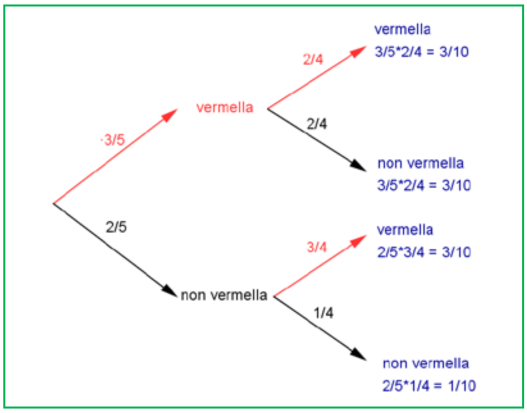
Exemplo:

- Temos unha bolsa con 3 bólas vermellas e 2 bólas negras. Cal é a probabilidade de *sacar unha bóla vermella*? Se sacamos dúas bólas, cal é a probabilidade de *sacar dúas bólas vermellas*?

A probabilidade de sacar unha bóla vermella é 3/5. Pero a de sacar dúas bólas vermellas, depende!

Depende de se volvemos meter na bolsa a primeira bóla vermella, ou se a deixamos fóra.

No primeiro caso dicimos que é **con substitución** e no segundo, **sen substitución**.



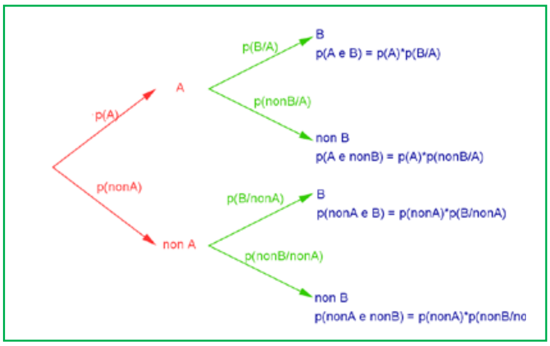
Se a volvemos meter, a probabilidade de sacar bóla vermella volverá ser 3/5 e a probabilidade de sacar dúas bólas vermellas é 3/5 · 3/5 = 9/25. A probabilidade desta segunda bóla non *depende* do que xa teñamos sacado e, neste caso, a probabilidade obtense multiplicando.

Se os sucesos A e B son **independentes: p(A e B) = p(A) · p(B)**.

Pero se a deixamos fóra, agora na bolsa só hai 4 bólas e delas só quedan 2 bólas vermellas, logo a probabilidade de que esa

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

segunda bóla sexa vermella é 2/4, e está condicionada polo que antes teñamos sacado. Escríbese: p(Vermella/Vermella) e lese *"probabilidade de vermella condicionado a ter sacado vermella"*. A probabilidade de sacar dúas bólas vermellas é agora: 3/5 · 2/4 = 6/20 = 3/10.



Observa o diagrama de árbore e comproba que a probabilidade de sacar primeiro unha bóla vermella e logo unha bóla negra (non vermella) é 3/5 · 2/4 = 3/10 pois despois de sacar unha bóla vermella na bolsa quedan só 4 bólas e delas 2 son negras. A probabilidade de sacar primeiro unha bóla negra e logo bóla vermella é 2/5 · 3/4 = 6/20 = 3/10, e a de sacar dúas bólas negras é: 2/5 · 1/4 = 2/20 = 1/10. Pero observa máis cousas.

Por exemplo, 3/5 + 2/5 = 1; 2/4 + 2/4 = 1; 3/4 + 1/4 = 1; 3/10 + 3/10 + 3/10 + 1/10 = 1.

Os sucesos non son independentes. O que ocorra A, ou non ocorra A, afecta á probabilidade de B. Por iso se di que B **está condicionado** a A.

Se os sucesos A e B son **dependentes** entón: **p(A e B) = p(A) · p(B/A)**

Actividades resoltas

- *Sacamos dúas cartas dunha baralla de 40 cartas sen substitución. Cal é a probabilidade de sacar dous ases?*

Se fose con substitución a probabilidade sería 4/40 · 4/40, pero ao ser sen substitución a probabilidade do segundo *as* vén condicionada por que teñamos sacado un as previamente. Agora na baralla xa non quedan 40 cartas senón 39, e non quedan 4 ases senón só 3, logo a probabilidade é:

4/40 · 3/39 = 1/130.

Observa que:

Se dous sucesos son **dependentes** entón: **p(B/A) ≠ p(B)**.

Pero se dous sucesos son **independentes** entón: **p(B/A) = p(B/non A) = p(B)**.

Actividades propostas

33. No teu caderno fai un diagrama en árbore similar ao anterior cos sucesos A e B: A = *sacar un as* na primeira extracción (non A = non sacalo), e B = *sacar un as* na segunda extracción (non B = non sacalo). Cal é a probabilidade de *sacar as* na segunda extracción condicionado a non telo sacado na primeira? E a de non sacar as na segunda extracción condicionado a non telo sacado na primeira? Cal é a probabilidade de *sacar dous ases*? E a de sacar un só as?

34. No diagrama de árbore anterior indica cal é a probabilidade de "*non saen 2 ases*" e a de "*non b sae ningún as*".

35. No experimento "sacar tres cartas seguidas", cal é a probabilidade de *sacar tres ases*? Primeiro con substitución, e logo sen substitución.

36. Ao tirar dúas veces un dado calcula a probabilidade de que saia un seis dobre.

37. Ao tirar dúas veces un dado calcula a probabilidade de sacar polo menos un 6. *Axuda*: Quizais che sexa máis doado calcular a probabilidade de non *sacar ningún* 6, e utilizar o suceso contrario.

Sucesos compatibles e incompatibles

Exemplo:

- Cal é a probabilidade de, nunha baralla de 40 cartas, sacar unha copa ou un ouro?

Hai 10 copas e 10 ouros e ningunha carta é á vez copa e ouro, logo a probabilidade é 20/40.

- Cal é a probabilidade de, nunha baralla de 40 cartas, sacar un as ou un ouro?

Hai 4 ases e hai 10 ouros pero hai o *as de ouros*, logo as cartas que son ou ben un as ou ben un ouro son 13, logo a probabilidade é 13/40.

Chamamos sucesos incompatibles aos que, como copa e ouro, non poden realizarse á vez, e sucesos compatibles aos que, como as e ouro, poden realizarse á vez.

Designamos p(A ou B) á probabilidade do suceso "*verifícase A ou ben verifícase B*". Vimos no exemplo que se os sucesos son incompatibles a súa probabilidade é igual á suma das probabilidades.

p(A ou B) = p(A) + p(B), se A e B son incompatibles.

Pero se A e B se poden verificar á vez haberá que restar eses casos, esas veces nas que se verifican A e B á vez.

p(A ou B) = p(A) + p(B) – p(A e B), se A e B son compatibles.

Esta segunda expresión é máis xeral que a primeira, xa que no caso no que A e B son incompatibles entón p(A e B) = 0.

Resumo:

Suceso **contrario**: **p(X) + p(non X) = 1.**

Sucesos **dependentes: p(A e B) = p(A) · p(B/A).**

Sucesos **compatibles: p(A ou B) = p(A) + p(B) – p(A e B).**

Actividades resoltas

- *Calcula a probabilidade dos sucesos seguintes: a) Sacar un rei ou unha figura; b) non sae un rei ou sae un rei; c) Sacar un basto ou unha figura.*

a) Hai 4 reis e hai 4 · 4 = 16 figuras (as, sota, cabalo e rei), pero os catro reis son figuras, polo tanto p(Rei ou Figura) = 4/40 + 16/40 – 4/40 = 16/40 = 0,4.

b) Hai 40 – 4 = 36 cartas que non son reis e hai 4 reis, logo p(non rei ou rei) = 36/40 + 4/40 = 1. Esta conclusión é máis xeral. Sempre:

p(non A ou A) = 1,

pois un suceso e o seu contrario xa vimos que verificaban que p(A) + p(non A) = 1.

c) Hai 10 bastos e hai 12 figuras pero hai 4 figuras que son á vez bastos (as, sota, cabalo e rei), logo p(Basto ou Figura) = 10/40 + 16/40 – 4/40 = 22/40 = 11/20.

Actividades propostas

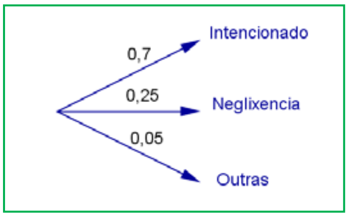
38. Lanzamos dous dados que non estean trucados e anotamos os números da súa cara superior. Consideramos o suceso A que a suma das dúas caras sexa 8, e o suceso B que eses números difiran en dúas unidades. a) Comproba que p(A) = 5/36 (2 + 6; 3 + 5; 4 + 4; 5 + 3; 6 + 2) e que p(B) = 8/36 ((1,3), (2, 4), …). b) Calcula as probabilidades de: p(A e B); p(A ou B); p(A e non B); p(non A e B); p(non A e non B). c) Calcula p(A/B); p(A/non B); p(non A/B).

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

3.4. Experiencias compostas: táboas de continxencia e diagramas de árbore

Diagramas de árbore

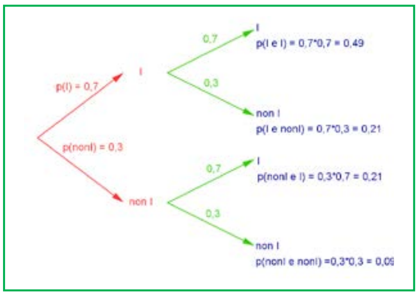
Exemplo:



- Faise un estudo sobre os incendios e compróbase que nunha determinada zona o 70 % dos incendios son intencionados, un 25 % débense a neglixencias e o 5 % a causas naturais como raios ou outras causas. Representa esta situación cun diagrama de árbore.

Actividades resoltas

- Se consideramos que a probabilidade de que un incendio sexa intencionado é 0,7, cal é a probabilidade de que ao considerar dous incendios polo menos un teña sido intencionado?



Chamamos I ao suceso "ser intencionado" e nonI ao suceso "non ser intencionado". Representamos a situación nun diagrama de árbore. Como que un incendio sexa intencionado é independente de cómo sexa o segundo, temos que:

$$p(I \text{ e } I) = 0.7 \cdot 0.7 = 0.49$$

$$p(I \text{ e nonI}) = 0.7 \cdot 0.3 = 0.21$$

xa que é a probabilidade de que o primeiro incendio sexa intencionado e o segundo non.

$$p(\text{nonI e } I) = 0.3 \cdot 0.7 = 0.21$$

$$p(\text{nonI e nonI}) = 0.3 \cdot 0.3 = 0.09$$

A probabilidade de que polo menos un teña sido intencionado podemos calculala sumando as probabilidades de (I e I), (I e nonI), e (nonI e I) que é 0.49 + 0.21 + 0.21 = 0.91. Pero máis sinxelo é calcular a probabilidade do suceso contrario p(nonI e nonI) = 0.09 e restala de 1:

$$p(\text{polo menos un intencionado}) = 1 – 0.09 = 0.91.$$

Actividades propostas

39. Debuxa no teu caderno un diagrama en árbore para tres incendios e calcula a probabilidade de que polo menos un teña sido intencionado sendo p(I) = 0.7.

40. Nunha aeronave instaláronse tres dispositivos de seguridade: A, B e C. Se falla A ponse B en funcionamento e, se tamén falla B, empeza a funcionar C. As probabilidades de que funcione correctamente cada dispositivo son: p(A) = 0.95; p(B) = 0.97 e p(C) = 0.98. a) Calcula a probabilidade de que fallen os tres dispositivos. b) Calcula a probabilidade de que todo vaia ben.

41. Unha fábrica de bonecas desbota normalmente o 0.5 % da súa produción por fallos debidos ao azar. Calcula a probabilidade de que: a) ao coller dúas bonecas ao azar haxa que desbotar ambas as dúas. b) ao coller dúas bonecas ao azar haxa que desbotar só unha. c) ao coller dúas bonecas ao azar non haxa que desbotar ningunha d) Verificamos 4 bonecas, calcula a probabilidade de desbotar unicamente a terceira boneca elixida.

42. Lanzamos unha moeda ata que apareza dúas veces seguidas do mesmo lado. Calcula as probabilidades de que: A) A experiencia termine ao segundo lanzamento. B) Termine ao terceiro lanzamento. C) Termine no cuarto. D) Termine como moito no cuarto lanzamento (é dicir, que

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

termine no segundo ou no terceiro ou no cuarto lanzamentos).

Táboas de continxencia

Exemplo:

- Estudáronse 500 enfermos do fígado analizando, por un procedemento novo, se as lesións son benignas ou malignas. Logo volveron ser analizados polo procedemento usual determinando que diagnósticos foran correctos e cales incorrectos. Os valores obtidos represéntanse na táboa:

	Diagnóstico correcto	**Diagnóstico incorrecto**	**Totais**
Lesión maligna	206	12	218
Lesión benigna	268	14	282
Totais	474	26	500

Determinamos a táboa de frecuencias relativas:

	Diagnóstico correcto (C)	**Diagnóstico incorrecto (I)**	**Totais**
Lesión maligna (M)	0.412	0.024	0.436
Lesión benigna (B)	0.536	0.028	0.564
Totais	0.948	0.052	1

Actividades resoltas

- *Imaxina que estas frecuencias relativas puideran tomarse como probabilidades. Interpretamos entón o significado de cada un destes valores.*

0.412 sería a probabilidade de que o diagnóstico de lesión maligna fose correcto: p(M e C).

0.024 = p(M e I); 0.536 = p(B e C); 0.028 = p(B e I).

E 0.436? o número de lesións malignas é 218, logo 0.436 = p(M).

Do mesmo modo: 0.564 = p(B); 0.948 = p(C); 0.052 = p(I).

Observa que p(M) + p(B) = 1 e que p(C) + p(I) = 1. Son sucesos contrarios.

- *Son dependentes ou independentes os sucesos M e C?*

Recorda que p(M e C) = p(M) · p(C/M), polo tanto: 0.412 = 0.436 · p(C/M), de onde p(C/M) = 0.412/0.436 = 0.945 que é distinto de 0.948 que é a probabilidade de C. Pódese afirmar que M e C son dependentes xa que p(C/M) ≠ p(C).

En xeral denomínase **táboa de continxencias** a:

	A	**No A**	
B	P(A e B)	P(non A e B)	P(B)
Non B	P(A e non B)	P(non A e non B)	P(non B)

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

	P(A)	P(non A)	1

Nunha táboa de continxencia figuran todas as probabilidades ou continxencias dos sucesos compostos.

Observa que, como sabemos pola probabilidade do suceso contrario:

p(A) + p(non A) = 1 e p(B) + p(non B) = 1.

Observa tamén que:

p(A) = p(A e B) + p(A e non B), do mesmo modo que p(B) = p(A e B) + p(non A e B)

pois obtéñense sumando respectivamente a primeira columna e a primeira fila.

Tamén:

p(non A) = p(non A e B) + p(non A e non B) e p(non B) = p(A e non B) + p(non A e non B).

Actividades propostas

43. Fíxose un estudo estatístico sobre accidentes de tráfico e determináronse as seguintes probabilidades reflectidas na táboa de continxencia:

	Accidente en estrada (C)	**Accidente en zona urbana (U)**	**Totais**
Accidente con vítimas (V)	0.27		0.56
Accidente con só danos materiais (M)			
Totais	0.58		1

a) Copia a táboa no teu caderno e complétaa.

b) Determina as seguintes probabilidades: p(V e C); p(V e U); p(M e C); p(M e U); p(V); p(M); p(C) e p(U).

c) Calcula p(U/V); p(C/V); p(V/U); p(V/C). Son dependentes ou independentes os sucesos: accidente con vítimas e accidente en estrada?

44. Inventa unha táboa de continxencia considerando que os accidentes poidan ser de estrada (C) ou urbanos (U) pero que agora clasificamos en leves (L), graves (G) ou mortais (M). *Observa que* o fundamental para confeccionar a táboa é que os sucesos sexan incompatibles dous a dous.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Diagramas de árbore e táboas de continxencia

Os diagramas de árbore e as táboas de continxencia están relacionados. Dada unha árbore podes obter a táboa de continxencia e viceversa. Ten interese esta relación pois cos datos do problema ás veces é máis sinxelo construír un deles e dar a solución pasando ao outro.

Actividades resoltas

- Dada a táboa de continxencia, obter o diagrama de árbore que comeza con A e non A.

	A	**Non A**	
B	2/9	5/9	7/9
Non B	1/9	1/9	2/9
	3/9 = 1/3	6/9 = 2/3	1

Coñecemos a p(A) = 3/9 = 1/3, p(non A) = 6/9 = 2/3, p(B) = 7/9 e p(non B) = 2/9.

Tamén coñecemos p(A e B) = 2/9; p(A e non B) =1/9; p(non A e B) = 5/9 e p(non A e non B) = 1/9.

Fáltanos coñecer p(B/A) que podemos obter dividindo p(A e B) entre p(A):

p(B/A) = p(A e B)/p(A) = 2/9 : 3/9 = 2/3.

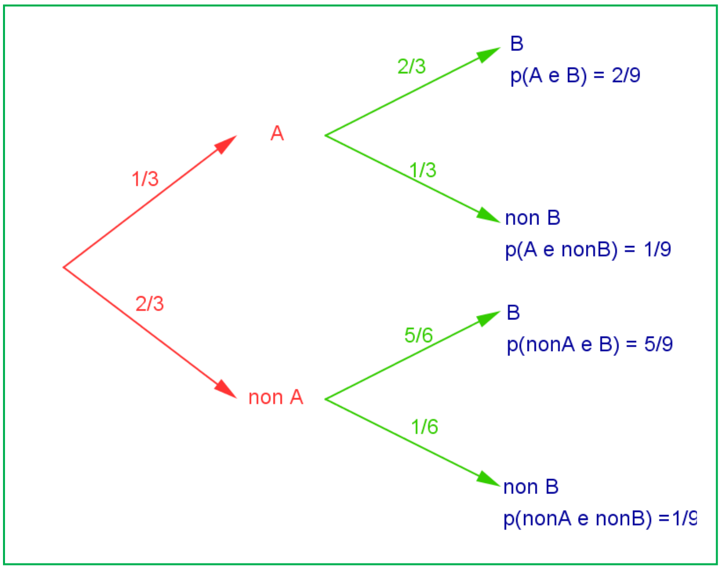
Do mesmo modo calculamos:

p(non B/A) = p(A e non B)/p(A) = 1/9 : 3/9 = 1/3

p(B/non A) = p(non A e B)/p(non A) = 5/9 : 6/9 = 5/6

p(non B/non A) = p(non A e non B)/p(non A) = 1/9 : 6/9 = 1/6.

A árbore é:



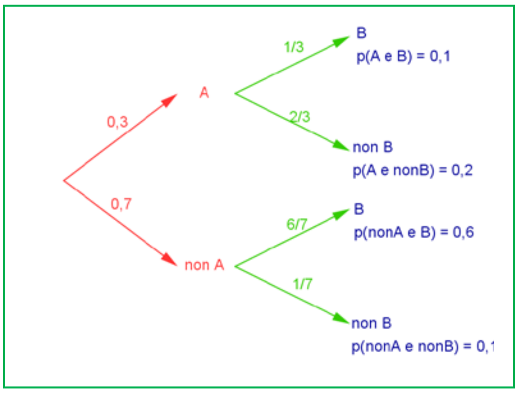
Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Actividades resoltas

- Reciprocamente, dado o diagrama de árbore obter o diagrama de continxencia:

Agora coñecemos p(A) = 0,3 e p(non A) = 0,7. Ademais coñecemos p(B/A) = 1/3; p(B/non A) = 6/7; p(non B/A) = 2/3 e p(non B/non A) = 1/7.

Calculamos, multiplicando: p(A e B) = 0.3·(1/3) = 0.1; p(A e non B) = 0.3·(2/3) = 0.2; p(non A e B) = 0.7·(6/7) = 0.6 e p(non A e non B) = 0.7·(1/7) = 0.1 que poñemos tamén na árbore.



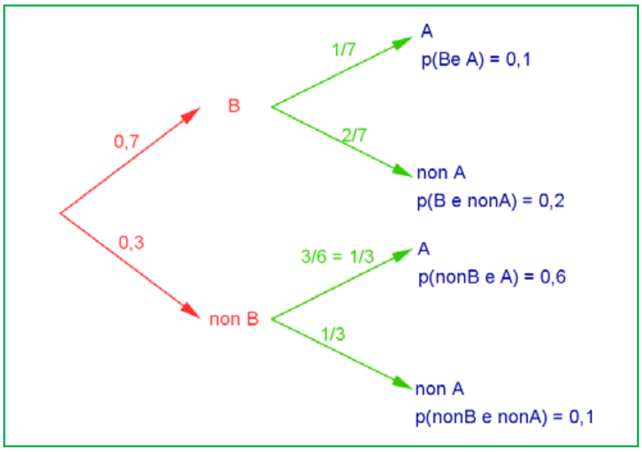
Enchemos con estes datos, unha táboa de continxencia:

	A	Non A	
B	0.1	0.6	
Non B	0.2	0.1	
	0.3	0.7	1

Calculamos, sumando, as casas que nos faltan, p(B) = 0.1 + 0.6 = 0.7 e p(non B) = 0.2 + 0.1 = 0.3.

	A	Non A	
B	0.1	0.6	0.7
Non B	0.2	0.1	0.3
	0.3	0.7	1

Pode ser moi interesante pasar dun diagrama de árbore á táboa de continxencia e desta, ao outro diagrama de árbore, co que podemos coñecer p(A/B) = 0.1/0.7 = 1/7; p(non A/B) = 0.2/0.7 = 2/7; p(A/non B) = 0.3/0.6 = 3/6 = 1/2 e p(non A/non B) = 0.1/0.3 = 1/3.



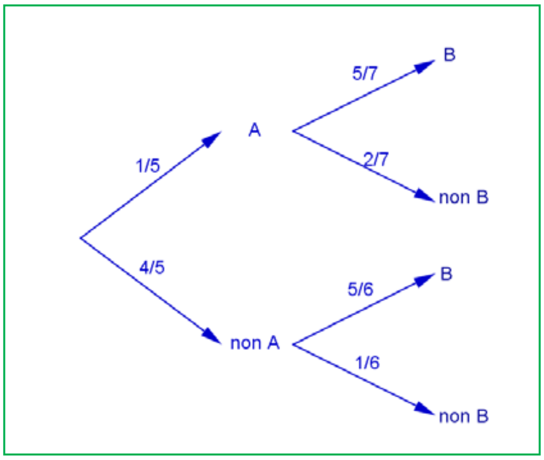
Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Actividades propostas

45. Dada a táboa de continxencia, constrúe dous diagramas de árbore.

	A	Non A	
B	0.4	0.2	0.6
Non B	0.15	0.25	0.4
	0.55	0.45	1

46. Dado o diagrama de árbore, constrúe a táboa de continxencia, e despois o outro diagrama de árbore.



47. Temos dúas urnas, A e B. A primeira con 8 bólas brancas e 2 bólas negras. A segunda con 4 bólas brancas e 6 bólas negras. Sácase unha bóla ao azar dunha das dúas urnas, elixida tamén ao azar, e resulta ser negra. Cal é a probabilidade de que proceda da urna A?

48. Estase estudando un tratamento cun novo medicamento, para o que se seleccionan 100 enfermos. 60 son tratados co medicamento e 40 cun placebo. Os valores obtidos represéntanse na táboa adxunta

	Medicamento (M)	Placebo (non M)	
Curados (C)	50	30	80
Non curados (no C)	10	10	20
	60	40	100

Utilízanse eses valores para asignar probabilidades. Calcula:

a) A probabilidade de que un enfermo curado teña sido tratado co medicamento. *Axuda*: p(M/C)

b) A probabilidade de que un enfermo curado teña sido tratado co placebo. *Axuda*: p(non M/C).

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

CURIOSIDADES E REVISTA

Estatística

O nome de Estatística provén do s. XIX, porén xa se utilizaban representacións gráficas e outras medidas en peles, rochas, paus de madeira e paredes de covas para controlar o número de persoas, animais ou certas mercadorías desde a Prehistoria. Os babilonios usaban xa envases de arxila para recompilar datos sobre a produción agrícola. Os exipcios analizaban os datos da poboación e a renda do país moito antes de construíren as pirámides. Os antigos gregos realizaban censos cuxa información se utilizaba cara ao 600 a.C.

O inicio da Teoría da Probabilidade, como sabes, foron os xogos de azar.

Cabaleiro da Meré

Ao *Cabaleiro da Meré* gustáballe xogar e era un gran xogador por iso sabía que era favorable apostar, ao tirar un dado, "sacar polo menos un 6 en 4 tiradas dun dado" e que non o era, ao tirar dous dados, o "sacar polo menos un 6 dobre en 24 xogadas".

Vese que xogara moito para saber que as frecuencias relativas lle dicían que o primeiro suceso tiña unha probabilidade superior a 0.5, e o segundo a tiña inferior. Pero non o comprendía. Non era matemático e só sabía a regra de tres. Isto non é unha proporcionalidade! Dixo 6 : 4 = 36 : 24. Pero as frecuencias relativas dicíanlle que non era así polo que escribiu a Pascal para que lle solucionara o problema.

Ti xa sabes o suficiente para solucionarllo. Antes de seguir lendo, intenta resolvelo.

En lugar de calcular a probabilidade de *sacar polo menos un 6* en 4 tiradas, calcula a probabilidade de non *sacar un 6*, que é o seu suceso contrario, e é $\left(\frac{5}{6}\right)^4$. Polo tanto a probabilidade de *sacar polo menos un 6* en 4 tiradas é:

$$1-\left(\frac{5}{6}\right)^4 = 0.5177 > 0.5.$$

Calculamos do mesmo modo a probabilidade de sacar polo *menos un seis dobre* ao tirar dous dados 24 veces, calculando a do seu suceso contrario, a de non *sacar ningún seis dobre*: $\left(\frac{35}{36}\right)^{24}$, polo que sacar polo menos un 6 dobre é:

$$1-\left(\frac{35}{36}\right)^{24} = 0.4914 < 0.5.$$

Canto debeu xogar o *Cabaleiro da Meré* para darse conta desa pequena diferenza nas probabilidades!

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

Galileo

No século XVI formulou o seguinte problema: ao tirar tres dados, por que é máis probable obter que a suma das caras superiores sexa 10, a que sexa 9?

Continuaba a reflexión coas posibles descomposicións nesas sumas:

9 = 3 + 3 + 3	10 = 4 + 3 + 3
9 = 4 + 3 + 2	10 = 4 + 4 + 2
9 = 4 + 4 + 1	10 = 5 + 3 + 2
9 = 5 + 2 + 2	10 = 5 + 4 + 1
9 = 5 + 3 + 1	10 = 6 + 2 + 2
9 = 6 + 2 + 2	10 = 6 + 3 + 1

En ambos os casos hai 6 descomposicións posibles, porén, tirando moitas veces os 3 dados, comprobaba que é máis probable sacar un 10.

Se fas un diagrama en árbore comprobarás que todas esas descomposicións non son igualmente probables.

Por exemplo: (3, 3, 3) ten unha probabilidade de 1/216, mentres que a suma 6 + 2 + 2 pode saír con tres sucesos (6, 2, 2), (2, 6, 2) e (2, 2, 6), logo a súa probabilidade é 3/216.

- Calcula as probabilidades de cada unha das sumas e a de sacar 10 e de sacar 9.

Se queres saber máis, busca:

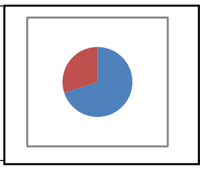
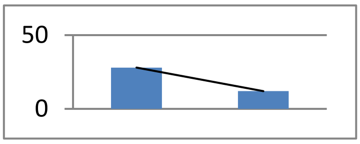
http://www.miscelaneamatematica.org/Misc34/caballero.pdfhttp://www.miscelaneamatematica.org/Misc34/caballero.pdf

A ruleta

William Jaggers chegou a Montecarlo cuns poucos francos no peto e, durante un mes, anotou os números que saían en cada ruleta e en catro días gañou dous millóns catrocentos mil francos. *Jaggers* conseguiu quebrar a banca en *Montecarlo* analizando as frecuencias relativas de cada número da ruleta e observando que se desgastara algo do mecanismo dunha delas co que todos os valores non tiñan igual probabilidade. Apostou aos números máis probables e gañou.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

RESUMO

Noción	*Definición*	*Exemplos*
Poboación e mostra	**Poboación:** Todo o conxunto de individuos sobre o que se fai o estudo. **Mostra:** Unha parte desa poboación.	Para coñecer a intención de voto, a poboación é todo o país, e selecciónase unha mostra.
Frecuencia absoluta, relativa e acumulada	**Frecuencia absoluta:** Número de veces que se obtivo ese resultado. **Frecuencia relativa:** Obtense dividindo a frecuencia absoluta polo número total. **Frecuencia acumulada:** Obtense sumando as frecuencias anteriores.	| Fr. Absoluta | Fr. Relativa | Fr. Acumulada Absoluta A | 28 | 0.7 | 28 B | 12 | 0.3 | 40
Gráficos estatísticos	Diagrama de barras Diagrama de liñas Diagrama de sectores 	
Media	Media = m = ($x_1 + x_2 + \ldots + x_n$)/n	Con: 8, 4, 6, 10 e 10. Media = 38/5 = 7.6
Moda	É o valor máis frecuente.	10
Mediana	Deixa por debaixo a metade.	4 < 6 < 8 < 10 = 10. Me = 8.
Varianza e Desviación típica	Varianza = $\frac{\sum_{i=1}^{n}(x_i-m)^2}{n} = \frac{\sum_{i=1}^{n}x_i^2}{n} - m^2$. s = $\sqrt{\frac{\sum_{i=1}^{n}(x_i-m)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n} - m^2}$	Varianza = 5.4. s = 2.33.
Cuartís	Q1 deixa por debaixo a cuarta parte. Q3 deixa por debaixo as tres cuartas partes. Percorrido intercuartílico = Q3 – Q1.	Q1 = 6; Q3 = 10; Percorrido intercuartílico = Q3 – Q1 = 4.
Histograma	A área de cada rectángulo é proporcional á frecuencia.	
Correlación	O **coeficiente de correlación**, ρ, mide a relación entre dúas variables. É un número entre –1 e 1.	ρ = 1 → correlación perfecta positiva. ρ = –1 → correlación perfecta negativa. ρ = 0 → correlación nula. $\rho \in (0, 1)$ → correlación positiva. $\rho \in (-1, 0)$ → correlación negativa.
Suceso	Ao realizar un experimento aleatorio existen varios posibles resultados ou **sucesos posibles**. Un **suceso** é un subconxunto do conxunto de posibles resultados.	Tiramos un dado. Posibles resultados = {1, 2, 3, 4, 5, 6} Suceso *obter múltiplo de* 3 = {3, 6}
Probabilidade	Límite ao que tenden as frecuencias relativas. Se os sucesos elementais son equiprobables entón: p = casos favorables / casos posibles.	P(5) = 1/6. P(sacar múltiplo de 3) = 2/6
Asignación de probabilidades	Suceso contrario: p(X) + p(non X) = 1. Sucesos dependentes: p(A e B) = p(A) · p(B/A). Sucesos compatibles: P(A ou B) = p(A) + p(B) – p(A e B).	P(no 5) = 1 – 1/6 = 5/6. P(5 ou múl. 3) = 1/6 + 2/6 = 3/6 P sacar primeiro un 5 e logo múltiplo de 3 = 1/6 · 2/6 = 2/36

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

EXERCICIOS E PROBLEMAS

Estatística

1. Nunha clase mírase a cor dos ollos de cada alumno e alumna e obtense o seguinte:

N := negro; A := azul e V := verde.

N, N, A, V, N, V, A, N, A, N, V, A, A, N, N, N, V, A, N, N, A, N, V, N, N, A, N, A, N, N.

Fai unha táboa de frecuencias absolutas, representa os valores nun diagrama de sectores e calcula a moda.

2. As notas dun conxunto de alumnos de 4º son:

2, 10, 7, 8, 1, 0, 3, 5, 6, 9, 2, 4, 1, 6, 9, 10, 5, 6, 7, 8, 3, 1, 0, 1, 5, 9, 10, 9, 8, 7.

a) Fai unha táboa de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas e frecuencias relativas acumuladas.
b) Calcula a media, a mediana e a moda.
c) Calcula a desviación típica e os cuartís.

3. Preguntouse a 40 alumnos polo número de irmáns que tiñan e obtívose

Número de irmáns	0	1	2	3	4	5	6 ou máis
Número de veces	5	15	7	6	4	2	1

a) Representa un diagrama de barras de frecuencias absolutas e un diagrama de liñas de frecuencias relativas.
b) Calcula a media, a mediana e a moda.

4. Lanzáronse catro moedas 100 veces e anotouse o número de veces que saíu cara. Os resultados están reflectidos na táboa seguinte:

Número de caras	0	1	2	3	4
Número de veces	7	25	36	26	6

a) Escribe no teu caderno unha táboa de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas e frecuencias relativas acumuladas. b) Representa un diagrama de barras de frecuencias absolutas acumuladas, un diagrama de liñas de frecuencias relativas e un diagrama de sectores de frecuencias absolutas. c) Calcula a media e a desviación típica. d) Calcula a mediana e os cuartís.

5. Para coñecer a distribución nun certo país das persoas segundo a súa idade recolleuse unha mostra de dez mil persoas e os valores obtidos veñen reflectidos na táboa seguinte:

Idades	[0, 5)	[5, 10)	[10, 15)	[15, 25)	[25, 35)	[35, 45)	[45, 55)	[55, 65)	[65,100)
Número de persoas	900	1 000	900	1 500	1 300	1 200	1 300	900	1 000

a) Utiliza as marcas de clase e escribe no teu caderno unha táboa de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas e frecuencias relativas acumuladas. b) Representa un histograma de frecuencias absolutas. *Coidado*: os intervalos non son todos iguais. *Recorda: a* área dos rectángulos debe ser proporcional ás frecuencias. c) Calcula a media e a desviación típica. d) Calcula a mediana e os cuartís de forma gráfica usando un histograma de frecuencias absolutas acumuladas.

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

6. Cos datos do problema anterior calcula o intervalo [media – desviación típica, media + desviación típica]. Cantas persoas están nese intervalo? Que porcentaxe? Calcula tamén o intervalo:

[media – 2*desviación típica, media + 2*desviación típica] e

[media – 3*desviación típica, media + 3*desviación típica].

Se a distribución fose normal habería no primeiro intervalo un 68 % da mostra, no segundo un 95 % e no terceiro máis dun 99.7 %. Compara os teus resultados con estes.

7. Cos mesmos datos calcula o percorrido intercuartílico, e indica cantas persoas están nese intervalo e que porcentaxe.

8. Unha compañía de seguros desexa establecer unha póliza de accidentes. Para iso, selecciona ao azar a 200 propietarios e pregúntalles cantos euros gastaron en reparacións do automóbil. Agrupáronse en intervalos os valores da variable obtidos:

Euros	[0, 100)	[100, 200)	[200, 400)	[400, 600)	[600, 800)	[800, 3000)
Número de persoas	40	30	20	40	50	20

a) Calcula as marcas de clase e escribe no teu caderno unha táboa de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas e frecuencias relativas acumuladas.
b) Representa un histograma de frecuencias relativas. *Coidado*: os intervalos non son todos iguais.
c) Calcula a media e a desviación típica.
d) Calcula a mediana e os cuartís de forma gráfica usando un histograma de frecuencias absolutas acumuladas.

9. Dous fabricantes de baterías de coches ofrecen o seu produto a unha fábrica ao mesmo prezo. A fábrica quere elixir a mellor. Para iso escolle unha mostra de 60 baterías de cada marca e obtén de cada unha os meses que funcionou sen avariarse. Obtén a seguinte táboa:

Vida da batería en meses	20	22	24	26	28	30	32
Marca A	2	7	13	16	12	8	2
Marca B	1	4	17	20	15	3	0

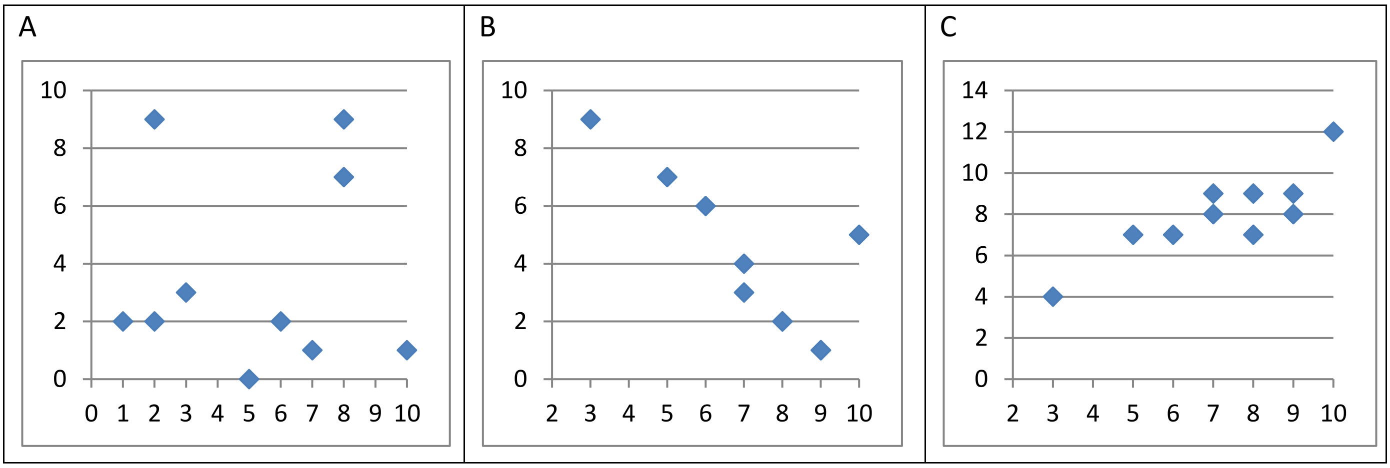
Que marca cres que elixirá?
Para tomar a decisión, calcula a media, a moda e a mediana para cada marca.
Se aínda non te decides, calcula o percorrido, a desviación típica, o intervalo [m – s, m + s] e o percorrido intercuartílico.

10. Fai un traballo. Pasa unha enquisa aos teus compañeiros e compañeiras da clase. Failles unha pregunta con datos numéricos, como por exemplo, canto mide a súa man, que número de zapato calzan, o número de libros que le nun mes, o número de horas que ve a televisión á semana, diñeiro que gasta ao mes en comprar música… Representa os datos obtidos nunha táboa e fai un estudo completo. Podes utilizar o ordenador:

a) Escribe no teu caderno unha táboa de frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas e frecuencias relativas acumuladas.
b) Debuxa un diagrama de barras, un diagrama de liñas e un diagrama de sectores.
c) Calcula a media, a mediana e a moda.
d) Calcula a varianza e a desviación típica.
e) Calcula os cuartís e o percorrido intercuartílico.
f) Reflexiona sobre os resultados e escribe un informe.

Coeficiente de correlación

11. Andrés calculou os coeficientes de correlación das tres nubes de puntos adxuntas e obtivo: –0.8, 0.85 e 0.03, pero agora non recorda cal é de cada unha. Podes axudalo a decidir que coeficiente corresponde con cada nube?



Probabilidade

12. Nun colexio selecciónase un grupo de 200 estudantes dos cales todos estudan francés ou inglés. Deles 150 estudan inglés e 70 estudan francés. Cantos estudan francés e inglés? Noutro centro escolar estúdanse varios idiomas: francés, inglés, alemán, italiano. Selecciónanse tamén 200 estudantes dos cales, 150 estudan inglés, 70 francés e 40 ambos os idiomas, cantos estudantes dese centro non estudan nin francés nin inglés?

13. Lanzamos un dado. Calcula a probabilidade de: a) Sacar un número impar. b) Non sacar un 3. c) Sacar un número maior que 3. d) Sacar un número maior que 3 e que sexa impar. e) Sacar un número maior que 3 ou ben que sexa impar.

14. Nunha clase hai 24 alumnos e 14 alumnas. A metade das alumnas e a terceira parte dos alumnos teñen os ollos azuis. Elíxese un estudante ao azar. A) Calcula a probabilidade de que sexa mozo e teña os ollos azuis. B) Calcula a probabilidade de que sexa mozo ou teña os ollos azuis.

15. Antón, Xoán e Xurxo teñen unha proba de natación. Antón e Xoán teñen a mesma probabilidade de gañar, é o dobre da probabilidade de Xurxo. Calcula a probabilidade de que gañen Xoán ou Xurxo.

16. Lanzamos dúas moedas distintas, unha de 50 céntimos e outra dun euro. Calcula a probabilidade de que: A) na moeda dun euro saia cara. B) Saia unha cara. C) Saia polo menos unha cara. D) non saia ningunha cara. E) Saian unha cara e unha cruz.

17. Lanzamos tres moedas. Calcula as probabilidades de: A) non saia ningunha cara. B) Saia polo menos unha cara. C) Saian dúas caras e unha cruz.

18. Lanzamos dous dados e anotamos os valores das caras superiores. Calcula as probabilidades de que a suma sexa 1, sexa 2, sexa 3, …. sexa 12.

19. Que é máis probable ao tirar tres dados, que a suma das súas caras superiores sexa 9 ou sexa 10? Escribe o suceso "sexa 9" e o suceso "sexa 10" e calcula as probabilidades dos seus sucesos elementais. Sabes xa máis que *Galileo*!

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF

20. Lanzamos á vez unha moeda e un dado. Chama A ao suceso "Saia cara e un número par", B ao suceso "Saia cruz e un número primo" e C ao suceso "saia un número primo". Calcula as probabilidades de A, B e C. Como son estes sucesos? Indica cales deles son compatibles e cales son incompatibles.

21. Lanzamos unha moeda 50 veces, que é máis probable, obter 50 caras seguidas ou obter nas primeiras 25 tiradas cara e nas 25 seguintes cruz? Razoa a resposta.

22. Unha moeda está trucada. A probabilidade de obter cara é dobre que a de obter cruz. Calcula as probabilidades dos sucesos obter cara e obter cruz ao tirar a moeda.

23. Tres mozos e dúas mozas xogan un torneo de xadrez. Todos os mozos teñen idéntica probabilidade de gañar e todas as mozas, tamén. Pero a probabilidade de gañar unha moza é o dobre da de gañar un mozo. Calcula a probabilidade de que un mozo gañe o torneo.

24. Sete parellas de noivos están nunha habitación. Selecciónanse dúas persoas ao azar. Calcula a probabilidade de: a) Sexan un mozo e unha moza. b) Sexan unha parella de noivos. Agora elíxense 4 persoas ao azar. Calcula a probabilidade de: c) Haxa polo menos unha parella de noivos. d) Non haxa ningunha parella de noivos.

25. Temos un dado trucado de forma que os números impares teñen unha probabilidade dobre á dos números pares. Calcula as probabilidades de: A) Saia un número impar. B) Saia un número primo. C) Saia un número primo impar. D) Saia un número que sexa primo ou sexa impar.

26. Nun grupo de 12 amigas hai 3 louras. Elíxense dúas mozas ao azar. Calcula a probabilidade de que: A) Ambas sexan louras. B) Polo menos unha sexa loura. C) Ningunha sexa loura. D) Unha sexa loura e a outra non.

27. Lanzamos dous dados e anotamos os valores das caras superiores. Calcula as probabilidades de que: A) os números obtidos sexan iguais. B) os números obtidos difiran en 3 unidades. C) os números obtidos sexan pares.

28. Lanzamos unha moeda ata que saia cara. Calcula a probabilidade de que: A) Saia cara antes do cuarto lanzamento. B) Saia cara despois do oitavo lanzamento.

29. Un lote de 20 artigos ten 2 defectuosos. Sácanse 4 ao azar, cal é a probabilidade de que ningún sexa defectuoso?

30. Lánzanse dous dados e a suma das caras superiores é 7. Cal é a probabilidade de que nun dos dados saíse un 3?

31. Téñense 3 caixas, A, B e C. A caixa A ten 10 bólas das cales 4 son negras. A caixa B ten 6 bólas cunha bóla negra. A caixa C ten 8 bólas con 3 negras. Cóllese unha caixa ao azar e desa caixa sácase unha bóla, tamén ao azar. Comproba que a probabilidade de que a bóla sexa negra é 113/360.

32. Temos unha moeda trucada cuxa probabilidade de obter cara é 3/5 e a de cruz é 2/5. Se sae cara escóllese ao azar un número do 1 ao 8 e, se sae cruz, escóllese un número do 1 ao 6. Calcula a probabilidade de que o número escollido sexa impar.

33. Nun proceso de fabricación de móbiles detéctase que o 2 % saen defectuosos. Utilízase un dispositivo para detectalos que resulta que detecta o 90 % dos móbiles defectuosos pero sinala como defectuoso un 1 % que non o é. A) Calcula a probabilidade de que sexa correcto un móbil que o dispositivo cualificou como defectuoso. B) Calcula a probabilidade de que sexa defectuoso un móbil que o dispositivo cualificou como correcto. *Axuda*: Utiliza primeiro un diagrama en árbore e logo unha táboa de continxencia.

AUTOAVALIACIÓN

Cos datos seguintes, 1, 5, 2, 8, 9, 4, 7, 7, 5, 7, calcula:

1. A media:

a) 5 b) 5.5 c) 6 d) 7

2. A mediana:

a) 5 b) 5.5 c) 6 d) 7

3. A moda:

a) 5 b) 5.5 c) 6 d) 7

4. A desviación típica:

a) 2 b) 2.3 c) 2.5 d) 2.6

5. O percorrido intercuartílico

a) 3 b) 2.75 c) 4 d) 2

6. Ao tirar dous dados, a probabilidade de sacar polo menos un 5 é:

a) 5/6 b) 11/36 c) 25/36 d) 30/36

7. Ao tirar 3 moedas, a probabilidade de sacar exactamente dúas caras é:

a) 1/2 b) 3/4 c) 3/8 d) 5/8

8. Ao tirar 3 moedas, a probabilidade de sacar polo menos dúas caras é:

a) 1/2 b) 3/4 c) 3/8 d) 5/8

9. Sacamos unha carta dunha baralla de 40 cartas, a probabilidade de que sexa un ouro ou un múltiplo de 2 é:

a) 22/40 b) 19/40 c) 36/40 d) 3/4

10. Indica cal das afirmacións seguintes é **sempre** correcta:

a) p(A) + p(non A) = 1

b) p(A e B) = p(A) · p(B)

c) p(A ou B) = p(A) + p(B)

Autores: María Molero e Andrés García
Tradutora: Mª Teresa Seara Domínguez
Ilustracións: Banco de Imaxes de INTEF