

4t A ESO

Capítol 7: Estadística. Atzar i probabilitat.

Propiedad Intelectual

El presente documento se encuentra depositado en el registro de Propiedad Intelectual de Digital Media Rights con ID de obra AAA-0181-02-AAA-042255

Fecha y hora de registro: 2014-05-08 18:17:57.0

Licencia de distribución: CC by-nc-sa



Queda prohibido el uso del presente documento y sus contenidos para fines que excedan los límites establecidos por la licencia de distribución.

Más información en <http://www.dmrighs.com>



LibrosMareaVerde.tk

www.apuntesmareaverde.org.es



Autores: María Molero i Andrés García Mirantes

Revisores: Raquel Caro i Nieves Zuasti

Il·lustracions: Banc d'Imatges d'INTEF

Traducció: Pedro Podadera, IES Juan de Garay

Índex

1. ESTADÍSTICA

- 1.1. MOSTRES. ESTUDIS ESTADÍSTICS
- 1.2. VARIABLE DISCRETA. TAULES I GRÀFICS
- 1.3. PARÀMETRES DE CENTRALITZACIÓ I DISPERSIÓ
- 1.4. DIAGRAMA DE CAIXES
- 1.5. VARIABLE CONTÍNUA: INTERVALS I MARQUES DE CLASSE. HISTOGRAMES

2. DADES BIDIMENSIONALS

- 2.1. IDEES GENERALS
- 2.2. FREQÜÈNCIES CONJUNTES
- 2.3. DIAGRAMA DE DISPERSIÓ I RECTA DE REGRESSIÓ
- 2.4. INTERPRETACIÓ DE LA RECTA DE REGRESSIÓ. INTRODUCCIÓ A LA CORRELACIÓ

3. ATZAR I PROBABILITAT

- 3.1. EXPERIMENT ALEATORI I SUCCÉS
- 3.2. FREQÜÈNCIA I PROBABILITAT
- 3.3. ASSIGNACIÓ DE PROBABILITATS. PROBABILITAT A PRIORI I A POSTERIORI. LLEI DE LAPLACE
- 3.4. EXPERIÈNCIES COMPOSTES: TAULES DE CONTINGÈNCIA I DIAGRAMES ARBRE. TEOREMA DE BAYES.

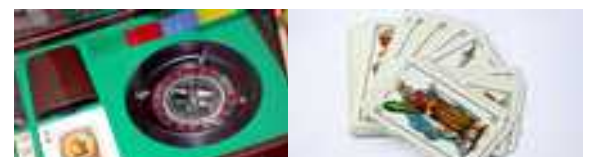
Resum

L'Estadística s'ocupa d'interpretar gran nombre de dades. L'Institut Nacional d'Estadística arreplega estudis de qualsevol tipus sobre la població espanyola. Entra en Internet escrivint INE i tindràs un muntó d'informació al teu abast sobre: a) Entorn físic i medi ambient; b) Demografia i població; c) Societat; d) Economia...

En un estudi estadístic conflueixen distintes parts de l'Estadística, la Teoria de Mostres que indica sobre la forma de seleccionar una mostra perquè siga representativa de la població, l'Estadística Descriptiva que utilitza taules, gràfics i paràmetres estadístics com la mitjana i la desviació típica per a descriure les dades, i la Inferència Estadística que utilitza la Teoria de Probabilitats per a obtenir conclusions.

Com sabràs, en temps de Jesucrist ja l'emperador August va fer censos per a conèixer la població a l'Imperi Romà.

La Teoria de la Probabilitat va tindre els seus inicis molt lligats als jocs d'atzar, i és sorprenent que amb aqueix inici haja resultat de tanta utilitat en la Ciència. Es preguntaven què és més probable en tirar dos daus, que la suma de les seues cares superiors siga 9 o siga 10. Analitzant jocs com aquest va ser avançant la Ciència.



1. ESTADÍSTICA

1.1. Mostres. Estudis estadístics

Si volem fer un estudi estadístic tenim que:

- Arreplegar les dades
- Descriure aqueixes dades amb taules i gràfiques, càlcul de paràmetres estadístics....
- Extraure conclusions.

Per a arreplegar les dades i determinar els valors de la variable es pot utilitzar tota la població, tot l'univers sobre el qual es realitza l'estudi, o fer una mostra. Moltes vegades no és convenient arreplegar valors de tota la població, perquè és complicat o massa costós, o inclús perquè és impossible com al cas d'un control de qualitat en què es destrüisca l'objecte a analitzar. La part de l'Estadística que s'ocupa de com seleccionar adequadament les mostres es denomina Teoria de Mostres.

Població o univers és tot el conjunt d'individus sobre el qual es realitza l'estudi.

Una **mostra** és un subconjunt representatiu d'aqueixa població.






Cada un dels elements de la població és un **individu**.

Les característiques de la població que s'estudien es denominen **variables estadístiques**, que es classifiquen en **quantitatives** i **qualitatives** segons que els valors que prenguen siguin o no numèrics. Les variables quantitatives que prenen valors aïllats es denominen **variables discretes** i les que poden prendre qualsevol valor d'un interval de la recta real, **variables contínues**.

La part de l'Estadística que ordena, analitza i representa un conjunt de dades per a descriure les seues característiques es denomina **Estadística Descriptiva**.

Per a extraure conclusions s'utilitzen les probabilitats i la part de l'Estadística que s'ocupa d'això és la **Inferència Estadística**.

Exemples:

-  Si volem conèixer les preferències en esports de l'alumnat de 4º, és possible preguntar a tota la població (alumnat de 4t), encara que és adequat triar una mostra representativa, seleccionant alguns estudiants.
-  En aquest estudi sobre preferències esportives, la variable utilitzada és qualitativa.
-  Per a conèixer la intenció de vot davant d'unes eleccions europees, municipals, autonòmiques... s'utilitzen mostres, perquè preguntar a tota la població seria molt costós (i això ja es fa a les eleccions). La variable en aquest cas també és qualitativa.
-  Per a estudiar el que més preocupa a una població: desocupació, terrorisme, corrupció... també s'utilitzen mostres. En aquest cas seria molt costós preguntar a tota la població, encara que seria factible. La variable en aquest cas també és qualitativa.
-  Però si una fàbrica vol conèixer les hores de vida útil d'una pereta, una nevera, un camió... no pot posar a funcionar a tota la població, (totes les peretes o neveres o camions...) fins que s'espatllen perquè es queda sense producció. En aquest cas és imprescindible seleccionar una mostra. La variable en aquest cas és quantitativa, i el temps presa qualsevol valor, és una

variable quantitativa contínua.

✚ Si preguntem pel nombre de germans és una variable quantitativa discreta.

✚ En *control de qualitat* es fan estudis estadístics i es prenen mostres.

Activitats proposades

1. Volem realitzar un estudi estadístic sobre el temps dedicat a l'estudi per l'alumnat d'ESO de Madrid. Per a això se seleccionen adequadament 100 alumnes. Indica quina és la població, quina la mostra, quina grandària té la mostra i qui seria un individu.
2. Vols passar una enquesta per a conèixer, el mateix que al problema anterior, el temps dedicat a l'estudi, en aquest cas el dels companys i companyes del teu centre escolar. Se la passaries només a les xiques? Només als xics? Preguntaries als millors de la classe? Als de pitjors notes? Indica el criteri que seguiries per a seleccionar la mostra a què preguntar.

1.2. Variable discreta. Taules i gràfics

Taules

En fer un estudi estadístic o realitzar un experiment aleatori la informació obtinguda es resumeix en una taula o distribució de freqüències.

Exemple:

✚ Preguntem a 40 estudiants de 4t si els agrada, o no, el futbol. A la taula del marge reflectim els resultats.

És una taula de freqüències absolutes.

En dividir la freqüència absoluta entre el nombre total tenim la freqüència relativa, així la freqüència relativa de què els agrada el futbol és $28/40 = 0,7$, i la de què no els agrada el futbol és $12/40 = 3/10 = 0,3$.

| Possibles resultats | Freqüència absoluta |
|---------------------|---------------------|
| Els agrada | 28 |
| No els agrada | 12 |
| Total | 40 |

| Possibles resultades | Freqüències relatives | Percentatge |
|----------------------|-----------------------|-------------|
| Les gusta | 0,7 | 70 |
| No les gusta | 0,3 | 30 |
| Suma total | 1 | 100 |

La **freqüència absoluta** és el nombre de vegades que s'ha obtingut aqueix resultat.

La **freqüència relativa** s'obté dividint la freqüència absoluta entre el nombre total de dades.

La suma de les freqüències relatives és sempre igual a 1.

Multiplicant per 100 s'obtenen els percentatges.

Activitat resolta



S'han obtingut les dades sobre el nombre de visites que s'han fet dels Textos Marea Verda de Matemàtiques als mesos indicats, i s'han reflectit en una taula. Fes una taula de freqüències absolutes, relatives i percentatges, de freqüències acumulades absolutes i de freqüències relatives acumulades.

| Marea verda | Freqüències absolutes | Freqüències relatives | Percentatges | Freqüències acumulades absolutes | Freqüències acumulades relatives |
|-------------|-----------------------|-----------------------|--------------|----------------------------------|----------------------------------|
| Setembre | 1834 | 0,51 | 51 | 1834 | 0,52 |
| Octubre | 956 | 0,26 | 26 | 2790 | 0,77 |
| Novembre | 432 | 0,12 | 12 | 3222 | 0,89 |
| Desembre | 389 | 0,11 | 11 | 3611 | 1 |
| TOTAL | 3611 | 1 | 100 | | |

| Resultats | Freqüències absolutes |
|-----------|-----------------------|
| 1 | 17 |
| 2 | 12 |
| 3 | 17 |
| 4 | 15 |
| 5 | 21 |
| 6 | 14 |

Observa que les **freqüències acumulades** s'obtenen sumant la freqüència anterior i indica, en aquest exemple, el nombre de visites fins a aqueix moment.

Activitats proposades

- Copia al teu quadern i completa la següent taula de freqüències absolutes dels valors obtinguts en tirar un dau, amb les freqüències relatives i percentatges, i amb freqüències acumulades absolutes i freqüències relatives acumulades.

Gràfics estadístics

Les representacions gràfiques ajuden a comprendre el significat de les dades.

Donada una taula de freqüències (absolutes, relatives, percentatges, acumulades absolutes o acumulades relatives) per a representar un **diagrama de rectangles o de barres** es traça per a cada valor de la variable un rectangle o barra d'altura proporcional a la freqüència que s'estiga representant.

Si s'uneixen els punts mitjans dels extrems superiors de les barres tenim un **polígon de freqüències o diagrama de línies**.

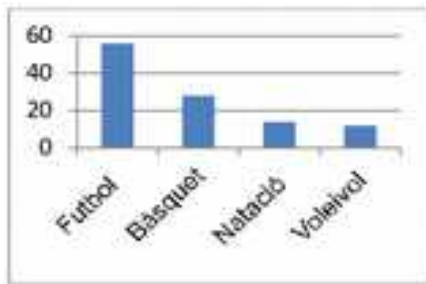
En un **diagrama de sectors** es dibuixa un cercle que es divideix en sectors d'amplituds proporcionals a les freqüències.

Activitat resolta

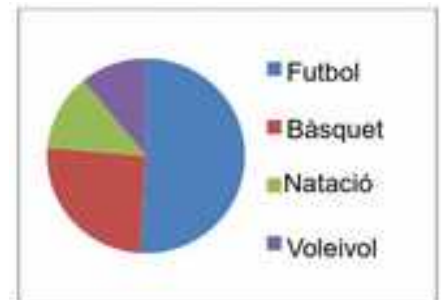
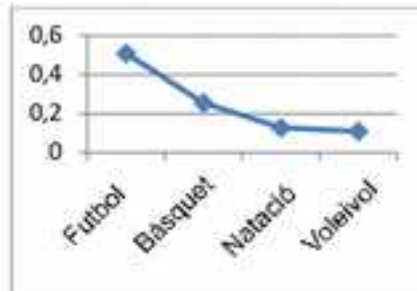
- ✚ Tenim un estudi estadístic sobre les preferències esportives de l'alumnat de 4t d'un determinat centre escolar. Representa'ls en un diagrama de barres de freqüències absolutes, en un polígon de freqüències relatives i en un diagrama de sectors.

| Esports | Freqüència Absoluta |
|----------|---------------------|
| Futbol | 56 |
| Bàsquet | 28 |
| Natació | 14 |
| Voleivol | 12 |

Diagrama de barres de freqüències absolutes



Polígon de freqüències relatives o diagrama de línies



Activitats proposades

- Amb la taula de valors de l'exercici anterior, dibuixa al teu quadern el diagrama de freqüències relatives, el polígon de freqüències absolutes acumulades i el diagrama de sectors.
- Fes un estudi estadístic preguntant als teus companys i companyes de classe sobre el nombre de llibres que lligen al mes. Confecciona una taula i representa-la en un diagrama de rectangles, un polígon de freqüències i un diagrama de sectors.
- Selecciona una mostra entre els teus companys i companyes i realitza un estudi estadístic sobre l'esport que més li agrada a cada u. Fes la representació que siga més senzilla d'interpretar.

Utilitza l'ordinador

Els fulls de càlcul són una ferramenta molt útil per a treballar l'Estadística. Sumen, multipliquen, i dibuixen els gràfics amb gran facilitat. Per a l'activitat resolta anterior, copiem la taula amb les dades al full de càlcul a partir de la casella A1. Calculem la suma total a la casella B6, simplement estrenyent la tecla: Σ , o bé escrivint =SUMA(B2:B5) que vol dir que volem sumar el que hi ha des de la casella B2 a la B5.

Per a calcular les freqüències relatives escrivim en C1: Freqüència relativa, i en C2, escrivim el signe igual, (amb el que estem dient al full que calcularem alguna cosa), punxem a la casella B2, escrivim: /, i punxem en B6: =B2/B6, ens ix 0,50909... La casella B2 anirà variant quan calculem C3, C4..., però volem que la casella B6 es quede fixa. Per a dir això, posem el símbol \$: =B2/\$B\$6. I ara arrosseguem fins a la casella C5. (Si arrosseguem abans de posar el \$ ens ix

| | A | B | C |
|---|-------------|---------------------|---------------------|
| 1 | Deportes | Frecuencia absoluta | Frecuencia relativa |
| 2 | Futbol | 56 | 0,509090909 |
| 3 | Baloncesto | 28 | 0,254545455 |
| 4 | Natación | 14 | 0,127272727 |
| 5 | Balón volea | 12 | 0,109090909 |
| 6 | SUMA | 110 | |
| 7 | | | |

un error, perquè està dividint per zero en anar modificant la casella). Tenim les freqüències relatives calculades.

Per a dibuixar els gràfics només hem de seleccionar les files i columnes que ens interessin i al menú de "Inserir" seleccionar el tipus de gràfic desitjat: Columna, Línia, Circular...

1.3. Paràmetres de centralització i dispersió

Paràmetres de centralització

Ja saps que els paràmetres de centralització ens donen informació sobre el "centre" d'un conjunt de dades. Estudiem la mitja aritmètica, la moda i la mitjana.

Activitat resolta

✚ *Neus ha tingut en Matemàtiques les notes següents: 8, 4, 6, 10 i 10. Calcula la seua mitja, la seua moda i la seua mitjana.*

La seua nota mitja es calcula sumant totes les notes: $8 + 4 + 6 + 10 + 10 = 38$, i dividint la suma entre el nombre total de notes que és 5: $38/5 = 7,6$.

La moda és 10 perquè és el valor més freqüent.

Una forma de calcular la mitjana és ordenar els valors de menor a major, i si el nombre de dades és imparell, el valor central és la mitjana. Si el nombre de dades és parell, la mitjana és la mitja de les dos dades centrals.

Al nostre cas: $4 \leq 6 \leq 8 \leq 10 \leq 10$, per la qual cosa la mitjana és 8.

Per a calcular la **mitja (m)** de x_1, x_2, \dots, x_n , se sumen tots i es divideix pel nombre total de dades (n).

$$\text{Mitja} = m = (x_1 + x_2 + \dots + x_n)/n$$

Què és el que està de moda? El que més es porta.

La **moda (mo)** d'una distribució de freqüències és el valor més freqüent.

La **mitjana (me)** és el valor central que deixa per davall el mateix nombre de valors de la variable que per damunt.

Utilitza l'ordinador

Per a calcular la mitja, la mitjana i la moda amb el full de càlcul, copiem en la casella B2, B3... les dades: 8, 4, 6, 10 i 10. Escrivim en la casella A7, Mitja, i per a calcular la mitja escrivim un signe igual en B7. Busquem, desplegant les possibles funcions, la funció MITJA, i escrivim

$$=PROMEDIO(B2:B6),$$

que significa que calcule la mitja dels valors que hi ha en les caselles des de B2 fins a B6.

| | A | B | C | D | E |
|---|---------|-------|---|---|---|
| 1 | | Datos | | | |
| 2 | | 8 | | | |
| 3 | | 4 | | | |
| 4 | | 6 | | | |
| 5 | | 10 | | | |
| 6 | | 10 | | | |
| 7 | Media | 7,6 | | | |
| 8 | Mediana | 8 | | | |
| 9 | Moda | 10 | | | |

De la mateixa manera calculem la mitjana buscant en les funcions o escrivint $=MITJANA(B2:B6)$ i la moda buscant a les funcions o escrivint $=MODA(B2:B6)$.

Activitats proposades

7. Donades les temperatura en una ciutat a una hora determinada el dia 1 de cada mes es té la taula següent:

| | | | | | | | | | | | | |
|-------------|-------|--------|------|-------|------|------|--------|-------|----------|---------|----------|----------|
| | Gener | Febrer | Març | Abril | Maig | Juny | Juliol | Agost | Setembre | Octubre | Novembre | Desembre |
| Temperatura | -2 | 5 | 8 | 9 | 11 | 13 | 27 | 33 | 21 | 14 | 9 | 4 |

a) Calcula la temperatura mitja, la moda i la mitjana.

b) Utilitza l'ordinador per a comprovar el resultat.

8. Calcula la mitja, la mitjana i la moda de les distribucions següents:

a) 2, 3, 4, 5, 7, 9, 9, 1000 b) 2, 3, 4, 5, 7, 9, 9, 10 c) 0, 0, 4, 5, 7, 9, 9, 100, 200

Utilitza l'ordinador per a comprovar els resultats.

Observa en cada cas com influeixen els valors extrems. Influeixen la moda? I a la mitja? I a la mitjana?

Activitat resolta

En una classe de 40 alumnes les qualificacions han sigut:

| | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|----|-------------|
| 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 l'anomenem x_i i a la freqüència absoluta d'aqueixa nota: f_i . Açò vol dir que hi ha hagut un zero, dos uns, cap 2... i 3 deus.

Per a calcular la mitja aritmètica afegim a la taula una fila amb els productes $x_i \cdot f_i$ i sumem aqueixa fila:

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

En ser 40 el nombre total d'estudiants la mitja és: **Mitja** = $m = 251 / 40 = 6,275$.

La **moda** és la nota més freqüent, que és $m_o = 5$ perquè és la de major freqüència.

Per a calcular la **mitjana** afegim una nova fila, la de les freqüències acumulades:

| | | | | | | | | | | | |
|-------------------------------|---|---|---|---|---|----|----|----|----|----|----|
| x_i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Freqüències acumulades | 1 | 3 | 3 | 4 | 6 | 14 | 21 | 27 | 33 | 37 | 40 |

La mitat de les dades és $40/2 = 20$, i com $14 < 20 < 21$, la mitja és 6.

Si la variable pren els valors x_1, x_2, \dots, x_n , amb una freqüència absoluta f_1, f_2, \dots, f_n , per a calcular la **mitja** es multiplica cada valor per la seua freqüència absoluta, es sumen els dits productes i es divideix per n el total de valors de la variable:

$$m = \text{Mitja} = (x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n) / (f_1 + f_2 + \dots + f_n)$$

La **moda** és la freqüència més alta.

Pot ocórrer que una distribució de freqüències tinga més d'una moda. Per exemple, la distribució:

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

té 3 modes, 1, 3 i 6, ja que el valor més alt de la freqüència absoluta és 10 als tres casos.

La moda permet classificar els conjunts de dades en *unimodals*, *bimodals* o *plurimodals*, segons el nombre de modes que tinguen.

Per a obtenir la **mitjana** es calculen les freqüències acumulades i es busca el valor de la variable que ocupa el lloc central: $n/2$.

Utilitza l'ordinador

Copiem les dades de l'activitat resolta en un full de càlcul, escrivint x_i a la casella B1, f_i a la C1. En B2 escrivim 0, i en B3, 1. Seleccionem aquestes dues caselles i arrosseguem fins a la casella B12. Copiem les freqüències a la columna C. En A13 escrivim SUMA. Calculem la suma de les freqüències amb la tecla: Σ i s'obté 40 a la casella C13. A la columna D1 escrivim $x_i \cdot f_i$. A D2 escrivim = i punxem a B2, escrivim * i punxem a C2 ($=B2*C2$). Seleccionem D2 i arrosseguem fins a D12. Calculem la suma (251) i dividim el valor de la casella D12 entre el de la casella C12.

| | A | B | C | D | E |
|----|--------|-------|-------|-----------------|---------|
| 1 | | x_i | f_i | $x_i \cdot f_i$ | 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 | Màximo | | 8 | | |

Podem calcular el valor màxim de les freqüències, que en aquest cas es veu a ull, però si haguera molts més valors, moltes més files, es pot utilitzar la funció MAX.

Per a calcular les freqüències acumulades utilitzem la columna E. En E2 escrivim =C2. En E3 escrivim =E2+C3. Per què? I seleccionant E3 arrosseguem fins a E12.

Activitats proposades

9. S'ha llançat un dau 100 vegades i s'ha confeccionat la següent taula de freqüències absolutes:

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

- Calcula la mitja, moda i mitjana.
- Utilitza l'ordinador per a comprovar els resultats.

10. Llancem 2 daus i sumem els valors obtinguts. Repetim l'experiment 1000 vegades i obtenim les següent taula de freqüències absolutes.

| | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|----|----|----|
| 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 |
|-------|----|----|----|----|-----|-----|-----|----|----|----|----|

- Calcula la mitja, la mitjana i la moda.
- Utilitza l'ordinador per a comprovar els resultats.
- Repeteix tu els llançaments, ara només deu vegades, i calcula novament la mitja, mitjana i moda.

11. Utilitza l'ordinador per a calcular la mitja, la mitjana i la moda de la següent taula de freqüències absolutes, que indica el nombre de fills que tenen 200 famílies entrevistades:

| | | | | | | | | | | | |
|-------|----|----|----|----|---|---|---|---|---|---|----|
| 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 |

Paràmetres de dispersió

Ens donen una mesura del “dispersos” que estan les dades.

La primera mesura ens la dóna el **recorregut**, o el valor màxim menys el valor mínim.

Les més utilitzades són la **variància** i la **desviació típica** (o desviació estàndard) que mesura la distància de les dades respecte de la mitja.

Ja saps que la mitjana ens indica el valor de la variable que ocupa el lloc central. Es denomina **primer quartil (Q1)** al valor de la variable que deixa menors o iguals que ell a la quarta part de les dades, (o un 25 %), (sent per tant les tres quartes parts majors o iguals que ell). La mitjana és el segon quartil, que deixa per davall la meitat de les dades o un 50 %. El **tercer quartil (Q3)** és el valor de la variable que deixa menors o iguals que ell les tres quartes parts de les dades o un 75 % (i majors o iguals la quarta part). S'anomena interval **interquartílic** (o recorregut interquartílic) a la distància entre el tercer i el primer quartil (**Q3 – Q1**). Pel que hem dit, en aqueix interval estan la meitat de les dades.

Activitat resolta

Seguim amb la mateixa activitat anterior.

- ✚ Neus ha tingut en Matemàtiques les notes següents: 8, 4, 6, 10 i 10. Calcula el seu recorregut, la variància, la desviació típica, els quartils i l'interval interquartílic.

La major qualificació ha sigut un 10 i la menor un 4, per tant el **recorregut** és $10 - 4 = 6$.

$$\text{Recorregut} = \text{Màxim} - \text{Mínim.}$$

La mitja ja l'hem calculada i és 7,6. Volem analitzar com les observacions es separen de la mitja. Si a cada valor li restem la mitja, uns ixen positius i altres negatius, i si sumem tots, es compensen, per la qual cosa ix 0. És possible superar aqueixa dificultat calculant aqueixes diferències en valor absolut, o elevant-les al quadrat. Si les elevem al quadrat, sumem tot i dividim pel nombre total de valors de la variable menys 1, obtenim la variància.

Es divideix per $n - 1$ per a millorar les propietats de l'estadístic: Variància.

Si després calculem l'arrel quadrada, s'obté la desviació típica. Estem avaluant la distància dels valors de la variable a la mitja.

| | | | |
|--|-------|----------------------|--------------------------|
| | x_i | $x_i - \text{mitja}$ | $(x_i - \text{mitja})^2$ |
|--|-------|----------------------|--------------------------|

| | | | |
|-------------|----|------|-------------|
| 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 |
| Mitja = 7'6 | | | Suma = 27'2 |

Si dividim 27'2 entre 5 (n) s'obté 5'44 que és la variància.

Calculem l'arrel quadrada: 2'33 que és la desviació típica.

$$\text{Variància} = ((x_1 - \text{mitja})^2 + (x_2 - \text{mitja})^2 + \dots + (x_n - \text{mitja})^2)/n = \frac{\sum_{i=1}^n (x_i - m)^2}{n}$$

$$s = \text{Desviació típica} = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n}}$$

Es pot demostrar, fent operacions una fórmula més còmoda per a calcular la variància i la desviació típica:

$$\text{Variància} = \frac{\sum_{i=1}^n x_i^2}{n} - m^2 \quad s = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - m^2}$$

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

| | x_i | x_i^2 |
|-----------|-----------|------------|
| | 8 | 64 |
| | 4 | 16 |
| | 6 | 36 |
| | 10 | 100 |
| | 10 | 100 |
| $m = 7'6$ | Suma = 38 | Suma = 316 |

$$\text{Variància} = (316/5) - (7'6)^2 = 63'2 - 57'76 = 5'44.$$

La desviació típica és l'arrel quadrada de la variància, és a dir, $s = 2'33$.

Per a calcular els quartils hem d'ordenar les dades; $4 \leq 6 \leq 8 \leq 10 \leq 10$.

| | | | | |
|---|----------|---|-----------|----|
| 1 | 2 | 3 | 4 | 5 |
| 4 | 6 | 8 | 10 | 10 |





El primer quartil deixa per davall la quarta part o el 25 % de les dades. Hi ha 5 dades i $5/4 = 1'25$, com $1 < 1'25 < 2$, el primer quartil és 6. $Q1 = 6$. El tercer quartil deixa per davall les tres quartes parts o el 75 % de les dades: $3(5/4) = 3'75$. Com $3 < 3'75 < 4$, llavors $Q3 = 10$.

Interval interquartílic = $Q3 - Q1$.

A l'exemple, l'interval interquartílic = $Q3 - Q1 = 10 - 6 = 4$.

Utilitza l'ordinador

Igual que hem calculat la mitja, la mitjana i la moda, el full de càlcul es pot utilitzar per a obtenir:

-  El recorregut calculant $MAX - MIN$.
-  La variància utilitzant $VARP$.
-  La desviació típica usant $DESVESTP$.
-  Els quartils, ($QUARTIL$), sent el quartil 0 el mínim; el quartil 1, $Q1$; el quartil 2, la mitja; el quartil 3, $Q3$; i el quartil 4, el màxim.



| | A | B | C | D |
|----|-----------|------|--------------------------|----|
| 1 | | 4 | 6 | 8 |
| 2 | | 6 | 8 | 10 |
| 3 | | 8 | 10 | 10 |
| 4 | | 10 | 10 | |
| 5 | | 10 | | |
| 6 | | 10 | | |
| 7 | MAX | 10 | Recorrido = | 6 |
| 8 | MIN | 4 | | |
| 9 | VARP | 5,44 | | |
| 10 | DESVESTP | 2,31 | | |
| 11 | CUARTIL 1 | 6 | Intervalo intercuartil = | 4 |
| 12 | CUARTIL 3 | 10 | | |

Activitats proposades

12. Donades les temperatures en una ciutat d'un exercici anterior:

| Mesos | Gener | Febrer | Març | Abril | Maig | Juny | Juliol | Agost | Setembre | Octubre | Novembre | Desembre |
|-------------|-------|--------|------|-------|------|------|--------|-------|----------|---------|----------|----------|
| Temperatura | -2 | 5 | 8 | 9 | 11 | 13 | 27 | 33 | 21 | 14 | 9 | 4 |

- Calcula el recorregut, la variància, la desviació típica, els quartils i l'interval interquartílic.
- Utilitza l'ordinador per a comprovar els resultats.

13. Calcula el recorregut, la variància, la desviació típica, els quartils i l'interval interquartílic de les distribucions següents:

a) 2, 3, 4, 5, 7, 9, 9, 1000

b) 2, 3, 4, 5, 7, 9, 9, 10

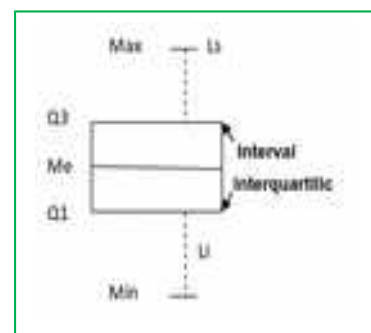
c) 0, 0, 4, 5, 7, 9, 9, 100, 200

Utilitza l'ordinador per a comprovar els resultats.

1.4. Diagrama de caixes

El diagrama de caixes és una representació gràfica en què s'utilitzen els quartils, la mitjana, els valors màxims i mínims... intentant visualitzar tot el conjunt de dades.

Es forma un rectangle (o caixa) els costats del qual són els quartils i on s'assenyala al centre, la mitjana. S'afegien dos braços (o bigots) on s'assenyalen els valors màxim i mínim.



Es poden calcular, a més, uns límits superior i inferior. L'inferior, L_1 ; és $Q_1 - 1,5$ per l'interval interquartílic, i el superior L_s és $Q_3 + 1,5$ per l'interval interquartílic.

El diagrama de caixa és el de la figura del marge.

A l'exemple anterior, una vegada ordenats les dades: $4 \leq 6 \leq 8 \leq 10 \leq 10$, hem calculat que:

Mitjana = $E_m = 8$. $Q_1 = 6$. $Q_3 = 10$. Interval interquartílic = 4.

Els bigots ens indiquen:

Màx = 10. Mín = 4. $L_s = Q_3 + 4 \cdot 1,5 = 16$. $L_i = Q_1 - 4 \cdot 1,5 = 0$.

En aquest exemple el màxim és igual a 10, que és menor que el possible extrem superior, igual a 16. El mínim és 4, major que l'extrem inferior, per tant no hi ha valors *atípics* que siguin majors que el límit superior o menors que el límit inferior. Els extrems dels bigots, al nostre exemple són 10 i 4.

1.5. Variable contínua: intervals i marques de classe. Histogrames

Recorda que les variables poden ser qualitatives, si no són numèriques, o quantitatives, que al seu torn poden ser discretes o contínues.

Per exemple: Si es fa un estudi estadístic sobre la població d'estudiants, es pot preguntar sobre la professió dels seus pares i mares, que és una variable qualitativa, sobre el nombre de germans, que és una variable quantitativa discreta (ningú té 3,7 germans), o sobre l'edat, l'estatura, la qualificació mitjana... que són variables quantitatives contínues.

Amb les variables quantitatives contínues té sentit agrupar els valors en intervals.

Al valor central de l'interval se'l denomina **marca de classe**.

La representació gràfica més adequada és l'**histograma** que és un diagrama de rectangles en què l'àrea de cada rectangle és proporcional a la freqüència. Té l'avantatge que d'aqueixa forma la freqüència de cada succés ve representada per l'àrea.

Activitat resolta



Realitza un estudi estadístic sabent que la taula de freqüències absolutes, amb intervals, dels pesos de 40 estudiants d'un centre escolar, és:

| Pes | [34, 40) | [40, 46) | [46, 52) | [52, 58) | [58, 64) | [64, 70) | [70, 76) |
|------------|----------|----------|----------|----------|----------|----------|----------|
| Estudiants | 2 | 10 | 12 | 9 | 4 | 2 | 1 |

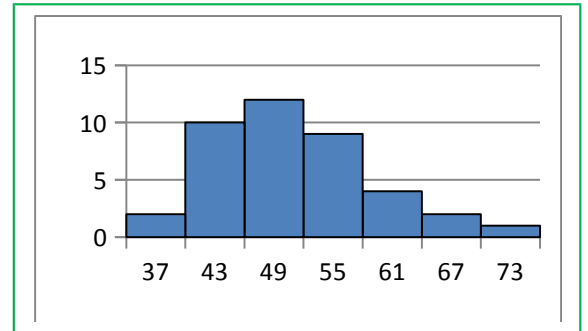
La taula ens diu que hi ha 2 estudiants el pes de la qual és major o igual a 34 i és menor que 40.

Calculem les marques de classe, buscant el punt mitjà de cada interval: $(40 - 34)/2 = 3$ i $34 + 3 = 37$. Tots els intervals en aquest exemple tenen una longitud de 6. Reescrivim la taula amb les marques de classe i les freqüències absolutes:

| x_i | 37 | 43 | 49 | 55 | 61 | 67 | 73 |
|-------|----|----|----|----|----|----|----|
|-------|----|----|----|----|----|----|----|

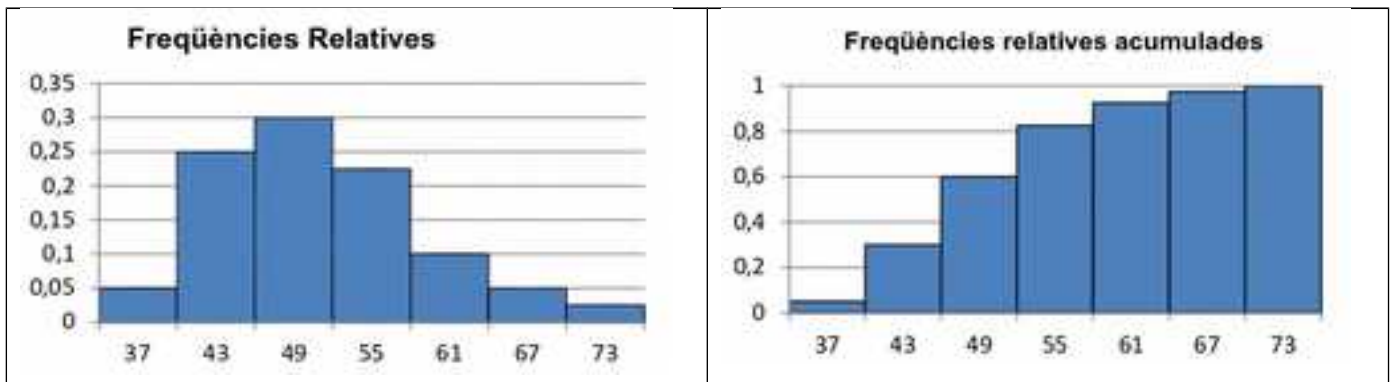
| | | | | | | | |
|-------|---|----|----|---|---|---|---|
| f_i | 2 | 10 | 12 | 9 | 4 | 2 | 1 |
|-------|---|----|----|---|---|---|---|

En aquest cas l'histograma de les freqüències absolutes és molt senzill perquè tots els intervals tenen la mateixa longitud. Si no fóra així, caldria calcular amb atenció les altures dels rectangles perquè les àrees foren proporcionals a les freqüències.



Representarem també l'histograma de les freqüències relatives i de les freqüències relatives acumulades:

| | | | | | | | |
|---|------|------|-----|-------|-------|-------|-------|
| x_i | 37 | 43 | 49 | 55 | 61 | 67 | 73 |
| Freqüències relatives | 0'05 | 0'25 | 0'3 | 0'225 | 0'1 | 0'05 | 0'025 |
| Freqüències relatives acumulades | 0'05 | 0'3 | 0'6 | 0'825 | 0'925 | 0'975 | 1 |



Càlcul de la mitja i la desviació típica:

Procedim de la forma que ja coneixem, calculant el producte de les marques de classe per les freqüències:

| | | | | | | | | |
|-----------------|----|-----|-----|-----|-----|-----|----|------|
| 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 | 2038 |

La **mitja** és igual a $2038/40 = 50'95$

Per a calcular la **desviació típica** restem a cada marca de classe, la mitja, elevem al quadrat i multipliquem per la freqüència 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 | 2619'9 |
|-------------------------|--------|---------|-------|--------|--------|---------|----------|--------|

La suma de les diferències de la mitja al quadrat per les freqüències relatives és 2619'9. Ara dividim entre n que al nostre cas és 40, i s'obté 65'5 que és la variància. Calculem l'arrel quadrada. La desviació típica és 8'09.

Activitat resolta

Utilitzem l'altra fórmula:
$$\text{Variància} = \frac{\sum_{i=1}^n (x_i^2 \cdot f_i)}{n} - m^2$$

| | | | | | | | | |
|-------------------|------|-------|-------|-------|-------|------|------|--------|
| x_i | 37 | 43 | 49 | 55 | 61 | 67 | 73 | Suma |
| f_i | 2 | 10 | 12 | 9 | 4 | 2 | 1 | 40 |
| x_i^2 | 1369 | 1849 | 2401 | 3025 | 3721 | 4489 | 5329 | 22183 |
| $x_i^2 \cdot f_i$ | 2738 | 18490 | 28812 | 27225 | 14884 | 8978 | 5329 | 106456 |

$\text{Variància} = (106456/40) - (50'95)^2 = 2661'4 - 2595'9 = 65'5$ i $\text{desviació típica} = s = 8'09$.

Vegem un altre exemple de càlcul de la mitja i la desviació típica utilitzant l'altra fórmula:

$$\text{Variància} = \frac{\sum_{i=1}^n (x_i^2 \cdot f_i)}{n} - m^2$$

| | | | | | | | | | | | | | |
|-------------------|------|------|------|-------|-------|--------|-------|-------|-------|-------|------|------|--------|
| 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 | 1518 | 1120 | 852 | 576 | 219 | 74 | 75 | 5577 |
| x_i^2 | 4096 | 4225 | 4356 | 4489 | 4624 | 4761 | 4900 | 5041 | 5184 | 5329 | 5476 | 5625 | |
| $x_i^2 \cdot f_i$ | 4096 | 0 | 8712 | 22445 | 41616 | 104742 | 78400 | 60492 | 41472 | 15987 | 5476 | 5625 | 389063 |

$n = 80$.

La mitja és igual a $m = 5577/80 = 69'7$.

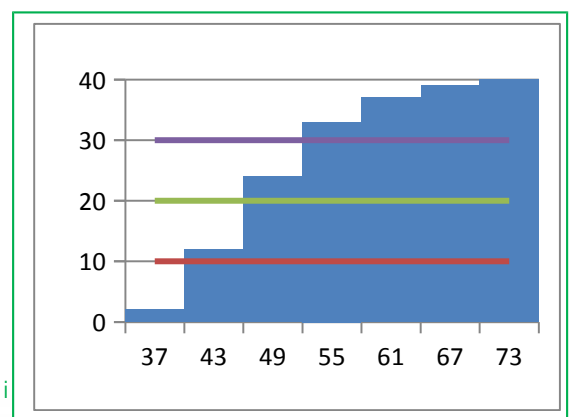
La variància és igual a
$$\frac{\sum_{i=1}^n (x_i^2 \cdot f_i)}{n} - m^2 = \frac{389063}{80} - 69,7^2 = 4863'2875 - 4858'09 = 5'1975$$

La desviació típica és igual a l'arrel quadrada de la variància, $s = 2'28$.

Càlcul de la mitjana i els quartils.

Representem l'histograma de freqüències absolutes acumulades, i tallem per les línies $n/2$ per a la mitjana, $n/4$ per al primer quartil, i $3n/4$ per al segon. Al nostre cas per 20, 10 i 30.

Observem, veient on les rectes horitzontals, $y = 20$, $y = 10$ i $y = 30$ tallen a l'histograma, que la mitjana està a l'interval



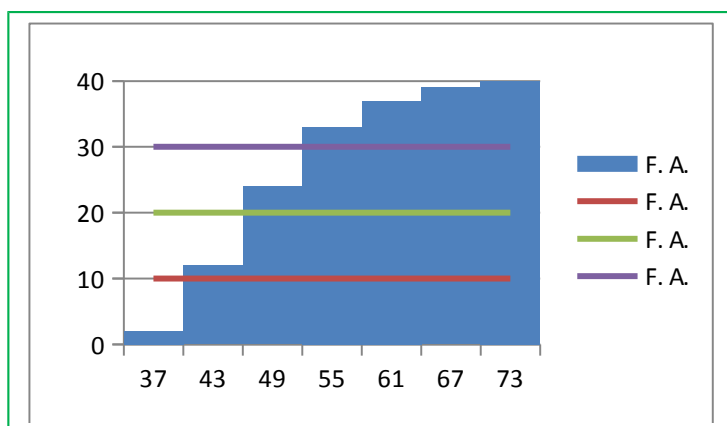
[46, 52) la marca de classe de la qual és 49, el primer quartil a l'interval [40, 46) la marca de classe del qual és 43, i el tercer quartil a [52, 58) la marca de classe del qual és 55.

| 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 |

Podem ajustar-lo més fent una interpolació lineal, és a dir, aproximant amb una recta.

Per a la mitjana tracem la recta que passa pels punts (46, 12) i (52, 24) ($y = 2x - 80$) i calculem on talla a la recta $y = 20$. Talla en $x = 50$. Per tant la mitjana és $Em = 50$.

El tercer quartil està a l'interval [52, 58). Calculem l'equació de la recta que passa pels punts (52, 24) i



(58, 33), que és $y = (3/2)x - 54$. Calculem on talla a $y = 30$, que és en $x = 56$. Per tant $Q_3 = 56$.

El primer quartil està a l'interval [40, 46). La recta que passa pels punts:

$$(40, 2) \text{ i } (46, 12)$$

té per equació $y = (5/3)x - 64,6666$, que talla a $y = 10$ en $x = 44,79999\dots$ $Q_1 = 44,8$.

Utilitza l'ordinador

Per a dibuixar histogrames amb l'ordinador utilitzant un full de càlcul ens trobem amb la dificultat que aquest dibuixa els rectangles separats. Dibuixa un diagrama de rectangles. Per a arreglar-lo al cas que la longitud de tots els intervals siga la mateixa, has d'assenyalar un dels rectangles, entrar en "donar format a la sèrie de dades" i, en "Opcions de sèrie" seleccionar en "Ample de l'interval" un ample del 0 %, és a dir, "sense interval". Si les longituds són distintes s'ha de calcular prèviament les altures dels rectangles.

Activitats proposades





- Utilitza l'ordinador per a dibuixar histogrames i repetir els càlculs de l'activitat resolta anterior.
- Es coneixen les quantitats de residus sòlids arreglats en m^3 /setmana durant 12 setmanes d'una urbanització: 23, 27, 30, 34, 38, 21, 30, 33, 36, 39, 32, 24. Escriu al teu quadern una taula de freqüències absolutes amb quatre intervals: [20, 25), [25, 30), [30, 35) i [35, 40). Calcula les marques de classe. Dibuixa l'histograma de freqüències absolutes. Calcula la mitja i la desviació típica. Calcula gràficament la mitjana i els quartils.
- Fes un estudi estadístic preguntant als teus companys i companyes de classe sobre el nombre de llibres que lliges al mes. Confecciona una taula i representa-la en un diagrama de rectangles, un polígon de freqüències i un diagrama de sectors.

2. DADES BIDIMENSIONALS

2.1. Idees generals

Possiblement, l'aplicació més important de l'estadística no siga l'estudi d'una variable aïllada sinó l'anàlisi de les relacions entre variables. Si tenim dues mesures que es donen juntes, és lògic voler saber en quina mesura una influeix en l'altra. Vegem alguns exemples.

Exemples:




-  En una botiga de camises, volem saber quantes vendrem (generalment) en funció del preu.
-  Si sabem l'altura del pare d'un xiquet, quina serà l'altura del fill?
-  Si a un grup d'alumnes li donem una paga i mesurem les seues qualificacions. Les alumnes que reben més diners trauen millors notes? Quant més? Aquest mateix estudi pot fer-se amb els treballadors d'una empresa. Si se'ls paga més augmenta la producció?
-  Són més intel·ligents els hòmens que les dones? O viceversa?

Pot parèixer que algun d'aquests casos és elemental. És obvi que els pares alts tenen fills alts i que si baixa el preu, veng més. Però l'important és QUANT. Si jo tinc una botiga, el que jo vull és guanyar diners. I per descomptat que si pose les camises a 0 € vendré molt... però no guanyaré res. El que vull és una estimació de quant veng a cada preu per a poder saber el preu que m'interessa posar.

2.2. Variables bidimensionals. Freqüències conjuntes

Una **variable bidimensional** són dos variables que es mesuren conjuntament. Si X i Y són les variables, la variable bidimensional és (X, Y) .

Exemples:

-  El preu a què posem les camises (X) i el preu anterior (Y). L'altura d'un pare (X) i l'altura del fill (Y)
-  El color del pèl (X) i el color dels ulls (Y).
-  El sexe d'una persona (X) i el seu coeficient d'intel·ligència (Y).

Dóna't compte que les variables bidimensionals poden ser qualitatives o quantitatives i inclús cada una d'un tipus. Així mateix podríem tindre les dades agrupades, i llavors el que hi hauria seria parelles d'interval.

La representació de forma de taula de freqüències és exactament igual que al cas unidimensional amb l'excepció que ara tenim parelles. Anem primer amb un exemple i després introduïrem els conceptes.

Exemple:



Tenim una mostra de 8 persones i mirem el seu color d'ulls i cabell. Hi ha 4 morenos d'ulls marrons, 1 moreno d'ulls verds, dos rossos d'ulls blaus i un ros d'ulls verds.

Encara no hem definit les freqüències però creiem que ho pots entendre igual. La taula és:

| Individu | Freqüències absolutes | Freqüències relatives |
|-------------------|-----------------------|-----------------------|
| (Moreno, marrons) | 4 | $0'5 = 4/8$ |
| (Moreno, verds) | 1 | $0'125 = 1/8$ |
| (Ros, blaus) | 2 | $0'25 = 2/8$ |
| (Ros, verds) | 1 | $0'125 = 1/10$ |
| TOTAL | 8 | 1 |

Com pots veure, per a que dos elements siguin iguals, han de ser iguals les dues components. La variable X és el color del cabell i la variable Y el color dels ulls. Es té $X = \{\text{"Moreno"}, \text{"Ros"}\}$ i $Y = \{\text{"Marrons"}, \text{"Verds"}, \text{"Blaus"}\}$. No té per què haver-hi el mateix nombre de valors en cada variable.

Les definicions són les mateixes.

La freqüència **absoluta** és el nombre de vegades que s'ha obtingut aqueixa parella de resultats (dues parelles són iguals si les seues dues components són iguals).

La freqüència **relativa** és la freqüència absoluta dividida entre el nombre total de dades.

Taula de freqüències conjunta:

De vegades, en compte de mostrar les dades en parelles, es posen en una **taula de doble entrada o taula de contingència**. S'anomena així perquè la X està en vertical i la Y en horitzontal. Als encreuaments es posen les freqüències, ja siguin absolutes o relatives. Si es posen les absolutes es diu **taula de doble entrada de freqüències absolutes** i si es posen les relatives aleshores **taula de doble entrada de freqüències relatives**.

La taula anterior, amb (x_i, y_j) no té un nom especial universalment acceptat. Podem anomenar-la **taula de freqüències de parelles**.

Exemple:



Tenim la mateixa mostra d'abans: 4 morenos d'ulls marrons, 1 moreno d'ulls verds, dos rossos d'ulls blaus i un ros d'ulls verds. Anem a col·locar-los en taules de doble entrada de freqüències absolutes i després relatives.

Ens limitem a posar a la primera columna els dos valors que tenim de la X , que és el color de cabell ("Moreno" i "Ros") i a la primera fila els de la Y , que és el color dels ulls ("Marrons", "Verds" i "Blaus").

| Y | Marrons | Verds | Blaus |
|---|---------|-------|-------|
|---|---------|-------|-------|

| X | | | |
|--------|---|---|---|
| Moreno | 4 | 1 | 0 |
| Ros | 0 | 1 | 2 |

Observa que en aquesta taula poden aparèixer zeros, que representen que no hi ha ningú amb aqueixa parella de característiques.

Si dividim les freqüències absolutes pel nombre total de dades (que en aquest cas és 8) obtenim la taula de doble entrada de freqüències relatives.

| Y | Marrons | Verds | Blaus |
|--------|-----------|-----------|----------|
| X | | | |
| Moreno | 0'5 = 4/8 | 0'125=1/8 | 0 |
| Ros | 0 | 0'125=1/8 | 0'25=2/8 |

Activitats proposades

17. Amb la taula de valors de l'exemple, construeix la taula de freqüències absolutes i relatives de la variable X ("Color de cabell") i la variable Y ("Color d'ulls") per separat, com a variables unidimensionals.
18. Completa la següent taula i expressa-la en forma de taula de doble entrada, primer amb freqüències relatives i després amb freqüències absolutes.

| (x_i, y_i) | Freqüència absoluta | Freqüència relativa |
|--------------|---------------------|---------------------|
| (0, 1) | 12 | |
| (1, 2) | 14 | |
| (2, 3) | 14 | |

19. Completa la següent taula de freqüències conjunta i expressa-la en freqüències de parelles (x_i, y_i) , tant amb freqüències relatives com absolutes.

2.3. Diagrama de dispersió i recta de regressió

Un **diagrama de dispersió**, també anomenat **núvol de punts** per la seua aparença, és un gràfic que s'obté representant cada parella com un punt del pla cartesià. S'usa principalment amb variables quantitatives i dades sense agrupar (si estigueren agrupades prendríem les marques de classe).

És molt simple de dibuixar. N'hi ha prou amb posar un punt en cada parella. De vegades si hi ha valors repetits es posen els punts més grossos però també és comú posar-los tots igual.

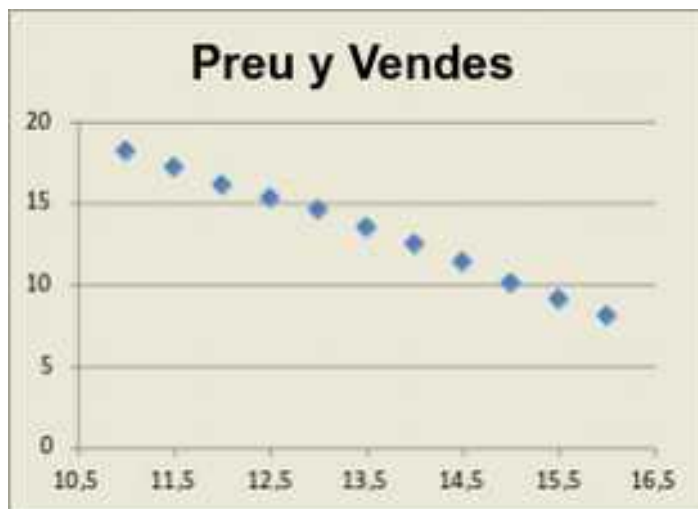
Exemple:



Tenim una botiga i volem estudiar les vendes d'una camisa en funció del preu. Per a això, provem cada setmana amb un preu distint i calculem les vendes mitges. Obtenim així una taula com la que segueix

| | | | | | | | | | | | |
|------------------------|------|------|------|------|------|------|------|------|------|------|-----|
| Preu | 11 | 11'5 | 12 | 12'5 | 13 | 13'5 | 14 | 14'5 | 15 | 15'5 | 16 |
| Vendes (mitges) | 18'2 | 17'2 | 16'1 | 15'3 | 14'6 | 13'5 | 12'5 | 11'4 | 10'1 | 9'1 | 8'1 |

Si copiem les dades a un full de càlcul i li donem a dibuixar un diagrama de dispersió, obtenim un diagrama com el següent:



que és el típic gràfic que pot veure's per a fer un estudi de resultats en qualsevol empresa.

La recta de regressió

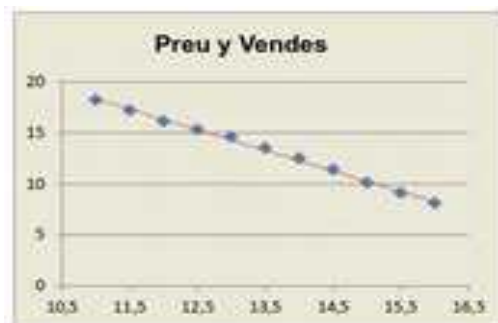
El problema amb el núvol de punts és que simplement descriu el que passa. Açò certament és important en si mateix, però el que és realment interessant és PREDIR què passarà. A l'exemple anterior, les nostres dades arriben a preus de 16 €. Què passaria si pugem el preu a 17 €? O l'abaixem a 9 €? I amb els preus intermedis, com 12'25 €?

Com hi ha infinits preus, no podem tindre en compte infinit preus. L'interessant és tindre un model matemàtic que ens diga, per a un preu donat, quin és el valor esperat de les vendes. O, en general, per a un valor de X quin és el valor esperat de Y .

El més fàcil és fer una recta que s'aproxime. Es pot dibuixar pràcticament a mà, però hi ha una fórmula matemàtica que la calcula. Aqueixa fórmula és complicada i està fora de l'abast d'aquest curs però sí que anem a ensenyar-te com fer-la amb ordinador.

Abans de res, anem a mostrar-te a l'exemple anterior la línia de tendència.

Observa que no passa per tots els punts, sinó que uns queden dalt i altres baix. De fet és impossible que una recta passe per tots i, al món real, l'ajust mai és exacte. La recta passa pel mig dels punts.



Utilitza l'ordinador

La següent taula són dades de l'altura d'un pare i de la del seu fill amb 15 anys d'edat. Les altures estan en metres.

| | | | | | | | | |
|-------------|------|-----|-----|-----|------|------|-----|------|
| Pare | 1'7 | 2 | 1'6 | 1'7 | 1'65 | 1'9 | 1'9 | 1'81 |
| Fill | 1'75 | 1'9 | 1'7 | 1'8 | 1'6 | 1'88 | 2 | 1'95 |

Primer, farem el diagrama de dispersió. Copiem les dades en un full de càlcul. Els posarem en vertical perquè es veja millor, però es podria fer exactament igual en horitzontal.

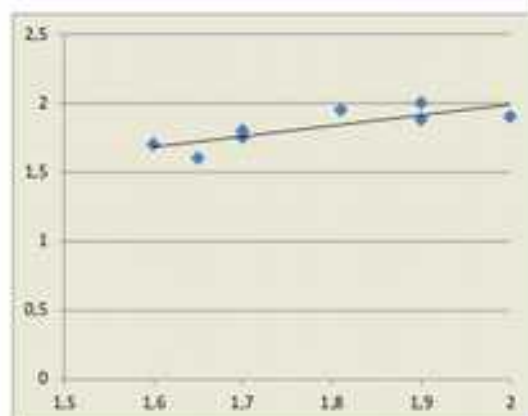
Després, assenyallem les dues sèries i li donem a *inserir gràfic de dispersió*.

Automàticament ens apareix el diagrama de dispersió (núvol de punts). Si jugues un poc amb les opcions pots modificar el títol, el format, l'escala dels eixos...



Més encara, la recta de regressió és molt fàcil de dibuixar. Basta amb què seleccions el gràfic i li dones a *anàlisi i a línia de tendència*. Triant una tendència lineal, ja tens la recta de regressió.

Al final, si ho has fet bé, el dibuix ha de ser més o menys quelcom semblant a açò:



I fixa't, la recta té tots els valors possibles. Per a veure quin valor correspondria a una altura del pare de 1'75 m, ho busquem a la recta.

2.4. Interpretació de la recta. Introducció a la correlació

Una vegada hem dibuixat la recta de regressió, podem veure com és la relació entre les dues variables. En essència el tipus de relació ve donada pel pendent de la recta.

1. Si la recta de regressió té pendent positiva (més informalment, “si va cap amunt”) es diu que la relació entre les variables és **positiva**.
2. Si la recta de regressió té pendent zero (més informalment, “si queda horitzontal”) es diu que la relació entre les variables és **nul·la** o que **no hi ha relació lineal**.
3. Si la recta de regressió té pendent negativa (més informalment “si va cap avall”) es diu que la relació entre les variables és **negativa**.

La qüestió és, per tant, senzilla. Basta dibuixar la recta i veure cap a on va. Però també ens interessa veure si els punts estan prop de la recta o lluny. En altres paraules, mirar si la recta *ajusta bé* o *ajusta malament*.

Per a calcular açò, s'obté el que s'anomena **coeficient de correlació**. Es defineix com:

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

Ja veus, molt complicat! Però, com abans, n'hi ha prou amb usar Excel o qualsevol full de càlcul. L'orde en Excel és COEF.DE.CORREL(serie1;serie2).

El coeficient de correlació ens mesura si la relació és positiva, negativa o nul·la. I TAMBÉ ens diu si l'ajust és bo. Veurem en un quadre els detalls.

El **coeficient de correlació, ρ** , mesura la relació entre dues variables. És un nombre entre -1 i 1 (pot ser exactament -1 o exactament 1).

Si el coeficient de correlació és exactament 1 la relació és **perfecta positiva**. La recta va cap amunt i TOTS els punts estan sobre ella.

Si el coeficient de correlació està a l'interval $(0, 1)$ la relació és **positiva**. La recta va cap amunt però no passa per tots els punts.

Si el coeficient de correlació és exactament 0 , la relació és **nul·la** (no hi ha relació lineal).

Si el coeficient de correlació està en $(-1, 0)$ la relació és **negativa**. La recta va cap avall però no passa per tots els punts.

Si el coeficient de correlació és exactament -1 la relació és **perfecta negativa**. La recta va cap avall i TOTS els punts estan sobre ella.

Açò és el que és objectiu. En algunes ocasions, es parla de correlació positiva forta (si està pròxima a 1) o positiva dèbil (si està entre 0 i 1 però pròxima a 0) i el mateix negativa. Però clar, això depèn de la interpretació de cada u. Així, una correlació de $0,96$ és positiva forta i una de $-0,02$ és negativa dèbil.

Resum

- $\rho = 1 \rightarrow$ correlació perfecta positiva
- $\rho = -1 \rightarrow$ correlació perfecta negativa
- $\rho = 0 \rightarrow$ correlació nul·la
- $\rho \in (0, 1) \rightarrow$ correlació positiva
- $\rho \in (-1, 0) \rightarrow$ correlació negativa

Però, i 0'55? això depèn del que consideres. El que sí que és objectiu és si és perfecta o nul·la, positiva o negativa.

Utilitza l'ordinador

Amb les dades de l'activitat anterior, calcularem el coeficient de correlació.

L'única cosa que cal fer és posar, en la casella corresponent =COEF.DE.CORR. al nostre exemple és la casella D2.

Automàticament ens dona a triar dos matrius i triem primer de la X i després de la Y.

Ens dona el coeficient de correlació, que en aquest cas resulta ser 0'81. És una relació positiva forta com ja imaginàvem pel núvol de punts i la recta de regressió.



Utilitza l'ordinador

Preguntem a 10 alumnes de 4t ESO per les seues qualificacions en Matemàtiques, pel nombre de minuts diaris que veuen la televisió, pel nombre d'hores setmanals que dediquen a l'estudi, i per la seua estatura en centímetres. Les dades s'arreglen a la taula adjunta. Volem dibuixar els núvols de punts que els relacionen amb les qualificacions de Matemàtiques, el coeficient de correlació i la recta de regressió.

| | | | | | | | | | | |
|--------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Qualificacions de Matemàtiques | 10 | 3 | 7 | 8 | 5 | 9 | 9 | 8 | 6 | 7 |
| Minuts diaris que veu la TV | 0 | 90 | 30 | 20 | 70 | 10 | 15 | 25 | 60 | 25 |
| Hores setmanals d'estudi | 15 | 2 | 9 | 12 | 7 | 14 | 13 | 11 | 7 | 8 |
| Estatura (en cm) | 177 | 168 | 157 | 159 | 163 | 179 | 180 | 175 | 169 | 170 |

Per a fer-ho, entrem en Excel, i copiem les dades. Seleccionem la primera i la segona fila, després la primera i la tercera i finalment la primera fila i la quarta.

Amb la primera i segona files seleccionades, inserirem, *Dispersió* i triem el *núvol de punts*. Podem aconseguir que l'eix d'abscisses vaja de 0 a 10 en "*Donar format a l'eix*". Punxem sobre un punt del núvol, i triem "*Agregar línia de tendència*". Perquè dibuixi l'ordinador la recta de regressió la línia de tendència ha de ser *Lineal*. A la pantalla que apareix marquem la casella que diu: "*Presentar equació en el gràfic*" i la casella que diu "*Presentar el valor de R quadrat en el gràfic*".



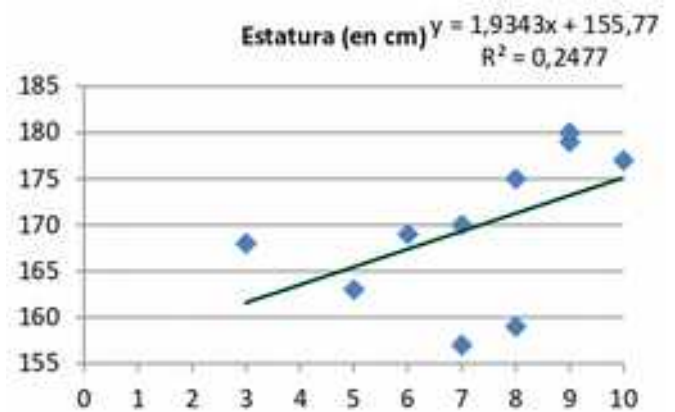
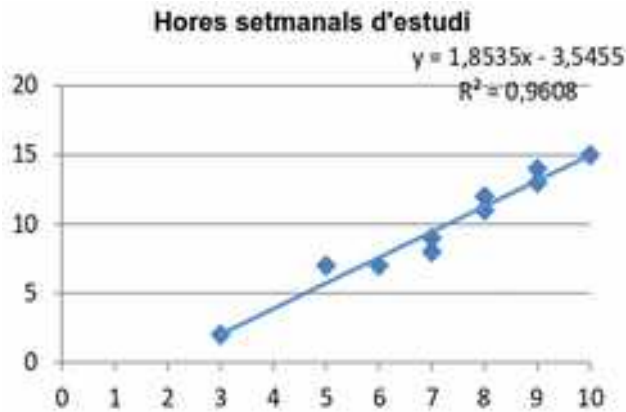
Observa, la recta de regressió, en color roig, és decreixent i la seua equació és aproximadament:

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

El quadrat del coeficient de correlació és $\rho^2 = 0.95$. La correlació és negativa i alta:

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

Fem el mateix amb la primera i tercera fila i amb la primera i quarta fila. Obtenim els gràfics:



Observa que en ambdós casos el pendent de la recta de regressió és positiva però al primer el coeficient de correlació, positiu, és pròxim a 1, $\rho = \sqrt{0,96} = 0,98$. La correlació és alta i positiva.

Al segon $\rho = \sqrt{0,25} = 0,5$.

Activitats resoltes



El propietari d'una instal·lació mixta solar-eòlica està realitzant un estudi del volum d'energia que és capaç de produir la instal·lació. Per a això, mesura la dita energia al llarg d'un total de $N = 16$ dies que considera prou representatius. L'energia (en kWh) produïda als dits dies per les instal·lacions solar i eòlica es troba arreplegada a la taula següent:

| | | | | | | | | | | | | | | | | |
|----------------------------|------|------|------|------|------|------|-----|-----|------|------|------|------|------|------|------|------|
| Generació 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 |
| Generació eòlica (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 |

Realitzarem una activitat resolta completa utilitzant les fórmules de la mitjana, la desviació típica i de la correlació perquè puguem servir-te de models si necessites alguna vegada calcular-les sense ajuda de l'ordinador.

Denotem a la generació solar com a variable **X** i la generació eòlica com a variable **Y**. Afegim noves files a la nostra taula, els quadrats de x , de y i els productes d'ambdues:

| | | | | | | | | | | | | | | | | |
|----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Generació 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 |
| Generació eòlica (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 |

Càlcul de les mitges:

Sumant la primera fila i dividint per $N = 16$, obtenim la mitja de la Generació Solar en Kwh.

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

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{13'1 + 10'5 + 4'1 + 14'8 + 19'5 + 11'9 + 18 + 8'6 + 5'7 + 15'9 + 11'2 + 6'8 + 14'2 + 8'2 + 2'6 + 9'7}{16} = 10'925 \text{ Kwh}$$

Sumant la segona fila i dividint per $N = 16$ obtenim la mitja de la Generació Eòlica en Kwh:

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N} = \frac{8'5 + 14'3 + 24'7 + 4 + 2'3 + 6'4 + 3'6 + 9'2 + 13'5 + 1'4 + 7'6 + 12'8 + 10'3 + 16'5 + 21'4 + 10'9}{16} = 10'463 \text{ Kwh}$$

Les mitges són:

$$\bar{x} = 10'925 \text{ Kwh i } \bar{y} = 10'463 \text{ Kwh,}$$

Molt paregudes.

Càlcul de les desviacions típiques:

A la tercera fila hem calculat els quadrats dels valors de la primera variable i els utilitzem per a calcular la variància:

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

$$s_x^2 = \frac{\sum_{i=1}^N x_i^2}{N} - \bar{x}^2 = \frac{13'1^2 + 10'5^2 + 4'1^2 + 14'8^2 + 19'5^2 + 11'9^2 + 18^2 + 8'6^2 + 5'7^2 + 15'9^2 + 11'2^2 + 6'8^2 + 14'2^2 + 8'2^2 + 2'6^2 + 9'7^2}{16} - 10'9^2 = \frac{141'5}{16} - 10'9^2 = 22'16$$

A la quarta fila els quadrats dels valors de la segona variable i calculem la seua variància:

$$s_y^2 = \frac{\sum_{i=1}^N y_i^2}{N} - \bar{y}^2 = \frac{8'5^2 + 14'3^2 + 24'7^2 + 4^2 + 2'3^2 + 6'4^2 + 3'6^2 + 9'2^2 + 13'5^2 + 1'4^2 + 7'6^2 + 12'8^2 + 10'3^2 + 16'5^2 + 21'4^2 + 10'9^2}{16} - 10'5^2 = \frac{150'48}{16} - 10'5^2 = 41'01$$

La desviació típica és l'arrel quadrada de la variància, per tant:

$$s_x = \sqrt{22'16} = 4'71 \text{ i}$$

$$s_y = \sqrt{41'01} = 6'4$$

Càlcul del coeficient de correlació:

Per a calcular el coeficient de correlació calculem en la cinquena fila els productes de la variable x per la variable y . Així, $13'1 \cdot 8'5 = 111'4$.

Volem calcula el terme: $\frac{\sum_{i=1}^N (x_i \cdot y_i)}{N}$.

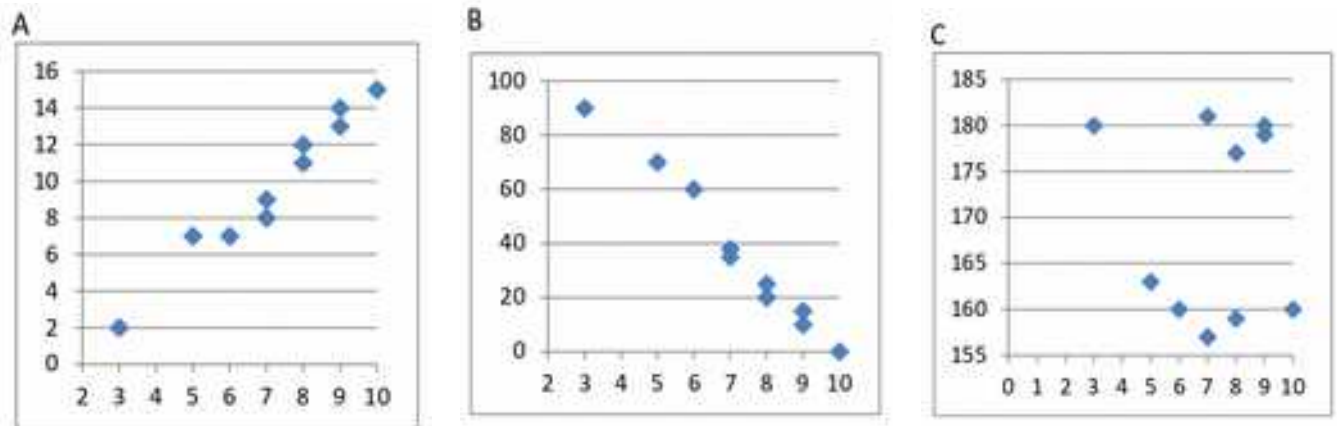
En sumar aqueixa fila obtenim $1401'2$, que dividim entre 16, li restem el producte de les mitges i dividim pel producte de les desviacions típiques:

$$\rho = \frac{\sum_{i=1}^N (x_i \cdot y_i) - \bar{x} \cdot \bar{y}}{s_x \cdot s_y} = \frac{1401'2 - (10'9 \cdot 10'5)}{4'71 \cdot 6'4} = \frac{-26'728}{4'71 \cdot 6'4} = -0'887$$

Aquest coeficient de correlació negatiu i pròxim a -1 ens indica que la relació entre les dues variables és negativa i prou important.

Activitats proposades

20. Maria ha calculat els coeficients de correlació dels tres núvols de punts adjunts, i ha obtingut: $-0,05$, $0,98$ i $-0,99$, però ara no recorda quin és de cada una. Pots ajudar-la a decidir quin coeficient correspon amb cada núvol?



21. Fes una enquesta entre els teus companys de classe. Amb ella realitzaràs un treball d'investigació i presentaràs un informe. Tria amb atenció les preguntes. Preguntaràs a cada un dels teus companys seleccionats, la mostra, dues preguntes, com per exemple el que mesura la seua mà i la seua nota en llengua, però a tu poden interessar-te altres qüestions molt distintes.

- La primera cosa que faràs és tabular les respostes i confeccionar dues taules de freqüències absolutes. Després completa aqueixes mateixes taules amb les freqüències relatives i les freqüències acumulades. Fes representacions gràfiques d'aqueixes freqüències: de barres, de línies, de sectors.
- Calcula les mitges, modes i mitjanes així com recorregut, desviació típica, quartils, interval interquartílic... Representa les dades en una taula de doble entrada i dibuixa el núvol de punts. Calcula el coeficient de correlació. Presenta un informe d'aquest treball.

3. ATZAR I PROBABILITAT

3.1. Experiment aleatori i succés

Un **fenomen o experiment aleatori** és aquell que, mantenint els mateixes condicions en l'experiència, no és pot predir el resultat.

- Són experiments aleatoris:
 - a) Llançar una moneda i anotar si ix cara o creu.
 - b) Llançar un dau i anotar el nombre de la cara superior.
 - c) Llançar dos daus o dues monedes.
 - d) Si a una urna hi ha boles blanques i roges, traure una a l'atzar i anotar el color.
 - e) Traure una carta d'una baralla.
 - f) Traure, sense reemplaçament, dues cartes de la baralla.
 - g) Obrir un llibre i anotar la pàgina per la qual s'ha obert.

No obstant això, calcular el cost d'una mercaderia, sabent el pes i el preu per kg, no és un experiment aleatori. Tampoc ho és calcular el cost del rebut de la llum sabent el gasto.

- No són experiments aleatoris
 - a) Eixir al carrer sense paraigües quan plou i veure si et mulles.
 - b) El preu de mig quilo de rosquilles si les rosquilles costen a 3 € el quilo.
 - c) Soltar un objecte i veure si cau.

Activitats proposades

22. Indica si són, o no, fenòmens aleatoris:

- a) La superfície de les comunitats autònomes espanyoles.
- b) Anotar el sexe del pròxim bebè nascut en una clínica determinada.
- c) L'àrea d'un quadrat del que es coneix el costat.
- d) Tirar tres daus i anotar la suma dels valors obtinguts.
- e) Saber si l'any que ve és bixest.

En realitzar un experiment aleatori hi ha diversos possibles resultats o **successos possibles**.

En realitzar un experiment aleatori sempre s'obindrà un dels **possibles resultats**.

S'anomena **succés elemental** a cada un dels possibles resultats d'un experiment aleatori.

El conjunt dels possibles resultats d'un experiment aleatori es denomina **espai mostral**.

Un **succés** és un subconjunt del conjunt de possibles resultats, és a dir, de l'espai mostral.

Activitat resolta

- **Per exemple** els possibles resultats en tirar una moneda són que isca *cara* o isca *creu*. El conjunt de successos elementals és {cara, creu}.
- *El conjunt de possibles resultats dels experiments aleatoris següents:*
 - a) Extraure una bola d'una bossa amb 9 boles blanques i 7 negres és {blanca, negra}.

- b) Traure una carta d'una baralla espanyola és $\{AO, 2O, 3O, \dots, SO, CO, RO, AC, \dots, RC, AB, \dots, RB, AE, \dots, RE\}$.
- c) Tirar dues monedes és: $\{(cara, cara), (cara, creu), (creu, cara), (creu, creu)\}$.
- En llançar un dau, el conjunt de possibles resultats és $\{1, 2, 3, 4, 5, 6\}$, el succés obtindre parell és $\{2, 4, 6\}$, el succés obtindre imparell és $\{1, 3, 5\}$, el succés obtindre múltiple de 3 és $\{3, 6\}$, traure un nombre menor que 3 és $\{1, 2\}$.
 - En llançar dues monedes el conjunt de possibles resultats és $\{(C, C), (C, +), (+, C), (+, +)\}$. El succés traure zero cares és $\{(+, +)\}$, traure una cara és $\{(C, +), (+, C)\}$ i traure dues cares $\{(C, C)\}$.

Activitats proposades

23. Escriu el conjunt de possibles resultats de l'experiment aleatori: "Escriure en cinc targetes cada una de les vocals i traure una a l'atzar".
24. Escriu el conjunt de possibles resultats de l'experiment aleatori: "Tirar una xinxeta i anotar si cau de punta o no".
25. Inventa dos successos de l'experiment aleatori: Tirar dues monedes.
26. Al joc de loteria, indica dos successos respecte a la xifra de les unitats del primer premi.
27. Escriu tres successos aleatoris de l'experiment aleatori traure una carta d'una baralla espanyola.

3.2. Freqüència i probabilitat

No definirem "probabilitat", perquè hi ha diverses definicions possibles. Hi ha una axiomàtica deguda a Kolmogorov relativament recent (1930), però abans ja s'havia sigut usat aquest concepte per exemple per Fermat i Pascal al segle XVII que es van escriure cartes reflexionant sobre el que ocorria als jocs d'atzar. Quan no comprenien com assignar una determinada probabilitat, jugaven moltes vegades al joc que fóra i veien a quin valor s'aproximaven les freqüències relatives. Així, la **probabilitat d'un succés** podria definir-se com el **límit a què tendeixen les freqüències relatives** d'aqueix succés quan el nombre d'experiments és molt alt. Per tant:

Per a calcular probabilitats s'usen dues tècniques, una experimental, analitzant les freqüències relatives de què ocorrega el succés, i l'altra per simetria, quan se sap que els successos elementals són **equiprobables**, és a dir, que **tots ells tenen la mateixa probabilitat**, llavors es **divideix el nombre de casos favorables pel nombre de casos possibles**.

Açò últim, quan es pot usar, simplifica la forma d'assignar probabilitats i es coneix com a **Regla de Laplace** que diu que: "Si els successos elementals són equiprobables, la probabilitat d'un succés és el nombre de casos favorables dividit pel nombre de casos possibles".

Activitat resolta

- La probabilitat que isca cara en tirar una moneda és $1/2$, perquè només hi ha dos casos possibles $\{cara, creu\}$, un únic cas favorable, *cara*, i suposem que la moneda no està trucada. Si sospitàrem que la moneda estiguera trucada per a assignar aqueixa probabilitat caldria tirar la moneda un muntó de vegades per a observar cap a quin valor s'acosta la freqüència relativa d'obtindre cara.
- La probabilitat de traure un 5 en tirar un dau és $1/6$ perquè hi ha sis casos possibles $\{1, 2, 3, 4, 5, 6\}$, un únic cas favorable, 5, i suposem que el dau no està trucat, després tots ells són equiprobables.

- La probabilitat que en creuar el carrer t'atropelle un cotxe NO és $1/2$, encara que només hi ha dos casos possibles, que t'atropelle el cotxe i que no t'atropelle, perquè ja t'hauria agarrat un muntó de vegades. Per a calcular aqueixa probabilitat s'arrepleguen dades de vianants atropellats i es calcula utilitzant les freqüències relatives.
- La probabilitat de traure una bola roja d'una bossa amb 7 boles roges i 3 boles blanques és $7/10$.
- La probabilitat que un bebè siga xiqueta és aproximadament 0,5, però en fer l'estudi amb les freqüències relatives s'ha vist que és 0,49.
- Si considerem una baralla espanyola de 40 cartes i triem una carta, alguns dels successos que poden ocórrer són "traure un or", o "traure un as", o "traure el cavall de copes"... Com per endavant no sabem el que ocorrerà diem que aquests successos són *aleatoris* o *d'atzar*. Abans de traure cap carta totes elles són igualment factibles, i com pot eixir una qualsevol de les 40 cartes diem que la probabilitat, de per exemple, *traure el cavall de copes* és $1/40$, la de *traure un or* és $10/40$, i la d'un *as* és $4/40$.



Quina és la probabilitat de traure el rei de copes? I de traure un rei? I una copa?

La probabilitat de traure *l'as de copes* és $1/40$. Però el succés *traure un as* es compleix si ix l'as d'ors, o de copes, o de bastos o d'espases. És a dir, no és un succés simple, està format, en aquest cas per 4 successos elementals, per tant la seua probabilitat és $4/40 = 1/10$. El mateix li ocorre a *traure una copa*. És un succés compost, i com hi ha 10 copes la seua probabilitat és $10/40 = 1/4$.

Activitats proposades

28. Calcula la probabilitat que en traure una carta de la baralla siga una espasa.

29. Per a saber la probabilitat que un xiquet de bolquers siga esquerrà, et basaries a l'estudi de les freqüències relatives o l'assignaries per simetria?

3.3. Assignació de probabilitats

Succés contrari

Activitats resoltes



Quina és la probabilitat de traure un as en la baralla de 40 cartes? I de **no** traure un as? I de traure una copa? I de **no** traure una copa?

El succés *no traure un as* és el succés **contrari** al de *traure un as*. Cartes que no són asos hi ha 36, per tant la probabilitat de no traure as és $36/40 = 9/10$. Observa que s'obté que $p(\text{as}) + p(\text{no as}) = 1/10 + 9/10 = 10/10 = 1$.

La probabilitat de *traure copa* és $10/40$, i hi ha 30 cartes que no són copes, per tant la probabilitat de **no traure copa** és $30/40$, i $10/40 + 30/40 = 1$.

Si designem per $p(X)$ a la probabilitat d'un succés X i per $p(\text{no}X)$ a la probabilitat del seu **succés contrari** resulta que:

$$p(X) + p(\text{no}X) = 1.$$

La probabilitat d'un succés més la probabilitat del seu succés contrari és igual a 1.

Activitats proposades

30. Quina és la probabilitat de **no** traure un 5 en tirar un dau? I de **no** traure un múltiple de 3? I de **no** traure un nombre menor que 2?

31. En tirar una moneda dues vegades, quina és la probabilitat de **no** traure cap cara? I de traure

almenys una cara? Observa que traure almenys una cara és el succés contrari de no traure cap cara.

Successos dependents i independents

Exemple:



Tenim una bossa amb 3 boles roges i 2 boles negres. Quina és la probabilitat de *traure una bola roja*? Si traiem dues boles, quina és la probabilitat de *traure dues boles roges*?

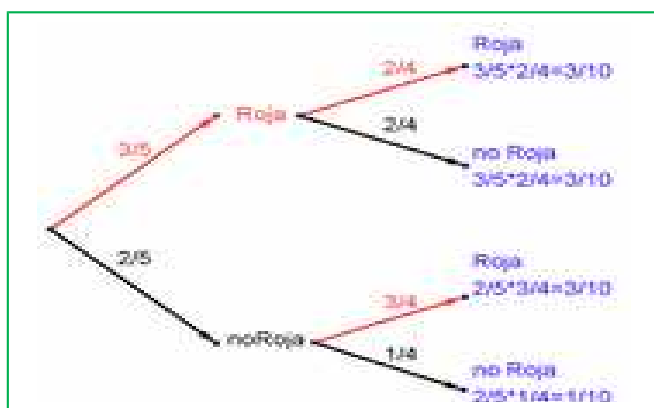
La probabilitat de traure una bola roja és $3/5$. Però la de traure dues boles roges, depèn!

Depèn de si tornem a ficar en la bossa la primera bola roja, o si la deixem fora.

Al primer cas diem que és **amb reemplaçament** i al segon, **sense reemplaçament**.

Si la tornem a ficar, la probabilitat de traure bola roja tornarà a ser $3/5$, i la probabilitat de traure dues boles roges és $3/5 \cdot 3/5 = 9/25$. La probabilitat d'aquesta segona bola *no depèn* del que ja hagem tret, i en aquest cas la probabilitat s'obté multiplicant.

Si els successos A i B són independents: $p(A \text{ i } B) = p(A) \cdot p(B)$.



Però si la deixem fora, ara a la bossa només hi ha 4 boles i d'elles només queden 2 boles roges, per tant la probabilitat que aqueixa segona bola siga roja és $2/4$, i està condicionada pel que abans hagem tret. S'escriu: $p(\text{Roja}/\text{Roja})$ i es llig "*probabilitat de roja condicionat a haver tret roja*". La probabilitat de traure dues boles roges és ara: $3/5 \cdot 2/4 = 6/20 = 3/10$.

Observa el diagrama d'arbre i comprova que la probabilitat de traure primer una bola roja i després una bola negra (no roja) és $3/5 \cdot 2/4 = 3/10$ perquè

després de traure una bola roja a la bossa queden només 4 boles i d'elles 2 són negres. La probabilitat de traure primer una bola negra i després bola roja és $2/5 \cdot 3/4 = 6/20 = 3/10$, i la de traure dues boles negres és: $2/5 \cdot 1/4 = 2/20 = 1/10$. Però observa més coses.

Per exemple, $3/5 + 2/5 = 1$; $2/4 + 2/4 = 1$; $3/4 + 1/4 = 1$; $3/10 + 3/10 + 3/10 + 1/10 = 1$.

Els successos no són independents. El que ocorregui A, o no ocorregui A, afecta la probabilitat de B. Per això es diu que B està **condicionat** a A.

Si els successos A i B són **dependents** llavors:

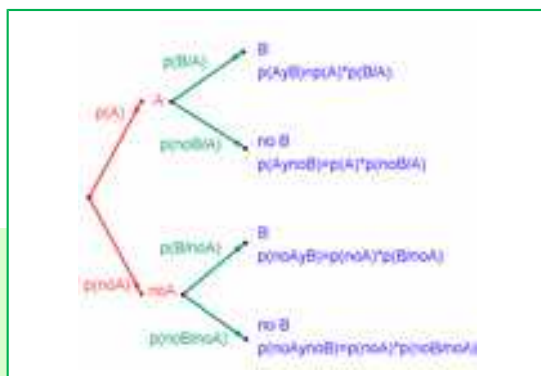
$$p(A \text{ i } B) = p(A) \cdot p(B/A)$$

Activitats resoltes



Traiem dues cartes d'una baralla de 40 cartes sense reemplaçament. Quina és la probabilitat de traure dos asos?

Si fóra amb reemplaçament la probabilitat seria $4/40 \cdot 4/40$, però en ser sense reemplaçament la probabilitat del segon as ve condicionada per que hagem tret un as prèviament. Ara a la baralla ja no



queden 40 cartes sinó 39, i no queden 4 asos sinó només 3, per tant la probabilitat és:

$$4/40 \cdot 3/39 = 1/130.$$

Observa que:

Si dos successos són **dependents** llavors: $p(B/A) \neq p(B)$.

Però si dos successos són **independents** llavors: $p(B/A) = p(B/\text{no}A) = p(B)$.

Activitats proposades

- 32.** Al teu quadern fes un diagrama en arbre semblant a l'anterior amb els successos A i B: A = *traure un as* a la primera extracció (noA = no traure'l), i B = *traure un as* a la segona extracció (no B = no traure'l). Quina és la probabilitat de traure *as* a la segona extracció condicionat a no haver-lo tret a la primera? I la de no traure *as* a la segona extracció condicionat a no haver-lo tret a la primera? Quina és la probabilitat de traure *dos asos*? I la de traure un sol *as*?
- 33.** Al diagrama d'arbre anterior indica qual és la probabilitat de "no ixen 2 asos" i la de "no ix cap as".
- 34.** A l'experiment "traure tres cartes seguides", quina és la probabilitat de traure *tres asos*? Primer amb reemplaçament, i després sense reemplaçament.
- 35.** En tirar dues vegades un dau calcula la probabilitat que isca un sis doble.
- 36.** En tirar dues vegades un dau calcula la probabilitat de traure almenys un 6. *Ajuda:* Potser et siga més fàcil calcular la probabilitat de *no traure cap 6*, i utilitzar el succés contrari.

Successos compatibles i incompatibles

Exemple:



Quina és la probabilitat de, en una baralla de 40 cartes, traure una copa o un or? Hi ha 10 copes i 10 ors, i cap carta és al mateix temps copa i or, per tant la probabilitat és 20/40.



Quina és la probabilitat de, en una baralla de 40 cartes, traure un as o un or? Hi ha 4 asos i hi ha 10 ors, però hi ha *l'as d'ors*, per tant les cartes que són o bé un as o bé un or són 13, per tant la probabilitat és 13/40.

Anomenem successos incompatibles als què, com a copa i or, no poden realitzar-se al mateix temps, i successos compatibles a què, com a as i or, poden realitzar-se al mateix temps.

Disseminem $p(A \text{ o } B)$ a la probabilitat del succés "es verifica A o bé es verifica B". Hem vist a l'exemple que si els successos són incompatibles la seua probabilitat és igual a la suma de les probabilitats.

$$P(A \text{ o } B) = p(A) + p(B), \text{ si } A \text{ i } B \text{ són incompatibles.}$$

Però si A i B si poden verificar-se al mateix temps caldrà restar aqueixos casos, aqueixes vegades en què es verifiquen A i B al mateix temps.

$$P(A \text{ o } B) = p(A) + p(B) - p(A \text{ y } B), \text{ si } A \text{ i } B \text{ són compatibles.}$$

Aquesta segona expressió és més general que la primera, ja que en el cas en què A i B són incompatibles llavors $p(A \text{ i } B) = 0$.

Activitats resoltes



Calcula la probabilitat dels successos següents: a) Traure un rei o una figura; b) No ix un rei o ix

un rei; c) Traure un bast o una figura.

- a) Hi ha 4 reis i hi ha $4 \cdot 4 = 16$ figures (as, sota, cavall i rei), però els quatre reis són figures, per tant $p(\text{Rey o Figura}) = 4/40 + 16/40 - 4/40 = 16/40 = 0,4$.
- b) Hi ha $40 - 4 = 36$ cartes que no són reis, i hi ha 4 reis, per tant $p(\text{no rei o rei}) = 36/40 + 4/40 = 1$. Aquesta conclusió és més general. Sempre:

$$p(\text{noA o A}) = 1,$$

ja que un succés i el seu contrari ja vam veure que verificaven que $p(A) + p(\text{noA}) = 1$.

- c) Hi ha 10 bastos i hi ha 12 figures, però hi ha 4 figures que són al mateix temps bastos (as, sota, cavall i rei), per tant $p(\text{Basto o Figura}) = 10/40 + 16/40 - 4/40 = 22/40 = 11/20$.

Resum:

Succés contrari: $p(X) + p(\text{noX}) = 1$.

Successos dependents: $p(A \text{ i } B) = p(A) \cdot p(B/A)$.

Successos compatibles: $P(A \text{ o } B) = p(A) + p(B) - p(A \text{ i } B)$.

Activitats proposades

37. Llancem dos daus que no estiguen trucats i anotem els nombres de la seua cara superior. Considerem el succés A que la suma de les dues cares siga 8, i el succés B que aqueixos nombres diferisquen en dues unitats. a) Comprova que $p(A) = 5/36$ (2 + 6; 3 + 5; 4 + 4; 5 + 3; 6 + 2) i que $p(B) = 8/36$ ((1,3), (2, 4), ...). b) Calcula les probabilitats de: $p(A \text{ i } B)$; $p(A \text{ o } B)$; $p(A \text{ i no } B)$; $p(\text{no } A \text{ i } B)$; $p(\text{no } A \text{ i no } B)$. c) Calcula $p(A/B)$; $p(A/\text{no } B)$; $p(\text{no } A/B)$.

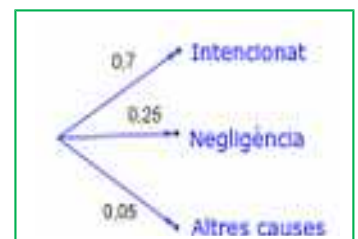
3.4. Experiències compostes: taules de contingència i diagrames d'arbre

Diagrames d'arbre

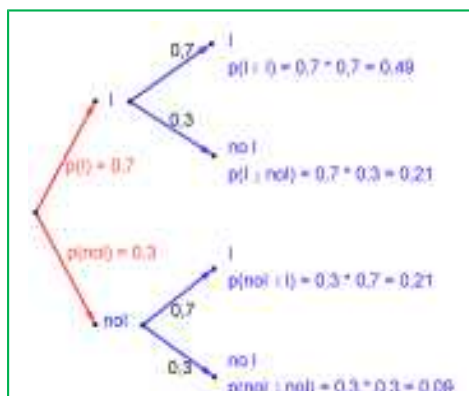
Exemple:



Es fa un estudi sobre els incendis i es comprova que en una determinada zona el 70 % dels incendis són intencionats, un 25% es deuen a negligències i 5 % a causes naturals com a rajos o a altres causes. Representa aquesta situació amb un diagrama d'arbre.



Activitats resoltes



Si considerem que la probabilitat que un incendi siga intencionat és 0,7, quina és la probabilitat que en considerar dos incendis, almenys un haja sigut intencionat?

Anomenem I al succés "ser intencionat" i noI al succés "no ser intencionat". Representem la situació en un diagrama d'arbre. Com el que un incendi siga intencionat és independent de com siga el segon, tenim que:

$$P(I \text{ i } I) = 0,7 \cdot 0,7 = 0,49$$

$$P(I \text{ i noI}) = 0,7 \cdot 0,3 = 0,21$$

ja que és la probabilitat que el primer incendi siga intencionat i el

segon no.

$$P(\text{noI i I}) = 0,3 \cdot 0,7 = 0,21$$

$$P(\text{noI i noI}) = 0,3 \cdot 0,3 = 0,09$$

La probabilitat que almenys un haja sigut intencionat la podem calcular sumant les probabilitats de (I i I), (I i noI), i (noI i I) que és $0,49 + 0,21 + 0,21 = 0,91$. Però més senzill és calcular la probabilitat del succés contrari $p(\text{noI i noI}) = 0,09$ i restar-la d'1:

$$p(\text{al menys un intencionat}) = 1 - 0,09 = 0,91.$$

Activitats proposades

- 38.** Dibuixa al teu quadern un diagrama en arbre per a tres incendis, i calcula la probabilitat que almenys un haja sigut intencionat sent $p(I) = 0,7$.
- 39.** En una aeronau s'han instal·lat tres dispositius de seguretat: A, B i C. Si no funciona A es posa B en funcionament, i si també B deixa de funcionar comença a funcionar C. Les probabilitats que funcione correctament cada dispositiu són: $p(A) = 0,95$; $p(B) = 0,97$ i $p(C) = 0,98$. a) Calcula la probabilitat que no funcionen els tres dispositius. b) Calcula la probabilitat que tot vaja bé.
- 40.** Una fàbrica de nines rebutja normalment el 0,5 % de la seua producció per errades degudes a l'atzar. Calcula la probabilitat que: a) En agafar dues nines a l'atzar haja que rebutjar ambdues. b) En agafar dues nines a l'atzar haja que rebutjar només una. c) En agafar dues nines a l'atzar no haja que rebutjar cap d) Verifiquem 4 nines, calcula la probabilitat de rebutjar únicament la tercera nina triada.
- 41.** Llancem una moneda fins que aparega dues vegades seguides del mateix costat. Calcula les probabilitats que: A) L'experiència acabe al segon llançament. B) Acabe al tercer llançament. C) Acabe al quart. D) Acabe com a màxim al quart llançament (és a dir, que acabe al segon o al tercer o al quart llançament).

Taules de contingència

Exemple:



S'han estudiat 500 malalts del fetge analitzant per un procediment nou si les lesions són benignes o malignes. Després se'ls va tornar a analitzar pel procediment usual determinant quins diagnòstics havien sigut correctes i quins incorrectes. Els valors obtinguts es representen a la taula:

| | Diagnòstic correcte | Diagnòstic incorrecte | Totals |
|---------------|---------------------|-----------------------|--------|
| Lesió maligna | 206 | 12 | 218 |
| Lesió benigna | 268 | 14 | 282 |
| Totals | 474 | 26 | 500 |

Determinem la taula de freqüències relatives:

| | Diagnòstic correcte (C) | Diagnòstic incorrecte (I) | Totals |
|-------------------|-------------------------|---------------------------|--------|
| Lesió maligna (M) | 0,412 | 0,024 | 0,436 |
| Lesió benigna (B) | 0,536 | 0,028 | 0,564 |
| Totals | 0,948 | 0,052 | 1 |

Activitats resoltes



Imagina que aquestes freqüències relatives pogueren prendre's com a probabilitats. Interpretem llavors el significat de cada un d'aquests valors.

0,412 seria la probabilitat que el diagnòstic de lesió maligna fóra correcte: $p(M \text{ i } C)$.

$0,024 = p(M \text{ i } I)$; $0,536 = p(B \text{ i } C)$; $0,028 = p(B \text{ i } I)$.

$0,436$? El nombre de lesions malignes és 218, per tant $0,436 = p(M)$.

De la mateixa manera: $0,564 = p(B)$; $0,948 = p(C)$; $0,052 = p(I)$.

Observa que $p(M) + p(B) = 1$ i que $p(C) + p(I) = 1$. Són successos contraris.



Són dependents o independents els successos M i C ?

Recorda que $p(M \text{ i } C) = p(M) \cdot p(C/M)$, per tant: $0,412 = 0,436 \cdot p(C/M)$, d'on $p(C/M) = 0,412/0,436 = 0,945$ que és diferent de $0,948$ que és la probabilitat de C . Es pot afirmar que M i C són dependents ja que $p(C/M) \neq p(C)$.

En general es denomina **taula de contingències** a:

| | A | No A | |
|------|------------------------------|---------------------------------------|-----------------|
| B | $P(A \text{ i } B)$ | $P(\text{no}A \text{ i } B)$ | $P(B)$ |
| No B | $P(A \text{ i } \text{no}B)$ | $P(\text{no}A \text{ i } \text{no}B)$ | $P(\text{no}B)$ |
| | $P(A)$ | $P(\text{no}A)$ | 1 |

En una taula de contingències figuren totes les probabilitats o contingències dels successos compostos.

Observa que, com sabem per la probabilitat del succés contrari:

$$p(A) + p(\text{no}A) = 1 \text{ i } p(B) + p(\text{no}B) = 1.$$

Observa també que:

$$p(A) = p(A \text{ i } B) + p(A \text{ i } \text{no}B), \text{ de la mateixa manera que } p(B) = p(A \text{ i } B) + p(\text{no}A \text{ i } B)$$

ja que s'obtenen sumant respectivament la primera columna i la primera fila.

També:

$$p(\text{no}A) = p(\text{no}A \text{ i } B) + p(\text{no}A \text{ i } \text{no}B) \text{ i } p(\text{no}B) = p(A \text{ i } \text{no}B) + p(\text{no}A \text{ i } \text{no}B).$$

Activitats proposades

42. S'ha fet un estudi estadístic sobre accidents de tràfic i s'han determinat les següents probabilitats reflectides a la taula de contingència:

| | Accident en carretera (C) | Accident en zona urbana (U) | Totals |
|--|---------------------------|-----------------------------|--------|
| Accident amb víctimes (V) | 0,27 | | 0,56 |
| Accident amb només danys materials (M) | | | |
| Totals | 0,58 | | 1 |

- Còpia la taula al teu quadern i completa-la.
- Determina les probabilitats següents: $p(V \text{ i } C)$; $p(V \text{ i } U)$; $p(M \text{ i } C)$; $p(M \text{ i } U)$; $p(V)$; $p(M)$; $p(C)$ i $p(U)$.
- Calcula $p(U/V)$; $p(C/V)$; $p(V/U)$; $p(V/C)$. Són dependents o independents els successos: accident

amb víctimes i accident en carretera?

43. Inventa una taula de contingència considerant que els accidents puguen ser de carretera (C) o urbans (U), però que ara els classifiquem en lleus (L), greus (G) o mortals (M). *Observa que* el més important per a confeccionar la taula és que els successos siguen incompatibles dos a dos.

Diagrames d'arbre i taules de contingència

Els diagrames d'arbre i les taules de contingència estan relacionats. Donat un arbre pots obtenir la taula de contingència, i viceversa. Té interès aquesta relació perquè amb les dades del problema a vegades és més senzill construir un d'ells i donar la solució passant a l'altre.

Activitats resoltes

✚ Donada la taula de contingència, obtindre el diagrama d'arbre que comença amb A i noA.

| | A | No A | |
|------|-----------|-----------|-----|
| B | 2/9 | 5/9 | 7/9 |
| No B | 1/9 | 1/9 | 2/9 |
| | 3/9 = 1/3 | 6/9 = 2/3 | 1 |

Coneixem la $p(A) = 3/9 = 1/3$, $p(\text{noA}) = 6/9 = 2/3$, $p(B) = 7/9$ i $p(\text{noB}) = 2/9$.

També coneixem $p(A \text{ i } B) = 2/9$; $p(A \text{ i noB}) = 1/9$; $p(\text{noA i } B) = 5/9$ i $p(\text{noA i noB}) = 1/9$.

Ens falta conèixer $p(B/A)$ que podem obtenir dividint $p(A \text{ i } B)$ entre $p(A)$:

$$p(B/A) = p(A \text{ i } B)/p(A) = 2/9 : 3/9 = 2/3.$$

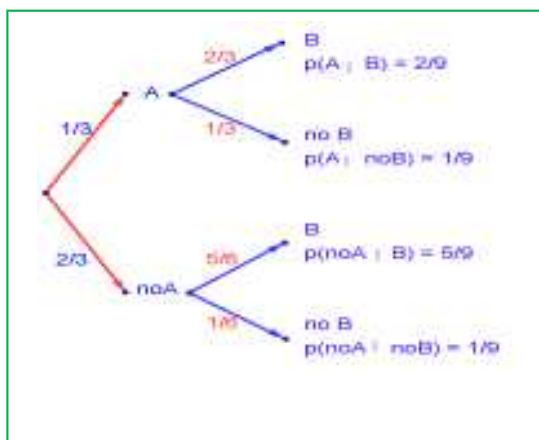
De la mateixa manera calculem:

$$p(\text{noB}/A) = p(A \text{ i noB})/p(A) = 1/9 : 3/9 = 1/3$$

$$p(B/\text{noA}) = p(\text{noA i } B)/p(\text{noA}) = 5/9 : 6/9 = 5/6$$

$$p(\text{noB}/\text{noA}) = p(\text{noA i noB})/p(\text{noA}) = 1/9 : 6/9 = 1/6.$$

L'arbre és :



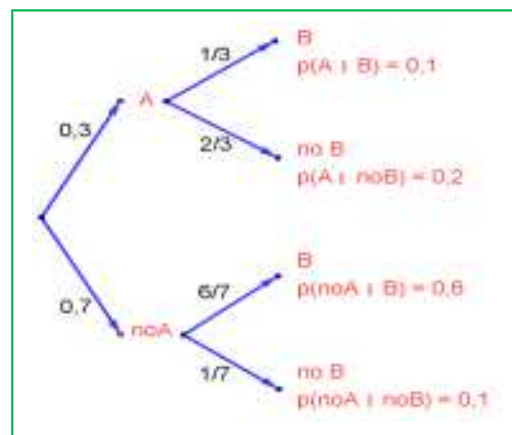
Activitats resoltes



Recíprocament, donat el diagrama d'arbre obtindre el diagrama de contingència:

Ara coneixem $p(A) = 0,3$ i $p(\text{no}A) = 0,7$. A més coneixem $p(B/A) = 1/3$; $p(B/\text{no}A) = 6/7$; $p(\text{no}B/A) = 2/3$ i $p(\text{no}B/\text{no}A) = 1/7$.

Calculem, multiplicant: $p(A \text{ i } B) = 0,3 \cdot (1/3) = 0,1$; $p(A \text{ i } \text{no}B) = 0,3 \cdot (2/3) = 0,2$; $p(\text{no}A \text{ i } B) = 0,7 \cdot (6/7) = 0,6$ i $p(\text{no}A \text{ i } \text{no}B) = 0,7 \cdot (1/7) = 0,1$ que posem també a l'arbre.



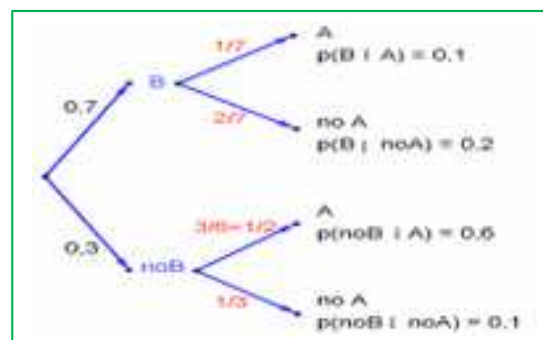
Omplim amb aquestes dades, una taula de contingència:

| | A | No A | |
|------|-----|------|---|
| B | 0,1 | 0,6 | |
| No B | 0,2 | 0,1 | |
| | 0,3 | 0,7 | 1 |

Calculem, sumant, les caselles que ens falten, $p(B) = 0,1 + 0,6 = 0,7$ i $p(\text{no}B) = 0,2 + 0,1 = 0,3$.

| | A | No A | |
|------|-----|------|-----|
| B | 0,1 | 0,6 | 0,7 |
| No B | 0,2 | 0,1 | 0,3 |
| | 0,3 | 0,7 | 1 |

Pot ser molt interessant passar d'un diagrama d'arbre a la taula de contingència i d'aquesta, a l'altre diagrama d'arbre, amb el que podem conèixer $p(A/B) = 0,1/0,7 = 1/7$; $p(\text{no}A/B) = 0,2/0,7 = 2/7$; $p(A/\text{no}B) = 0,3/0,6 = 3/6 = 1/2$ i $p(\text{no}A/\text{no}B) = 0,1/0,3 = 1/3$.

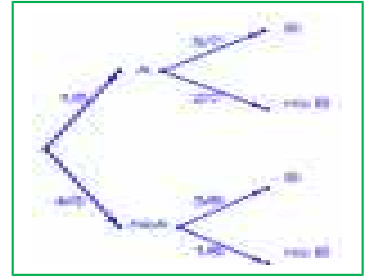


Activitats proposades

44. Donada la taula de contingència, construeix dos diagrames d'arbre.

| | A | No A | |
|------|------|------|-----|
| B | 0,4 | 0,2 | 0,6 |
| No B | 0,15 | 0,25 | 0,4 |
| | 0,55 | 0,45 | 1 |

45. Donat el diagrama d'arbre, construeix la taula de contingència, i després l'altre diagrama d'arbre.
46. Tenim dues urnes, A i B. La primera amb 8 boles blanques i 2 boles negres. La segona amb 4 boles blanques i 6 boles negres. Es trau una bola a l'atzar, d'una de les dues urnes, també a l'atzar i resulta ser negra. Quina és la probabilitat de què procedisca de l'urna A?
47. S'està estudiant un tractament amb un nou medicament, per al que se seleccionen 100 malalts. A 60 se'ls tracta amb el medicament i a 40 amb un placebo. Els valors obtinguts es representen a la taula adjunta



| | Medicament (M) | Placebo (no M) | |
|------------------|----------------|----------------|-----|
| Curats (C) | 50 | 30 | 80 |
| No curats (no C) | 10 | 10 | 20 |
| | 60 | 40 | 100 |

S'utilitzen aqueixos valors per a assignar probabilitats. Calcula:

- La probabilitat que un malalt curat haja sigut tractat amb el medicament. *Ajuda:* $p(M/C)$
- La probabilitat que un malalt curat haja sigut tractat amb el placebo. *Ajuda:* $p(\text{no}M/C)$.

CURIOSITATS I REVISTA

Estadística

El nom d'Estadística prové del s. XIX, no obstant això ja s'utilitzaven representacions gràfiques i altres mesures en pells, roques, pals de fusta i parets de coves per a controlar el nombre de persones, animals o certes mercaderies des de la Prehistòria. Els babilonis usaven ja envasos d'argila per a recopilar dades sobre la producció agrícola. Els egipcis analitzaven les dades de la població i la renda del país molt abans de construir les piràmides. Els antics grecs realitzaven censos la informació dels quals s'utilitzava cap a 600 aC.

El inicio de la Teoría de la Probabilidad, como sabes, fueron los juegos de azar.

Cavaller de la Meré

Al *Cavaller de la Meré* li agradava jugar i era un gran jugador, per això sabia que era favorable apostar, en tirar un dau "traure almenys un 6 en 4 tirades d'un dau" i que no ho era en tirar dos daus el "traure almenys un 6 doble en 24 jugades".

Es veu que havia jugat molt per a saber que les freqüències relatives li deien que el primer succés tenia una probabilitat superior a 0,5, i el segon la tenia inferior. Però no ho comprenia. No era matemàtic i sols sabia la regla de tres. Açò no és una proporcionalitat! Va dir $6 : 4 = 36 : 24$. Però les freqüències relatives li deien que no era així, per la qual cosa va escriure a Pascal per a que li solucionara el problema.

Tu ja saps suficient per a resoldre'l. Abans de continuar llegint, tracta de resoldre'l.

En lloc de calcular la probabilitat de *traure al menys un 6* en 4 llançaments, calcula la probabilitat de *no traure un 6*, que és el seu succés contrari, i és $\left(\frac{5}{6}\right)^4$. Per tant la probabilitat de *traure al menys un 6* en 4 tirades és:

$$1 - \left(\frac{5}{6}\right)^4 = 0,5177 > 0,5.$$

Calculem de la mateixa manera la probabilitat de *traure al menys un sis doble* en llançar dos daus 24 vegades, calculant la del seu succés contrari, la de *no traure cap sis doble*: $\left(\frac{35}{36}\right)^{24}$, per tant *traure al menys un 6 doble* és:

$$1 - \left(\frac{35}{36}\right)^{24} = 0,4914 < 0,5.$$

Quantes vegades va haver de jugar el Cavaller de la Meré per a donar-se compte de eixa xicoteta diferència en les probabilitats!

Si vols saber més, has de buscar:

<http://www.misclaneamatemati.ca.org/Misc34/caballero.pdf>
<http://www.misclaneamatemati.ca.org/Misc34/caballero.pdf>

Galileu,

Al segle XVI va plantejar el problema següent: En tirar tres daus, per què és més probable obtenir que la suma de les cares superiors siga 10, que siga 9?

Continuava la reflexió amb les possibles descomposicions en aqueixes sumes:

$$9 = 3 + 3 + 3 \quad 10 = 4 + 3 + 3$$

$$9 = 4 + 3 + 2 \quad 10 = 4 + 4 + 2$$

$$9 = 4 + 4 + 1 \quad 10 = 5 + 3 + 2$$

$$9 = 5 + 2 + 2 \quad 10 = 5 + 4 + 1$$

$$9 = 5 + 3 + 1 \quad 10 = 6 + 2 + 2$$

$$9 = 6 + 2 + 2 \quad 10 = 6 + 3 + 1$$

En ambdós casos hi ha 6 descomposicions possibles, no obstant això, tirant moltes vegades els 3 daus comprovava que és més probable traure un 10.

Si fas un diagrama en arbre comprovaràs que totes aqueixes descomposicions no són igualment probables.

Per exemple: (3, 3, 3) té una probabilitat d'1/216,

mentres que la suma 6 + 2 + 2, pot eixir amb tres successos (6, 2, 2), (2, 6, 2) i (2, 2, 6), per tant la seua probabilitat és 3/216.

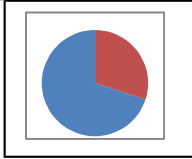

Calcula les probabilitats de cada una de les sumes i la de traure 10 i de traure 9.

La ruleta

William Jagers va arribar a Montecarlo amb uns pocs francs a la butxaca i, durant un mes va anotar els nombres que eixien en cada ruleta, i en quatre dies va guanyar dos milions quatre-cents mil francs. *Jagers* va aconseguir trencar la banca a *Montecarlo* analitzant les freqüències relatives de cada nombre de la ruleta i observant que s'havia desgastat alguna cosa del mecanisme d'una d'elles, amb la qual cosa tots els valors no tenien la mateixa probabilitat. Va apostar als nombres més probables i va guanyar.



RESUM

| | | Exemples | | | | | | | | | | | | |
|--|---|--|------------------------|--------------|--------------|------------------------|---|----|-----|----|---|----|-----|----|
| Població i mostra | Població: Tot el conjunt d'individus sobre el qual es fa l'estudi. Mostra: Una part d'aqueixa població. | Per a conèixer la intenció de vot, la població és tot el país, i se selecciona una mostra | | | | | | | | | | | | |
| Freqüència absoluta, relativa i acumulada | Freqüència absoluta: Nombre de vegades que s'ha obtingut aqueix resultat. Freqüència relativa: S'obté dividint la freqüència absoluta pel nombre total. Freqüència acumulada: S'obté sumant les freqüències anteriors. | <table border="1"> <thead> <tr> <th></th> <th>Fr. Absoluta</th> <th>Fr. Relativa</th> <th>Fr. Acumulada Absoluta</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>28</td> <td>0,7</td> <td>28</td> </tr> <tr> <td>B</td> <td>12</td> <td>0,3</td> <td>40</td> </tr> </tbody> </table> | | Fr. Absoluta | Fr. Relativa | Fr. Acumulada Absoluta | A | 28 | 0,7 | 28 | B | 12 | 0,3 | 40 |
| | Fr. Absoluta | Fr. Relativa | Fr. Acumulada Absoluta | | | | | | | | | | | |
| A | 28 | 0,7 | 28 | | | | | | | | | | | |
| B | 12 | 0,3 | 40 | | | | | | | | | | | |
| Gràfics estadístics | Diagrama de barres Diagrama de línies Diagrama de sectors |   | | | | | | | | | | | | |
| Mitja | Mitja = $m = (x_1 + x_2 + \dots + x_n)/n$ | Amb: 8, 4, 6, 10 i 10. Mitja = $38/5 = 7,6$ | | | | | | | | | | | | |
| Moda | És el valor més freqüent | 10 | | | | | | | | | | | | |
| Mitjana | Deixa per davall la meitat | $4 < 6 < 8 < 10 = 10$. Me = 8. | | | | | | | | | | | | |
| Variància i Desviació típica | $\text{Variància} = \frac{\sum_{i=1}^n (x_i - m)^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - m^2 \cdot s = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - m^2}$ | Variància = 5,4. s = 2,33. | | | | | | | | | | | | |
| Quartils | Q1 deixa per davall la quarta part. Q3 deixa per davall les tres quartes parts. Interval interquartilic = $Q3 - Q1$. | Q1 = 6; Q3 = 10; Interval interquartilic = $Q3 - Q1 = 4$. | | | | | | | | | | | | |
| Histograma | L'àrea de cada rectangle és proporcional a la freqüència. | | | | | | | | | | | | | |
| Correlació | El coeficient de correlació , ρ , mesura la relació entre dos variables. És un nombre entre -1 i 1. | $\rho = 1 \rightarrow$ correlació perfecta positiva $\rho = -1 \rightarrow$ correlació perfecta negativa $\rho = 0 \rightarrow$ correlació nul·la $\rho \in (0, 1) \rightarrow$ correlació positiva $\rho \in (-1, 0) \rightarrow$ correlació negativa | | | | | | | | | | | | |
| Succés | En realitzar un experiment aleatori hi ha diversos possibles resultats o successos possibles . Un succés és un subconjunt del conjunt de possibles resultats. | Tirem un dau. Possibles resultats = {1, 2, 3, 4, 5, 6} Succés obtenir <i>múltiple de 3</i> = {3, 6} | | | | | | | | | | | | |
| Probabilitat | Limit a què tendeixen les freqüències relatives. Si els successos elementals són equiprobables llavors: $p = \text{casos favorables} / \text{casos possibles}$. | $P(5) = 1/6$. $P(\text{traure múltiple de 3}) = 2/6$ | | | | | | | | | | | | |
| Assignació de probabilitats | Succés contrari: $p(X) + p(\text{no}X) = 1$. Successos dependents: $p(A \text{ i } B) = p(A) \cdot p(B/A)$. Successos compatibles: $P(A \text{ o } B) = p(A) + p(B) - p(A \text{ y } B)$. | $P(\text{no } 5) = 1 - 1/6 = 5/6$. $P(5 \text{ o } \text{múl. } 3) = 1/6 + 2/6 = 3/6$ $P(\text{traure primer un 5 i després múltiple de } 3) = 1/6 \cdot 2/6 = 2/36$ | | | | | | | | | | | | |

EXERCICIS I PROBLEMES.

Estadística

1. En una classe es mira el color dels ulls de cada alumne i alumna i s'obté el següent:

N := negre; A := blau i V := verd.

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.

Fes una taula de freqüències absolutes, representa els valors en un diagrama de sectors i calcula la moda.

2. Les notes d'un conjunt d'alumnes de 4t són:

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.

- Fes una taula de freqüències absolutes, freqüències relatives, freqüències acumulades absolutes i freqüències relatives acumulades.
- Calcula la mitja, la mitjana i la moda.
- Calcula la desviació típica i els quartils.

3. S'ha preguntat a 40 alumnes pel nombre de germans que tenia, i s'ha obtingut

| Nombre de germans | 0 | 1 | 2 | 3 | 4 | 5 | 6 o més |
|-------------------|---|----|---|---|---|---|---------|
| Nombre de vegades | 5 | 15 | 7 | 6 | 4 | 2 | 1 |

- Representa un diagrama de barres de freqüències absolutes i un diagrama de línies de freqüències relatives.
- Calcula la mitja, la mitjana i la moda.

4. S'han llançat quatre monedes 100 vegades i anotat el nombre de vegades que ha eixit cara. Els resultats estan reflectits a la taula següent:

| Nombre de cares | 0 | 1 | 2 | 3 | 4 |
|-------------------|---|----|----|----|---|
| Nombre de vegades | 7 | 25 | 36 | 26 | 6 |

- Escriu al teu quadern una taula de freqüències absolutes, freqüències relatives, freqüències acumulades absolutes i freqüències relatives acumulades.
- Representa un diagrama de barres de freqüències absolutes acumulades, un diagrama de línies de freqüències relatives i un diagrama de sectors de freqüències absolutes.
- Calcula la mitja i la desviació típica.
- Calcula la mitjana i els quartils.

5. Per a conèixer la distribució d'un cert país de les persones segons la seua edat s'ha arreplegat una mostra de deu mil persones i els valors obtinguts vénen reflectits a la taula següent:

| Edats | [0, 5) | [5, 10) | [10, 15) | [15, 25) | [25, 35) | [35, 45) | [45, 55) | [55, 65) | [65,100) |
|--------------------|--------|---------|----------|----------|----------|----------|----------|----------|----------|
| Nombre de persones | 900 | 1000 | 900 | 1500 | 1300 | 1200 | 1300 | 900 | 1000 |

- Utilitza les marques de classe i escriu al teu quadern una taula de freqüències absolutes, freqüències relatives, freqüències acumulades absolutes i freqüències relatives acumulades.

- b) Representa un histograma de freqüències absolutes. *Atenció:* Els intervals no són tots iguals. Recorda: *L'àrea* dels rectangles ha de ser proporcional a les freqüències.
- c) Calcula la mitja i la desviació típica.
- d) Calcula la mitjana i els quartils de forma gràfica usant un histograma de freqüències absolutes acumulades.
6. Amb les dades del problema anterior calcula l'interval [mitja – desviació típica, mitja + desviació típica]. Quantes persones estan al dit interval? Quin percentatge? Calcula també l'interval [mitja – 2*desviació típica, mitja + 2*desviació típica] i [mitja – 3*desviació típica, mitja + 3*desviació típica]. Si la distribució fora normal hi hauria al primer interval un 68 % de la mostra, al segon un 95 % i al tercer més d'un 99'7 %. Compara els teus resultats amb aquests.
7. Amb les mateixes dades calcula l'interval interquartílic, i indica quantes persones estan al dit interval i quin percentatge.
8. Una companyia d'assegurances desitja establir una pòlissa d'accidents. Per a això, selecciona a l'atzar a 200 propietaris i els pregunta quants euros han gastat en reparacions de l'automòbil. S'han agrupat en intervals els valors de la variable obtinguts:

| Euros | [0, 100) | [100, 200) | [200, 400) | [400, 600) | [600, 800) | [800, 3000) |
|--------------------|----------|------------|------------|------------|------------|-------------|
| Nombre de persones | 40 | 30 | 20 | 40 | 50 | 20 |

- a) Calcula les marques de classe i escriu al teu quadern una taula de freqüències absolutes, freqüències relatives, freqüències acumulades absolutes i freqüències relatives acumulades.
- b) Representa un histograma de freqüències relatives. *Atenció:* Els intervals no són tots iguals.
- c) Calcula la mitja i la desviació típica.
- d) Calcula la mitjana i els quartils de forma gràfica usant un histograma de freqüències absolutes acumulades.
9. Dos fabricants de bateries de cotxes ofereixen el seu producte a una fàbrica al mateix preu. La fàbrica vol triar la millor. Per a això tria una mostra de 60 bateries de cada marca i obté de cada una els mesos que ha funcionat sense espatllar-se. Obté la taula següent:

| Vida de la bateria en mesos | 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 |

Quina marca creus que triarà?

Per a prendre la decisió, calcula la mitja, la moda i la mitjana per a cada marca.

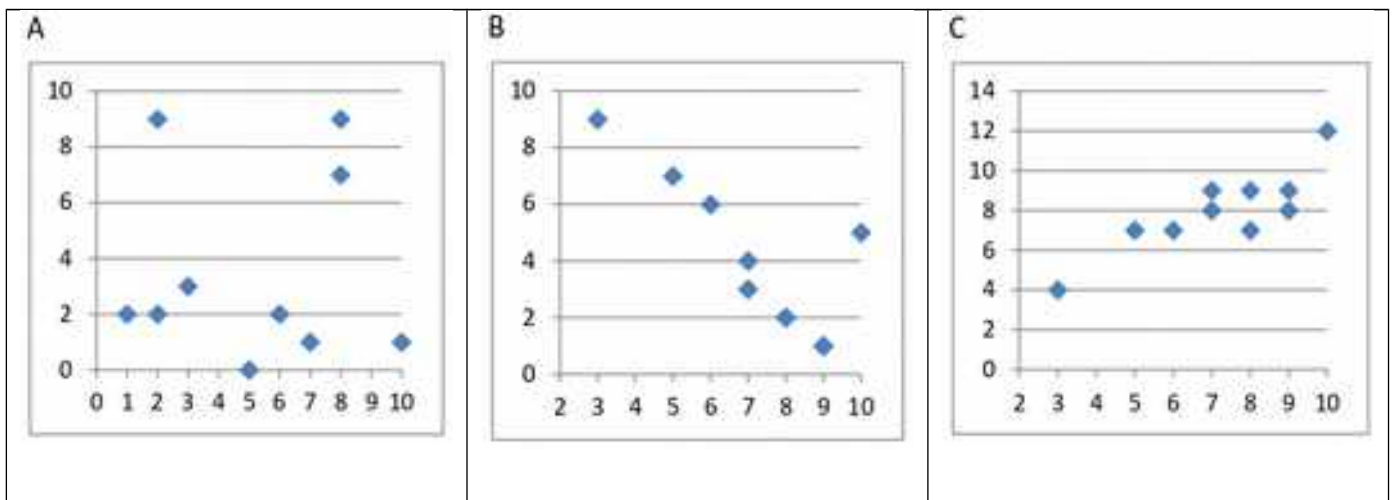
Si encara no et decideixes, calcula el recorregut, la desviació típica, l'interval $[m - s, m + s]$ i l'interval interquartílic.

10. Fes un treball. Passa una enquesta als teus companys i companyes de classe. Fes-los una pregunta amb dades numèriques, com per exemple, quant mesura la seua mà, quin nombre de sabata calcen, el nombre de llibres que lligen en un mes, el nombre d'hores que veuen la televisió a la setmana, diners que gasten al mes en comprar música... Representa les dades obtingudes en una taula. I fes un estudi complet. Pots utilitzar l'ordinador:

- Escriu al teu quadern una taula de freqüències absolutes, freqüències relatives, freqüències acumulades absolutes i freqüències relatives acumulades.
- Dibuixa un diagrama de barres, un diagrama de línies i un diagrama de sectors.
- Calcula la mitja, la mitjana i la moda
- Calcula la variància i la desviació típica
- Calcula els quartils i l'interval interquartílic.
- Reflexiona sobre els resultats i escriu un informe.

Coeficient de correlació

11. Andrés ha calculat els coeficients de correlació dels tres núvols de punts adjunts, i ha obtingut: $-0,8$, $0,85$ i $0,03$, però ara no recorda quin és de cada una. Pots ajudar a decidir quin coeficient correspon amb cada núvol?



Probabilitat

- En un col·legi se selecciona un grup de 200 estudiants dels quals tots estudien francès o anglès. D'ells 150 estudien anglès i 70 estudien francès. Quants estudien francès. i anglès? En un altre centre escolar s'estudien diversos idiomes: francès., anglès, alemany, italià. Se seleccionen també 200 estudiants dels quals, 150 estudien anglès, 70 francès. i 40 ambdós idiomes, quants estudiants d'aqueix centre no estudien ni francès. ni anglès?
- Llancem un dau. Calcula la probabilitat de: a) Traure un nombre imparell. b) No traure un 3. c) Traure un nombre més gran que 3. d) Traure un nombre més gran que 3 i que siga imparell. e) Traure un nombre més gran que 3 o bé que siga imparell.
- En una classe hi ha 24 xics i 14 xiques. La meitat de les xiques i la tercera part dels xics tenen els ulls blaus. Es tria un estudiant a l'atzar. A) Calcula la probabilitat que siga xic i tinga els ulls blaus. B) Calcula la probabilitat que siga xic o tinga els ulls blaus.
- Antoni, Joan i Jordi tenen una prova de natació. Antoni i Joan tenen la mateixa probabilitat de guanyar, i doble a la probabilitat de Jordi. Calcula la probabilitat que guanye Joan o Jordi.
- Llancem dues monedes distintes, una de 50 cèntims i una altra d'un euro. Calcula la probabilitat que: A) En la moneda d'un euro isca cara. B) Isca una cara. C) Isca almenys una cara. D) No isca cap cara. E) Isca una cara i una creu.
- Llancem tres monedes. Calcula les probabilitats de: A) No isca cap cara. B) Isca almenys una cara. C) Isquen dues cares i una creu.
- Llancem dos daus i anotem els valors de les cares superiors. Calcula les probabilitats de que la suma

sigués 1, sigués 2, sigués 3, sigués 12.

19. Què és més probable en tirar tres daus, que la suma de les seues cares superiors sigui 9 o sigui 10? Escribe el succés "sigués 9" i el succés "sigués 10" i calcula les probabilitats dels seus successos elementals. Saps ja més que *Galileu*!
20. Llancem al mateix temps una moneda i un dau. Anomena A al succés "Isca cara i un nombre parell". B al succés "Isca creu i un nombre primer" i C al succés "isca un nombre primer". Calcula les probabilitats de A, B i C. Com són aquests successos? Indica quins d'ells són compatibles i quins són incompatibles.
21. Llancem una moneda 50 vegades, què és més probable, obtindre 50 cares seguides o obtindre en les primeres 25 tirades cara i en les 25 següents creu? Raona la resposta.
22. Una moneda està trucada. La probabilitat d'obtindre cara és doble que la d'obtindre creu. Calcula les probabilitats dels successos obtindre cara i d'obtindre creu en tirar la moneda.
23. Tres xics i dues xiques juguen un torneig d'escacs. Tots els xics tenen idèntica probabilitat de guanyar, i totes les xiques, també. Però la probabilitat de guanyar una xica és doble de la de guanyar un xic. Calcula la probabilitat que un xic guanye el torneig.
24. Set parelles de nòvios estan en una habitació. Se seleccionen dues persones a l'atzar. Calcula la probabilitat de: a) Siguen un xic i una xica. b) Siguen una parella de nòvios. Ara es trien 4 persones a l'atzar. Calcula la probabilitat de: c) Hi haja almenys una parella de nòvios. d) No hi haja cap parella de nòvios.
25. Tenim un dau trucat de manera que els nombres imparells tenen una probabilitat doble a la dels nombres parells. Calcula les probabilitats de: A) Isca un número imparell. B) Isca un nombre primer. C) Isca un nombre primer imparell. D) Isca un nombre que sigui primer o sigui imparell.
26. En un grup de 12 amigues hi ha 3 rosses. Es trien dues xiques a l'atzar. Calcula la probabilitat que: A) Ambdues siguin rosses. B) Almenys una sigui rossa. C) Cap sigui rossa. D) Una sigui rossa i l'altra no.
27. Llancem dos daus i anotem els valors de les cares superiors. Calcula les probabilitats que: A) Els nombres obtinguts siguin iguals. B) Els nombres obtinguts diferisquen en 3 unitats. C) Els nombres obtinguts siguin parells.
28. Llancem una moneda fins que isca cara. Calcula la probabilitat que: A) Isca cara abans del quart llançament. B) Isca cara després del huitè llançament.
29. Un lot de 20 articles té 2 defectuosos. Es trauen 4 a l'atzar, quina és la probabilitat que cap sigui defectuós?
30. Es llancen dos daus i la suma de les cares superiors és 7. Quina és la probabilitat que en un dels daus haja eixit un 3?
31. Es tenen 3 caixes, A, B i C. La caixa A té 10 boles de les quals 4 són negres. La caixa B té 6 boles amb una bola negra. La caixa C té 8 boles amb 3 negres. S'agafa una caixa a l'atzar i d'aqueixa caixa es trau una bola, també a l'atzar. Comprova que la probabilitat que la bola sigui negra és $113/360$.
32. Tenim una moneda trucada la probabilitat de de la qual obtindre cara és $3/5$ i la de creu és $2/5$. Si ix cara es tria a l'atzar un nombre de l'1 al 8, i si ix creu, es tria un nombre de l'1 al 6. Calcula la probabilitat que el nombre triat sigui imparell.
33. En un procés de fabricació de mòbils es detecta que el 2 % ixen defectuosos. S'utilitza un dispositiu per a detectar-los que resulta que detecta el 90 % dels mòbils defectuosos, però assenyala com defectuosos un 1 % que no ho són. A) Calcula la probabilitat que sigui correcte un mòbil que el dispositiu ha qualificat com defectuós. B) Calcula la probabilitat que sigui defectuós un mòbil que el dispositiu ha qualificat com correcte. *Ajuda:* Utilitza primer un diagrama en arbre i després una taula de contingència.

AUTOAVALUACIÓ

Amb les dades següents, 1, 5, 2, 8, 9, 4, 7, 7, 5, 7, calcula:

1. La mitja:

- a) 5 b) 5'5 c) 6 d) 7

2. La mitjana:

- a) 5 b) 5'5 c) 6 d) 7

3. La moda:

- a) 5 b) 5'5 c) 6 d) 7

4. La desviació típica:

- a) 2 b) 2,27 c) 2,46 d) 2,65

5. L'interval interquartílic

- a) 3 b) 2,75 c) 4 d) 2

6. En tirar dos daus, la probabilitat de traure almenys un 5 és:

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

7. En tirar 3 monedes, la probabilitat de traure exactament dues cares és:

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

8. En tirar 3 monedes, la probabilitat de traure almenys dues cares és:

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

9. Traiem una carta d'una baralla de 40 cartes, la probabilitat que siga un or o un múltiple de 2 és:

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

10. Indica quina de les afirmacions següents és **sempre** correcta:

- a) $P(A) + P(\text{no}A) = 1$
 b) $P(A \text{ i } B) = P(A) \cdot P(B)$
 c) $P(A \text{ o } B) = P(A) + P(B)$