Lliurament 4:
Estadística i probabilitat

Com podem predir els efectes del canvi climàtic?

Matemàtiques I

Josep Mulet Pol

Àmbit científic

IEDIB

https://iedib.net/

Índex

1. Estadística descriptiva univariant

L'estadística parteix d'una gran quantitat de dades i intenta extreure i resumir-ne la informació mésrellevant. Per exemple,al butlletíde notes a final de curs no apareixen totes les notes de les tasques del curs; tan sols un paràmetre(la mitjana) que resumeixl'evolució de l'alumne.

Font de la imatge. [https://matematicasiesoja.wordpress.com/estadistica/]

Els **paràmetres estadístics** són aquells valors que extreuen informació rellevantde la distribució de dades. Coneixem molt bé la mitjana aritmètica, però, n'existeixen molts d'altres paràmetres, cadascun amb el seu significat iutilitat. Podem agrupar els paràmetresestadístics en dues categories:

- Paràmetres de **posició** : De forma semblant a la mitjana, ens indiquen quina posició ocupen elsindividus dins del rang de valors de la distribució.

- Paràmetres de **dispersió** : Ens informa de com diferents (variabilitat) són les dades ambrespecte la mesura de posició.

Mesures de posició

Donada una variable estadística x_i amb cada valor repetit f_i vegades(freqüència absoluta), es defineixen els següents **paràmetres estadístics de posició**

- **Nombre de dades:** $N = \sum_i f_i$
- **Mitjana aritmètica:** $\bar{x} = \dfrac{\sum_i f_i x_i}{N}$
- **Moda:** El valor de x més freqüent.
- **Mediana:** Valor de x pel qual la freqüència acumulada assoleix el 50%.

El símbol $\sum_i$ significa la suma de tots els elements enumerats per l'índex $i = 1, 2, 3, \cdots$

Anem a veure un exemple que posa de manifest que la mitjana no dóna tota la informació.Considereu dos alumnes: l'Andreu que ha obtingut les notes 5, 4 i 6; i na Maria ha tret 2, 9 i 4.Fixeu-vos que en ambdós casos lamitjana és 5, però està clar que les notes de l'Andreu varien poc mentre que les de na Maria tenen unagran dispersió. En el primer cas, el rang=6-4=2 i, en el segon cas, rang=9-2=7. Tot i que el rang és unamesura de dispersió fàcil d'obtenir, generalment s'utilitza la desviació típica que passam a definir.

Mesures de dispersió

- **Rang:** La diferència entre els valors major i menor de x.
- **Variància:** $Var = \dfrac{\sum_i f_i x_i^2}{N} - \bar{x}^2$
- **Desviació típica:** $\sigma = \sqrt{Var}$
- **Coeficient de variació:** $CV = \dfrac{\sigma_x}{\bar{x}}$

La **mitjana** és el **centre de gravetat** de la distribució; si el diagrama de barresestigues construït amb barres de fusta, la mitjana és el valor de x pel qual les barres queden en equilibri.

En canvi, la **desviació típica** ens informa de la **dispersió** , és a dir, ens diu com d'allunyades estan les dadesrespecte de la mitjana.Podem pensar que una dada serà *normal* si es troba dins l'interval de x $(\bar{x}-\sigma, \bar{x}+\sigma)$.

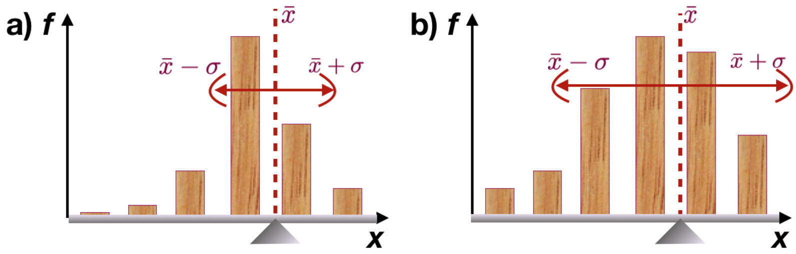


Figura 1: Significat de la mitjana i la desviació típica. Les dues distribucions tenen igual mitjana però a) té menor dispersió que b).

Tot seguit farem un exemple de com es realitzen els càlculs d'aquests paràmetres enel cas d'una variable quantitativa discreta.

Exemple 1

Hem demanat pel nombre d'assignatures suspeses durant la primeraavaluació a un grup de 15 alumnesi aquestes han estat les respostes:

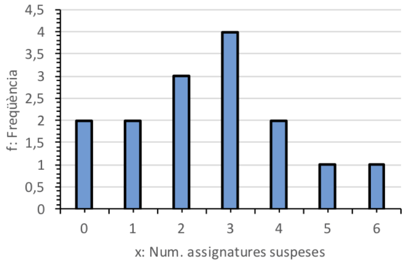
$$\begin{array}{cccccccc} 1 & 0 & 3 & 2 & 0 & 6 & 2 & 5 \\ 3 & 2 & 4 & 3 & 4 & 1 & 3 & \end{array}$$

a) Fes un recompte i dibuixa un diagrama de barres.
b) Calcula la mitjana i la desviació típica.

La variable estadística és x_i =N. d'assignatures suspeses ies tracta d'una variable quantitativa discreta. El rang de lavariable és de 0 a 6.
Començam fent un recompte quan freqüent (f_i) és cada valor (x_i) de la variable:

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

Amb aquesta taula podem dibuixar un diagrama de barres



Per calcular els paràmetres estadístics necessitamcalcular dues columnes més $f \cdot x$ i $f \cdot x^2$:

x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
0	2	0	0
1	2	2	2
2	3	6	12
3	4	12	36
4	2	8	32
5	1	5	25
6	1	6	36
$SUMES$	15	39	143

Número total de dades $N = 15$
La mitjana s'obté de $\bar{x} = \frac{39}{15} = 2.6$
La desviació típica $\sigma = \sqrt{\frac{143}{15} - 2.6^2} = 1.67$
El coeficient de variació és $CV = \frac{1.67}{2.6} = 0.64$, aproximadament un 64 %.

Exercicis

1. S'ha realitzat un estudi a 100 dones majors de 15 anys i se'ls hi ha demanatel nombre de fills que tenien. S'han presentat el resultats en forma de taula de freqüències:

N. fills (x_i)	0	1	2	3	4	5	6
N. dones (f_i)	13	20	25	20	11	7	4

a) Identifica la variable estadística i indica el seu tipus.

b) Representa un diagrama de barres.

c) Calcula la mitjana del nombre de fills.

d) Calcula la desviació típica del nombre de fills i el coeficient de variació.

2. Estadística bidimensional

Dependència funcional vs. correlació

En nombroses ocasions ens interessa estudiar simultàniament dos o més caràcters de les dades. Enparticular,volem saber quinarelació existeix entre aquestes variables. Aquest és l'objecte d'estudi de **l'estadísticabidimensional** .

Considerem dos exemples:

- x =kg de pomes, y =preu en €

x (kg)	0	1	2	3	4	5	6	7	8	9
y (€)	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5

- x =nota matemàtiques, y=nota física d'un grup de 10 alumnes.

x (mates)	3	4	5	6	7	7	7	8	9	10
y (física)	2	5	5	6	7	6	7	9	8	10

Existeixen dos tipus de relació entre variables $\{x_i, y_i\}$:

- **Dependència funcional** : Existeix una llei (una fórmula) que relaciona les dues variables $y_i = f(x_i)$. *Quan més quilos de pomes, major preu* $y = 0.5 \cdot x$.En aquest cas, el preu per quilogram és 0,5 €/kg.

- **Correlació estadística** : No existeix cap llei exacta sinó que vindrà donada per unatendència. Perexemple, *si un alumne és bo en matemàtiques, cal esperar que també ho sigui en física* , però, noexisteix capfórmula a partir de la qual es pugui predir nota que obtindrà l'alumne.

Així mateix, la correlació estadística la podem classificar segons el tipus (positiva-negativa) o segonsla intensitat(forta-feble):

- **Correlació positiva** : Quan x_i augmenta, també ho fa y_i. Per exemple, *les notesde l'examen de física i les notes de l'examen de matemàtiques.*

- **Correlació negativa** : Quan x_i augmenta, y_i disminueix. Per exemple, *a majordistància de la cistella de basquet, menor el nombre d'encerts.*

Cadascuna d'elles pot ésser correlació **forta** o correlació **feble** . Començarem l'estudimesurant de forma qualitativa aquesta correlació. A mesura que avancem en el tema,però, aprendrem com descriure-la de forma quantitativa.

Exemple 2

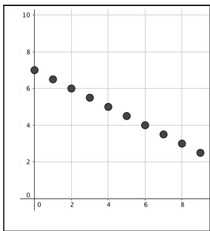
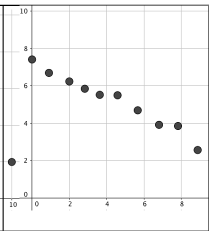
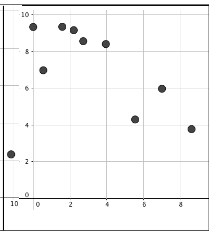
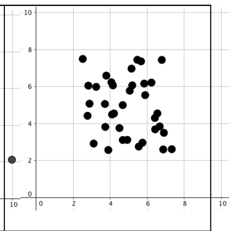
Per a cadascun dels casos següents analitza quin és el tipus de relacióentre les variables (funcional o correlació). En cas de correlació indica si éspositiva o negativa.

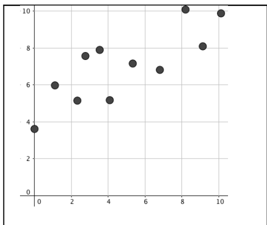
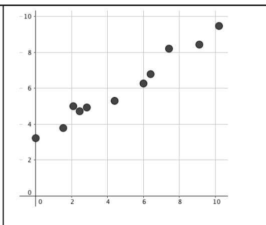
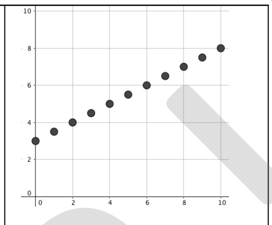
1. El radi d'una esfera – el costat d'aquesta
2. El nombre d'encerts d'un jugador de basquet – La distància a la cistella.
3. Les notes de l'examen de matemàtiques – Notes de l'examen de física.
4. La distància d'un trajecte en tren – El preu el bitllet.
5. El pes dels alumnes de 1r de batxillerat – La seva altura.
6. El nombre de membres de la família – El preu del rebut d'aigua mensual.

1. Funcional $D = 2R$, 2. Correlació negativa, 3. Correlació positiva, 4. Correlaciópositiva, 5. Correlació positiva, 6. Correlació positiva

2.1 Núvol de punts

Un núvol de punts s'obté de dibuixar els punts corresponents als parells (x_i, y_i).Aquest núvol ens permet identificar visualment el tipus de relació entre les variables:

			
Funcional negativa	Correlació negativa forta	Correlació negativa feble	Sense correlació

		
Correlació positiva feble	Correlació positiva forta	Funcional positiva

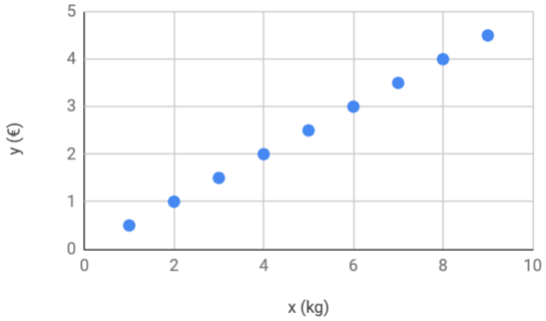
Com més concentrats estiguin els punts al voltant d'una línia recta diem quemajor és la correlació. Aquesta línia s'anomena **recta de regressió** . Siaquesta recta és creixent es diu que la correlació és positiva, mentre quesi és decreixent la correlació és negativa.

El **centre de gravetat** del núvol és a les mitjanes de les variables $G(\bar{x}, \bar{y})$. Es compleix que la recta de regressió passa semprepel centre de gravetat.

Exemple 3

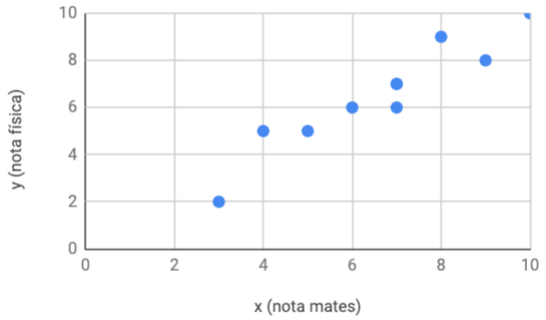
Representeu els núvols de punts per a les dades dels dos exemples de l'apartat anterior.
a) x =kg de pomes, y =preu en €
b) x =nota matemàtiques, y=nota física d'un grup de 10 alumnes.
A partir del gràfic, interpretau quin tipus de relació es dona.

a) x =kg de pomes, y =preu en €



Els punts estan exactament sobre una línia recta creixent,aleshores es tracta d'una relació funcional positiva.

b) x =nota matemàtiques, y=nota física d'un grup de 10 alumnes.



Els punts estan bastant concentrats al voltant d'una recta creixent,aleshores existeix correlació positiva forta.

Exercicis

2. S'ha fet un estudi sobre com afecta el nombre de cigarretsconsumits en un dia a l'índex de mortalitat. S'han obtingut aquests resultats:

Núm. cigarrets	3	6	8	20	25
Índex de mortalitat	0,2	0,4	0,5	1,2	1,7

a) Representau un núvol de punts.

b) Indicau el tipus de correlació.

3. En l'estudi estadístic bidimensional s'han recollit les següents dades:

x_i	0	4	6	2	4	8	0	2	6	8
y_i	1	1	1	3	3	3	5	5	5	5

a) Representau un núvol de punts.

b) Indicau el tipus de correlació.

2.2 Covariància. Coeficient de correlació

A l'apartat anterior, el núvol de punts ens proporciona una forma de classificar la correlació entre lesvariables $\{x_i, y_i\}$ de forma qualitativa.

Anomenam **covariància** al paràmetre estadístic que relaciona **quantitativament** la relació entre lesvariables:

$$\sigma_{xy} = \frac{\sum_i x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y}$$

on N és el nombre de punts (x_i, y_i) i $\bar{x}$, $\bar{y}$ les mitjanes de cada variable.

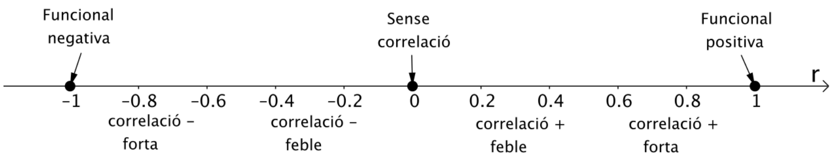
La covariància pot ésser negativa, zero o positiva segons els tipus de correlació que existeixi.La covariància té l'inconvenient, però, que depèn de l'escala de les dades que utilitzem, per la qual cosaes fa difícil tenir una idea de quan forta o feble és la relació.

Per poder quantificar millor la intensitat de la correlació empram el **coeficient de correlació lineal** definit mitjançant

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

σ_{xy}: covariància, σ_x: desviació típica de la variable x_i, σ_y: desviació típica de la variable y_i

Aquest coeficient té el mateix signe que la covariància i, per tant, ens indicarà si es tracta d'una correlaciópositiva o negativa. A més a més, compleix que els seus valors sempre es troben entre $-1 \leq r \leq 1$.



Vegem quin tipus de dependència existeix entre les variables x_i i y_i segons els valors de r.

- **Si** $r = -1$, tots els valors de la variable bidimensional (x_i, y_i) es troben situats sobre una recta decreixent;conseqüentment, satisfan l'equació d'una recta. Llavors x_i, y_i presenten una **dependència funcional negativa** .

- **Si** $-1 < r < 0$, la **correlació és negativa** i serà més forta com més proper a -1 sigui el valorde r i més feblecom més s'aproximi a zero.

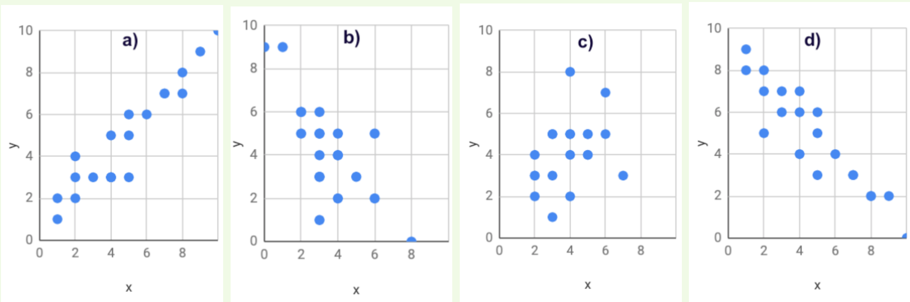
- **Si** $r = 0$, llavors no existeix cap tipus de relació entre les variables, diem que estan **descorrelacionades** .

- **Si** $0 < r < 1$, la **correlació és positiva** i serà més forta com més proper a 1 sigui el valorde r i més feble com més s'aproximi a zero.

- **Si** $r = 1$, tots els valors de la variable bidimensional (x_i, y_i) es troben situats sobre una recta creixent;conseqüentment, satisfan l'equació d'una recta. Llavors x_i, y_i presenten una **dependència funcional positiva** .

Exemple 4

Ordenau de menor a major segons el valor del coeficient de correlació lineal de les següents distribucionsbidimensionals. Assigna el valor correcte de r -0.92, -0.75, 0.34, 0.94 a cada apartat.



a) $r = 0.94$, b) $r = -0.75$, c) $r = 0.34$, d) $r = -0.92$.

De menor a major valor de r

$$\text{d)} < \text{b)} < \text{c)} < \text{a)}$$

Exemple 5

Calculau la covariància i el coeficient de correlació lineal dels dos exemples de l'apartat anterior.
a) x =kg de pomes, y =preu en €
b) x =nota matemàtiques, y=nota física d'un grup de 10 alumnes.

a) x =kg de pomes, y =preu en €

	x	y	x^2	y^2	$x \cdot y$
	0	0	0	0	0
	1	0.5	1	0.25	0.5
	2	1	4	1	2
	3	1.5	9	2.25	4.5
	4	2	16	4	8
	5	2.5	25	6.25	12.5
	6	3	36	9	18
	7	3.5	49	12.5	24.5
	8	4	64	16	32
	9	4.5	81	20.25	40.5
$Sumes$	45	22.5	285	71.25	142.5

Número de parelles $N = 10$
Mitjana x: $\bar{x} = \dfrac{45}{10} = 4.5$
Variància x: $Var_x = \dfrac{285}{10} - 4.5^2 = 8.25$
Desviació típica x: $\sigma_x = \sqrt{Var_x} = 2.87$
Mitjana y: $\bar{y} = \dfrac{22.5}{10} = 2.25$
Variància y: $Var_y = \dfrac{71.25}{10} - 2.25^2 = 2.06$
Desviació típica y: $\sigma_y = \sqrt{Var_y} = 1.436$

Covariància: $\sigma_{xy} = \dfrac{142.5}{10} - 4.5 \cdot 2.25 = 4.125$
Coeficient de correlació lineal: $r = \dfrac{4.125}{2.87 \cdot 1.436} = 1$
Donat que $r = 1$, tenim una relació funcional positiva.

b) x =nota matemàtiques, y=nota física d'un grup de 10 alumnes.

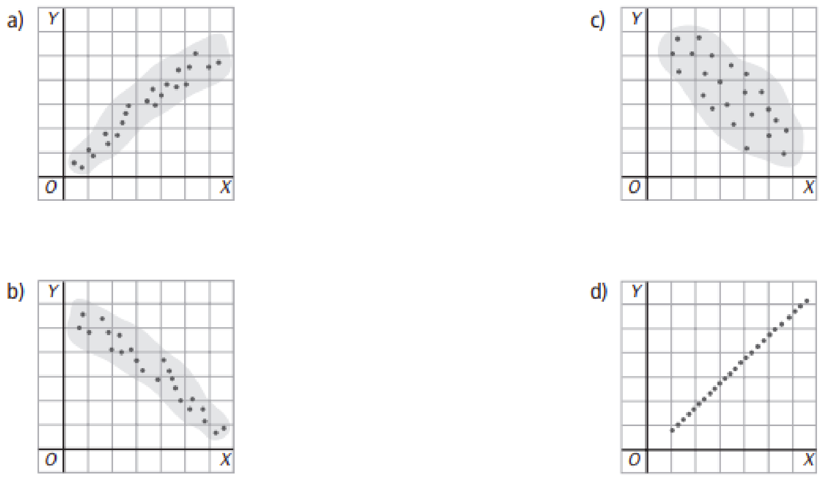
	x	y	x^2	y^2	$x \cdot y$
	3	2	9	4	6
	4	5	16	25	20
	5	5	25	25	25
	6	6	36	36	36
	7	7	49	49	49
	7	6	49	36	42
	7	7	49	49	49
	8	9	64	81	72
	9	8	81	64	72
	10	10	100	100	100
$Sumes$	66	65	478	469	471

Número de parelles $N = 10$
Mitjana x: $\bar{x} = \dfrac{66}{10} = 6.6$
Variància x: $Var_x = \dfrac{478}{10} - 6.6^2 = 4.24$
Desviació típica x: $\sigma_x = \sqrt{Var_x} = 2.06$
Mitjana y: $\bar{y} = \dfrac{65}{10} = 6.5$
Variància y: $Var_y = \dfrac{469}{10} - 2.25^2 = 4.65$
Desviació típica y: $\sigma_y = \sqrt{Var_y} = 2.156$

Covariància: $\sigma_{xy} = \dfrac{471}{10} - 6.6 \cdot 6.5 = 4.2$
Coeficient de correlació lineal: $r = \dfrac{4.2}{2.06 \cdot 2.156} = 0.946$
Donat que $r = 0.946$ és positiu i molt proper a 1, tenim una correlació positiva forta.

Exercicis

4. Associa cada núvol de punts amb el seu coeficient de correlació: $r = 1$; $r = 0,92$; $r = -0,25$; $r = -0,78$;



5. En la següent variable bidimensional:

X	1	2	3	4	3	7	6	3	4	5
Y	45	30	30	25	25	10	20	15	10	15

a) Representa un núvol de punts. Indica el tipus de correlació.

b) Calcula el centre de masses ($\bar{X}, \bar{Y}$) i la covariància.

c) Determina el coeficient de correlació lineal.

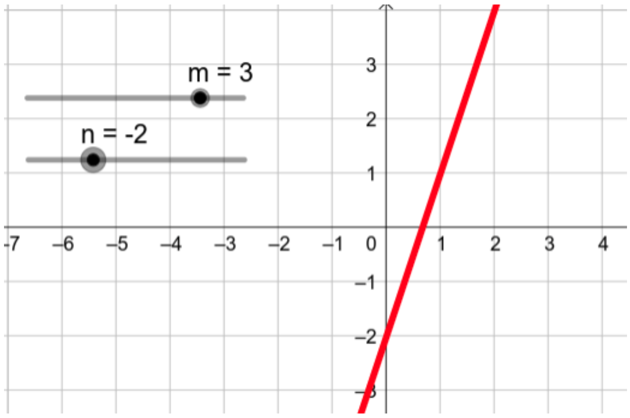
2.3 Repàs de la funció lineal

En aquest apartat farem un petit repàs de l'equació de la recta que haurem d'utilitzar per obtenir la rectade regressió.

L'expressió d'una funció lineal és $y = mx + n$, essent m el pendent i n l'ordenada a l'origen (el punt detall de la recta amb l'eix Y).

- Si $m < 0$: La recta és decreixent

- Si $m = 0$: es diu que la funció és constant i la gràfica és una recta horitzontal
- Si $m > 0$: La recta és creixent



Simulació 1: https://www.geogebra.org/m/GnGvayd4 : *La funció lineal y=mx+n. Canvia els paràmetres m i n i observa com canvia la recta.*

Alternativament, si ens donen un punt (x_0, y_0) i el pendent m podem escriure directament l'equació punt-pendent:

$$y - y_0 = m(x - x_0) \tag{1}$$

Exemple 6

Calcula l'equació de la recta que passa pel punt $(5,3)$ i té pendent $m = -\frac{2}{3}$. Representa-la gràficament.

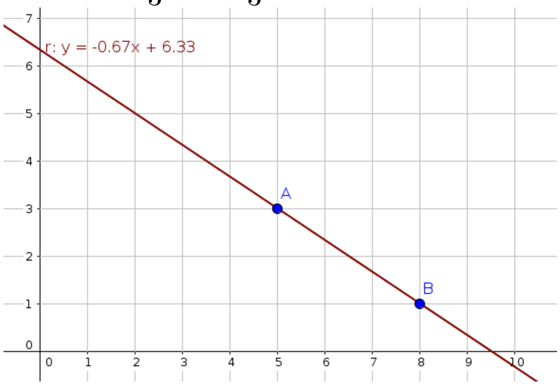
Escrivim l'equació punt-pendent $y - 3 = -\frac{2}{3}(x-5)$

Eliminam els parèntesi i passam tot al membre de la dreta: $y = 3 - \frac{2}{3}x + \frac{2}{3}5$

Obtenim l'equació de la recta: $y = -\frac{2}{3}x + \frac{19}{3}$

La recta talla a l'eix Y en el punt $n = \frac{19}{3}$

Per representar una recta basta amb dos punts, un dels quals ja el tenim $(5,3)$. Si prenem $x = 8$, obtenim $y = -\frac{2}{3} \cdot 8 + \frac{19}{3} = 1$; llavors l'altre punt és $(8,1)$



Exercicis

6. a) Calculeu la recta que passa pels punt $x = 3$, $y = -1$ i que té pendent $m = 4$.

b) Trobeu dos punts més d'aquesta recta.

2.4 Recta de regressió lineal

Suposem que tenim un diagrama núvol de punts com el que mostra la figura; araintentem construir una línia recta que s'aproximi tant com es pugui al núvol de punts.

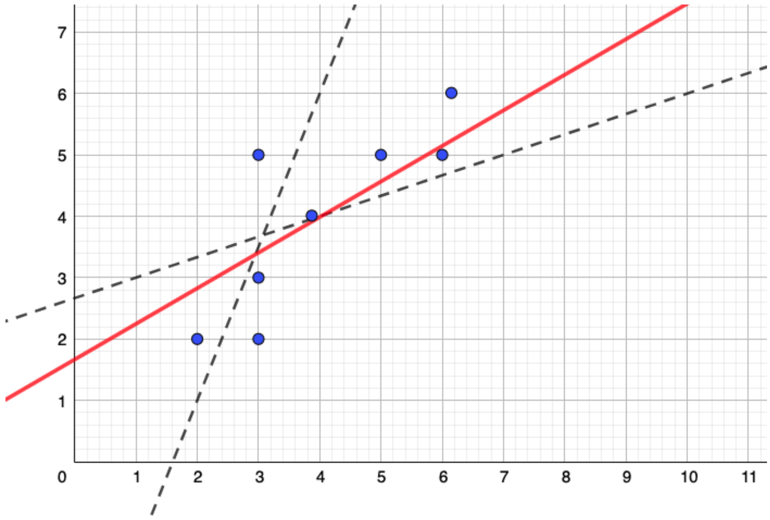


Figura 2: Núvol de punts i diverses rectes. La recta que millor s'aproxima als punts està en vermell. Aquesta recta no té perquè passar per cap punt de les dades.

De totes les rectes possibles, sembla que n'existeix una que compleix la condició. Aquestarecta l'anomenem **recta de regressió lineal** .

Ara bé, el mètode per aconseguir aquesta recta no és fàcil; com a primera aproximació es pot traçar a ull.Però, si necessitam més precisió, caldrà utilitzar un mètode analític.

Exemple complet fent els càlculs a mà

Vídeo 4.1: *Recta de regressió i prediccions*
https://www.youtube.com/watch?v=8mY_IKyMTDQ

Exemple fent els càlculs amb full de càlcul Excel

Vídeo 4.2: *Recta de regressió i prediccions amb EXCEL*
https://www.youtube.com/watch?v=-mMyvX2ubsI

Exemple fent els càlculs amb full de càlcul de Google

Vídeo 4.3: *Recta de regressió i prediccions amb fulls de Google*
https://www.youtube.com/watch?v=XF9gyepL5BU

Mètode dels mínims quadrats

Volem trobar la recta $y = mx + n$ que millor ajusta al núvol de punts. Existeixen molts de mètodes per determinarels valors de m i n. El més utilitzat es coneix com el mètode del mínims quadrats. Consisteix en fer mínimala suma dels quadrats de les diferències entre els valors experimentals i teòrics obtinguts amb l'equació. Els càlculs perl'obtenció de la recta queden fora de l'abast d'aquests apunts i simplement ens limitarem a donar el resultat.

Si s'aplica aquest mètode, s'obté que la recta de regressió passa pel **centre de gravetat** del núvol de punts $(\bar{x}, \bar{y})$. Llavors, la seva equació ha d'ésser de la forma:

$$y - \bar{y} = m(x - \bar{x})$$

El pendent de la recta és $m = \dfrac{\sigma_{xy}}{\sigma_x^2}$.

Aquesta es coneix com la recta de regressió de y sobre x. Amb aquesta recta podem fer prediccions del valor de y conegut el valor de x.Si, en canvi, ens donen una y com a dada, es pot estimar la x aïllant-la de l'equació. Tot seguit farem exemples de com es fan les estimacionsbasant-nos en la recta de regressió.

La recta de regressió per fer estimacions

La recta de regressió s'utilitza per fer prediccions d'una variable coneguda l'altra.Ara bé, quina fiabilitat podem concedir als càlculs obtinguts a través de les rectes de regressió?La predicció serà més bona o fiable com més s'apropi a ± 1 el coeficient de correlació r.Així doncs:

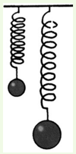
- Si r és proper a 0, no té sentit fer cap mena d'estimacions o prediccions.
- Si r és pròxim a -1 o $+1$, probablement els valors reals seran pròxims a les estimacions que feim.
- Si r és proper 1 o -1, les estimacions fetes coincidiran amb els valors reals.

Existeixen dos tipus de prediccions: **interpolacions** i **extrapolacions** . En les interpolacions,es vol determinar un valor que es troba dins del rang de les dades. En les extrapolacions, en canvi,cau fora del rang de dades.Independentment del grau de correlació, cal tenir en compte que la validesa d'una extrapolaciónomés serà certa si no ens allunyam massa del rang de dades. Per exemple, és possible predir el tempsmeteorològic de demà amb certa precisió però és impossible el predir amb fiabilitat el temps que farà la setmana o el més que vé.

Exemple 7

La llei de Hooke relaciona l'allargament d'una molla amb la força que hi aplicam $F = kx$ on k ésla constant d'elàstica. Hem anant penjat a la molla pesos de diferent valors i hem anotat l'allargament

Allargament x (m)	0	0.06	0.2	0.3	0.6
Pes F (N)	0	0.45	1.1	1.47	2.95



a) Calcula la recta de regressió i estima el valor de la constant elàstica de la molla.
b) Estima el pes necessari per produir un allargament de 0.5 m.

Començarem calculant els paràmetres de cada distribució per separat així com la covariància i el coeficientde correlació lineal.

	x	y	x^2	y^2	$x \cdot y$
	0	0	0	0	0
	0.06	0.45	0.0036	0.202	0.027
	0.2	1.1	0.04	1.21	0.22
	0.3	1.47	0.09	2.16	0.441
	0.6	2.95	0.36	8.7	1.77
$Sumes$	1.16	5.97	0.494	12.28	2.46

Número de parelles $N = 5$
Mitjana x: $\bar{x} = \dfrac{1.16}{5} = 0.232$
Variància x: $Var_x = \dfrac{0.4936}{5} - 0.232^2 = 0.0449$
Desviació típica x: $\sigma_x = \sqrt{Var_x} = 0.212$
Mitjana y: $\bar{y} = \dfrac{5.97}{5} = 1.194$
Variància y: $Var_y = \dfrac{12.28}{5} - 1.194^2 = 1.103$
Desviació típica y: $\sigma_y = \sqrt{Var_y} = 1.015$

Covariància: $\sigma_{xy} = \dfrac{2.46}{5} - 0.242 \cdot 1.194 = 0.2146$
Coeficient de correlació lineal: $r = \dfrac{0.2146}{0.212 \cdot 1.015} = 0.998$
Donat que $r \approx 1$, tenim una correlació forta positiva.

El pendent de la recta y sobre x és

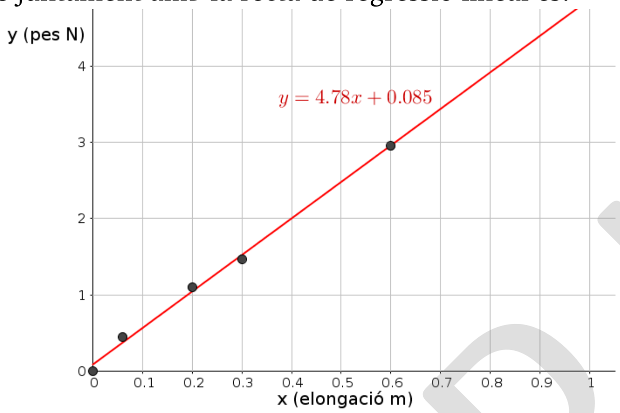
$$m = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{0.2146}{0.212^2} = 4.78 \tag{2}$$

Podem aproximar la constant de la molla a $k = 4.78$ N/m.
L'equació de la recta de regressió és

$$y - 1.194 = 4.78(x - 0.232) \tag{3}$$

que podem simplificar com $y = 4.78x + 0.085$.

b) L'estimació s'obté a partir de la recta de regressió $y = 4.78 \cdot 0.5 + 0.085 = 2.48$ N.Aquesta predicció és bastant fiable ja que $r \approx 1$.
El núvol de punts juntament amb la recta de regressió lineal és:



del qual també es desprèn una correlació forta positiva.

Exemple 8

Es determina la pèrdua d'efectivitat d'un medicament al llarg del temps is'obtenen els resultats següents

Temps x (mesos)	1	2	3	4	5
% d'efectivitat restant	90	78	42	35	21

a) Quin tant per cent d'efectivitat quedarà al cap de 6 mesos?
b) Quan de temps haurà de passar perquè quedi el 50% d'efectivitat restant?

Començarem calculant els paràmetres de cada distribució per separat així com la covariància i el coeficientde correlació lineal.

	x	y	x^2	y^2	$x \cdot y$
	1	90	1	8100	90
	2	78	4	6084	156
	3	42	9	1764	126
	4	35	16	1225	140
	5	21	25	441	105
$Sumes$	15	266	55	17614	617

Número de parelles $N = 5$
Mitjana x: $\bar{x} = \dfrac{15}{5} = 3$
Variància x: $Var_x = \dfrac{55}{5} - 3^2 = 2$
Desviació típica x: $\sigma_x = \sqrt{Var_x} = 1.41$
Mitjana y: $\bar{y} = \dfrac{266}{5} = 53.2$
Variància y: $Var_y = \dfrac{17614}{5} - 53.2^2 = 692.56$
Desviació típica y: $\sigma_y = \sqrt{Var_y} = 26.32$

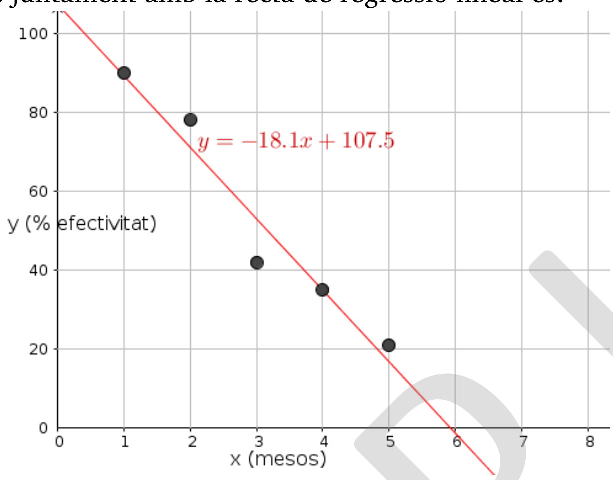
Covariància: $\sigma_{xy} = \dfrac{617}{5} - 3 \cdot 53.2 = -36.2$
Coeficient de correlació lineal: $r = \dfrac{-36.2}{1.41 \cdot 26.32} = -0.972$
Donat que $r \approx -1$, tenim una correlació forta negativa.
El pendent de la recta y sobre x és

$$m = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{-36.2}{1.41^2} = -18.1 \tag{4}$$

L'equació de la recta de regressió és $y - 53.2 = -18.1(x - 3)$que podem simplificar com $y = -18.1x + 107.5$.

Les prediccions que ens demanen seran
a) $y = -18.1 \cdot 6 + 107.5 = -1.1$ %. En aquest cas l'extrapolació fa que obtinguem un percentatgenegatiu cosa que no té sentit. Això indica que el medicament haurà deixat de fer efecte passats 6 mesos.

b) Per fer l'altra predicció, substituïm el valor de y dins l'equació de la recta $50 = -18.1 \cdot x + 107.5$ i aïllam $x = \dfrac{50 - 107.5}{-18.1} = 3.17$ mesos.
Les dues prediccions són bastant fiables ja que $r \approx -1$.
El núvol de punts juntament amb la recta de regressió lineal és:



Exercicis

7. Considereu les dades de l'exercici proposat 6. Es demana que:

a) Calculeu la recta de regressió.

b) Estimeu el valor de y quan $x = 5$

c) Estimeu el valor de x quan $y = 100$

d) Justifiqueu la validesa de les prediccions b) i c).

3. Probabilitat

En aquesta secció farem un resum dels conceptes bàsics de probabilitat apresos durant l'ESO.

Dos tipus d'experiències

- **Experiència aleatòria** és aquella en què hi intervé l'atzar. Per exemple:
  - Llançar una moneda i mirar si surt C o X
  - Treure una carta d'una baralla espanyola i mirar si és un rei
  - Temps que tardam en anar de casa a la feina

- **Experiència determinista** és aquella on és possible predir a priori el resultat de l'experiment.
  - Treure una bolla d'una bossa que només conté bolles vermelles i mirar quin color té (cert i segur que sortirà vermella).
  - Mesurem la quantitat d'aigua que cap en un envàs buit.

Figura: Font de la imatge Pinterest.
[https://www.pinterest.es/pin/803188914805247225/]

Espai mostral i esdeveniments

Anomenam **Espai Mostral** Ω al conjunt de tots els possibles resultats d'una experiència aleatòria

Exemple 9

Escriu l'espai mostral de les experiències aleatòries:
a) llançar una moneda
b) llançar un dau
c) llançar una moneda i un dau simultàniament

a) llançar una moneda: $\Omega = \{C, X\}$

b) llançar un dau: $\Omega = \{1, 2, 3, 4, 5, 6\}$

c) llançar una moneda i un dau simultàniament: $\Omega = \{C\text{-}1, C\text{-}2, C\text{-}3, C\text{-}4, C\text{-}5, C\text{-}6, X\text{-}1, X\text{-}2, X\text{-}3, X\text{-}4, X\text{-}5, X\text{-}6\}$

Anomenam **Esdeveniment elemental** a cadascun dels elements de l'espai mostral; per exemple: "treureun 3 al dau $\{3\}$"és un esdeveniment elemental.

Anomenam **Esdeveniment** a qualsevol subconjunt de l'espai mostral; per exemple: "treureun número major que 3"al dau $\{4, 5, 6\}$ és un exemple d'esdeveniment.

Anomenam **Esdeveniment Impossible** al subconjunt buit $\varnothing$;per exemple: "treure un nombre major que 6 en llançar un dau cúbic".

Anomenam **Esdeveniment Segur** al subconjunt que coincideix amb l'espai mostral Ω;per exemple: "treure entre 1 i 6 en llançar un dau cúbic".

3.1 Probabilitat elemental

La **probabilitat** és una mesura de quant freqüent és un esdeveniment.Es quantifica amb un nombre que com a mínim val 0 (**esdeveniment impossible**)i com a màxim 1 (**esdeveniment segur**).

$$P(\varnothing) = 0 \qquad P(\Omega) = 1 \tag{5}$$

A més, donats dos **esdeveniments incompatibles** ($A \cap B = \varnothing$), es compleix

$$P(A \cup B) = P(A) + P(B) \tag{6}$$

Aquesta relació es generalitza amb

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{7}$$

quan existeix una intersecció entre els dos esdeveniments.

Si S' i S són **esdeveniments contraris** , es compleix que $P(S') = 1 - P(S)$.

Es considerà falta molt greu tot aquell qui afirmi en un examen o treball que la probabilitatés negativa o major que 1.

Si ens parlem d'una probabilitat del 70%, en realitat, cal expressar-la en tant per u, P=0,7

Llei dels grans nombres

Considerem l'experiment aleatori de llançar un dau cúbic i anotar la puntuació de la cara superior.

Regla de Laplace

Si tots els possibles resultats d'un experiment aleatori (**esdeveniments elementals**)tenen la mateixa probabilitat (són **equiprobables**),podem aplicar la **regla de Laplace** :

$$P(S) = \frac{\text{Nombre de casos favorables a S}}{\text{Nombre de casos totals}} \tag{8}$$

La regla de Laplace no es pot aplicar sempre. Només en experiències regulars.

- Si una moneda no és perfectament simètrica, cada cara tindrà una probabilitat diferent de sortir que l'altra.

Exemple 10

Justifica si es pot aplicar la regla de Laplace i, si és possible, calcula la probabilitat de cada esdeveniment:
a) Treure creu en una moneda regular
b) Treure un número major que 4 en dau cúbic regular
c) Treure un número major que 4 en dau trucat
d) No treure As d'una baralla espanyola de 40 cartes
NOTA: Quan treballem amb baralles de cartes espanyoles, consideram que els 8 i 9 s'han eliminat i, per tant, hi haun total de 40 cartes.

a) $P(X) = \frac{1}{2}$
b) $P(>4) = \frac{2}{6} = \frac{1}{3}$
c) Si el dau està trucat no totes les cares tindran igual probabilitat de sortir i, per tant, no es pot aplicarla regla de Laplace. S'hauria d'obtenir la probabilitat experimentalment.
d) "No treure As" és el esdeveniment contrari de "treure As". Sabem que "treure As" té una probabilitat de $P(As) = \frac{4}{40} = \frac{1}{10}$. Per tant,el esdeveniment contrari "no As" té probabilitat $P(As') = 1 - P(As) = 1 - \frac{1}{10} = \frac{9}{10}$.

Exercicis

8. En un grup hi ha 20 alumnes, dels quals 16 són dones i la resta homes. Si agafam un alumne a l'atzar, trobau la probabilitat quesigui dona i la probabilitat que sigui home.

9. Consideram l'experiment llançar un dau. Siguin els esdeveniments:

A="Treure un nombre primer"; B="Treure múltiple de 3"; C="Treure parell"

Calculau les probabilitats $P(A)$, $P(B)$, $P(C)$, $P(A \cap B)$ i $P(B \cup C)$. *Recorda: La intersecció $\cap$ significa que passin els dos esdeveniments alhora i la unió $\cup$ que passi almenys un d'ells.*

3.2 Probabilitat d'experiments composts

Entenem com **experiment compost** aquell format per dos o més experiments simples. Per exemple,si consideram els experiments simples llançar un dau i treure una carta;podríem formar els següents experiments composts:

Considerem dos esdeveniments A i B. Classificam els experiments composts en:

- **Independents** : Si A i B són independents es compleix que

$$P(A \cap B) = P(A) \cdot P(B)$$

  El símbol $\cap$ s'anomena intersecció. $A \cap B$ vol dir que passi A i B alhora.

- **Dependents** : La probabilitat de B depèn del resultat que s'hagi obtingut d' A. En aquest cas, necessitaremintroduir el concepte de **probabilitat condicionada**

Exemple 11

Classifica els següents esdeveniments en independents o dependents. En el cas que siguin independents,calcula'n la probabilitat a partir de $P(A \cap B) = P(A) \cdot P(B)$.
a) Obtenir creu en llançar una moneda 2 vegades.
b) Obtenir dos asos en treure dues cartes amb reemplaçament (la treim, miram i la tornam a introduir a la baralla).
c) Obtenir dos asos en treure dues cartes sense reemplaçament (la treim, i no la tornam a introduir).
d) Obtenir cara en una moneda, parell a un dau i oros en extreure simultàniament una carta.

a) Els llançaments són independents. $P(C_1 \cap C_2) = P(C) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

b) Si les extraccions són amb reemplaçament, són esdeveniments independents.
$P(As_1 \cap As_2) = P(As) \cdot P(As) = \frac{4}{40} \cdot \frac{4}{40} = \frac{1}{100}$

c) Si les extraccions són sense reemplaçament, els esdeveniments són dependents. La probabilitat que la segona carta surti As depèn si la primera ha sortit As o no.

d) Els tres esdeveniments són independents.

$$P(Cara \cap Parell \cap Oros) = P(Cara) \cdot P(Parell) \cdot P(Oros) = \frac{1}{2} \cdot \frac{3}{6} \cdot \frac{10}{40} = \frac{1}{16}$$

Producte Cartesià

La tècnica del producte Cartesià permet escriure tots els esdeveniments d'un experiment compost. Ens construïm unataula on posam a les capçaleres de les files i columnes, els esdeveniments elementals de cada experiment simple. Vegem com s'aplica mitjançant un exemple.

Exemple 12

Llançam dos daus i en sumam les puntuacions. Calculau la probabilitat que:
a) la suma sigui igual 4
b) la suma sigui major o igual a 10

Ens construïm una taula on en la fila escrivim els resultats del primer dau i, en columna, els resultats delsegon dau. A aquesta taula l'anomenam producte cartesià.

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Ens adonam que existeixen $6 \times 6 = 36$ combinacions, totes elles amb igual probabilitat. Llavors podem aplicar la regla de Laplace:
a) $P(x = 4) = \frac{3}{36} = \frac{1}{12}$
b) $P(x \geq 10) = \frac{6}{36} = \frac{1}{6}$

Exercicis

10. Llançam dos daus i restam les seves puntuacions (en valor absolut). Anomenam els esdeveniments A=l·la diferència és menor o igual a 3”, B=l·la diferència és major o igual a 2”.Calcula la probabilitat dels esdeveniments:

a) $P(A)$

b) $P(B)$

c) $P(A \cap B)$