

3. Estadística descriptiva bivariant

Universitat de Barcelona



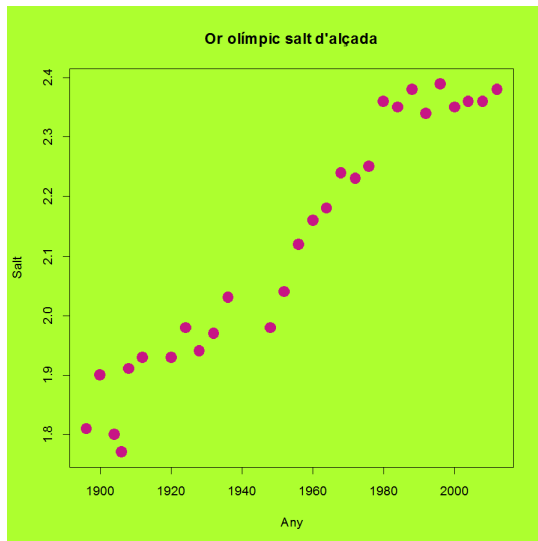
UNIVERSITAT_{DE}
BARCELONA

Marca dels medallistes d'or

Dades de la marca dels medallistes d'or de salt d'alçada-homes de tots els jocs olímpics, des de 1896 fins a 2012.

Com evoluciona aquesta marca amb el temps?

Gràfica de les marques dels medallistes d'or amb l'any

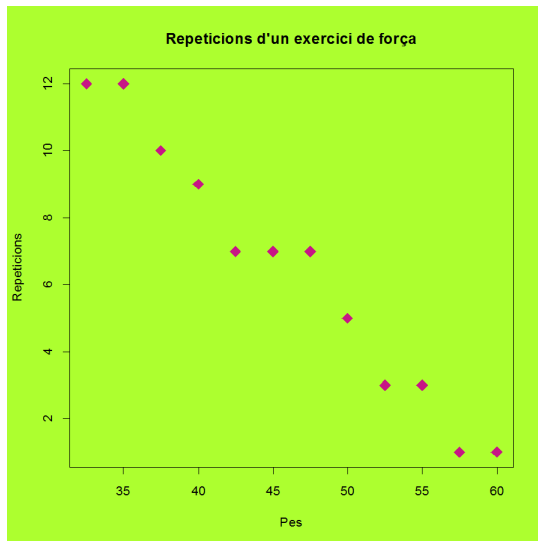


Nombre màxim de repeticions segons el pes

En un exercici de força, quan s'augmenta el pes a aixecar, es poden fer menys repeticions de l'exercici.

Anem augmentant el pes i registrem el nombre màxim de repeticions que poden fer diferents atletes. Veiem un exemple dels resultats.

Gràfica del nombre màxim de repeticions amb el pes

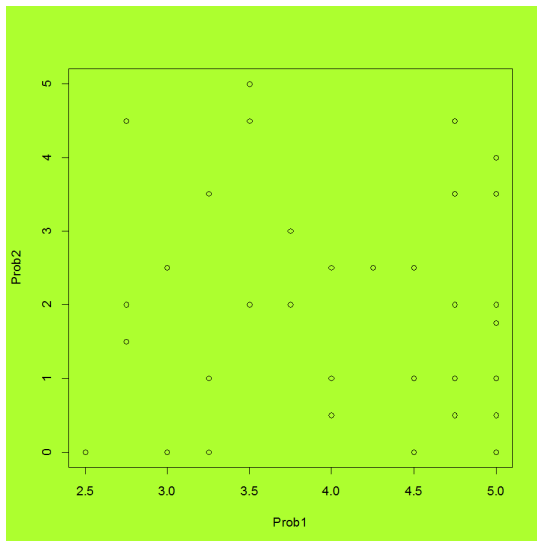


Notes de dos problemes

Els alumnes d'un curs de probabilitats han fet un examen amb dos problemes.

Comparem les notes obtingudes.

Gràfica de les notes

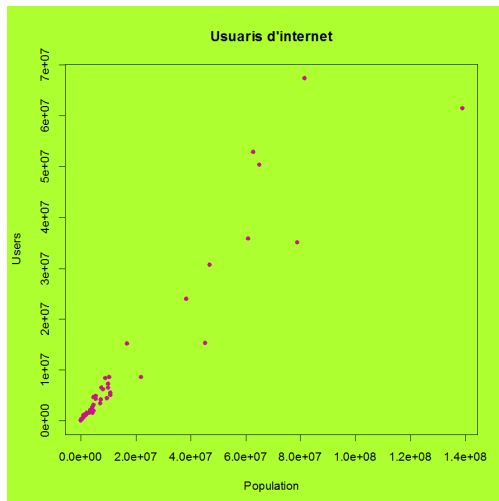


Usuaris d'internet i de Facebook

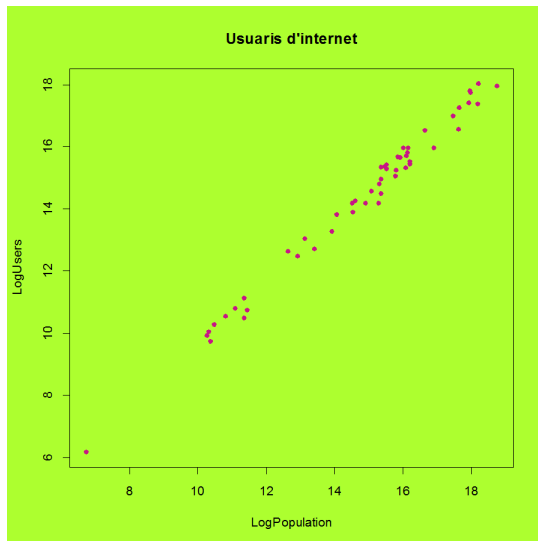
Per països, dades de 2011 i 2012,

- Nombre d'habitants,
- Nombre d'usuaris d'internet i
- Nombre d'usuaris de Facebook

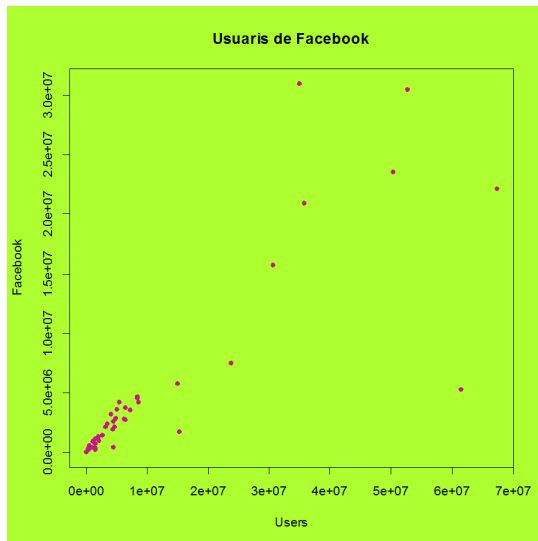
Gràfica de nombre d'usuaris d'internet per nombre d'habitants



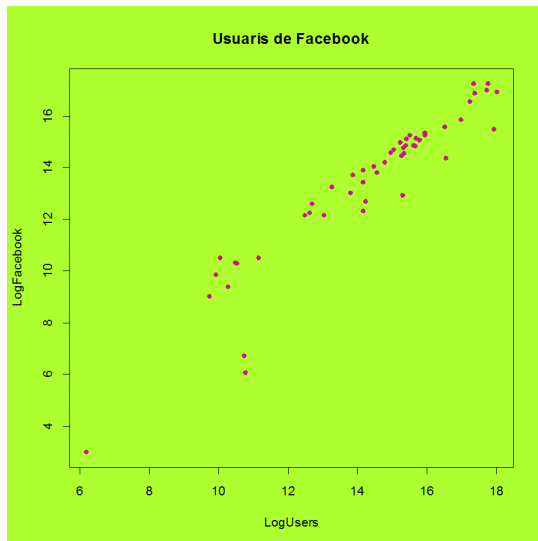
Gràfica dels logaritmes



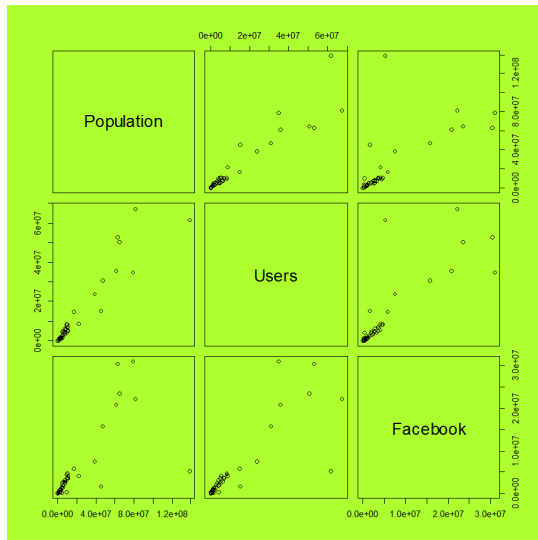
Usuaris de Facebook per usuaris d'internet



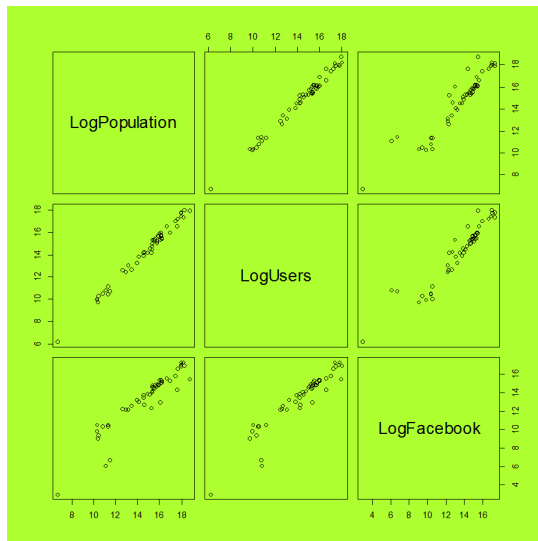
Gràfica dels logaritmes



Matriu de diagrames de dispersió (*scatterplot matrix*)



Matriu de diagrames de dispersió dels logaritmes



Taules de contingència

- Tenim n observacions bivariants $(x_1, y_1), \dots, (x_n, y_n)$.
- Per variables principalment qualitatives.
- Si les variables són quantitatives contínues agruparem per intervals.
- Taula de contingència: taula bidimensional.

Exemple: Color dels ulls i dels cabells. Format data frame

Eyes	Hair	Freq
Blue	Fair	326
Light	Fair	688
Medium	Fair	343
Dark	Fair	98
Blue	Red	38
Light	Red	116
Medium	Red	84
Dark	Red	48
Blue	Medium	241
Light	Medium	584
Medium	Medium	909
Dark	Medium	403
Blue	Dark	110
Light	Dark	188
Medium	Dark	412
Dark	Dark	681
Blue	Black	3
Light	Black	4
Medium	Black	26
Dark	Black	85

Exemple: Color dels ulls i dels cabells. Format taula

Dades de color d'ulls i dels cabells de 5387 nens de Caithness (Escòcia).

Files = colors dels ulls; Columnes = colors dels cabells.

	Fair	Red	Medium	Dark	Black
Blue	326	38	241	110	3
Light	688	116	584	188	4
Medium	343	84	909	412	26
Dark	98	48	403	681	85

Marginals d'una taula de contingència \mathbf{N} ($k \times l$)

$$\mathbf{N} = (n_{i,j}), \quad 1 \leq i \leq k, \quad 1 \leq j \leq l$$

- $n_{i,j}$ és la freqüència absoluta de la parella (x_i, y_j) .

- La suma

$$n = \sum_{i=1}^k \sum_{j=1}^l n_{i,j}$$

- La freqüència absoluta de x_i

$$n_{i,\cdot} = \sum_{j=1}^l n_{i,j}$$

- La freqüència absoluta de y_j

$$n_{\cdot,j} = \sum_{i=1}^k n_{i,j}$$

Amb les dades Caithness

	Fair	Red	Medium	Dark	Black	
Blue	326	38	241	110	3	718
Light	688	116	584	188	4	1580
Medium	343	84	909	412	26	1774
Dark	98	48	403	681	85	1315
	1455	286	2137	1391	118	5387

Taula de freqüències relatives i les seves marginals

$$\mathbf{F} = \mathbf{N}/n$$

- $f_{i,j} = \frac{n_{i,j}}{n}$ és la freqüència relativa de la parella (x_i, y_j) i pren valors entre 0 i 1.

- La suma

$$1 = \sum_{i=1}^k \sum_{j=1}^l f_{i,j}$$

- Les freqüències relatives de x_i i de y_j

$$f_{i,\cdot} = \frac{n_{i,\cdot}}{n} \qquad f_{\cdot,j} = \frac{n_{\cdot,j}}{n}$$

- La freqüència relativa de y_j dins dels x_i i la freqüència relativa de de x_i dins dels y_j

$$\frac{n_{i,j}}{n_{i,\cdot}} \qquad \frac{n_{i,j}}{n_{\cdot,j}}$$

Amb les dades Caithness

	Fair	Red	Medium	Dark	Black	
Blue	0.061	0.007	0.045	0.020	0.001	0.133
Light	0.128	0.022	0.108	0.035	0.001	0.293
Medium	0.064	0.016	0.169	0.076	0.005	0.329
Dark	0.018	0.009	0.075	0.126	0.016	0.244
	0.270	0.053	0.397	0.258	0.022	1.000

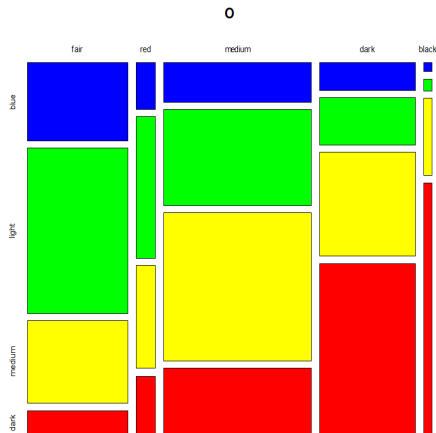
- La proporció de nens de cabells rossos en les dades és, en tant per u, 0.270.
- La proporció de nens d'ulls blaus és, en tant per u, 0.133.
- La proporció dels que tenen a la vegada ulls blaus i cabells rossos és 0.061.

Diagrama de mosaics

Una columna de rectangles per a cada columna de la taula.

Rectangles d'alçada proporcional a cada freqüència.

Diagrama de mosaics per a les dades Caithness



Si no hi hagués relació entre els dos caràcters

Si la proporció d'ulls blaus a la columna de cabells rossos fos la mateixa que a la població total,

$$e_{blue,fair}/0.270 = 0.133,$$

aleshores la proporció dels que tenen a la vegada ulls blaus i cabell ros hauria de ser (*frequència esperada*):

$$e_{blue,fair} = 0.133 \times 0.270 = 0.0359,$$

Però en realitat la *frequència observada* és:

$$o_{blue,fair} = n_{blue,fair} = 0.061,$$

Si no hi hagués relació entre els dos caràcters

En general, la taula de *freqüències relatives esperades* $\mathbf{e} = (e_{i,j})$, si els dos caràcters fóssin independents (*freqüències relatives esperades*), és el producte:

$$e_{i,j} = p_i \times q_j.$$

La taula de *freqüències absolutes esperades*, $\mathbf{E} = n \mathbf{e}$, és diferent de la taula observada $\mathbf{O} = \mathbf{N}$.

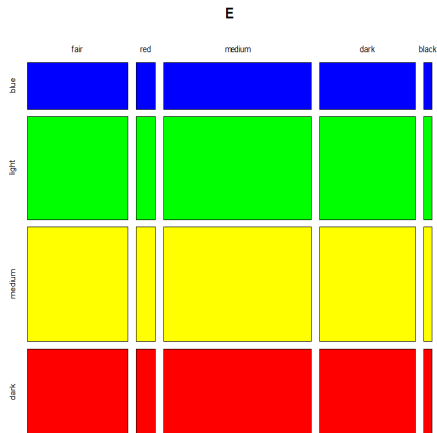
K. Pearson (1904) anomenà *contingència* aquesta diferència.

Amb les dades Caithness

La taula de freqüències relatives esperades és:

	Fair	Red	Medium	Dark	Black	
Blue	0.036	0.007	0.053	0.034	0.003	0.133
Light	0.079	0.016	0.116	0.076	0.006	0.293
Medium	0.089	0.017	0.131	0.085	0.007	0.329
Dark	0.066	0.013	0.097	0.063	0.005	0.244
	0.270	0.053	0.397	0.258	0.022	1.000

Mosaïc de la taula de freqüències esperades



L'estadístic χ^2

És una mesura de la diferència entre les freqüències observades **O** i les esperades **E** en cas de no associació.

$$\chi^2 = \sum \frac{(\mathbf{O} - \mathbf{E})^2}{\mathbf{E}}.$$

Amb les dades Caithness, $\chi^2 = 1240.04$

Mesures de dependència

Un conjunt de n parelles de valors:

$$(x_i, y_i) \in \mathbb{R}^2, \quad 1 \leq i \leq n.$$

Cada parella: dues variables, X , Y , mesurades sobre un mateix individu.

Per a cada individu i -èsim,

$$(x_i - \bar{x})(y_i - \bar{y})$$

és positiu o negatiu segons si les diferències de x_i , y_i respecte les seves mitjanes van o no en el mateix sentit.

Covariància

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Compta si, en promig, les dues variables es mouen en la mateixa direcció o en direccions oposades.

Alternativa

$$s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$$

on $\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$.

Propòsit de la covariància

Objectiu: Mesurar si hi ha relació o associació entre les variables X , Y .

Interpretació:

- Covariància gran positiva \rightarrow en promig creixen o decreixen juntes.
- Covariància gran negativa \rightarrow en promig una creix quan l'altra disminueix.

Problema: Depèn de les unitats de mesura.

El coeficient de correlació de Pearson

$$r_{xy} = \frac{s_{xy}}{s_x s_y},$$

És la covariància de les variables estandarditzades:

$$r_{xy} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right).$$

Alternativa

$$r_{xy} = \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}}{s_x s_y}$$

Desigualtat de Cauchy-Schwarz

$$|s_{xy}| \leq s_x \cdot s_y$$

Com a conseqüència,

$$-1 \leq r_{xy} \leq 1$$

r_{xy} no depèn de les unitats de mesura.

Com afecten les transformacions lineals?

Covariància

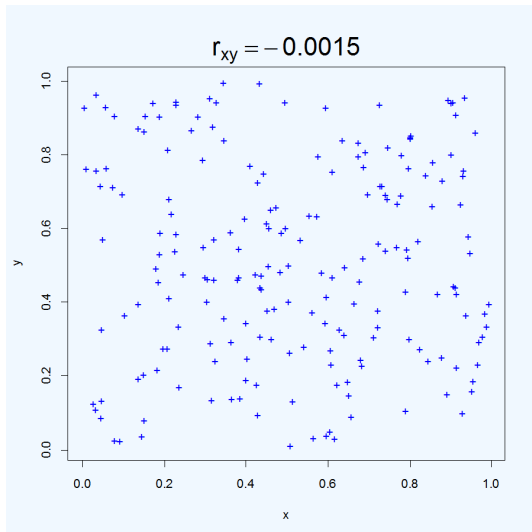
$$s_{a+bx, c+dy} = b d s_{x,y}$$

Correlació

No varia per desplaçament o canvis d'escala de les variables.

$$|r_{a+bx, c+dy}| = |r_{x,y}|$$

Punts a l'atzar sense relació entre X , Y



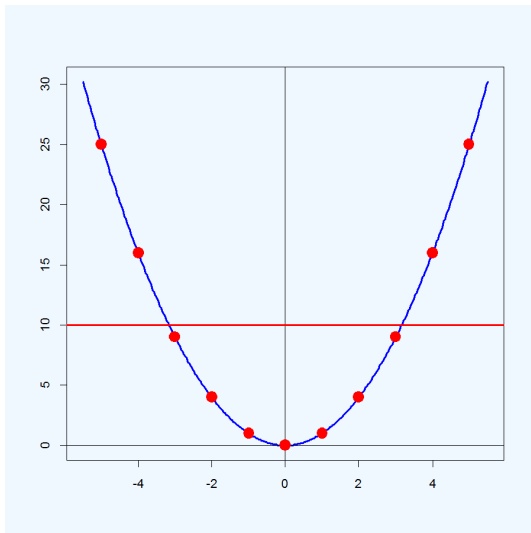
Exemple: Y funció de X però $r_{xy} = 0$

X	-5	-4	-3	-2	-1	0	1	2	3	4	5
Y	25	16	9	4	1	0	1	4	9	16	25

$$r_{xy} = 0.$$

No detecta la dependència quadràtica de les variables.

Exemple: Y funció de X però $r_{xy} = 0$

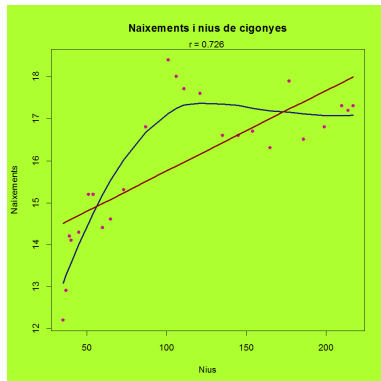


Què és el que mesura r_{xy} ?

- Punts a l'atzar sense relació entre X , Y , tenen r_{xy} pròxim a 0.
- Però variables *molt associades* poden tenir també r_{xy} pròxim a 0 o, fins i tot, exactament igual a 0.
- $|r_{xy}|$ és pròxim a 1 si els punts estan pròxims a estar situats en línia recta. **Només detecta l'associació o dependència lineal.**

Nombre de nius de cigonyes i nombre de naixements

Dades de Dinamarca, al període 1953–1977, referents a freqüència de nius de cigonyes.



Hi ha alguna relació de causa a efecte?

Un problema de geometria

Dades:

$$n \text{ punts } (x_i, y_i) \in \mathbb{R}^2,$$

Volem calcular:

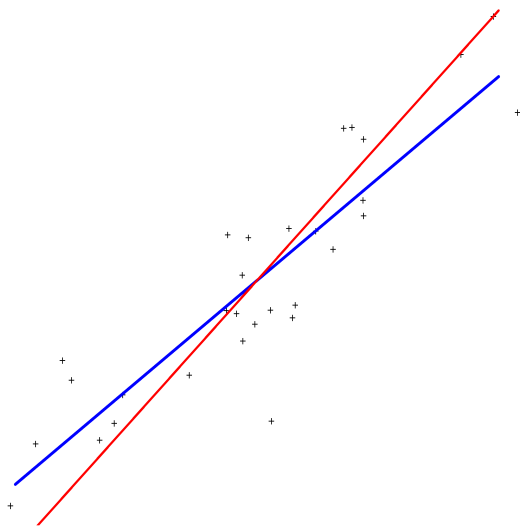
$$a, b \in \mathbb{R}$$

de manera que la recta d'equació

$$y = a + bx$$

sigui *una bona aproximació* al conjunt de punts.

Què és *aproximació d'una recta a un conjunt de punts*



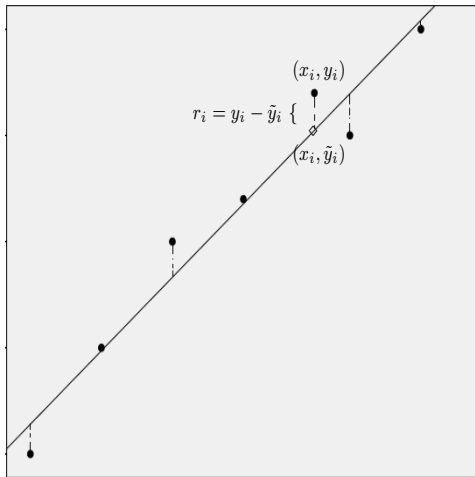
Ajust per mínims quadrats

La recta de regressió *per mínims quadrats* és la que té els paràmetres a i b que fan mínima la funció

$$\sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - (a + b x_i))^2, \quad a, b \in \mathbb{R},$$

la mitjana de les distàncies verticals al quadrat (equivalentment, la suma).

Ajust per mínims quadrats



$$r_i = y_i - (a + b x_i)$$

Un problema d'optimització

$\sum_{i=1}^n r_i^2$ és un polinomi de segon grau en a i b

Determinem el mínim igualant a 0 les derivades

$$\frac{\partial}{\partial a} \sum_{i=1}^n r_i^2 = -2 \sum_{i=1}^n (y_i - a - b x_i) = 0$$

$$\frac{\partial}{\partial b} \sum_{i=1}^n r_i^2 = -2 \sum_{i=1}^n x_i (y_i - a - b x_i) = 0$$

Resolem el sistema i comprovem que obtenim un màxim per

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2}$$

$$\hat{a} = \bar{y} - b \bar{x}.$$

Propietats de la recta de regressió

La recta de regressió per mínims quadrats

$$y = \hat{a} + \hat{b}x,$$

- Passa pel centre de gravetat (\bar{x}, \bar{y}) del núvol de punts:

$$y - \bar{y} = \hat{b}(x - \bar{x}).$$

- El pendent de la recta té el mateix signe que s_{xy} i que r_{xy} :
 - $r > 0$: Pendent positiu,
 - $r < 0$: Pendent negatiu.

Valors ajustats

Per cada x_i , tenim el i -èsim *valor ajustat*,

$$\hat{y}_i = \hat{a} + \hat{b} x_i,$$

la ordenada del punt situat en la vertical de x_i , sobre la recta de regressió.

Observació: La mitjana aritmètica dels \hat{y}_i és igual a la mitjana aritmètica dels y_i .

Residus

Per cada x_i tenim el i -èsim *residu*,

$$e_i = y_i - \hat{y}_i.$$

Els *residus* són una estimació dels errors d'ajustament als punts observats.

Propietats:

$$\sum_{i=1}^n e_i = 0$$

$$\sum_{i=1}^n x_i e_i = 0$$

El coeficient de determinació

El **coeficient de determinació** R^2 és

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

Interpretació: proporció de la variabilitat de les dades que queda *explicada* pel model de regressió.

Observació Com que $s_{\hat{y}}^2 = \frac{s_{xy}^2}{s_x^2}$, tenim que:

$$R^2 = r^2$$

Per tant, tenim que $R^2 \leq 1$.

Alternativa pel càlcul de R^2

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT}$$

on

$$SQT = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{suma de quadrats totals}$$

$$SQE = \sum_{i=1}^n e_i^2 \quad \text{suma de quadrats dels errors}$$

$$SQR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{suma de quadrats de la regressió}$$

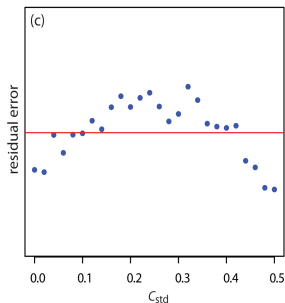
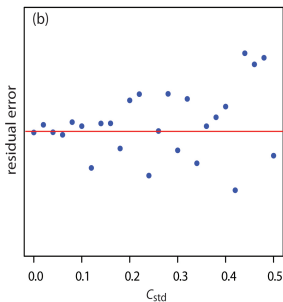
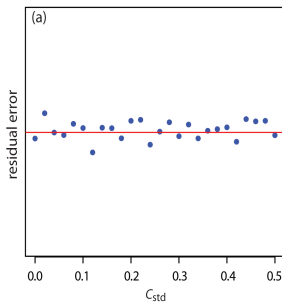
Es compleix que

$$SQT = SQR + SQE$$

Diagrama de residus

El *diagrama de residus* és la gràfica dels punts (\hat{y}_i, e_i) , residus respecte a valors ajustats.

Serveix per diagnosticar si la regressió és adequada.

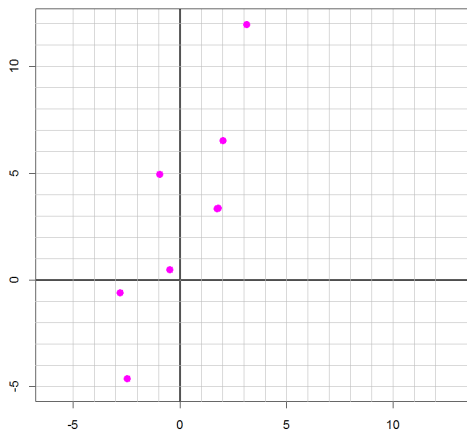


El conjunt de punts

Un conjunt de $n = 8$ punts obtinguts per simulació:

x	1.74	-2.81	-2.47	1.80	3.13	-0.46	2.02	-0.94
y	3.34	-0.58	-4.61	3.37	11.93	0.46	6.51	4.94

El conjunt de punts



Dades

$$\bar{x} = 0.25,$$

$$\bar{y} = 3.17,$$

$$s_x^2 = 4.35,$$

$$s_y^2 = 21.65,$$

$$\tilde{s}_x^2 = 4.97,$$

$$\tilde{s}_y^2 = 24.75,$$

$$s_x = 2.09,$$

$$s_y = 4.65,$$

$$\tilde{s}_x = 2.23,$$

$$\tilde{s}_y = 4.97,$$

$$s_{xy} = 8.03,$$

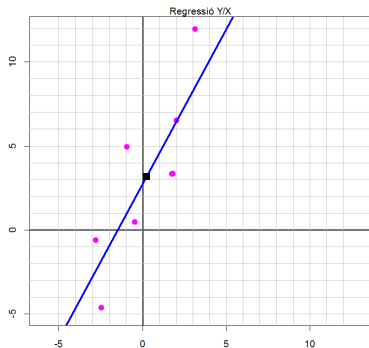
$$\tilde{s}_{xy} = 9.18,$$

$$r_{xy} = 0.83,$$

$$r_{xy}^2 = 0.69.$$

Regressió de Y/X

$$\begin{aligned}
 b &= \frac{\tilde{s}_{xy}}{\tilde{s}_x^2} = \frac{s_{xy}}{s_x^2} = \frac{9.18}{4.97} = \frac{8.03}{4.35} = 1.85, \\
 a &= \bar{y} - b\bar{x} = 3.17 - 1.85 \times 0.25 = 2.71.
 \end{aligned}$$



Valors ajustats i residus

$$\hat{y}_i = a + b x_i, \quad \tilde{y}_i = e_i = y_i - \hat{y}_i, \quad 1 \leq i \leq n.$$

x	1.74	-2.81	-2.47	1.80	3.13	-0.46	2.02	-0.94
y	3.34	-0.58	-4.61	3.37	11.93	0.46	6.51	4.94
\hat{y}	5.92	-2.49	-1.86	6.03	8.50	1.86	6.45	0.97
\tilde{y}	-2.58	1.90	-2.75	-2.66	3.44	-1.40	0.06	3.98

Punts ajustats sobre el diagrama de regressió

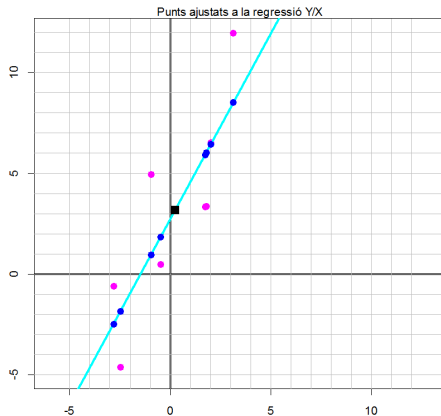
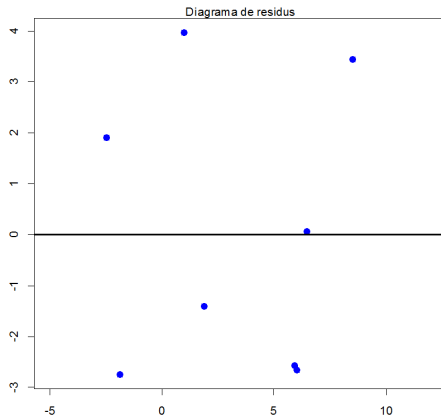


Diagrama dels residus front als valors ajustats



Descomposició de la variabilitat

Expressada en termes de les variàncies (corregides):

$$\text{Variància de } y = \tilde{s}_y^2 = 24.75,$$

$$\text{Variància de } \hat{y} = \tilde{s}_{\hat{y}}^2 = 16.96,$$

$$\text{Variància dels residus} = \tilde{s}_{\tilde{y}}^2 = 7.78.$$

Coefficient de determinació:

$$R^2 = \frac{\tilde{s}_{\hat{y}}^2}{\tilde{s}_y^2} = r_{xy}^2 = 0.69.$$

Descomposició de la variabilitat

Expressada en termes de les sumes de quadrats dels valors centrats:

$$SS(Total)_0 = (n - 1) \tilde{s}_y^2 = 173.22,$$

$$SS(Reg)_0 = (n - 1) \tilde{s}_{\hat{y}}^2 = 118.75,$$

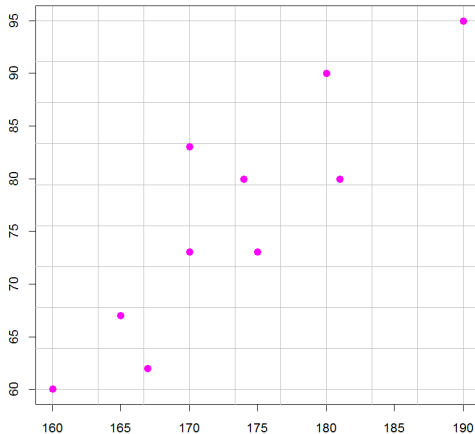
$$RSS = (n - 1) \tilde{s}_{\tilde{y}}^2 = 54.47.$$

El conjunt de punts

Un conjunt de $n = 10$ punts:

x	180	170	165	190	170	181	175	174	167	160
y	90	73	67	95	83	80	73	80	62	60

El conjunt de punts



Mitjanes, variàncies, desviacions estàndard

$$\bar{x} = 173.2,$$

$$\bar{y} = 76.3,$$

$$s_x^2 = 69.36,$$

$$s_y^2 = 118.81,$$

$$s_x = 8.33,$$

$$s_y = 10.9,$$

$$\tilde{s}_x^2 = 77.07,$$

$$\tilde{s}_y^2 = 132.01,$$

$$\tilde{s}_x = 8.78,$$

$$\tilde{s}_y = 11.49,$$

$$\tilde{s}_{xy} = 89.16,$$

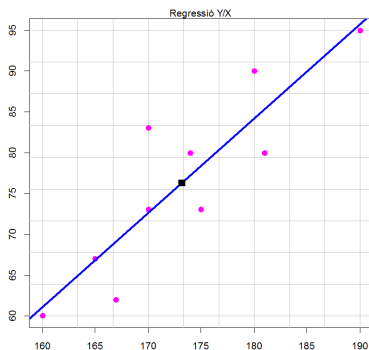
$$s_{xy} = 80.24,$$

$$r_{xy} = 0.88,$$

$$r_{xy}^2 = 0.78.$$

Gràfica de la regressió de Y/X

$$\begin{aligned} b &= \frac{\tilde{s}_{xy}}{\tilde{s}_x^2} = \frac{89.16}{77.07} = 1.16, \\ a &= \bar{y} - b\bar{x} = 76.3 - 1.16 \times 173.2 = -124.07. \end{aligned}$$



Valors ajustats i residus

$$\hat{y}_i = a + b x_i, \quad \tilde{y}_i = e_i = y_i - \hat{y}_i, \quad 1 \leq i \leq n.$$

x	180	170	165	190	170	181	175	174	167	160
y	90	73	67	95	83	80	73	80	62	60
\hat{y}	84.17	72.60	66.81	95.74	72.60	85.32	78.38	77.23	69.13	61.03
\tilde{y}	5.83	0.40	0.19	-0.74	10.40	-5.32	-5.38	2.77	-7.13	-1.03

Punts ajustats sobre el diagrama de regressió

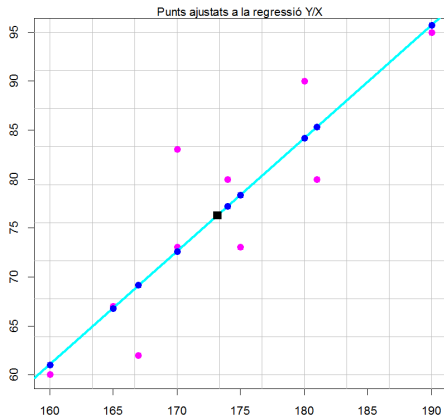
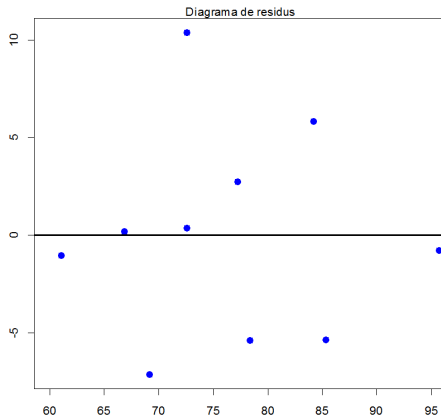


Diagrama dels residus front als valors ajustats



Descomposició de la variabilitat

Expressada en termes de les variàncies (corregides):

$$\text{Variància de } y = \tilde{s}_y^2 = 132.01,$$

$$\text{Variància de } \hat{y} = \tilde{s}_{\hat{y}}^2 = 103.14,$$

$$\text{Variància dels residus} = \tilde{s}_{\tilde{y}}^2 = 28.87.$$

Coefficient de determinació:

$$R^2 = \frac{\tilde{s}_{\hat{y}}^2}{\tilde{s}_y^2} = r_{xy}^2 = 0.78.$$

Descomposició de la variabilitat

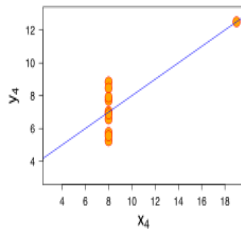
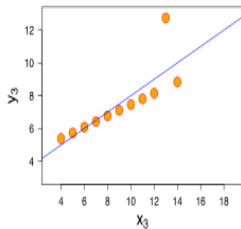
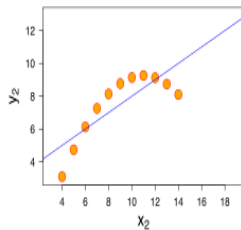
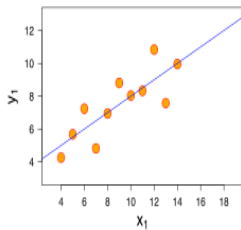
Expressada en termes de les sumes de quadrats dels valors centrats:

$$SS(Total)_0 = (n - 1) \tilde{s}_y^2 = 1188.10,$$

$$SS(Reg)_0 = (n - 1) \tilde{s}_{\hat{y}}^2 = 928.27,$$

$$RSS = (n - 1) \tilde{s}_{\tilde{y}}^2 = 259.83.$$

```
data(anscombe)
```



data(roller)

```
require(DAAG)
data(roller)
roller.lm <- lm(depression ~ weight, data = roller)
summary(roller.lm )
plot(depression ~ weight, data = roller)
abline(roller.lm)
anova(roller.lm)
```

```
data(ironslag)
```

```
require(DAAG)
data(ironslag)
ironslag.lm <- lm(chemical ~ magnetic, data = ironslag)
summary(ironslag.lm)
plot(chemical ~ magnetic, data = ironslag)
abline(ironslag.lm)
anova(ironslag.lm)
```

Regressió del pes sobre el volum dels llibres

data(softbacks)

```
require(DAAG)
data(softbacks)
softbacks.lm <- lm(weight ~ volume, data=softbacks)
summary(softbacks.lm)
plot(weight ~ volume, data=softbacks)
abline(softbacks.lm)
anova(softbacks.lm)
```


Dades de cromatografia de gasos. Plantejament del problema

La cromatografia de gasos és una tècnica per a detectar i identificar substàncies a partir de quantitats molt petites.

S'han pres quatre mesures de cadascuna de cinc mostres amb diferents quantitats conegudes d'un producte.

La finalitat és calibrar l'instrument.

Dades per a la regressió

amount	response	amount	response
0.25	6.55	5.00	211
0.25	7.98	5.00	204
0.25	6.54	5.00	212
0.25	6.37	5.00	213
0.25	7.96	5.00	205
1.00	29.7	20.00	929
1.00	30.0	20.00	905
1.00	30.1	20.00	922
1.00	29.5	20.00	928
1.00	29.1	20.00	919

Resultat de la regressió

```
lm(formula = response ~ amount, data = chrom)
```

Coefficients:

	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	-14.4107	2.6142	-5.512	3.11e-05
amount	46.6287	0.2533	184.086	$< 2e - 16$

Residual standard error: 9.023 on 18 degrees of freedom

Multiple R-Squared: 0.9995.

Gràfica de la regressió

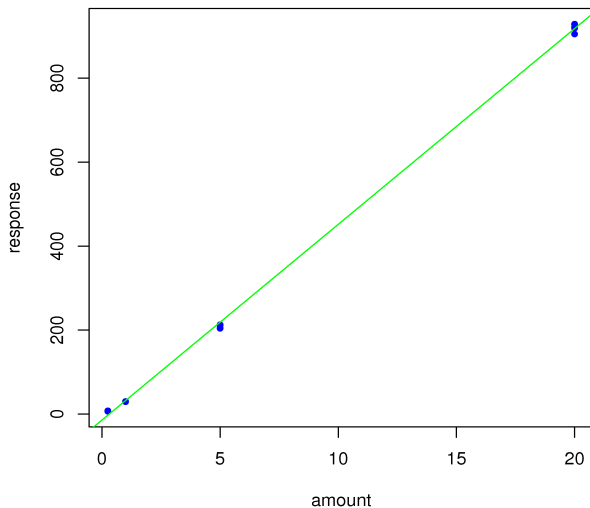
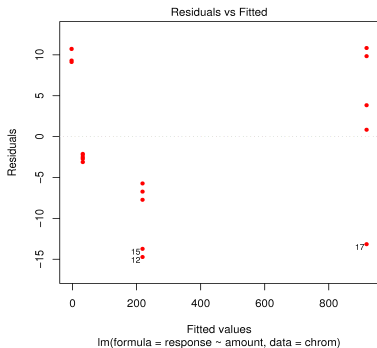


Diagrama de residus



La variabilitat dels residus augmenta amb els resultats ajustats.

Això suggereix una inadequació del model.

No s'aconsegueix explicar la variabilitat de les dades.

Preus del peix. Plantejament del problema

Dades de preus de diverses espècies de peix a l'origen (pescadors) en els anys 1970 i 1980.

Es tracta de predir, per a una determinada espècie, el preu a l'any 1980 a partir del preu a l'any 1970.

Dades per a la regressió

Type_Fish	Price_1970	Price_1980
Cod	13.1	27.3
Flounder	15.3	42.4
Haddock	25.8	38.7
Menhaden	1.8	4.5
Ocean perch	4.9	23
Salmon, Chinook	55.4	166.3
Salmon, Coho	39.3	109.7
Tuna, Albacore	26.7	80.1
Clams, Soft-shelled	47.5	150.7
Clams, Blue Hard-shelled	6.6	20.3
Lobsters, american	94.7	189.7
Oysters, eastern	61.1	131.3
Sea Scallops	135.6	404.2
Shrimp	47.6	149

Resultats de la regressió

```
lm(formula = Price_1980 ~ Price_1970)
```

Coefficients:

	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	-1.2338	11.2579	-0.11	0.915
Price_1970	2.7016	0.2053	13.16	1.72e-08

Residual standard error: 27.88 on 12 degrees of freedom

Multiple R-Squared: 0.9352.

Gràfica de la regressió

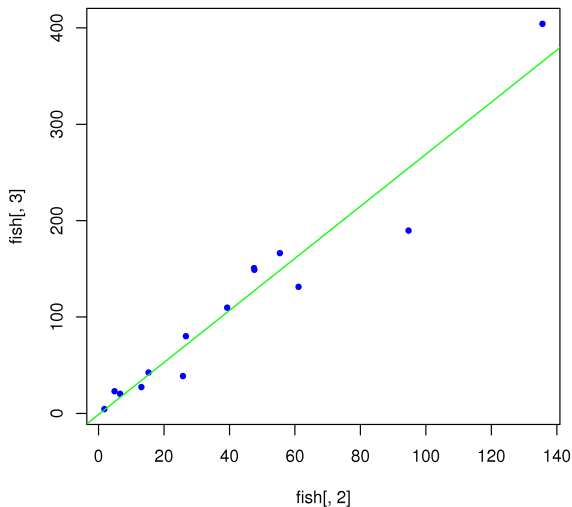
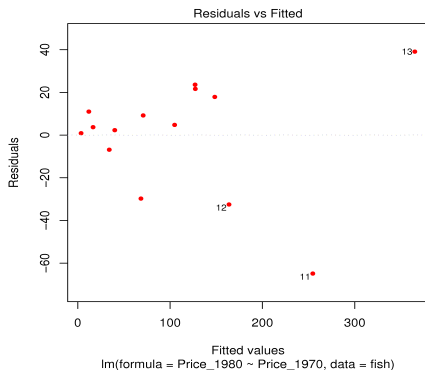


Diagrama de residus



Els residus augmenten amb els valors ajustats.

Un procediment per a corregir aquest efecte pot ser transformar les dades.

Transformació logarítmica: Resultats de la regressió

Una transformació amb bones propietats és la logarítmica (logaritme de la resposta sobre el logaritme de la variable explicativa). Transforma una relació potencial, $y = a x^b$, en una relació lineal $\log y = \log a + b \log x$.

```
lm(formula = log80 ~ log70)
```

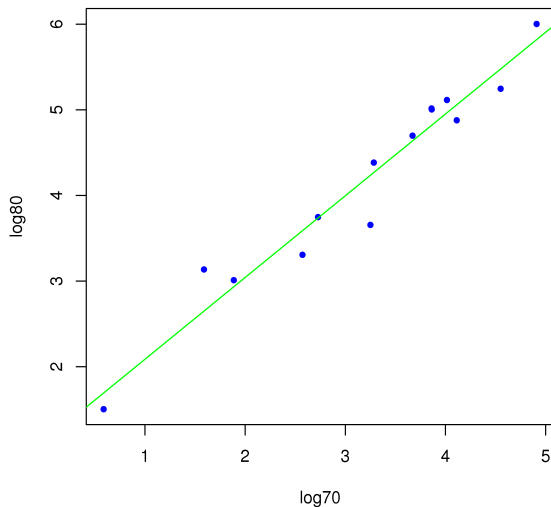
Coefficients:

	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	1.13231	0.21664	5.227	0.000212
log70	0.95465	0.06349	15.036	3.78e-09

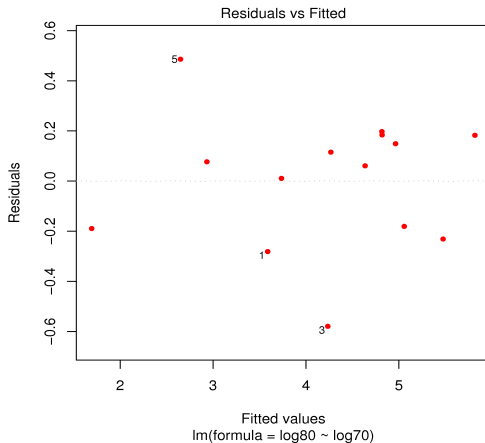
Residual standard error: 0.2776 on 12 degrees of freedom

Multiple R-Squared: 0.9496.

Transformació logarítmica: Gràfica de la regressió



Transformació logarítmica: Diagrama de residus



Ara els residus no presenten regularitat.