Erregresioaren analisia

Xekiko Yren erregresio lineal anizkoitza

- *Y*ren erregresio ez bada nahiko ona, saiatu aldagaien eraldaketa edo transformazioak (polinomiala, lerromakurra), bai *Y*rena baita X_j-ena ere.
- *Y*ren erregresioa ez bada nahiko ona, saiatu X_j aldagaien elkarreraginak ereduan txertatzen.
- *Y*ren erregresioa aztertu bitarra denean, baita X_j aldagaiak bitarrak direnean ere.

*X*ekiko *Y*ren erregresio lineal anizkoitza

- *Y*ren erregresio ez bada nahiko ona, saiatu aldagaien eraldaketa edo transformazioak (polinomiala, lerromakurra), bai *Y*rena baita X_j-ena ere.
- *Y*ren erregresioa ez bada nahiko ona, saiatu X_j aldagaien elkarreraginak ereduan txertatzen.
- *Y*ren erregresioa aztertu bitarra denean, baita X_j aldagaiak bitarrak direnean ere.

*X*ekiko *Y*ren erregresio ez-lineala

- *Y* menpeko aldagaiaren erregresioa aldagai askeen konbinazio lineala ez den funtzio baten bitartez egiten da:

$$Y=f(X_1,X_2,\ldots,X_p)+E$$

*X*ekiko *Y*ren erregresio ez-lineala

- *Y* menpeko aldagaiaren erregresioa aldagai askeen konbinazio lineala ez den funtzio baten bitartez egiten da:

$$Y=f(X_1,X_2,\ldots,X_p)+E$$

- Datuen doiketa hurrenez-hurreneko hurbilketen bitartez egiten da. Funtzioa *Taylor-en serie*aren bitartez garatu egiten da, polinomiala bihurtzen da, eta erregresio linealaren teknika aplikatzen zaio.

*X*ekiko *Y*ren erregresio ez-lineala

•*Y* menpeko aldagaiaren erregresioa aldagai askeen konbinazio lineala ez den funtzio baten bitartez egiten da:

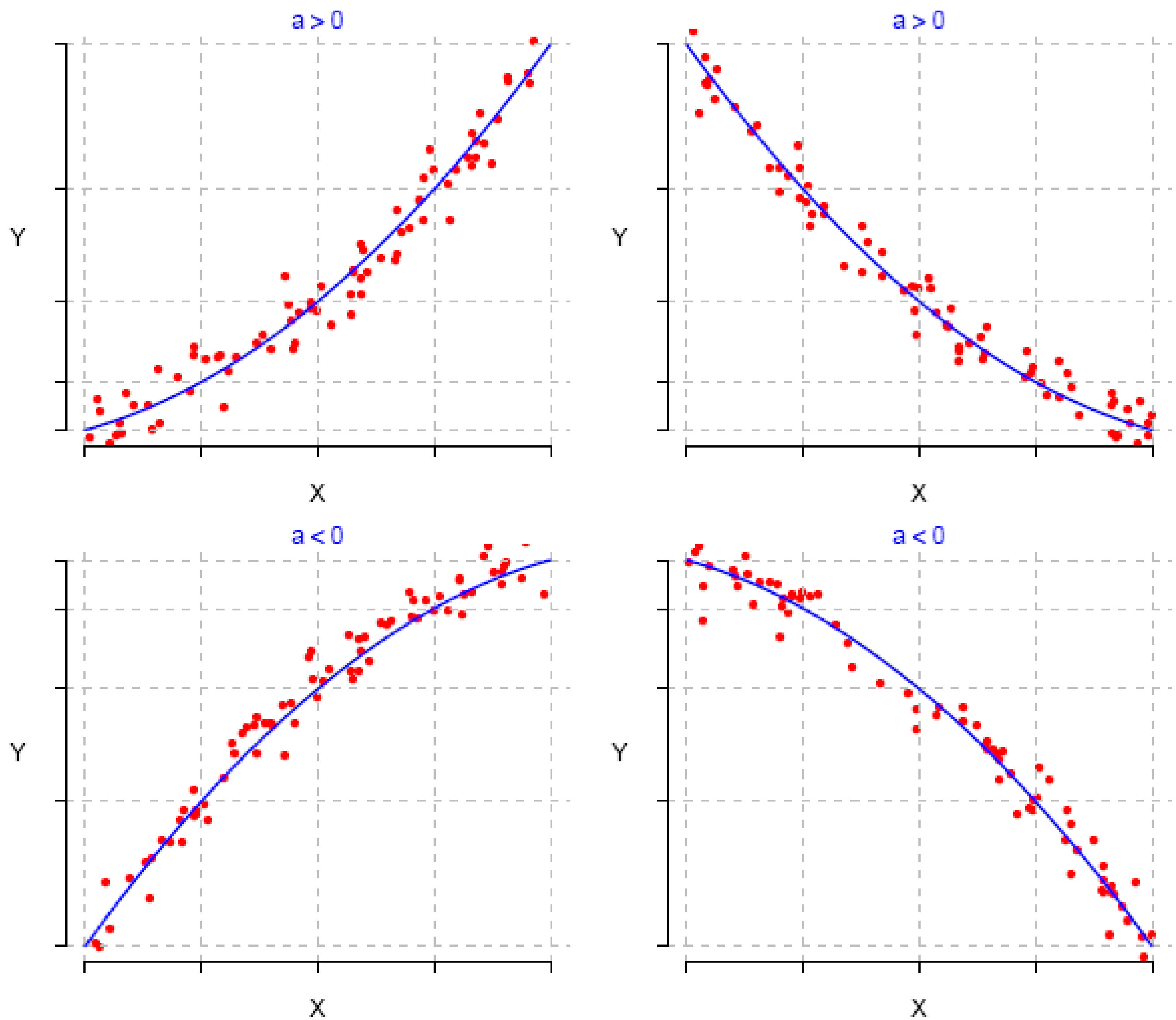
$$Y=f(X_1,X_2,\ldots,X_p)+E$$

•Datuen doiketa hurrenez-hurreneko hurbilketen bitartez egiten da. Funtzioa *Taylor-en serie*aren bitartez garatu egiten da, polinomiala bihurtzen da, eta erregresio linealaren teknika aplikatzen zaio.

•Funtzio batzuren erregresioa, berriz, *esponentziala* eta *logaritmikoa* kasu, eraldatu daitezke erregresioa lineala izan dadin. Orduan erregresio linealaren teknika aplika daiteke zuzenean, baina tentuz ibili behar da interpretazioaren garaian.

*X*ekiko *Y*ren erregresio karratikoa

$$Y = aX^2 + bX + c + E$$



*X*ekiko *Y*ren erregresio karratikoa

http://lib.stat.cmu.edu/DASL/Datafiles/tvadsdat.html

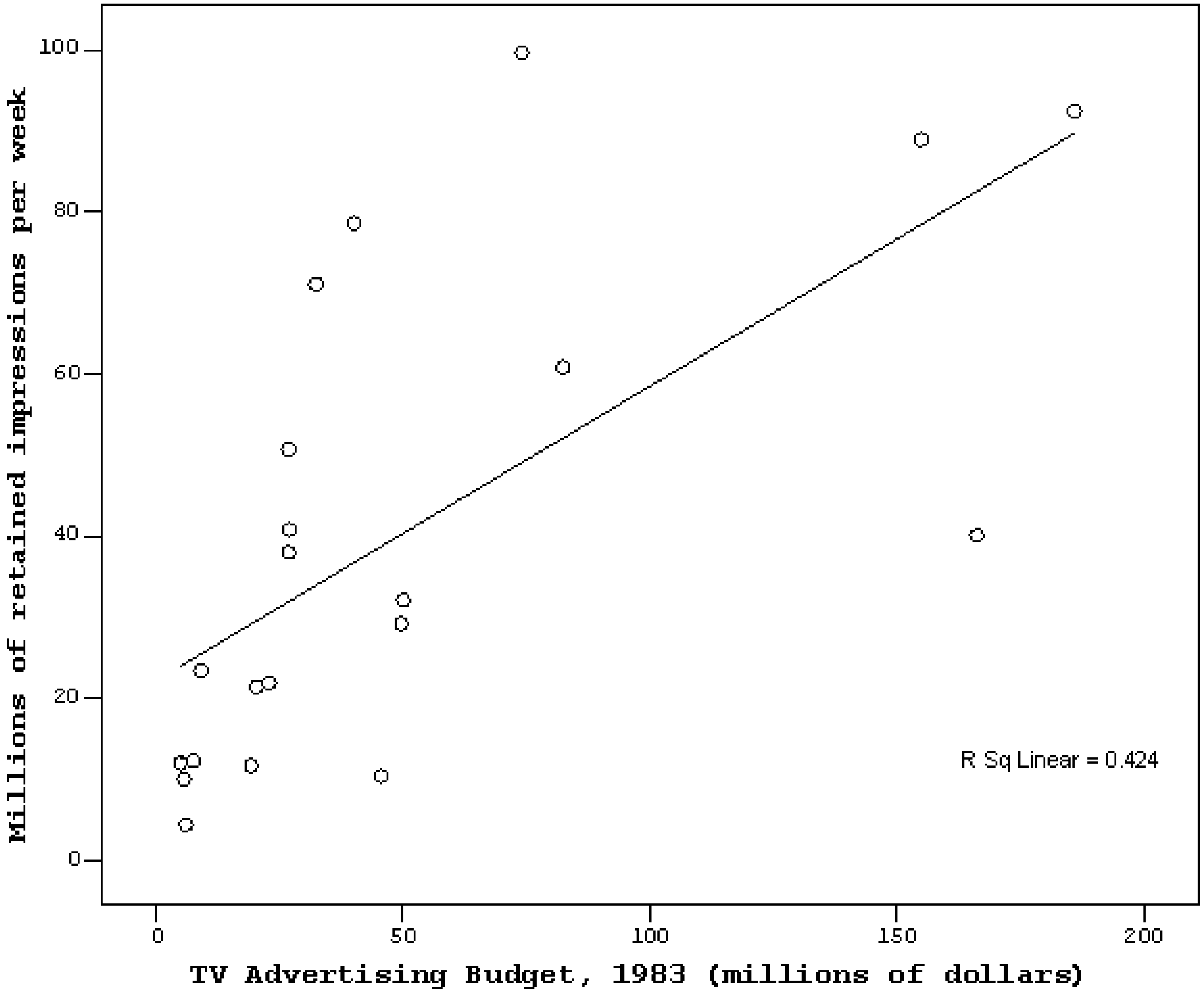
Reference: Wall Street Journal, March 1, 1984

Variable Names:

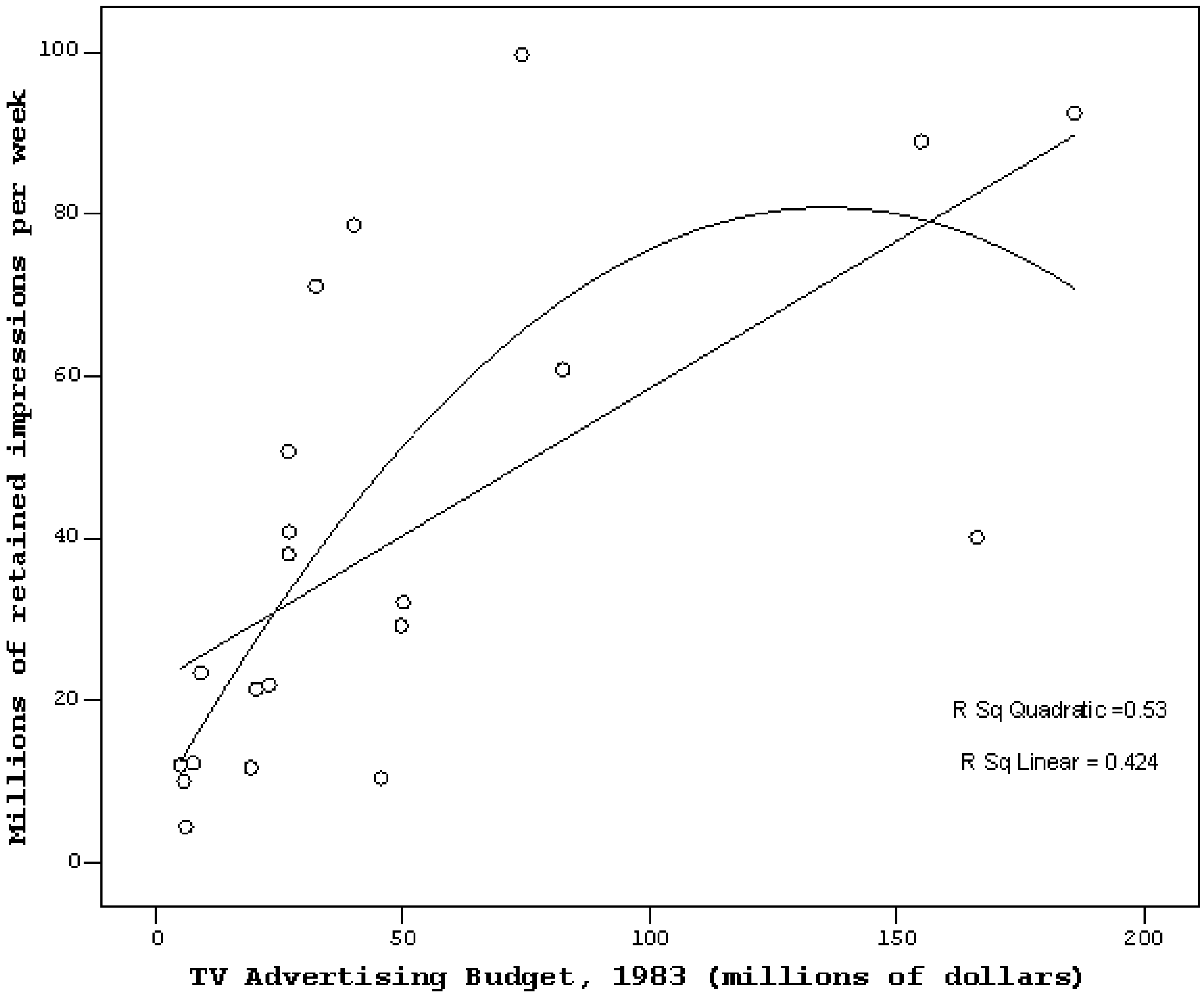
1. FIRM: Firm name
2. SPEND: TV advertising budget, 1983 ($ millions)
3. MILIMP: Millions of retained impressions per week

FIRM	SPEND	MILIMP
MILLER LITE	50.1	32.1
PEPSI	74.1	99.6
STROH'S	19.3	11.7
FED'L EXPRESS	22.9	21.9
BURGER KING	82.4	60.8
COCO-COLA	40.1	78.6
MC DONALD'S	185.9	92.4
MCI	26.9	50.7
DIET COLA	20.4	21.4
FORD	166.2	40.1
LEVI'S	27.0	40.8
BUD LITE	45.6	10.4
ATT/BELL	154.9	88.9
CALVIN KLEIN	5.0	12.0
WENDY'S	49.7	29.2
POLAROID	26.9	38.0
SHASTA	5.7	10.0
MEOW MIX	7.6	12.3
OSCAR MEYER	9.2	23.4
CREST	32.4	71.1
KIBBLES 'N BITS	6.1	4.4

*X*ekiko *Y*ren erregresio karratikoa



*X*ekiko *Y*ren erregresio karratikoa



*X*ekiko *Y*ren erregresio karratikoa

$$\boldsymbol{Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E}$$

$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2 + E$$

$$Y = c + b \cdot X + a \cdot X^2 + E$$

$$\mathrm{Min}_{a,b,c}\, G(a,b,c) = \sum_{i=1}^{4} e_i^2 = \sum_{i=1}^{4} \left(y_i - \left(a \cdot x_i^2 + b \cdot x_i + c \right) \right)^2$$

*X*ekiko *Y*ren erregresio karratikoa

$$\boldsymbol{Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E}$$

$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2 + E$$

$$Y = c + b \cdot X + a \cdot X^2 + E$$

$$\text{Min}_{a,b,c}\, G(a,b,c) = \sum_{i=1}^{4} e_i^2 = \sum_{i=1}^{4} \left(y_i - \left(a \cdot x_i^2 + b \cdot x_i + c \right) \right)^2$$

Ariketa: Bilatu datozen datuentzat

X	-1	1	0	2
Y	0	2	-1	1

akats (*E*) karratuen batura txikiena egiten duen funtzio karratikoa.

*X*ekiko *Y*ren erregresio karratikoa

$$G(a,b,c)=\sum_{i=1}^{4}e_i^2=\sum_{i=1}^{4}\left(y_i-\left(a\cdot x_i^2+b\cdot x_i+c\right)\right)^2$$

$$\begin{aligned}
G(a,b,c) &= (a(-1)^2+b(-1)+c-0)^2+(a(1)^2+b(1)+c-2)^2\\
&\quad +(a(0)^2+b(0)+c-(-1))^2+(a(2)^2+b(2)+c-1)^2\\
&= (a-b+c)^2+(a+b+c-2)^2\\
&\quad +(c+1)^2+(4a+2b+c-1)^2
\end{aligned}$$

*X*ekiko *Y*ren erregresio karratikoa

$$G(a,b,c) = \sum_{i=1}^{4} e_i^2 = \sum_{i=1}^{4} \left(y_i - \left(a \cdot x_i^2 + b \cdot x_i + c \right) \right)^2$$

$$\begin{aligned} G(a,b,c) &= (a(-1)^2 + b(-1) + c - 0)^2 + (a(1)^2 + b(1) + c - 2)^2 \\ &\quad + (a(0)^2 + b(0) + c - (-1))^2 + (a(2)^2 + b(2) + c - 1)^2 \\ &= (a - b + c)^2 + (a + b + c - 2)^2 \\ &\quad + (c + 1)^2 + (4a + 2b + c - 1)^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial S}{\partial a} &= 2[(a - b + c) + (a + b + c - 2) + 4(4a + 2b + c - 1)] \\ &= 2[18a + 8b + 6c - 6] \\ \frac{\partial S}{\partial b} &= 2[(-1)(a - b + c) + (a + b + c - 2) + 2(4a + 2b + c - 1)] \\ &= 2[8a + 6b + 2c - 4] \\ \frac{\partial S}{\partial c} &= 2[(a - b + c) + (a + b + c - 2) + (c + 1) + (4a + 2b + c - 1)] \\ &= 2[6a + 2b + 4c - 2] \end{aligned}$$

$$\frac{\partial S}{\partial a} = 0 \quad \frac{\partial S}{\partial b} = 0 \quad \frac{\partial S}{\partial c} = 0$$

*X*ekiko *Y*ren erregresio karratikoa

$$G(a,b,c)=\sum_{i=1}^{4} e_i^2=\sum_{i=1}^{4}\left(y_i-\left(a\cdot x_i^2+b\cdot x_i+c\right)\right)^2$$

$$\begin{aligned} G(a,b,c) &= (a(-1)^2+b(-1)+c-0)^2+(a(1)^2+b(1)+c-2)^2 \\ &\quad +(a(0)^2+b(0)+c-(-1))^2+(a(2)^2+b(2)+c-1)^2 \\ &= (a-b+c)^2+(a+b+c-2)^2 \\ &\quad +(c+1)^2+(4a+2b+c-1)^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial S}{\partial a} &= 2[(a-b+c)+(a+b+c-2)+4(4a+2b+c-1)] \\ &= 2[18a+8b+6c-6] \\ \frac{\partial S}{\partial b} &= 2[(-1)(a-b+c)+(a+b+c-2)+2(4a+2b+c-1)] \\ &= 2[8a+6b+2c-4] \\ \frac{\partial S}{\partial c} &= 2[(a-b+c)+(a+b+c-2)+(c+1)+(4a+2b+c-1)] \\ &= 2[6a+2b+4c-2] \end{aligned}$$

$$\frac{\partial S}{\partial a}=0 \quad \frac{\partial S}{\partial b}=0 \quad \frac{\partial S}{\partial c}=0$$

$$a=0,\ b=\tfrac{3}{5},\ c=\tfrac{1}{5}$$

*X*ekiko *Y*ren erregresio karratikoa

$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2 + E$$

$$G(a_0, a_1, a_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \left(a_2 \cdot x_i^2 + a_1 \cdot x_i + a_0 \right) \right)^2$$

<u>Ekuazio linealak</u>:

$$a_0 \cdot \sum_{i=1}^{n} x_i^0 + a_1 \cdot \sum_{i=1}^{n} x_i^1 + a_2 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$a_0 \cdot \sum_{i=1}^{n} x_i^1 + a_1 \cdot \sum_{i=1}^{n} x_i^2 + a_2 \cdot \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

$$a_0 \cdot \sum_{i=1}^{n} x_i^2 + a_1 \cdot \sum_{i=1}^{n} x_i^3 + a_2 \cdot \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i$$

*X*ekiko *Y*ren erregresio karratikoa

Ekuazio linealak:

$$c\cdot\sum_{i=1}^{n} x_i^0 + b\cdot\sum_{i=1}^{n} x_i^1 + a\cdot\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$c\cdot\sum_{i=1}^{n} x_i^1 + b\cdot\sum_{i=1}^{n} x_i^2 + a\cdot\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

$$c\cdot\sum_{i=1}^{n} x_i^2 + b\cdot\sum_{i=1}^{n} x_i^3 + a\cdot\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i$$

Soluzioa:

$$c\cdot 4 + b\cdot 2 + a\cdot 6 = 2$$

$$c\cdot 2 + b\cdot 6 + a\cdot 8 = 4$$

$$c\cdot 6 + b\cdot 8 + a\cdot 18 = 6$$

$$\boxed{a = 0, b = \frac{3}{5}, c = \frac{1}{5}}$$

*X*ekiko *Y*ren erregresio karratikoa

$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2 + E$$

$$G(a_0, a_1, a_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \left(a_2 \cdot x_i^2 + a_1 \cdot x_i + a_0 \right) \right)^2$$

Ekuazio linealak:

$$\boldsymbol{X^T \cdot X \cdot a = X^T \cdot Y}$$

$$\boldsymbol{a = (X^T \cdot X)^{-1} \; X^T \cdot Y}$$

Soluzioa:

$$a = \left(\begin{pmatrix} +1, 1, 1, 1 \\ -1, 1, 0, 2 \\ +1, 1, 0, 4 \end{pmatrix} \begin{pmatrix} 1, -1, 1 \\ 1, +1, 1 \\ 1, +0, 0 \\ 1, +2, 4 \end{pmatrix} \right)^{-1} \begin{pmatrix} +1, 1, 1, 1 \\ -1, 1, 0, 2 \\ +1, 1, 0, 4 \end{pmatrix} \begin{pmatrix} +0 \\ +2 \\ -1 \\ +1 \end{pmatrix}$$

*X*ekiko *Y*ren erregresio karratikoa

$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2 + E$$

$$G(a_0, a_1, a_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \left(a_2 \cdot x_i^2 + a_1 \cdot x_i + a_0 \right) \right)^2$$

Ekuazio linealak:

$$\boldsymbol{X^T \cdot X \cdot a = X^T \cdot Y}$$

$$\boldsymbol{a = (X^T \cdot X)^{-1} \; X^T \cdot Y}$$

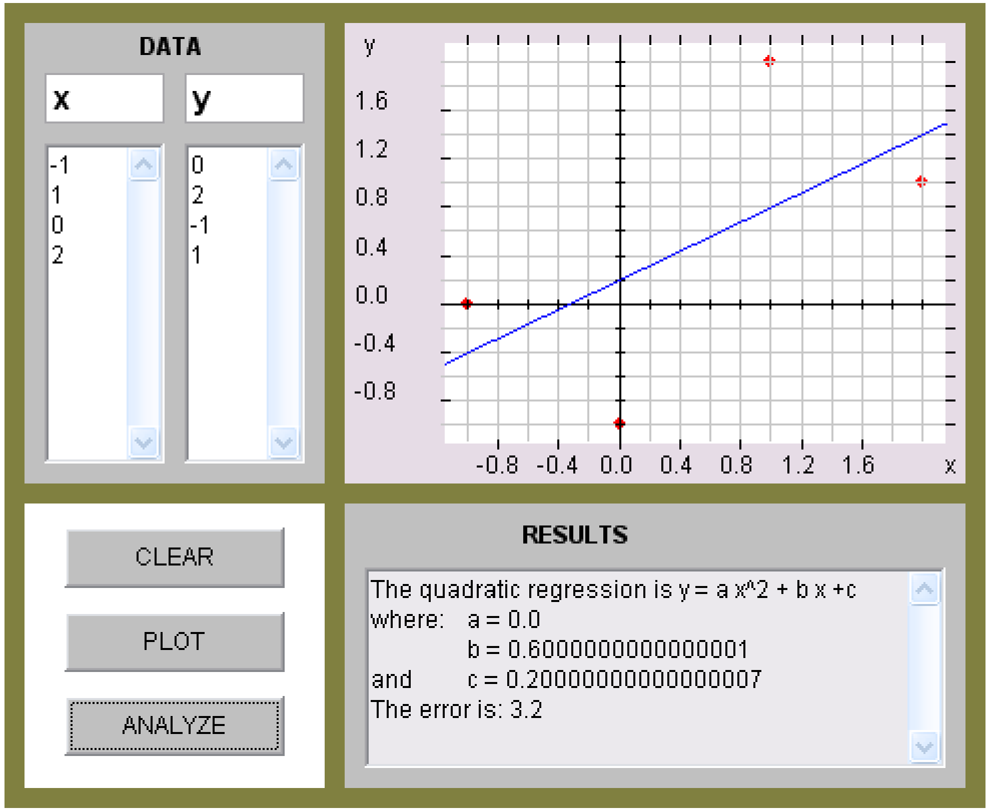
Soluzioa:

$$a = \begin{pmatrix} 4,2,6 \\ 2,6,8 \\ 6,8,18 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \frac{1}{80} \begin{pmatrix} +44,+12,-20 \\ +12,+36,-20 \\ -20,-20,+20 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \frac{1}{80} \begin{pmatrix} 16 \\ 48 \\ 0 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$$

*X*ekiko *Y*ren erregresio karratikoa

http://earthmath.kennesaw.edu/main_site/tool_chest/quad_reg.htm

QUADRATIC REGRESSION APPLET



The error is obtained with the formula $\sum_{i=1}^{n} |y_i - y_i'|$

*X*ekiko *Y*ren erregresio karratikoa

$$Y = c + b \cdot X + a \cdot X^2 + E$$

```
> x <- c(-1,1,0,2)
> y <- c(0,2,-1,1)
> x2 <- x^2
> Qregr <- lm(y ~ x+x2)
> Qregr$coefficients
 (Intercept)            x           x2
 2.00000e-01  6.00000e-01 -7.77373e-17
> round(var(Qregr$fitted.values)/var(y),2)
[1] 0.36
```

$$Y = 2 + 0.6 \cdot X + E$$

*X*ekiko *Y*ren erregresio karratikoa

$$Y = c + b \cdot X + a \cdot X^2 + E$$

```
> x <- c(-1,1,0,2)
> y <- c(0,2,-1,1)
> x2 <- x^2
> Qregr <- lm(y ~ x+x2)
> Qregr$coefficients
 (Intercept)            x            x2
 2.00000e-01  6.00000e-01 -7.77373e-17
> round(var(Qregr$fitted.values)/var(y),2)
[1] 0.36
```

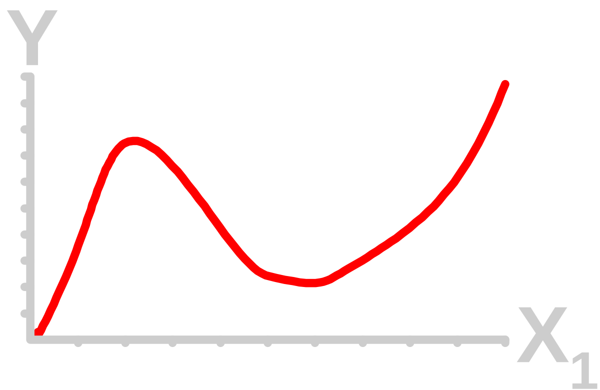
$$Y = 2 + 0.6 \cdot X + E$$

```
> Lregr <- lm(y ~ x)
> Lregr$coefficients
(Intercept)           x
        0.2         0.6
> round(cor(x,y)^2,digits=2) # round(var(Lregr$fitted.values)/var(y),2)
[1] 0.36
```

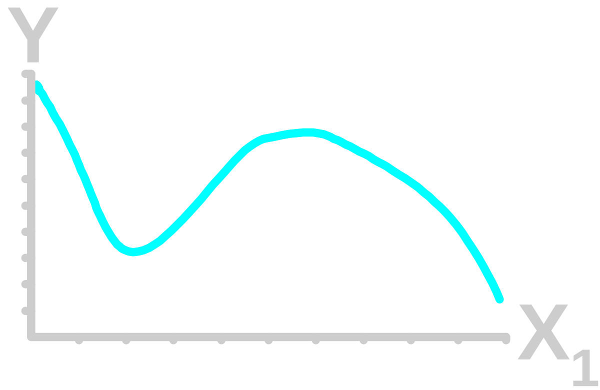
*X*ekiko *Y*ren erregresio kubikoa

$$Y = d + c \cdot X + b \cdot X^2 + a \cdot X^3 + E$$

a > 0



a < 0



*X*ekiko *Y*ren erregresio polinomikoa

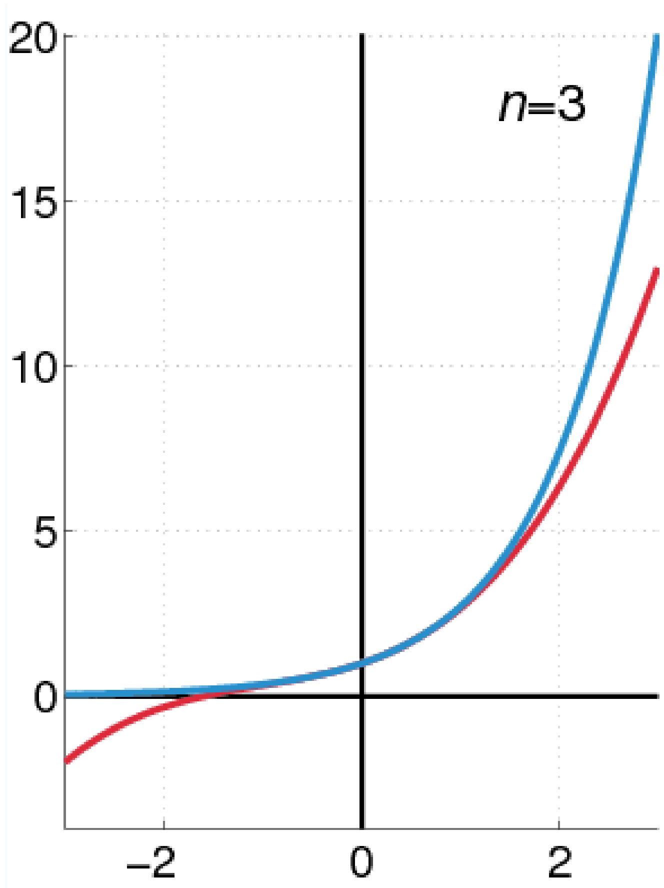
$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2 + ... + a_p \cdot X^p + E$$

*X*ekiko *Y*ren erregresio esponentziala

$$Y = b \cdot e^{a \cdot X} + E$$

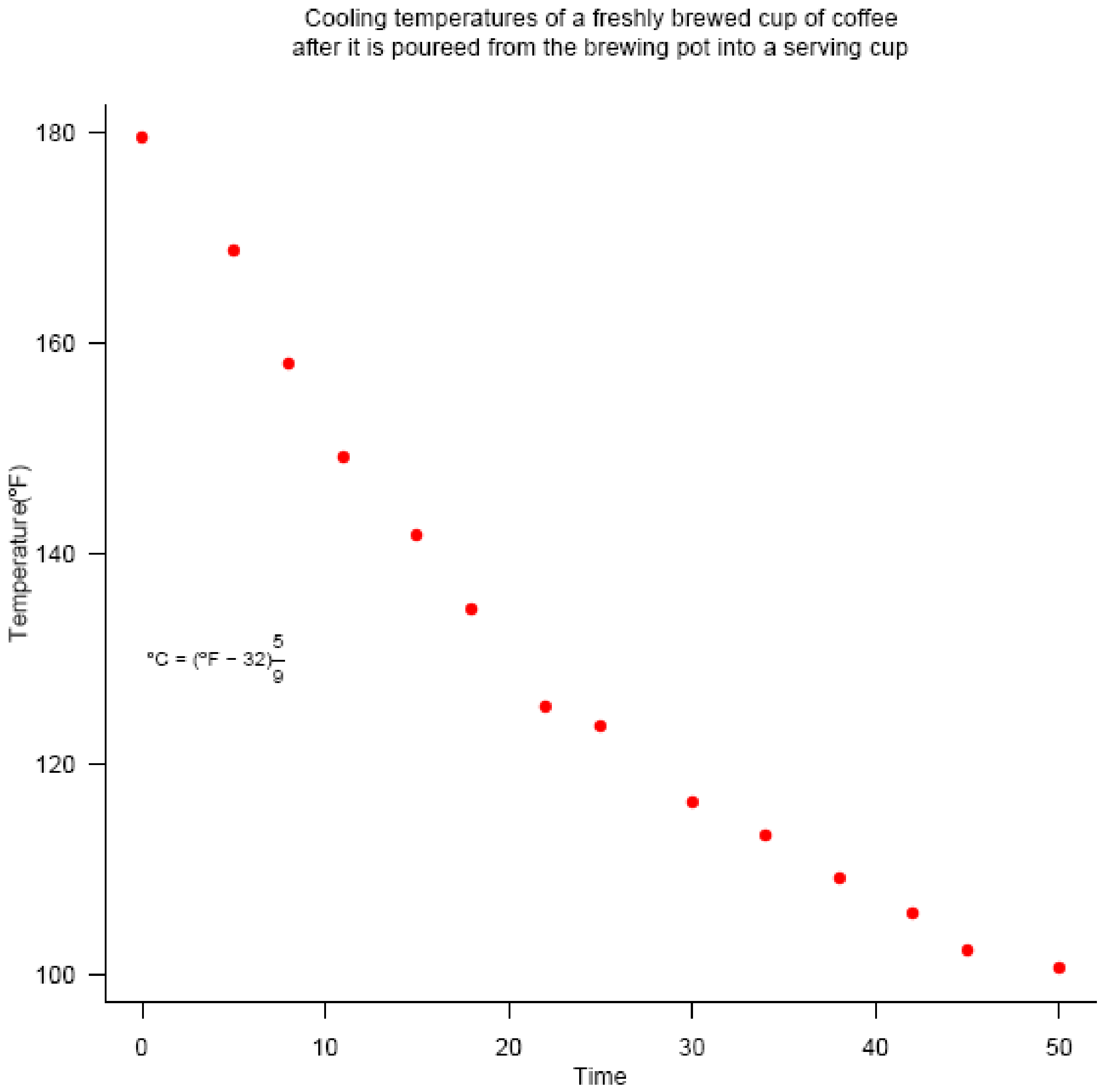
$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

http://en.wikipedia.org/wiki/File:Exp_series.gif



*X*ekiko *Y*ren erregresio esponentziala

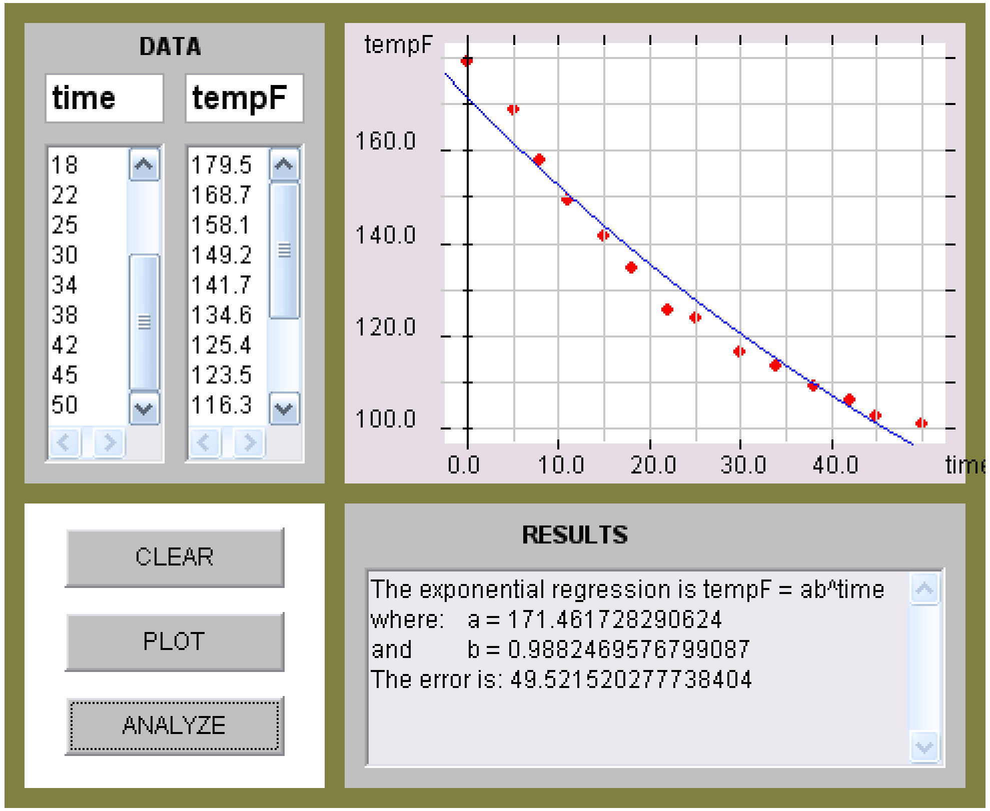
Time (mins)	Temp (° F)
0	179.5
5	168.7
8	158.1
11	149.2
15	141.7
18	134.6
22	125.4
25	123.5
30	116.3
34	113.2
38	109.1
42	105.7
45	102.2
50	100.5



*X*ekiko *Y*ren erregresio esponentziala

http://earthmath.kennesaw.edu/main_site/tool_chest/exp_reg.htm

EXPONENTIAL REGRESSION APPLET



The error is obtained with the formula $\sum_{i=1}^{n} |y_i - y_i'|$

*X*ekiko *Y*ren erregresio esponentziala

$$Y = b \cdot e^{a \cdot X} + E$$

$$Y' = b \cdot e^{a \cdot X} \qquad b^* = \ln b$$

$$\ln Y' \equiv Y'^* = b^* + a \cdot X$$

*X*ekiko *Y*ren erregresio esponentziala

$$Y = b \cdot e^{a \cdot X} + E$$

$$Y' = b \cdot e^{a \cdot X} \qquad b^{*} = \ln b$$

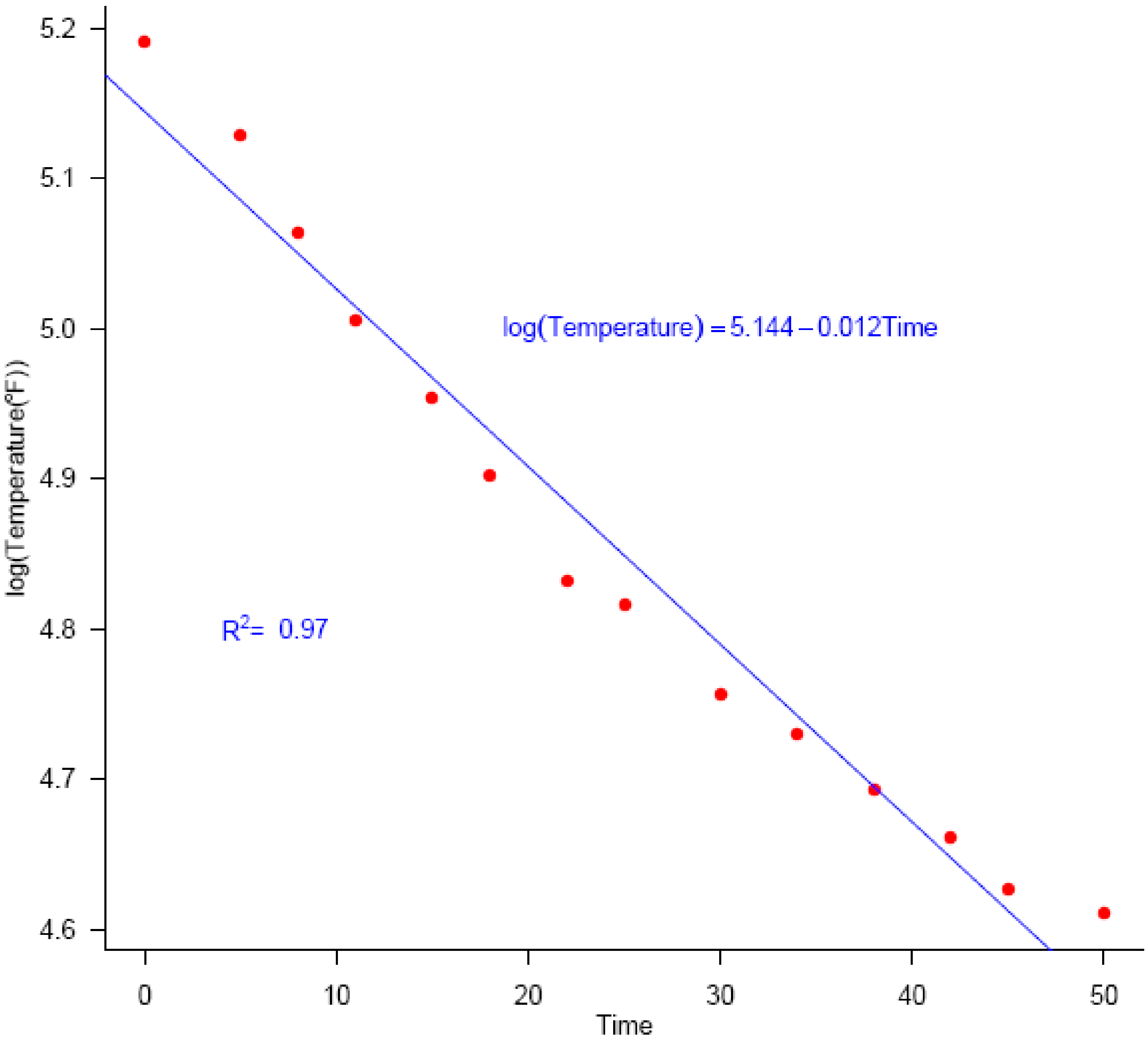
$$\ln Y' \equiv Y'^{*} = b^{*} + a \cdot X$$

```
> lm(tempf ~ time)$coefficients
(Intercept)        time
 168.861042   -1.563716
> cor(tempf,time)^2
[1] 0.9399495
>
> logtempf <- log(tempf)
> lm(logtempf ~ time)$coefficients
(Intercept)        time
 5.14436008 -0.01182266
> cor(logtempf,time)^2
[1] 0.9701377
```

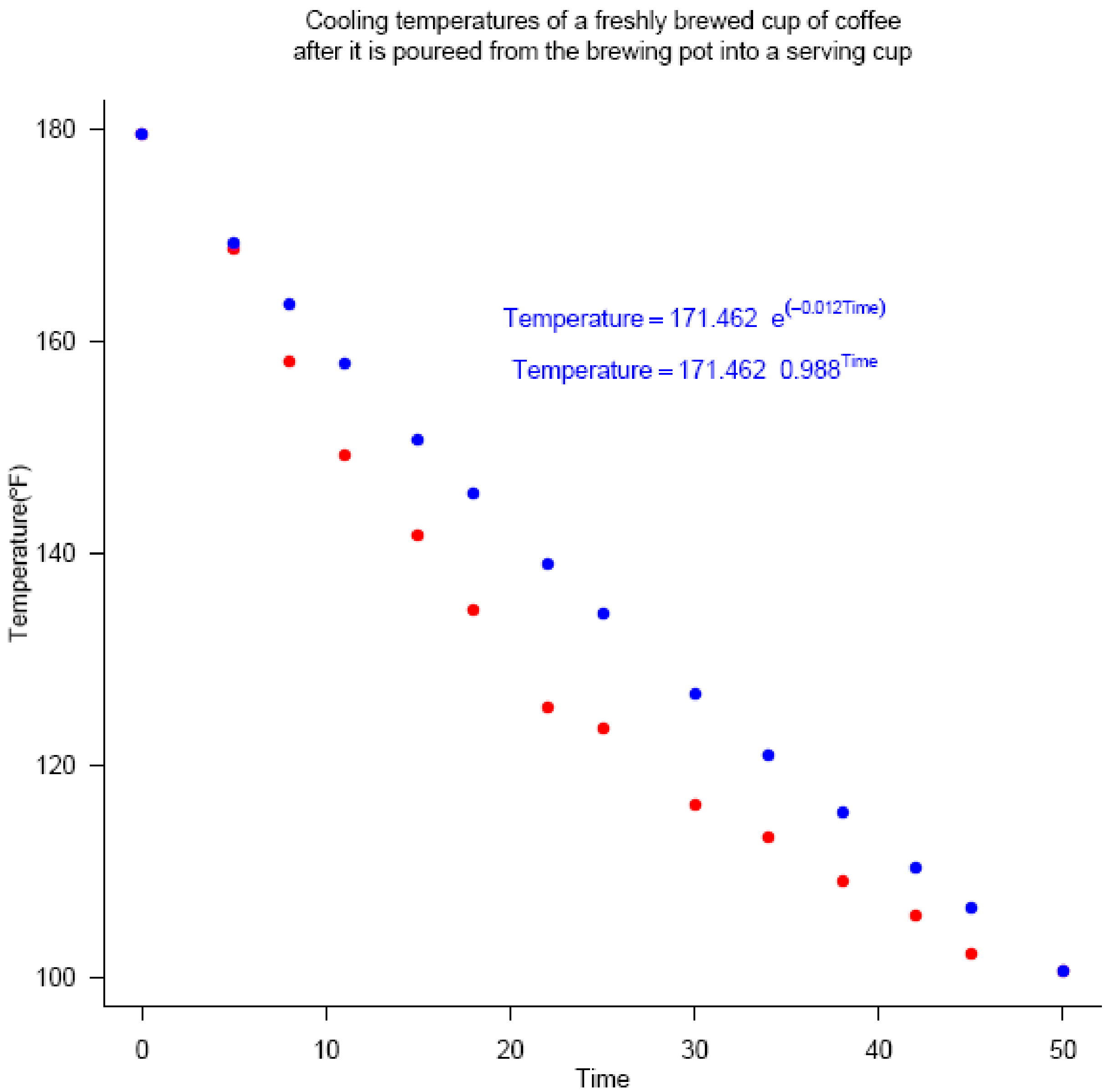
$$\ln Y' = 5.14 - 0.012 \cdot X$$

$$Y' = 171.462 + \mathrm{e}^{-0.012 \cdot X} \equiv 171.462 + 0.988^{X}$$

*X*ekiko *Y*ren erregresio esponentziala



*X*ekiko *Y*ren erregresio esponentziala



*X*ekiko *Y*ren erregresio esponentziala

```
> tempprime <- exp(lm(logtempf ~ time)$coefficients[[1]])*
+              exp(lm(logtempf ~ time)$coefficients[[2]]*time)
> tempf; round(tempprime,1)
 [1] 179.5 168.7 158.1 149.2 141.7 134.6 125.4 123.5 116.3 113.2 109.1 105.7
[13] 102.2 100.5
 [1] 171.5 161.6 156.0 150.6 143.6 138.6 132.2 127.6 120.3 114.7 109.4 104.4
[13] 100.7  94.9
> sum(abs(tempf-tempprime))
 [1] 49.52152
> var(tempprime)/var(tempf); var(tempf-tempprime)/var(tempf)
 [1] 0.8956339
 [1] 0.03027372
```

*X*ekiko *Y*ren erregresio esponentziala

```
> tempprime <- exp(lm(logtempf ~ time)$coefficients[[1]])*
+              exp(lm(logtempf ~ time)$coefficients[[2]]*time)
> tempf; round(tempprime,1)
 [1] 179.5 168.7 158.1 149.2 141.7 134.6 125.4 123.5 116.3 113.2 109.1 105.7
[13] 102.2 100.5
 [1] 171.5 161.6 156.0 150.6 143.6 138.6 132.2 127.6 120.3 114.7 109.4 104.4
[13] 100.7  94.9
> sum(abs(tempf-tempprime))
 [1] 49.52152
> var(tempprime)/var(tempf); var(tempf-tempprime)/var(tempf)
 [1] 0.8956339
 [1] 0.03027372
```

$$s_y^2 \neq s_{y'}^2 + s_e^2$$

```
ExpReg
 y=a*b^x
 a=171.4617283
 b=.9882469577
 r²=.9701377262
 r=-.9849556976
```

The exponential regression equation is

$y = (171.462) \cdot 0.988^x$

The correlation coefficient, r, is -.9849556976 which places the correlation into the "strong" category. (0.8 or greater is a "strong" correlation)

The coefficient of determination, r^2, is .9701377262 which means that 97% of the total variation in y can be explained by the relationship between x and y.

Yes, it is a very "good fit".

RESULTS

The exponential regression is tempF = ab^time
where: a = 171.461728290624
and b = 0.9882469576799087
The error is: 49.521520277738404

*X*ekiko *Y*ren erregresio esponentziala

```
> tempprime <- exp(lm(logtempf ~ time)$coefficients[[1]])*
+              exp(lm(logtempf ~ time)$coefficients[[2]]*time)
> tempf; round(tempprime,1)
 [1] 179.5 168.7 158.1 149.2 141.7 134.6 125.4 123.5 116.3 113.2 109.1 105.7
[13] 102.2 100.5
 [1] 171.5 161.6 156.0 150.6 143.6 138.6 132.2 127.6 120.3 114.7 109.4 104.4
[13] 100.7  94.9
> sum(abs(tempf-tempprime))
 [1] 49.52152
> var(tempprime)/var(tempf); var(tempf-tempprime)/var(tempf)
 [1] 0.8956339
 [1] 0.03027372
```

$$s_y^2 \neq s_{y'}^2 + s_e^2$$

```
ExpReg
 y=a*b^x
 a=171.4617283
 b=.9882469577
 r²=.9701377262
 r=-.9849556976
```

The exponential regression equation is

$y = (171.462)\cdot 0.988^x$

The correlation coefficient, r, is -.9849556976 which places the correlation into the "strong" category. (0.8 or greater is a "strong" correlation)

The coefficient of determination, r^2, is .9701377262 which means that 97% of the total variation in y can be explained by the relationship between x and y.

Yes, it is a very "good fit". **!!**

RESULTS

The exponential regression is tempF = ab^time
where: a = 171.461728290624
and b = 0.9882469576799087
The error is: 49.521520277738404

*X*ekiko *Y*ren erregresio logaritmikoa

$$Y = b + a \cdot \ln X + E$$

*X*ekiko *Y*ren berreketa-erregresioa

$$Y = b \cdot X^{a} + E$$

$$\ln Y \equiv Y^{*} = b^{*} + a \cdot \ln X + E^{*}$$

Xekiko Yren erregresio lineal anizkoitza

- Yren erregresio ez bada nahiko ona, saiatu aldagaien eraldaketa edo transformazioak (polinomiala, lerromakurra), bai Yrena baita X_j-ena ere.
- Yren erregresioa ez bada nahiko ona, saiatu X_j aldagaien elkarreraginak ereduan txertatzen.
- Yren erregresioa aztertu bitarra denean, baita X_j aldagaiak bitarrak direnean ere.

*X*ekiko *Y*ren erregresio lineal anizkoitza

elkarreragina

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_{12} \cdot X_1 \cdot X_2 + E$$

Ω	Y	X_1	X_2	$X_1 X_2$
ω_1	1	1	3	3
ω_2	4	8	5	40
ω_3	1	3	2	6
ω_4	3	5	6	30
…	…	…	…	…

Yren erregresio linealaren a_0, a_1, a_2 eta a_{12} ohi bezala kalkulatu

*X*ekiko *Y*ren erregresio lineal anizkoitza

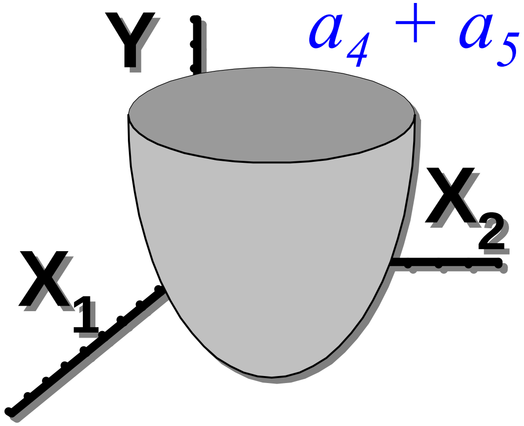
elkarreragina

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_1 \cdot X_2 + a_4 \cdot X_1^2 + a_5 \cdot X_2^2 + E$$

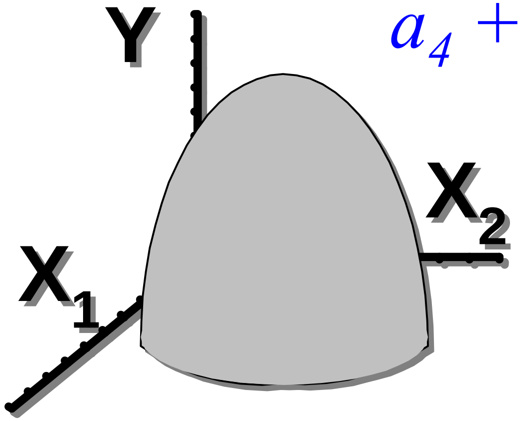
*X*ekiko *Y*ren erregresio lineal anizkoitza

elkarreragina

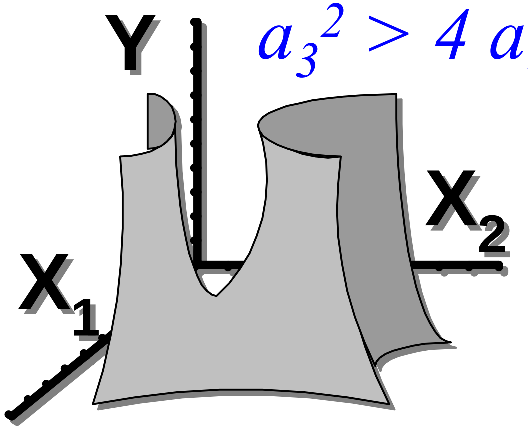
$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_1 \cdot X_2 + a_4 \cdot X_1^2 + a_5 \cdot X_2^2 + E$$



$a_4 + a_5 > 0$



$a_4 + a_5 > 0$



$a_3^2 > 4\, a_4\, a_5$

*X*ekiko *Y*ren erregresio lineal anizkoitza

- *Y*ren erregresio ez bada nahiko ona, saiatu aldagaien eraldaketa edo transformazioak (polinomiala, lerromakurra), bai *Y*rena baita X_j-ena ere.

- *Y*ren erregresioa ez bada nahiko ona, saiatu X_j aldagaien elkarreraginak ereduan txertatzen.

- *Y*ren erregresioa aztertu bitarra denean, baita X_j aldagaiak bitarrak direnean ere.

***X*ekiko *Y*ren erregresio lineal bakuna**

Aldagai aske bat kualitatibo bitarra

$\boldsymbol{Y = b + a \cdot X + E}$

X aldagaia, kualitatibo bitarra da:

bati *1* zenbakia egokitu, eta besteari *0*a.

Dummy aldagaia esaten zaio.

(sasi-aldagaia, aldagaiarena egiten duena)

*X*ekiko *Y*ren erregresio lineal bakuna

Aldagai aske bat kualitatibo bitarra

$$Y = b + a \cdot X + E$$

Zenbait oztopo duen espazio batetik nabigatzeko diseinatu den **robot** baten **jokabidea** ebaluatu nahi da. Robotak informazioa jasotzen du sentsore ezberdinetatik, eta programa baten bitartez prozesatzen du. Funtsean robotaren jokabidea kontrolatzen duena **programa** bat da, eta bat baino gehiago burutu daiteke.

Esperimentu bat egin zen. **Bi programen** agindupean (**A** eta **B**) begiratu zen ea zenbat **denbora** (**X**) behar zuen robotak argi-igorgailu batera iristeko. Emaitzak ondorengoak izan ziren

A: 31 16 37 24 25 20 31 32 36 12

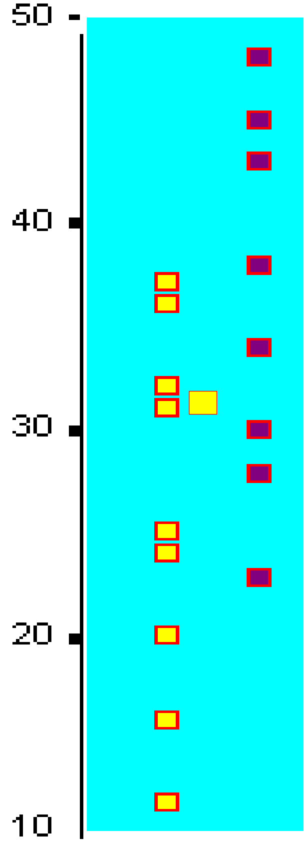
B: 34 28 43 23 38 48 45 30

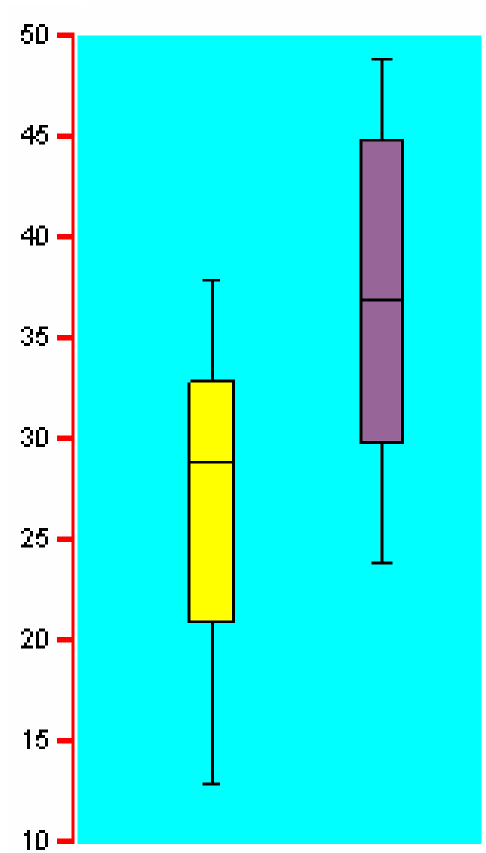
*X*ekiko *Y*ren erregresio lineal bakuna

Aldagai aske bat kualitatiboa

$$Y = b + a \cdot X + E$$

x	programa
48	B
45	B
43	B
38	B
37	A
36	A
34	B
32	A
31	A
31	A
30	B
28	B
25	A
24	A
23	B
20	A
16	A
12	A





*X*ekiko *Y*ren erregresio lineal bakuna

Aldagai aske bat kualitatibo bitarra

$$Y = b + a \cdot X + E$$

```
> denbora <- c(31,16,37,24,25,20,31,32,36,12,34,28,43,23,38,48,45,30)
> programa <- c(rep(1,10),rep(0,8))
> sum(denbora); sum(denbora[1:10]); sum(denbora[11:18])
[1] 553
[1] 264
[1] 289
> sum(denbora^2); sum(denbora[1:10]^2); sum(denbora[11:18]^2)
[1] 18603
[1] 7612
[1] 10991
>
```

Xekiko Yren erregresio lineal bakuna

Aldagai aske bat kualitatibo bitarra

$$Y = b + a \cdot X + E$$

```
> denbora <- c(31,16,37,24,25,20,31,32,36,12,34,28,43,23,38,48,45,30)
> programa <- c(rep(1,10),rep(0,8))
> sum(denbora); sum(denbora[1:10]); sum(denbora[11:18])
[1] 553
[1] 264
[1] 289
> sum(denbora^2); sum(denbora[1:10]^2); sum(denbora[11:18]^2)
[1] 18603
[1] 7612
[1] 10991
>
```

$$\bar{x} = \frac{10}{18} \qquad s_x^2 = \frac{10}{18}\frac{8}{18}$$

$$\bar{y} = \frac{553}{18} \qquad s_y^2 = \frac{18603}{18} - \left(\frac{553}{18}\right)^2$$

$$s_{xy} = \frac{264}{18} - \frac{10}{18}\frac{553}{18} = \frac{10}{18}\frac{8}{18}\left(\frac{264}{10} - \frac{289}{8}\right) \qquad (\bar{y}_1 - \bar{y}_0)$$

$$a = \frac{s_{xy}}{s_x^2} = \frac{264}{10} - \frac{289}{8} = -9.725$$

$$b = \frac{553}{18} + 0.975\frac{10}{18} = 36.125$$

$$Y' = 36.125 - 9.725 \cdot X$$

*X*ekiko *Y*ren erregresio lineal bakuna

Aldagai aske bat kualitatibo bitarra

$$Y = b + a \cdot X + E$$

```
> denbora <- c(31,16,37,24,25,20,31,32,36,12,34,28,43,23,38,48,45,30)
> programa <- c(rep(1,10),rep(0,8))
> sum(denbora); sum(denbora[1:10]); sum(denbora[11:18])
[1] 553
[1] 264
[1] 289
> sum(denbora^2); sum(denbora[1:10]^2); sum(denbora[11:18]^2)
[1] 18603
[1] 7612
[1] 10991
> cov(denbora,programa)*17/18
[1] -2.401235
> mean(denbora[1:10]); mean(denbora[11:18])
[1] 26.4
[1] 36.125
> var(denbora)*17/18
[1] 89.64506
> lm(denbora~programa)$coefficients
(Intercept)    programa
     36.125      -9.725
> var(lm(denbora~programa)$fitted.values)/var(denbora)
[1] 0.2604941
```

***X*ekiko *Y*ren erregresio lineal bakuna**

Aldagai aske bat kualitatibo bitarra

$$Y = b + a \cdot X + E$$

$$\boxed{Y = 36.125 - 9.725 \cdot X + E}$$

$y_i' = \overline{y_1} = 26.400 \qquad x_i = 1$

$y_i' = \overline{y_0} = 36.125 \qquad x_i = 0$

*X*ekiko *Y*ren erregresio lineal bakuna

Aldagai aske bat kualitatibo bitarra

$$Y = b + a \cdot X + E$$

$$Y = 36.125 - 9.725 \cdot X + E$$

$$y_i' = \overline{y_1} = 26.400 \qquad x_i = 1$$

$$y_i' = \overline{y_0} = 36.125 \qquad x_i = 0$$

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = p(1-p)\frac{(\overline{y}_1 - \overline{y}_0)^2}{s_y^2} = p(1-p)\left(\frac{\overline{y}_1 - \overline{y}}{s_y} - \frac{\overline{y}_0 - \overline{y}}{s_y}\right)^2 = \frac{10}{18}\frac{8}{18}\left(\frac{\frac{264}{10} - \frac{553}{18}}{9.47} - \frac{\frac{289}{10} - \frac{553}{18}}{9.47}\right)^2 = 0.26$$

*X*ekiko *Y*ren erregresio lineal bakuna
Aldagai aske bat kualitatibo bitarra

$$Y = b + a\cdot X + E$$

$$Y = 36.125 - 9.725\cdot X + E$$

$$y_i' = \overline{y_1} = 26.400 \leftarrow x_i = 1$$

$$y_i' = \overline{y_0} = 36.125 \leftarrow x_i = 0$$

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = p(1-p)\frac{(\bar{y}_1 - \bar{y}_0)^2}{s_y^2} = p(1-p)\left(\frac{\bar{y}_1 - \bar{y}}{s_y} - \frac{\bar{y}_0 - \bar{y}}{s_y}\right)^2 = \frac{10}{18}\frac{8}{18}\left(\frac{\frac{264}{10} - \frac{553}{18}}{9.47} - \frac{\frac{289}{10} - \frac{553}{18}}{9.47}\right)^2 = 0.26$$

Bariantz-analisia

Programa	n	$\bar{y}$	s^2		$\sum (y_i - \bar{y})^2$
Total	18	30.72	89.65		1613.61
A	10	26.40	64.24	642.41	1193.27
B	8	36.13	68.86	550.86	
A vs. B		**-9.73**	0.037		420.34

$$R^2 = \eta^2 = \frac{420.34}{1613.61} = 0.26$$

***X*ekiko *Y*ren erregresio lineal anizkoitza**

Aldagai aske bat kualitatibo bitarra

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E$$

X_2 aldagaia, kualitatibo bitarra da:

bati *1* zenbakia egokitu, eta besteari *0*a.

Dummy aldagaia esaten zaio.

(sasi-aldagaia, aldagaiarena egiten duena)

*X*ekiko *Y*ren erregresio lineal anizkoitza

Aldagai aske bat kualitatibo bitarra

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E$$

Ω	Y	X_1	X_2
ω_1	1	1	1
ω_2	4	8	0
ω_3	1	3	1
ω_4	3	5	1
…	…	…	…

*X*ekiko *Y*ren erregresio lineal anizkoitza

Aldagai aske bat kualitatibo bitarra

$$\boldsymbol{Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E}$$

$\boldsymbol{X_2}$ aldagaia, kualitatibo bitarra da:
bati *1* zenbakia egokitu, eta besteari *0*a.
Dummy aldagaia esaten zaio.
(sasi-aldagaia, aldagaiarena egiten duena)

Yren erregresio linealaren a_0, a_1, eta a_2 ohi bezala kalkulatuta:

$X_2(\omega_i)=1$ dutenentzako erregresio lineala hauxe da:

$$\boldsymbol{Y' = (a_0 + a_2) + a_1 \cdot X_1}$$

$X_2(\omega_i)=1$ dutenentzat:

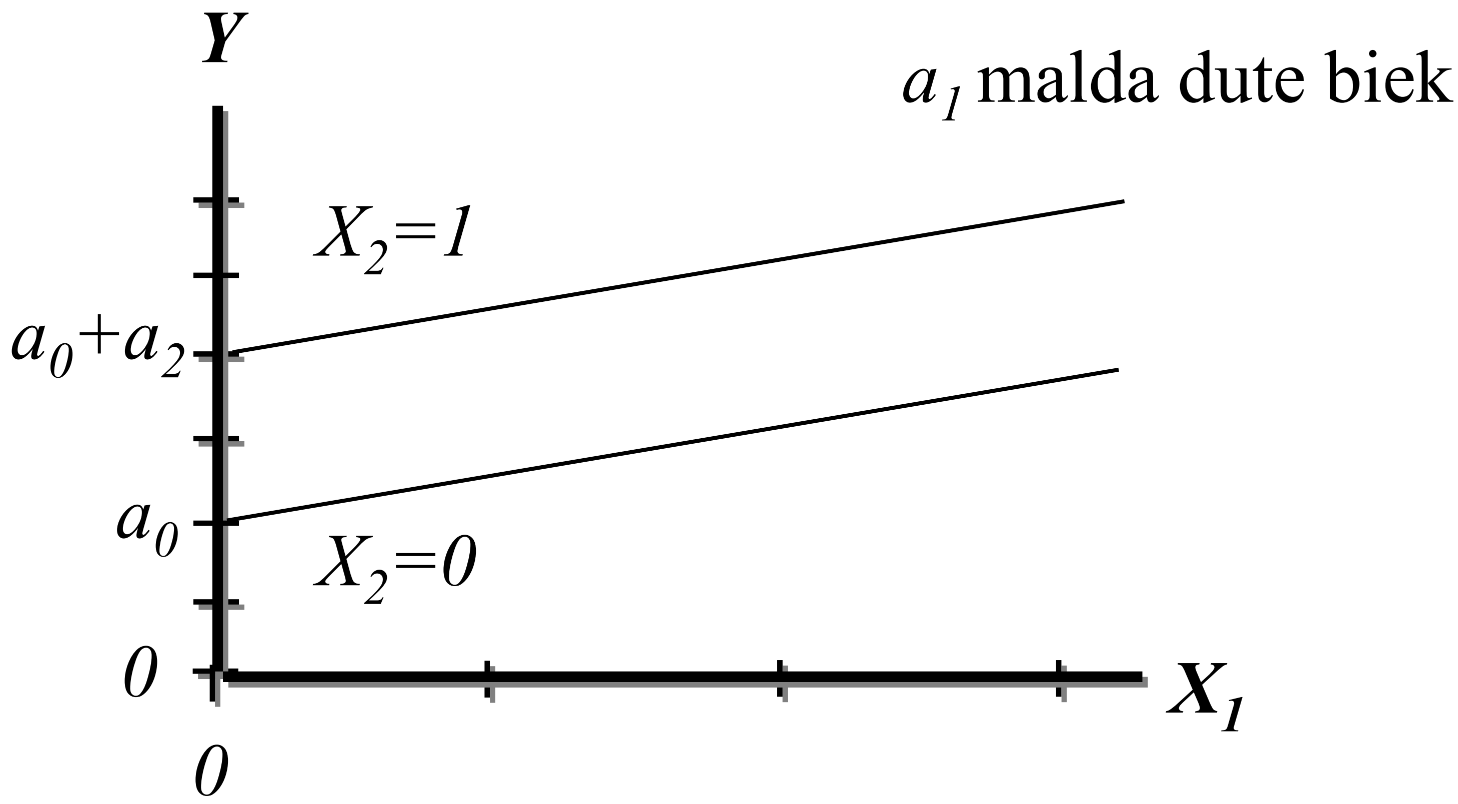
$$\boldsymbol{Y' = a_0 \qquad + a_1 \cdot X_1}$$

*X*ekiko *Y*ren erregresio lineal anizkoitza

Aldagai aske bat kualitatibo bitarra

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E$$

Erregresio-zuzenak bi dira, eta paraleloak, gainera



***X*ekiko *Y*ren erregresio lineal anizkoitza**

Aldagai aske bat kualitatiboa

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E$$

X_2 aldagaia, kualitatiboa da: edozein modalitate-kopurua.

*X*ekiko *Y*ren erregresio lineal anizkoitza

Aldagai aske bat kualitatiboa

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E$$

X_2 aldagaia, kualitatiboa da: edozein modalitate-kopurua.

Dummy aldagai bat modalitate bat modalitate bakoitzeko.

Denen artean baturaz *1* eman behar dute:

sasi-aldagai bat bat soberan dago.

X_2	X_{21}	X_{22}	X_{23}
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

$$Y = a_0 + a_1 \cdot X_1 + a_{21} \cdot X_{21} + a_{22} \cdot X_{22} + a_{23} \cdot X_{23} + E$$

a_0, a_1, a_{21}, a_{22}, eta a_{23} ohi bezala kalkulatu

Xekiko Yren erregresio lineal anizkoitza

Aldagai aske bat kualitatiboa

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E$$

$$Y = a_0 + a_1 \cdot X_1 + a_{21} \cdot X_{21} + a_{22} \cdot X_{22} + a_{23} \cdot X_{23} + E$$

Erregresio-zuzenak paraleloak