

Erregresio linealaren analisia

Datuak prozesatzeko sistema baten oinarrizko egitura-elementuak hiru dira:

- **Fitxategiak.** Sisteman dauden erregistro-multzo iraunkor bat osatzen dute.
- **Fluxuak.** Sistema eta inguruaren arteko datu-interfazeak dira.
- **Prozesuak.** Funtzionalki definitutako datuen manipulazio logikoak.

Hiru elementu hauei buruzko ikerketa bati egiten zaio aurre software-n garapenaren **kostuak** ikertzeko asmoz.

Datuak hauek izan ziren:

Kostua	(Y)	22.6	15.0	78.1	28.0	80.5	24.5	20.5	147.6	4.2	48.2	20.5
Fitxategiak	(X_1)	4	2	20	6	6	3	4	16	4	6	5
Fluxuak	(X_2)	44	33	80	24	227	20	41	187	19	50	48
Prozesuak	(X_3)	18	15	80	21	50	18	13	137	15	21	17

```
> Kostua <- c(22.6,15.0,78.1,28.0,80.5,24.5,20.5,147.6,4.2,48.2,20.5)
> Fitxategiak <- c(4,2,20,6,6,3,4,16,4,6,5)
> Fluxuak <- c(44,33,80,24,227,20,41,187,19,50,48)
> Prozesuak <- c(18,15,80,21,50,18,13,137,15,21,17)
> datuak <- data.frame(Kostua,Fitxategiak,Fluxuak,Prozesuak)
> attach(datuak)
```

Xekiko Yren erregresio lineal anizkoitza

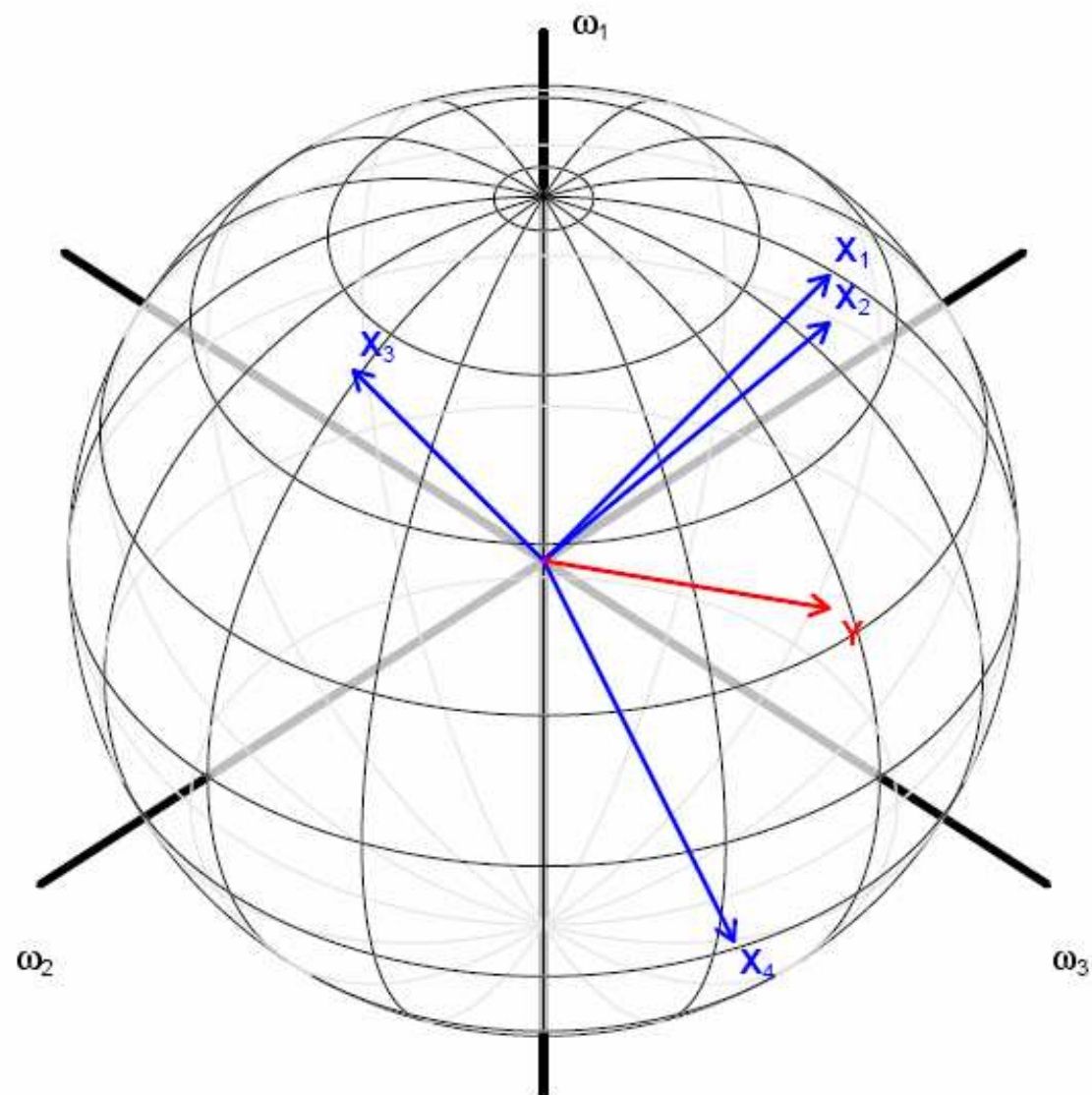
Hobetuko al da *Kostuaren* erregresioa hiru aldagaiak (*Fitxategiak*, *Fluxuak*, *Prozesuak*) linealki konbinatzen badira? Zenbateraino? Hiru aldagaiak beharrezkoak al dira?

```
> cor(datuak)
>
> Kostua      Fitxategiak      Fluxuak      Prozesuak
> Kostua      1.0000000      0.7784743  0.8303919  0.9598421
> Fitxategiak 0.7784743      1.0000000  0.4589818  0.8545609
> Fluxuak      0.8303919      0.4589818  1.0000000  0.7204369
> Prozesuak    0.9598421      0.8545609  0.7204369  1.0000000

> eval(expression(paste("r2 =", round(cor(Kostua,Fitxategiak)^2,digits=2))))
[1] "r2 = 0.61"
> eval(expression(paste("r2 =", round(cor(Kostua,Fluxuak)^2,digits=2))))
[1] "r2 = 0.69"
> eval(expression(paste("r2 =", round(cor(Kostua,Prozesuak)^2,digits=2))))
[1] "r2 = 0.92"
```

Korrelazio lineal koefizienteak

Ω	X_1	X_j	X_p	Y
ω_1	x_{11}	x_{1j}	x_{1p}	y_1
ω_i	x_{i1}	x_{ij}	x_{ip}	y_i
ω_n	x_{n1}	x_{nj}	x_{np}	y_n



Xekiko Yren erregresio lineal anizkoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_j X_j + \dots + a_p X_p + E$$

$$Y(\omega_i) = a_0 + a_1 X_1(\omega_i) + a_2 X_2(\omega_i) + \dots + a_j X_j(\omega_i) + \dots + a_p X_p(\omega_i) + E(\omega_i)$$

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_j x_{ij} + \dots + a_p x_{ip} + e_i \quad i=1,2,\dots,n$$

Optimizazio-problema:

Kalkulatu a_j ($j=0,1,2,\dots,p$) *hoberenak*, **errorea** (oro har) ahalik eta txikiena izan dadin.

Hoberenak, **errore karratu txikien**en zentzuan:

$$\frac{1}{n} \sum_{i=1}^n e_i^2$$

$\frac{1}{n}$ -k ez du eraginik,
baina batezbestekoaren
ulerkera ematen dio

Xekiko Yren erregresio lineal bikoitza

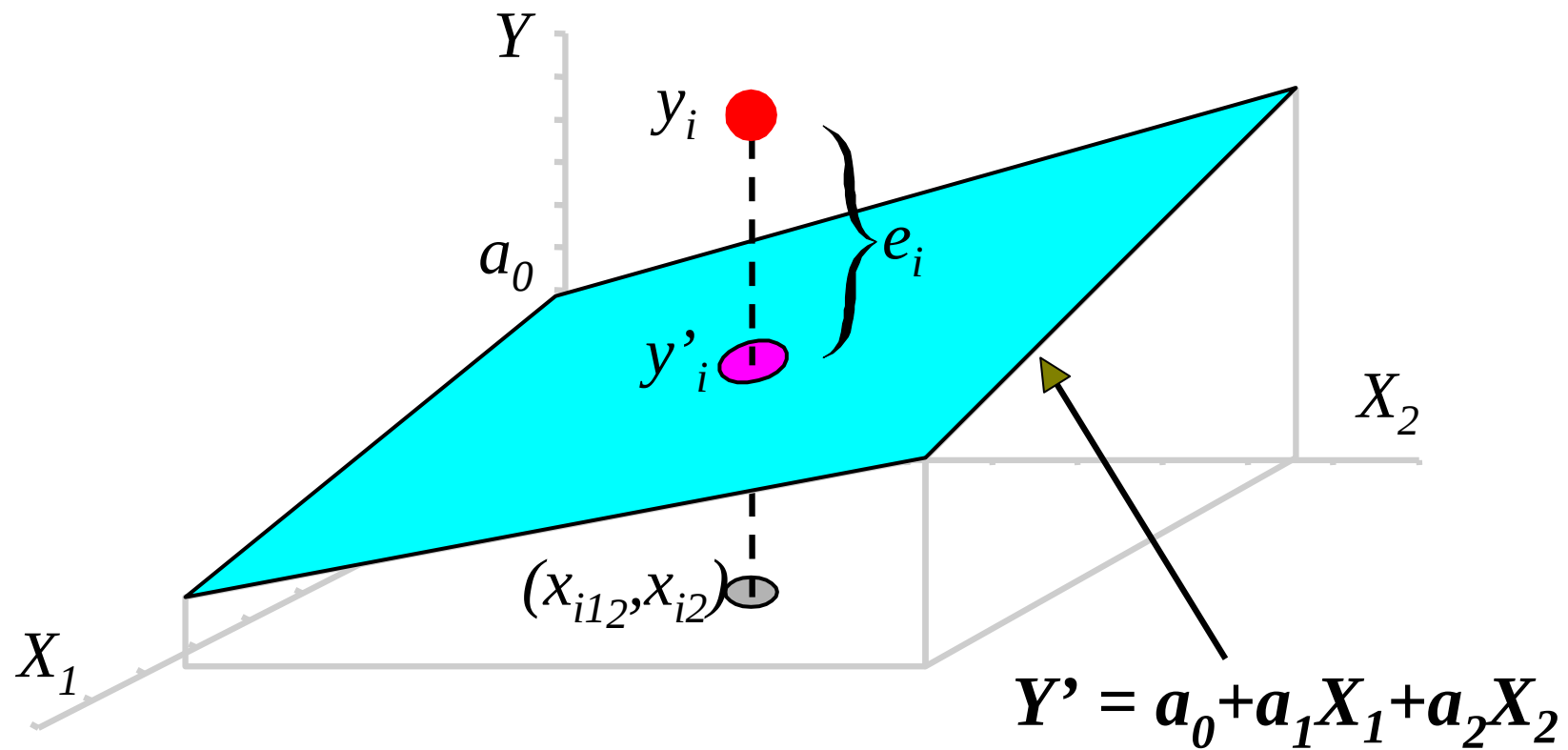
$$p=2 \quad Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$
$$y'_i = a_0 + a_1 x_{i1} + a_2 x_{i2} \quad i=1,2,\dots,n$$

Xekiko Yren erregresio lineal bikoitza

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

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

$$y'_i = a_0 + a_1 x_{i1} + a_2 x_{i2} \quad i=1,2,\dots,n$$



Xekiko Yren erregresio lineal bikoitza

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

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

$$y'_i = a_0 + a_1 x_{i1} + a_2 x_{i2} \quad i=1,2,\dots,n$$

Optimizazio-problema:

$$\begin{aligned} \underset{a_0, a_1, a_2}{\text{Min}} \quad G(a_0, a_1, a_2) &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y'_i)^2 = \\ &= \sum_{i=1}^n (y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i))^2 \end{aligned}$$

Xekiko Yren erregresio lineal bikoitza

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

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

$$y'_i = a_0 + a_1 x_{i1} + a_2 x_{i2} \quad i=1,2,\dots,n$$

Optimizazio-problema:

$$\begin{aligned} \underset{a_0, a_1, a_2}{\text{Min}} \quad G(a_0, a_1, a_2) &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y'_i)^2 = \\ &= \sum_{i=1}^n (y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i))^2 \end{aligned}$$

Ekuazio linealak:

$$\delta G / \delta a_0 = \sum_{i=1}^n 2 \cdot (y_i - a_0 - a_1 \cdot x_{i1} - a_2 \cdot x_{i2}) (-1) = 0$$

$$\delta G / \delta a_1 = \sum_{i=1}^n 2 \cdot (y_i - a_0 - a_1 \cdot x_{i1} - a_2 \cdot x_{i2}) (-x_{i1}) = 0$$

$$\delta G / \delta a_2 = \sum_{i=1}^n 2 \cdot (y_i - a_0 - a_1 \cdot x_{i1} - a_2 \cdot x_{i2}) (-x_{i2}) = 0$$

Xekiko Yren erregresio lineal bikoitza

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

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

Ekuazio linealak:

$$a_0 \cdot n + a_1 \sum_{i=1}^n x_{i1} + a_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i$$

$$a_0 \cdot \sum_{i=1}^n x_{i1} + a_1 \sum_{i=1}^n x_{i1}^2 + a_2 \sum_{i=1}^n x_{i2} x_{i1} = \sum_{i=1}^n x_{i1} y_i$$

$$a_0 \cdot \sum_{i=1}^n x_{i2} + a_1 \sum_{i=1}^n x_{i2} x_{i1} + a_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n x_{i2} y_i$$

Xekiko Yren erregresio lineal bikoitza

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

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

Soluzioa:

Xekiko Yren erregresio lineal bikoitza

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

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fitxategiak+Fluxuak)
> regKostua$coefficients
>(Intercept) Fitxategiak      Fluxuak
   -6.595274    3.732831    0.360353
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6 4.2 48.2 20.5
> round(regKostua$fitted.values, 1)         # Y'
   24.2 12.8 96.9 24.5 97.6 11.8 23.1 120.5 15.2 33.8 29.4
> round(regKostua$residuals, 1)             # E
   -1.6 2.2 -18.8 3.5 -17.1 12.7 -2.6 27.1 -11.0 14.4 -8.9
> var(regKostua$fitted.values)/var(Kostua)  # R^2
[1] 0.8895653
> var(regKostua$residuals)/var(Kostua)     # 1-R^2
[1] 0.1104347
```

$$s_y^2 = s_{y'}^2 + s_e^2$$

$$SST = SSR + SSE$$

$$R^2 = \frac{s_{y'}^2}{s_y^2} \quad R^2 = \frac{SSR}{SST}$$

Determinazio-koefizientea

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fitxategiak+Fluxuak)
> regKostua$coefficients
>(Intercept) Fitxategiak      Fluxuak
   -6.595274    3.732831    0.360353
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6 4.2 48.2 20.5
> round(regKostua$fitted.values, 1)         # Y'
      24.2 12.8 96.9 24.5 97.6 11.8 23.1 120.5 15.2 33.8 29.4
> round(regKostua$residuals, 1)             # E
      -1.6 2.2 -18.8 3.5 -17.1 12.7 -2.6 27.1 -11.0 14.4 -8.9
> var(regKostua$fitted.values)/var(Kostua)  # R^2
[1] 0.8895653
> var(regKostua$residuals)/var(Kostua)     # 1-R^2
[1] 0.1104347

> eval(expression(paste("r2 =", round(cor(Kostua,Fitxategiak)^2,digits=2))))
[1] "r2 = 0.61"
> eval(expression(paste("r2 =", round(cor(Kostua,Fluxuak)^2,digits=2))))
[1] "r2 = 0.69"
```

Elkarrekin hobe, bakoitzak bere aldetik baino

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fitxategiak+Prozesuak)
> regKostua$coefficients
(Intercept) Fitxategiak    Prozesuak
   8.849290   -1.148343    1.184276
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6 4.2 48.2 20.5
> round(regKostua$fitted.values, 1)          # Y'
   25.6 24.3 80.6 26.8 61.2 26.7 19.7 152.7 22.0 26.8 23.2
> round(regKostua$residuals, 1)              # E
   -3.0 -9.3 -2.5  1.2 19.3 -2.2  0.8 -5.1 -17.8 21.4 -2.7
> var(regKostua$fitted.values)/var(Kostua)   # R^2
[1] 0.927765

> eval(expression(paste("r2 =", round(cor(Kostua,Fitxategiak)^2,digits=2))))
[1] "r2 = 0.61"
> eval(expression(paste("r2 =", round(cor(Kostua,Prozesuak)^2,digits=4))))
[1] "r2 = 0.9213"
```

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fitxategiak+Prozesuak)
> regKostua$coefficients
(Intercept) Fitxategiak    Prozesuak
   8.849290   -1.148343    1.184276
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6 4.2 48.2 20.5
> round(regKostua$fitted.values, 1)         # Y'
   25.6 24.3 80.6 26.8 61.2 26.7 19.7 152.7 22.0 26.8 23.2
> round(regKostua$residuals, 1)             # E
   -3.0 -9.3 -2.5  1.2 19.3 -2.2  0.8 -5.1 -17.8 21.4 -2.7
> var(regKostua$fitted.values)/var(Kostua)  # R^2
[1] 0.927765

> eval(expression(paste("r2 =", round(cor(Kostua,Fitxategiak)^2,digits=2))))
[1] "r2 = 0.61"
> eval(expression(paste("r2 =", round(cor(Kostua,Prozesuak)^2,digits=4))))
[1] "r2 = 0.9213"
```

Elkarrekin ezer gutxi hobetu dute erregresioa

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fitxategiak+Prozesuak)
> regKostua$coefficients
(Intercept) Fitxategiak    Prozesuak
   8.849290   -1.148343    1.184276
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6  4.2 48.2 20.5
> round(regKostua$fitted.values, 1)          # Y'
   25.6 24.3 80.6 26.8 61.2 26.7 19.7 152.7 22.0 26.8 23.2
> round(regKostua$residuals, 1)              # E
   -3.0 -9.3 -2.5  1.2 19.3 -2.2  0.8 -5.1 -17.8 21.4 -2.7
> var(regKostua$fitted.values)/var(Kostua)   # R^2
[1] 0.927765
> cor(Fitxategiak,Prozesuak)
[1] 0.8545609
> eval(expression(paste("r2 =", round(cor(Kostua,Fitxategiak)^2,digits=2))))
[1] "r2 = 0.61"
> eval(expression(paste("r2 =", round(cor(Kostua,Prozesuak)^2,digits=4))))
[1] "r2 = 0.9213"
```

Elkarrekin ezer gutxi hobetu dute erregresioa

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fluxuak+Prozesuak)
> regKostua$coefficients
(Intercept)      Fluxuak      Prozesuak
  2.3033865    0.1736161    0.8152037
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6  4.2 48.2 20.5
> round(regKostua$fitted.values, 1)          # Y'
      24.6 20.3 81.4 23.6 82.5 20.4 20.0 146.5 17.8 28.1 24.5
> round(regKostua$residuals, 1)              # E
      -2.0 -5.3 -3.3  4.4 -2.0  4.1  0.5  1.1 -13.6 20.1 -4.0
> var(regKostua$fitted.values)/var(Kostua)   # R^2
[1] 0.9614019
> cor(Fluxuak, Prozesuak)
[1] 0.7204369
```

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fluxuak+Prozesuak)
> regKostua$coefficients
(Intercept)      Fluxuak      Prozesuak
  2.3033865    0.1736161    0.8152037
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6  4.2 48.2 20.5
> round(regKostua$fitted.values, 1)          # Y'
      24.6 20.3 81.4 23.6 82.5 20.4 20.0 146.5 17.8 28.1 24.5
> round(regKostua$residuals, 1)              # E
      -2.0 -5.3 -3.3  4.4 -2.0  4.1  0.5  1.1 -13.6 20.1 -4.0
> var(regKostua$fitted.values)/var(Kostua)   # R^2
[1] 0.9614019
> cor(Fluxuak, Prozesuak)
[1] 0.7204369

> eval(expression(paste("r2 =", round(cor(Kostua, Fluxuak)^2, digits=2))))
[1] "r2 = 0.69"
> eval(expression(paste("r2 =", round(cor(Kostua, Prozesuak)^2, digits=2))))
[1] "r2 = 0.92"
```

Elkarrekin pixka bat hobetu dute erregresioa

Xekiko Yren erregresio lineal hirukoitza

$$p=3 \quad Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + E$$

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fitxategiak+Fluxuak+Prozesuak)
> regKostua$coefficients
(Intercept) Fitxategiak      Fluxuak      Prozesuak
  1.9617795    0.1177586    0.1767263    0.7964477
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6  4.2 48.2 20.5
> round(regKostua$fitted.values, 1)         # Y'
  24.5 20.0 82.2 23.6 82.6 20.2 20.0 146.0 17.7 28.2 24.6
> round(regKostua$residuals, 1)             # E
  -1.9 -5.0 -4.1  4.4 -2.1  4.3  0.5  1.6 -13.5 20.0 -4.1
> var(regKostua$fitted.values)/var(Kostua)  # R^2
[1] 0.961457
```

Kostuaren erregresioa bi aldagairekin (Fluxuak, Prozesuak) egin edo hirurekin oso antzekoa da ontasunari begira, 0.961402tik 0.961457ra soilik igaro baita.

Xekiko Yren erregresio lineal hirukoitza

$$p=3 \quad Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + E$$

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fitxategiak+Fluxuak+Prozesuak)
> regKostua$coefficients
(Intercept) Fitxategiak      Fluxuak      Prozesuak
  1.9617795    0.1177586    0.1767263    0.7964477
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6  4.2 48.2 20.5
> round(regKostua$fitted.values, 1)         # Y'
      24.5 20.0 82.2 23.6 82.6 20.2 20.0 146.0 17.7 28.2 24.6
> round(regKostua$residuals, 1)             # E
      -1.9 -5.0 -4.1  4.4 -2.1  4.3  0.5  1.6 -13.5 20.0 -4.1
> var(regKostua$fitted.values)/var(Kostua)  # R^2
[1] 0.961457
```

Kostuaren erregresioa bi aldagairekin (Fluxuak, Prozesuak) egin edo hirurekin oso antzekoa da ontasunari begira, 0.961402tik 0.961457ra soilik igaro baita:

Aldagaien hautaketa-problema

Xekiko Yren erregresio lineal hirukoitza

Aldagaien hautaketa-problema

Oinarrizko araua: Ahalik eta aldagai aske gutxien erabili.

Aldagai-azpimultzorik hoberena:

- Kalkulatu aldagaien multzoaren azpimultzo guztien erregresioaren ontasuna (R^2 handia edo $1-R^2$ (erroreak) txikia), eta hautatu hoberena.
- Aztertu azpimultzo hoberenen artean aldagai-azpimultzoaren kopurua (edo beste irizpideren bat) eta beren ontasunaren arteko erlazioa. Azterketa hau egiteko bada zenbait formula.
- Aldagai-kopurua (p) handia denean, konputazio-konplexotasun handiko problema izan daiteke (2^p-1 aldagai-azpimultzo): hoberena bilatzeko heuristikoren bat erabili.

Xekiko Yren erregresio lineal hirukoitza

Aldagaien hautaketa-problema

Oinarrizko araua: Ahalik eta aldagai aske gutxien erabili.

Pausokako erregresioa:

- Aurreraka. Hasieran aldagai hautatuen multzoa hutsa dago. Hautatu Yrekin korrelatuen dagoen X_j aldagaia (r_{xy}^2 ontasun handiena duena) eta multzoan ipini. Gehitu berekin ontasun handien lortzen duen X_j aldagai. Gelditu hobekuntzak, txikia izateagatik, merezi ez duenean.
- Atzeraka. Hasieran aldagai guztiak daude hautatuen multzoan. Hautatu guztien arteko ontasuna gutxien txikiagotzen duen X_j aldagaia, eta kendu aldagai-multzotik. Pausoka jardun era berean ontasunaren beherakada oso nabarmena ez den bitartean.
- Aurreraka eta atzeraka. Hasieran hautatuen multzoa edozein aldagai egon daiteke. Aurreko bi estrategiak konbinatu.

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

```
> regKostua <- lm(Kostua~Fluxuak+Prozesuak)
> regKostua$coefficients
(Intercept)      Fluxuak      Prozesuak
  2.3033865    0.1736161    0.8152037
> Kostua                                     # Y
[1] 22.6 15.0 78.1 28.0 80.5 24.5 20.5 147.6  4.2 48.2 20.5
> round(regKostua$fitted.values, 1)         # Y'
  24.6 20.3 81.4 23.6 82.5 20.4 20.0 146.5 17.8 28.1 24.5
> round(regKostua$residuals, 1)             # E
  -2.0 -5.3 -3.3  4.4 -2.0  4.1  0.5  1.1 -13.6 20.1 -4.0
> var(regKostua$fitted.values)/var(Kostua)  # R^2
[1] 0.9614019
```

$$Kostua = 2.30 + 0.17 \cdot Fluxuak + 0.82 \cdot Prozesuak + E$$

Xekiko Yren erregresio lineal bikoitza

- *Kostuaren* erregresio lineala balio tipikoak erabiliz egina izango balitz, emaitza berbera lortuko litzateke.

Xekiko Yren erregresio lineal anizkoitza



n oinarritutako kalkulagailua:

http://www.wessa.net/rwasp_multipleregression.wasp

Send output to:	
Browser Blue - Charts White	
Data X:	
<pre>99.0 98.1 100.1 100.0100.0100.0 111.6104.990.6 122.2104.986.5 117.6109.589.7 121.1110.890.6 136.0112.382.8 154.2109.370.1 153.6105.365.4 158.5101.761.3 140.695.4 62.5 136.296.4 63.6 168.097.6 52.6 154.3102.459.7 149.0101.659.5 165.5103.861.3</pre>	
Names of X columns:	
Cons Inc Price	
Sample Range: (leave blank to include all observations)	
From:	
To:	
Column Number of Endogenous Series (?)	
1	
Fixed Seasonal Effects	
Do not include Seasonal Dummies	
Type of Equation	
No Linear Trend	
Chart options	
Width:	600
Height:	400
Compute	

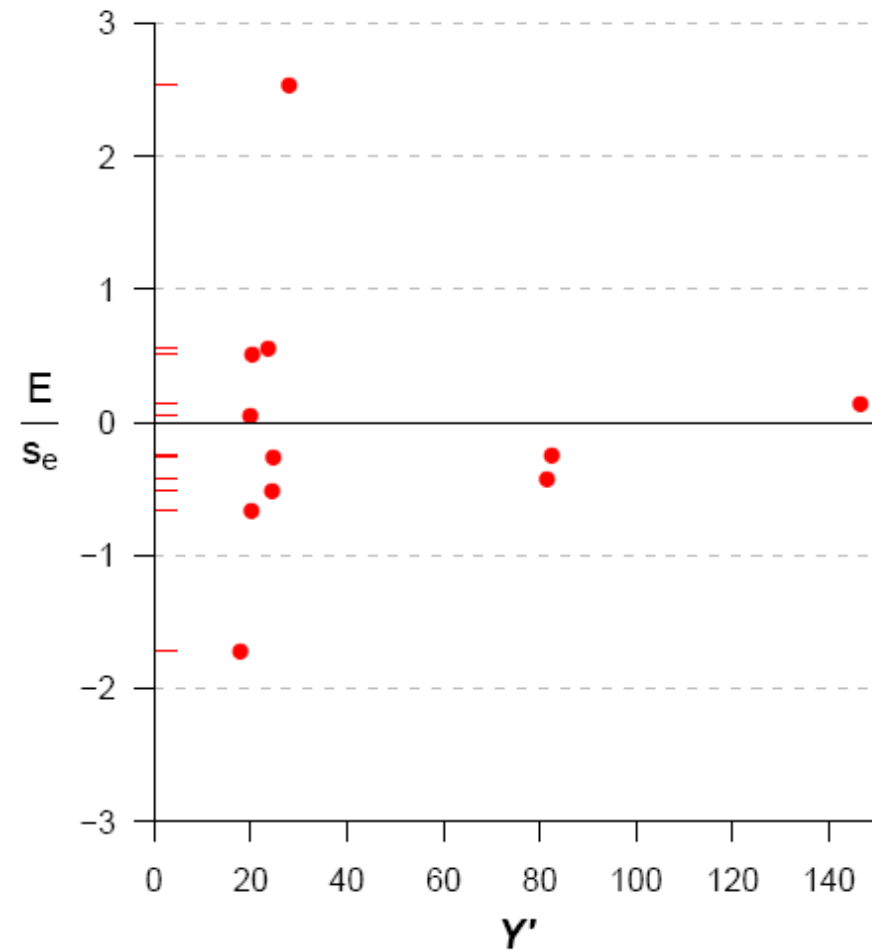
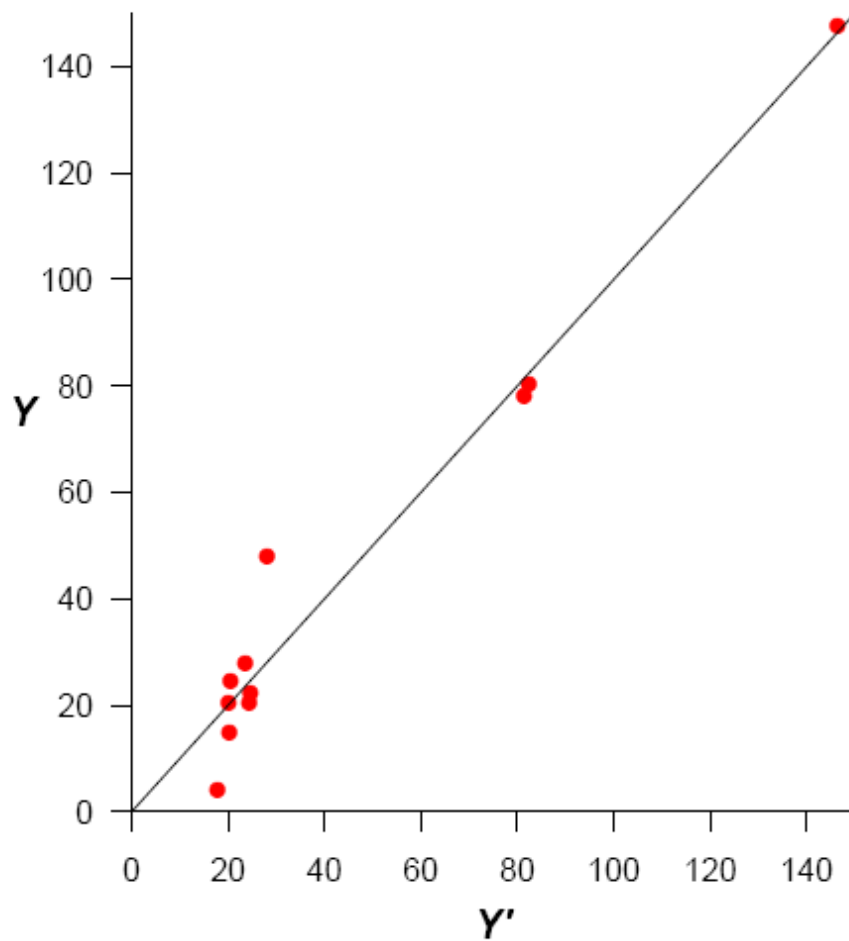
[Click here to edit the underlying code of this R Module.](#)

Xekiko Yren erregresio lineal anizkoitza

- Yren erregresioa egiteko, hautatu Yrekin estuki korrelatuak eta beren artean inkorrelatuak dauden X_j aldagaiak
- 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.

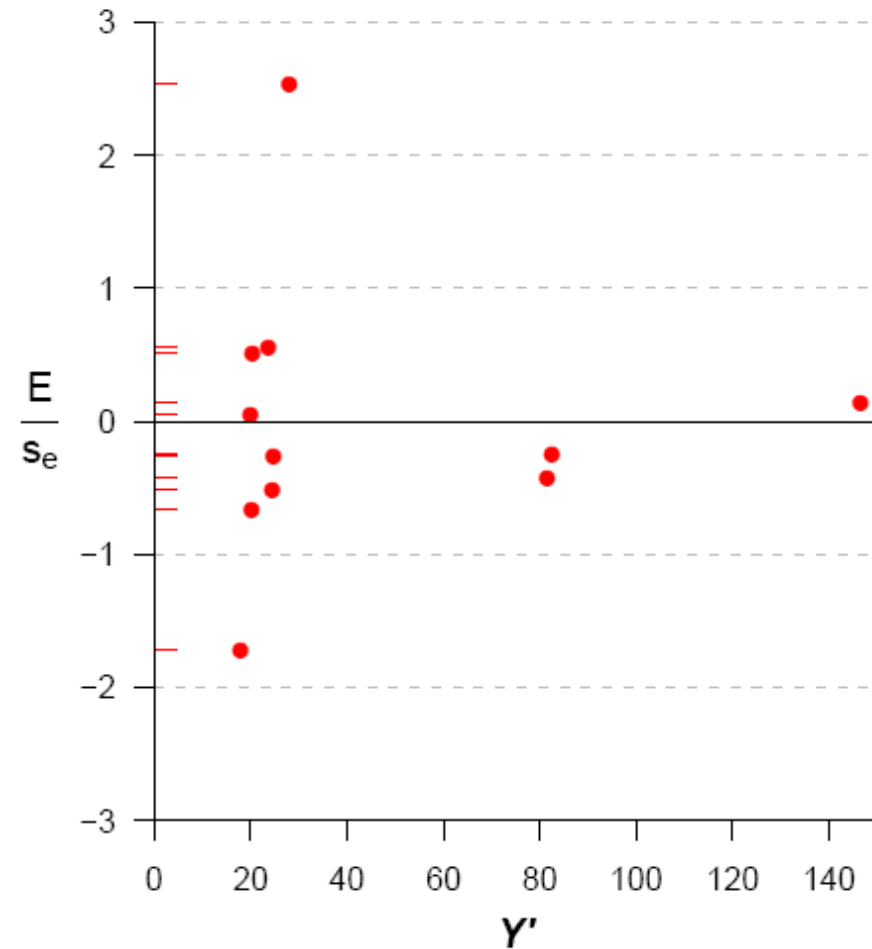
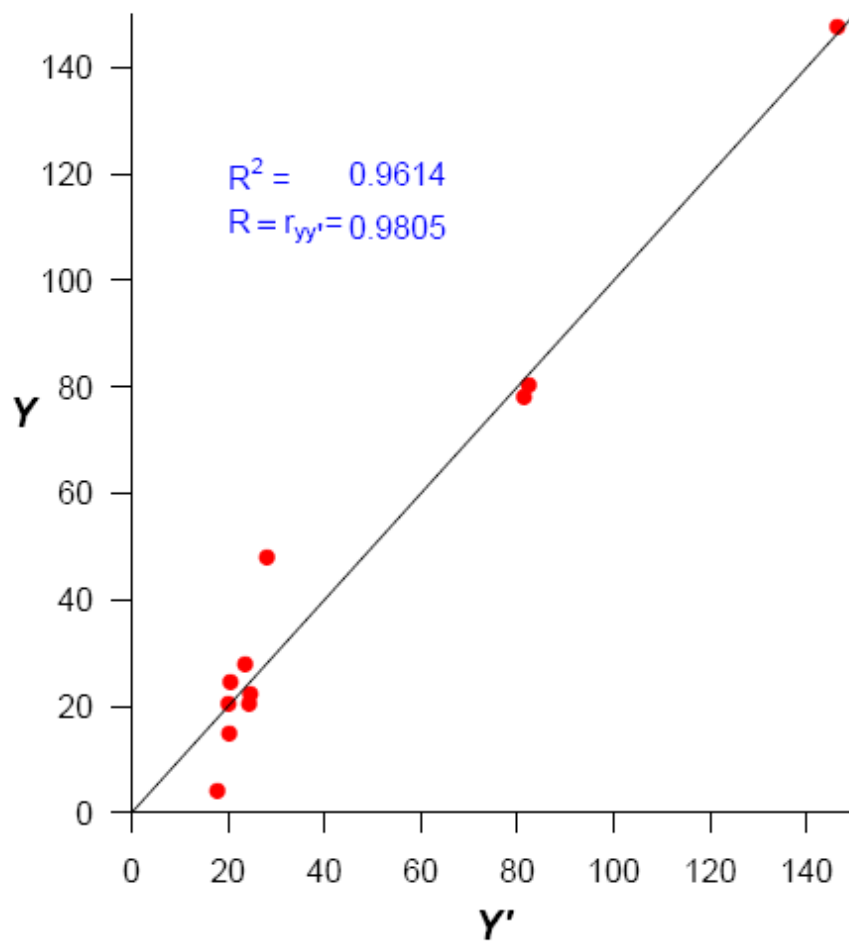
Xekiko Yren erregresio lineal hirukoitza

Hondarren azterketa



Xekiko Yren erregresio lineal hirukoitza

Hondarren azterketa



Xekiko Yren erregresio lineal hirukoitza

Hondarren azterketa

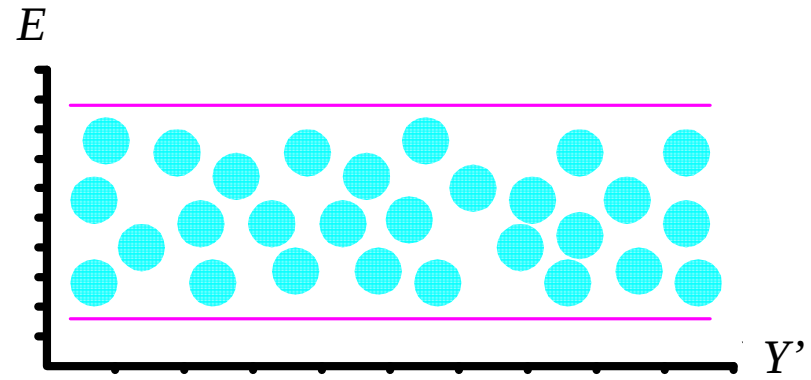
```
> regKostua <- lm(Kostua~Fluxuak+Prozesuak)
> regKostua$coefficients
> y <- Kostua # Y
> yprime <- regKostua$fitted.values # Y'
> e <- regKostua$residuals # E
> par(mfrow=c(1,2))
> plot(yprime, y, xlim=c(0,150), ylim=c(0,150), las=1,
      xaxs="i", yaxs="i", pch=19, col="red", axes=FALSE, ann=FALSE)
> abline(a=0, b=1)
> axis(1, at=seq(0,140,20) )
> mtext("Y'", 1, line=2.5, cex=1.25, font=4)
> axis(2, at=seq(0,140,20), las=1 )
> mtext("Y", 2, line=2.5, cex=1.25, font=4, las=1)
> box(bty="l")
> e <- e/(sqrt(var(e)*(length(e)-1)/length(e)))
> plot(yprime, e, xlim=c(0,150),
      ylim=c(-3,3), las=1,
      xaxs="i", yaxs="i", pch=19, col="red", axes=FALSE, ann=FALSE)
> axis(1, at=seq(0,140,20) )
> mtext("Y'", 1, line=2.5, cex=1.25, font=4)
> rug(e, ticksize = 0.03, side=2, col="red")
> axis(2, at=seq(-3,3,1), las=1 )
> mtext(expression(frac(E,s[e])), 2, line=2.5, cex=1.25, font=4, las=1)
> abline(h=seq(-3,3,1), col="gray", lty="dashed")
> abline(h=0)
> box(bty="l")
> par(mfrow=c(1,1))
```

Xekiko Yren erregresio lineal hirukoitza

Hondarren azterketa

Hondarren banaketa 'egokia':

Y 'ren edozein balioarentzat,
era berean

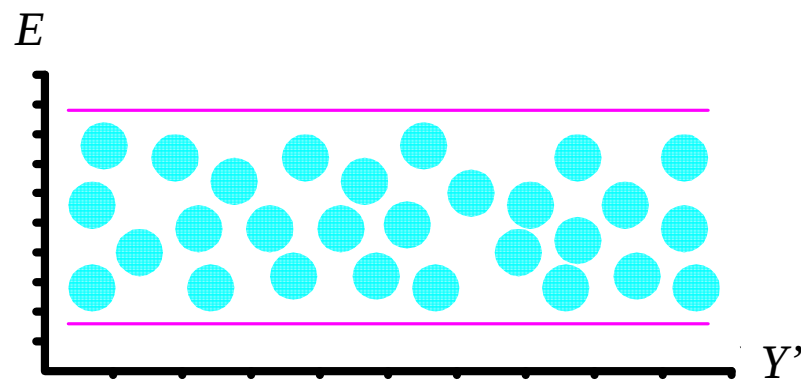


Xekiko Yren erregresio lineal hirukoitza

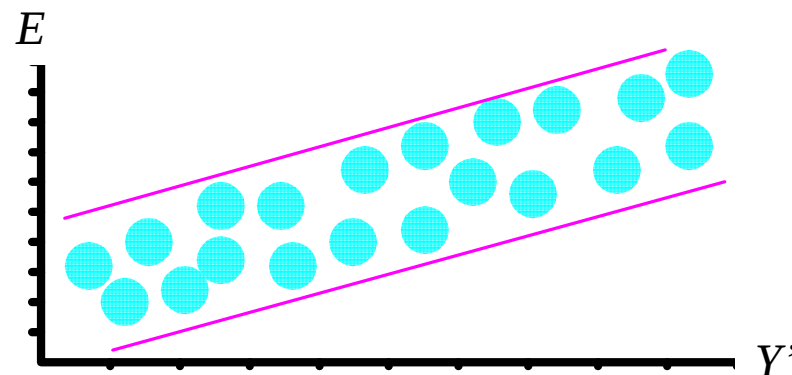
Hondarren azterketa

Hondarren banaketa 'egokia':

Y 'ren edozein balioarentzat, era berean banatzen dira.



Hondarren banaketak Y 'ren balioen menpekotasun lineala azaltzen dute.

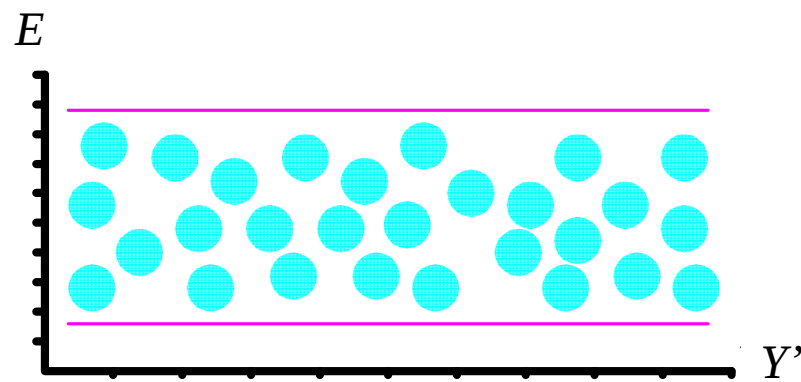


Xekiko Yren erregresio lineal hirukoitza

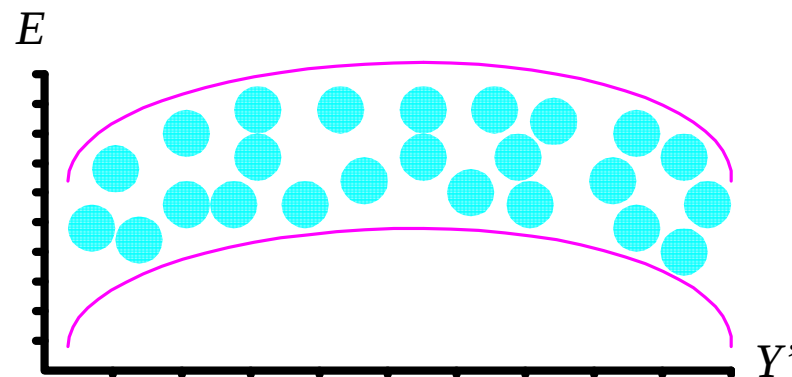
Hondarren azterketa

Hondarren banaketa 'egokia':

Y 'ren edozein balioarentzat,
era berean



Hondarrek 2. graduko erlazio
funtzional bat erakusten dute:

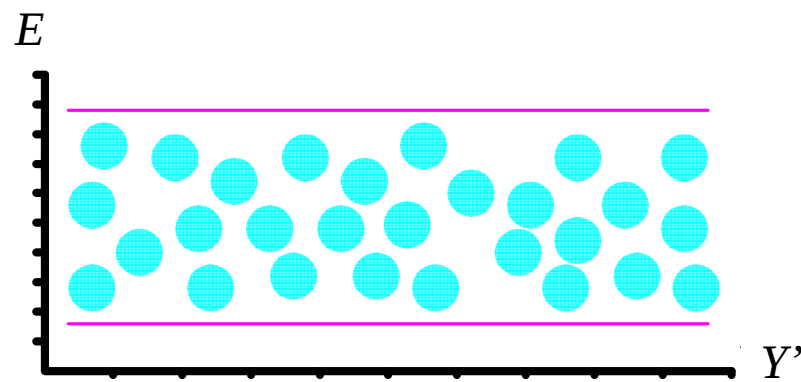


Xekiko Yren erregresio lineal hirukoitza

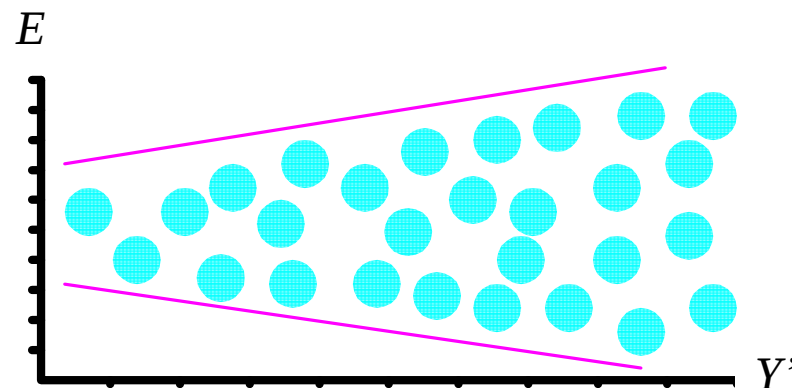
Hondarren azterketa

Hondarren banaketa 'egokia':

Y' ren edozein balioarentzat, era berean banatzen dira.

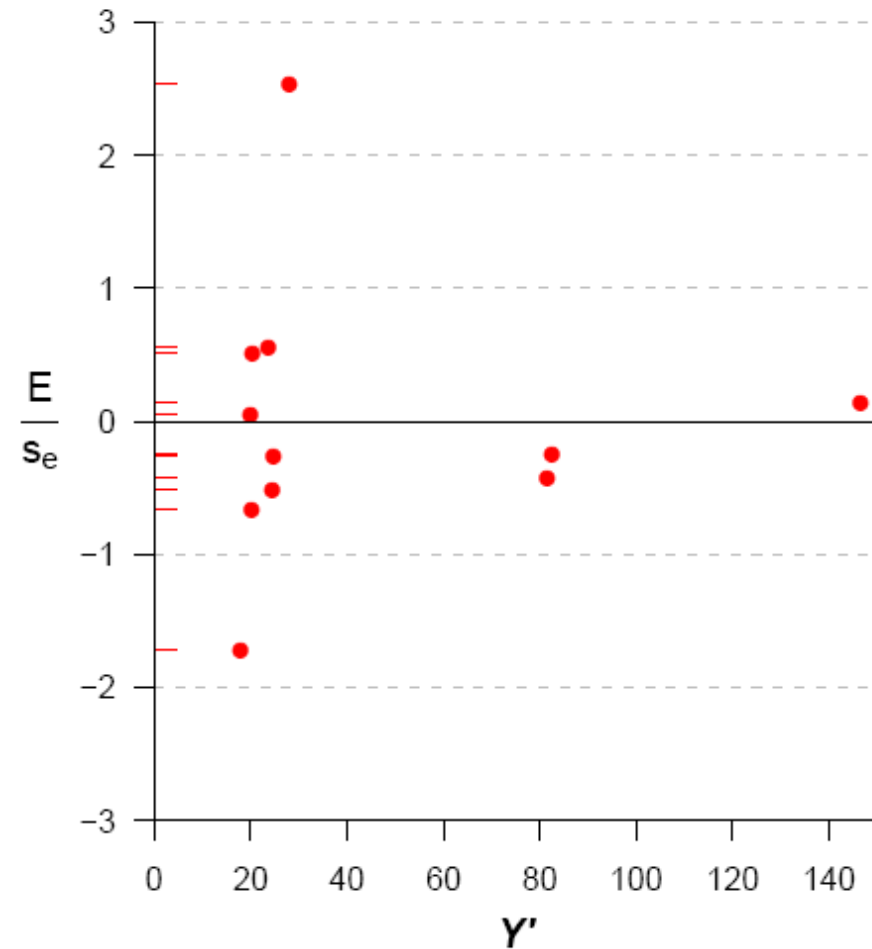
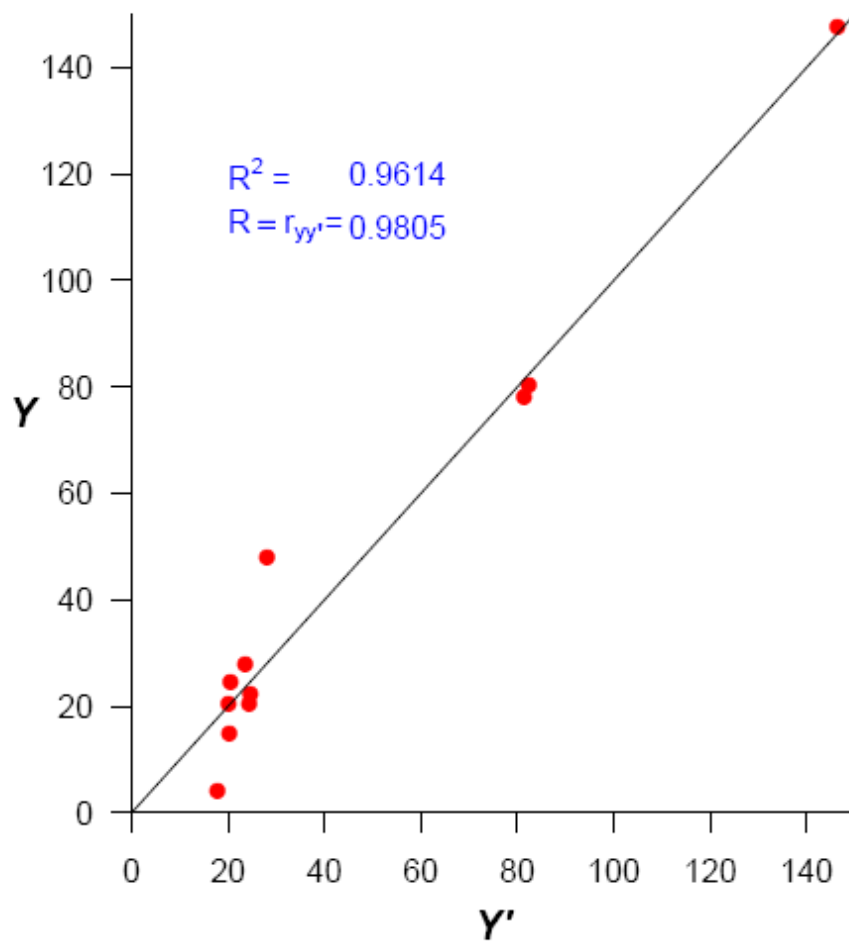


Hondarrek banaketa ezberdina azaltzen dute Y' ren balioen arabera.



Xekiko Yren erregresio lineal hirukoitza

Hondarren azterketa



Xekiko Yren erregresio lineal anizkoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_j X_j + \dots + a_p X_p + E$$

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_j x_{ij} + \dots + a_p x_{ip} + e_i \quad i=1,2,\dots,n$$

Matrizetaz baliatuta: $\mathbf{Y} = \mathbf{X} \cdot \mathbf{a} + \mathbf{E}$, eta $\mathbf{Y}' = \mathbf{X} \cdot \mathbf{a}$

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_i \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1, x_{11}, x_{12}, \dots, x_{1j}, \dots, x_{1p} \\ 1, x_{21}, x_{22}, \dots, x_{2j}, \dots, x_{2p} \\ \dots \\ 1, x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{ip} \\ \dots \\ 1, x_{n1}, x_{n2}, \dots, x_{nj}, \dots, x_{np} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_j \\ \dots \\ a_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_i \\ \dots \\ e_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{a} + \mathbf{E}$$

Xekiko Yren erregresio lineal bikoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + E$$
$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

Ekuaizio linealak:

$$a_0 \cdot n + a_1 \sum_{i=1}^n x_{i1} + a_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i$$

$$a_0 \cdot \sum_{i=1}^n x_{i1} + a_1 \sum_{i=1}^n x_{i1}^2 + a_2 \sum_{i=1}^n x_{i2} x_{i1} = \sum_{i=1}^n x_{i1} y_i$$

$$a_0 \cdot \sum_{i=1}^n x_{i2} + a_1 \sum_{i=1}^n x_{i2} x_{i1} + a_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n x_{i2} y_i$$

Adierazi matrizeen bitartez

Xekiko Yren erregresio lineal bikoitza

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

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + e_i \quad i=1,2,\dots,n$$

Ekuazio linealak:

$$a_0 \cdot n + a_1 \sum_{i=1}^n x_{i1} + a_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i$$

$$a_0 \cdot \sum_{i=1}^n x_{i1} + a_1 \sum_{i=1}^n x_{i1}^2 + a_2 \sum_{i=1}^n x_{i2} x_{i1} = \sum_{i=1}^n x_{i1} y_i$$

$$a_0 \cdot \sum_{i=1}^n x_{i2} + a_1 \sum_{i=1}^n x_{i2} x_{i1} + a_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n x_{i2} y_i$$

$$X^T \cdot X \cdot a = X^T \cdot Y$$

Xekiko Yren erregresio lineal anizkoitza

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_j X_j + \dots + a_p X_p + E$$

$$Y = X \cdot a + E, \text{ eta } Y' = X \cdot a$$

Ekuazio linealak:

$$X^T \cdot X \cdot a = X^T \cdot Y$$

Soluzioa:

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

Xekiko Yren erregresio lineal anizkoitza

http://en.wikipedia.org/wiki/Linear_regression

http://en.wikipedia.org/wiki/Regression_analysis

http://en.wikipedia.org/wiki/Robust_regression

http://en.wikipedia.org/wiki/Least_absolute_deviations

http://www.math.wpi.edu/Course_Materials/SAS/lablets/7.3/73_choices.html