Erregresioaren analisia

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

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

$$Y' = f(X_1, X_2, ..., X_p)$$

*X*ekiko *Y*ren erregresio lineal anizkoitza

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

- *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
Analisi bereizle lineala

- ***Y*** menpeko aldagaia kualitatibo bitarra da. Bere bi modalitateak *0* eta *1* zenbakiez kodetu.
  - y_i balioak *1*k ala *0*k dira.
  - x_{ij} balioak *1*k ala *0*k dira.
- Kasu honetan erregresio anizkoitza bi sail bereizteko *analisi bereizle lineala* esaten zaiona da.

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Y aldagaia, kualitatibo bitarra da:
bati *1* zenbakia egokitu, eta besteari *0*a.

X aldagaia, kualitatibo bitarra da:
bati *1* zenbakia egokitu, eta besteari *0*a.

Ω	X	Y
ω_1	x_1	y_1
ω_i	x_i	y_i
ω_n	x_n	y_n

$x_i \in \{0,1\}$

$y_i \in \{0,1\}$

X \ *Y*	*1*	*0*	
1	n_{11}	n_{10}	$n_{1\cdot}$
0	n_{01}	n_{00}	$n_{0\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 0}$	$n_{\cdot\cdot}$

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Adibidea:

X \ Y	*1*	*0*
1	40	10
0	5	45

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Adibidea:

X \ Y	*1*	*0*	
1	40	10	50
0	5	45	50
	45	55	100

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Y aldagaia, kualitatibo bitarra da:

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

$$a = \frac{s_{xy}}{s_x^2}$$

$$b = \bar{y} - \frac{s_{xy}}{s_x^2}\bar{x}$$

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Y aldagaia, kualitatibo bitarra da:

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

$$a = \frac{s_{xy}}{s_x^2} \qquad \bar{x} = \frac{n_{1\cdot}}{n_{\cdot\cdot}}$$

$$b = \bar{y} - \frac{s_{xy}}{s_x^2}\bar{x} \qquad \bar{y} = \frac{n_{\cdot 1}}{n_{\cdot\cdot}}$$

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Y aldagaia, kualitatibo bitarra da:

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

$$a = \frac{s_{xy}}{s_x^2} \qquad \bar{x} = \frac{n_{1\cdot}}{n_{\cdot\cdot}} \qquad s_x^2 = \frac{n_{1\cdot}}{n_{\cdot\cdot}}\left(1 - \frac{n_{1\cdot}}{n_{\cdot\cdot}}\right) = \frac{n_{1\cdot}}{n_{\cdot\cdot}}\frac{n_{0\cdot}}{n_{\cdot\cdot}}$$

$$b = \bar{y} - \frac{s_{xy}}{s_x^2}\bar{x} \qquad \bar{y} = \frac{n_{\cdot 1}}{n_{\cdot\cdot}} \qquad s_y^2 = \frac{n_{\cdot 1}}{n_{\cdot\cdot}}\left(1 - \frac{n_{\cdot 1}}{n_{\cdot\cdot}}\right) = \frac{n_{\cdot 1}}{n_{\cdot\cdot}}\frac{n_{\cdot 0}}{n_{\cdot\cdot}}$$

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Y aldagaia, kualitatibo bitarra da:
bati *1* zenbakia egokitu, eta besteari *0*a.

$$a = \frac{s_{xy}}{s_x^2} \qquad \bar{x} = \frac{n_{1\cdot}}{n_{..}} \qquad s_x^2 = \frac{n_{1\cdot}}{n_{..}}\left(1 - \frac{n_{1\cdot}}{n_{..}}\right) = \frac{n_{1\cdot}}{n_{..}}\frac{n_{0\cdot}}{n_{..}}$$

$$b = \bar{y} - \frac{s_{xy}}{s_x^2}\bar{x} \qquad \bar{y} = \frac{n_{\cdot 1}}{n_{..}} \qquad s_y^2 = \frac{n_{\cdot 1}}{n_{..}}\left(1 - \frac{n_{\cdot 1}}{n_{..}}\right) = \frac{n_{\cdot 1}}{n_{..}}\frac{n_{\cdot 0}}{n_{..}}$$

$$s_{xy} = \frac{1}{n_{..}}\sum_{i=1}^{n_{..}}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = \frac{1}{n_{..}}\sum_{i=1}^{n_{..}} x_i y_i - \overline{xy} =$$

$$= \frac{n_{11}}{n_{..}} - \frac{n_{1\cdot}}{n_{..}}\frac{n_{\cdot 1}}{n_{..}} = \frac{n_{..}n_{11} - n_{1\cdot}n_{\cdot 1}}{n_{..}^2}$$

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Adibidea:

X \ Y	1	0	
1	40	10	50
0	5	45	50
	45	55	100

$$\bar{x} = \frac{50}{100} \qquad s_x^2 = \frac{50^2}{100^2}$$

$$\bar{y} = \frac{45}{100} \qquad s_y^2 = \frac{45 \cdot 55}{100^2}$$

$$s_{xy} = \frac{100 \cdot 40 - 50 \cdot 45}{100^2}$$

$$a = \frac{100 \cdot 40 - 45 \cdot 50}{50 \cdot 50} = 0.7$$

$$b = 0.45 - 0.7 \cdot 0.5 = 0.1$$

$$Y' = 0.1 + 0.7 \cdot X$$

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Adibidea:

$$Y' = 0.1 + 0.7 \cdot X$$

X \ Y	1	0	
1	40	10	50
0	5	45	50
	45	55	100

```
> x <- c(rep(1,50),rep(0,50))
> y <-  c(rep(1,40),rep(0,10),rep(1,5),rep(0,45))
> datuak <- data.frame(x,y)
> table(datuak); table(datuak)[1:4]
   y
x    0  1
  0 45  5
  1 10 40
> lm(y~x)$coefficients
(Intercept)           x
        0.1         0.7
```

$$Y' = 0.1 + 0.7 \cdot X$$

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Adibidea:

$$Y' = 0.1 + 0.7 \cdot X$$

X \ Y	*1*	*0*	
1	40	10	50
0	5	45	50
	45	55	100

```
> x <- c(rep(1,50),rep(0,50))
> y <-  c(rep(1,40),rep(0,10),rep(1,5),rep(0,45))
> datuak <- data.frame(x,y)
> table(datuak); table(datuak)[1:4]
   y
x     0  1
  0 45  5
  1 10 40
> lm(y~x)$coefficients
(Intercept)           x
        0.1         0.7
```

$$Y' = 0.1 + 0.7 \cdot X$$

X \ Y	*1*	*0*		$Y=1$ in $X=x_i$
1	40	10	50	0.8
0	5	45	50	0.1
	45	55	100	

*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

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

Adibidea:

$$Y' = 0.1 + 0.7 \cdot X$$

X \ Y	1	0	
1	40	10	50
0	5	45	50
	45	55	100

```
> x <- c(rep(1,50),rep(0,50))
> y <-  c(rep(1,40),rep(0,10),rep(1,5),rep(0,45))
> datuak <- data.frame(x,y)
> table(datuak); table(datuak)[1:4]
   y
x     0  1
  0 45  5
  1 10 40
> lm(y~x)$coefficients
(Intercept)           x
        0.1         0.7
```

$$Y' = 0.1 + 0.7 \cdot X$$

X \ Y	1	0		$Y=1$ in $X=x_i$
1	40	10	50	0.8
0	5	45	50	0.1
	45	55	100	

```
> xyt <- table(datuak)
> y1prop <- c(xyt[4]/(xyt[2]+xyt[4]),
+             xyt[3]/(xyt[1]+xyt[3]))
> y1prop <- c(rep(y1prop[1],50),
+             rep(y1prop[2],50))
> lm(y1prop~x)$coefficients
(Intercept)           x
        0.1         0.7
> y1regrx <- lm(y1prop~x)$fitted.values
> var(y1regrx)/var(y1prop)
[1] 1
```

*X*ekiko *Y*ren erregresio logistiko bakuna
Analisi bereizle lineala

$$\ln \frac{Y'}{1-Y'} = b + a \cdot X$$ Erregresio logistikoa

Adibidea:

X \ *Y*	*1*	*0*	
1	40	10	50
0	5	45	50
	45	55	100

Y=1 *in* ***X=x$_i$***	***Odds:#(Y=1)/#(Y=0)*** *in* ***X=x$_i$***	***Log(Odds)*** *in* ***X=x$_i$***
0.8	4.000	1.386
0.1	0.111	-2.197

*X*ekiko *Y*ren erregresio logistiko bakuna
Analisi bereizle lineala

$$\ln\frac{Y'}{1-Y'} = b + a\cdot X$$ Erregresio logistikoa

Adibidea:

Y / X	*1*	*0*	
1	40	10	50
0	5	45	50
	45	55	100

Y=1 *in* ***X=x$_i$***	***Odds:#(Y=1)/#(Y=0)*** *in* ***X=x$_i$***	***Log(Odds)*** *in* ***X=x$_i$***
0.8	4.000	1.386
0.1	0.111	-2.197

$s_{xy} = 0.5\cdot\log 4 - 0.5\cdot(\log 4 - 0.5\cdot\log 9) = 0.25\cdot\log 36$

$s_x^2 = 0.25$

$a = \log 36 = 3.583519$

$b = 0.5\cdot\log(4/9) - 0.5\cdot\log 36 = -\log 9 = -2.197225$

*X*ekiko *Y*ren erregresio logistiko bakuna
Analisi bereizle lineala

$$\ln\frac{Y'}{1-Y'} = b + a\cdot X$$ Erregresio logistikoa

Adibidea:

Y / X	1	0	
1	40	10	50
0	5	45	50
	45	55	100

Y=1 in X=x_i	Odds:#(Y=1)/#(Y=0) in X=x_i	Log(Odds) in X=x_i
0.8	4.000	1.386
0.1	0.111	-2.197

$s_{xy} = 0.5\cdot\log 4 - 0.5\cdot(\log 4 - 0.5\cdot\log 9) = 0.25\cdot\log 36$

$s_x^2 = 0.25$

$a = \log 36 = 3.583519$

$b = 0.5\cdot\log(4/9) - 0.5\cdot\log 36 = -\log 9 = -2.197225$

$$Y' = \frac{1}{1+e^{-(-\log 9 + X\cdot\log 36)}} = \frac{1}{1+9\cdot 36^{-X}}$$

$$0.5 = \frac{1}{1+9\cdot 36^{-x_0}}$$

$$x_0 = \frac{\log 3}{\log 3 + \log 2} = 0.6131472$$

```
> odds <- y1prop/(1-y1prop)
> lm(log(odds) ~ x)$coefficients
(Intercept)           x
  -2.197225    3.583519
```

*X*ekiko *Y*ren erregresio logistiko bakuna
Analisi bereizle lineala

$$\ln\frac{Y'}{1-Y'} = b + a\cdot X$$ Erregresio logistikoa

Adibidea:

Y / X	*1*	*0*	
1	40	10	50
0	5	45	50
	45	55	100

Y=1 ***in X=x_i***	***Odds:#(Y=1)/#(Y=0)*** ***in X=x_i***	***Log(Odds)*** ***in X=x_i***
0.8	4.000	1.386
0.1	0.111	-2.197

$$s_{xy} = 0.5\cdot\log 4 - 0.5\cdot(\log 4 - 0.5\cdot\log 9) = 0.25\cdot\log 36$$

$$s_x^2 = 0.25$$

$$a = \log 36 = 3.583519$$

$$b = 0.5\cdot\log(4/9) - 0.5\cdot\log 36 = -\log 9 = -2.197225$$

$$Y' = \frac{1}{1+e^{-(-\log 9 + X\cdot\log 36)}} = \frac{1}{1+9\cdot 36^{-X}}$$

$$0.5 = \frac{1}{1+9\cdot 36^{-x_0}}$$

$$x_0 = \frac{\log 3}{\log 3 + \log 2} = 0.6131472$$

*X*ekiko *Y*ren erregresio logistiko bakuna
Analisi bereizle lineala

$$\ln\frac{Y'}{1-Y'} = b + a\cdot X$$ Erregresio logistikoa

Adibidea:

Y / X	1	0	
1	40	10	50
0	5	45	50
	45	55	100

Y=1 in X=x_i	Odds:#(Y=1)/#(Y=0) in X=x_i	Log(Odds) in X=x_i
0.8	4.000	1.386
0.1	0.111	-2.197

$s_{xy} = 0.5\cdot\log 4 - 0.5\cdot(\log 4 - 0.5\cdot\log 9) = 0.25\cdot\log 36$

$s_x^2 = 0.25$

$a = \log 36 = 3.583519$

b=0.5·log(4/9)-0.5·log36=-log9=- 2.197225

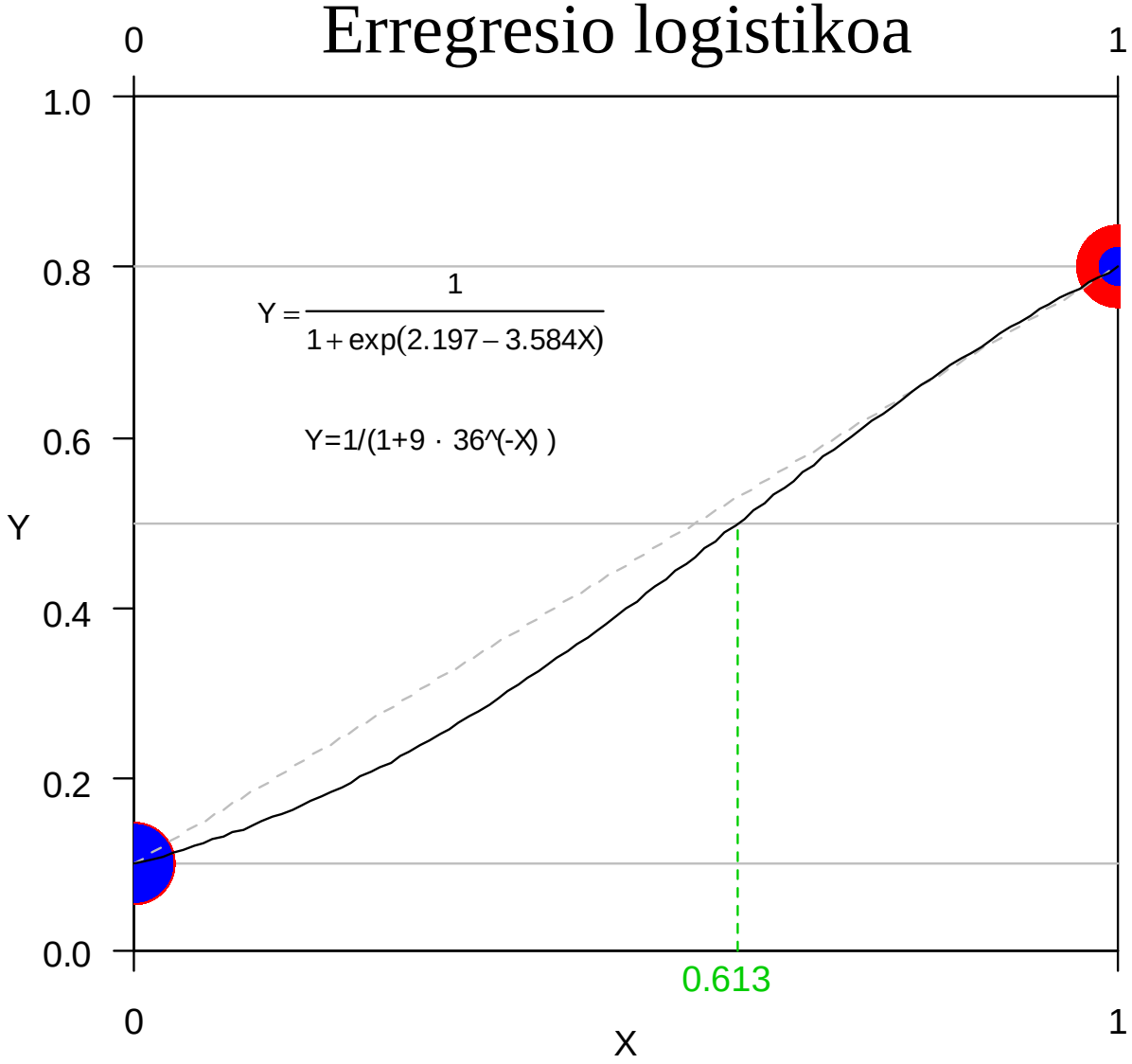
$$Y' = \frac{1}{1+e^{-(-\log 9 + X\cdot\log 36)}} = \frac{1}{1+9\cdot 36^{-X}}$$

$$0.5 = \frac{1}{1+9\cdot 36^{-x_0}}$$

$$x_0 = \frac{\log 3}{\log 3 + \log 2} = 0.6131472$$

```
> odds <- y1prop/(1-y1prop)
> lm(log(odds) ~ x)$coefficients
(Intercept)           x
  -2.197225    3.583519
```

*X*ekiko *Y*ren erregresio logistiko bakuna
Analisi bereizle lineala



*X*ekiko *Y*ren erregresio lineal bakuna
Analisi bereizle lineala

X aldagaia, kualitatiboa da:
p modalitate, $X_1, X_2, \ldots, X_j, \ldots, X_p$.

Y aldagaia, kualitatibo bitarra da:
bati *1* zenbakia egokitu, eta besteari *0*a.

Ω	X	Y
ω_1	x_1	y_1
ω_i	x_i	y_i
ω_n	x_n	y_n

$$x_i \in \{X_1, X_2, \ldots, X_p\}$$

$$y_i \in \{0,1\}$$

Y / *X*	*1*	*0*	
X_1	n_{11}	n_{10}	$n_{1\cdot}$
...	...	...	...
X_i	n_{i1}	n_{i0}	$n_{i\cdot}$
...	...	...	...
X_p	n_{p1}	n_{p0}	$n_{p\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 0}$	$n_{\cdot\cdot} = n$

_X_ekiko _Y_ren erregresio logistiko bakuna
Analisi bereizle lineala
Erregresio logistikoa

Y aldagaia, kualitatiboa da. Modalitate bakoitzeko aldagai bitar bat X_j, batean izan ezik, konbinazio linealik egon ez dadin:

$$X_1, X_2, \ldots, X_j, \ldots, X_{p-1}.$$

$$Y = \frac{1}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \ldots + a_{p-1} \cdot X_{p-1})}} + E$$

$$\ln \frac{Y'}{1 - Y'} = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \ldots + a_{p-1} \cdot X_{p-1}$$