

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

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

$$Y' = f(X_1, X_2, \dots, X_p)$$

Xekiko Yren erregresio lineal anizkoitza

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

- 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 eredian txertatzen.
- Yren erregresioa aztertu bitarra denean, baita X_j aldagaiak bitarrak direnean ere.

Xekiko Yren 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 edozein zenbaki dira.
- Kasu honetan erregresio anizkoitza bi sail bereizteko *analisi bereizle lineala* esaten zaiona da.

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

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

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

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

$$a = \frac{S_{xy}}{S_x^2}$$

$$b = \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x}$$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

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

$$a = \frac{S_{xy}}{S_x^2} \quad \bar{x} = \frac{n_1}{n} \bar{x}_1 + \frac{n_0}{n} \bar{x}_0$$
$$b = \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x} \quad \bar{y} = \frac{n_1}{n}$$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

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

$$a = \frac{s_{xy}}{s_x^2} \quad \bar{x} = \frac{n_1}{n} \bar{x}_1 + \frac{n_0}{n} \bar{x}_0$$

$$b = \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x} \quad \bar{y} = \frac{n_1}{n}$$

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

$$= \frac{n_1}{n} \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(1)} - \bar{x} \frac{n_1}{n} = \frac{n_1}{n} (\bar{x}_1 - \bar{x}) = \frac{n_1 n_0}{n^2} (\bar{x}_1 - \bar{x}_0)$$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X kuantitatiboa

Adibidea:

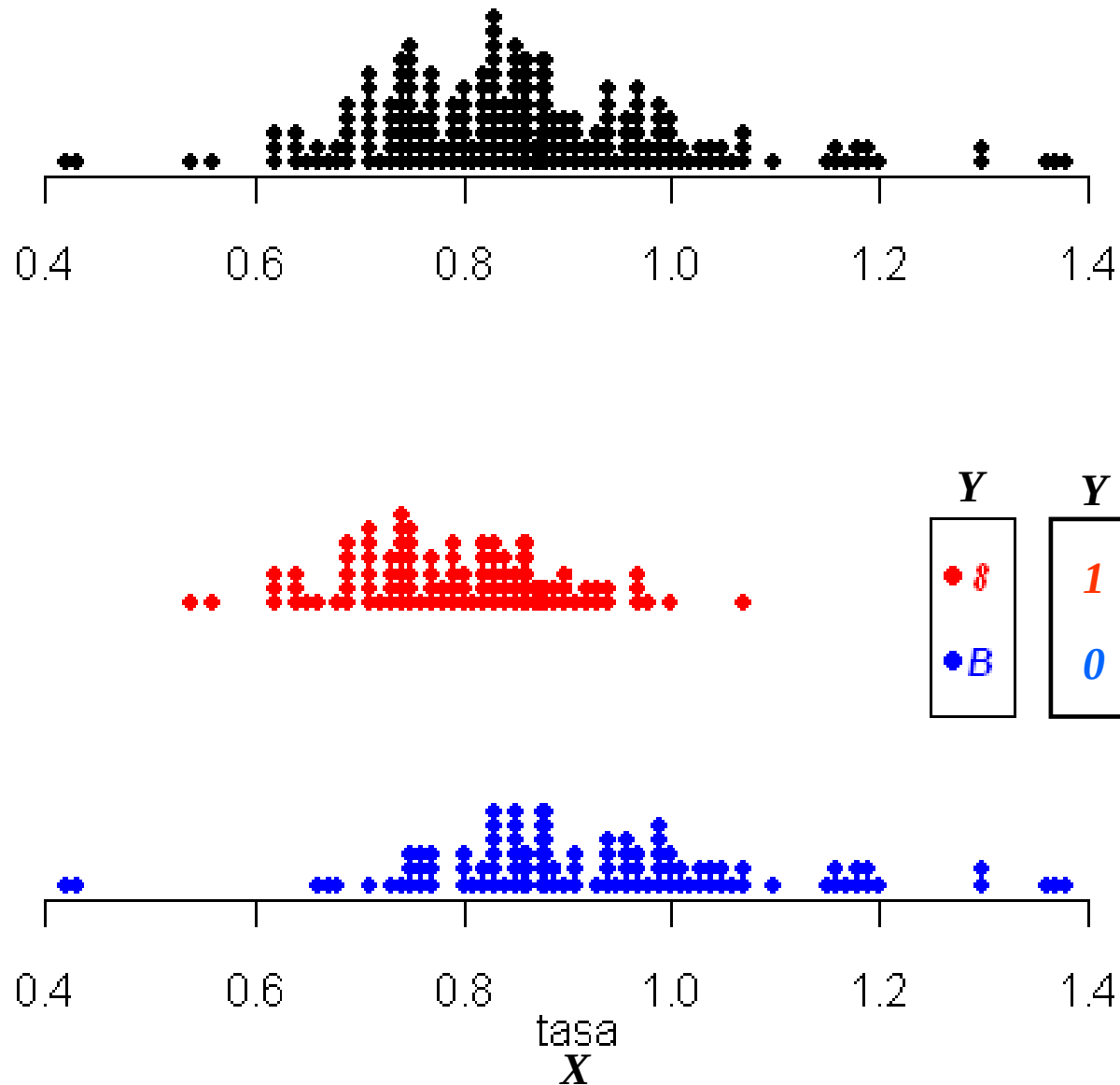
8 digituaren eta **B** letraren formen artean bereizteko metodo bat zuzentasun tasa bat (X) kalkulatzean datza. Tasa hori ikurraren altuera eta ezkerraldeko arkuaren arteko arrazoi edo tasa da.

Sistemara etortzen den ikur baten tasa neurtu ondoren, bi formen artean bereizi egin behar da.

- Zein izango da erabakitze-araua?

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> set.seed(2009)
> n1 <- 100
> n0 <- 100
> zortzi <- round(rnorm(n1, 0.80, 0.10), 2)
> bletra <- round(rnorm(n0, 0.96, 0.15), 2)
> datuak <- matrix(c(c(zortzi, bletra), c(rep(1, n1), rep(0, n0))),
+                 ncol=2, byrow=FALSE)
> datuak <- data.frame(datuak)
> names(datuak) <- c("tasa", "ikurra")
> attach(datuak)
> summary(datuak)
> ikurrai <- as.factor(ikurra)
> par(bty="n")
> par(mar=c(5, 4, 0, 2))
> #
> layout(matrix(c(1, 2), nrow=2))
> stripchart(round(tasa, 2), method="stack",
+           pch=19, cex=0.8, col="black", yaxt="n")
> stripchart(round(tasa, 2)~ikurrai, method="stack",
+           pch=19, cex=0.8, col=c("blue", "red"), yaxt="n")
> text(1.3, 2.1, "8", col="red", cex=0.8, font=4)
> text(1.3, 1.8, "B", col="blue", cex=0.8, font=4)
> points(1.27, 2.1, pch=19, col="red")
> points(1.27, 1.8, pch=19, col="blue")
> rect(1.25, 1.6, 1.33, 2.3)
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(ikurra ~ tasa)$coefficients
```

```
(Intercept)      tasa  
  1.714422    -1.408352
```

```
> table(round(lm(ikurra ~ tasa)$fitted.values,1))
```

```
-0.2 -0.1    0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9    1  1.1  
   3    2    3    6    9   22   24   47   33   35   12    1    1    2
```

```
> var(lm(ikurra ~ tasa)$fitted.values)/var(ikurra)
```

```
[1] 0.1937892
```

```
> detach(datuak)
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(ikurra ~ tasa)$coefficients
```

```
(Intercept)      tasa  
  1.714422    -1.408352
```

```
> table(round(lm(ikurra ~ tasa)$fitted.values,1))
```

-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1
3	2	3	6	9	22	24	47	33	35	12	1	1	2

```
> var(lm(ikurra ~ tasa)$fitted.values)/var(ikurra)
```

```
[1] 0.1937892
```

```
> detach(datuak)
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(ikurra ~ tasa)$coefficients
(Intercept)      tasa 
  1.714422    -1.408352 
> table(round(lm(ikurra ~ tasa)$fitted.values,1))

-0.2 -0.1   0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9   1  1.1
   3    2    3    6    9   22   24   47   33   35   12    1    1    2
> var(lm(ikurra ~ tasa)$fitted.values)/var(ikurra)
[1] 0.1937892
> 
> detach(datuak)
```

Bilatu $y_0=0.5$ erabakitzen duen *tasa*-balioa (x_0)

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(ikurra ~ tasa)$coefficients
(Intercept)      tasa 
  1.714422    -1.408352 
> table(round(lm(ikurra ~ tasa)$fitted.values,1))

-0.2 -0.1   0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9   1  1.1
   3    2    3    6    9   22   24   47   33   35   12    1   1   2
> var(lm(ikurra ~ tasa)$fitted.values)/var(ikurra)
[1] 0.1937892
> detach(datuak)
```

Bilatu $y_0=0.5$ erabakitzen duen *tasa*-balioa (x_0)

$$0.5 = 1.7144 - 1.4084 \cdot x_0 \quad x_0 = 0.8623$$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(ikurra ~ tasa)$coefficients
(Intercept)      tasa
  1.714422    -1.408352
> table(round(lm(ikurra ~ tasa)$fitted.values,1))

-0.2 -0.1    0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9    1  1.1
   3    2    3    6    9   22   24   47   33   35   12    1    1    2
> var(lm(ikurra ~ tasa)$fitted.values)/var(ikurra)
[1] 0.1937892
> detach(datuak)
```

Bilatu $y_0=0.5$ erabakitzen duen *tasa*-balioa (x_0)

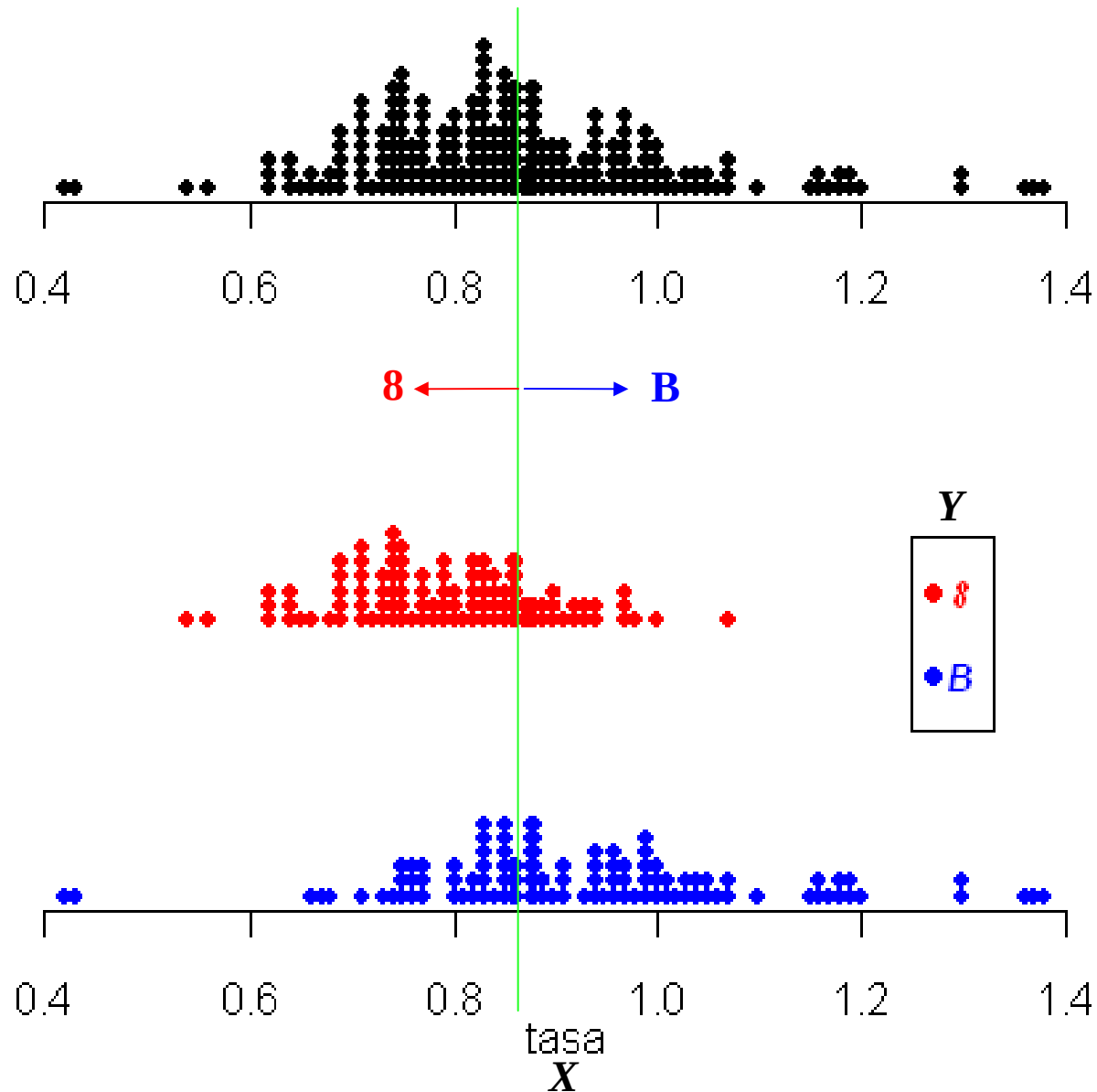
$$0.5 = 1.7144 - 1.4084 \cdot x_0 \quad x_0 = 0.8623$$

Erabakitze-araua:

Baldin $X > 0.8623$ bada, orduan **B**; bestela, **8**

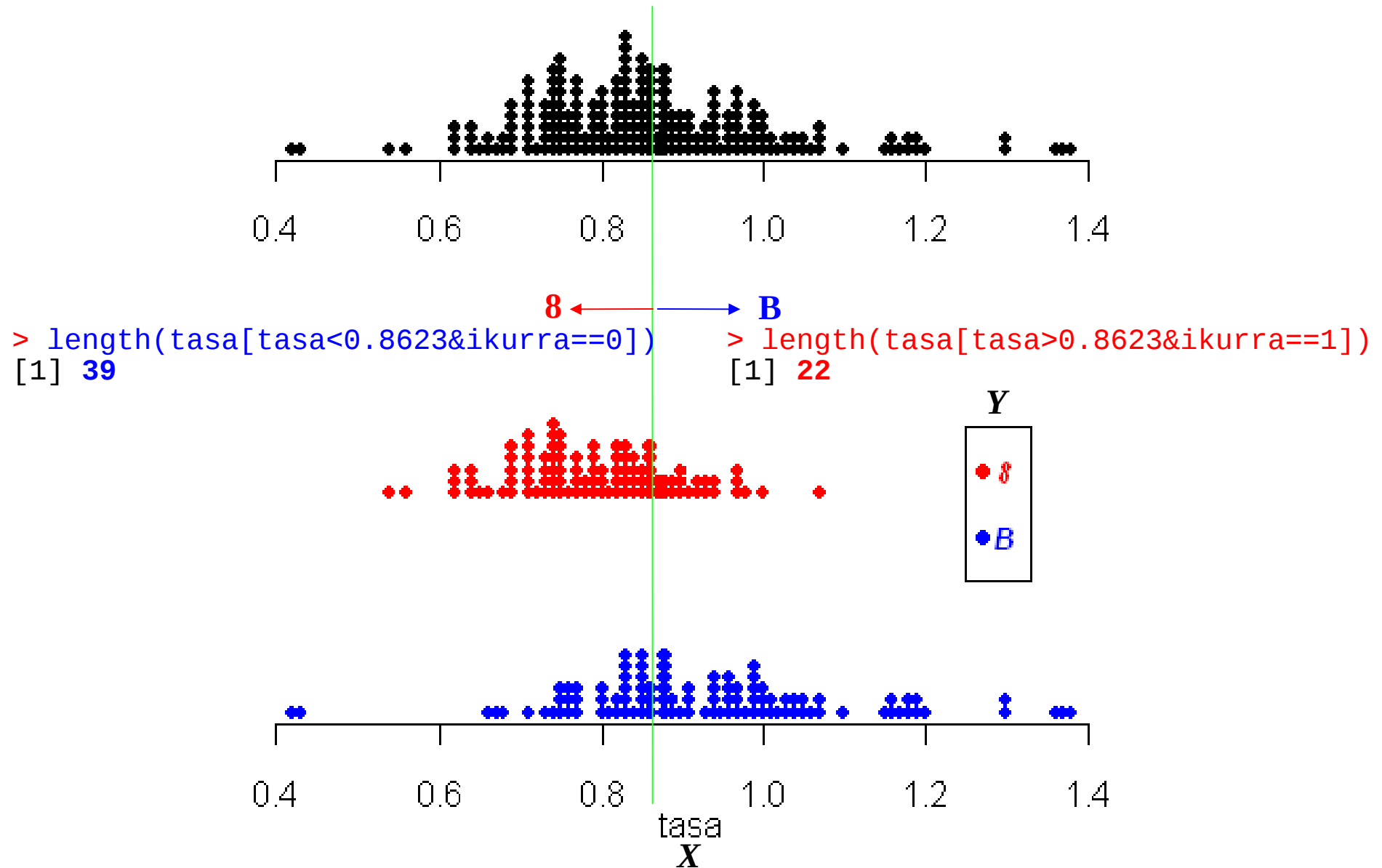
Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

Erabakitze araua:

Baldin bada $X < 0.8623$ orduan **8**; bestela **B**.

Erroreen proportzioak:

Natura Erabaki			
8	0.78	0.22	1.00
B	0.39	0.61	1.00

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

Erabakitze araua:

Baldin bada $X < 0.8623$ orduan **8**; bestela **B**.

Erroreen proportzioak:

Natura Erabaki			
8	0.78	0.22	1.00
B	0.39	0.61	1.00

Erroreen proportzioak hobetzea ez da izan optimizazio-problema ezartzeko erabili den irizpidea; beste bat da.

Xekiko Yren analisi bereizlea

Menpeko aldagaia kualitatibo bitarra

Adibidea:

8 digituaren eta **B** letraren formen artean bereizteko metodo bat zuzentasun tasa bat (X) kalkulatzear datza. Tasa hori ikurraren altuera eta ezkerraldeko arkuaren arteko arrazoi edo tasa da.

Sistemara etortzen den ikur baten tasa neurtu ondoren, bi formen artean bereizi egin behar da.

Bereizketa egitearren, probabilitate-eredu normal bana egokitzen zaie:

8ren kasuan $X \sim N(0.80, 0.10^2)$,

eta **B**renean $X \sim N(0.96, 0.15^2)$

Zein izango da erabakitze-araua bi errore-moten probabilitateak *berdinak* izatea nahi baldin bada?

Xekiko Yren *analisi bereizlea*

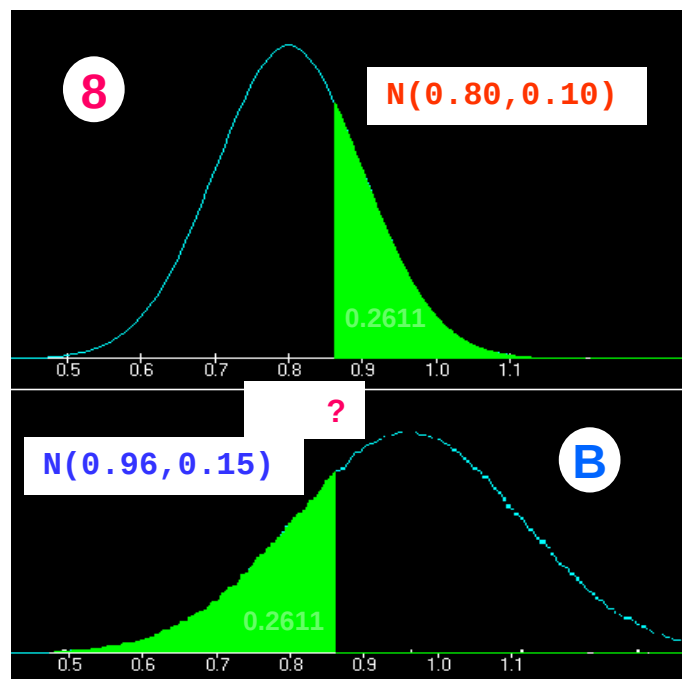
Menpeko aldagaia kualitatibo bitarra

Bereizketa egitearren, probabilitate-eredu normal bana egokitzen zaie:

8ren kasuan $X \sim N(0.80, 0.10^2)$,

eta **B**renean $X \sim N(0.96, 0.15^2)$

Zein izango da erabakitze-araua bi errore-moten probabilitateak *berdinak* izatea nahi baldin bada?



Xekiko Yren analisi bereizlea

Menpeko aldagaia kualitatibo bitarra

Bereizketa egitearren, ***probabilitate-eredu normal bana egokitzen*** zaie:

8ren kasuan $X \sim N(0.80, 0.10^2)$,
eta **B**renean $X \sim N(0.96, 0.15^2)$

Zein izango da erabakitze-araua bi errore-moten probabilitateak *berdinak* izatea nahi baldin bada?

$$\Pr_8(X \leq x_0) = \Pr_B(X > x_0)$$

$$\phi\left(\frac{x_0 - 0.80}{0.10}\right) = 1 - \phi\left(\frac{x_0 - 0.96}{0.15}\right)$$

$$\phi\left(\frac{x_0 - 0.80}{0.10}\right) = \phi\left(\frac{0.96 - x_0}{0.15}\right)$$

$$\frac{x_0 - 0.80}{0.10} = \frac{0.96 - x_0}{0.15}$$

$$x_0 = 0.864$$

Xekiko Yren *analisi bereizlea*

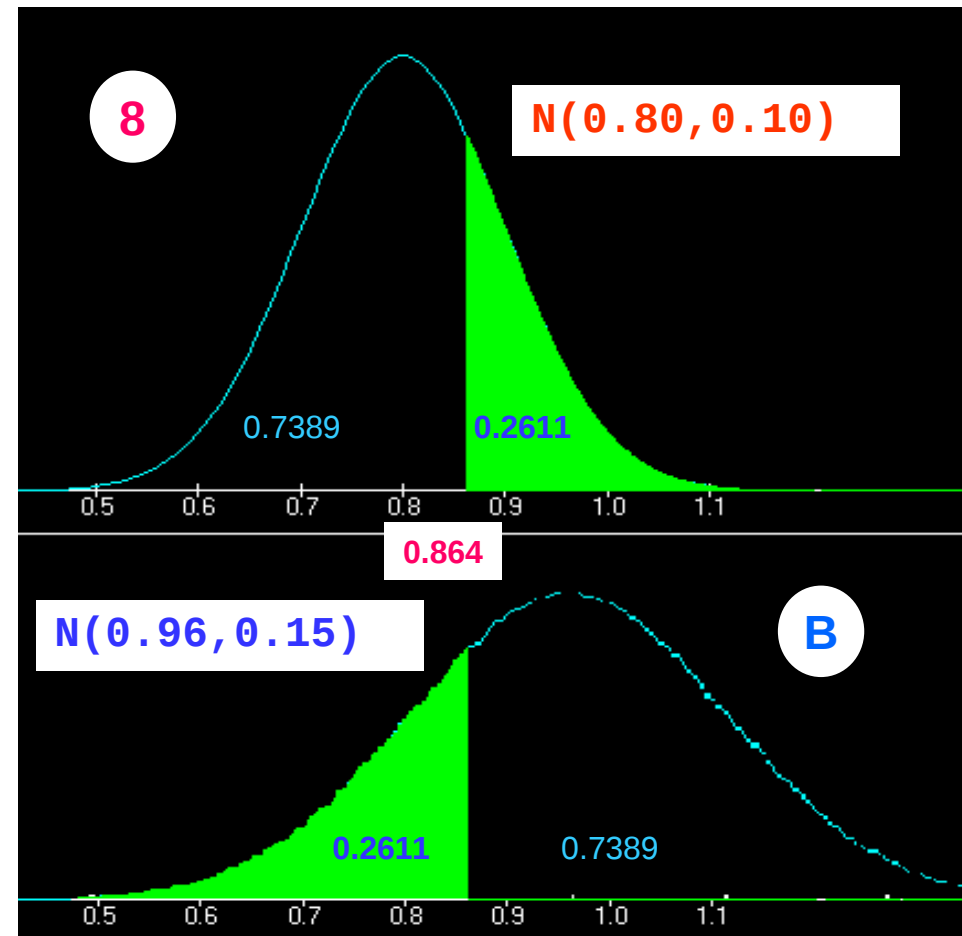
Menpeko aldagaia kualitatibo bitarra

Erabakitze araua:

Baldin bada $X < 0.864$ orduan **8**; bestela **B**.

Erroreen proportzioak:

Natura Erabaki	8	B	
8	0.7389	0.2611	1.00
B	0.2611	0.7389	1.00



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as	
	0	1
28	4	2
29	3	2
30	2	7
31	2	7
32	4	16
33	1	14

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as	
	0	1
28	4	2
29	3	2
30	2	7
31	2	7
32	4	16
33	1	14

```
> asteak <- 28:33
> bularra0 <- c(4,3,2,2,4,1)
> bularra1 <- c(2,2,7,7,16,14)
> pisuak <- bularra0 + bularra1
> asteak <- c(rep(asteak,bularra0),rep(asteak,bularra1))
> bularra <- c(rep(0,sum(bularra0)),rep(1,sum(bularra1)))
> table(asteak , bularra)
```

	bularra	
asteak	0	1
28	4	2
29	3	2
30	2	7
31	2	7
32	4	16
33	1	14

Xekiko Yren erregresio lineal bakuna

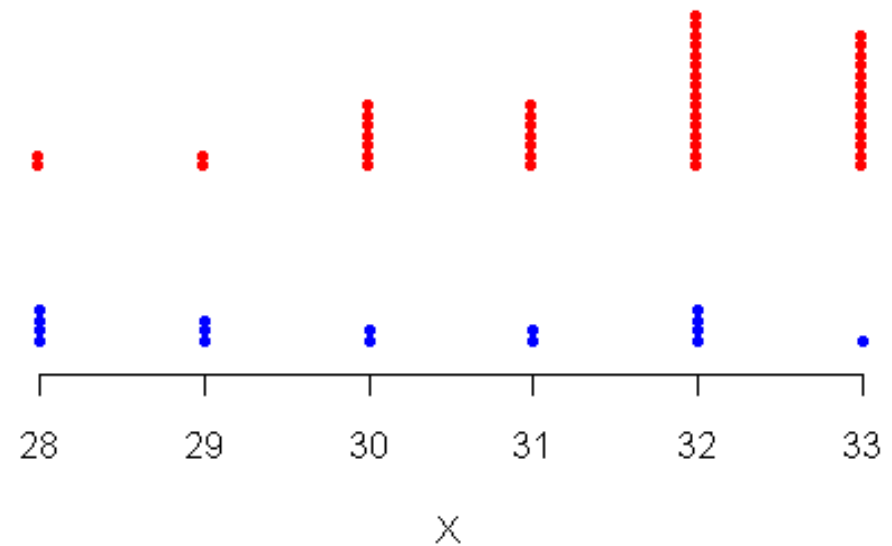
Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as		Total
	0	1	
28	4	2	6
29	3	2	5
30	2	7	9
31	2	7	9
32	4	16	20
33	1	14	15



Xekiko Yren erregresio lineal bakuna

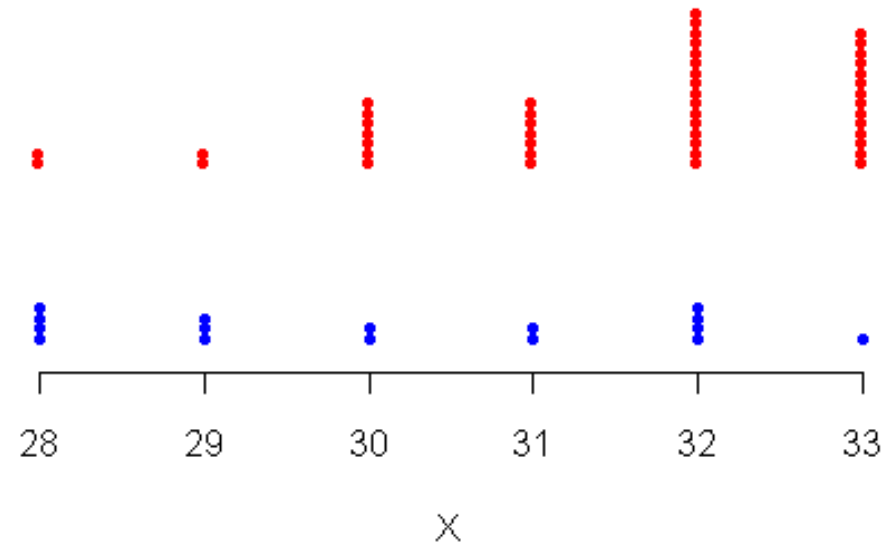
Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333



Xekiko Yren analisi bereizlea

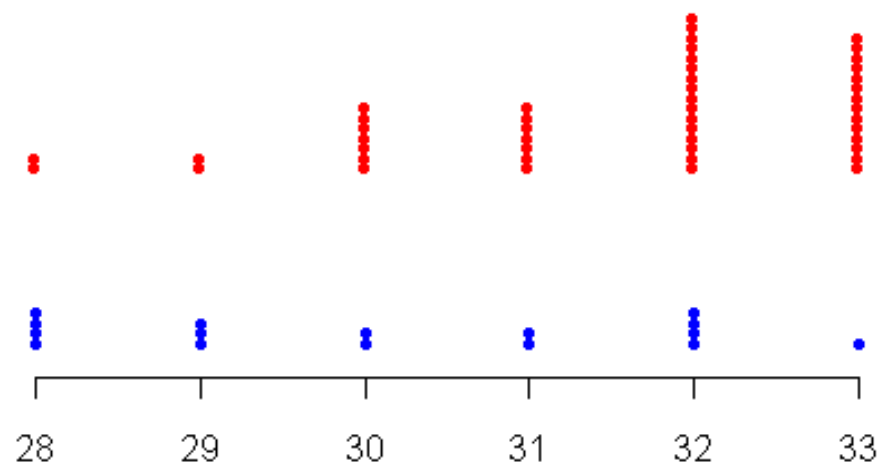
Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as		Total	Freq Y=1 / $X=x_i$
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333



Erabakitze-arau *inozoa* (eredu espliziturik gabe, begiz jota):

Xekiko Yren *analisi bereizlea*

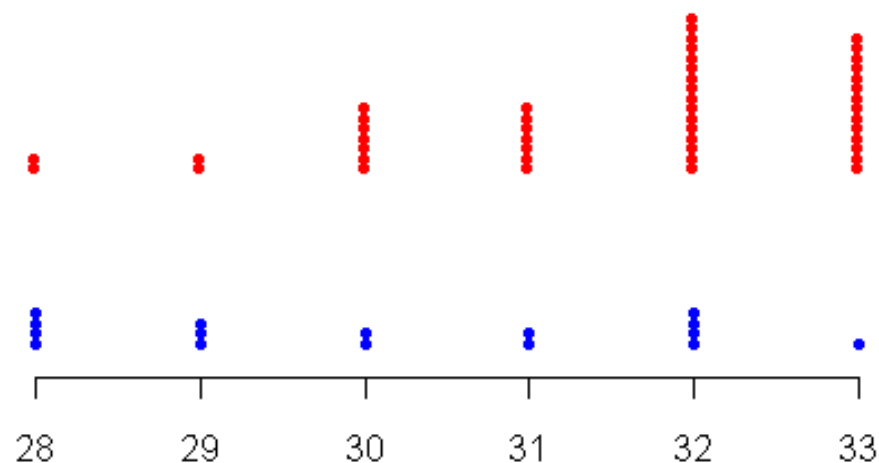
Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as		Total	Freq Y=1 / $X=x_i$
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333



Erabakitze-arau *inozoa* (eredu espliziturik gabe, begiz jota):

Xekiko Yren *analisi bereizlea*

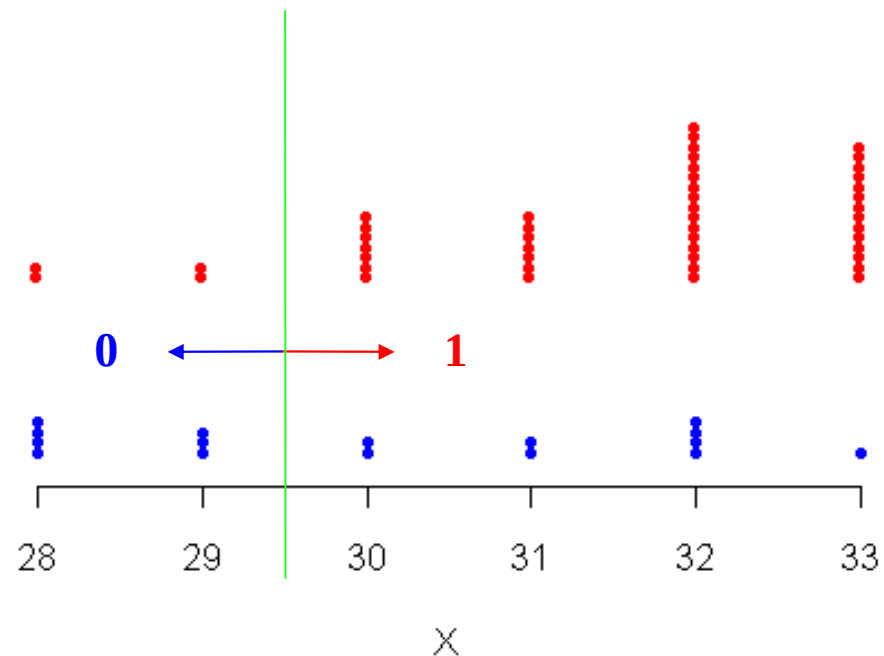
Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as		Total	Freq Y=1 / $X=x_i$
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333



Erabakitze-arau inozoa (eredu espliziturik gabe, begiz jota):

Baldin $X > 29.50$ bada, orduan $Y=1$; bestela, $Y=0$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(bularra~asteak)$coefficients
(Intercept)      asteak
-2.6065478      0.1075709
> var(lm(bularra~asteak)$fitted.values)/var(bularra)
[1] 0.1546331
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(bularra~asteak)$coefficients  
(Intercept)      asteak  
-2.6065478      0.1075709  
> var(lm(bularra~asteak)$fitted.values)/var(bularra)  
[1] 0.1546331
```

Bilatu $y_0=0.5$ erabakitzen duen *tasa*-balioa (x_0)

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(bularra~asteak)$coefficients  
(Intercept)      asteak  
-2.6065478      0.1075709  
> var(lm(bularra~asteak)$fitted.values)/var(bularra)  
[1] 0.1546331
```

Bilatu $y_0=0.5$ erabakitzen duen *tasa*-balioa (x_0)

$$0.5 = -2.6065 + 0.1076 \cdot x_0 \quad x_0 = 28.8708$$

Erabakitze-araua:

Baldin $X > 28.87$ bada, orduan $Y=1$; bestela, $Y=0$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

```
> lm(bularra~asteak)$coefficients
(Intercept)      asteak
  -2.6065478    0.1075709
> var(lm(bularra~asteak)$fitted.values)/var(bularra)
[1] 0.1546331
> table(round(lm(bularra~asteak)$fitted.values,3))
0.405 0.513 0.621 0.728 0.836 0.943
   6     5     9     9    20    15
> table(asteak)
X
28 29 30 31 32 33
 6  5  9  9 20 15
```

Bilatu $y_0=0.5$ erabakitzen duen *tasa*-balioa (x_0)

$$0.5 = -2.6065 + 0.1076 \cdot x_0 \quad x_0 = 28.8708$$

Erabakitze-araua:

Baldin $X > 28.87$ bada, orduan $Y=1$; bestela, $Y=0$

Xekiko Yren erregresio lineal bakuna

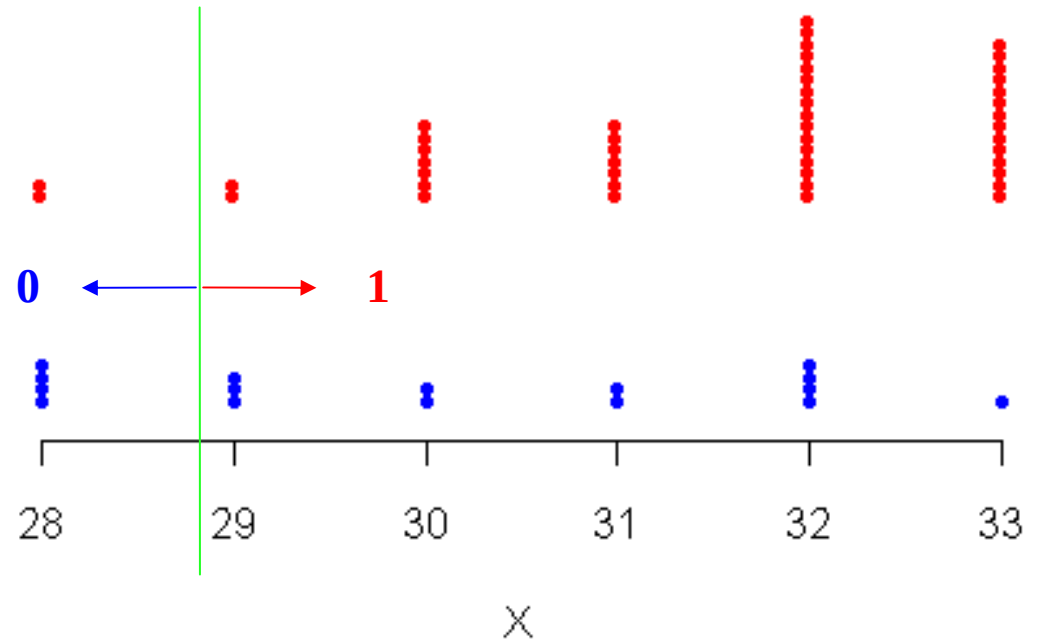
Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as		Total ii+iii
	0	1	
28	4	2	6
29	3	2	5
30	2	7	9
31	2	7	9
32	4	16	20
33	1	14	15



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

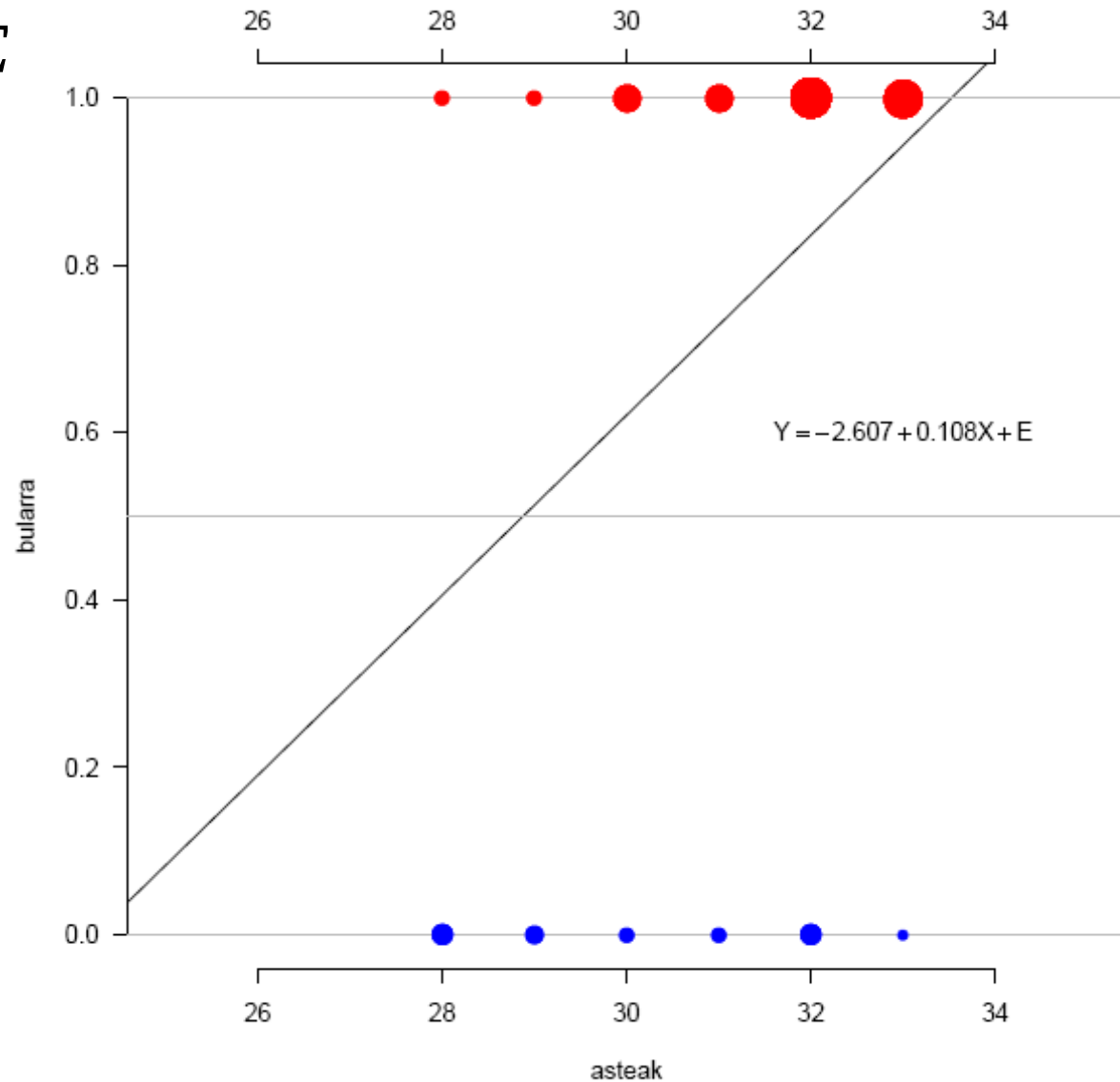
$$Y = b + a \cdot X + E$$

```
> par(bty="n")
> plot(asteak,bularra, xlim=c(25,35), ylim=c(0,1), las=1, pch="")
> axis(3)
> abline(a=lm(bularra~asteak)$coefficients[1],
+        b=lm(bularra~asteak)$coefficients[2])
> abline(h=c(0,0.5,1), col="gray")
> points(asteak,bularra, pch=rep(19,sum(pisuak)),
+        cex=c(rep(7*sqrt(bularra0/sum(pisuak)),bularra0),
+              rep(7*sqrt(bularra1/sum(pisuak)),bularra1)),
+        col=c(rep("red",sum(bularra0)),rep("blue",sum(bularra1))))
> text(33,0.6,expression(Y== -2.607+0.108*X+E))
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

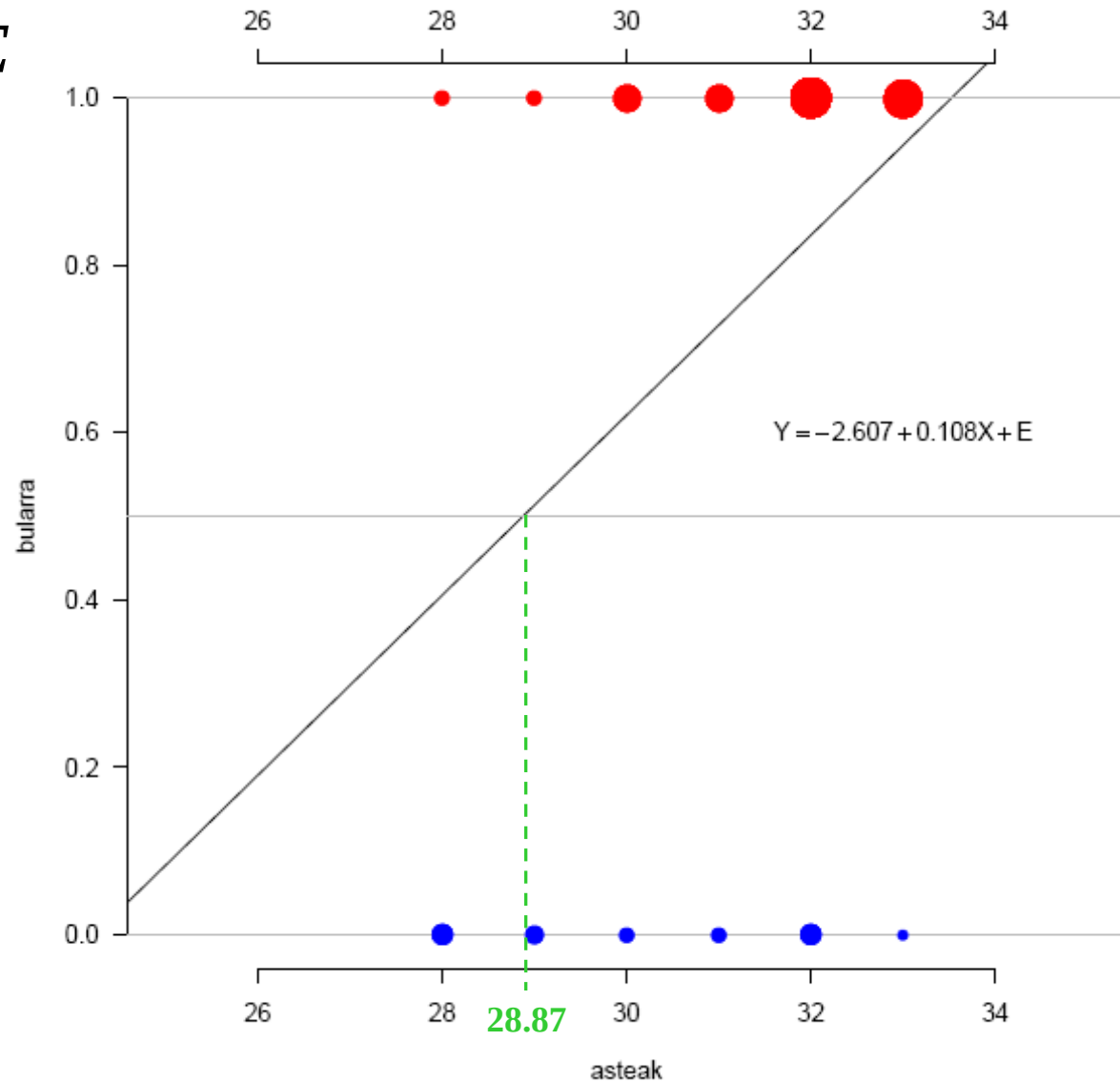
$$Y = b + a \cdot X + E$$



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333

```
> asteak <- 28:33
> bular1prop <- bularra1/pisuak
> asteak <- rep(asteak, pisuak)
> bular1prop <- rep(bular1prop, pisuak)
> lm(bular1prop~asteak)$coefficients
(Intercept)      asteak
-2.6065478    0.1075709
> var(lm(bular1prop~asteak)$fitted.values)/var(bular1prop)
[1] 0.835019
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333

```
> asteak <- 28:33
> bular1prop <- bularra1/pisuak
> asteak <- rep(asteak, pisuak)
> bular1prop <- rep(bular1prop, pisuak)
> lm(bular1prop~asteak)$coefficients
(Intercept)      asteak
-2.6065478      0.1075709
> var(lm(bular1prop~asteak)$fitted.values)/var(bular1prop)
[1] 0.835019
```

```
> lm(bularra~asteak)$coefficients
(Intercept)      asteak
-2.6065478      0.1075709
> var(lm(bularra~asteak)$fitted.values)/\
[1] 0.1546331
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

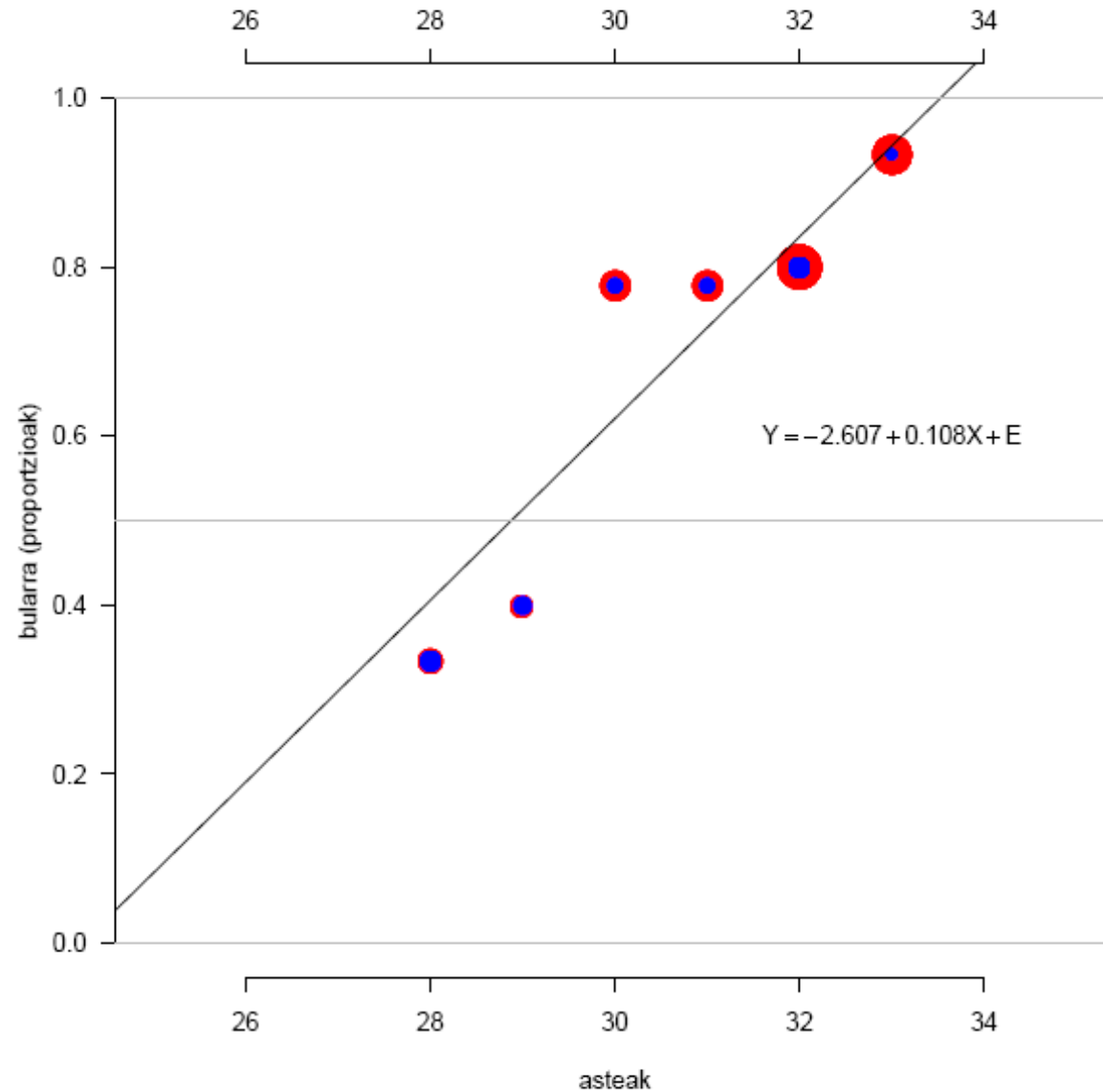
```
> par(bty="n")
> plot(asteak,bular1prop, xlim=c(25,35), ylim=c(0,1), las=1, ylab="", pch="")
> points(asteak,bular1prop, pch=rep(19,length(asteak)),
+       cex=rep(7*sqrt(pisuak/sum(pisuak)),pisuak),
+       col=rep("red",length(asteak)))
> points(asteak,bular1prop, pch=rep(19,length(asteak)),
+       cex=rep(7*sqrt(bularra0/sum(pisuak)),pisuak),
+       col=rep("blue",length(asteak)))
> mtext("bularra (proportzioak)", 2, line=2.5)
> axis(3)
> abline(a=lm(bular1prop~asteak)$coefficients[1],
+       b=lm(bular1prop~asteak)$coefficients[2])
> abline(h=c(0,0.5,1), col="gray")
> text(33,0.6,expression(Y== -2.607+0.108*X+E))
```

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333

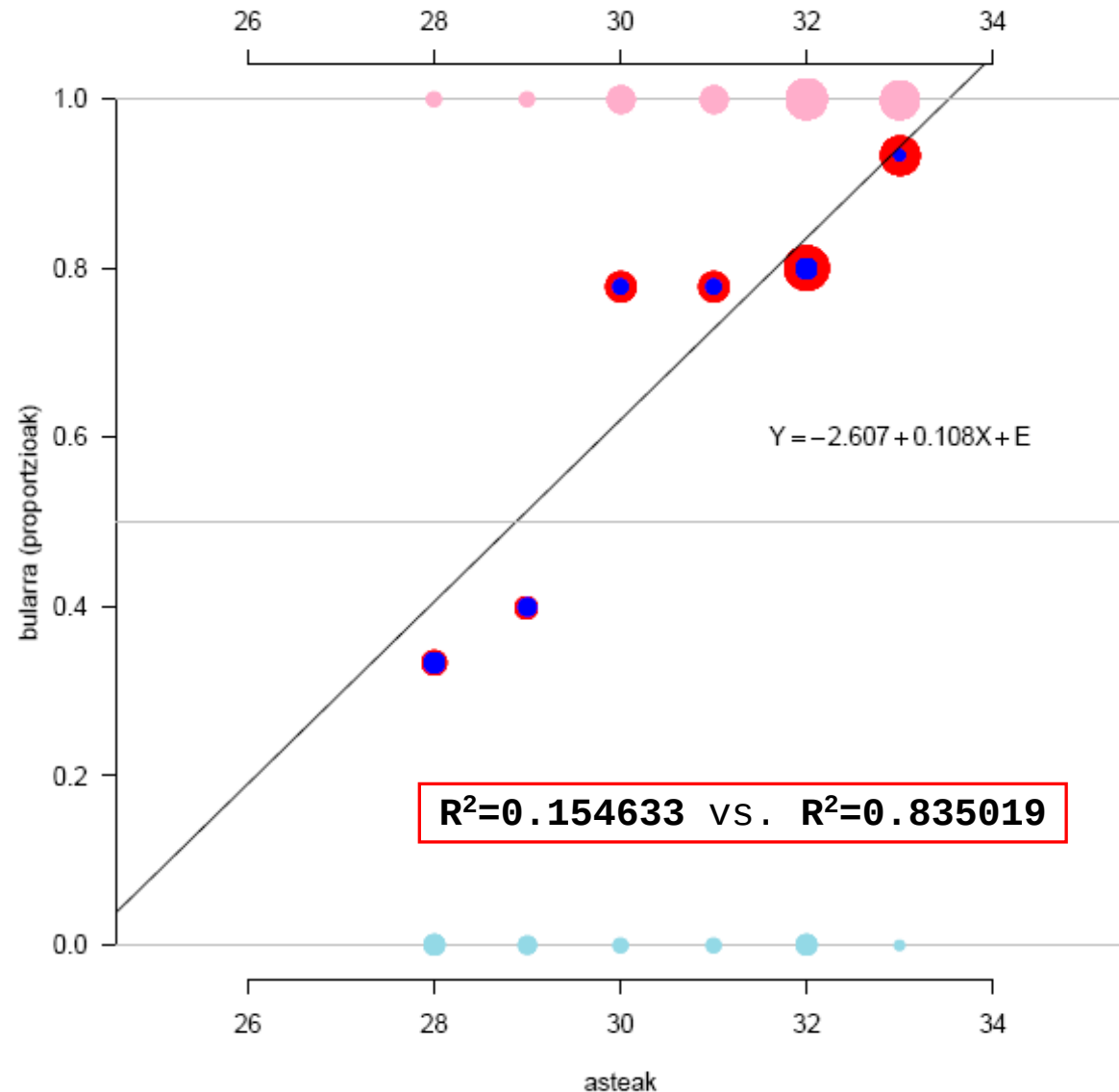


Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333



Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
X				
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333

```
> lm(bularra~asteak)$coefficients
(Intercept)      asteak
-2.6065478      0.1075709
```

```
> lm(bular1prop~asteak)$coefficients
(Intercept)      asteak
-2.6065478      0.1075709
```

$$a = \frac{S_{xy}}{S_x^2}$$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
X				
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333

```
> lm(bularra~asteak)$coefficients
(Intercept)      asteak
-2.6065478      0.1075709
```

```
> lm(bular1prop~asteak)$coefficients
(Intercept)      asteak
-2.6065478      0.1075709
```

$$a = \frac{S_{xy}}{S_x^2}$$

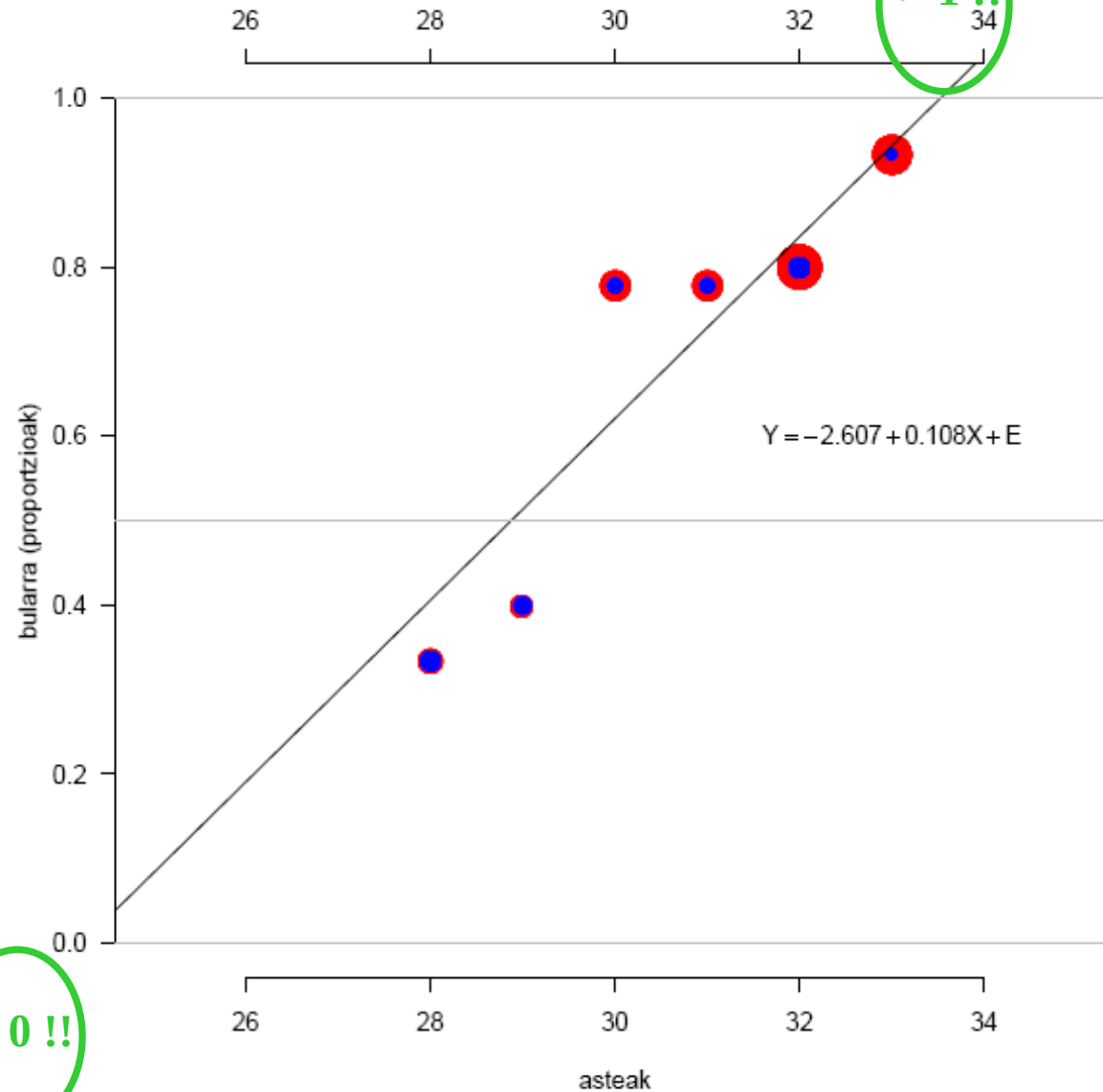
$$\begin{aligned}
 S_{xy} &= \frac{1}{n} \sum_{i=1}^n (X(\omega_i) - \bar{x})(Y(\omega_i) - \bar{y}) = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})(y_j - \bar{y}) = \\
 &= \sum_{i=1}^p \sum_{j=1}^q f_{ij} (x_i - \bar{x})(y_j - \bar{y}) = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q f_i \frac{f_{ij}}{f_i} (x_i - \bar{x})(y_j - \bar{y}) = \\
 &= \frac{1}{n} \sum_{i=1}^p f_i (x_i - \bar{x}) \sum_{j=1}^q \frac{f_{ij}}{f_i} (y_j - \bar{y}) = \frac{1}{n} \sum_{i=1}^p f_i (x_i - \bar{x}) (\bar{y}_i^j - \bar{y})
 \end{aligned}$$

Xekiko Yren erregresio lineal bakuna

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 / X=x _i
	0	1		
28	4	2	6	.3333
29	3	2	5	.4000
30	2	7	9	.7778
31	2	7	9	.7778
32	4	16	20	.8000
33	1	14	15	.9333



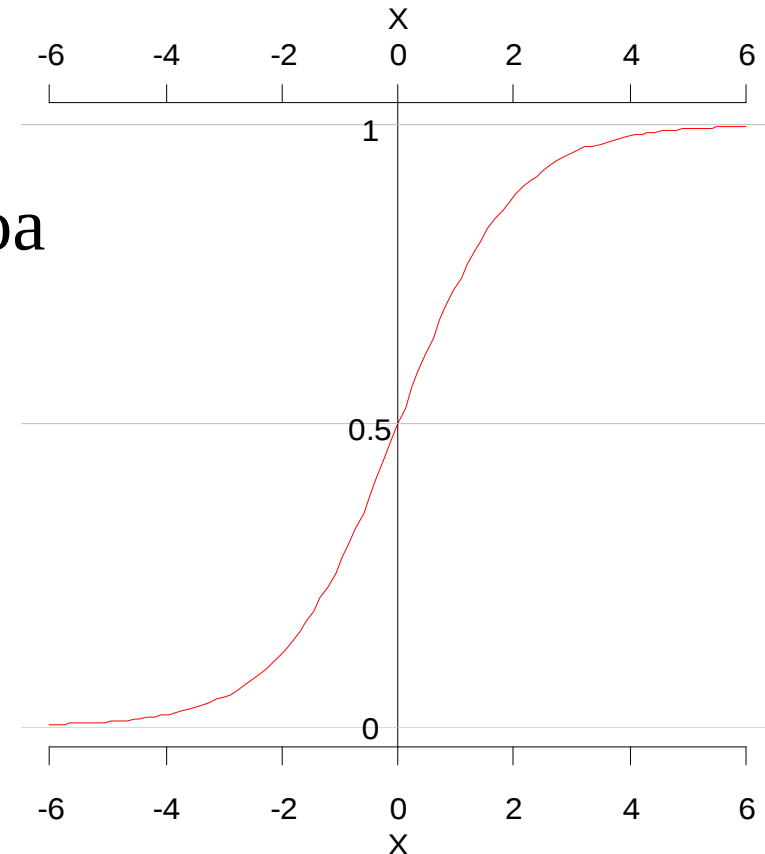
Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

Funtzio *logistikoa*.

Forma dela eta, *sigmoide* funtzioa ere esaten zaio:

$$f(x) = \frac{1}{1 + e^{-x}}$$



```
> curve(1/(1+exp(-x)), from = -6, to = +6, n = 101,  
+       ann=FALSE, yaxt="n", col="red")  
> axis(3)  
> abline(v=0)  
> abline(h=c(0,0.5,1), col="gray")  
> mtext("X",1,line=2)  
> mtext("X",3,line=2)  
> text(rep(-0.5,3), c(0,0.5,1), c(0,0.5,1))
```

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

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

$$Y = \frac{1}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}} + E$$

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

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

$$Y = \frac{1}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}} + E$$

$$Y' = \frac{1}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}}$$

$$1 - Y' = \frac{e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}}$$

$$\frac{1 - Y'}{Y'} = e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}$$

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

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

$$Y = \frac{1}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}} + E$$

$$Y' = \frac{1}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}}$$

$$1 - Y' = \frac{e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}}{1 + e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}}$$

$$\frac{1 - Y'}{Y'} = e^{-(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)}$$

$$\ln \frac{1 - Y'}{Y'} = -(a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p)$$

$$\ln \frac{Y'}{1 - Y'} = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_p \cdot X_p$$

Logit ereduak ere esaten zaio.

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

X	Instances of Y Coded as		Total	Freq Y=1 in $X=x_i$	Odds: $\frac{\#(Y=1)}{\#(Y=0)}$ in $X=x_i$	log(Odds) in $X=x_i$
	0	1				
28	4	2	6	.3333	.5000	-.6931
29	3	2	5	.4000	.6667	-.4055
30	2	7	9	.7778	3.5000	1.2528
31	2	7	9	.7778	3.5000	1.2528
32	4	16	20	.8000	4.0000	1.3863
33	1	14	15	.9333	14.0000	2.6391

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

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

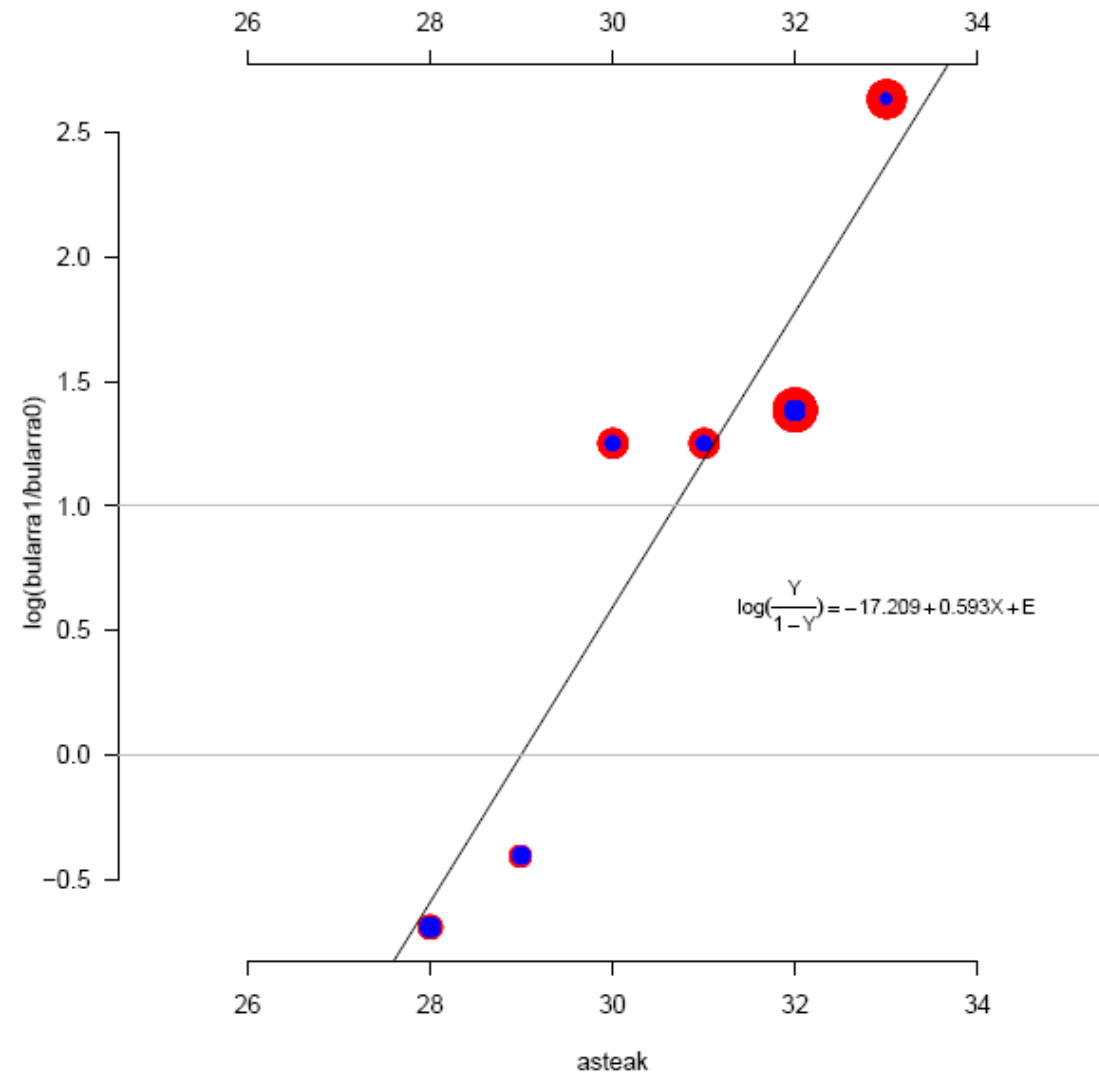
```
> asteak <- 28:33
> asteak <- rep(asteak,pisuak)
> odds <- bularra1/bularra0
> odds <- rep(odds,pisuak)
> lm(log(odds) ~ asteak)$coefficients
(Intercept)      asteak
-17.2085565    0.5934015
> var(lm(log(odds) ~ asteak)$fitted.values)/var(log(odds))
[1] 0.862605
>
> par(bty="n")
> plot(asteak,log(odds), xlim=c(25,35), las=1, ylab="", pch="")
> points(asteak,log(odds), pch=rep(19,length(asteak)),
+       cex=rep(7*sqrt(pisuak/sum(pisuak)),pisuak),
+       col=rep("red",length(asteak)))
> points(asteak,log(odds), pch=rep(19,length(asteak)),
+       cex=rep(7*sqrt(bularra0/sum(pisuak)),pisuak),
+       col=rep("blue",length(asteak)))
> mtext("log(bularra1/bularra0)", 2, line=2.5)
> axis(3)
> abline(a=lm(log(odds)~asteak)$coefficients[1],
+       b=lm(log(odds)~asteak)$coefficients[2] )
> abline(h=c(0,1), col="gray")
> text(33,0.6,expression(log(frac(Y,1-Y))=-17.209+0.593*X + E), cex=0.8)
```

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

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

X	Instances of Y Coded as		Total	Freq Y=1 in X=x _i	Odds: #(Y=1) #(Y=0) in X=x _i	log(Odds) in X=x _i
	0	1				
28	4	2	6	.3333	.5000	-.6931
29	3	2	5	.4000	.6667	-.4055
30	2	7	9	.7778	3.5000	1.2528
31	2	7	9	.7778	3.5000	1.2528
32	4	16	20	.8000	4.0000	1.3863
33	1	14	15	.9333	14.0000	2.6391



Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{-(b+a \cdot X)}} + E$$

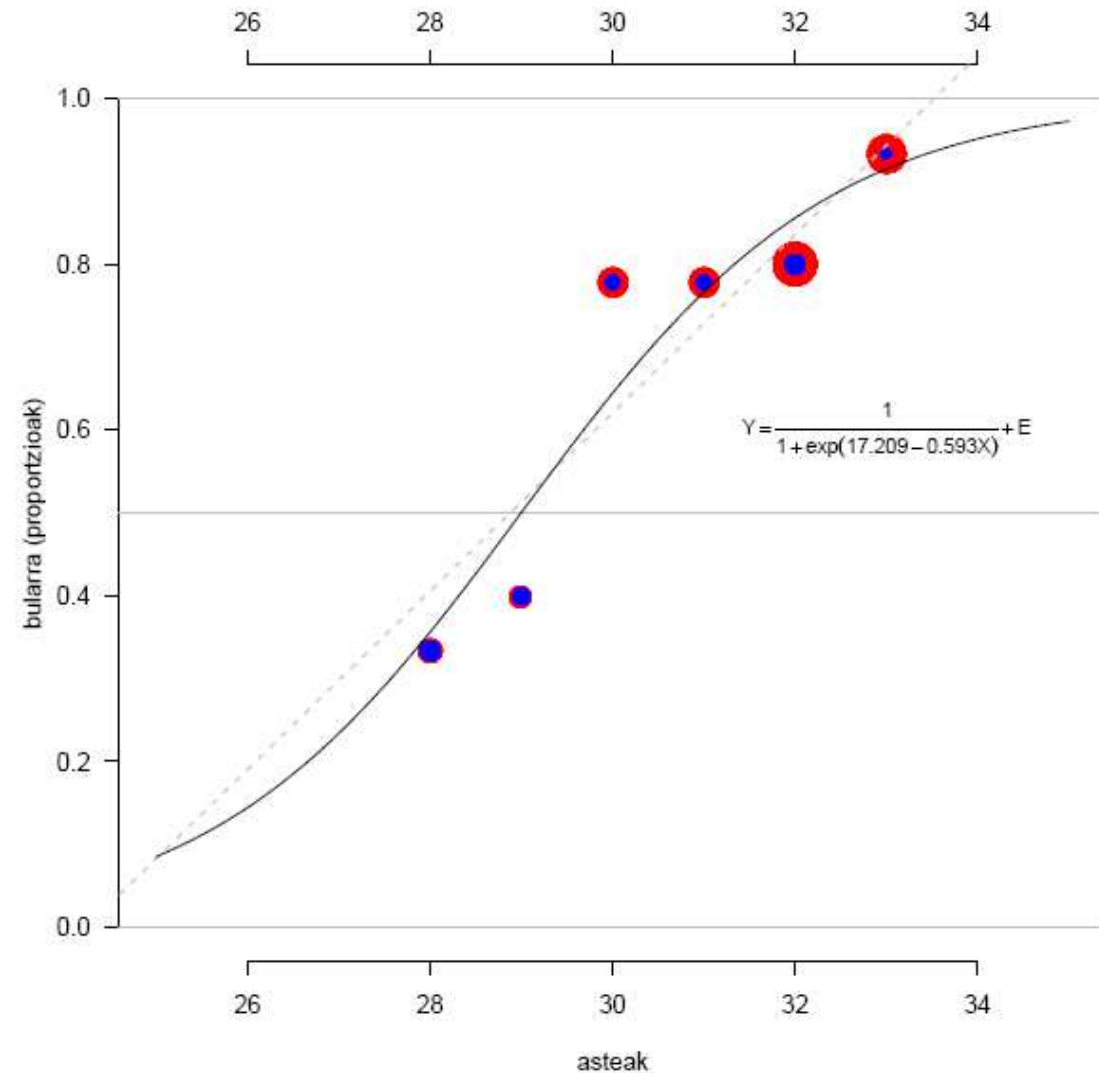
```
> par(bty="n")
> plot(asteak,bular1prop, xlim=c(25,35), ylim=c(0,1), las=1, ylab="", pch="")
> points(asteak,bular1prop, pch=rep(19,length(asteak)),
+       cex=rep(7*sqrt(pisuak/sum(pisuak)),pisuak),
+       col=rep("red",length(asteak)) )
> points(asteak,bular1prop, pch=rep(19,length(asteak)),
+       cex=rep(7*sqrt(bularra0/sum(pisuak)),pisuak),
+       col=rep("blue",length(asteak)) )
> mtext("bularra (proportzioak)", 2, line=2.5)
> axis(3)
> abline(a=lm(bular1prop~asteak)$coefficients[1],
+       b=lm(bular1prop~asteak)$coefficients[2], col="gray", lty="dashed" )
> abline(h=c(0,0.5,1), col="gray")
> x <- lm(log(odds) ~ asteak)$coefficients[1]+
+       lm(log(odds) ~ asteak)$coefficients[2]*seq(25,35, length.out=101)
> lines(seq(25,35, length.out=101), 1/(1+exp(-x)) )
> text(33,0.6,expression(Y==frac(1,1+exp(17.209-0.593*X))+ E), cex=0.8)
```

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{-(b+a \cdot X)}} + E$$

X	Instances of Y Coded as		Total	Freq Y=1 in X=x _i	Odds # (Y=1) # (Y=0) in X=x _i	log(Odds) in X=x _i
	0	1				
28	4	2	6	.3333	.5000	-.6931
29	3	2	5	.4000	.6667	-.4055
30	2	7	9	.7778	3.5000	1.2528
31	2	7	9	.7778	3.5000	1.2528
32	4	16	20	.8000	4.0000	1.3863
33	1	14	15	.9333	14.0000	2.6391



Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{-(b+a \cdot X)}} + E$$

```
> bular1esp <- 1/(1+exp(17.209-0.593*asteak))  
> var(bular1esp)/var(bular1prop)  
[1] 0.896202  
> var(bular1prop-bular1esp)/var(bular1prop)  
[1] 0.1273692  
>  
> var(lm(bular1prop~asteak)$fitted.values)/var(bular1prop)  
[1] 0.835019  
> var(lm(bular1prop~asteak)$fitted.values-bular1prop)/var(bular1prop)  
[1] 0.164981  
>
```

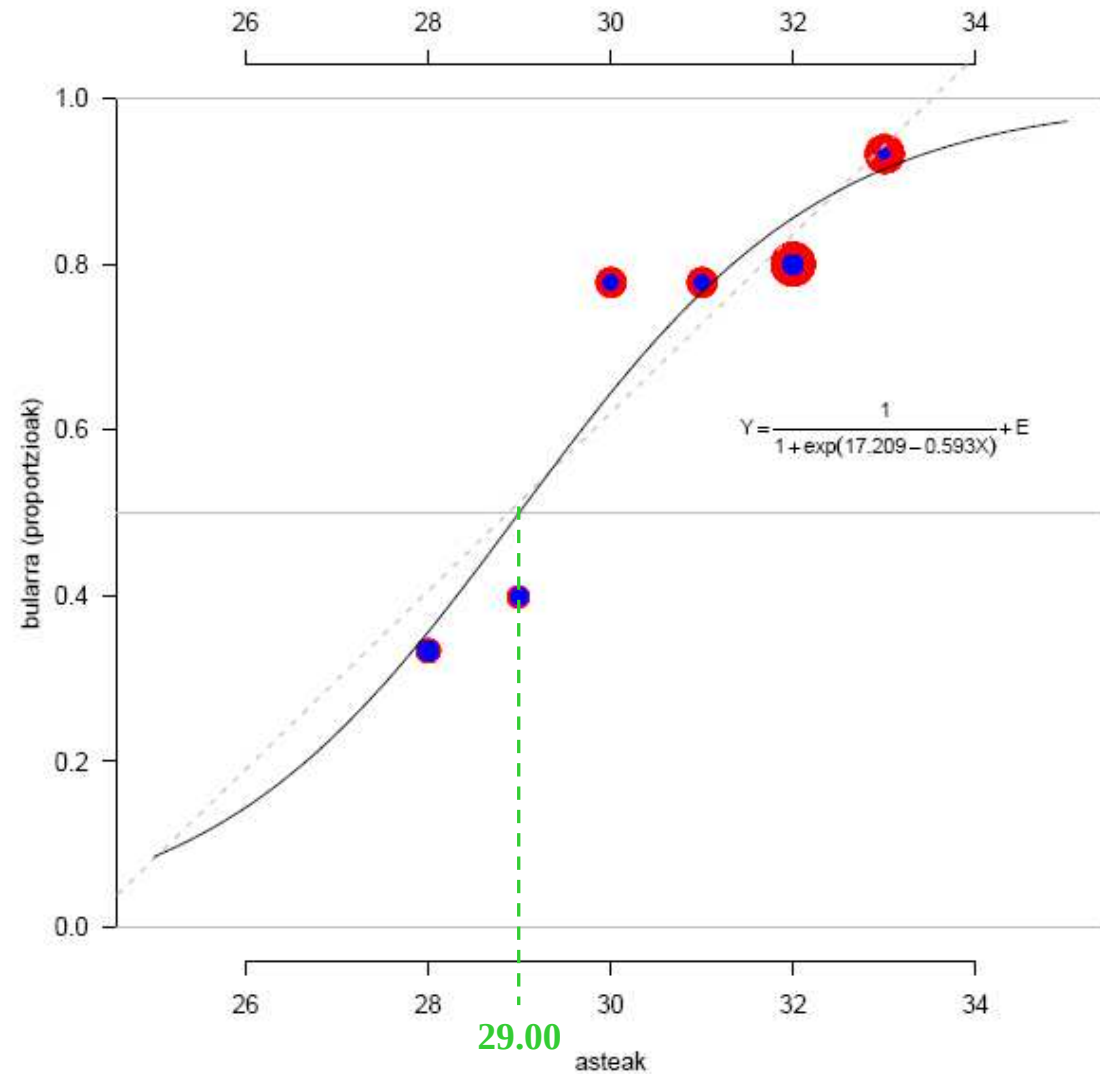


Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{-(b+a \cdot X)}} + E$$

X	Instances of Y Coded as		Total	Freq / X=x _i	Y=1 Odds Y=1/Y=0 / X=x _i	ln(Odds) / X=x _i
	0	1				
28	4	2	6	.3333	.5000	-.6931
29	3	2	5	.4000	.6667	-.4055
30	2	7	9	.7778	3.5000	1.2528
31	2	7	9	.7778	3.5000	1.2528
32	4	16	20	.8000	4.0000	1.3863
33	1	14	15	.9333	14.0000	2.6391



Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{-(b+a \cdot X)}} + E$$

$$Y' = \frac{1}{1 + e^{-(b+a \cdot X)}}$$

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

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{b+a \cdot X}} + E$$

$$Y' = \frac{1}{1 + e^{b+a \cdot X}}$$

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

Bilatu $y_0=0.5$ erabakitzen duen ***aste-kopurua*** (x_0)

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{b+a \cdot X}} + E$$

$$Y' = \frac{1}{1 + e^{b+a \cdot X}}$$

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

Bilatu $y_0 = 0.5$ erabakitzen duen ***aste-kopurua*** (x_0)

$$\ln \frac{0.5}{1 - 0.5} = b + a \cdot x_0 \quad x_0 = -\frac{b}{a} \quad x_0 = 28.99985$$

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

$$Y = \frac{1}{1 + e^{b+a \cdot X}} + E$$

$$Y' = \frac{1}{1 + e^{b+a \cdot X}}$$

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

Bilatu $y_0=0.5$ erabakitzen duen ***aste-kopurua*** (x_0)

$$\ln \frac{0.5}{1 - 0.5} = b + a \cdot x_0 \quad x_0 = -\frac{b}{a} \quad x_0 = 28.99985$$

Erabakitze-araua:

Baldin $X < 29$ bada, orduan $Y=0$; bestela, $Y=1$

Xekiko Yren erregresio logistikoa

Menpeko aldagaia kualitatibo bitarra

Adibidea: 64 jaioberri.

X: Jaio aurretik amaren sabelean izandako aste-kopurua.

Y: Bularra eman zioten ($Y=1$) ala ez ($Y=0$).

Hiru erabakitze-arau, datuen *maiztasun-taularen* aurrean:

- Begiz jota.
- Erregresio lineala.
- Erregresio logistikoa.