BLOC

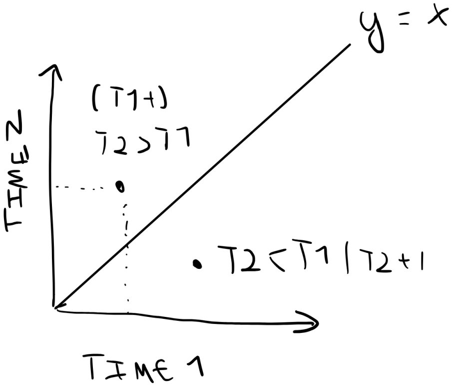
D

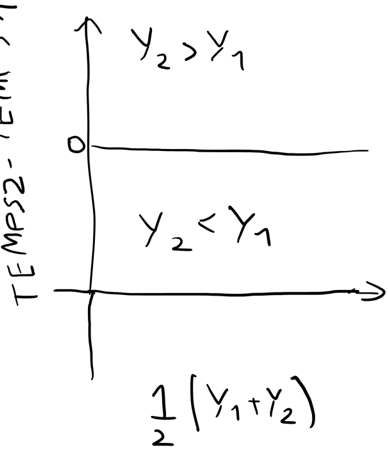
5/12 => PS1

13/12 => PS2

• Mostres aparellades — mar → Plant - Altman

→ Bland - Altman





Si hay diferència, los punts per sobre o per sota

• Proves d'hipòtesi (PH):

Exemple embotelladora: $\bar{X} = 997$ c.c , $n = 50$

X: volum $S = 10$ c.c

H_0 (hip nul·la): $\mu = 1000$ c.c (el que ens diu el fabricant)
(l'acceptarem o no?)

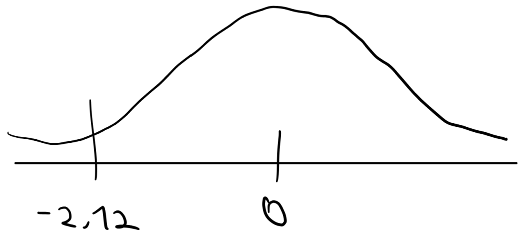
H_0 vs H_1 (hip alternativa) (tot en lo contrari) $H_1: \mu \neq 1000$ c.c

És probable que la mitjana de 50 ampolles sigui 997 cc si és correcte H_0?

Decidirem nivell de confiança: $1-\alpha$, per ex el 95%
nivell de significació: α, ex 5%

$$TCL: \bar{X}_n \overset{H_0}{\sim} N\left(\underset{\substack{\downarrow \\ 1000\ H_0}}{\mu}, \frac{\sigma}{\sqrt{n}}\right)$$

Estandarditzem i usem $\hat{\sigma}$



$$\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

$\boxed{\text{estadístic de contrast}}$

per veure si hip és correcta

Con Ho, 1000 tendría que estar cerca de 0, cuanto más lejos menos me lo creo

- p-valor: probabilitat sota Ho de que la variable t

$t = -2{,}12$

$$p = P_{H_0}(T \leq t \cup T \geq -t)$$

?

$$p = pt(-2{,}12, 49) \cdot \underbrace{2}_{\text{simetria}} = \underline{0{,}04} < \alpha$$

(0,02)

0,04 < 0,05 => Rechazamos Ho

=> $\mu < 1000$ c.i.

Afirmo amb el 95% que la màquina no funciona bé

Ara ara $\bar{X} = 997$ c.c, $n=30$, $S=10$, $1-\alpha=0.95$

$t = -1,64$ (quantil)

$p = P_{H_0}(-1,64,\ 29) \cdot 2 = 0,11 > 0,05$

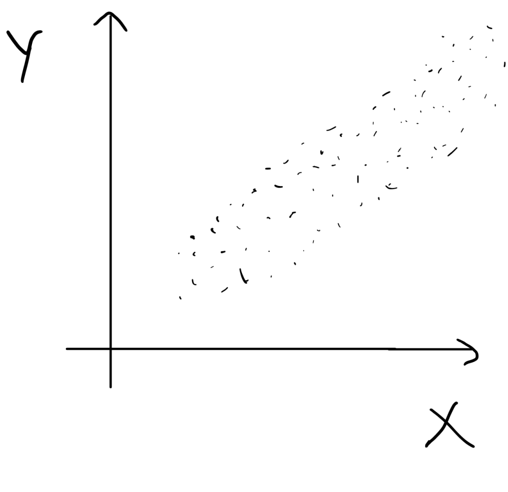
$\Downarrow$

<u>NO</u> rechazamos Ho, pero <u>no</u> lo podemos afirmar con confianza de 95%

<u>Model lm(Y ~ X). Consum i alcohol</u>

(regresión lineal simple) (RLS)

¿ Y ~ X ? (existeix relació entre X i Y?)



$$\left[\; Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, \dots, n \;\right]$$

- $\beta_0 + \underline{\beta_1} X_i \equiv$ és la recta, la part determinista

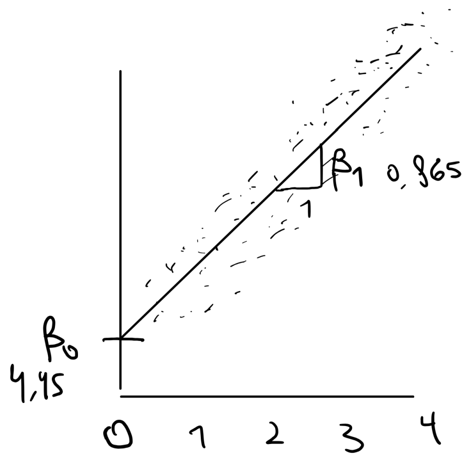
$\underline{\beta_1} \downarrow$ pendent

- $+ \varepsilon_i \equiv$ és la part aleatòria

Premisses per buscar β_0 i β_1

- $\varepsilon_i \sim N(0, \sigma)$
- cal homocedasticitat
- cal independència

$$E(Y_i | X_i) = \beta_0 + \beta_1 X_i$$



Interpretació de les paràmetres

$$\begin{cases} \beta_0 = E(Y_i | X_i = 0) \\ \\ \beta_1 = E(Y_i | X_i = X+1) - E(Y_i | X = X) \end{cases}$$ (pendent, és la pendent)

$$\left[\cdot\hat{\beta}_1 = r_{xy}\cdot\frac{S_y}{S_x} = \frac{S_{xy}}{S_x^2} \qquad \cdot\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}\right]$$

correlación entre X i Y

covariancia entre X i Y

variància de X

· Ej: Bundesliga

Y: Puntos

X: Goles a favor

$E(Y) = 46.67$

$E(X) = 48,8$

$\sigma(Y) = 13,2$

$\sigma(X) = 12,97$

(with) Correlació$(X,Y) = 0,8496$

$$\beta_1 = 0.8496 \cdot \frac{13,2}{12,97} = 0.865$$ per cada gol cuantos puntos aumentan

$$\beta_0 = 46,67 - 0.865 \cdot 48,8 = 4,458$$

si no marquen cap gol, el valor esperat és de 4,458

$$IC(\beta_1, 1-\alpha) = b_1 \pm t_{n-2,\, 1-\frac{\alpha}{2}} \cdot se_{\hat{}}$$

R:

σ_{b_1} = Std error

$\times$

$\boxed{X_2}$

$1-\alpha = 0{,}95$

n = nivells de libertad + 2 = 378

qt(0,975, 376) = 1,966

$\shortparallel$

qnorm(0.975) = 1,96

si n molt gran, es pot usar la Normal

$IC = [0{,}87,\ 0{,}92]$

R

Intercept = β_0

el de abajo = β_1 (pendent)

Donem l'expressió

Point₂ ~ GF

$$Y = \underline{\beta_0 + \beta_1 X} + \underline{\varepsilon}$$

$\uparrow$ Point₂

$\downarrow$ Part determinista

$\uparrow$ GF

$\searrow$ part aleatòria

$\varepsilon \sim N(0, \sigma)$

Std error

S_{β_0}

$\boxed{S_{\beta_1}}$

confint(lmod) calcula IC!!!

Regresión lineal múltiple (RLM)

$$\boxed{\begin{array}{l} Y = \beta_0 + \beta_1 \underset{\text{Sexo}}{X} + \beta_2 \underset{\text{Edad}}{Z} + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma) \\[2em] X = \begin{cases} 1 & \text{Male} \\ 0 & \text{Female} \end{cases} \end{array}}$$

$\beta_0 =$ Intercept (fila 0) $E[Y \mid X=0, Z=0]$

$\beta_1 = E(Y \mid X=1, Z) - E(Y \mid X=0, Z)$ (dif entre hom i dona) i mismo edad

$$\beta_2 = \underbrace{E(Y \mid X, Z = Z+1)}_{\beta_0 + \beta_1 X + \beta_2 (Z+1)} - \underbrace{E(Y \mid X, Z = Z)}_{\beta_0 + \beta_1 X + \beta_2 X}$$

por cada año cuanto más se gana por los trab

- p-value. dir que edat o sexe no influeix, si p-value $\simeq 0$, si que afecta!!
- Residual standard error $= \hat{\sigma} = S$ $\quad \boxed{\varepsilon \sim (N(0, \sigma))}$
- p-valor final, hay una afecta a v

Valor R-squared:

Exemple:

$\hat{Y} \mid$ Mujer, 50 años

$$\hat{Y}_1 = 56740.86 + 948.53 \cdot \underset{50}{X} + 9279.32 \cdot \underset{0}{Z}$$

$\hat{Y} \mid$ Hombre, 40 años

$$\hat{Y}_2 = 56740.86 + 948.53 \cdot \underset{40}{X} + 9279.32 \cdot \underset{1}{Z}$$

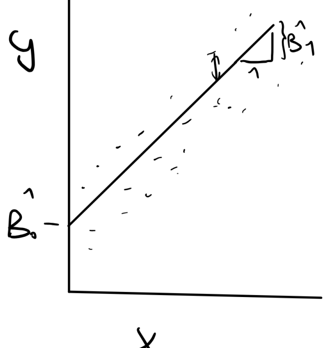
$$\boxed{\begin{aligned} \hat{Y}_1 &= 104167,36 \\ \hat{Y}_2 &= 103961,38 \end{aligned}}$$

- No se usa lo ε porque en este caso es 0 ($N(0, \sigma)$) $\rightarrow \varepsilon$: tratamos con una estimación puntual.

– R-squared:

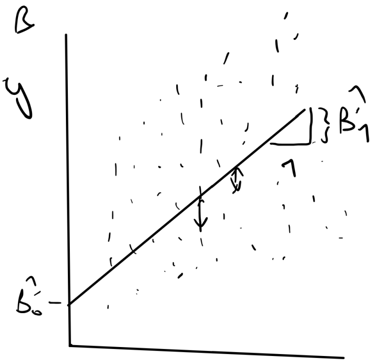
$\hat{B}_0 \simeq \hat{B}_0'$. $\hat{B}_1 \simeq \hat{B}_1'$

A B



"X predice Y bastante bien"

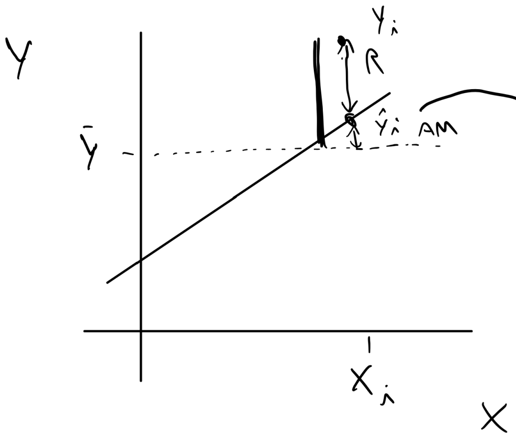
B



AM

"Habrá más predictores de Y"

$\sum (\hat{Y}_i - Y_i)^2$ serà petita $\sum (\hat{Y}_i - Y_i)^2$ serà més gran



Este explica el model

B

Variabilidad total de Y

$$SQ_T = \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2$$

AM

Variabilidad de Y explicada por X (el modelo)

$$SQ_E = \sum (\hat{Y}_i - \bar{Y})^2$$

R

Variable residual

$$SQ_R = \sum (Y_i - \hat{Y}_i)^2$$

$$SQ_T = SQ_E + SQ_R$$

$R^2 \equiv$ coeficiente de determinación

$$R^2 = \frac{SQ_E}{SQ_T}$$

Indica como de bien la LM se ajusta a los datos!!!

$R^2 \gg$ mejor, más se ajustan!!

$0 \le R^2 \le 1$

$R^2 \ll$ peor, menos se ajustan!!

En caso de RLS, $R^2 = r^2$

↳ correlación!!

$R^2 = r^2$

$0{,}2849 = r^2$

$r = \pm\sqrt{0{,}2849} = +0{,}534$

porqué la recta tiene pendiente — !!

$R^2 = 0{,}3787$

Frase: el model explica un 37,87% de la variabilidad de Y

Model per estimar μ

- $Y_i = \mu + \varepsilon_i \; , \quad \varepsilon_i \sim N(0, \sigma)$

$i = 1, \dots, n$

- $\hat{\mu} = \bar{Y}$

$$\underset{\substack{\text{resid}\\ \text{est}\\ \text{error}}}{\hat{\sigma}} = \underset{\substack{\text{est}\\ \text{error}}}{0,10607} \cdot \sqrt{n}$$

$$Y \sim \beta_0 + \beta_1 \cdot \log(x) + \varepsilon$$

$$\downarrow$$

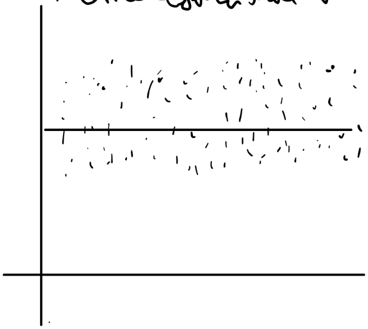
Que significa augmentar $\log(x)$ en 1? ($\log(x)+1 =$

$$\log(x) + 1 = \log(x) + \log(e^1) = \boxed{\log(e \cdot x)}$$

al augmentar x i mult per e esperem un augment de β_1

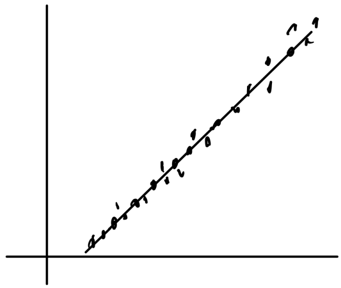
Linealitat ✓

Homocedasticitat ✓



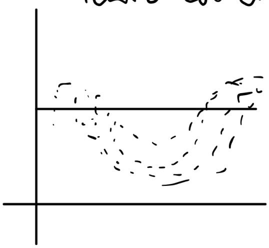
Q-Q plot

Normal ✓

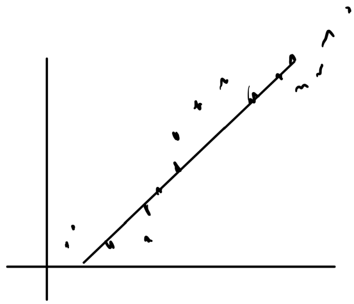


Linealitat ✗

Homocedasci ✓

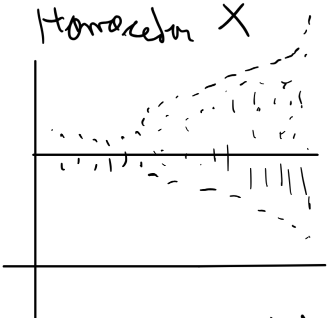


Normal ✗



Linealitat ✓

Homocedar ✗



Linealitat = una recta

Homocedar = sempre mateixa variabilitat