2023-12-13

Παρένθεση

Έλεγχοι Υποθέσεων για τη μέση τιμή κανονικού πληθυσμού

$Y \sim N(\mu, \sigma^2)$, μ : άγνωστο ← Εκτίμηση
σ^2 = γνωστό   Έλεγχος υποθέσεων

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ αμφίπλευρος

Ερώτηση: Υπάρχει σημαντική ένδειξη ότι $\mu \neq \mu_0$?

Αν απορρ. H_0 : ναι υπάρχει

Αν δεχτούμε H_0 (δεν μπορούμε να απορρ.) : όχι δεν υπάρχει σημ. ένδειξη ότι $\mu \neq \mu_0$.

Στη Στατιστική έχουμε δει τον έλεγχο με λόγο πιθανοφάνειας (probability ratio test)

$\alpha = P(\text{σφάλμα τύπου I}) = P(\text{απορ. } H_0 \mid H_0) =$ λάθος απόρριψη H_0 (false alarm)

$\beta = P(\text{σφάλμα τύπου II}) = P(\text{δεχ } H_0 \mid H_1) =$ μη ανίχνευση

Για δεδομένο α το test ελαχιστοποιεί β.

α = επίπεδο σημαντικότητας (significance level)

Τυπικές τιμές του α : 1%, 5%, 10%

Θέτουμε $\alpha = P(\text{απορ. } H_0 \mid H_0)$ ←

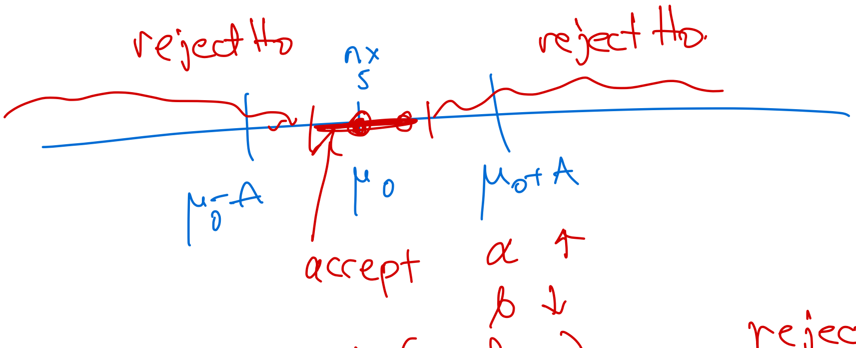
$$\boxed{A = ?}$$

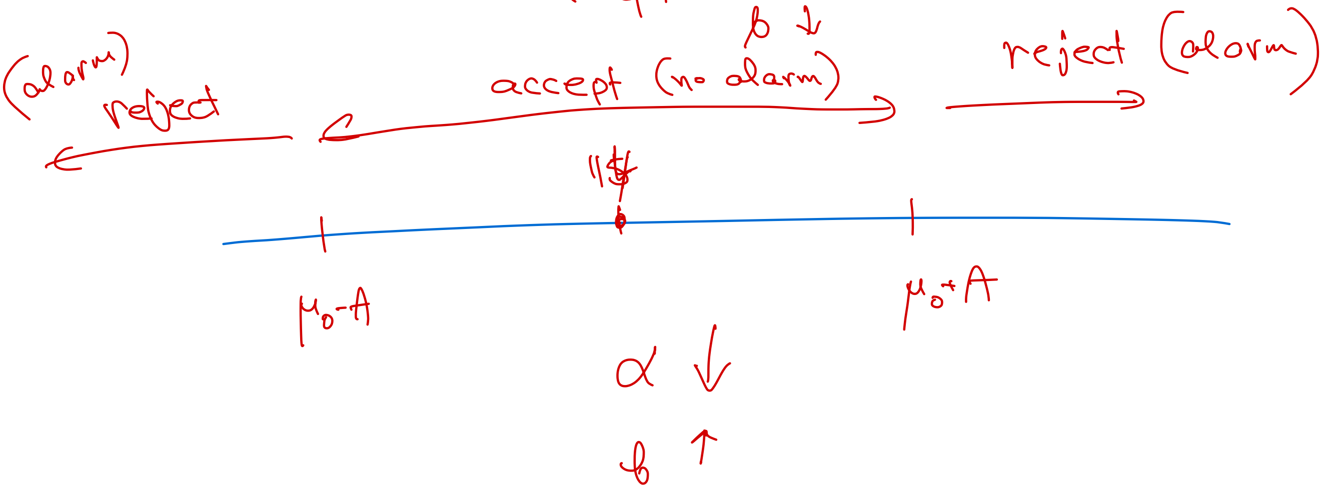
Ο έλεγχος είναι της μορφής

$Y = (y_1, \dots, y_n)$

$\bar{Y}_n = \dfrac{y_1 + \cdots + y_n}{n}$

$|\bar{Y}_n - \mu_0|$ → $\leq A \Rightarrow$ accept ; $> A \Rightarrow$ reject





Ζητάμε εκείνο το A : $P(\text{απορ. } H_0 \mid H_0) = \alpha \iff$

$$\iff \boxed{P(\text{accept } H_0 \mid H_0) = 1 - \alpha}$$

$$\iff P\left(|\bar{Y} - \mu_0| \leq A \;\middle|\; \mu = \mu_0\right) = 1 - \alpha$$

Όμως όταν $Y \sim N(\mu_0, \sigma^2)$ $\quad (\mu = \mu_0)$

$$\Rightarrow \bar{Y}_n \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \Rightarrow \bar{Y}-\mu_0 \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$|\bar{Y}-\mu_0| \le A \Leftrightarrow -A \le \bar{Y}-\mu_0 \le A$$

$$P(-A \le \bar{Y}-\mu_0 \le A) = 1-\alpha$$

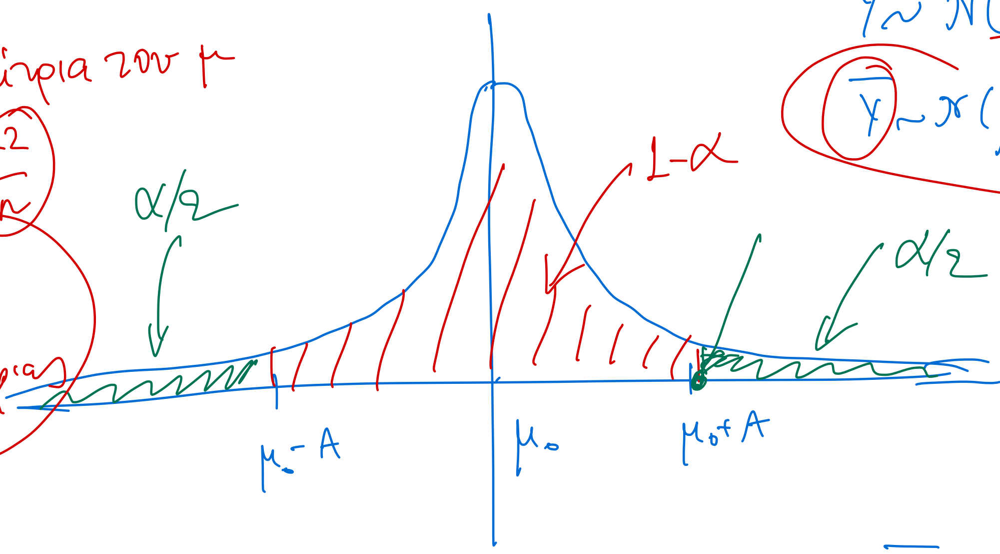
$\bar{Y}$: εκτιμήτρια του μ

$\sigma^2(\bar{Y}) = \frac{\sigma^2}{n}$

τυπικό σφάλμα εκτιμήτριας

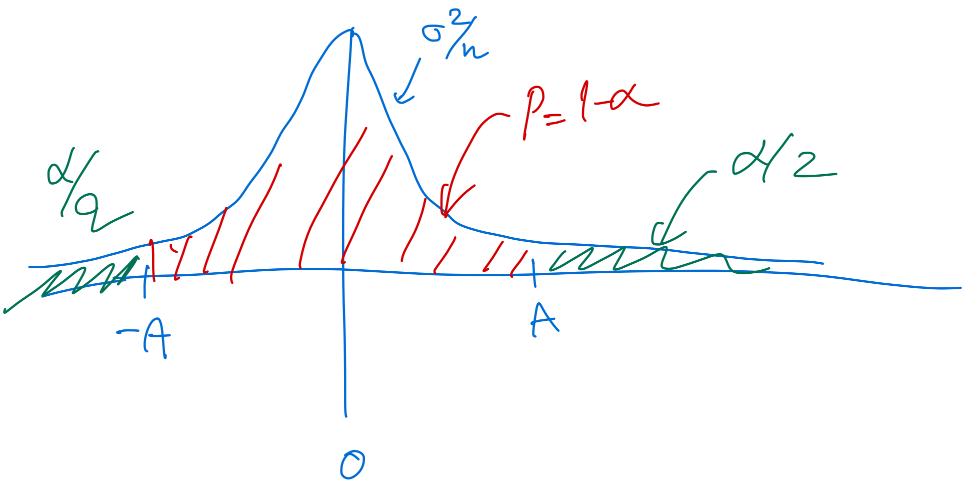
$Y \sim N(\mu_0, \sigma^2)$

$\bar{Y} \sim N(\mu_0, \sigma^2/n)$



$$\bar{Y}-\mu_0 \sim N\left(0, \frac{\sigma^2}{\sqrt{n}}\right)$$

$$\frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}} = Z \sim N(0,1)$$



$$A: \quad P(\bar{Y}-\mu_0 > A) = \alpha/2 \Rightarrow P(\bar{Y}-\mu_0 \le A) = 1-\frac{\alpha}{2}$$

$$\Rightarrow P\left(\frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}} \le \frac{A}{\sigma/\sqrt{n}}\right) = 1-\frac{\alpha}{2} \Rightarrow$$

$$\Rightarrow P\left(Z \le \frac{A}{\sigma/\sqrt{n}}\right) = 1-\frac{\alpha}{2} \qquad \underline{Z \sim N(0,1)}$$

$$\Rightarrow \Phi\left(\frac{A}{\sigma/\sqrt{n}}\right) = 1-\frac{\alpha}{2} \Rightarrow$$

$$\Rightarrow \frac{A}{\sigma/\sqrt{n}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right) = z_{1-\alpha/2} \Rightarrow$$

$$\Rightarrow A = \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}$$

επ. σημ. α

accept αν $|\bar{Y}-\mu_0| \le \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}$

reject H_0 $|\bar{Y}-\mu_0| > \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}$

στατιστικό ελέγχου

κρίσιμη τιμή

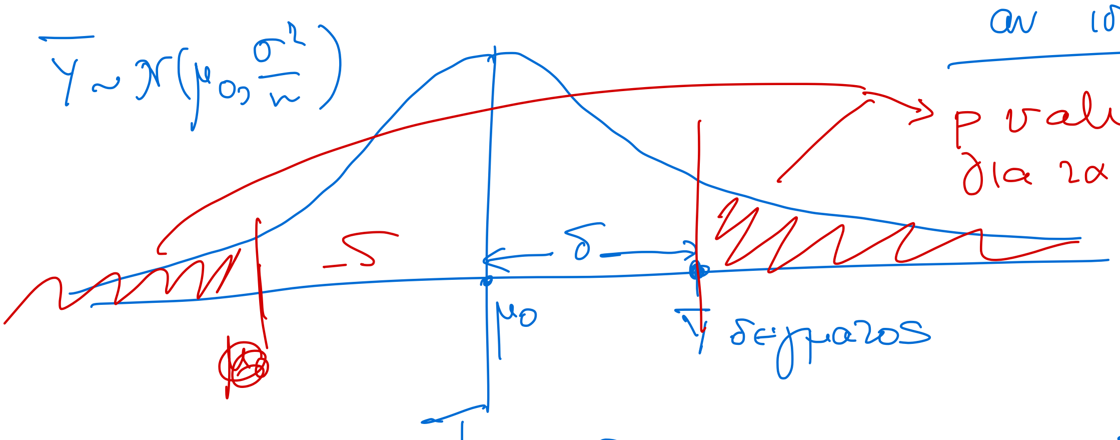
ισοδύναμα $\left|\frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}}\right| \le z_{1-\alpha/2}$ accept

$\frac{\bar{Y}-\mu_0}{s(Y)}$

$> z_{1-\alpha/2}$ reject

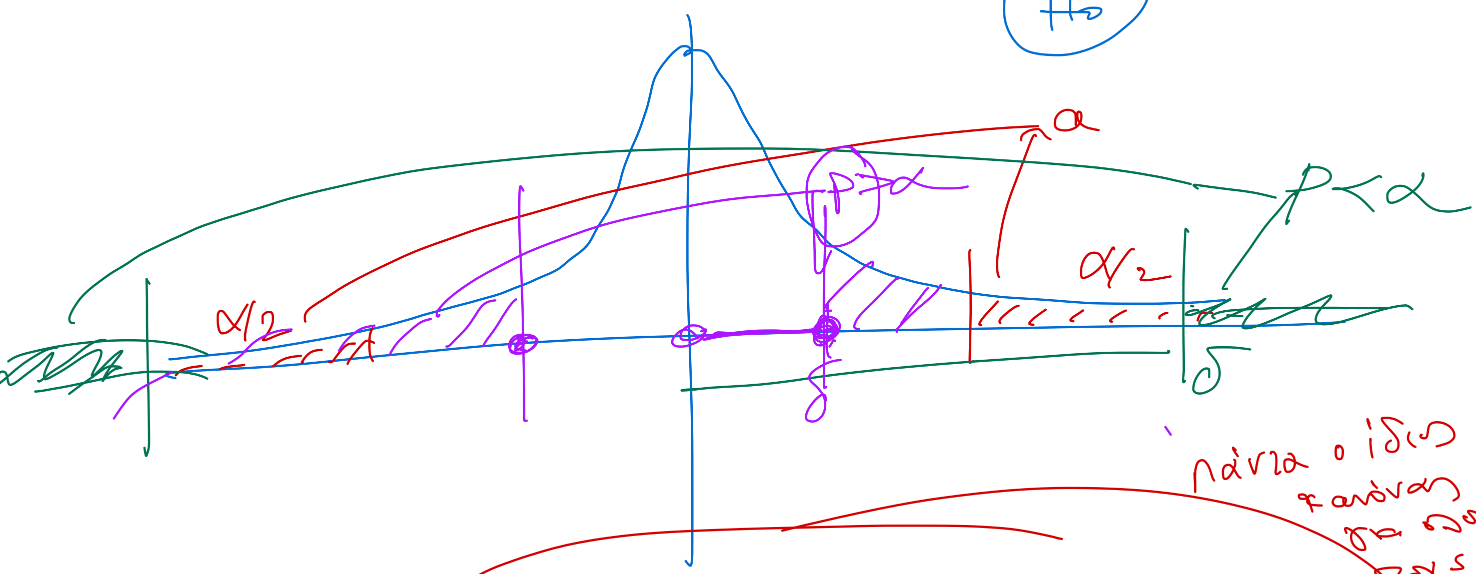
π.χ. $H_0: \mu \le \mu_0 \qquad H_1: \mu > \mu_0$ (μονόπλευρο δεξιά) p-value : P ← συνάρτηση των δεδομένων

Έστω ότι στο δείγμα μας το $|\bar{Y}-\mu_0|=\delta$ αν ισχύει η Ho



τότε θα μπορούσα να ρωτήσω:

"Αν ίσχυε η Ho τότε ποια θα ήταν η πιθανότητα η μέση τιμή του δείγματος να απέχει δ ή περισσότερο από το μο?" (μπορώ να το υπολογίσω)



πάντα ο ίδιος κανόνας για όλους τους ελέγχους

Κανόνας
accept Ho αν $p \le \alpha$
reject Ho αν $p > \alpha$

Επιστροφή στο Μονοπαραγοντικό Μοντέλο

$$Y = b_0 + b_1 X + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2)$$

$\hat{\beta}_0$, $\hat{\beta}_1$: Εκτιμήτριες LSE

$$SST = SSR + SSE$$

- SST: συνολική μεταβλ. της Y στο δείγμα
- SSR: μεταβλ. που "εξηγείται" από τη X τώρα συσχέτισης με την Y
- SSE: παραμένει ανεξήγητη από τη X (μπορεί να οφείλεται σε άλλους παράγοντες)

$$R^2 = \frac{SSR}{SST}$$ = % μεταβ. της Y που εξηγείται από το μοντέλο

Output (Stata) (κ' του R)

Πίνακας Ανάλυσης Διασποράς

	Sum of Squares	degrees of freedom	Mean Square
Μοντέλο	SSR	df_{mod} 1	MSR/1
Error	SSE	df_{er} n−2	$MSE = \frac{SSE}{n-2}$
Σύνολο	SST	n−1	

$$\boxed{R^2 = \frac{SSR}{SST}}$$

(F-test : θα το δούμε)

$MSE = \hat{\sigma}^2$ αμερόληπτη εκτιμήτρια

$E(MSE) = \sigma^2$

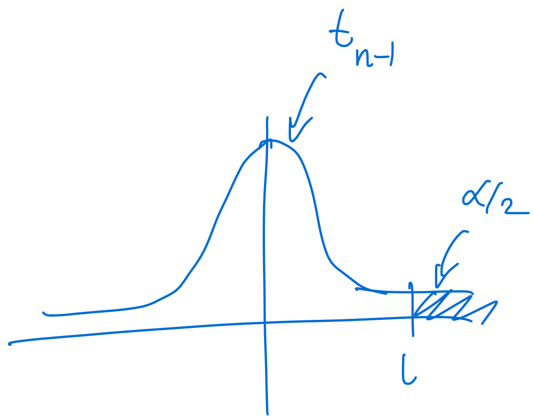
Εκτιμητική

Παράμετρος	Εκτίμηση	Τυπικό σφάλμα $s(\hat{b})$	t-statistic	p-value
Constant	$\hat{b}_0$	$s(\hat{b}_0)$	$\frac{\hat{b}_0}{s(\hat{b}_0)}$	p : $H_0: b_0 = 0$, $H_1: b_0 \neq 0$
X	$\hat{b}_1$	$s(\hat{b}_1)$	$\hat{b}_1 / s(\hat{b}_1)$	p : $H_0: b_1 = 0$, $H_1: b_1 \neq 0$

$$H_0: b_1 = 0 \quad vs \quad H_1: b_1 \neq 0$$

στατ. ελέγχου $\left| \frac{\hat{b}_1 - 0}{s(\hat{b}_1)} \right|$

Κρίσιμη τιμή $t_{\alpha/2, n-1}$

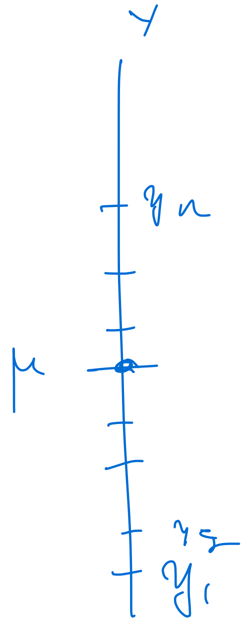


Παρένθεση: βαθμοί ελευθερίας

Έστω $Y \sim N(\mu, \sigma^2)$ μ: γνωστό

σ^2: άγνωστο

$$\sigma^2 = E\left((Y-\mu)^2\right)$$



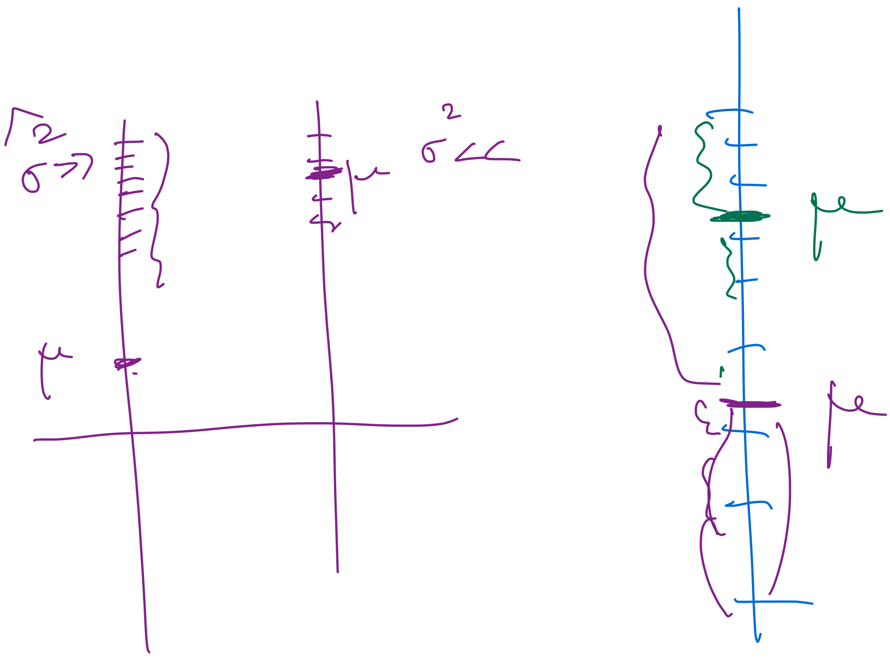
$$(y_1-\mu)^2, (y_2-\mu)^2, \dots, (y_n-\mu)^2$$

$$S^2 = \frac{\sum_{j=1}^{n}(y_j-\mu)^2}{n} = \hat{\sigma}^2$$

$$E(S^2) = \sigma^2 \quad \text{αμερόληπτη}$$

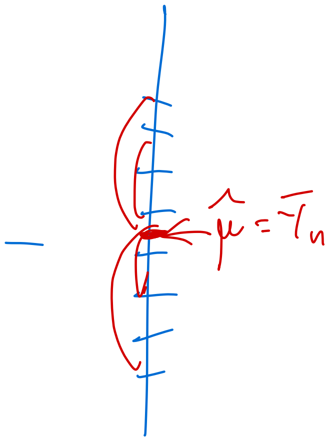
Τι αλλάζει όταν $Y \sim \mathcal{N}(\mu, \sigma^2)$ $\quad$ μ: άγνωστο, σ^2: άγνωστο

στην εκτίμηση του σ^2?



$$\hat{\mu} = \overline{Y}_n = \frac{y_1 + \dots + y_n}{n}$$

$$S^2 = \frac{(y_1-\hat{\mu})^2 + \dots + (y_n-\hat{\mu})^2}{n}$$



$$s^2 = \frac{\sum_{j=1}^{n} (y_j - \hat{\mu})^2}{n}$$ <u>όχι αμερόληπτη</u>

Από αποδείξεις $E(s^2) = \frac{\sigma^2 \cdot (n-1)}{n} \neq \sigma^2$

$(y_1 - \hat{\mu})^2, (y_2 - \hat{\mu})^2$ δεν είναι ανεξάρτητα

$$\parallel$$

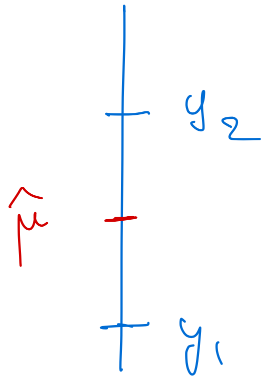
$$\left(y_1 - \frac{y_1 + \dots y_n}{n}\right)^2$$

$$\tilde{S}^2 = \frac{\sum_{j=1}^{n} (y_j - \bar{y})^2}{n-1} = S^2 \cdot \frac{n}{n-1}$$

$n-1$ → βαθμοί ελευθερίας

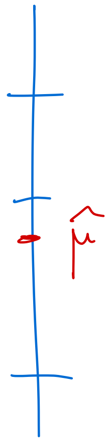
$$E(\tilde{S}^2) = E(S^2) \cdot \frac{n}{n-1} = \sigma^2 \cdot \cancel{\frac{n-1}{n} \cdot \frac{n}{n-1}} = \sigma^2$$

$$\sum (y_j - \hat{\mu})^2 \sim \chi^2_{n-1}$$



$$(y_1-\hat{\mu})^2 = (y_2-\hat{\mu})^2$$

$$\frac{(y_1-\hat{\mu})^2}{1} \qquad 1 = 2-1$$



$$(y_1-\bar{y}) + (y_2-\bar{y}) + (y_3-\bar{y}) = 0$$

$$y_1+y_2+y_3 - 3\bar{y} = 0$$

$$s^2 = \frac{(y_1-\bar{y})^2 + (y_2-\bar{y})^2 - (y_3-\bar{y})^2}{2}$$

Πίσω στο $\quad Y = b_0 + b_1 X + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$

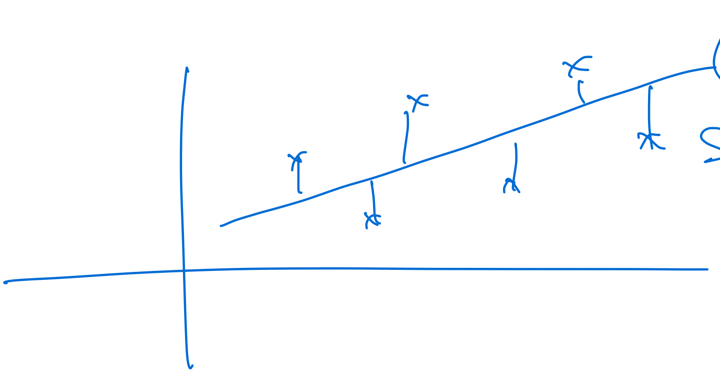
άγνωστο

$$Y | X = x \sim N(b_0 + b_1 x, \sigma^2)$$

μαζί $b_0 + b_1 X$

$$\sigma^2 \approx SSE \qquad \frac{SSE}{n}$$

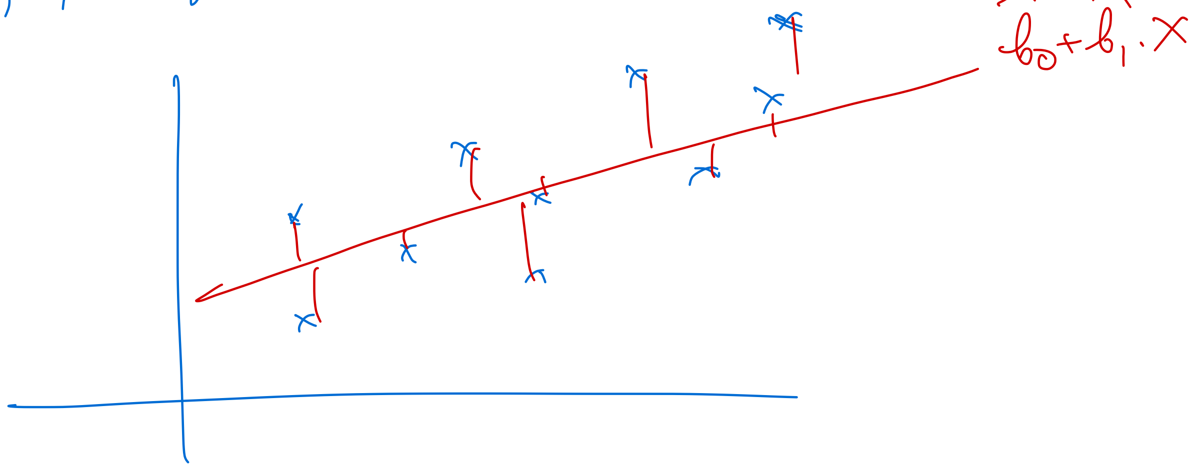
γνωστή



$$SSE \sim \chi^2_n$$

$$SSE = \sum_j (y_j - b_0 - b_1 x_j)^2$$

Όταν b_0, b_1 : άγνωστα



$$SSE = \sum \left(y_i - \hat{b}_0 - \hat{b}_1 \cdot x_i\right)^2$$

1 2

2 εκτ. παράμετροι

$$SSE \sim \chi^2_{n-2}$$

$$\hat{\sigma}^2 = \frac{SSE}{n-2} \qquad E(\hat{\sigma}^2) = \sigma^2$$

π.χ. $n=10$

$$Y = \underbrace{b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_7 X_7}_{8 \text{ άγνωστες}}$$

$$df_{error} = 10 - 8 = 2$$

$$\sigma^2 = \frac{SSE}{2}$$

Πρόβλεψη για νέες τιμές του X

Δείγμα

x	y
x_1	y_1
$\vdots$	$\vdots$
x_n	y_n

x: χρόνος άσκησης, y: score

$\longrightarrow$ $\hat{b}_0$, $\hat{b}_1$. — — —

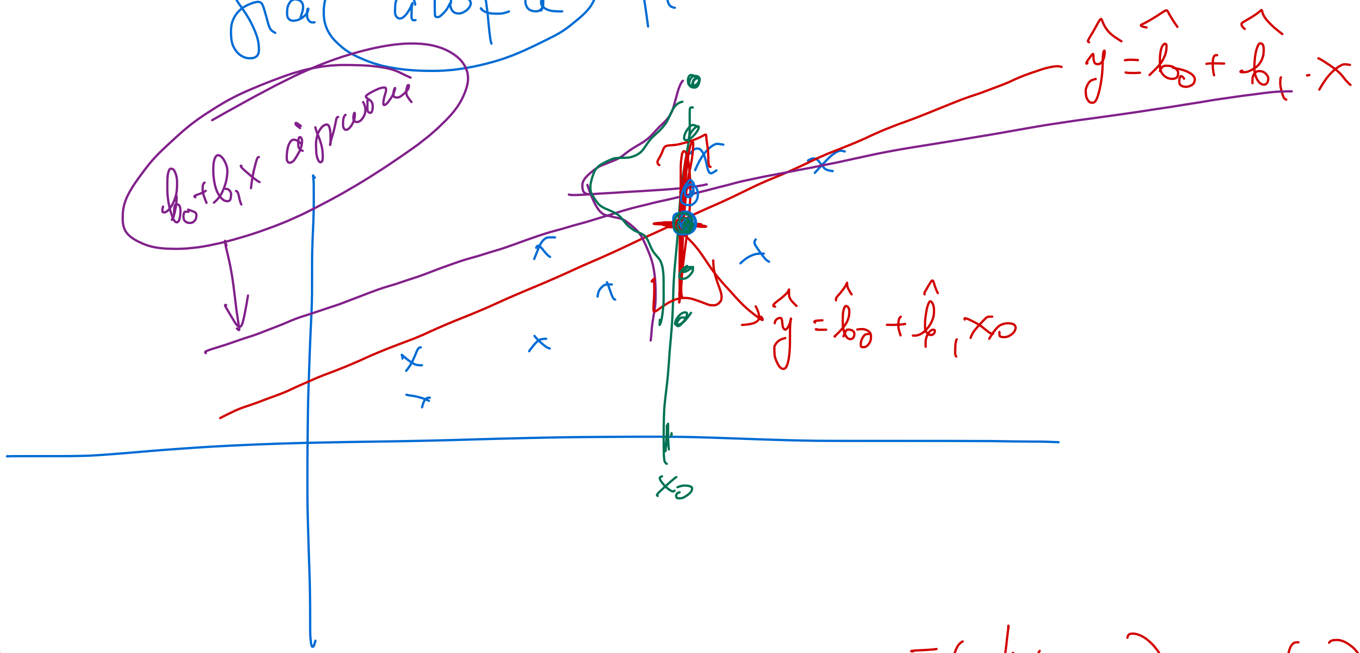
συμπερασματολογία

Έρχεται ένα νέο άτομο με χρόνο άσκησης x_0

① Τι μπορώ να πω για το y αυτού του ατόμου;

πρόβλεψη

② Τι μπορώ να πω για τη μέση τιμή του Y για άτομα με $x = x_0$



② Διάστημα εμπιστοσύνης για $E(Y \mid X = x_0) = \mu(x_0)$

$$\hat{\mu}(x_0) = \hat{y}(x_0) = \hat{b}_0 + \hat{b}_1 \cdot x_0$$

ΔΕ $\hat{y}(x_0) \pm \underbrace{S(\hat{y}(x_0))}_{\text{stata}\ \text{stdp}} \cdot t_{\alpha/2,\,n-1}$

$x = 614$ $1975 \pm 586 \cdot t_{\alpha/2,\,n-1}$ πίνακας της t

1 Διάστημα Πρόβλεψης

Αν περιμένω να εμφανίζεται το Y ενός μόνο ατόμου που έχει $x = x_0$?

$$\hat{y}(x_0) \pm (\text{stdf}) \times t_{\alpha/2,\,n-1}$$

$\hat{y}(x_0) = \hat{b}_0 + \hat{b} \cdot x_0$

stdf: std error forecast