Δευτέρα 2/11/20 Μάθημα 5ο.

Το F-τεστ για την πολυδρόμηση ή ~~έλεγχο~~ F-τεστ για έλεγχο $H_0: \beta_1=0$ έναντι $H_a: \beta_1 \neq 0$

Αποδείξαμε σε προηγούμενα μαθήματα ότι $E(MSres)=\sigma^2$.

Θ.δ.ο $E(MSreg)=\sigma^2+\beta_1^2\sum(x_i-\bar{x})^2$

Αποδ

$$MSreg = \frac{SSreg}{1} = SSreg = \sum(\hat{y}_i-\bar{y}) = \hat{\beta}_1\sum(x_i-\bar{x})^2$$

$$E(MSreg) = E(\hat{\beta}_1^2\sum(x_i-\bar{x})^2) = \sum(x_i-\bar{x})^2 E(\hat{\beta}_1^2) =$$

$$= \sum(x_i-\bar{x})^2[Var(\hat{\beta}_1)+[E(\hat{\beta}_1)]^2] = \sum(x_i-\bar{x})^2\left[\frac{\sigma^2}{\sum(x_i-\bar{x})^2}+\beta_1^2\right]$$

$$= \sigma^2+\beta_1^2\sum(x_i-\bar{x})^2$$

Αν η $H_0: \beta_1=0$ είναι αληθής τότε $E(MSreg)=E(MSres)$.

» » » τότε $MSreg \approx MSres$

Ισχύει Αν πρόταση A ⇒ Πρόταση B τότε $\sim B \Rightarrow \sim A$ (άρνηση της A) } ⇒

Αν $MSreg \not\approx MSres$ τότε $H_0: \beta_1=0$ πρέπει να απορριφθεί.

διαφορετικό

Άρα ένα κριτήριο για την απόρριψη της $H_0: \beta_1=0$ πρέπει να στηριχτεί στη σύγκριση του MSreg με το MSres [Για να συγκρίνω ~~θα~~ θα πάρω ή τη διαφορά ή το πηλίκο]

και ειδικότερα η κ.π να στηριχτεί στο πηλίκο $\frac{MSreg}{MSres}$.

Αποδείξαμε ότι $\frac{SSres}{\sigma^2} \sim \chi^2_{n-2}$

Θ.δ.ο υπό τις υποθέσεις για τα σφάλματα τότε $\frac{SSreg}{\sigma^2} \sim \chi_1^2$.

Αποδ $SSreg = \hat{\beta}_1^2\sum(x_i-\bar{x})^2$.

Ξέρω $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum(x_i-\bar{x})^2}) \xrightarrow[H_0: \beta_1=0]{\text{υπό την}} \hat{\beta}_1 \sim N(0, \frac{\sigma^2}{\sum(x_i-\bar{x})^2}$

(2)

$$\Rightarrow \frac{\hat{\beta}_1}{\sigma/\sqrt{\Sigma(x_i-\bar{x})^2}} \sim N(0,1) \Rightarrow \frac{\hat{\beta}_1^2 \Sigma(x_i-\bar{x})^2}{\sigma^2} \sim N^2(0,1) \equiv \chi_1^2.$$

Θεωρώ το πηλίκο $F = \frac{MSreg}{MSres} = \frac{SSreg/1}{SSres/(n-2)}$.

$$\Rightarrow F = \frac{\frac{SSreg}{\sigma^2}/1}{SSres/\sigma^2(n-2)} \sim \frac{\chi_1^2/1}{\chi_{n-2}^2/n-2} \equiv F_{1,n-2} \text{ (αν } \chi_1^2, \chi_{n-2}^2 \text{ είναι ανεξάρτητες)}$$

επειδή ~~Αν~~ SSreg είναι αξεξάρτητο SSres.

Άρα για τον έλεγχο της υπόθεσης $H_0: \beta_1=0$ η ΣΣΤ είναι $F = \frac{MSreg}{MSres} \sim F_{1,n-2}$ υπό $H_0: \beta_1=0$

Μορφή της κ.π

Η H_0 απορ. για μεγάλες τιμές του F, δηλαδή $F \geq c$

Προσδιορισμός κρίσιμου σημείου

$$\alpha = P(\text{Απορ} H_0 \mid H_0 \text{ αληθής}) = P(F \geq c \mid F \sim F_{1,n-2}) = P(F_{1,n-2} \geq c)$$

$\Rightarrow \boxed{c = F_{1,n-2,\alpha}}$ από τον πίν. των εκατοστιαίων σημείων.

και κρίσιμη περιοχή μεγέθους α. την $F \geq F_{1,n-2,\alpha}$

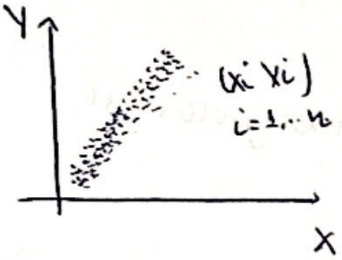
Παρατήρηση:

Το F-τεστ είναι ισοδύναμο με το t-τεστ. για τον έλεγχο της $H_0: \beta_1=0$.

Πράγματι, $F = \frac{MSreg}{MSres} = \frac{\hat{\beta}_1^2 \Sigma(x_i-\bar{x})^2}{MSres} = \left[\frac{\hat{\beta}_1 [\Sigma(x_i-\bar{x})^2]^{1/2}}{\sqrt{MSres}}\right]^2 = \frac{\hat{\beta}_1^2}{\hat{Var}(\hat{\beta}_1)} = t^2$

Συντελεστής Συσχέτισης Pearson:

X	x_1 .... x_n
Y	y_1 .... y_n

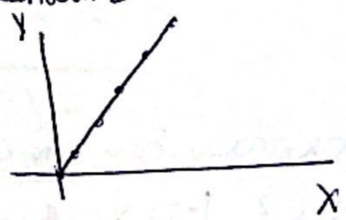


← Εικόνα ή φωτογραφία του συντελεστή συσχέτισης Pearson $r(x,y)$

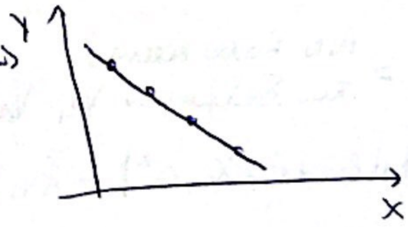
Όπου $$r_{x,y} = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum (x_i-\bar{x})^2 \sum (y_i-\bar{y})^2}}$$

Ιδιότητες του $r_{x,y}$: (i) καθαρός αριθμός (χωρίς μονάδες μέτρησης)
(ii) $-1 \le r_{x,y} \le 1$ (Ανισότητα Cauchy-Schwarz)
(iii) συμμετρικός, δηλ. $r_{x,y} = r_{y,x}$

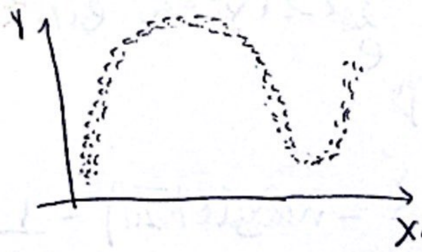
• Αν $r_{x,y}=1$ τότε τέλεια γραμμική σχέση μεταξύ X, Y, θετικής κατεύθυνσης



• Αν $r_{x,y}=-1$ τότε τέλεια γραμμική σχέση μεταξύ X, Y αρνητικής κατεύθυνσης



• Αν $r_{x,y}=0$ τότε ∄ γραμμική σχέση μεταξύ X, Y., μπορεί να υπάρχει άλλου τύπου σχέση



Π.χ)

X	-3	-1	0	1	3
Y	9,54	995	10	9,95	994

Για αυτά $r_{x,y}=0$. ∄ υπάρχει γραμμική σχέση όμως τα x, y βρίσκονται σε κύκλο ακτίνας 10 γιατί ικανοποιούν τη σχέση $x^2+y^2=10$

Αν $|r(x,y)| \geq 0{,}6$ ή $0{,}7$ τότε υπάρχει ισχυρή γραμμική σχέση μεταξύ X, Y

Άσκηση:

Έστω το μοντέλο της α.γ.π $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \ i = 1 \dots n$

με $\varepsilon_i \sim N(0, \sigma^2), \ i = 1 \dots n$ ασυσχέτιστα.

Ν.δ.ο οι εκτιμητές μέγιστης πιθανοφάνειας των β_0, β_1 ταυτίζονται με τους Ε.Ε.Τ.

Έστω τ.δ $W_1, \dots W_n$ από $f(\cdot, \theta)$. Οι ΕΜΠ ορίζονται – Πιθανοφάνειας

$L = L(\theta) =$ Από κοινού κατανομή των $W_1, \dots W_n = f_{W_1, \dots W_n}(w_1, \dots w_n, \theta) =$

$= \prod_{i=1}^{n} f_{W_i}(w_i, \theta)$ ο ΕΜΠ της θ συμβολίζεται με $\hat{\theta}$ και είναι το σημείο που μεγιστοποιεί την πιθανοφάνεια L, $L(\hat{\theta}) = \max_{\theta} L(\theta)$ ή μεγιστοποιεί το $\log L(\theta)$

$\hat{\theta} = \arg\max_{\theta} L(\theta)$

Οι ΕΕΤ προκύπτουν από την ελαχιστοποίηση ως προς β_0, β_1

της $S = \sum \varepsilon_i^2 = \boxed{\sum (Y_i - \beta_0 - \beta_1 X_i)^2}$

$L = L(\beta_0, \beta_1) =$ Από κοινού κατανομή των δεδομένων $Y_1, \dots Y_n = \prod_{i=1}^{n} f_{Y_i}(y_i, \beta_0, \beta_1)$

Αλλά $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \Rightarrow f_{Y_i}(Y_i, \beta_0, \beta_1) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(Y_i - \beta_0 - \beta_1 X_i)^2}$

$$\Rightarrow L = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum (y_i - \beta_0 - \beta_1 x_i)^2}$$

$$\log(L) = -n\log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum (y_i - \beta_0 - \beta_1 x_i)^2$$

Οι Ε.Μ.Π των β_0, β_1 θα προκύψουν από μεγιστοποίηση της L

» » » » του $-\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

$\Leftrightarrow$ $\boxed{\text{ελαχιστοποίηση του } \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2}$

Άρα οι ΕΕΤ, ΕΜΠ προκύπτουν από ελαχιστοποίηση ~~ως~~ προς b_0 και b_1 της ίδιας ποσότητας $\sum_{i=1}^{n}(Y_i-b_0-b_1X_i)^2$. Άρα ταυτίζονται

β) Για τον ~~ΕΕΤ~~ ΕΜΠ της σ^2 μεγιστοποιώ ως προς σ^2 την πιθανοφάνεια

$$L = L(\hat{b}_0,\hat{b}_1) = \frac{1}{(\sigma^2 2\pi)^{n/2}}\, e^{-\frac{1}{2\sigma^2}\sum(Y_i-\hat{b}_0-\hat{b}_1X_i)^2}$$

Άσκηση

Ασκ.2

Έστω μοντέλο α.γ.π. $Y_i=b_0+b_1X_i+\varepsilon_i$, $i=1,\dots n$. με $\varepsilon_i \sim N(0,\sigma^2)$, $i=1,\dots n$ ασυσχέτιστα

Ν.δ.ο

$$\hat{b}_0 \sim N\left(b_0,\ \sigma^2\left\{\frac{1}{n}+\frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right\} = \frac{\sigma^2}{n}+\frac{\sum X_i^2\sigma^2}{\sum(X_i-\bar{X})^2}\right)$$

$$E(\hat{b}_0) = E(\bar{Y}-\hat{b}_1\bar{X}) = E(\bar{Y})-\bar{X}E(\hat{b}_1) = E(\bar{Y})-\bar{X}b_1$$

$$E(\bar{Y}) = E\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n}\sum E(Y_i) = \frac{1}{n}\sum(b_0+b_1X_i) = b_0+b_1\bar{X}$$

$$E(\hat{b}_0) = E(\bar{Y})-\bar{X}b_1 = b_0+b_1\bar{X}-\bar{X}b_1 \Rightarrow \boxed{E(\hat{b}_0)=b_0}$$

$$Var(\hat{b}_0) = Var(\bar{Y}-\hat{b}_1\bar{X}).$$

Αλλά $Var\left(\sum_{i=1} a_iW_i\right) = \sum a_i^2 Var(W_i) + \sum_{i=1}\sum_{\substack{j=1\\ i\neq j}} a_ia_j Cov(W_i,W_j) \Big\} \Rightarrow$

$$\Rightarrow Var(\hat{b}_0) = Var(\bar{Y}) + \bar{X}^2 Var(\hat{b}_1) - \bar{X}\,Cov(\bar{Y},\hat{b}_1).$$

Αποδείξαμε όταν βρήκαμε την $Var(Y_i)$ ότι $Cov(\bar{Y},\hat{b}_1)=0$

Άρα $Var(\hat{b}_0) = Var(\bar{Y})+\bar{X}^2 Var(\hat{b}_1)$

$Var(\bar{Y}) = Var\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n^2}\sum Var(Y_i) = \frac{1}{n^2}\sum\sigma^2 \Rightarrow Var(\bar{Y}) = \frac{\sigma^2}{n}$

Επίσης αποδείξαμε ότι $Var(\hat{b}_1) = \frac{\sigma^2}{\sum(X_i-\bar{X})^2}$ $\Big\} \Rightarrow$

$$Var(\hat{b}_0) = \frac{\sigma^2}{n} + \frac{\bar{X}^2\cdot\sigma^2}{\sum(X_i-\bar{X})^2} = \sigma^2\left\{\frac{1}{n}+\frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right.$$

Απομένει να δείξω οτι $\hat{\beta_0} \sim Normal$.

$\hat{\beta_0} = \bar{Y} - \hat{\beta_1}\bar{X}$ δηλαδή γραμμικός συνδυασμός των $\bar{Y}$ και $\hat{\beta_1}$ που έχουν κανονική και είναι και ασυσχέτιστα Άρα $\boxed{\hat{\beta_0} \sim Normal.}$

Ασκ. 3

Έστω μοντέλο α.γ.π $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i=1,..n$ με $\varepsilon_i \sim N(0,\sigma^2)$ $i=1,...n$ και ασυσχέτιστα. Ν.δ.ο $Cov(\hat{\beta_0},\hat{\beta_1}) = -\sigma^2 \dfrac{\bar{X}}{\Sigma(X_i-\bar{X})^2}$

$$\left.\begin{array}{l} Cov(\hat{\beta_0},\hat{\beta_1}) = Cov(\bar{Y} - \hat{\beta_1}\bar{X}, \hat{\beta_1}) \\ Cov(\Sigma a_i X_i, \Sigma b_i Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j Cov(X_i, Y_j) \end{array}\right\} \Rightarrow$$

$$Cov(\hat{\beta_0},\hat{\beta_1}) = \cancel{Cov(\bar{Y},\hat{\beta_1})}^{0} - \bar{X} Cov(\hat{\beta_1},\hat{\beta_1}) = -\bar{X} Cov(\hat{\beta_1},\hat{\beta_1}) =$$

$$= -\bar{X} Var(\hat{\beta_1}) = -\bar{X}\frac{\sigma^2}{\Sigma(X_i-\bar{X})^2}.$$

Άσκηση 4

Τα παρακάτω δεδομένα δείχνουν την επίδραση της θερμοκρασίας στην απόδοση μιας χημικής αντίδρασης

Θερμοκρασία	-5	-4	-3	-2	-1	0	1	2	3	4	5
Απόδοση	1	5	4	7	10	8	9	13	14	13	8

, $n=11$

α) αφού οριστεί η ανεξάρτητη και η εξαρτημένη μεταβλητή να βρεθούν οι ΕΕΤ.

β) Να κατασκευαστεί ο πίνακας ανάλυσης διασποράς και να γίνει ένα αρχικό τεστ. για τον έλεγχο $H_0: \beta_1 = 0$

γ) Κάνοντας τις κατάλληλες υποθέσεις σε επίπεδο σημαντικότητας $a=0,05$
i) Να ελεγχθεί η $H_0: \beta_1 = 0$
ii) Να βρεθεί το 95% δ.ε για τη β_1

α) Θερμοκρασια. ⟶ Ανεξάρτητη ⇝ X
Απόδοση χημ. αντιδρασης ⇝ Εξαρτημένη ⇝ Y } Μοντέλο $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i=1,..n.$

ΕΕΤ: $\hat{\beta_0} = \bar{Y} - \hat{\beta_1}\bar{X} = 9,27.$

$\beta_1 = \frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2} = 1,44$ → Δείχνει τη μεταβολή της Y σε μοναδιαία μεταβολή της X.

Εκτιμώμενο μοντέλο: $\hat{Y} = 9,27 + 1,44X$

β)

Πηγή μεταβλητοτητας	SS	β.Ε	MS	F- πηλίκο
Μοντέλο	$SS_{reg} = 226,95$	1	$MS_{reg} = 226,95$	$F = \frac{MS_{reg}}{MS_{res}} = 96,17$
Υπόλοιπα	$SS_{res} = 21,23$	9	$MS_{res} = 2,38$	
Ολική	$SS_{tot} = 248,18$	10		

δ) i) κ.σ $F_{1,n-2,a} = F_{1,9,0,05} = 5,12$

$F = 96,17 >> 5,12.$ ⇝ $H_0 = \beta_1 = 0$ απορ.

ii) Το δ.ε. είναι $\hat{\beta_1} \pm t_{n-2,\frac{a}{2}} \sqrt{\hat{Var}(\hat{\beta_1})}$, $\hat{Var}(\hat{\beta_1}) = \frac{MS_{res}}{\Sigma(X_i-\bar{X}^2)}$

Το 95% δ.ε. για τη β1 ειναι $\underbrace{(1,11, 1,77)}$.

Το διαστημα που κινειται η εξαρτημένη Y αν μεταβάλλεις την ανεξάρτητη X κατά μια μονάδα.