15o online μάθημα

14/5/2020

Άσκηση: $Y_i = \beta X_i + \varepsilon_i \quad i=1,\dots,n$

$E(\varepsilon_i)=0$, $Var(\varepsilon_i) = \sigma^2 X_i$

$$\frac{Var(\varepsilon_i)}{X_i} = \sigma^2 \Rightarrow Var\left(\frac{\varepsilon_i}{\sqrt{X_i}}\right) = \sigma^2$$

$$Y_i = \beta X_i + \varepsilon_i \xrightarrow[\sqrt{X_i}]{\text{Διαιρώ με}} \frac{Y_i}{\sqrt{X_i}} = \beta\sqrt{X_i} + \frac{\varepsilon_i}{\sqrt{X_i}}$$

Συμβολίζω $Y_i' = \frac{Y_i}{\sqrt{X_i}}$, $X_i' = \sqrt{X_i}$, $\varepsilon_i' = \frac{\varepsilon_i}{\sqrt{X_i}}$

Τότε $Y_i' = \beta X_i' + \varepsilon_i'$. Θα βρω Ε.Ε.Τ.

Πως ελαχιστοποιώ την $Q = \sum_{i=1}^{n} (Y_i' - \beta X_i')^2 = 0$ ως προς β.

$$\frac{\partial Q}{\partial \beta} = 0 \Rightarrow \frac{\partial}{\partial \beta} \sum_{i=1}^{n} (Y_i' - \beta X_i')^2 = 0$$

$$\Rightarrow -2\sum_{i=1}^{n} \left[(Y_i' - \beta X_i') X_i' \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i' X_i' - \beta \sum_{i=1}^{n} (X_i')^2 = 0$$

Άρα

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i' X_i'}{\sum (X_i')^2}$$

Συνάμαι ότι $Y_i' = \frac{Y_i}{\sqrt{X_i}}$ και $X_i' = \sqrt{X_i}$

Άρα $\hat{\beta} = \dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$ είναι ο Ε.Ε.Τ.

Είναι αμερόληπτος;
Αρκεί ν.δ.ο. $E(\hat{\beta}) = \beta$

$$E(\hat{\beta}) = E\left(\frac{\sum Y_i}{\sum X_i}\right) = \frac{1}{\sum X_i} E(\sum Y_i) =$$

$$= \frac{1}{\sum X_i} \sum E Y_i = \frac{1}{\sum X_i} \sum_{i=1}^{n} \beta X_i = \beta$$

γιατί $E Y_i = E(\beta X_i + \varepsilon_i) = \beta X_i + E(\varepsilon_i) = \beta X_i$

$Y_i = \beta X_i + \varepsilon_i$ δεν είναι στοχαστικό δηλαδή του συμπεριφέρομαι ως σταθερά

$$E\left(\frac{\sum Y_i}{\sum X_i}\right) = \frac{1}{\sum X_i} E(\sum Y_i)$$

Για το μοντέλο: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i=1,\dots,n$

(i) $\sum \hat{Y}_i = \sum Y_i$ και $\sum e_i = 0$

Θ.δ.ο. $\sum \hat{Y}_i = \sum Y_i$ ή ισοδύναμα $\sum (Y_i - \hat{Y}_i) = 0$ το οποίο είναι ισοδύναμο με $\sum e_i = 0$

Εμείς ξέρουμε ότι τα $\hat{\beta}_0$, $\hat{\beta}_1$ προκύπτουν από τις λεγόμενες κανονικές εξισώσεις δηλαδή ικανοποιούν το σύστημα των εξισώσεων.

$$\begin{cases} \dfrac{\partial}{\partial \beta_0} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 = 0 \\ \dfrac{\partial}{\partial \beta_1} \sum (Y_i - \beta_0 - \beta_1 X_1)^2 = 0 \end{cases}$$

ή ισοδύναμα

$$\begin{cases} -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0 & (1) \\ -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) X_i = 0 \end{cases}$$

Επομένως από την (1)

$\sum(Y_i - \hat{b}_0 - \hat{b}_1 X_i) = 0$

(ii) $\bar{\hat{Y}} = \bar{Y}$ δηλαδή $\frac{\sum \hat{Y}_i}{n} = \frac{\sum Y_i}{n}$

Προφανώς αφού $\sum \hat{Y}_i = \sum Y_i$

(iii) $\bar{Y} = \hat{b}_0 + \hat{b}_1 \bar{X}$

~~Ένα~~

1ος τρόπος: $\hat{b}_0 + \hat{b}_1 \bar{X} \xrightarrow[\hat{b}_0,\ \hat{b}_1]{\text{αντικατασταση}}$ πράξεις

$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$

οπότε το δείξαμε

2ος τρόπος: Από την (1) των κανονικών εξισώσεων

$\sum(Y_i - \hat{b}_0 - \hat{b}_1 X_i) = 0$

$\Rightarrow \sum Y_i - n\hat{b}_0 - \hat{b}_1 \sum X_i = 0$

$\Rightarrow \sum Y_i = n\hat{b}_0 + \hat{b}_1 \sum X_i$

$\xRightarrow{:n} \bar{Y} = \hat{b}_0 + \hat{b}_1 \bar{X}$

(iv) Νδο $\sum \hat{Y}_i X_i = \sum Y_i X_i$

$\sum e_i \hat{Y}_i = 0$, $\sum e_i X_i = 0$

$$\begin{cases} \sum(Y_i - \hat{b}_0 - \hat{b}_1 X_i) = 0 & (1) \\ \sum(Y_i - \hat{b}_0 - \hat{b}_1 X_i) X_i = 0 & (2) \end{cases}$$

$$\begin{cases} ① \ \sum(Y_i - \hat{Y}_i) = 0 = \sum_{i=1}^{n} e_i \\ ② \ \sum(Y_i - \hat{Y}_i) X_i = 0 = \sum e_i X_i \end{cases}$$

$\sum y_i x_i - \sum \hat{y}_i x_i = 0 \Rightarrow \sum y_i x_i = \sum \hat{y}_i x_i$

Μένει ν.δ.ο. $\sum e_i \hat{y}_i = 0$

Θυμάμαι ότι $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$

$e_i \hat{y}_i = e_i \hat{b}_0 + \hat{b}_1 e_i x_i$

$\sum e_i y_i = \hat{b}_0 \cancelto{0}{\sum e_i} + \hat{b}_1 \cancelto{0}{\sum e_i x_i} = 0$

$y_i = b_0 + b_1 x_i + \varepsilon_i$ ν.δ.ο. $r^2_{y,\hat{y}} = R^2$

Υπενθύμιση: ΘΕΩΡΙΑ δείξαμε ότι $r^2_{XY} = R^2$

Λύση:

$$r^2_{y,\hat{y}} = \frac{\left[\sum (y_i - \bar{x})(\hat{y}_i - \bar{\hat{y}})\right]^2}{\sum (y_i - \bar{x})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}$$

ενώ

$$r^2_{y,x} = \frac{\left[\sum (y_i - \bar{y})(x_i - \bar{x})\right]^2}{\sum (y_i - \bar{y})^2 \sum (x_i - \bar{x})^2}$$

Παρατηρήσεις: Δείχνω όπως πριν ότι $\sum y_i = \sum \hat{y}_i$

δηλαδή (1) κανονικές εξισώσεις

Αυτό συμβαίνει $\bar{\hat{y}} = \bar{y}$

Επίσης $\hat{y}_i - \bar{\hat{y}} = \hat{b}_0 + \hat{b}_1 x_i - \bar{y}$

Όμως $\bar{y} = \hat{b}_0 - \hat{b}_1 \bar{x}$

$\hat{y}_i - \bar{\hat{y}} = \hat{b}_1 (x_i - \bar{x})$

$\bar{\hat{y}} = \hat{b}_1 (x_i - \bar{x})$

$$r^2_{y,y} = \frac{\left[\hat{b}_1 \sum (y_i - \bar{y})(x_i - \bar{x})^2\right]^2}{\hat{b}_1 \sum (y_i - \bar{x})^2 \sum (x_i - \bar{x})^2} = r^2_{y,x}$$

Άσκηση 4.10: (i) ΔΕ για μ κανονικού πληθυσμού όταν σ^2 άγνωστη

$\bar{x} \pm t_{n-1,\alpha/2}\, S/\sqrt{n}$

Εδώ $n=10$, $\alpha=0,05$, $t_{9,0.025}=2.262$

$\bar{X} = \frac{1058}{10} = 105.8$, $S^2 = 78.622$

$(L,U) = (99.626,\ 111.974)$

(ii) Περιοχή απόρριψης

$H_0: \mu=\mu_0 \quad (\mu_0=100) \qquad H_a: \mu \neq \mu_0$

$|t| \geq t_{n-1,\alpha/2}$

όπου $t = \frac{\bar{X} - \overbrace{100}^{\mu_0}}{S/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$

Είναι $t = \frac{105.8 - 100}{8.8668/\sqrt{10}} = 2.068$

$t_{0.025,9} = 2.262$

Άρα δεν ~~επ~~ απορρίπτεται η H_0, επομένως η μέση ~~τιμή πίεσης~~ συστολική πίεση του πληθυσμού δε διαφέρει στατιστικά σημαντικά από την τιμή 100 με επίπεδο σημαντικότητας 5%

Υπάρχει άλλος τρόπος να γίνει ο έλεγχος εδώ;

Άσκηση 4.11:

Είναι διαθέσιμο ένα τ.δ. μεγέθους $n=6$ $X_1 \dots X_6$ από $N(\mu, \sigma^2)$. Θέλουμε να ελέγξουμε την

$H_0: \mu = 9.75$

έναντι της

$H_a: \mu \neq 9.75$

Καθώς σ^2 είναι άγνωστη:

Είναι $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - 9.75}{S/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$

Περιοχή απόρριψης: $|t| \geq t_{n-1,\alpha/2}$

Με εφαρμογή προκύπτει $|t| = |1.239| \ngeq 2.571 = t_{0.025,5}$

Άρα δεν απορρίπτεται η Ho, επομένως ο μέσος όρος πειραματος σε μήνες δε διαφέρει στατιστικά σημαντικά από την τιμή 9.75 με επίπεδο σημαντικότητας 5%

Άσκηση 4.14:

Δ.Ε για τη διαφορά $\mu_A - \mu_B$ κανονικών πληθυσμών με άγνωστες ~~*~~ αλλά ίσες διακυμάνσεις $\sigma_A^2 = \sigma_B^2 = \sigma^2$ με ανεξάρτητα δείγματα

$90\% \Rightarrow \quad a = 10\%$

Είναι

$$(\bar{x} - \bar{y}) \pm t_{n+m-2,\, a/2}\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

όπου

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \overset{n=8}{=\!=} \frac{15}{8} = 1.875$$

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{x})^2 = 0.696$$

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^{m} Y_i \overset{m=8}{=\!=} \frac{21}{8} = 2.625$$

$$S_Y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2 = 0.839$$

$$S_p^2 = \frac{(n-1) S_x^2 + (m-1) S_Y^2}{n+m-2} = 0.7675$$

$$\Downarrow$$

$$S_p = 0.876$$

$t_{14,0.05} = 1.761$

$(L, U) = (-1.521,\ 0.0213)$