3ο μάθημα 18/10/21 ①

Θεώρημα: Αν ισχύουν οι υποθέσεις για τα σφάλματα, τότε $\boxed{E(MSres)=\sigma^2}$ (MSres αμερόληπτος εκτιμητής της σ^2).

Απόδειξη: (Σκιαγράφηση)

$$MSres = \frac{SSres}{n-2}, \quad SSres = \sum_{1}^{n} (Y_i - \hat{Y}_i)^2, \quad SSres = SStot - SSreg.$$

$$E(SStot) = E.\Sigma(Y_i - \bar{Y})^2 = \sum^{n} E(Y_i - \bar{Y})^2 =$$

$$= \Sigma . E\left(b_0 + b_1 X_i + \varepsilon_i - \frac{1}{n} . \Sigma(b_0 + b_1 . X_i + \varepsilon_i)\right)^2 =$$

= κάνω πράξεις λαμβάνοντας υπόψιν ότι $E(\varepsilon_i - \bar{\varepsilon}) = 0$.

$$E(SStot) = b_1^2 . \sum_{1}^{n} (X_i - \bar{X})^2 + \sum_{1}^{n} E(\varepsilon_i - \bar{\varepsilon})^2 \qquad (1)$$

Αλλά $\sum E(\varepsilon_i - \bar{\varepsilon})^2 = \ldots$ πράξεις $\ldots = (n-1)\cdot\sigma^2$ (2)

Από (1), (2): $E(SS_{tot}) = \beta_1^2 \cdot \sum (X_i - \bar{X})^2 + (n-1)\cdot\sigma^2$ (3)

$SS_{reg} = \hat{\beta}_1^2 \cdot \sum_1^n (X_i - \bar{X})^2$

$E(SS_{reg}) = \left(\sum (X_i - \bar{X})^2\right)\cdot E(\hat{\beta}_1^2) \overset{E(\hat{\beta}_1^2) = Var(\hat{\beta}_1)}{=\!=\!=\!=}$ πράξεις

$E(SS_{reg}) = \sigma^2 + \beta_1^2 \cdot \sum_1^n (X_i - \bar{X})^2$ (4)

$E(MS_{res}) = \dfrac{1}{n-2}\cdot E(SS_{res}) = \dfrac{1}{n-2}\left(E(SS_{tot}) - E(SS_{reg})\right)$

$\overset{(3),(4)}{=\!=} \sigma^2$

<u>Θεώρημα</u>: Αν ισχύουν οι υποθέσεις για τα σφάλματα τότε: $\boxed{\dfrac{SS_{res}}{\sigma^2} \sim X^2_{n-2}}$

<u>Απόδειξη</u>: Από τις υποθέσεις: Y_i ασυσχέτιστα και ακολουθούν $N(\beta_0 + \beta_1 X_i, \sigma^2) \Leftrightarrow Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$ με $i = 1, \ldots, n$ ανεξάρτητες.

$\dfrac{Y_i - (\beta_0 + \beta_1 X_i)}{\sigma} \sim N(0,1)$ ή $\dfrac{(Y_i - (\beta_0 + \beta_1 X_i))^2}{\sigma^2} \sim X_1^2$

ή $\dfrac{\sum (Y_i - (\beta_0 + \beta_1 X_i))^2}{\sigma^2} \underset{\text{ανεξ}}{\overset{Y_i}{\sim}} X^2_{\sum_1^n 1} = X^2_n$

$\Leftrightarrow \dfrac{\sum_1^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2}{\sigma^2} \sim X^2_{n-2}$

γιατί θα έχω 2 έξτρα εξισώσεις με το $\hat{\beta}_0, \hat{\beta}_1$ (plug-in διαδικασία).

• Βιβλίο (Γκαζέλα - Μπεργκερ 2002)

Θεώρημα: Οι ΕΕΤ $\hat{b_0}, \hat{b_1}$ είναι ανεξ/τοι των SSres

Απόδειξη: ?? (Δεν έχουμε τις απαραίτητες γνώσεις)

▸ Συμπερασματολογία για τις παραμέτρους του μοντέλου της α.γ.π. $Y_i = b_0 + b_1 X_i + \varepsilon_i$, $i = 1, \dots, n$ - Στατιστικά τεστ και δ.ε. για b_0, b_1.

Τεστ Σημαντικότητας ~ Wald

Z-test: Έστω $X_1, \dots, X_n$ τ.δ. από $N(\mu, \sigma^2)$ με σ^2 γνωστό. Για τον έλεγχο της: $H_0: \mu = \mu_0$, μ_0 γνωστό η ΣΣΤ: $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$ υπό H_0,

η κ.π. διαμορφώνεται από την H_a και το κρίσιμο σημείο υπολογ. $a = P(\text{απορ. } H_0 \mid H_a \text{: αληθής})$

Wald: Η ΣΣΤ βασίζεται σε έναν εκτιμητή της παραμέτρου που διατυπώνει την H_0.

$$\text{ΣΣΤ} = \frac{\text{Εκτιμητής} - E(\text{Εκτιμητή υπό } H_0)}{\text{Τυπική απόκλιση του Εκτιμητή}}$$

Χρειάζεται η κατανομή της ΣΣΤ υπό την H_0, για να αξιοποιηθεί στο τελευταίο βήμα για τον προσδιορισμό του κ.σ.

Στατιστικό Τεστ για τον έλεγχο της $H_0: b_0 = b_0^*$ (b_0^* γνωστό), $H_a: b_0 \neq b_0^*$

Χρησιμοποιούμε την Ε.Ε.Τ : $\hat{b_0} \sim N\left(b_0, \frac{\sigma^2 \Sigma X_i^2}{n\Sigma(X_i-\bar{X})^2}\right)$

Υπό την Ho : $\hat{b_0} \sim N\left(b_0^*, \frac{\sigma^2 \cdot \Sigma X_i^2}{n\Sigma(X_i-\bar{X})^2}\right)$

$$\frac{\hat{b_0} - b_0^*}{\sqrt{\frac{\sigma^2 \Sigma X_i^2}{n\Sigma(X_i-\bar{X})^2}}} \sim N(0,1) \text{ υπό την Ho}$$

μεγάλες τιμές σημαίνει μεγάλος αριθμητής, άρα μεγάλη απόκλιση b_0, b_0^*, άρα θα έχω συμπέρασμα ότι Απορ. Ha

$$\frac{SSres}{\sigma^2} \sim X^2_{n-2}$$

Θεωρώ τη Σ.Σ.Τ. $t = \frac{\frac{\hat{b_0} - b_0^*}{\sqrt{\sigma^2 \Sigma X_i^2 / \Sigma(X_i-\bar{X})^2}}}{\sqrt{\frac{SSres}{\sigma^2} / (1-a)}} = \frac{\hat{b_0} - b_0^*}{\sqrt{\frac{\Sigma X_i^2}{n\Sigma(X_i-\bar{X})^2} \cdot MSres}}$

ή $t = \frac{\hat{b_0} - b_0^*}{\sqrt{\hat{Var}(\hat{b_0})}}$, όπου $\hat{Var}(\hat{b_0}) = \frac{\Sigma X_i^2}{\Sigma(X_i-\bar{X})^2} \cdot MSres$

- Κατανομή της t υπό την Ho: $b_0 = b_0^*$

$$t \equiv \frac{N(0,1)}{\sqrt{X^2_{n-2}/(n-2)}} \overset{\hat{b_0} \text{ ανεξ.}}{\underset{SSres}{\sim}} t_{n-2} \text{, υπό την Ho.}$$

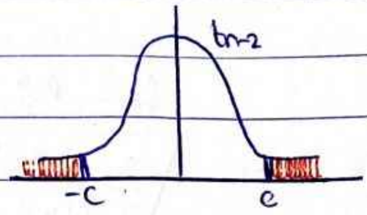
Εύρεση της μορφής της κ.π.

Η Ho απορρίπτεται για μεγάλες τιμές της Σ.Σ.Τ t, γιατί μεγάλες τιμές σημαίνουν ότι ο $\hat{b_0}$ είναι πολύ διαφορετικός από b_0^*, οπότε και b_0 πολύ διαφορετικός από b_0^*, άρα η Ho πρέπει να απορρ. Επομένως η $\boxed{|t| \geq c}$, αφού το t μπορεί να είναι και αρνητικό.

Υπολογισμός του κ.σ. c

$\alpha = P(\text{απορ. } H_0 \mid H_0: \text{αληθής}) = P(|t| \geq c \mid t \sim t_{n-2})$
$= P(t \geq c \text{ ή } t \leq -c \mid t \sim t_{n-2}) = 2 \cdot P(t \geq c \mid t \sim t_{n-2})$

$\Rightarrow P(t_{n-2} \geq c) = \frac{\alpha}{2} \Rightarrow \boxed{c = t_{n-2, \frac{\alpha}{2}}}$



Συγκεντρωτικά:

Για τον έλεγχο $H_0: \beta_0 = \beta_0^*$ έναντι $H_a: \beta_0 \neq \beta_0^*$, η ΣΣΤ είναι η: $t = \frac{\hat{\beta_0} - \beta_0^*}{\sqrt{\hat{Var}(\hat{\beta_0})}} \sim t_{n-2}$, υπό την H_0

, $\hat{Var}(\hat{\beta_0}) = \frac{\sum X_i^2}{n \cdot \sum (X_i - \bar{X})^2} \cdot MS_{res}$

$|t| \geq t_{n-2, \frac{\alpha}{2}}$.

Σπίτι) Την ίδια διαδικασία του τεστ για $H_0: \beta_1 = \beta_1^*$, με $\hat{\beta_1} \sim N(\beta_1, \frac{\sigma^2}{\sum (X_i - \bar{X})^2})$

(Αποτέλεσμα: ΣΣΤ η $t = \frac{\hat{\beta_1} - \beta_1^*}{\sqrt{\hat{Var}(\hat{\beta_1})}} \sim t_{n-2}$ υπό την H_0

και τα υπόλοιπα ίδια).

Πρακτική Αξία του τεστ:

α) Αν $\beta_1^* = 0$, τότε μη απορρ. της $H_0: \beta_1 = 0$ σημαίνει ότι $\nexists$ γραμμική σχέση μεταξύ της Y και X

β) Αποδοχή της $H_0: \beta_1 = \beta_1^*$ σημαίνει ότι μπορώ να δέχτω ότι η μεταβολή της Y σε μοναδιαία μεταβολή της X είναι ίση με b_1^*.

Διάστημα Εμπιστοσύνης (δ.ε.) για b_0 και b_1

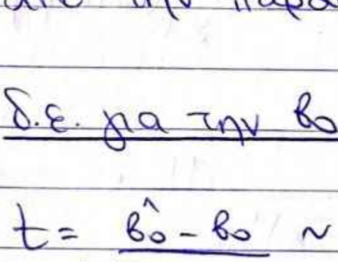
Αντιστρεπτή Ποσότητα: Ποσότητα που περιέχει την παράμετρο και η κατανομή της δεν εξαρτάται από την παράμετρο.

δ.ε. για την b_0: Θεωρώ ως αντιστρεπτή ποσότητα την

$$t = \frac{\hat{b}_0 - b_0}{\sqrt{\hat{Var}(\hat{b}_0)}} \sim t_{n-2} \quad \text{με} \quad \hat{Var}(\hat{b}_0) = \frac{\sum X_i^2}{n \cdot \sum (X_i - \bar{X})^2} \cdot MS_{res}$$

Αφού η t είναι αντιστρεπτή $\exists\, q_1, q_2$ (ανεξ/τα από την b_0) με $q_1 < q_2$ τ.ω: $1-a = P(q_1 \le t \le q_2) =$

$$= P\left(q_1 \le \frac{\hat{b}_0 - b_0}{\sqrt{\hat{Var}(\hat{b}_0)}} \le q_2\right) = P\left(\hat{b}_0 - q_2 \sqrt{\hat{Var}(\hat{b}_0)} \le b_0 \le \hat{b}_0 - q_1 \cdot \sqrt{\hat{Var}(\hat{b}_0)}\right)$$



Από τον ορισμό του δ.ε. ένα $100\% \cdot (1-a)$ δ.ε. για την b_0 είναι το $\left(\hat{b}_0 - q_2 \sqrt{\hat{Var}(\hat{b}_0)},\ \hat{b}_0 - q_1 \cdot \sqrt{\hat{Var}(\hat{b}_0)}\right)$

Όπως για την περίπτωση δ.ε. για το μ κανονικού πληθυσμού, έτσι και εδώ το δ.ε. ελαχ. μήκους ή το δ.ε. ίσων ουρών για τη b_0 προκύπτει για:

$$\boxed{q_1 = t_{n-2, \frac{a}{2}}, \quad q_2 = t_{n-2, \frac{a}{2}}}$$

<u>Συγκεντρωτικά</u>: Το δ.ε. για την β_0 με β.ε. $1-a$, με $0<a<1$ είναι $\left(\hat{\beta_0} - t_{n-2,\frac{a}{2}}\sqrt{\hat{Var}(\hat{\beta_0})}, \hat{\beta_0} + t_{n-2}\ldots\right)$

όπου $\hat{Var}(\hat{\beta_0}) = \dfrac{\Sigma X_i^2}{\Sigma(X_i-\bar{X})^2} \cdot MSres$

- Το ίδιο και για τη β_1 με δ.ε.
$\left(\hat{\beta_1} - t_{n-2,\frac{a}{2}}\sqrt{\hat{Var}(\hat{\beta_1})}, \hat{\beta_1} + t_{n-2,\frac{a}{2}}\sqrt{\hat{Var}(\hat{\beta_1})}\right)$

όπου: $\hat{Var}(\hat{\beta_1}) = \dfrac{1}{\Sigma(X_i-\bar{X})^2} \cdot MSres$

<u>Πρακτική Αξία του δ.ε. για τη β_1</u>:

Το δ.ε. για την β_1 δείχνει τα όρια ανάμεσα στα οποία κινείται η μεταβολή της Υ σε μοναδιαία μεταβολή της Χ, με πιθανότητα $1-a$, $0<a<1$.