ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΟΜΕΑΣ ΠΙΘΑΝΟΤΗΤΩΝ-ΣΤΑΤΙΣΤΙΚΗΣ
& ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

ΠΑΛΙΝΔΡΟΜΗΣΗ & ΑΝΑΛ. ΔΙΑΚΥΜΑΝΣΗΣ
ΦΥΛΛΟ ΑΣΚΗΣΕΩΝ 1
Ακαδ. Ετος: 2016-2017

ΑΣΚΗΣΕΙΣ ΣΤΗΝ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

ΑΣΚΗΣΗ 1. Εστω το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, i=1,…,n, με τα σφάλματα $\varepsilon_i \sim N(0,\sigma^2)$ και ασυσχέτιστα ανά δύο. Να δειχθεί ότι οι εκτιμητές ελαχίστων τετραγώνων των β_0 και β_1 του μοντέλου αυτού, συμπίπτουν με τους εκτιμητές μέγιστης πιθανοφάνειάς τους.

ΑΣΚΗΣΗ 2. Εστω το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, i=1,…,n, με $E(\varepsilon_i)=0$, $Var(\varepsilon_i)=\sigma^2$ και τα ε_i, i=1,…,n, είναι ασυσχέτιστα. Αν $\hat{\beta}_0$ και $\hat{\beta}_1$ οι εκτιμητές ελαχίστων τετραγώνων των β_0 και β_1, να δειχθούν τα ακόλουθα:

α) $E(\hat{\beta}_0)=\beta_0$, $E(\hat{\beta}_1)=\beta_1$, $Var(\hat{\beta}_0)=\sigma^2\left\{\frac{1}{n}+\frac{\overline{X}^2}{\Sigma(X_i-\overline{X})^2}\right\}=\frac{\sigma^2\Sigma X_i^2}{n\Sigma(X_i-\overline{X})^2}$, $Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$.

β) $\sum_{i=1}^{n} X_i e_i = 0$, με $e_i = Y_i - \hat{Y}_i$, i=1,…,n, τα υπόλοιπα.

γ) $Cov(\hat{\beta}_0,\hat{\beta}_1) = -\frac{\sigma^2\overline{X}}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$.

ΑΣΚΗΣΗ 3. Τα παρακάτω δεδομένα αναφέρονται στην 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	18

α) Αφού ορισθεί η εξαρτημένη και η ανεξάρτητη μεταβλητή να βρεθούν οι εκτιμητές ελαχίστων τετραγώνων των παραμέτρων του μοντέλου.
β) Να κατασκευαστεί ο πίνακας ΑΝΑΔΙΑ και να γίνει ένα αρχικό τεστ για τον έλεγχο της υπόθεσης H_0: β_1=0.
γ) Κάνοντας τις κατάλληλες υποθέσεις
 ι) Να ελεγχθεί η υπόθεση H_0: β_1=0 σε επίπεδο σημαντικότητας α=0.05.
 ιι) Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για το β_1.

ΑΣΚΗΣΗ 4. Για το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, i=1,…,n, να δειχθεί ότι:

α) $r_{X,Y}^2 = R^2$, με $r_{X,Y}$ το δειγματικό συντελεστή συσχέτισης του Pearson και R^2 το συντελεστή προσδιορισμού.

β) Αν $r_{Y,\hat{Y}}$ είναι ο δειγματικός συντελεστής συσχέτισης του Pearson μεταξύ της παρατηρούμενης και της εκτιμώμενης τιμής του Y, να δειχθεί ότι $r_{Y,\hat{Y}}^2 = R^2$.

ΑΣΚΗΣΗ 5. Τα παρακάτω δεδομένα συγκεντρώθηκαν σε μια μελέτη που σκοπό είχε να εξετασθεί ποιοι παράγοντες επιδρούν στο βάρος ενός νεογέννητου μωρού.

Βάρος κατά τη γέννηση (gr)	1361	1588	1815	2087	2268	2404	3402	3629	3765	4083
Κοινωνικο-οικο-νομική κατάσταση	8	7	4	5	5	4	3	3	2	1
Σειρά γεννήσεως	4	3	4	3	2	2	2	1	1	1

ι) Να ορισθεί η εξαρτημένη και οι ανεξάρτητες μεταβλητές.

ιι) Να βρεθούν οι πίνακες $\mathbf{X'X}$, $\mathbf{(X'X)^{-1}}$, $\mathbf{X'Y}$.

ιιι) Να βρεθούν οι εκτιμητές ελαχίστων τετραγώνων του μοντέλου $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$, και να δοθεί ο πίνακας διακυμάνσεων-συνδιακυμάνσεών τους. Να υπολογισθεί ο δειγματικός συντελεστής συσχέτισης μεταξύ των $\hat{\beta}_1$ και $\hat{\beta}_2$.

iv) Να δοθεί ο πίνακας ΑΝΑΔΙΑ και να βρεθεί ένας αμερόληπτος εκτιμητής για τη διακύμανση σ^2 των σφαλμάτων του μοντέλου.

v) Υποθέτοντας ότι τα σφάλματα ακολουθούν κανονική κατανομή να ελεγχθούν οι υποθέσεις,

α) $\beta_1=\beta_2=0$, β) $\beta_0=0$, γ) $\beta_1=0$, δ) $\beta_2=0$.

vi) Να δοθεί η φυσική ερμηνεία του εκτιμώμενου μοντέλου.

ΑΣΚΗΣΗ 6. Το γραφείο υποδοχής ασθενών ενός νοσοκομείου, θέλοντας να εξετάσει ποιοι παράγοντες μπορεί να επιδρούν στη διάρκεια παραμονής των ασθενών, που υποβάλονται σε κάποιο είδος εγχείρησης στο νοσοκομείο, συγκέντρωσε τα παρακάτω δεδομένα για 20 τέτοιους ασθενείς.

Α/Α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ημέρες παραμονής Μετά την εγχείρηση	6	6	11	9	16	16	4	8	11	13	13	9	17	17	12
Αριθ. Ιατρικών Προβλημάτων	1	2	2	1	3	1	1	3	2	3	1	1	3	2	4
Ημέρες Παραμονής Πρίν την εγχείρηση	1	1	2	3	3	5	1	1	2	2	4	2	3	4	1

16	17	18	19	20
6	5	12	8	9
1	1	3	1	2
1	1	2	2	2

Κάνοντας τις κατάλληλες υποθέσεις να αναλυθούν τα δεδομένα και να διατυπωθούν τα συμπεράσματα.

ΑΣΚΗΣΗ 7. Έστω Y_1 και Y_2 δύο ανεξάρτητες τυχαίες μεταβλητές με μέσες τιμές θ και 2θ αντίστοιχα, με θ πραγματικό αριθμό. Χρησιμοποιώντας το μοντέλο της γραμμικής παλινδρόμησης να προσδιορισθεί ο εκτιμητής ελαχίστων τετραγώνων της παραμέτρου θ και να υπολογισθεί το άθροισμα τετραγώνων των υπολοίπων.

ΑΣΚΗΣΗ 8. Για το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης να δειχθεί ότι

$$(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}).$$

ΑΣΚΗΣΗ 9. Αν σε ένα γραμμικό μοντέλο υπάρχει σταθερός όρος, τότε το άθροισμα των υπολοίπων ισούται με το μηδέν.

ΑΣΚΗΣΗ 10. Θεωρούμε το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$, με $Rank(\mathbf{X}) = p+1$. Να δειχθεί ότι,

α) $\sum_{i=1}^{n} \hat{Y}_i(Y_i - \hat{Y}_i) = 0$, και β) $\sum_{i=1}^{n} Var(\hat{Y}_i) = \sigma^2(p+1)$.

ΑΣΚΗΣΗ 11. Για το μοντέλο $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$, $i=1,\ldots,n$, με ε_i ανεξάρτητες $N(0,\sigma^2)$, δίνονται οι κανονικές εξισώσεις:

$$10\hat{\beta}_0 + 2\hat{\beta}_1 - 6\hat{\beta}_2 = 4$$
$$2\hat{\beta}_0 + 2\hat{\beta}_1 = 6$$
$$-6\hat{\beta}_0 + 5\hat{\beta}_2 = 7\,.$$

α) Αν n=10 και $\sum_{i=1}^{n} Y_i^2 = 107$, να υπολογισθούν ο εκτιμητής $\hat{\boldsymbol{\beta}}$ και το MS_{res}.

β) Να ελεγχθεί με τη βοήθεια ενός F - τεστ η μηδενική υπόθεση $H_0: \beta_1 = 2\beta_2$, σε επίπεδο a =5%.

γ) Να κατασκευαστεί ένα t-τεστ για τον έλεγχο της υπόθεσης $H_0: \beta_1 = 2\beta_2$, σε επίπεδο a =5%.

ΑΣΚΗΣΗ 12. Έστω το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, με $\sigma^2 = Var(\varepsilon_i)$, $v_i = E(Y_i)$, και $q_i = \left(Y_i - \hat{Y}_i\right) - \left(v_i - E\left(\hat{Y}_i\right)\right)$, $i=1,\ldots,n$. Να δειχθεί ότι

$$E\left(\sum_{i=1}^{n} q_i^2\right) = (n-2)\sigma^2\,.$$

ΑΣΚΗΣΗ 13. Έστω το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$, με $Rank(\mathbf{X}) = p+1$.

α) Να δειχθεί ότι, $SS_{reg} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2$.

β) Υπό την υπόθεση ότι το διάνυσμα των σφαλμάτων $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$, να δειχθεί ότι οι εκτιμητές ελαχίστων τετραγώνων $\hat{\boldsymbol{\beta}}$ των παραμέτρων $\boldsymbol{\beta}$ του μοντέλου, ταυτίζονται με τους εκτιμητές μέγιστης πιθανοφάνειας.

γ) Υπό την υπόθεση $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ να δειχθεί ότι $\frac{SS_{res}}{\sigma^2} \sim \chi^2_{n-p-1}$.

ΑΣΚΗΣΗ 14. Έστω το μοντέλο παλινδρόμησης

$$Y_i = \beta X_i + \varepsilon_i,\ i = 1,\ldots,n,$$

για το οποίο ισχύει ότι τα σφάλματα ε_i ακολουθούν κανονική κατανομή με $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2 X_i$ και $Cov(\varepsilon_i, \varepsilon_j) = 0$, για $i, j = 1,\ldots,n$, $i \neq j$.

α) Να βρεθεί ο εκτιμητής μέγιστης πιθανοφάνειας (ή ο εκτιμητής ελαχίστων τετραγώνων) και να δειχθεί ότι είναι αμερόληπτος εκτιμητής του β.

β) Να δείξετε ότι οι $\beta^{*}=\dfrac{Y_1+Y_2}{X_1+X_2}$, $\beta^{**}=\dfrac{Y_n-Y_1}{X_n-X_1}$ και $\beta^{***}=\dfrac{X_1Y_1+X_2Y_2}{X_1^2+X_2^2}$ είναι επίσης αμερόληπτοι εκτιμητές του β.

γ) Γιατί οι εκτιμητές του προηγούμενου ερωτήματος δεν είναι αποτελεσματικοί;

ΑΣΚΗΣΗ 15. Για το μοντέλο της απλής γραμμικής παλινδρόμησης

$$Y_i=\beta_0+\beta_1X_i+\varepsilon_i,\ i=1,\ldots,n,$$

να δείξετε ότι

i) $\sum_{i=1}^{n}\widehat{Y}_i=\sum_{i=1}^{n}Y_i$, και $\sum_{i=1}^{n}e_i=0$, ii) $\overline{\widehat{Y}}=\overline{Y}$, iii) $\overline{Y}=\widehat{\beta}_0+\widehat{\beta}_1\overline{X}$, iv) $\sum_{i=1}^{n}\widehat{Y}_iX_i=\sum_{i=1}^{n}Y_iX_i$, και $\sum_{i=1}^{n}e_i\widehat{Y}_i=0$,

v) $\sum_{i=1}^{n}e_iX_i=0$.

ΑΣΚΗΣΗ 16. Έστω ότι $\widehat{\beta}_0,\widehat{\beta}_1,\ldots,\widehat{\beta}_p$ είναι οι εκτιμητές ελαχίστων τετραγώνων του μοντέλου πολλαπλής παλινδρόμησης

$$Y_i=\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}+\varepsilon_i,\ i=1,\ldots,n.$$

Να βρεθούν συναρτήσει των $\widehat{\beta}_0,\widehat{\beta}_1,\ldots,\widehat{\beta}_p$ οι εκτιμητές ελαχίστων τετραγώνων του μοντέλου

$$Y_i=\beta_0+\beta_1\mathrm{Z}_{i1}+\ldots+\beta_p\mathrm{Z}_{ip}+\varepsilon_i,\ i=1,\ldots,n,$$

όπου $\mathrm{Z}_{ij}=c_jX_{ij},\ j=1,\ldots,p$, με $c_j,\ j=1,\ldots,p$, σταθερές και $i=1,\ldots,n$.

ΑΣΚΗΣΗ 17. Για το μοντέλο της απλής γραμμικής παλινδρόμησης

$$Y_i=\beta_0+\beta_1X_i+\varepsilon_i,\ i=1,\ldots,n,$$

να δείξετε ότι οι εκτιμητές ελαχίστων τετραγώνων των β_0,β_1 είναι ασυσχέτιστοι αν $\sum_{i=1}^{n}X_i=0.$

ΑΣΚΗΣΗ 18. Έστω το μοντέλο παλινδρόμησης

$$Y_i=\beta X_i+\varepsilon_i,\ i=1,\ldots,n,$$

για το οποίο ισχύει ότι τα σφάλματα ε_i ακολουθούν κανονική κατανομή με $E(\varepsilon_i)=0,\ Var(\varepsilon_i)=\dfrac{\sigma^2}{X_i}$ και $Cov(\varepsilon_i,\varepsilon_j)=0$, για $i,j=1,\ldots,n,\ i\neq j$. Να βρεθεί ο γραμμικός αμερόληπτος εκτιμητής του β με την μικρότερη διακύμανση.

ΑΣΚΗΣΗ 19. Για το κλασικό μοντέλο απλής γραμμικής παλινδρόμησης

$$Y_i=\beta_0+\beta_1X_i+\varepsilon_i,\ i=1,\ldots,n,$$

να βρείτε τις συνθήκες που πρέπει να ικανοποιούνται έτσι ώστε ένας γραμμικός εκτιμητής του β_1 να είναι αμερόληπτος. Πότε αυτός είναι ο BLUE;

(Υπενθύμιση: Ένας εκτιμητής είναι γραμμικός όταν είναι της μορφής $\sum_{i=1}^{n} d_i Y_i$).

ΑΣΚΗΣΗ 20. Στο μοντέλο της πολλαπλής γραμμικής παλινδρόμησης $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, με $Rank(\mathbf{X}) = p+1$, τα μαθηματικοποιημένα υπόλοιπα ορίζονται

$$t_i = \frac{e_i}{\sqrt{MS_{res}}\sqrt{1-\rho_{ii}}}, \quad i=1,\ldots,n,$$

με ρ_{ij} να συμβολίζει το στοιχείο της (i,j) –θέσης του πίνακα $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Να δειχθεί ότι

$$t_i \sim t_{n-p-1}.$$

ΑΣΚΗΣΗ 21. Έστω οι παρατηρήσεις $Y_1 = \alpha_1 + \varepsilon_1$, $Y_2 = 2\alpha_1 - \alpha_2 + \varepsilon_2$ και $Y_3 = \alpha_1 + 2\alpha_2 + \varepsilon_3$, με $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N_3(\mathbf{0}, \sigma^2 \mathbf{I}_3)$. Με τη θεωρία των γραμμικών μοντέλων να ελεγχθεί η υπόθεση $H_0 : \alpha_1 = \alpha_2$.

7/11/16

1

Παρασκευή	11/11/16
Μάθημα	15:00

Πολυσυγγραμμικότητα και αναλ

▶ Το F-test για έλεγχο πολυσυγγραμμικότητας

ή το F-test για έλεγχο της $H_0: B_1=0$ εναντι $H_a: B_1 \neq 0$

Για τον έλεγχο αυτόν κατασκευάζουμε t-test ($H_0: B_0=B_1^*$, B_1^* γνωστό)

▶ **Ιδέα** Γνωρίζουμε ότι $E(MS_{res})=\sigma^2$

Ισχύει $E(MS_{reg})=\sigma^2+B_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2$

[Πράγματι: $E(MS_{reg})=E\left(\frac{SS_{reg}}{1}\right)=E(SS_{reg})$

$=E\left(\hat{B}_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$

Εφόσον: $Var(w)=E(w^2)-E(w)^2$ τα χαρακτηριστικά σταθερά

$\longrightarrow$ $=\sum_{i=1}^{n}(x_i-\bar{x})^2 E(\hat{B}_1^2)$

$=\sum_{i=1}^{n}(x_i-\bar{x})^2\left[Var(\hat{B}_1)+E(B_1)^2\right]$

$=\sum_{i=1}^{n}(x_i-\bar{x})^2\left[\sigma^2+B_1^2\sum(x_i-\bar{x})^2\right]$]

οπότε Υπο την συνθήκη $H_0: B_1=0$, αν δηλαδή η μηδενική υπόθεση ισχύει

τότε $E(MS_{reg})=E(MS_{res})$ η $MS_{reg}\cong MS_{res}$

απο προτασιακό λογισμό: Αν $A \Rightarrow B$ τότε $\sim B \Rightarrow \sim A$

Αν MS_{res} πολυ διαφορετικό MS_{reg} (οπότε έχω βρει αμέσως μία συνθήκη)

τότε η $H_0: B_1=0$ δεν ισχύει

άρα απορρ

Άρα ένα τεστ για τον έλεγχο της $H_0: B_1 = 0$ μπορεί να ~~[illegible]~~

στηρίχτεί στην σύγκριση των MSreg και MSres - (Και διαλέγω να συγκρίνω βάση το πηλίκο και όχι την διαφορά τους διότι ξέρω την κατανομή)

Και γνωρίζοντας ότι $SSres/σ^2 \sim χ^2_{n-1}$ συγκρίνω τα MSreg και MSres

με βάση το πηλίκο τους: $MSreg/MSres = \frac{SSreg/1}{SSres(n-2)}$

οπότε απομένει να βρω κατανομή για SSreg.

άρα $SSreg = \hat{B}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$

Υπό τις υποθέσεις για σφάλματα

$\hat{B}_1 \sim N\left(B_1, \frac{σ^2}{\sum(x_i-\bar{x})^2}\right)$, υπό την $H_0: B_1 = 0$ έχω

$\hat{B}_1 \sim N\left(0, \frac{σ^2}{\sum(x_i-\bar{x})^2}\right)$

$\Downarrow$

$\frac{\hat{B}_1 - 0}{\sqrt{\frac{σ^2}{\sum(x_i-\bar{x})^2}}} \sim N(0,1)$

$\Downarrow$

$\frac{\hat{B}_1 \sum(x_i-\bar{x})^2}{σ^2} \sim N(0,1)$

$\Downarrow$ και άρα

$\boxed{\frac{SSreg}{σ^2} \sim χ_1^2}$ υπό την συνθήκη $H_0: B_1 = 0$

Τελικά θεωρώ την στατιστική συνάρτηση

$F = \frac{MSreg}{MSres} = \frac{SSreg/1}{SSres/(n-2)} = \frac{(SSreg/σ^2)/1}{(SSres/σ^2)/(n-2)} = \frac{χ_1^2/1}{χ^2_{n-2}/(n-2)} \sim F_{1,n-2}$ υπό $H_0: B_1 =$

οπότε κατάφερα στηριζόμενος στα MSreg και MSres να βρώ την κατανομή

▶ Άρα ένα τεστ για τον έλεγχο της $H_0: B_1=0$ μπορεί να βασιστεί

στηρίζεται στην σύγκριση των MSreg και MSres. (Και θέλω να συγκρίνω με βάση το πηλίκο και όχι την διαφορά τους διότι ξέρω την κατανομή)

Και γνωρίζοντας ότι $SSres/σ^2 \sim χ^2_{n-1}$ συγκρίνω τα MSreg και MSres

με βάση το πηλίκο τους: $MSreg/MSres = \dfrac{SSreg/1}{SSres/(n-2)}$

▶ οπότε απομένει να βρω κατανομή για SSreg.

άρα $SSreg = \hat{B}_1^2 \sum_{i=1}^{n} (x_i-\bar{x})^2$

Υπό τις υποθέσεις για σφάλματα

$\hat{B}_1 \sim N\left(B_1, \dfrac{σ^2}{\sum(x_i-\bar{x})^2}\right)$, υπό την $H_0: B_1=0$ έχω

$\hat{B}_1 \sim N\left(0, \dfrac{σ^2}{\sum(x_i-\bar{x})^2}\right)$

$\Downarrow$

$\dfrac{\hat{B}_1 - 0}{\sqrt{\dfrac{σ^2}{\sum(x_i-\bar{x})^2}}} \sim N(0,1)$

$\Downarrow$

$\dfrac{\hat{B}_1 \sum(x_i-\bar{x})^2}{σ^2} \sim N(0,1)$

$\Downarrow$ και άρα

$\boxed{\dfrac{SSreg}{σ^2} \sim χ_1^2}$ υπό την συνθήκη $H_0: B_1=0$

Τελικά θεωρώ την στατιστική συνάρτηση

$F = \dfrac{MSreg}{MSres} = \dfrac{SSreg/1}{SSres/(n-2)} = \dfrac{(SSreg/σ^2)/1}{(SSres/σ^2)/(n-2)} = \dfrac{χ_1^2/1}{χ^2_{n-2}/(n-2)} \sim F_{1,n-2}$ υπό $H_0: B_1=$

οπότε κατάφερα στηριζόμενος στα MSreg και MSres να βρω την κατανομή

▶ Για τον έλεγχο της $H_0: \beta_1 = 0$ η στατιστική συνάρτηση τεστ είναι $F = \frac{MSreg}{MSres}$ με κατανομή $F_{1,n-2}$ υπό H_0

και Κ.Π της μορφής μεγάλης τιμής του F δηλ $F \geq C$: κρίσιμο σημείο

Και για να ολοκληρώσω βρίσκω το κρίσιμο σημείο οπότε

<u>Υπολογισμός</u>

$$a = P(\text{απορ}H_0 / H_0 \text{ αληθ}) = P(F \geq C / F \sim F_{1,n-2})$$

$$= P(F_{1,n-2} \geq C) \Rightarrow \boxed{C = F_{1,n-2,a}}$$

δηλαδή $F \geq F_{1,n-2,a}$

<u>Παρατήρηση</u> : Το F-test ισοδύναμο με t-test <u>διατί</u>:

- $$F = \frac{MSreg}{MSres} = \frac{SSreg}{MSres} = \frac{\hat{\beta}_1^2 \Sigma(x_i - \bar{x})^2}{MSres}$$

$$= \left(\frac{\hat{\beta}_1}{\sqrt{\frac{MSres}{\Sigma(x_i - \bar{x})^2}}} \right)^2 = \left(\frac{\hat{\beta}_1 - 0}{\sqrt{v\hat{a}r(\hat{\beta}_1)}} \right) = t^2$$

(δηλ παρόλου που είχα διαφορετική προσέγγιση στα δύο τεστ τελικά αυτά είναι ισοδύναμα)

Συντελεστής Συσχέτισης Pearson

Πληθυσμιακή μορφή : Αν X, Y είναι τ.μ. τότε $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$

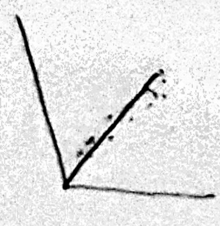
Δειγματική μορφή : αν έχω τ.δ.: $x_1, \dots, x_n$, $y_1, \dots, y_n$ τότε $r(x,y) = \dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}}$

Ιδιότητες: 1. είναι καθαρός αριθμός

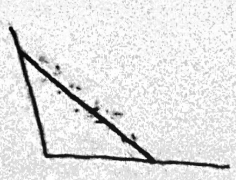
2. συμμετρικός $r_{x,y} = r_{y,x}$

3. $-1 \le r_{x,y} \le 1$

Το $r_{x,y}$ σχετίζεται με διαγράμματα διασποράς

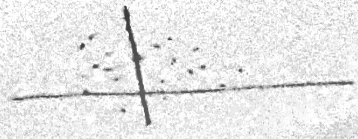


$r_{x,y} = 1$



$r_{x,y} = -1$

$r_{x,y} = 0 \Leftrightarrow \nexists$ γραμμική σχέση μεταξύ των x, y



• Αν $r_{X,Y}=0$ τότε ∄ γραμμική σχέση μπορεί όμως

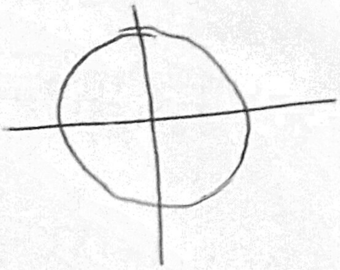
άλλου είδους σχέση

Παράδειγμα

X	-3	1	0	1	3
Y	9.54	9.95	10	9.95	9.54

$r_{X,Y}=0$ Άρα ∄ γραμμική σχέση αλλά

$x^2+y^2=100$



Ασκήσεις από φυλλ

Ασκ 1 $\quad Y_i=\beta_0+\beta_1 X_i+\varepsilon_i \quad i=1,\dots,n$

Ισχύουν οι υποθέσεις σφάλματα $\varepsilon_i \sim N(0,\sigma^2)$

Ν.Δ.Ο ΕΕΤ=ΕΜΠ

Να βρεθεί ΕΜΠ της σ^2 (Άσκηση για σπίτι)

[Στατιστική Συμπερασματολογία: Έστω $x_1,\dots,x_n$ τ.δ από πληθυσμό με κατανομή $f(x,\theta)$

Εκτιμητής για το θ

Εκτιμητής Μέγιστης Πιθανοφάνειας (ΕΜΠ) για το θ

Πιθανοφάνεια: $L=L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$

Ο ΕΜΠ $\hat{\theta}$ είναι εκείνος ο οποίος μεγιστοποιεί ως προς θ την L ή την log L

για μεγ της L ή $\log L$: $\frac{\partial L}{\partial \theta}=0$ λύνω... $\rightsquigarrow \hat{\theta}=\dots$]

...ίσ Λύση

• ΕΕΤ προκύπτουν από Ελαχισ των B_0, B_1 : $S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum (Y_i - B_0 - B_1 x_i)^2 \quad (*)$

• Πιθανοφάνεια : $L = L(B_0, B_1) = \prod f_{Y_i}(y_i)$

$$Y_i \sim N(B_0 + B_1 x_i, 6^2)$$

$$F_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi 6^2}} e^{-1/26^2 (y_i - B_0 - B_1 x_i)^2}$$

$$\circ\ L(B_0, B_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi 6^2}} e^{-1/26^2 (y_i - B_0 - B_1 x_i)^2}$$

$$= \frac{1}{(2\pi 6^2)^{n/2}} e^{1/26^2 \sum_{i=1}^{n} (y_i - B_0 - B_1 x_i)^2}$$

• $\log L(B_0, B_1) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log 6^2 - \frac{1}{26^2} \sum (Y_i - B_0 - B_1 x_i)^2$

κοιτάω ποιοι όροι έχουν B_0, B_1 άρα αρκεί να μεγιστοποιήσω ως προς B_0, B_1 το $\log L$ δηλ του όρο $-\frac{1}{26^2} \sum (Y_i - B_0 - B_1 x_i)^2$

Άρα αρκεί να μεγιστοποιήσω το $-\sum_{i=1}^{n} (Y_i - B_0 - B_1 x_i)^2$

Άρα αρκεί να ελαχιστοποιήσω το $\sum_{i=1}^{n} (Y_i - B_0 - B_1 x_i)^2 \quad (**)$

από $(*)$ και $(**)$ $\longrightarrow$ ΕΕΤ = ΕΜΠ

(Για 2ο παίρνω $\frac{\partial}{\partial 6^2} \log L(\hat{B_0}, \hat{B_1}) = 0$

$\Downarrow$

$\vdots$

$\Downarrow$

$\hat{6}^2 = \frac{1}{n} SS_{res} = \frac{n-2}{n} MS_{res}$. Παρατήρηση ο ΕΜΠ $\hat{6}^2$ της 6^2 δεν είναι αμερόληπτος)

Ασκηση 2

α. $\hat{B}_1 \sim N(..)$, $\hat{B}_0 \sim N\left(B_0, \sigma^2\left\{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right\} = \sigma^2\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}\right)$

↳ Το έχουμε αποδείξει

Λύση

$\hat{B}_0 = \bar{Y} - \hat{B}_1\bar{X}$. Ψάχνω αρχικά μ.τ και διακύμανση

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

$$= E\left(\frac{1}{n}\sum Y_i\right) - \bar{X}B_1$$

$$= \frac{1}{n}E(Y_i) - \bar{X}B_1$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(B_0 + B_1 X_i + \varepsilon_i) - \bar{X}B_1$$

$$= \frac{1}{n}\sum_{i=1}^{n}(B_0 + B_1 X_i) - \bar{X}B_1$$

$$= B_0 + B_1\bar{X} - \bar{X}B_1 = B_0$$

- $Var(\hat{B}_0) = Var(\bar{Y} - \hat{B}_1 x)$

[Επειδή $Var\left(\sum_{i=1}^{n} a_i w_i\right) = \sum_{i=1}^{n} a_i^2 Var(w_i) + \underbrace{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} a_i a_j (w_i, w_j)}_{0}$]

άρα

$$Var(\hat{B}_0) = Var(\bar{Y}) + \bar{X}^2 Var(\hat{B}_1) - \bar{X}\cancel{Cov(\bar{Y}, B_1)}$$

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

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

[Επειδή $Var\,\bar{Y} = Var\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n^2}Var\left(\sum Y_i\right)$

Yi ασυσχέτιστα $\quad \frac{1}{n^2}\sum Var(Y_i) = \frac{1}{n^2}\sum\sigma^2$

$$= \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n}\ ]$$

άρα

$$Var(\hat{B}_0) = \frac{\sigma^2}{n} + \bar{x}^2 Var(\hat{B}_1) = \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}$$

$$= \sigma^2 \left\{ \frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i-\bar{x})^2} \right\}$$

Το $\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{x}$ είναι γραμμικός συνδυασμός των ασυσχέτιστων κανονικών $\bar{Y}$ και $B_1 \Rightarrow \hat{B}_0 \sim N(E(\hat{B}_0), Var(\hat{B}_0))$

(β) να δείξω $\sum_{i=1}^{n} x_i e_i = 0$ $(\underset{\sim}{x} \perp \underset{\sim}{e})$

• $\sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} x_i(y_i - \hat{y}_i) = \sum_{i=1}^{n} x_i y_i - \Sigma x_i \hat{y}_i$

$= \Sigma x_i y_i - \Sigma x_i (\hat{B}_0 + \hat{B}_1 x_i)$

$= \Sigma x_i y_i - \hat{B}_0 \Sigma x_i - \hat{B}_1 \Sigma x_i^2$ (το οποίο εκ πρώτης όψεως δεν κάνει '0')

όμως <u>κανονική εξίσωση</u> : $B_0 \Sigma x_i + B_1 \Sigma x_i = \Sigma x_i y_i$

τό οποίο επαληθεύεται από τα $\hat{B}_0, \hat{B}_1$,

Από $\underline{\Sigma x_i y_i - \hat{B}_0 \Sigma x_i - \hat{B}_1 \Sigma x_i^2 = 0}$

(γ) $Cov(\hat{B}_0, \hat{B}_1) = Cov(\bar{Y} - \bar{x}\hat{B}_1, \hat{B}_1)$

<u>Λύση</u> επειδή $Cov(\hat{B}_0, \hat{B}_1) = \cancel{Cov(\bar{Y}, \hat{B}_1)}^{\,0} - \bar{x}\, Cov(\hat{B}_1, \hat{B}_1)$

$= -\bar{x}\, Var(\hat{B}_1)$

$= -\bar{x} \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}$