Άσκηση (ΕΒΟ) (sos)

Η ελληνική Βιομηχανία Όπλων (ΕΒΟ) απασχολεί μέρος του εργατικού δυναμικού της κατασκευάζοντας κάννες ενός τύπου όπλων. Η απασχόληση σε ώρες εργασίας ανα μήνα διαφέρει από μήνα σε μήνα ανάλογα με το μέγεθος της παραγγελίας που πρέπει να εκτελεστεί. Στον παρακάτω πίνακα δίνονται οι παραγγελίες προς την ΕΒΟ των 10 τελευταίων μηνών μαζί με τις αντίστοιχες ώρες εργασίας που απαιτήθηκαν για την εκτέλεση τους.

Μήνας Παραγωγής (i)	Μέγεθος Παραγγελίας (X_i)	Ώρες Εργασίας (Y_i)
1	30	73
2	20	50
3	60	128
4	80	170
5	40	87
6	50	108
7	60	135
8	30	69
9	70	148
10	60	132

i) Να κατασκευαστεί το διάγραμμα διασποράς και να σχεδιάσετε "με το μάτι" μια ευθεία που να δείχνει τη στατιστική σχέση μεταξύ μεγέθους παραγγελίας και ωρών εργασίας

ii) Να υπολογίσετε τους εκτιμητές $\hat{\beta}_0$ και $\hat{\beta}_1$ των ελαχίστων τετραγώνων των παραμέτρων και την εκτιμώμενη συνάρτηση παλινδρόμησης

iii) α. Χρησιμοποιώντας την εκτιμώμενη συνάρτηση παλινδρόμησης ($\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$) να υπολογίσετε για κάθε μια από τις τιμές της ανεξάρτητης μεταβλητής, την εκτιμώμενη τιμή της εξαρτημένης μεταβλητής ($\hat{Y}_i$)

β. Να υπολογιστούν τα υπόλοιπα (ή τις κατακόρυφες αποκλίσεις των παρατηρήσεων Y_i από την ευθεία της επιτιμώμενης συνάρτησης παλινδρόμησης). Τι παρατηρείτε για τα υπόλοιπα που υπολογίσατε;

iv) α. Να ελέγξετε την υπόθεση ότι $\beta_1=0$ με επίπεδο σημαντικότητας $a=0{,}05$.

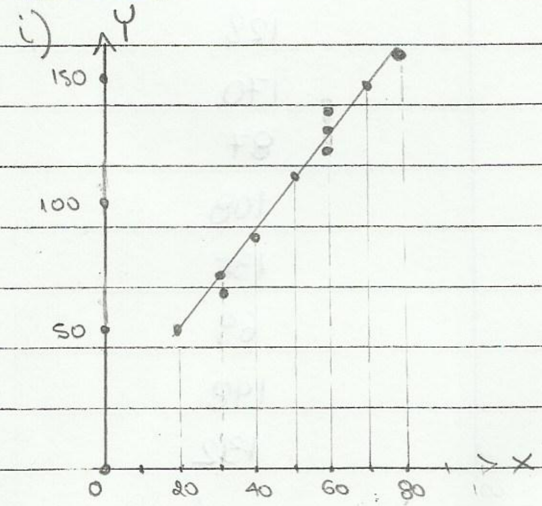
β. Να βρεθεί ένα 95% Διάστημα εμπιστοσύνης για το β_1. Υπάρχει κάποια σχέση μεταξύ των X και Y;

v) Να βρείτε ένα 90% Διάστημα εμπιστοσύνης για τη μέση τιμή του Y_0 όταν $X_0=55$

vi) Να ελέγξετε την υπόθεση $H_0: \rho=0$ σε επίπεδο σημαντικότητας $a=0{,}01$, όπου ρ: ο συντελεστής συσχέτισης

ΛΥΣΗ

i)



ii) Είχαμε αποδείξει για τη συνάρτηση

$$Q(\beta_0,\beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 X_i)^2$$

μέσω μερικής παραγώγισης ότι

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 X_i) \text{ και}$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n}X_i(Y_i-\beta_0-\beta_1 X_i)$$

Δηλαδή,

$$\begin{cases} \sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \cdot \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} X_i Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \end{cases}$$

Και επιλύοντας το σύστημα βρίσκαμε ότι:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \text{και} \quad \hat{\beta}_1 = \frac{\Sigma(x_i-\bar{x})\cdot(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}$$

όπου $\hat{\beta}_0$ και $\hat{\beta}_1$ οι ζητούμενοι εκτιμητές ελαχίστων τετραγώνων (Οι τύποι μπορούν να χρησιμοποιούνται δίχως την παραπάνω απόδειξη)

Με τη χρήση του ακόλουθου πίνακα και του δοσμένου πίνακα εξαρτημένων και ανεξάρτητων τυχαίων μεταβλητών έχαμε τους παρακάτω υπολογισμούς:

$Y_i \cdot X_i$	X_i^2
2190	900
1000	400
7680	3600
13600	6400
3480	1600
5400	2500
8100	3600
2070	900
10360	4900
7920	3600
ΣΥΝΟΛΟ 61800	28400

Συνεπώς, έχουμε:

$$\hat{\beta}_1 = \frac{61800 - \frac{500 \cdot 1100}{10}}{28400 - \frac{500^2}{10}} = 2.0$$

και

$$\hat{\beta}_0 = \frac{1100}{10} - 2.0\,\frac{500}{10} = 10.0$$

Άρα, η εκτιμώμενη συνάρτηση παλινδρόμησης είναι η εξής: $\hat{Y} = 2.0 \cdot X + 10.0$

(Παρατήρηση: $\hat{\beta}_1 = 2.0 > 0$ δηλ. ο συντελεστής διευθύνσεως της ευθείας είναι θετικός και άρα η ευθεία που σχετίζει τα X με τα Y θα έχει θετική κλίση όπως φαίνεται και στο ερώτημα (i))

iii) α. Είναι: $\hat{y} = 2.0\, x_i + 10.0$

για καθεμία τιμή των X_i στον πίνακα (για $i=1,2,\ldots,10$) παίρναμε με αντικατάσταση στην ευθεία τις αντίστοιχες τιμές των $\hat{y}_i$

$\hat{y}_i$	70	50	130	170	90	110	130	70	150	130	$\xrightarrow{\text{ΣΥΝΟΛΟ}}$	$\boxed{1100.}$

β. Εξ ορισμού των υπολοίπων ορίζονται από τον τύπο

$$e_i = Y_i - \hat{Y}_i \quad , i=1,2,\ldots,10$$

Έτσι αντίστοιχα, $\forall i=1,2,\ldots,10$ παίρναμε τα εξής:

e_i	3	0	-2	0	-3	-2	5	-1	-2	2

Αυτό που παρατηρούμε είναι ότι τα υπόλοιπα (ή αποκλίσεις) αν τα αθροίσουμε θα μας δώσουν αποτέλεσμα 0.

Δηλαδή $\sum_{i=1}^{n=10} e_i = 0$.

iv) α. Από τη θεωρία γνωρίζουμε ότι

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}\right)$$

Άρα,

$$B_1 = \frac{\hat{\beta}_1 - \beta_1}{S\Big/\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \underset{\text{Αληθ.}}{\overset{H_0}{\sim}} t_{n-2} \qquad (1)$$

Έχουμε,

$$(\hat{\sigma}^2 =) S^2 = \frac{1}{10-2}\sum_{i=1}^{10}(y_i - 10.0 - 2.0x_i)^2 = \frac{60}{8} = 7.5 \Rightarrow S = \sqrt{7.5}$$

Ενώ,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 3400$$

Επομένως,

$$B_1 = \frac{2.0 - 0}{\sqrt{7.5}\Big/\sqrt{3400}} = \frac{2.0}{0.0497} = 42.6$$

Έπειτα,

$t_{0,025,8} = 2,306 \leftarrow$ (Από πίνακα Student)

και λόγω ότι έχουμε αμφίπλευρη κρίσιμη περιοχή (δηλ. $H_0: \beta_1 = 0$ v $H_a: \beta_1 \neq 0$) όπου $C_{κρ.περ} = (-\infty, -t_{a/2,n-2}) \cup (+t_{a/2,n-2}, +\infty)$

Καθώς, $B_1 = 42,6 > t_{0,025,8} = 2,306$.
(δηλ. $B_1 \in C_{κρ.περ.}$) συνεπώς H_0 απορρίπτεται
Άρα, σε επίπεδο σημαντικότητας $a=0,05$ το $\beta_1 \neq 0$
Παρατήρηση: Υπάρχει γραμμική εξάρτηση του Y από το X

β. Με βάση τα παραπάνω ένα 95% Δ.Ε είναι το

$$\hat{\beta}_1 \pm t_{a/2,n-2} \cdot \frac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \Rightarrow 2.0 \pm 2,306 \cdot \frac{\sqrt{7.5}}{\sqrt{3400}} \Rightarrow$$

$\Rightarrow$ Δ.Ε $[1.89, 2.11]$.

v) Για συγκεκριμένες τιμές τώρα ($\pi\chi$ $X_0=55$) η διαδικασία που εκτελούμε είναι παρόμοια με την διαδικασία του ερωτήματος (vi).

Γνωστό από θεωρία:

$$E(\hat{Y}_0) = E(\hat{\beta}_0+\hat{\beta}_1 X_0) = \hat{\beta}_0+\hat{\beta}_1 X_0$$

και

$$Var(\hat{Y}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(X_0-\bar{X})^2}{\Sigma(X_i-\bar{X})^2}\right)$$

Ενώ, ισχύει:

$$\frac{\hat{Y}_0 - E(\hat{Y}_0)}{\sqrt{\widehat{Var}(\hat{Y}_0)}} \sim t_{n-2} \quad \text{και Δ.Ε} \quad \hat{Y}_0 \pm t_{a/2,n-2}\sqrt{\widehat{Var}(\hat{Y}_0)}$$

Συνεπώς, συγκεκριμένα στην άσκησή μας έχουμε:

$$\hat{Y}_0 = \hat{\beta}_0+\hat{\beta}_1 \cdot X_0 = 10.0+2.0\cdot 55 = 120$$

και $t_{a/2,n-2} = t_{0,05,8} = 1,860$

Ενώ,

$$\widehat{Var}(\hat{Y}_0) = 7,5\left(\frac{1}{10} + \frac{(55-50)^2}{3400}\right) = 0,80515$$

Και άρα, ένα 90% Δ.Ε είναι το ακόλουθο

$$120 \pm 1,860\cdot\sqrt{0,80515} \Rightarrow \text{Δ.Ε το } [118,3,\ 121,7].$$

vi) Από τη θεωρία μας έχουμε:

$$r(x,y) = \frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\Sigma(X_i-\bar{X})^2\,\Sigma(Y_i-\bar{Y})^2}} = \frac{6800}{\sqrt{3400\cdot 13660}} = 0,9978$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \overset{H_0}{\underset{\alpha\lambda\eta\theta.}{\sim}} t_{n-2} \quad (\text{με } H_0: \rho=0 \ v\ H_a: \rho\neq 0)$$

$$t = \frac{0,9978\sqrt{8}}{\sqrt{1-0,9956}} = 42,6$$

, Αφού $t=42,6 > t_{0,005,8} = 3,335$ η $H_0: \rho=0$ Απορρίπτεται για $a=0,01$.

β' τρόπος για το ερώτημα (iv) 1α.

Μέσω του Πίνακα ΑΝΑΔΙΑ (ANOVA).

Προέλευση μεταβλητότητας	Β.Ε	SS	MS
Παλινδρόμηση	1	13600	13600
Υπόλοιπο (ή απόκλιση)	8	60	7.5
ολική μεταβλητότητα	9	13660	

Στατιστική συνάρτηση:

$$F = \frac{MS_{res}}{MS_{res}} = \frac{SS_{res}/1}{SS_{res}/n-2} = \frac{13600}{7.5} = 1813,3$$

σε επίπεδο σφάλματος $\alpha = 0,05$

το $F_{0,05,1,8} = 5.32$

$F > F_{0.05,1,8} \Rightarrow$

$\Rightarrow$ Η $H_0: \beta_1 = 0$ απορρίπτεται όπως δείξαμε και μέσω του t-τεστ

(επίσης $R^2 = \frac{13600}{13660} = 0,9956$ ← Συντελεστής προσδιορισμού

Δηλ. το 99,56% (ποσοστό) της μεταβλητότητας του Y οφείλεται στην παλινδρόμηση).