Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρηματικά Προβλήματα

Φροντιστήριο #5

Παραδείγματα συνέχεια...

(2) Από όλες τις βιοτεχνίες που παράγουν ένα προϊόν, ένα δείγμα $n=8$ βιοτεχνιών έδωσε (δειγματική) παλινδρόμηση της παραγωγής Y σε τόνους πάνω στον αριθμό των εργαζόμενων X την $\hat{Y} = 3.5 + 0.75X$. Επίσης $\sum_{i=1}^{8} x_i^2 = 300$, $\frac{1}{7}\sum_{i=1}^{8}(x_i - \bar{x})^2 = 1.71$, $\frac{1}{7}\sum_{i=1}^{8}(y_i - \bar{y})^2 = 1.71$ (βλ. Παράδειγμα 1).

(i) Να ελεγχθεί σε $\alpha = 5\%$ η $H_0: B_1 = 0$ vs $H_1: B_1 \neq 0$ και να ερμηνευθεί το αποτέλεσμα. Δίνεται $t_{6,0.025} = 2.447$.

(ii) Να ελεγχθεί σε $\alpha = 5\%$ η $H_0: \rho = 0$ vs της κατάλληλης εναλλακτικής της (δίνεται $t_{6,0.05} = 1.94$).

(iii) Να δοθεί ο πίνακας ανάλυσης διακύμανσης της εξαρτημένης Y και να ελεγχθεί η $H_0: B_0 = B_1 = 0$ έναντι της H_1: $\exists$ ένα τουλάχιστον i με $B_i \neq 0, i = 1,2$ σε $\alpha = 5\%$ (δίνεται $\tilde{F}^{u}_{1,6,0.05} = 5.99$).

(iv) Να υπολογισθεί ο R^2.

<u>Λύση</u> (i) Για τον έλεγχο: $H_1: B_1 = 0$ vs $H_1: B_1 \neq 0$

χρησιμοποιούμε τη σ.σ. ελέγχου:

$$T_{n-2} = \frac{|\hat{B}_1 - B_1^*|}{S_{\hat{B}_1}} \overset{H_0}{\sim} t_{n-2}. \quad B_1^* = 0$$

$$T_{n-2} = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$

Όμως $s_{\hat{\beta}_1} = \sqrt{\dfrac{s^2}{\sum_{i=1}^{8}(x_i-\bar{x})^2}}$

με $s^2 = \dfrac{1}{n-2}\left(\sum_{i=1}^{8}(y_i-\bar{y})^2 - \hat{\beta}_1^2 \sum_{i=1}^{8}(x_i-\bar{x})^2\right)$

$(SST) - (SSR)$

$$= \frac{1}{6}\left(7\cdot 1.71 - 0.75^2\cdot 7\cdot 1.71\right)$$

$$= \frac{7}{6}\left(1.71 - 0.75^2\cdot 1.71\right) = 0.873$$

Άρα $s_{\hat{\beta}_1} = \sqrt{\dfrac{0.873}{7\cdot 1.71}} = 0.27$

Άρα $|T_{n-2}| = \dfrac{|\hat{\beta}_1|}{s_{\hat{\beta}_1}} = \dfrac{0.75}{0.27} = 2.78 > t_{6,0.025} = 2.442$

συνεπώς, σε $\alpha = 5\%$ η H_0 απορρίπτεται, άρα τα δεδομένα παρέχουν ισχυρές ενδείξεις ότι ο αριθμός των εργαζόμενων έχει μία γραμμική σχέση με την παραγωγή του προϊόντος.

(ii) Ο δειγματικός συντελεστής συσχέτισης r είναι:

$$r = \hat{\beta}_1 \cdot \frac{\sqrt{\frac{1}{7}\sum_{i=1}^{8}(x_i-\bar{x})^2}}{\sqrt{\frac{1}{7}\sum_{i=1}^{8}(y_i-\bar{y})^2}} = 0.75\sqrt{\frac{1.71}{1.71}} = 0.75$$

Γνωρίζουμε ότι $T_{n-2} = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} \overset{H_0}{\sim} t_{n-2}$

όπου $H_0: \rho = 0$.

Επειδή $r = 0.75 > 0$ ο έλεγχος της $H_0: \rho = 0$ είναι λογικό να είναι μονόπλευρος με $H_1: \rho > 0$.

Έχουμε $T_{n-2} = \dfrac{0.75\sqrt{8-2}}{\sqrt{1-0.75^2}} = 2.78 > t_{8,0.05} = 1.94$

Άρα, σε $\alpha = 5\%$ υπάρχουν ισχυρές ενδείξεις απόρριψης της H_0, συνεπώς φαίνεται ότι υπάρχει μία έντονη θετική γραμμική συσχέτιση μεταξύ των μεταβλητών X και Y.

(iii) Επειδή $\sum_{i=1}^{8}(y_i - \bar{y})^2 = 7 \cdot 1.71 = 12 = SST$

και $\sum_{i=1}^{8}(\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{8}(x_i - \bar{x})^2 = 0.75^2 \cdot 7 \cdot 1.71$
$= 6.75 = SSR$.

Άρα $SSE = SST - SSR = \sum_{i=1}^{8}\hat{\varepsilon}_i^2 = 12 - 6.75 = 5.25$

Κατασκευάζουμε τον πίνακα ανάλυσης διακύμανσης για την εξαρτημένη μεταβλητή Y

Μεταβλητότητα	Άθροισμα τετραγώνων	Β.ε	Μέσο άθροισμα τετραγώνων	Πηλίκο
Παλινδρόμηση	$SSR = 6.75$	1	$MSR = 6.75$	F
Σφάλματα	$SSE = 5.25$	$n-2=6$	$MSE = 0.873$	
Σύνολο	$SST = 12$	$n-1=7$		

όπου $F = \dfrac{MSR}{MSE} = \dfrac{(n-2)R^2}{1-R^2} = \dfrac{6 \cdot 0.75^2}{1-0.75^2}$

διότι $R^2 = r^2 = 0.75^2$ $= 7.71$

δηλ. $F = 7.71 > F^{u}_{1,6,0.05} = 5.99$ απορρίπτουμε σε $\alpha = 5\%$ την H_0 άρα τα δεδομένα παρέχουν ενδείξεις απόρριψης της H_0 συνεπώς η εξίσωση παλινδρόμησης της Y πάνω στην X φαίνεται να είναι στατιστικά σημαντική και ο αριθμός των εργαζόμενων φαίνεται να είναι στατιστικά σημαντικός για την ερμηνεία της μεταβλητής της παραγωγής Y.

(iv) Υπολογισμός R^2

Έχουμε $R^2 = \dfrac{SSR}{SST} = \dfrac{\sum_{i=1}^{8}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{8}(y_i-\bar{y})^2} = \dfrac{6.75}{12} = 0.5625$

ή $R^2 = r^2 = 0.75^2 = 0.5625$

ή από $F = 7.71 = \dfrac{(n-2)R^2}{1-R^2}$ ή

$$R^2 = \hat{\beta}_1^2 \cdot \frac{\sum_{i=1}^{8}(x_i-\bar{x})^2}{\sum_{i=1}^{8}(y_i-\bar{y})^2} = 0.75^2 \frac{1.71}{1.71} = 0.5625.$$

③ Τυχαίο δείγμα $n=8$ βιοτεχνιών που παράγουν ένα προϊόν ερευνήθηκε κατά το προηγούμενο έτος ως προς τις δαπάνες τους X για διαφήμιση και τις εισπράξεις τους Y σε χιλιάδες ευρώ έδωσε:

$$\hat{Y}=3.5+0.75X,\ \bar{X}=6,\ S_X^2=\frac{1}{n-1}\sum_{i=1}^{8}(x_i-\bar{x})^2=1.71$$

$$\text{και } S_Y^2=\frac{1}{n-1}\sum_{i=1}^{8}(y_i-\bar{y})^2=1.71.$$

(i) Αν μία βιοτεχνία προγραμματίζει $X_0=10$ χιλιάδες ευρώ να προβλεφθούν οι εισπράξεις της και να βρεθεί το 0.95 δ.ε για αυτές (δίνεται $t_{6,0.025}=2.447$).

(ii) Αν όλες οι βιοτεχνίες ξοδέψουν για διαφήμιση από 10 χιλιάδες ευρώ να προβλεφθούν οι μέσες εισπράξεις τους και να βρεθεί για αυτές το 0.95 δ.ε.

Λύση

(i) Πρόβλεψη: $\hat{Y}_0=3.5+0.75\cdot 10=11$ χιλ. ευρώ

$s^2=0.873$ (από παράδειγμα ②)

Άρα, το 95% δ.ε. για τις εισπράξεις αν $X_0=10$ είναι

$$\hat{Y}_0 \pm t_{n-2,\alpha/2}\, s\sqrt{1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum_{i=1}^{8}(x_i-\bar{x})^2}}$$

$$=11 \pm 2.447\cdot\sqrt{0.873}\sqrt{1+\frac{1}{8}+\frac{(10-6)^2}{7\cdot 1.71}}$$

$= 11 \pm 3.68 \Rightarrow$ 95% δ.ε. για τη μελλοντική τιμή της Y, $Y_0 \in (7.32, 14.68)$ χιλιάδες ευρώ.

<u>Σημ.</u> με επίπεδο εμπιστοσύνης ίσο με 95% η πραγματική μελλοντική τιμή των εισπράξεων των βιοτεχνιών αν οι δαπάνες τους για διαφήμιση είναι $x_0 = 10$ χιλιάδες ευρώ θα κυμαίνεται (ή αναμένεται να κυμαίνεται) περίπου μεταξύ των τιμών 7.32 και 14.68 χιλιάδων ευρώ.

(ii) Σημειακή πρόβλεψη για τη μελλοντική μέση τιμή: $\widehat{E(Y_0)} = 3.5 + 0.75 \cdot 10 = 11$ χιλ. ευρώ

Επειδή $s^2 = 0.873$ το 95% δ.ε. για την πρόβλεψη των μελλοντικών μέσων εισπράξεων είναι:

$$\widehat{E(Y_0)} \pm t_{n-2,\alpha/2}\, s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{8}(x_i - \bar{x})^2}}$$

$$= 11 \pm 2.447 \cdot \sqrt{0.873} \sqrt{\frac{1}{8} + \frac{(10-6)^2}{7 \cdot 1.71}}$$

$$= 11 \pm 2.76 = (8.24, 13.76)$$

$\Rightarrow$ 95% δ.ε. για τη μελλοντική <u>μέση</u> τιμή της Y, $E(Y_0 | X = x_0)$ θα είναι $E(Y_0 | X = x_0) \in (8.24, 13.76)$

χιλιάδες ευρώ,

δηλαδή, με επίπεδο εμπιστοσύνης ίσο με 95% η πραγματική μελλοντική μέση τιμή των εισπράξεων των βιοτεχνιών (αν οι δαπάνες τους για διαφήμιση είναι $x_0 = 10$ χιλιάδες ευρώ) αναμένεται να κυμαίνεται περίπου μεταξύ των τιμών 8.24 και 13.76 χιλιάδων ευρώ.

(4) Από τυχαίο δείγμα $n = 100$ ζευγών της τιμής X ενός προϊόντος σε ευρώ και της προσφοράς του Y σε τόνους από τον παραγωγό υπολογίστηκαν:

$$\sum_{i=1}^{100} x_i = 36, \quad \sum_{i=1}^{100} x_i^2 = 110, \quad \sum_{i=1}^{100} y_i = 25, \quad \sum_{i=1}^{100} y_i^2 = 366,$$

$$\sum_{i=1}^{100} x_i y_i = 30.$$

(i) Να εκτιμηθεί η παλινδρόμηση της Y πάνω στην X και να ερμηνευθούν οι συντελεστές $\hat{\beta}_0$ και $\hat{\beta}_1$ της γραμμικής παλινδρόμησης.

(ii) Να ελεγχθεί σε $\alpha = 5\%$ αν υπάρχει στατιστικά σημαντική γραμμική σχέση (συσχέτιση) μεταξύ της τιμής X και της προσφοράς Y του προϊόντος. (δίνεται $z_{0.025} = 1.96$).

Λύση (i) Έχουμε

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{100} x_i y_i - \left(\sum_{i=1}^{100} x_i\right)\left(\sum_{i=1}^{100} y_i\right)}{n\sum_{i=1}^{100} x_i^2 - \left(\sum_{i=1}^{100} x_i\right)^2}$$

$$= \frac{100 \cdot 30 - 36 \cdot 25}{100 \cdot 110 - 36^2} = 0.22$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{25}{100} - 0.27 \frac{36}{100} = 0{,}171$$

Άρα η εκτιμώμενη ευθεία παλινδρόμησης είναι: $\hat{Y} = 0.171 + 0.22 X$.

$\underline{\hat{\beta}_1 = 0.22}$ Αν η τιμή του προϊόντος αυξηθεί κατά 1 ευρώ η προσφορά του αναμένεται να αυξηθεί κατά (περίπου) 0.22 τόνους.

$\underline{\hat{\beta}_0 = 0.171}$ Αν η τιμή του προϊόντος είναι μηδενική η προσφορά του αναμένεται να είναι περίπου 0.171 τόνους (χωρίς ιδιαίτερη οικονομική ερμηνεία).

(ii) Ελέγχουμε $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ σε $\alpha = 5\%$

$$\frac{1}{n-1}\left(\sum_{i=1}^{100} x_i^2 - n\bar{x}^2\right) = 0.98 = S_X^2$$

$$S_Y^2 = \frac{1}{n-1}\left(\sum_{i=1}^{100} y_i^2 - n\bar{y}^2\right) = 3.63 \quad (n = 100)$$

$$S^2 = \frac{n-1}{n-2}\left(S_Y^2 - \hat{\beta}_1^2 S_X^2\right) = \frac{99}{98}\left(3.63 - 0.22^2 \cdot 0.98\right)$$

$$= 3.62$$

$$S_{\hat{\beta}_1}^2 = \frac{S^2}{\left(\sum_{i=1}^{100} x_i^2 - n\bar{x}^2\right)} = \frac{3.62}{110 - 100 \cdot 0.36^2} = 0.037.$$

Επειδή $n>30$, μπορούμε να θεωρήσουμε ότι

$$T_{n-2} \simeq Z = \frac{|\hat{\beta}_1|}{S_{\hat{\beta}_1}} \quad \text{όπου } Z \text{ τμ. } \sim N(0,1)$$

προσεγγιστικά καθώς $n \to \infty$

λόγω του ΚΟΘ.

Όμως $Z = \dfrac{0.22}{0.037} = 1.14 < z_{0.025} = 1.96$

άρα, τα δεδομένα παρέχουν ισχυρές ενδείξεις μη-απόρριψης της H_0 σε $\alpha = 5\%$. Συνεπώς, <u>δεν</u> φαίνεται να υπάρχει <u>στατιστικά</u> <u>σημαντική</u> <u>γραμμική</u> <u>σχέση</u> μεταξύ της τιμής X και της προσφοράς Y του προϊόντος.

ξ