Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρηματικά Προβλήματα

Φροντιστήριο #4

Έλεγχος στατιστικής σημαντικότητας του συντελεστή παλινδρόμησης β_1

Θέλουμε να ελέγξουμε σε ε.σ.σ. α την

$H_0: \beta_1 = \beta_1^*$ δηλ. αν ο συντελεστής παλινδρόμησης του πληθυσμού ισούται με την τιμή β_1^*

vs

$H_1: \beta_1 \neq \beta_1^*$ δηλ. ο συντελεστής παλινδρόμησης του πληθυσμού είναι σημαντικά διάφορος της τιμής β_1^*

Για τη διενέργεια του ελέγχου, χρησιμοποιούμε τη σ.σ. ελέγχου

$$T_{n-2} = \frac{|\hat{\beta}_1 - \beta_1^*|}{S_{\hat{\beta}_1}}$$

, όπου $S_{\hat{\beta}_1}$ είναι το τυπικό σφάλμα εκτίμησης της παραμέτρου β_1.

όπου $$S_{\hat{\beta}_1} = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

, όπου s είναι το τυπικό σφάλμα εκτίμησης των σφαλμάτων.

$$s = \sqrt{\frac{SSE}{n-2}}$$

Αν $|T_{n-2}| > t_{n-2,\alpha/2}$ τότε σε εσσ α απορρίπτουμε την H_0, όπου $t_{n-2,\alpha/2}$ είναι το

$\alpha/2$ - ποσοστιαίο σημείο της κατανομής t_{n-2} δηλ. της κατανομής t με $n-2$ β.ε.

Με παρόμοιο τρόπο μπορούμε να διενεργήσουμε και τους μονόπλευρους ελέγχους υποθέσεων δηλ τους: $H_1: B_1 > B_1^*$ και $H_1: B_1 < B_1^*$. Τότε, αντίστοιχα:

Απορρίπτουμε την H_0 σε εσσ α αν

$$T_{n-2} > t_{n-2,\alpha} \text{ και } T_{n-2} < -t_{n-2,\alpha}.$$

<u>Παρατήρηση:</u> Τονίζουμε ότι στην απλή γραμμική παλινδρόμηση, οι έλεγχοι υποθέσεων για τον συντελεστή γραμμικής συσχέτισης ρ, για τον συντελεστή προσδιορισμού R^2 και για τον συντελεστή παλινδρόμησης B_1 θεωρούνται ισοδύναμοι. Έχει γενικά, επικρατήσει ο έλεγχος υποθέσεων για την παράμετρο B_1.

<u>Διάστημα Εμπιστοσύνης για την παράμετρο B_1</u>

Μπορούμε να κατασκευάσουμε ένα $100(1-\alpha)\%$ δ.ε. για την παράμετρο B_1. Το διάστημα αυτό είναι:

$$\hat{B}_1 - S_{\hat{B}_1} t_{n-2,\alpha/2} < B_1 < \hat{B}_1 + S_{\hat{B}_1} t_{n-2,\alpha/2}$$

δηλ. $B_1 \in \left(\hat{B}_1 - S_{\hat{B}_1} t_{n-2,\alpha/2},\ \hat{B}_1 + S_{\hat{B}_1} t_{n-2,\alpha/2}\right)$

Εν συντομία, μπορούμε να γράψουμε:

δ.ε. για το β_1: $\boxed{\hat{\beta}_1 \pm s_{\hat{\beta}_1} \cdot t_{n-2,\alpha/2}}$

Προβλέψεις

Η αναμενόμενη μέση τιμή της μεταβλητής απόκρισης Y για $X = x_0$, δίνεται από:

$$E(Y_0) = \beta_0 + \beta_1 x_0$$

Προκειμένου να δημιουργήσουμε ένα δ.ε. για αυτήν χρησιμοποιούμε την εκτιμήτρια $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ η οποία ακολουθεί την κανονική κατανομή και για την οποία ισχύει:

$$E(\hat{Y}_0) = \beta_0 + \beta_1 x_0 \text{ και}$$

$$\operatorname{Var}(\hat{Y}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right), \text{ όπου}$$

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

Εισάγοντας στη έκφραση για τη $\operatorname{var}(\hat{Y}_0)$ την αμερόληπτη s^2 του σ^2 προκύπτει η ποσότητα:

$$s^2(\hat{Y}_0) = s^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$$

και μπορεί να δειχθεί ότι: $\dfrac{\hat{Y}_0 - (\beta_0 + \beta_1 x_0)}{s(\hat{Y}_0)} \sim t_{n-2}$.

Άρα, ένα δ.ε. με συντελεστή εμπιστοσύνης $1-\alpha$ για τη <u>μέση</u> πρόβλεψη $E(Y_0)=\beta_0+\beta_1 x_0$ για την τιμή $X=x_0$ της ανεξάρτητης μεταβλητής είναι το διάστημα

$$\left(\hat{Y}_0 - s(\hat{Y}_0)\, t_{n-2,\alpha/2},\ \hat{Y}_0 + s(\hat{Y}_0)\, t_{n-2,\alpha/2}\right) \quad (1)$$

όπου $s(\hat{Y}_0)=\sqrt{s^2(\hat{Y}_0)}$.

Αν μας ενδιαφέρει ένα διάστημα το οποίο να περιέχει τις τιμές που μπορεί να λάβει η μεταβλητή απόκρισης Y όταν $X=x_0$, τότε χρησιμοποιούμε το διάστημα πρόβλεψης (prediction interval). Αυτό το διάστημα με συντελεστή εμπιστοσύνης $1-\alpha$ $(0<\alpha<1)$ δίνεται ως εξής:

$$\left(\hat{Y}_0 - s(\tilde{Y}_0)\, t_{n-2,\alpha/2},\ \hat{Y}_0 + s(\tilde{Y}_0)\, t_{n-2,\alpha/2}\right) \quad (2)$$

όπου $s(\tilde{Y}_0)$ είναι η τετραγωνική ρίζα της παράστασης

$$s^2(\tilde{Y}_0) = s^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right).$$

Συγκριτικά, το δ.ε. για τη πρόβλεψη της μέσης τιμής (1) της τ.μ. Y έχει μεγαλύτερη πρακτική αξία από το διάστημα πρόβλεψης (2).

Ελαστικότητες

Ορίζουμε ως ελαστικότητα της Y ως προς τη X το λόγο μεταξύ της ποσοστιαίας μεταβολής της Y ως προς τη ποσοστιαία μεταβολή της X.

$$\eta_{Y|X} = \frac{dY}{dX} \cdot \frac{X}{Y} = \hat{\beta}_1 \cdot \frac{X}{Y}$$

Είναι ένας καθαρός αριθμός που δίνει τη ποσοστιαία μεταβολή της Y που αντιστοιχεί σε μία μοναδιαία ποσοστιαία μεταβολή της X (κατά 1%, 1‰, 10% κ.ο.κ.)

Π.χ. αν $\eta_{Y|X} = -1.5$ σημαίνει ότι αν η X αυξηθεί κατά 1%, η Y θα μειωθεί κατά 1.5% (αν η X αυξηθεί κατά 10% η Y θα μειωθεί κατά 15%).

Αντίστοιχα, ορίζεται η μέση ελαστικότητα:

$$\bar{\eta}_{Y|X} = \hat{\beta}_1 \cdot \frac{\bar{X}}{\bar{Y}}$$

Π.χ. αν $\hat{Y} = 615,28 + 46,92X$

με Y := πωλήσεις (χιλ. €)

X := # πωλητών $\quad \bar{\eta}_{Y|X} = \hat{\beta}_1 \frac{\bar{X}}{\bar{Y}} = 0,827.$

και αν $\hat{Y} = 677,4 + 3,11X$

με Y := πωλήσεις (χιλ. €), X := διαφήμιση (χιλ. €)

$$\bar{n}_{Y|X} = \hat{\beta}_1 \frac{\bar{X}}{\bar{Y}} = 0,809$$

<u>Συμπέρασμα</u> Τόσο ο αριθμός των πωλητών όσο και η διαφημιστική δαπάνη προκαλούν σε σχετικούς όρους την ίδια (περίπου) επίδραση στις πωλήσεις.

<u>Παραδείγματα</u>

(1) Στην άσκηση η του <u>Φρ. #3</u> να ελέγξετε αν η σχέση αποταμίευσης/εισοδημάτων είναι θετική.

<u>Λύση</u> Ελέγχουμε σε εσσ $\alpha = 5\%$ αν

$H_0: \beta_1 = 0$
vs
$H_1: \beta_1 > 0$

Χρησιμοποιούμε την ελεγχοσυνάρτηση:

$$T_{n-2} = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} \overset{H_0}{\sim} t_{n-2}$$

όπου $S_{\hat{\beta}_1} = \sqrt{\dfrac{S^2}{S_{xx}}}$

όμως

$$S^2 = 2,6625 \text{ και } S_{\hat{\beta}_1} = \sqrt{\frac{2,6625}{1000}} = 0,0515$$

Άρα $T_{n-2} = \dfrac{0,14}{0,0515} = 2,71844$

Απορρίπτουμε την Ho σε εσσ $\alpha = 5\%$ αν

$T_{n-2} > t_{n-2,\alpha} = t_{8,0.05} = 1,8595$ (από πίνακες της t_8)

Άρα, η H_0 απορρίπτεται σε εσσ $\alpha = 5\%$

συνεπώς, η παράμετρος β_1 είναι στατιστικά σημαντική για το μοντέλο και η X επηρεάζει στατιστικά σημαντικά την Y. Στη προκειμένη περίπτωση φαίνεται ότι τα ετήσια εισοδήματα ασκούν στατιστικά σημαντική θετική επίδραση στις ετήσιες αποταμιεύσεις.

② Δίνονται στοιχεία του αριθμού εργαζομένων X σε μία βιοτεχνία και της παραγωγής Y του προϊόντος της σε τόνους σε ένα τ.δ. $n=8$ εβδομάδων λειτουργίας της.

Εργαζόμενοι X:	5	6	7	5	7	4	6	8
Παραγωγή Y:	8	7	8	6	9	7	9	10

(i) Να εκτιμηθεί η ευθεία παλινδρόμησης για το γραμμικό υπόδειγμα $Y = \beta_0 + \beta_1 X + \varepsilon$ της παραγωγής Y επάνω στην εργασία X και να ερμηνευθούν οι συντελεστές.

(ii) Να επιβεβαιωθεί ότι η ευθεία παλινδρόμησης διέρχεται από το σημείο $(\bar{x}, \bar{y})$.

(iii) Αν για μία εβδομάδα οι εργαζόμενοι αυξηθούν κατά 3 έναντι της προηγούμενης εβδομάδας, πόσο θα αυξηθεί η παραγωγή;

(iv) Να υπολογισθεί και να ερμηνευτεί ο συντελεστής προσδιορισμού R^2.

(v) Να υπολογισθεί και να ερμηνευτεί ο συντελεστής γραμμικής συσχέτισης r.

(vi) Να υπολογισθεί η ελαστικότητα της παραγωγής ως προς την εργασία για $X_0=10$ εργαζομένους και η μέση ελαστικότητα της παραγωγής ως προς την εργασία και να ερμηνευθούν.

<u>Λύση</u> Υπολογίζουμε: $\sum_{i=1}^{8} x_i = 48$, $\bar{x}=6$

$\sum_{i=1}^{8} x_i^2 = 300$, $\frac{1}{7}\sum_{i=1}^{8}(x_i-\bar{x})^2 = 1.71$

$\sum_{i=1}^{8} y_i = 64$, $\bar{y}=8$, $\frac{1}{7}\sum_{i=1}^{8}(y_i-\bar{y})^2 = 1.71$

$\sum_{i=1}^{8} x_i y_i = 393$.

(i) $$\hat{\beta}_1 = \frac{n\sum_{i=1}^{8} x_i y_i - \sum_{i=1}^{8} x_i \sum_{i=1}^{8} y_i}{n\sum_{i=1}^{8} x_i^2 - \left(\sum_{i=1}^{8} x_i\right)^2} = \frac{3\cdot 393 - 48\cdot 64}{8\cdot 300 - 48^2}$$

$$= \frac{72}{96} = 0{,}75 \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 8 - 0{,}75\cdot 6 = 3.5.$$

Άρα, η εκτιμώμενη ευθεία παλινδρόμησης της Y πάνω στη X είναι:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X = 3.5 + 0.75X$$

Ο συντελεστής $\hat{\beta}_1 = 0.75$ σημαίνει ότι αύξηση των εργαζομένων κατά 1, αυξάνει την εβδομαδιαία παραγωγή κατά 0.75 τόνους (οριακή παραγωγικότητα της εργασίας). Ο συντελεστής $\hat{\beta}_0 = 3.5$ που θεωρητικά σημαίνει ότι με 0 εργαζόμενους η παραγωγή αναμένεται να είναι περίπου 3.5 τόνοι δεν φαίνεται να έχει οικονομική ερμηνεία.

(ii) Θέτουμε στην $\hat{Y} = 3.5 + 0.75X$ όπου X το $\bar{X} = 6$ και έχουμε ότι $\hat{Y} = 8 = \bar{Y}$. Άρα, το ζεύγος $(\bar{x}, \bar{y})$ επαληθεύει την εκτιμώμενη ευθεία παλινδρόμησης.

(iii) Αφού με κάθε επιπλέον εργαζόμενο η παραγωγή αυξάνει κατά 0.75 τόνους με τρεις επιπλέον εργαζόμενους αναμένεται να αυξηθεί κατά $3 \cdot 0.75 = 2.25$ τόνους.

(iv) Έχουμε $R^2 = \dfrac{SSR}{SST} = \dfrac{\hat{\beta}_1^2 \cdot \frac{1}{7} \sum_{i=1}^{8} (x_i - \bar{x})^2}{\frac{1}{7} \sum_{i=1}^{8} (y_i - \bar{y})^2}$

$$= 0{,}75^2 \cdot \frac{1{,}71}{1{,}71} = 0{,}5625$$

Ερμηνεία: Η Y "εξαρτάται" από την X κατά 56.25% και κατά το υπόλοιπο 43.7% από άλλους παράγοντες που δε λάβαμε υπόψη δηλ. η X φαίνεται ότι ερμηνεύει μόνο το 56.25% της συνολικής μεταβλητότητας της Y.

(v) $r = \sqrt{R^2} = \sqrt{0.5625} = 0.75$

άρα επιτηρείται σχετικά έντονη θετική γραμμική συσχέτιση των X και Y.

(vi) $n_{Y|X} = \hat{\beta}_1 \frac{X}{\hat{\beta}_0 + \hat{\beta}_1 X} \overset{X_0=10}{=} 0.75 \frac{10}{3.5+0.75 \cdot 10} = 0.68$

Στο επίπεδο των 10 εργαζομένων αύξηση τους κατά 1% αυξάνει την παραγωγή κατά 0.68%. Συνεπώς, η παραγωγή είναι ανελαστική ως προς την εργασία.

Η μέση ελαστικότητα της παραγωγής ως προς την εργασία είναι

$n_{\bar{Y}|\bar{X}} = \hat{\beta}_1 \frac{\bar{X}}{\bar{Y}} = 0.75 \cdot \frac{6}{8} = 0.5625$ άρα αύξηση του αριθμού εργαζομένων κατά 1% θα αυξήσει την παραγωγή κατά μέσο όρο κατά 0.5625%.