Ασκήσεις στο Μάθημα Στατιστική 2 - Απλή Γραμμική Παλινδρόμηση

Α. Λαδάς (*a_ladas@upatras.gr*)

Πανεπιστήμιο Πατρών
Τμήμα Οικονομικών Επιστημών

24/5/2022

Άσκηση 1

Εταιρεία πετρελαιοειδών έκανε το εξής πείραμα για να βελτιώσει την ποιότητα των καυσίμων της. Χρησιμοποίησε 10 αυτοκίνητα ίδιων χαρακτηριστικών (μάρκα, τύπος, κυβικά, κ.λ.π.), στα οποία έβαλε βενζίνη με κάποια ποσότητα βελτιωτικής ουσίας σε 7 διαφορετικά επίπεδα και μέτρησε τη μείωση των εκπεμπόμενων καυσαερίων.
Προέκυψαν τα αποτελέσματα, όπως φάινονται στον Πίνακα 1.

Συνέχεια

Επίπεδα Βελτιωτικής Ουσίας Χ	Ποσοστό Μείωσης Καυσαερίων Υ
1	2.1
1	2.5
2	3.1
3	3
4	3.8
4	3.2
5	4.3
6	3.9
6	4.4
7	4.8

Πίνακας 1: Δεδομένα για τις μεταβλητές Χ και Υ.

Συνέχεια

Να εκτιμηθούν τα εξής:

1. Η ευθεία γραμμικής παλινδρόμησης
2. Η εκτίμηση της διακύμανσης των καταλοίπων
3. Τα 95% διαστήματα εμπιστοσύνης για τους συντελεστές παλινδρόμησης
4. Είναι στατιστικά σημαντικοί οι συντελεστές παλινδρόμησης, σε επίπεδο σημαντικότητας 5% ;
5. Να υπολογιστεί και να ερμηνευτεί ο συντελεστής προσδιορισμού.

Λύση Άσκησης 1

1. Γνωρίζουμε ότι η ευθεία παλινδρόμησης είναι της μορφής:

Μορφή Ευθείας Παλινδρόμησης

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i + \hat{u}_i \tag{1}$$

με

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} \tag{2}$$

και

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \tag{3}$$

όπου

$S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$

$S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2$

$S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$

Συνέχεια

Όπως φαίνεται και από τους παραπάνω τύπους, χρειάζομαι να υπολογίσω το μέσο του X και το μέσο του Υ, χρησιμοποιώντας τους γνωστούς τύπους:

$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1+...+7}{10} = 3.9$

$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{2.1+...+4.8}{10} = 3.51$

Χρησιμοποιώ τους παραπάνω τύπους και υπολογίζω τα S_{XY}, S_{XX} και S_{YY}.

$S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) =$
$(1 - 3.9)(2.1 - 3.51) + ... + (7 - 3.9)(4.8 - 3.51) = 15.81$

$S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2 = (1 - 3.9)^2 + ... + (7 - 3.9)^2 = 40.9$

$S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = (2.1 - 3.51)^2 + ... + (4.8 - 3.51)^2 = 6.849$

Συνέχεια

Οπότε, από τον τύπο 2 υπολογίζω
$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{15.81}{40.9} = 0.3865$
και από τον τύπο 3
$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 3.51 - 0.3865 * 3.9 = 2.0024$
Άρα, η ευθεία παλινδρόμησης είναι της μορφής:
$\hat{Y}_i = 2.0024 + 0.3865X$
2. Η εκτίμηση της διακύμανσης των καταλοίπων δίνεται από τον τύπο:

Εκτίμηση Διακύμανσης Καταλοίπων

$$S^{*2}_{Y/X} = \frac{1}{n-2}\left(S_{YY} - \frac{S^2_{XY}}{S_{XX}}\right) \tag{4}$$

Συνέχεια

Από τον τύπο 4 μπορούμε να υπολογίσουμε την εκτίμηση της διακύμανσης των καταλοίπων, ως εξής:

$$S^{*2}_{Y/X} = \frac{1}{n-2}\left(S_{YY} - \frac{S^2_{XY}}{S_{XX}}\right) = \frac{1}{10-2}\left(6.85 - \frac{15.81^2}{40.9}\right) = 0.0922$$

Άρα το τυπικό σφάλμα των καταλοίπων είναι

$S_{Y/X} = \sqrt{S^{*2}_{Y/X}} = \sqrt{0.0922} = 0.3036$. Οι τιμές της εκτίμησης της διακύμανσης των καταλοίπων και του τυπικού σφάλματος των καταλοίπων είναι μικρές (κοντά στο μηδέν) και άρα η προσαρμογή της ευθείας παλινδρόμησης φαίνεται να είναι αρκετά καλή.

Συνέχεια

3. Ο τύπος υπολογισμού του (1-α)% διαστήματος εμπιστοσύνης για τους συντελεστές παλινδρόμησης, είναι:

Διάστημα Εμπιστοσύνης για το α

$$\hat{\alpha} \pm t_{n-2,1-\alpha/2}\sqrt{S^{*2}_{Y/X}\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}}\right)} \tag{5}$$

Διάστημα Εμπιστοσύνης για το β

$$\hat{\beta} \pm t_{n-2,1-\alpha/2}\sqrt{\frac{S^{*2}_{Y/X}}{S_{XX}}} \tag{6}$$

Συνέχεια

Άρα το διάστημα εμπιστοσύνης για το α είναι το εξής:

$$\left[\hat{\alpha} - t_{n-2,1-\alpha/2}\sqrt{S^{*2}_{Y/X}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right)},\right.$$

$$\left.\hat{\alpha} + t_{n-2,1-\alpha/2}\sqrt{S^{*2}_{Y/X}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right)}\right] =$$

$$\left[2.0024 - 2.306\sqrt{0.0922\left(\frac{1}{10} + \frac{3.9^2}{40.9}\right)},\right.$$

$$\left.2.0024 + 2.306\sqrt{0.0922\left(\frac{1}{10} + \frac{3.9^2}{40.9}\right)}\right] = [1.52, 2.48]$$

ΠΡΟΣΟΧΗ!! Τα άκρα των δ.ε. είναι και τα δυο θετικά, συνεπώς τα δ.ε. δεν περιέχουν το μηδέν.

Συνέχεια

Το διάστημα εμπιστοσύνης για το β είναι το εξής:

$$\left[\hat{\beta} - t_{n-2,1-\alpha/2}\sqrt{\frac{S^{*2}_{Y/X}}{S_{XX}}}, \hat{\beta} + t_{n-2,1-\alpha/2}\sqrt{\frac{S^{*2}_{Y/X}}{S_{XX}}}\right] =$$

$$\left[0.3865 - 2.3\sqrt{\frac{0.0922}{40.9}}, 0.3865 + 2.3\sqrt{\frac{0.0922}{40.9}}\right] = [0.277, 0.496]$$

ΠΡΟΣΟΧΗ!! Τα άκρα των δ.ε. είναι και τα δυο θετικά, συνεπώς τα δ.ε. δεν περιέχουν το μηδέν.

Συνέχεια

4. Για να εξετάσουμε εάν οι συντελεστές της παλινδρόμησης είναι στατιστικά σημαντικοί, θα πρέπει να ελέγξουμε τις εξής υποθέσεις:

1.Μηδενική και εναλλακτική υπόθεση του δίπλευρου ελέγχου στατιστικής σημαντικότητας των συντελεστών της παλινδρόμησης

- $H_0 : \alpha = 0$
- $H_1 : \alpha \neq 0$

και

- $H_0 : \beta = 0$
- $H_1 : \beta \neq 0$

Συνέχεια

2.Συνάρτηση του ελέγχου στατιστικής σημαντικότητας των συντελεστών της παλινδρόμησης

Για το α

$$TS = \frac{\hat{\alpha} - \alpha}{\sqrt{S^{*2}_{Y/X}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right)}} \tag{7}$$

Για το β

$$TS = \frac{\hat{\beta} - \beta}{\sqrt{\frac{S^{*2}_{Y/X}}{S_{XX}}}} \tag{8}$$

Κάνοντας χρήση του τύπου (7) υπολογίζουμε την τιμή της ελεγχοσυνάρτησης

$$TS = \frac{\hat{\alpha}-\alpha}{\sqrt{S^{*2}_{Y/X}(\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}})}} = \frac{2.0024}{\sqrt{0.0922(\frac{1}{10}+\frac{3.9^2}{40.9})}} = 9.57.$$

Συνέχεια

Γνωρίζουμε ότι $\alpha = 0.05$, άρα μπορούμε να υπολογίσουμε ότι $1 - \alpha/2 = 0.975$.

Από τις πληροφορίες που έχω διαθέσιμες, μπορώ να χρησιμοποιήσω τον εξής κανόνα απόρριψης:

3.Κανόνας απόρριψης του δίπλευρου ελέγχου στατιστικής σημαντικότητας των συντελεστών της παλινδρόμησης

$$TS > t_{n-2,1-\alpha/2} \tag{9}$$

ή

$$TS < -t_{n-2,1-\alpha/2} \tag{10}$$

$\Leftrightarrow$

$$|TS| > t_{n-2,1-\alpha/2} \tag{11}$$

Συνέχεια

Χρησιμοποιώντας τον κανόνα απόρριψης (11), συγκρίνουμε την τιμή που υπολογίσαμε από τον τύπο (7) με την κρίσιμη τιμή της $t - student$ κατανομής $t_{n-2,1-\alpha/2} = t_{8,1-0.05/2} = 2.306$. Μπορούμε να διαπιστώσουμε ότι $|TS| > t_{n-2,1-\alpha/2}$ και άρα μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι το α ισούται με το μηδέν. Άρα ο συντελεστής α είναι στατιστικά σημαντικός, σε επίπεδο σημαντικότητας 5%.

ΠΡΟΣΟΧΗ!! Το αποτέλεσμα που προέκυψε από τον έλεγχο υποθέσεων συμφωνεί με το συμπέρασμα που έχει προκύψει από το δ.ε. για το α.
Ο έλεγχος για το β αφήνεται για εξάσκηση.

Συνέχεια

5. Για τον υπολογισμό του συντελεστή προσδιορισμού R^2, κάνουμε χρήση του τύπου (12)

Τύπος υπολογισμού Συντελεστή Προσδιορισμού R^2

$$R^2 = \frac{S_{XY}^2}{S_{XX} S_{YY}} \tag{12}$$

Έχουμε διαδοχικά: $R^2 = \frac{S_{XY}^2}{S_{XX} S_{YY}} = \frac{15.81^2}{40.9*6.849} = 0.8923.$
Η μεταβλητικότητα του επιπέδου της βελτιώτικής ουσίας ερμηνεύει περίπου το 89.23% της μεταβλητικότητας του ποσοστού της μείωσης καυσαερίων. Το υπόλοιπο 10.77% οφείλεται σε τυχαίους παράγοντες (παραμένει ανερμήνευτο). Παρατηρούμε, επίσης, ότι η τιμή του συντελεστή προσδιορισμού βρίσκεται αρκετά κοντά στη μονάδα, άρα η ευθεία παλινδρόμησης προσαρμόζεται αρκετά καλά στα δεδομένα.