

Στην περίπτωση που δεν απορριψαμε την μηδενική υπόθεση μπορούμε να πούμε ότι τα δεδομένα προέρχονται από τον ίδιο αβιβασ πληθυσμό και μπορούν να θεωρηθούν ως μια ομάδα.

Ζήτημα Β

$n=60$, X_1, X_2, X_3, X_4, X_5

B1) Για την μεταβλητή X_2

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} = \frac{-33.831}{5.943} = -6.45$$

$$P\text{-Value} = (\text{εντολές στην R}) = 2 * pt(q = -6.45, df = 54)$$

$$\Rightarrow P\text{-value} \approx 3.21 \cdot 10^{-8} < 0.001$$

$$\begin{array}{l} n-k-1 \\ 60-5-1 \end{array}$$

Για την μεταβλητή X_4

$$t = \frac{-3.390}{3.541} = -0.95 \sim t_{54} \leadsto P\text{-value} = 0.347$$

$$2 * pt(q = -0.95, df = 54)$$

Για την μεταβλητή X_5

$$t = -0.04 \sim t_{54} \leadsto P\text{-value} = 0.97$$

$$R\text{-sq} = \frac{SSR}{SST} = \frac{944930}{1049009} = 0.9097 \approx 91\%$$

$$SSE = SST - SSR \Rightarrow SSE = 94079$$

$$\bar{R}^2 = 1 - \frac{n-1}{n-p} \frac{SSE}{SST} = 1 - \frac{59}{59} \frac{94079}{104909} =$$

$$\Rightarrow \bar{R}^2 = 0.903 \approx R^2 = 90,1\%$$

Παρατηρούμε ότι η Τιμή του σημειακού εκτιμητή του σταθερού όρου είναι 99.43 με τυπικό σφάλμα 19.68 ενώ το στατιστικό έλεγχο T για τον έλεγχο $H_0: \alpha=0$ έναντι της $H_1: \alpha \neq 0$ έδωσε Τιμή 5.05. Η P -Τιμή του εν λόγω ελέγχου είναι < 0.001 οπότε έχουμε πολύ σοβαρές ενδείξεις για να απορρίψουμε την μηδενική υπόθεση ότι $\alpha=0$.

Η τιμή του σημειακού εκτιμητή του συντελεστή b_5 του X_5 είναι περίπου -0.00116 με τυπικό σφάλμα 0.031 ενώ το στατιστικό έλεγχο για τον έλεγχο $H_0: b_5=0$ έναντι της $H_1: b_5 \neq 0$ έδωσε -0.04. Η P -Τιμή του εν λόγω ελέγχου είναι 0.94, οπότε σε επίπεδο σημαντικότητας 5% δεν έχουμε σοβαρές ενδείξεις για να απορρίψουμε την μηδενική υπόθεση ότι $b_5=0$.

Επίσης παρατηρούμε ότι ο συντελεστής προσδιορισμού και ο προσαρμοσμένος συντελεστής προσδιορισμού είναι κοντά στην μονάδα οπότε μπορούμε να θεωρήσουμε ότι η προσαρμογή του μοντέλου είναι ικανοποιητική.

Τέλος, λόγω του παραγινά-πληθυσμίου διακύμανσης (VIF), για τις μεταβλητές X_3 και X_5 φαίνεται να υπάρχει πρόβλημα πολυσυγχρονικότητας.

• B9)

i) Ο διορθωμένος συντελεστής προσδιορίσματος ορίζεται ως:

$$\bar{R}^2 = 1 - \frac{MSE}{MST} = 1 - \frac{n-1}{n-p} \frac{SSE}{SST}$$

και προτιμάται έναντι του συνθηματικού συντελεστή προσδιορίσματος R^2 διότι έχει την ιδιότητα να μην αυξάνεται η τιμή του όταν προστίθεται στο μοντέλο μια οποιαδήποτε επεξηγηματική μεταβλητή αλλά μόνο όταν αυτή βελτιώνει το μοντέλο.

ii) Αρχικά ελέγχουμε ποια μοντέλα είναι κατάλληλα με βάση τις τιμές C_p -Mallows θέλουμε $C_p \approx p$

Παρατηρούμε ότι τα καλύτερα μοντέλα που ικανοποιούν την συνθήκη είναι:

M_5	$X_1 + X_2 + X_3$	$C_p = 3,4$	$p = 3$
M_4	$X_1 + X_2 + X_3 + X_4$	$C_p = 4$	$p = 4$

Επειτα παρατηρούμε ότι και στα δύο αυτά εξετασμένα μοντέλα η τιμή του συντελεστή προσδιορίσματος και του διορθωμένου συντελεστή προσδιορίσματος είναι αρκετά κοντά στη μονάδα ενώ οι τιμές των κριτηρίων AIC είναι οι μικρότερες από όλα τα μοντέλα.

$$iii) F = \frac{SSE_5 - SSE_4}{SSE_4/(n-p)} = \frac{96431 - 94081}{94081} = 1,36 \rightarrow p\text{-value} = 0,246$$

Συνεπώς επιλέγω το μοντέλο M_5

B3)

α) Για την μεταβλητή X_1

$$t = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{-1.7598}{0.9659} = -2.024$$

$$P\text{-value} = (2 * pt(q = -2.024, df = 56)) = 0.48$$

Αρχικά παρατηρούμε ότι όλες οι p-τιμές των ελέγχων είναι υπέρ της απορρίψης της μηδενικής υπόθεσης. Επειτα, το VIF για όλες τις μεταβλητές είναι < 5 άρα δεν έχουμε ενδείξεις πολυσυγγρημίας. Παρόλο που ο συντελεστής προσδιορισμού μειώθηκε ο διορθωμένος αυξήθηκε άρα έχουμε υποψίες ότι ορθά αιωρεθόταν οι μεταβλητές X_4 και X_5 .

β) i) Παρατηρούμε στο παρακάτω σχήμα ότι οι τιμές των παρατηρήσεων 6, 15, 59, 60 υπερβαίνουν την κρίσιμη τιμή 0.13 αλλά για όλες τις παρατηρήσεις οι τιμές των αποστάσεων Cook είναι κάτω από μικρότερες των μονάδων εκτός των παρατηρήσεων 15 που πιθανόν να είναι σημείο επιρροής και θα πρέπει να αιωρεθεί.

ii) Στα διαγράμματα των μερικών υπολοίπων, παρατηρούμε ότι τα σημεία των γραφικών σχηματίζουν μια αρκετά καλά ορισμένη ευθεία άρα οδηγούμαστε στο συμπέρασμα ότι και οι 3 μεταβλητές ~~και~~ μας είναι χρήσιμες στο μόντελ.