

Ανάλυση και Ερμηνεία των Αποτελεσμάτων του Υποδείγματος Πολλαπλής Παλινδρόμησης

Το υπόδειγμα που εκτιμήθηκε είναι μια πολλαπλή γραμμική παλινδρόμηση, όπου η **εξαρτημένη μεταβλητή** είναι η **κατανάλωση καυσίμου σε λίτρα ανά 100 χιλιόμετρα (lt_100km)** και οι **ερμηνευτικές μεταβλητές** είναι οι **hp, disp, wt** και **qsec**.

Η εξίσωση που εκτιμάται είναι:

$$lt_100km_i = \beta_0 + \beta_1 hp_i + \beta_2 disp_i + \beta_3 wt_i + \beta_4 qsec_i + u_i$$

Από τα αποτελέσματα προκύπτει ότι ο σταθερός όρος εκτιμάται ίσος με 6.3178. Αυτός εκφράζει τη θεωρητική τιμή της κατανάλωσης όταν όλες οι ερμηνευτικές μεταβλητές είναι μηδέν, αν και στην πράξη αυτή η ερμηνεία συνήθως δεν έχει ουσιαστικό οικονομικό νόημα.

Ο συντελεστής της μεταβλητής **hp** είναι 0.008574. Αυτό σημαίνει ότι, κρατώντας τις υπόλοιπες μεταβλητές σταθερές, **αύξηση της ιπποδύναμης κατά 1 μονάδα συνδέεται με αύξηση της κατανάλωσης κατά περίπου 0.0086 λίτρα ανά 100 χιλιόμετρα**. Όμως το **p-value** είναι 0.36854, άρα η επίδραση αυτή δεν είναι στατιστικά σημαντική στα συνηθισμένα επίπεδα σημαντικότητας.

Ο συντελεστής της **disp** είναι 0.004271. Επομένως, για **κάθε επιπλέον μονάδα κυβισμού, η κατανάλωση αυξάνεται κατά περίπου 0.0043 λίτρα ανά 100 χιλιόμετρα**, με τις άλλες μεταβλητές σταθερές. Και εδώ όμως το **p-value** είναι 0.51339, συνεπώς η μεταβλητή δεν είναι στατιστικά σημαντική.

Ο συντελεστής της **wt** είναι 2.564876 και είναι στατιστικά σημαντικός, αφού το **p-value** είναι 0.00225. Αυτό σημαίνει ότι, **αν το βάρος αυξηθεί κατά 1 μονάδα, η κατανάλωση αυξάνεται κατά περίπου 2.56 λίτρα ανά 100 χιλιόμετρα**, κρατώντας τις άλλες μεταβλητές σταθερές. **Άρα το βάρος φαίνεται να έχει ουσιαστική και ισχυρή θετική επίδραση στην κατανάλωση**.

Ο συντελεστής της **qsec** είναι -0.227340. Αυτό δείχνει ότι **αύξηση του qsec κατά 1 μονάδα συνδέεται με μείωση της κατανάλωσης κατά περίπου 0.227 λίτρα ανά 100 χιλιόμετρα**, υπό την προϋπόθεση ότι οι υπόλοιπες μεταβλητές παραμένουν σταθερές. Ωστόσο το **p-value** είναι 0.42416, επομένως και αυτή η μεταβλητή δεν είναι στατιστικά σημαντική.

Συνολικά, μόνο η μεταβλητή **wt** φαίνεται να επηρεάζει στατιστικά σημαντικά την κατανάλωση στο συγκεκριμένο υπόδειγμα.

Όσον αφορά τη **συνολική προσαρμογή**, ο συντελεστής προσδιορισμού $R^2 = 0.8553$ δείχνει ότι το υπόδειγμα εξηγεί περίπου το 85.5% της διακύμανσης της κατανάλωσης. Ο διορθωμένος συντελεστής προσδιορισμού $Adjusted R^2 = 0.8339$ παραμένει επίσης πολύ υψηλός, κάτι που σημαίνει ότι το υπόδειγμα έχει καλή ερμηνευτική ικανότητα ακόμη και **αφού ληφθεί υπόψη ο αριθμός των ερμηνευτικών μεταβλητών**.

Το F-statistic είναι 39.9 με p-value 5.816×10^{-11} , **άρα το υπόδειγμα συνολικά είναι στατιστικά σημαντικό**. Δηλαδή, οι ερμηνευτικές μεταβλητές από κοινού συμβάλλουν σημαντικά στην εξήγηση της εξαρτημένης μεταβλητής.

Τέλος, το residual standard error είναι 1.575, που εκφράζει το τυπικό μέγεθος του σφάλματος πρόβλεψης του υποδείγματος. Όσο μικρότερο είναι αυτό, τόσο πιο κοντά βρίσκονται συνήθως οι προβλεπόμενες τιμές στις πραγματικές.

Συμπέρασμα: Το υπόδειγμα πολλαπλής παλινδρόμησης έχει πολύ καλή συνολική προσαρμογή, αφού εξηγεί περίπου το 85.5% της διακύμανσης της κατανάλωσης. Από τις τέσσερις ερμηνευτικές μεταβλητές, μόνο το βάρος (**wt**) εμφανίζεται στατιστικά σημαντικό και με θετική επίδραση στην κατανάλωση. Οι μεταβλητές **hp**, **disp** και **qsec** δεν είναι στατιστικά σημαντικές ατομικά, αν και το υπόδειγμα συνολικά είναι ισχυρά στατιστικά σημαντικό.

Έλεγχος Γραμμικών Περιορισμών

Χρησιμοποιούμε F test για να εξετάσουμε αν οι μεταβλητές **disp** και **qsec**, που προστίθενται στο πλήρες υπόδειγμα, είναι από κοινού στατιστικά σημαντικές. Η μηδενική υπόθεση είναι ότι οι αντίστοιχοι συντελεστές είναι ταυτόχρονα ίσοι με το μηδέν. Αν το **p-value** είναι μικρό, τότε το πλήρες υπόδειγμα υπερέχει στατιστικά του περιορισμένου. Αν είναι μεγάλο, τότε οι πρόσθετες μεταβλητές δεν βελτιώνουν σημαντικά την προσαρμογή και προτιμάται το απλούστερο υπόδειγμα.

- **restricted model:** $lt_100km \sim hp + wt$
- **unrestricted model:** $lt_100km \sim hp + disp + wt + qsec$

$$H_0: \beta_{disp} = \beta_{qsec} = 0$$

δηλαδή ότι οι δύο επιπλέον μεταβλητές **δεν προσθέτουν ερμηνευτική δύναμη**. Το F test βασίζεται στη σύγκριση των αθροισμάτων τετραγώνων καταλοίπων:

$$F = \frac{(RSS_r - RSS_u)/J}{RSS_u/(n - k)} \sim F(J, n - k)$$

όπου:

- RSS_r : άθροισμα τετραγώνων καταλοίπων του restricted model
- RSS_u : άθροισμα τετραγώνων καταλοίπων του unrestricted model
- J : αριθμός περιορισμών
- n : μέγεθος δείγματος
- k : αριθμός παραμέτρων του μεγάλου υποδείγματος

άρα

- οι περιορισμοί είναι 2
- άρα $J = 2$, επειδή ελέγχουμε μαζί τις disp και qsec