

ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ ΤΩΝ ΣΦΑΛΜΑΤΩΝ

Επιτυγχάνεται με τη βοήθεια των υπολοίπων

1. Γραφικοί τρόποι ελέγχου

Normal Probability plot, Histogram των τυποποιημένων υπολοίπων (Linear regression plots)

2. Γραφικοί και στατιστικοί τρόποι ελέγχου

Αποθηκεύουμε τα τυποποιημένα υπόλοιπα και έπειτα μέσω της διαδικασίας Explore έχουμε P-P plot, Q-Q plot, test Shapiro-Wilk.

ΔΙΟΡΘΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

- Μετασχηματισμός των τιμών της εξαρτημένης μεταβλητής έτσι ώστε να επιτευχθεί η κανονικότητα. Η μορφή του ιστογράμματος των υπολοίπων ίσως μας υποδεικνύει ποιος μετασχηματισμός είναι κατάλληλος.
- Μετασχηματισμός των Box και Cox (1964).

Συνηθισμένοι μετασχηματισμοί

- Αντίστροφος μετασχηματισμός

- Λογαριθμικός

Με βάση το 10, το 2 και το e . Ο πρώτος κυρίως όταν το εύρος είναι αρκετά μεγάλο.

- Ρίζα

Ιδιαίτερη προσοχή όταν υπάρχουν αρνητικές τιμές!!!

Χρησιμοποιείται όταν τα δεδομένα ακολουθούν Poisson κατανομή.

- Τα ποσοστά ίσως κανονικοποιούνται με τον \arcsine μετασχηματισμό.

Box-Cox μετασχηματισμός (Rawlings (1988) p.310-313)

- Χρησιμοποιείται η παραμετρική οικογένεια μετασχηματισμών

$$Y_i^{(\lambda)} = \begin{cases} \frac{Y_i^\lambda - 1}{\lambda \left(\bar{Y} \right)^{\lambda-1}}, & \lambda \neq 0 \\ \left(\bar{Y} \right) \ln(Y_i), \dots, \lambda = 0 \end{cases}$$

όπου

$$\left(\bar{Y} \right) = \exp \left(\sum_{i=1}^n \ln(Y_i) / n \right)$$

Box Cox Method

- Υποθέτει ότι για κάποια τιμή της παραμέτρου λ τα μετασχηματισμένα δεδομένα ικανοποιούν την υπόθεση της κανονικότητας. Υπό αυτή την υπόθεση οι Ε.Μ.Π. των παραμέτρων προσδιορίζονται.
- Για διάφορες επιλογές του λ υπολογίζουμε το $SS_{res}(\lambda)$ για τα μετασχηματισμένα δεδομένα.

Box-Cox

- Η εύρεση του Ε.Μ.Π. του λ ισοδυναμεί με την εύρεση του ελάχιστου ως προς λ της ποσότητας $SS_{res}(\lambda)$.
- Το SPSS δυστυχώς δεν μας υλοποιεί αυτόματα αυτό το μετασχηματισμό.
- Είναι διαθέσιμο μέσω της ιστοσελίδας του Tao Li ένα Syntax File το οποίο και θα χρησιμοποιούμε όπου χρειάζεται.
- <http://www.stat.tamu.edu/ftp/pub/mspeed/Syntax/>

ΣΥΝΕΠΕΙΕΣ

- Λάθος διαστήματα εμπιστοσύνης και μη σωστοί έλεγχοι υποθέσεων για τις παραμέτρους του μοντέλου.
- Οι εκτιμητές ελαχίστων τετραγώνων δεν είναι Α.Ο.Ε.Δ.
- Αν το μέγεθος του δείγματος είναι μεγάλο (λόγω του Κεντρικού Οριακού Θεωρήματος) χρησιμοποιούμε την κανονικότητα των σφαλμάτων προσεγγιστικά, με τη διαφοροποίηση ότι οι κρίσιμες πιθανότητες (οι p -τιμές δηλαδή) είναι προσεγγιστικές και όχι ακριβείς.

Παρατήρηση

- Κατά κανόνα ο μετασχηματισμός για να πετύχουμε την κανονικότητα συνήθως καταστρέφει την γραμμικότητα του μοντέλου ή την ύπαρξη σταθερής διακύμανσης. Με την προϋπόθεση ότι το μέγεθος του δείγματος είναι μεγάλο η υπόθεση της κανονικότητας δεν είναι τόσο απαραίτητη.