5/12/16 παλινδ. ①

Αναπλήρωση
Παρασκευή 9/12/16
12:00

Μοντέλα ανάλυσης και διακύμανσης

↓

Γραμμικά Μοντέλα

Μοντέλα παλινδρόμησης

1. Ποσοτικές μεταβλητές

2. προσπαθούμε να κατασκ. τη γραμμική σχέση που συνδέει την Y με τις $X_1, ..., X_p$, να ελέγξουμε την εγκυρότητα της σχέσης και να χρησιμοποιήσουμε αυτή την για προβλέψεις

Μοντέλα ανάλυσης διακύμανσης

1. Θεωρούμε μια εξαρτημένη ποσοτική μ. Y και μια ή περισσότερες κατα κανόνα ποιοτικές μεταβ.

2. Στα μοντέλα ανάλυσης διακύμανσης δεν ενδιαφερόμαστε τόσο για την κατασκευή κάποιας συγκεκριμένης σχέσης, αλλά ενδιαφερόμαστε στο να εντοπίσουμε ποία ή ποιες από τις κατηγορίες των ποιοτικών μεταβλητών ασκούν τη σημαντικότερη επίδραση στην Y.

→ Ορολογία και συμβολισμός

Παράγοντας → Ταυτόσημη με την ανεξάρτητη μεταβ.

Κάθε τιμή του παράγοντα (που είναι μια κατηγορία αφού ο παράγοντας είναι ποιοτική μεταβλητή) ονομάζεται επίπεδο του παράγοντα

Προβλήματα στα οποία υπεισέρχεται ένας παράγοντας ονομάζ
προβλήματα ανάλυσης διακυμ. κατα ένα παράγοντα

Προβλήματα στα οποίοι υπεισέρχονται δύο παράγοντες ονομ
προβλήματα ανάλυσης διακυμ. κατα δύο παράγοντες

Εαν έχουμε περισσότερους → Πειραματικούς Σχεδιασμούς

Θεραπεία → Συνδιασμό επιπέδων των παραγόντων

Παράδειγμα:

Ενδιαφέρει η επίδοση Y των μαθητών (ποσοτική)

Ενδιαφέρεται να διερευνήσει πως το Y επηρεάζεται από

το Επίπεδο Μόρφωσης (ΕΜΠ)

Το ΕΜΠ είναι ποιοτική μεταβλητή η οποία κατηγοριοποιείται

ως εξής
1. απόφοιτος Υποχρεωτικής (ΑΥ)
2. Λυκείου (Λ)
3. Πανεπιστημίου (Π)
4. Μεταπτυχιακού (Μ)
5. Διδακτορικού (Δ)

- Πρόβλημα ανάλυσης διακύμανσης κατά ένα παράγοντα

παράγοντας ΕΜΠ, Επίπεδα παράγοντα: ΑΥ, Λ, Π, Μ, Δ → Y

Αν είναι ΕΜΜ - Επίπεδο μόρφωσης Μητέρας τότε πάμε σε κατά δύο

παράγοντες

$1^{ος}$ παρ. ΕΜΠ

$2^{ος}$ παρ ΕΜΜ, Επίπεδα παράγοντα: ΑΥ, Λ, Π, Μ, Δ → Y

<u>Δοκιμασία</u>: όλοι οι συνδυασμοί των 2 επιπέδων παραγόντων.

π.χ (ΑΥ, ΑΥ), (ΑΥ, Δ) κ.ο.κ

- Αν έχω πρόβλημα αναλ. διακύμανσης κατά ένα παράγοντα δοκιμασία = επίπεδο

I) Ανάλυση διακύμανσης κατά ένα παράγοντα.

Θεωρώ μια ποσοτική Y και ένα παράγοντα ο οποίος απαντάται

σε I επίπεδα

όπου Y_{ij} η j παρατήρηση στο i-επίπεδο του παράγοντα

Μ.Ο του παρ

<u>Μορφή Δεδομένων</u>

Άθροισμα παρατηρήσεων

Παράγοντας:	Δεδομένα για Y	Σύνολα	Μ.Ο
1 επίπεδο	$Y_{11}, Y_{12}, \ldots, Y_{1J_1}$	$Y_{1\cdot}$	$\bar{Y}_{1\cdot}$
⋮		⋮	⋮
i επίπεδο	$Y_{i1}, Y_{i2}, \ldots, Y_{iJ_i}$	$Y_{i\cdot}$	$\bar{Y}_{i\cdot}$
⋮			⋮
I επίπεδο	$Y_{I1}, Y_{I2}, \ldots, Y_{IJ_I}$	$Y_{I\cdot}$	$\bar{Y}_{I\cdot}$

②

όπου $Y_{i\cdot} = \sum_{j=1}^{J_i} Y_{ij}$

$$\bar{Y}_{i\cdot} = \frac{1}{J_i}\sum_{j=1}^{J_i} Y_{ij}$$

Γενικός δειγματικός μέσος

$$\bar{Y}_{\cdot\cdot} = \frac{1}{N^o}\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}, \quad N = J_1+J_2+\dots+J_I$$

Από το πινακάκι, στα δεδομένα για την Y κάθε γραμμή είναι δεδομένα από τον πληθυσμό το αντίστοιχου επιπέδου

ας πούμε για 1ο επιπ= απόφοιτος Υποχρ. είναι τα δεδομένα από τον πληθυσμό, απόφοιτοι υποχρεωτικής και σίγουρα υπάρχει μια κατανομή που τα περιγράφει $N(\mu_1, \sigma_1^2)$

για το 2ο επίπεδο είναι τα δεδομένα από τον πληθυσμό απόφοιτοι λυκείου, και σίγουρα υπάρχει μια κατανομή που τα περιγράφει $N(\mu_2, \sigma_2^2)$

⋮

για το Ι-στο επίπεδο, δεδομένα από τον πληθυσμό του Ι-επιπέδου $N(\mu_I, \sigma_I^2)$

Μοντέλο ανάλυσης διασποράς κατά ένα παράγοντα

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad i=1,\dots,I \quad j=1,\dots,J_i$$

Εκτιμητές Ελαχίστων Τετραγώνων

$$S = \sum_{i=1}^{I}\sum_{j=1}^{J_i} \varepsilon_{ij}^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij}-\mu_i)^2$$

$$\frac{\partial S}{\partial \mu_i} = 0 \Rightarrow \dots \Rightarrow \text{Ε.Ε.Τ} \quad \boxed{\hat{\mu}_i = \bar{Y}_{i\cdot},\ i=1,\dots,I}$$

Ιδιότητες: Υποθέσεις για σφάλματα

1. $E(\varepsilon_{ij})=0$
2. $Var(\varepsilon_{ij})=\sigma^2$
3. $Cov(\varepsilon_{ij}, \varepsilon_{kl})=0, k\neq l$
4. $\varepsilon_{ij} \sim N(0,\sigma^2)$

- $E(\hat{\mu}_i)=\mu_i, \ i=1,..,I$
- $Var(\hat{\mu}_i)=\frac{\sigma^2}{J_i}, \ i=1,..,I$

$$E(\hat{\mu}_i) = E(\bar{Y}_i) = E\left(\frac{1}{J_i}\sum_{j=1}^{J_i} Y_{ij}\right) = \frac{1}{J_i}\sum_{j=1}^{J_i} E(Y_{ij}) = \frac{1}{J_i}\sum_{j=1}^{J_i}\mu_i = \frac{1}{J_i} J_i\mu_i=\mu_i$$ (αμερόληπτος

ανάλογα και το 2ο

Ισοδύναμο μοντέλο (μαθηματικά $\mu_i=\mu+a_i$)

$$Y_{ij}=\mu+a_i+\varepsilon_{ij} \quad i=1,..,I, \quad j=1,..,J_i$$

Εκτιμητές Ελαχίστων Τετραγώνων

$$S=\sum_{i=1}^{I}\sum_{j=1}^{J_i}\varepsilon_i^2=\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\mu-a_i)^2$$ πρώτη εξίσωση

$$\left.\begin{array}{l}\frac{\partial S}{\partial \mu}=0\\ \frac{\partial S}{\partial a_i}=0\end{array}\right\} \Rightarrow \left.\begin{array}{l}\sum_{i=1}^{I} a_i J_i+\mu\sum_{i=1}^{I} J_i=\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}\\ a_i J_i+\mu J_i=\sum_{j=1}^{J_i} Y_{ij}, \ i=1,..,I\end{array}\right\}$$ κανονικές εξισώσεις

Το σύστημα κανονικών εξισώσεων δεν οδηγεί σε μοναδική λύση, αφού αν αθροίσω την 2η εξίσωση προκύπτει η 1η ως προς a_i (γραμμ. εξαρτημένα)

(3)

Για να πετύχω μοναδική λύση θεωρώ <u>πλευρική συνθήκη</u>.

Το ερώτημα είναι ποιά θα είναι η <u>πλευρική συνθήκη</u>?

Έχουν προταθεί διάφορες. Απο τις διάφορες καλύτερη φυσική ερμηνεία έχει εκείνη που οδηγεί σε αποδεκτούς διαισθητικά εκτιμητές.

π.χ διαισθητικά αποδεκτός εκτιμήτης για το μ είναι $\hat{\mu} = \bar{Y}_{..}$ (αφού το μ σχετίζεται με όλες τις παρατηρήσεις δηλ διαισθητικά το δέχομαι)

γενικό δειγματικός μέσος

Για να πετύχω ως εκτιμητή του γενικ. δειγμ. μέσο αρκεί να υποθέσω ότι $\sum_{i=1}^{I} a_i J_i = 0$ (απο την πρώτη εξίσωση)

<u>Οπότε</u> <u>θεωρώ πλευρική συνθήκη</u>: $\sum_{i=1}^{I} a_i J_i = 0$

Υπό αυτή τη πλευρική συνθήκη οι ΕΣΤ είναι $\hat{\mu} = \bar{Y}_{..}$

$\hat{a}_i = \bar{Y}_{i\cdot} - \bar{Y}_{..}$

• <u>Πίνακας Ανάλ. του μοντέλου Ανάλυσης διακύμανσης κατά ένα παράγοντα</u>

○ ολική μεταβλητότητα = Δειγματική διακύμανση χωρίς τα μεγέθη δείγματος

○ Δειγματική διακύμανση $= \frac{1}{N-1} \boxed{\sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{..})^2}$

χωρίς μεγέθη δείγματος

Θεώρημα: $\underbrace{\sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{..})^2}_{SS_{tot}} = \underbrace{\sum_{i=1}^{I} J_i (\bar{Y}_{i\cdot} - \bar{Y}_{..})}_{SS_{treat}} + \underbrace{\sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{i\cdot})^2}_{SS_{res}}$

<u>Απόδ</u>: $\sum\sum (Y_{ij} - \bar{Y}_{..} + \bar{Y}_{i\cdot} - \bar{Y}_{i\cdot})^2 = \dots$ πράξεις

και αυτό διότι →

(Ποιο άθροισμα οφείλεται στα υπόλοιπα και ποιο στο μοντέλο)

<u>Υπόλοιπα</u>

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\hat{\mu} + \hat{a}_i) = Y_{ij} - (\bar{Y}_{..} + \bar{Y}_{i.} - \bar{Y}_{..})$$

$$= Y_{ij} - \bar{Y}_{j.}$$

~~οπότε ισο~~

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ

Πηγή μεταβλητότητας	SS	Β.Ε	MS	F-πηλίκο
Μοντέλο	$SS_{\cancel{reg}\,tr} = \sum_{i=1}^{I} J_i (\bar{Y}_{i.} - \bar{Y}_{..})$	$I-1$	$MS_{tr} = \frac{SS_{tr}}{I-1}$	$F = \frac{MS_{tr}}{MS_{res}}$
Υπόλοιπα	$SS_{res} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{i.})^2$	$N-I$	$MS_{res} = \frac{SS_{res}}{N-I}$	
Ολική	$SS_{tot} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{..})^2$	$N-1$		