19/11/2019

ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

Μοντέλα για τη διερεύνηση και μελέτη της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών

ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

α) Η εξαρτημένη και οι ανεξάρτητες μεταβλητές είναι ποσοτικές. (βάρος, ύψος κλπ)

β) Αναζητείται, κατασκευάζεται και ελέγχεται η ορθότητα της γραμμικής σχέσης που συνδέει την εξαρτημένη με τις ανεξάρτητες μεταβλητές. Η γραμμική σχέση που κατασκευάζεται χρησιμοποιείται για την πρόβλεψη της εξαρτημένης μεταβλητής για δεδομένες τιμές των ανεξάρτητων μεταβλητών.

ΜΟΝΤΕΛΑ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ

α) Η εξαρτημένη μεταβλητή είναι ποσοτική και οι ανεξάρτητες μεταβλητές είναι, κατά κανόνα, ποιοτικές. (φύλο, θρήσκευμα, κλπ)

β) Δεν κατασκευάζεται κάποια γραμμική σχέση μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών αλλά το ενδιαφέρον εστιάζεται στην ανάλυση των επιδράσεων μιας ή περισσότερων ανεξάρτητων μεταβλητών στην εξαρτημένη μεταβλητή. <u>Αναζητούνται οι τιμές των ανεξάρτητων μεταβλητών που ασκούν τη σημαντικότερη επίδραση στην εξαρτημένη μεταβλητή</u>.

$\boxed{\text{πχ}}$ Y = επίδοση

X ← δείκτης νοημοσύνης — χαμηλό / — φυσιολογικό / — υψηλό

$\cancel{Y = 5 + 3X + \varepsilon}$

Διάλεξη 7η 19/11/2019

Παλινδρόμηση και Ανάλυση Διακύμανσης.

Μοντέλα ανάλυσης διακύμανσης. (κοιτώ το φυλλάδιο που έδωσε)

Γραμμικά Μοντέλα

Μοντέλα Παλινδρόμησης   Μοντέλα Ανάλυση Διακύμανσης.

...

ΟΡΟΛΟΓΙΑ - ΣΥΜΒΟΛΙΣΜΟΙ.

Παράγοντας: Χρησιμοποιείται για να περιγράψει την ανεξάρτητη μεταβλητή (ποιοτική).

Επίπεδο του παράγοντα: Ενοούμε τιμή του παράγοντα
πχ Αν ο παράγοντας είναι ο δείκτης νοημοσύνης (δ.ν) τότε τα επίπεδά του είναι χαμηλός, φυσιολογικός, υψηλός.

Μοντέλα Ανάλυσης Διακύμανσης στα οποία υπεισέρχεται ένας παράγοντας ονομάζονται μοντέλα Αναλ. Διακ. κατά ένα παράγοντα.

Μοντέλα Αναλ Διακ στα οποία υπεισέρχονται δυο παράγοντες ονομάζονται μοντέλα Αναλ. Διακ κατά δύο παράγοντες.
« «

Ανάλογα κατά τρεις Παράγοντες.

Αν υπάρχουν περισσότεροι από τρεις → Πειραματικούς Συνδιασμούς (Experimental Designs)

Δοκιμασία: Κάθε συνδυασμός επιπέδων των παραγόντων του μοντέλου.

Παράδειγμα: Έστω η ποσοτική μεταβλητή
Y= Επίδοση Μαθητή.

Επίπεδα — Υποχ. Εκπ / Λύκειο / Παν/μιο / Μεταπτ. / Διδακτορι

$Y \xleftarrow[\text{σχετίζεται}]{\text{Εξαρτάται}}$ Επίπεδο Μόρφωσης Πατέρα.

↑

Ποιοτική Μεταβλητή.

↑

Παράγοντας

ΠΡΟΣΟΧΗ Εδώ δεν έχω Δοκιμασία.

Πρόβλημα Ανάλυσης Διακύμανσης κατά ένα παράγοντα.

$Y \xleftarrow[\text{σχετίζεται}]{\text{Εξαρτάται}}$
- Επίπεδο Μορφ Πατέρα — ΥΕ, Λ, Π, Μ, Δ
- Επίπεδο Μορφ Μητέρας — ΥΕ, Λ, Π, Μ, Δ

} Πρόβλημα Αναλ Διακ Κατά δυο παράγοντες

ΠΟΤΕ ΤΟΝ ΙΔΙΟ ΠΑΡΑΓΟΝΤΑ.

(Λ,Μ)
(Μ,Π) κλπ
} Δοκιμασία.

Ανάλυση Διακύμανσης κατά ένα παράγοντα

Έστω ποσοτική εξαρτ. μεταβλητή Y
και έστω ένας παράγοντας (ανεξ μτβλ) η οποία απαντάται σε I επίπεδα.

Μορφή Δεδομένων.

Παράγοντα / Επίπεδα παράγοντα	Εξαρτ. μτβ Y.	Σύνολα $Y_{i\cdot}$	Μέσος όρος $\bar{Y}_{i\cdot}$
1	$Y_{11}\ Y_{12} \dots Y_{1J_1}$	$Y_{1\cdot}$	$\bar{Y}_{1\cdot}$
2	$Y_{21}\ Y_{22} \dots Y_{2J_2}$	$Y_{2\cdot}$	$\bar{Y}_{2\cdot}$
⋮	⋮	⋮	⋮
I	$Y_{I1}\ Y_{I2} \dots Y_{IJ_I}$	$Y_{I\cdot}$	$\bar{Y}_{I\cdot}$

$Y_{11}\ Y_{12} \dots Y_{1J_1}$ → μ_1 $\quad Y_{1j} \simeq \mu_1$ ή $Y_{1j} = \mu_1 + \varepsilon_{1j}$

κλπ

- Y_{ij} ← η παρατήρηση της εξαρτ. μεταβ Y που έχει συλλεγεί στην J-μονάδα του i-επιπέδου.

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

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

- $Y_{\cdot\cdot} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}$ γενικό άθροισμα

- $\bar{Y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}$ γενικός μέσος

$= \frac{1}{N} Y_{\cdot\cdot}$ όπου $N = \sum_{i=1}^{I} J_i$ για $i=1,\dots,I$
κ $j=1,\dots,J_i$

ΜΟΝΤΕΛΟ ΑΝΑΛΥΣΗΣ ΔΙΑΚ. κατά ένα παράγοντα.

σφάλματα

$Y_{ij} = \mu_i + \varepsilon_{ij}$ $\quad i=1,2,\dots,I$

$j=1,2,\dots,J_i$

επίδραση του i-επιπέδου του παράγοντα στην Y

Εκτίμηση παραμέτρων

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

$\frac{\partial S}{\partial \mu_i} = 0 \Rightarrow \dots \Rightarrow \hat{\mu}_i = \bar{Y}_{i\cdot}$, $i=1,\dots,I$

Υποθέσεις Σφαλμάτων

1. $E(\varepsilon_{ij}) = 0$
2. $Var(\varepsilon_{ij}) = \sigma^2$
3. $Cov(\varepsilon_{ij}, \varepsilon_{kl}) = 0$

$\Rightarrow E(\hat{\mu}_i) = \mu_i$

$Var(\hat{\mu}_i) = \frac{\sigma^2}{J_i}$

$$E(\hat{\mu_i}) = E(\overline{Y_{i\cdot}}) = E\left(\frac{1}{J_i}\sum_{j=1}^{J_i} Y_{ij}\right)$$

$$= \frac{1}{J_i}\sum_{j=1}^{J_i} E(Y_{ij})$$

$$= \frac{1}{J_i}\sum_{j=1}^{J_i} \mu_i = \frac{1}{J_i} J_i \mu_i = \mu_i$$

Ανάλογα βγαίνει

$$Var(\hat{\mu_i}) = \frac{\sigma^2}{J_i}$$

Ισοδύναμο μοντέλο Αναλ Διακ κατά ένα παράγοντα.

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

(Αν $\mu_i = \mu + \alpha_i$ αλγεβρικά είναι ισοδύναμα)

Το μοντέλο $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ αποδέχεται και διατυπώνει ότι υπάρχει μια κοινή επίδραση μ όλων των επιπέδων στην Y και μια ατομική επίδραση α_i, $i=1,\dots,I$ του κάθε επιπέδου στην Y.

$$Y_{ij} = \underbrace{\mu}_{\text{Κοινή επίδραση επιπέδων στην Y}} + \overbrace{\alpha_i}^{\text{ατομική επίδραση}} + \underbrace{\varepsilon_{ij}}_{\text{Σφάλματα}} \qquad i=1,\dots,I, \quad j=1,\dots,J_i$$

Εκτίμηση Παραμέτρων

EET των παραμέτρων μ και α_i

Προκύπτουν από ελαχ. ως προς μ και α_i της $S = \sum_{i=1}^{I}\sum_{j=1}^{J_i} \varepsilon_{ij}^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - \mu - \alpha_i)^2$

$$\left.\frac{\partial S}{\partial \mu} = 0 \quad \frac{\partial S}{\partial \alpha_i} = 0 \right\} \text{Κανονικές Εξισώσεις}$$

$$\boxed{\begin{array}{ll}\displaystyle\sum_{i=1}^{I} a_i J_i + \mu \sum_{i=1}^{I} J_i = \sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij} & \\[2ex] \displaystyle a_i J_i + \mu J_i = \sum_{j=1}^{J_i} Y_{ij} \quad i=1,\dots,I & \end{array}}$$

$1^{η}$ εξίσωση

$2^{η}$ εξίσωση

Η $1^{η}$ εξίσωση προκύπτει από την άθροιση των I-υπόλοιπων εξισώσεων.
Άρα το σύστημα δεν έχει μοναδική λύση.

Για να πετύχουμε μοναδική λύση καταφεύγουμε στην τεχνική των πλευρικών συνθηκών.
Υιοθετούμε δηλ. κάποιες συνθήκες (περιορισμούς) που οδηγούν σε διαισθητικά αποδεκτούς εκτιμητές των παραμέτρων.
Ένας διαισθητικά αποδεκτός εκτιμητής για το μ είναι το $\bar{Y}_{..}$.
Για να οδηγηθούμε σε εκτιμητή του μ το $\bar{Y}_{..}$ από την $1^{η}$ εξίσωση αρκεί $\sum_{i=1}^{I} a_i J_i = 0$

Αν υιοθετηθεί η πλευρική συνθήκη

$$\sum_{i=1}^{I} a_i J_i = 0 \quad \text{από } 1^{η} \text{ εξίσωση}$$

$$\mu = \frac{1}{\sum J_i} \sum_i \sum_j Y_{ij} = \bar{Y}_{..}$$

Άρα ο ΕΕΤ του μ είναι $\hat{\mu} = \bar{Y}_{..}$

Από τις υπόλοιπες I εξισώσεις αν θέσω όπου μ το $\hat{\mu}$ και λύσω ως προς a_i προκύπτει $a_i = \bar{Y}_{i.} - \bar{Y}_{..}$

Άρα ο ΕΕΤ των a_i είναι $\hat{a}_i = \bar{Y}_{i.} - \bar{Y}_{..}$.

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑΣ του Μοντέλου Ανάλυσης.

Διακύμανση κατά παράγοντες

Ολική Μεταβλ $\longrightarrow$ Δειγματική διακύμανση των δεδομένων Y_{ij}

$$SS_{tot} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{..})^2$$

ΠΡΟΤΑΣΗ

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

SS_{tr} (treatments) $\qquad SS_{res}$

Αποδ

$$\sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{..})^2 = \ldots\ldots$$

$\pm \bar{Y}_{i\cdot}$

Αποδ $\sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{i.})(\bar{Y}_{i.} - \bar{Y}_{..}) = 0$

Υπόλοιπα: $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\hat{\mu} + \hat{\alpha}_i)$

$= Y_{ij} - \bar{Y}_{\cdot\cdot} - \bar{Y}_{i\cdot} + \bar{Y}_{..}$

$\Rightarrow e_{ij} = Y_{ij} - \bar{Y}_{i\cdot}$

$SS_{tot} = SS_{tr} + SS_{res}.$

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑΣ

Πηγή μεταβλητότητας	SS	β.ε	MS	F-πηλίκο
Δοκιμασίες	$SStr$	$I-1$	$MStr = \frac{SStr}{I-1}$	$F = \frac{MStr}{MSres}$
Υπόλοιπα	$SSres$	$N-I$	$MSres = \frac{SSres}{N-I}$	
ολική	$SStot$	$N-1$		

— o — o — o — o — o — o —

ΑΣΚΗΣΕΙΣ από ΦΥΛΛΑΔΙΟ.

ΑΣΚΗΣΗ 2 $\quad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , E(\varepsilon_i)=0 \quad , Var(\varepsilon_i)=\sigma^2$

β) ΝΔΟ

$$\sum_{i=1}^{n} X_i e_i = 0 \quad \text{με } e_i = Y_i - \hat{Y}_i \quad , i=1,\dots,n$$

$$\Updownarrow$$

$$\underline{X} \perp \underline{e}$$

Απ $\quad \sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i (Y_i - \hat{Y}_i)$

$$= \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \hat{Y}_i$$

$$= \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i (\hat{\beta}_0 + \hat{\beta}_1 X_i)$$

$$= \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \quad *$$

Αλλά οι κανονικές εξισώσεις του μοντέλου της αγπ είναι $\beta_0 \sum X_i + \beta_1 \sum X_i^2 = \sum X_i Y_i$

και προφανώς ικανοποιούνται από $\hat{\beta}_0, \hat{\beta}_1$ $\quad **$

Από $*$ και $**$ προφανώς

$$\sum X_i Y_i - \hat{\beta}_0 \sum X_i - \hat{\beta}_1 \sum X_i^2 = 0 \implies \text{Άρα } \sum X_i e_i = 0$$

ΝΔΟ

γ) $Cov(\hat{\beta_0}, \hat{\beta_1}) = -\dfrac{\sigma^2 \bar{X}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$$\left.\begin{aligned} &Cov(\hat{\beta_0}, \hat{\beta_1}) = Cov(\bar{Y}-\bar{X}\hat{\beta_1}, \hat{\beta_1}) \\ &Cov\left(\sum_{i=1}^{n} a_i X_i, \sum \beta_j Y_j\right) = \sum_i \sum_j a_i \beta_j Cov(X_i, Y_j) \end{aligned}\right\} \longrightarrow$$

$$Cov(\hat{\beta_0}, \hat{\beta_1}) = Cov(\bar{Y}, \hat{\beta_1}) - \bar{X}\, Cov(\hat{\beta_1}, \hat{\beta_1})$$

$$= Cov(\bar{Y}, \hat{\beta_1}) - \bar{X}\, Var(\hat{\beta_1})$$

↓αποδείξαμε.

$$= 0 - \bar{X}\, Var(\hat{\beta_1})$$

$$= -\frac{\sigma^2 \bar{X}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

↓SOS (σίγουρο θέμα)

ΑΣΚΗΣΗ 3 (αριθμητική άσκηση) ~ πολλές πράξεις.

Θερμοκρασία	-5	-4	-3	-2	-1	0	1	2	3	4	5
Απόδοση	1	5	4	7	10	8	9	13	14	13	18

α) ανεξ μεταβλ X=Θερμοκρασία
εξαρτ μεταβλ Y=Απόδοση

Μοντέλο $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ $\quad i=1,2,\dots n \; (n=11)$

$\hat{\beta_1} = \dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} = \dots = 1,44$ (Η Απόδοση αυξάνεται κατά 1,44 (περίπου) όταν η θερμοκρασία μεταβληθεί κατά 1°C)

$\hat{\beta_0} = \bar{Y} - \bar{X}\hat{\beta_1} = \dots = 9,27$

Εκτιμώμενο μοντέλο : $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

ΑΝΑΔΙΑ

	SS	β.Ε	MS	F
Παλινδρομηση	226,95	1	226,95	F=96,17
Υπολοιπα	21,23	9	2,36	
ολική	248,18	10		

Από τον πινακα ΑΝΑΔΙΑ βλέπω ότι θα απορρ η $H_0: \beta_1 = 0$.

γ) Υποθέσεις Σφαλματων

- - - - - -

Για τον έλεγχο $H_0: \beta_1 = 0$ Χρησιμοποιειται το $F = \frac{MS_{reg}}{MS_{res}}$ με $F_{1,9}$

i) κ.π $F \geq F_{1,9,\alpha}$

$\alpha = 0{,}05$ το $F_{1,9,0{,}05} = 5{,}12$

$F = 96{,}17 > 5{,}12 = F_{1,9,0.05}$

Απορρ $H_0: \beta_1 = 0$

ii) 95% δε για β_1.

το δε ~~......~~ $\left(\hat{\beta}_1 - t_{9,0{,}025}\sqrt{\hat{Var}(\hat{\beta}_1)},\ \hat{\beta}_1 + t_{9,0{,}025}\sqrt{\hat{Var}(\hat{\beta}_1)}\right)$

(κάνω απόδειξη)

το δ.ε $(1{,}11\ ,\ 1{,}77)$.