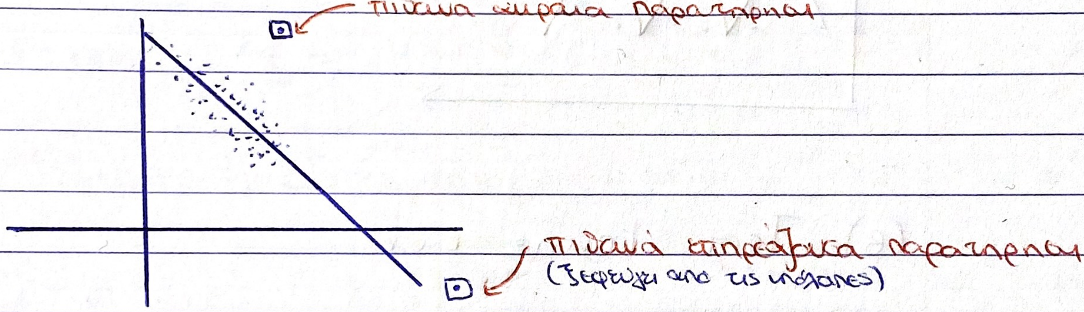
"8ο μάθημα" 22/11/21

Ακραίες και Επηρεάζουσες Παρατηρήσεις.

Μια παρατήρηση λέγεται ακραία αν απέχει σημαντικά από το κύριο σώμα των παρατηρήσεων που διαθέταμε. Ειδικότερα, στα γραμμικά μοντέλα, μια παρατήρηση είναι ακραία αν δεν προσαρμόζεται στο γραμμικό μοντέλο στο οποίο προσαρμόζονται οι υπόλοιπες παρατηρήσεις.



Μια παρατήρηση λέγεται επηρεάζουσα αν προσαρμόζεται στο μοντέλο (το γραμμικό), απέχει όμως από τις υπόλοιπες παρατηρήσεις και επηρεάζει τους εκτιμητές.

Διάγνωση ακραίων και επηρεαζουσών παρατηρήσεων

Ακραίες: Μια παρατήρηση Y_k είναι ακραία, αν το αντίστοιχο υπόλοιπο $e_k = Y_k - \hat{Y}_k$ έχει μεγάλη τιμή

Επηρεάζουσες: Τεστ που βασίζεται στην απόσταση Cook.

Μοντέλα Ανάλυσης Διακύμανσης

Ⓘ Ανάλυση Διακύμανσης κατά 1 παράγοντα.

Γραμμικά Μοντέλα: Μοντέλα, Μαθηματικές κατασκευές που διατυπώνονται για να διατυπώσουν τη σχέση (γραμμική) μεταξύ μιας εξαρτημένης και μίας ή περισσότερων ανεξάρτητων μεταβλητών.

Μοντέλα Παλινδρόμησης	Μοντέλα Ανάλυσης Διακύμανσης
α) Οι εξαρτημένες και οι ανεξάρτητες μεταβλητές είναι κατά κανόνα ποσοτικές (δηλ. μπορούν να μετρηθούν).	α) Η εξαρτημένη μεταβλητή είναι ποσοτική και οι ανεξάρτητες κατά κανόνα είναι ποιοτικές
β) Αναζητείται, κατασκευάζεται και ελέγχεται η ορθότητα της γραμμικής σχέσης μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών.	β) Δεν κατασκευάζεται κάποια γραμμική σχέση διερευνάται όμως ποιές τιμές της ανεξάρτητης μεταβλητής επηρεάζουν σημαντικά ή επιδρούν σημαντικότερα στην εξαρτημένη μεταβλητή

Περεταίρω, το γραμμικό μοντέλο παλινδρόμησης χρησιμοποιείται για προβλέψεις.

Ορολογία:

Παράγοντας: Ο όρος παράγοντας χρησιμοποιείται για να εκφράσει την ανεξάρτητη (-τες) μεταβλητή (-τες).

Επίπεδο του Παράγοντα: εννοούμε τιμή του παράγοντα

Δοκιμασία: Εννοούμε συνδυασμό επιπέδων των παραγόντων.

Προβλήματα Ανάλυσης Διακύμανσης στα οποία υπεισέρχεται ένας παράγοντας αφορούν την ανάλυση διακύμανσης κατά ένα παράγοντα

Δυο παράγοντες → Αναλ/δια κατά 2 παράγοντες

⋮

Πολλοί → Πειραματικός Σχεδιασμός.

Παράδειγμα:

Επίδοση Μαθητών = Y

↓

Επιδρά στην Y το επίπεδο μόρφωσης του πατέρα ??? (παράγοντας) A

- Y.E (Υποχ.Εκπ)
- Λ (λύκειο)
- Π (παν/ιο)
- Μ (μετ/κο)
- Δ (διδ/κο)

} Επίπεδα του Παράγοντα A

Επιδρά στην Y το επίπεδο μόρφωσης της μητέρας ??? (παράγοντας) B

- Y.E
- Λ
- Π
- Μ
- Δ

} Επίπεδα του παράγοντα B

(άρα πρόβλημα αναλ/δια κατά 2 παράγοντες).

Μορφή Δεδομένων στην Αναλ/δια κατά 1 παράγοντα:

Παράγοντας / Επίπεδα	(τοσοτική) Εξαρτ. Μεταβ/τη Y	$Y_{i\cdot}$	$\bar{Y}_{i\cdot}$	
1	$Y_{11}, Y_{12}, \ldots, Y_{1J_1}$	$Y_{1\cdot}$	$\bar{Y}_{1\cdot}$	μ_1
2	$Y_{21}, Y_{22}, \ldots, Y_{2J_2}$	$Y_{2\cdot}$	$\bar{Y}_{2\cdot}$	μ_2
⋮	⋮	⋮	⋮	
I	$Y_{I1}, Y_{I2}, \ldots, Y_{IJ_I}$	$Y_{I\cdot}$	$\bar{Y}_{I\cdot}$	μ_I

όπου: $Y_{i\cdot} = \sum_{j=1}^{J_i} Y_{ij}$, $\overline{Y}_{i\cdot} = \frac{1}{J_i} \cdot \sum_{j=1}^{J_i} Y_{ij}$

$$Y_{\cdot\cdot} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} Y_{ij} \quad , \quad \overline{Y}_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} Y_{ij} =$$

$$= \frac{1}{N} \cdot Y_{\cdot\cdot}$$

όπου $N = J_1 + \ldots + J_I = \sum_{i=1}^{I} J_i$

Συνεπώς, μπορώ να πω: $Y_{1J} \approx \mu_1 = \mu_1 + \varepsilon_{ij}$

$Y_{2J} \approx \mu_2 = \mu_2 + \varepsilon_{2j}$

$\vdots$

$Y_{IJ} \approx \mu_i + \varepsilon_{ij}$, με $i = 1, \ldots, I$

$j = 1, \ldots, J_i$

Μοντέλα Ανάλυσης Διακύμανσης

Ⓘ Ανάλυση Διακύμανσης κατά ένα Παράγοντα.

1^{η} μορφή μοντέλου ανάλυσης ένα παράγοντα: $\underset{\text{αποκρίσεις}}{Y_{ij}} = \underset{\text{παράμετροι}}{\mu_i} + \underset{\text{σφάλματα}}{\varepsilon_{ij}}$

, με $i = 1, \ldots, I$, $j = 1, \ldots, J_i$

Εκτιμητές Ελαχ. Τετραγώνων των παραμέτρων μ_i, $i = 1, \ldots, I$: $S = \sum_{i=1}^{I} \sum_{j=1}^{J_i} \varepsilon_{ij}^2 =$

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

$\frac{dS}{d\mu_I} = 0 \Rightarrow \ldots \Rightarrow \hat{\mu}_i = \overline{Y}_{i\cdot}$, με $i = 1, \ldots, I$

Υποθέσεις Σφάλματα:

$E(\varepsilon_i) = 0 \qquad i = 1, \dots, I$

$Var(\varepsilon_{ij}) = \sigma^2 \qquad j = 1, \dots, J_i$

$Cov(\varepsilon_{ij}, \varepsilon_{k\lambda}) = 0, \quad i \neq k, \ j \neq \lambda$

$\varepsilon_{ij} \sim N(0, \sigma^2)$

Ιδιότητες $\hat{\mu}_i$, $i = 1, \dots, I$

① Οι ΕΕΤ είναι αμερόληπτοι των παραμέτρων, δηλ.
$E(\hat{\mu}_i) = \mu_i, \quad i = 1, \dots, I$

② $Var(\hat{\mu}_i) = \dfrac{\sigma^2}{J_i}, \quad i = 1, \dots, I$

Απόδειξη προφανής

2η Μορφή Μοντέλου Αναλ/δια:

Παράμετροι

σφάλματα

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

αποκρίσεις

κοινή επίδραση όλων των επιπέδων πάνω στην Y.

Κωδικοποιεί και εκφράζει την ατομική επίδραση του i-επιπέδου στην Y.

ΕΕΤ των παραμέτρων μ και α_i, $i = 1, \dots, I$

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

$$\left.\begin{aligned} \frac{\partial S}{\partial \mu} &= 0 \\ \frac{\partial S}{\partial \alpha_i} &= 0 \end{aligned}\right\} \Rightarrow \text{Κανονικές Εξισώσεις} \Rightarrow$$

$$\sum_{i=1}^{I} \alpha_i J_i + \mu \cdot \sum_{i=1}^{I} J_i = \sum_{i=1}^{I} \cdot \sum_{j=1}^{J_i} Y_{ij}$$

$$\alpha_i J_i + \mu J_i = \sum_{j=1}^{J_i} Y_{ij} \quad , \quad \mu\epsilon \quad i=1,\dots,I$$

} Επειδή η 1η εξίσωση προκύπτει από τις υπόλοιπες I-εξισ. το σύστημα ΔΕΝ έχει μοναδική λύση ως προς μ και α_i, $i=1,\dots,I$.

(συμβαίνει όταν τα μοντέλα ανάλδια)

⟹ Για να πετύχω μοναδική λύση, υιοθετούμε τις πλευρικές συνθήκες όπως η $\sum_{i=1}^{I} \alpha_i J_i = 0$, η οποία οδηγεί από την 1η εξίσωση σε Ε.Ε.Τ. για την μ το $\bar{Y}_{..}$ κάτι που είναι στατιστικά αποδεκτό.

Για $\hat{\mu} = \bar{Y}_{..}$, τα $\hat{\alpha}_i = \frac{1}{J_i} \cdot \sum_{j=1}^{J_i} (Y_{ij} - \hat{\mu}) =$

$$= \bar{Y}_{i.} - \bar{Y}_{..}$$

Επομένως, οι ΕΕΤ των μ, α_i, με $i=1,\dots,I$ είναι :

$$\hat{\mu} = \bar{Y}_{..} \quad , \quad \hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..} \quad , \quad \forall\, i=1,\dots,I$$