Δευτέρα 30/11/2020.

Γραμμικά μοντέλα: Μαθηματικό μοντέλο, πιθανοθεωρητικό - στατιστικό το οποίο διατυπώνεται και αποτιμάται για να περιγράψει και να εκφράσει τη σχέση μεταξύ δύο ή περισσότερων τ.μ.

Μοντέλα Παλινδρόμησης.

α/ Διερευνούν τη γραμμική σχέση μεταξύ δύο ή περισσότερων ποσοτικών τ.μ.

β/ Διατυπώνουν και διερευνούν την εγκυρότητα της γραμ. σχέσης μεταξύ των εξαρτημένων και των ανεξάρτητων μεταβλητών, έτσι ώστε τελικά οι εξαρτημένες μεταβλητές να προβλέπονται για συγκεκριμένες τιμές των ανεξάρτητων.

Μοντέλα Ανάλυσης διακύμανσης:

1. Η εξαρτημένη μεταβλητή είναι κατά κανόνα ποσοτική και οι ~~εξ~~ ανεξάρτητες κατά κανόνα ποιοτικές.

2. Ενδιαφέρον παρουσιάζει ποιες από τις κατηγορίες της ποιοτικής ανεξάρτητης μεταβλητής ασκούν την σημαντικότερη επίδραση στην εξαρτημένη μεταβλητή.

Ορολογία - Συμβολισμοί

Παράγοντας: Ταυτόσημος με την ανεξάρτητη μεταβλητή. Χρησιμοποιείται για να εκφράσει την ανεξάρτητη μεταβλητή.

Επίπεδο του παράγοντα. εννοούμε κάθε τιμή, κάθε κατηγορία που αποτελεί τιμή της ποιοτικής μεταβλητής, δηλαδή του παράγοντα.

Ένα πρόβλημα στο οποίο διερευνάται η σχέση μιας εξαρτημένης μεταβλητής Y με ένα παράγοντα λέγεται πρόβλημα Ανάλυσης διακύμανσης κατά έναν παράγοντα.

Αν διερευνώ τη σχέση της Y με δύο παράγοντες έχω πρόβλημα ανάλυσης διακύμανσης κατά δύο παράγοντες.

Περισσότεροι παράγοντες → Πειραματικοί σχεδιασμοί.

Δευτέρα 30/11/2020

Διαδικασία : κάθε συνδυασμός επιπέδων του παράγοντα.

Παράδειγμα: Ενδιαφέρει η διερεύνηση της σχέσης μεταξύ της επίδοση των μαθητών και του επιπέδου μόρφωσης του πατέρα.

Μοντελοποίηση:

$Y \leftarrow$ εξαρτημένη $\leftarrow$ επίδοση των μαθητών

$X \leftarrow$ παράγοντας $\leftarrow$ επίπεδο μόρφωσης του πατέρα.

Τα επίπεδα του παράγοντα είναι:

1) Απόφοιτος υποχρεωτικής εκπ/ης. (Υ)
2) απόφοιτος λυκείου (Λ)
3) απόφοιτος πανεπιστημίου (Π)
4) μεταπτυχιακό (Μ)
5) διδακτορικό (Δ).

Αν εισάγαμε και το επίπεδο μόρφωσης της μητέρας τότε το πρόβλημα ανάγεται σε πρόβλημα ανάλυσης διακύμανσης κατά δυο παράγοντες

2ος παράγοντας = Ε.Μ.Μ. επίπεδο 2ου παράγοντα — ΑΥ, ΑΛ, ΑΠ, Μ, Δ → Y.

Δοκιμασία (ΑΥ,ΑΥ), (ΑΥ, Λ), ... κ.λ.π.

Αν έχω πρόβλημα ανάλυσης διακύμανσης κατά έναν παράγοντα.
$\Rightarrow$ Δοκιμασία $\equiv$ επίπεδο.

Ανάλυση Διακύμανσης κατά έναν παράγοντα

Μορφή Δεδομένων

Άθροισμα παρατηρήσεων της 1ης γραμμής ως προς τον δεύτερο δείκτη

Παράγοντας	Δεδομένα για την Y για κάθε επίπεδο του παράγοντα	Σύνολα	Μέσος όρος
1	$Y_{11}, Y_{12}, \dots Y_{1J_1}$	$Y_{1\cdot}$	$\bar{Y}_{1\cdot}$
2	$Y_{21}, Y_{22}, \dots Y_{2J_2}$	$Y_{2\cdot}$	$\bar{Y}_{2\cdot}$
$\vdots$			
I	Y_{I1}, Y_{I2}, Y_{IJ_I}	$Y_{I\cdot}$	$\bar{Y}_{I\cdot}$

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

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

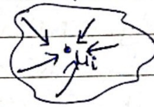
$$\bar{Y}_{i\cdot} = \frac{1}{J_i}\sum_{j=1}^{J_i} Y_{ij}.$$

$$Y_{\cdot\cdot} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}.$$ Άθροισμα όλων των παρατηρήσεων

$$\bar{Y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}$$ όπου $N = J_1 + \dots + J_I$: πλήθος παρατηρήσεων

Τα δεδομένα $Y_{i1}, Y_{i2}, \dots Y_{iJ_i}$ της i-γραμμής συνιστούν τ.δ. από πληθυσμό με μέση τιμή μ_i, $i=1,\dots I$

Σχηματικά $Y_{i1}, Y_{i2}, \dots Y_{iJ_i}$ είναι τ.δ από κάποιον πληθυσμό που αναφέρεται στο επίπεδο i του παράγοντα, $i=1,\dots I$ με μ_i η μέση τιμή του πληθυσμού, γύρω από το οποίο συγκεντρώνονται τα μέλη του πληθυσμού και επομένως $Y_{i1}, \dots Y_{iJ}$.



Άρα $Y_{ij} \simeq \mu_i$ ή καλύτερα $Y_{ij} = \mu_i + \varepsilon_{ij}$, $i=1,\dots I$
$j=1,\dots J_i$

Άρα το μοντέλο ανάλυσης διακύμανσης κατά ένα παράγοντα είναι

$$Y_{ij} = \overbrace{\mu_i}^{\text{παράμετρος μοντέλου}} + \underbrace{\varepsilon_{ij}}_{\text{σφάλματα μοντέλου}}, \quad i=1,\dots I, \quad j=1,\dots J_i$$

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

ΕΕΤ. των παραμέτρων μ_i, $i=1,\dots I$.

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

$\frac{\partial S^2}{\partial \mu_i} = 0$, $i=1,..I$ $\Rightarrow$ ΕΕΤ: $\hat{\mu_i} = \bar{Y}_{i\cdot}$, $i=1,...I$.

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

περιγράφει την ατομική επίδραση του i επιπέδου στην Y.

Ισοδύναμο μοντέλο: $Y_{ij} = \mu + a_i + \varepsilon_{ij}$, $i=1,...I$, $j=1,...J$.

περιγράφει την κοινή επίδραση όλων των επιπέδων Y.

μ, a_i, $i=1,...I$ παράμετροι του μοντέλου

Εκτιμητές ελαχίστων τετραγώνων παραμέτρων μ, a_i, $i=1,...I$.

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

$$\left.\begin{array}{l} \frac{\partial S^2}{\partial \mu} = 0 \\ \frac{\partial S^2}{\partial a_i} = 0 \end{array}\right\} \xrightarrow[\text{εξισώσεις}]{\text{Κανονικές}} \left\{\begin{array}{l} \sum_{i=1}^{I} a_i J_i + \mu \sum_{i=1}^{I} J_i = \sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij} \\ a_i J_i + \mu \cdot J_i = \sum_{j=1}^{J_i} Y_{ij}, \quad i=1,..I \end{array}\right\}$$

Το σύστημα των κανονικών εξισώσεων δεν οδηγεί σε μοναδική λύση ως προς μ και αi επειδή αν αθροίσω τη δεύτερη εξίσωση προκύπτει η πρώτη. Για να πετύχω μοναδική λύση θεωρώ κάποια συνθήκη την καλούμενη $\boxed{\text{πλευρική συνθήκη}}$

Πλευρικές συνθήκες διατυπώνονται έτσι ώστε να πετύχουμε διαισθητικά αποδεκτές λύσεις δηλαδή διαισθητικά -στατιστικά αποδεκτούς εκτιμητές.

π.χ) Αν το $\sum_{i=1}^{I} a_i J_i = 0$ τότε $\mu = \frac{\sum\sum Y_{ij}}{\sum J_i} = \frac{1}{N} Y_{\cdot\cdot} = \bar{Y}_{\cdot\cdot}$

Άρα ΕΕΤ του μ θα είναι το $\bar{Y}_{\cdot\cdot}$ αν $\sum a_i J_i = 0$, όπου αυτή είναι η πλευρική συνθήκη.

Θεωρώ πλευρική συνθήκη την $\sum_{i=1}^{I} a_i J_i = 0$

Τότε η λύση της πρώτης συνθήκης είναι $\mu = \bar{Y}_{..}$

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

από τις υπόλοιπες I εξισώσεις αν θέσω στη θέση του μ το $\bar{Y}_{..}$ και λύσω ως προς το αi προκύπτει $a_i = \bar{Y}_{i.} - \bar{Y}_{..}$

Άρα ΕΕΤ του αi είναι το $\boxed{\hat{a}_i = \bar{Y}_{i.} - \bar{Y}_{..}}$

Ανάλυση της ολικής μεταβλητότητας.

Η ολική μεταβλητότητα σχετίζεται με την δειγματική διακύμανση που είναι άθροισμα τετραγώνων αποκλίσεων των παρατηρήσεων από το δειγματικό μέσο.

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

Θεώρημα:

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

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

Απόδ)

Προκύπτει με πράξεις αν. από το SStot $\pm \bar{Y}_{i.}$ (Άσκηση).

Υπόλοιπα:

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

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

$$\boxed{SStr = \sum_{i=1}^{I} J_i (\bar{Y}_{i.} - \bar{Y}_{..})^2.}$$

Πίνακας Ανάλυσης του μοντέλου της Ανάλυσης Διακύμανσης.

Πηγή μεταβλ.	SS.	β.Ε.	MS.	F-πηλίκο
Μοντέλο (Δοκιμασίες)	SS_{tr}	$I-1$	$SS_{tr}/I-1$	
Υπόλοιπα	SS_{res}	$N-I$	$SS_{res}/N-I$	$F = \frac{MS_{tr}}{MS_{res}}$
Ολική μεταβλ.	SS_{tot}	$N-1$	–	