ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΑΝΟΙΚΤΑ ΑΚΑΔΗΜΑΪΚΑ ΜΑΘΗΜΑΤΑ

Εφαρμοσμένη Στατιστική

Ανάλυση διακύμανσης

Διδάσκων: Επίκουρος Καθηγητής
Κωνσταντίνος Μπλέκας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

Ανάλυση διασποράς (ANOVA)

Ερώτημα: Πως μπορούμε να συγκρίνουμε k>2 πληθυσμούς ως προς τις μέσες τιμές τους $\mu_1, \mu_2, \ldots, \mu_k$.

Πιθανή λύση: κάνοντας $c = \binom{k}{2} = \frac{k(k-1)}{2}$ t-tests (ελέγχους) ανά δύο.

Δυσκολίες:

- Μεγάλη πολυπλοκότητα
- Υπάρχει ο κίνδυνος να οδηγηθούμε σε λάθος συμπεράσματα καθώς η πιθανότητα του σφάλματος τύπου Ι αυξάνεται δραματικά. Γιατί; αν κάθε έλεγχος έχει πιθανότητα σφάλματος a και άρα πιθανότητα να μην κάνουμε σφάλμα 1-a, τότε η πιθανότητα να μην κάνουμε σφάλμα σε c ελέγχους είναι $(1-a)^c$ και άρα πιθανότητα να κάνουμε σφάλμα τύπου Ι είναι $1- (1-a)^c$.

Ανάλυση διασποράς (ANOVA) (συν.)

Λύση: Μέθοδος ANOVA (Fisher 1918) η οποία βασίζεται στην ιδέα ότι όλες οι διαφορές στους δειγματικούς μέσους κρίνονται στατιστικά σημαντικές ή όχι με βάση το μέτρο της μεταβλητότητας (διασποράς) εντός των δειγμάτων.

Έστω k ομάδες δειγμάτων $X_{(i)} = \{x_{i1}, x_{i2}, \ldots, x_{in_i}\}$ μεγέθους n_i . Ορίζουμε ως:

$$TSS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}\right)^2$$

Συνολικό άθροισμα τετραγώνων (total sum of squares)

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2$$

Άθροισμα τετραγώνων στο εσωτερικό ομάδων (sum of squares within groups)

$$SSB = \sum_{i=1}^{k} n_i \left(\bar{x}_i - \bar{x}\right)^2$$

Άθροισμα τετραγώνων μεταξύ των ομάδων (sum of squares between groups)

Ανάλυση διασποράς (ANOVA) (συν.)

Ισχύει ότι: TSS = SSB + SSW

Η μέθοδος ANOVA ορίζει τον στατιστικό έλεγχο:

H_0 : $\mu_1 = \mu_2 = \ldots = \mu_k$
H_1 : τουλάχιστον 2 από τα κέντρα διαφέρουν στατιστικά

Χρησιμοποιώντας το γνωστό αποτέλεσμα για την κατανομή X_2 παίρνουμε:

$$Y_W = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2}{\sigma^2} \sim X^2_{N-k} \qquad Y_B = \frac{\sum_{i=1}^{k} n_i \left(\bar{x}_i - \bar{x}\right)^2}{\sigma^2} \sim X^2_{k-1}$$

όπου $N = \sum_{i=1}^{k} n_i$

Ανάλυση διασποράς (ANOVA) (συν.)

Έτσι χρησιμοποιούμε τη μεταβλητή:

$$F = \frac{\dfrac{Y_B}{k-1}}{\dfrac{Y_W}{N-k}} = \frac{\dfrac{SSB}{k-1}}{\dfrac{SSW}{N-k}} \sim F_{k-1,N-k}$$

Και έτσι ο έλεγχος γίνεται πάνω στον παραπάνω λόγο, δηλ.

If $F = \dfrac{\dfrac{SSB}{k-1}}{\dfrac{SSW}{N-k}} > f_{k-1,N-k}(a)$ then H0 is rejected, δηλ. υπάρχει στατιστικά σημαντική μεταβλητότητα.

Ανάλυση διασποράς (ANOVA) (συν.)

Συνοψίζοντας, τα διάφορα βήματα περιγράφονται στον λεγόμενο πίνακα ανάλυσης διασποράς (**ANOVA – table**)

Πηγή μεταβλητότητας	Αθροισμα τετραγώνων	β.ε.	Μέσο άθροισμα τετραγώνων	Κριτήριο F
Μεταξύ ομάδων (B)	$SSB = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2$	*k-1*	$S_B^2 = \frac{SSB}{k-1}$	
Εσωτερικό ομάδων (W)	$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	*N-k*	$S_W^2 = \frac{SSW}{N-k}$	$F = \frac{S_B^2}{S_W^2}$
Ολικό (T)	$TSS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	*N-1*		

Ανάλυση διασποράς (ANOVA) (συν.)

Η προηγούμενη περίπτωση αφορούσε ανάλυση **ANOVA με 1 παράγοντα**.
Για ανάλυση **ANOVA με 2 παράγοντες**, τότε κάνουμε τα παρακάτω:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{h=1}^{n}(x_{ijh} - \bar{x})^2$$

Συνολική μεταβολή των παρατηρήσεων από τη μέση τιμή

$$SSA = mn\sum_{i=1}^{k}(\bar{x}_{i.} - \bar{x})^2$$

μεταβολή που οφείλεται στο παράγοντα A (β.ε. k-1)

$$SSB = kn\sum_{j=1}^{m}(\bar{x}_{.j} - \bar{x})^2$$

μεταβολή που οφείλεται στο παράγοντα B (β.ε. m-1)

$$SSAB = n\sum_{i=1}^{k}\sum_{j=1}^{m}(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2$$

μεταβολή που οφείλεται στην αλληλεπίδραση των A, B (β.ε. (k-1)(m-1))

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{h=1}^{n}(x_{ijh} - \bar{x}_{ij})^2$$

μεταβολή εξαιτίας σφαλμάτων (β.ε. km(n-1))

ισχύει ότι SST=SSA+SSB+SSAB+SSE

Ανάλυση διασποράς (ANOVA) (συν.)

Στην ανάλυση **ANOVA με 2 παράγοντες** διακρίνουμε τους ελέγχους:

- Επίδραση παράγοντα Α $$F_A = \frac{\frac{SSA}{k-1}}{\frac{SSE}{km(n-1)}}$$
- Επίδραση παράγοντα Β $$F_B = \frac{\frac{SSB}{m-1}}{\frac{SSE}{km(n-1)}}$$
- Αλληλεπίδραση Α, Β $$F_{A,B} = \frac{\frac{SSAB}{(k-1)(m-1)}}{\frac{SSE}{km(n-1)}}$$

Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Ιωαννίνων**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Έχουν προηγηθεί οι κάτωθι εκδόσεις:

- Έκδοση 1.0 διαθέσιμη εδώ. http://ecourse.uoi.gr/course/view.php?id=1052.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Ιωαννίνων, Διδάσκων: Επίκουρος Καθηγητής Κωνσταντίνος Μπλέκας. «Εφαρμοσμένη Στατιστική. Ανάλυση διακύμανσης». Έκδοση: 1.0. Ιωάννινα 2014. Διαθέσιμο από τη δικτυακή διεύθυνση: http://ecourse.uoi.gr/course/view.php?id=1052.

Σημείωμα Αδειοδότησης