



# Ανάλυση διακύμανσης (Μέρος 2<sup>ο</sup>)

31/3/2017

## Σχέδιο τυχαιοποιημένων πλήρων ομάδων (1)

- Αποτελεί ευθεία γενίκευση του σχεδίου που γνωρίσαμε όταν μιλήσαμε για τη **σύγκριση κατά ζεύγη** δύο μέσων  $\mu_1$  και  $\mu_2$
- Η διαφορά από τη σύγκριση κατά ζεύγη έγκειται μόνο στο ότι αντί να δημιουργούμε έναν αριθμό, έστω  $b$ , ζευγών πειραματικών μονάδων επί των οποίων κάνουμε δύο επεμβάσεις, δημιουργούμε έναν αριθμό  $b$  **ομάδων** αποτελούμενων από  $k > 2$  **πειραματικές μονάδες** (αντί για δύο) γιατί κάνουμε περισσότερες από δύο επεμβάσεις (ή αλλιώς, γιατί ο **παράγοντας του οποίου ελέγχουμε την επίδραση στη μεταβλητή απόκρισης**, έχει περισσότερες από δύο στάθμες)

## Σχέδιο τυχαιοποιημένων πλήρων ομάδων (2)

- ▶ Το επιλέγουμε όταν θεωρούμε/υποθέτουμε ότι οι πειραματικές μονάδες προσθέτουν μεταβλητότητα στις τιμές της μεταβλητής απόκρισης, με την έννοια ότι αυτή επηρεάζεται και από κάποιον άλλο, «εξωγενή/ενοχλητικό» παράγοντα, πέραν του παράγοντα του οποίου ελέγχουμε την επίδραση
- ▶ Οι ομάδες δημιουργούνται με κριτήριο να αποτελούνται από πειραματικές μονάδες που έχουν παρόμοια/ομοιογενή συμπεριφορά ως προς αυτόν τον «εξωγενή» παράγοντα
- ▶ Έτσι, επιτυγχάνουμε ο έλεγχος της επίδρασης του παράγοντα που μας ενδιαφέρει να γίνεται εντός κάθε ομάδας σε παρόμοιες συνθήκες και να απομονωθεί η μεταβλητότητα που οφείλεται στον «εξωγενή» παράγοντα

## Σχέδιο τυχαιοποιημένων πλήρων ομάδων (3)

- Αφού δημιουργήσουμε τις  $b$  ομάδες (με  $k$  πειραματικές μονάδες στην καθεμιά), αντιστοιχίζουμε, με μια τυχαία διαδικασία, μία προς μία τις  $k$  επεμβάσεις στις  $k$  πειραματικές μονάδες κάθε ομάδας
- Έτσι, συνολικά έχουμε  $n = b \cdot k$  παρατηρήσεις

## Υποθέσεις / παραδοχές στο σχέδιο των τυχαιοποιημένων πλήρων ομάδων (1)

- Οι  $b$  παρατηρήσεις, για κάθε μία από τις  $k$  επεμβάσεις, συγκεντρώνονται από τις  $b$  ομάδες (μία παρατήρηση από κάθε ομάδα)
- Επομένως, οι  $k$  παρατηρήσεις κάθε ομάδας **είναι προφανώς εξαρτημένες** (γιατί σε κάθε μία αντανακλώνται οι συνθήκες που ενυπάρχουν στη συγκεκριμένη ομάδα)
- Δηλαδή, η ανεξαρτησία των παρατηρήσεων, στο σχέδιο των τυχαιοποιημένων πλήρων ομάδων, δεν υφίσταται για όλες τις παρατηρήσεις
- Γι' αυτό, στο σχέδιο αυτό, παίζει πολύ σημαντικό ρόλο η τυχαία αντιστοίχιση των επεμβάσεων στις πειραματικές μονάδες κάθε ομάδας

## Υποθέσεις / παραδοχές στο σχέδιο των τυχαιοποιημένων πλήρων ομάδων (2)

- ▶ Οι υποθέσεις που ισχύουν είναι οι εξής:
  1. Καθένα από τα δείγματα προέρχεται από **κανονικό πληθυσμό**
  2. Οι  **$k$**  πληθυσμοί έχουν κοινή διακύμανση  **$\sigma^2$**  (ομοσκεδαστικότητα)
  3. Οι επεμβάσεις και οι ομάδες επιδρούν στη μεταβλητή απόκρισης προσθετικά ανεξάρτητα

## Υποθέσεις / παραδοχές στο σχέδιο των τυχαιοποιημένων πλήρων ομάδων (3)

- Οι υποθέσεις 1 και 2 είναι ισοδύναμες με τις υποθέσεις ότι τα **σφάλματα (θεωρητικά υπόλοιπα)** ακολουθούν κανονικές κατανομές με μέση τιμή **0** και κοινή διακύμανση  **$\sigma^2$** , δηλαδή ότι
  - $\varepsilon_{ij} \sim N(0, \sigma^2)$
- Τα  $\varepsilon_{ij}$  εκφράζουν το πειραματικό σφάλμα, δηλαδή τη μεταβλητότητα που δεν εξηγείται από τις επεμβάσεις
- Οι υποθέσεις 1 και 2 ελέγχονται όπως στο εντελώς τυχαιοποιημένο σχέδιο

## Έλεγχος ανάλυσης διακύμανσης στο σχέδιο τυχαιοποιημένων πλήρων ομάδων (1)

- ▶  $H_0: \mu_1 = \mu_2 = \dots = \mu_k$  ή ο παράγοντας **δεν επιδρά**
- ▶  $H_1: \mu_i \neq \mu_j$  για ένα τουλάχιστον ζεύγος  $(i, j)$  ή ο παράγοντας **επιδρά**
  
- ▶  $H_0: \mu_1 = \mu_2 = \dots = \mu_b$  ή οι ομάδες **δεν επιδρούν**
- ▶  $H_1: \mu_i \neq \mu_j$  για ένα τουλάχιστον ζεύγος  $(i, j)$  ή οι ομάδες **επιδρούν**



# Έλεγχος ανάλυσης διακύμανσης στο σχέδιο τυχαιοποιημένων πλήρων ομάδων (2)

Πηγή μεταβλητότητας	B.E.	Άθροισμα τετραγώνων SS	Μέσο άθροισμα τετραγώνων MS	Κριτήριο F	Περιοχή απόρριψης
Επεμβάσεις ή Παράγοντας ή Μεταξύ των δειγμάτων	$k-1$	$SST_r = \sum_{j=1}^k \frac{T_j^2}{b} - \frac{G^2}{n}$	$MST_r = \frac{SST_r}{k-1}$	$F_{Tr} = \frac{MST_r}{MSE}$	$F_{Tr} \geq F_{k-1;(b-1)(k-1);a}$
Ομάδες	$b-1$	$SSB = \sum_{i=1}^b \frac{B_i^2}{k} - \frac{G^2}{n}$	$MSB = \frac{SSB}{b-1}$	$F_B = \frac{MSB}{MSE}$	$F_B \geq F_{b-1;(b-1)(k-1);a}$
Σφάλμα ή Εντός των δειγμάτων	$(b-1)(k-1)$	$SSE = SST_{ot} - SST_r - SSB$	$MSE = \frac{SSE}{(b-1)(k-1)}$		
Ολική	$kb-1$	$SST_{ot} = SST_r + SSB + SSE$			

όπου  $k$  το πλήθος των επεμβάσεων και  $b$  το πλήθος των ομάδων

$G = \sum_{ij} y_{ij}$ , το άθροισμα όλων των παρατηρήσεων

$T_j, j = 1, 2, \dots, k$  το άθροισμα όλων των παρατηρήσεων από την  $j$  επέμβαση

$B_i, i = 1, 2, \dots, b$  το άθροισμα όλων των παρατηρήσεων στην  $i$  ομάδα

# Έλεγχοι πολλαπλών συγκρίσεων

- ▶ Όπως και στην περίπτωση του εντελώς τυχαιοποιημένου σχεδίου, όταν η μηδενική υπόθεση  $H_0: \mu_1 = \mu_2 = \dots = \mu_k$  απορριφθεί, δηλαδή όταν τα πειραματικά δεδομένα υποστηρίζουν ότι οι μέσοι  $\mu_1, \mu_2, \dots, \mu_k$  των  $k$  επεμβάσεων διαφέρουν στατιστικά σημαντικά, μπορούμε να κάνουμε ελέγχους πολλαπλών συγκρίσεων για να διερευνήσουμε ποιοι από τους  $k$  μέσους διαφέρουν στατιστικά σημαντικά

# Παράδειγμα 1 (1)

- Μια ομάδα ερευνητών προκειμένου να συγκρίνει την επίδραση τριών μεθόδων προετοιμασίας του εδάφους στην ανάπτυξη ενός είδους πεύκης κατά το πρώτο έτος από τη φύτευση, σχεδίασε το εξής πείραμα
- Επέλεξε τέσσερις εκτάσεις (ίδιου εμβαδού) σε τέσσερις διαφορετικές δασικές περιοχές και κάθε μια τη διαίρεσε σε τρία πειραματικά τεμάχια ίδιου εμβαδού
- Σε καθένα από τα τρία πειραματικά τεμάχια κάθε έκτασης αντιστοίχησε, με μια τυχαία διαδικασία, μια από τις τρεις μεθόδους προετοιμασίας του εδάφους
- Στη συνέχεια, σε κάθε πειραματικό τεμάχιο, αφού προετοίμασε το έδαφος με την αντίστοιχη μέθοδο, φύτεψε ένα δενδρύλλιο πεύκης

## Παράδειγμα 1 (2)

- Στον Πίνακα που ακολουθεί, δίνονται (σε cm) οι μετρήσεις που πήραν οι ερευνητές ένα έτος μετά τη φύτευση

		Περιοχή φύτευσης			
		E1	E2	E3	E4
Μέθοδος προετοιμασίας εδάφους	M1	11	13	16	10
	M2	15	17	20	12
	M3	10	15	13	10

## Παράδειγμα 1 (3)

- Οι ερευνητές θέλησαν να απομονώσουν τη μεταβλητότητα στις τιμές της μεταβλητής απόκρισης (**ανάπτυξη δενδρυλλίων πεύκης κατά το πρώτο έτος από τη φύτευση**) που οφείλεται στην **περιοχή φύτευσης**
- Γι' αυτό δημιούργησαν 4 **ομάδες** (καθόρισαν 4 διαφορετικές **περιοχές**) και σε κάθε μια από αυτές έκαναν και τις 3 **επεμβάσεις (μέθοδοι προετοιμασίας του εδάφους)**
- Αυτό γιατί θεώρησαν ότι η ανάπτυξη των δενδρυλλίων πεύκης, επηρεάζεται εκτός από τον παράγοντα **μέθοδος προετοιμασίας του εδάφους** (που τους ενδιαφέρει να μελετήσουν την επίδρασή του), και από την **περιοχή της φύτευσης**

## Παράδειγμα 1 (4)

- ▶ Ελέγξτε αν η επίδραση του παράγοντα **μέθοδος προετοιμασίας του εδάφους** είναι, σε επίπεδο σημαντικότητας 5%, στατιστικά σημαντική

## Παράδειγμα 2 (1)

- ▶ Ένας φοιτητής, στο πλαίσιο της πτυχιακής του εργασίας, προκειμένου να συγκρίνει τέσσερα είδη (A1, A2, A3, και A4) πρόσθετης ύλης ζωοτροφών που χρησιμοποιείται για αύξηση του βάρους νεογέννητων χοίρων, σχεδίασε και εκτέλεσε το εξής πείραμα
- ▶ Από πέντε διαφορετικές γέννες επέλεξε τυχαία 4 χοίρους από την κάθε μία
- ▶ Δημιούργησε έτσι, 5 τετράδες χοίρων (Γ1, Γ2, Γ3, Γ4 και Γ5) όπου οι χοίροι της ίδιας τετράδας προέρχονται από την ίδια γέννα, και στους χοίρους κάθε τετράδας αντιστοίχησε, με μια τυχαία διαδικασία, από μια πρόσθετη ύλη ζωοτροφών
- ▶ Αφού χορήγησε για τρεις μήνες στους χοίρους κάθε ομάδας τροφή με την αντίστοιχη πρόσθετη ύλη, κατέγραψε την αύξηση σε βάρος κάθε χοίρου

## Παράδειγμα 2 (2)

- Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές στη μέση αύξηση του βάρους των νεογέννητων χοίρων που να οφείλονται στα τέσσερα είδη πρόσθετης ύλης ζωοτροφών;

Ομάδα (γέννα)	Παράγοντας/Επεμβάσεις (πρόσθετη ύλη)			
	A1	A2	A3	A4
Γ1	78	69	78	80
Γ2	68	64	70	70
Γ3	76	70	73	83
Γ4	76	66	77	74
Γ5	61	66	69	70