# Κεφάλαιο 13

# Εισαγωγή στην Ανάλυση Διακύμανσης

1

## Η Ανάλυση Διακύμανσης

Από τα πιο συχνά χρησιμοποιούμενα στατιστικά κριτήρια στην κοινωνική έρευνα

Γιατί;

1. Ενώ αναφέρεται σε διαφορές μέσων όρων, όπως και το κριτήριο t, δεν έχει περιορισμούς στον αριθμό των μέσων όρων που είναι δυνατόν να συγκριθούν.
2. Μας επιτρέπει να μελετήσουμε ταυτόχρονα την επίδραση δύο ή περισσότερων ανεξάρτητων μεταβλητών. Έτσι, υπολογίζουμε όχι μόνο την επίδραση της καθεμίας ανεξάρτητης μεταβλητής στην εξαρτημένη, αλλά και τις αλληλεπιδραστικές συνέπειες των ανεξάρτητων μεταβλητών στην εξαρτημένη.

2

## Ένα παράδειγμα

Τι θα κάνατε στην περίπτωση που είχατε δεδομένα από τέσσερις ερευνητικές ομάδες και θέλατε να συγκρίνετε τις επιδόσεις τους;

**Μια πιθανή απάντηση**: να πραγματοποιήσουμε πολλαπλά κριτήρια t σε όλους τους πιθανούς συνδυασμούς μέσων όρων…

Κάτι τέτοιο δεν θα ήταν ενδεδειγμένο για δύο λόγους:

**Ο πρακτικός λόγος**: όσο περισσότερες ομάδες δεδομένων έχουμε, τόσο περισσότερα κριτήρια t θα πρέπει να πραγματοποιήσουμε (π.χ., με οκτώ ομάδες πρέπει να γίνουν 28 κριτήρια t!)

**Ο στατιστικός λόγος**: όσο περισσότερα κριτήρια t πραγματοποιούνται, τόσο αυξάνει η πιθανότητα να οδηγηθούμε σε σφάλμα Τύπου Ι…

3

## Η διακύμανση των ερευνητικών δεδομένων

Ας υποθέσουμε ότι έχουμε πραγματοποιήσει μια έρευνα στην οποία εξετάστηκε η επίδραση του φύλου στην επίδοση σε μια δοκιμασία.

Είναι αναμενόμενο ότι οι μετρήσεις θα διαφέρουν μεταξύ τους (έστω και ελάχιστα) και οι διαφορές αυτές θα είναι προς δύο κατευθύνσεις:

1. Θα διαφέρουν μεταξύ τους οι τιμές καθεμίας ερευνητικής ομάδας (π.χ., όλα τα αγόρια δεν θα έχουν την ίδια επίδοση) λόγω ατομικών διαφορών και τυχαίων σφαλμάτων.
2. Θα διαφέρουν μεταξύ τους οι μέσοι όροι των ερευνητικών συνθηκών (η μέση επίδοση των αγοριών και αυτή των κοριτσιών δεν θα είναι ακριβώς η ίδια) λόγω της επίδρασης της ανεξάρτητης μεταβλητής (του φύλου), ατομικών διαφορών και τυχαίων σφαλμάτων.

Αυτό που κάνει ένα παραμετρικό στατιστικό κριτήριο είναι να υπολογίζει το ποσοστό της συνολικής διακύμανσης των μετρήσεων που οφείλεται στην ανεξάρτητη μεταβλητή

4

## Η διακύμανση των ερευνητικών δεδομένων

- Το σύνολο της διακύμανσης οφείλεται τόσο στην επίδραση των ανεξάρτητων μεταβλητών όσο και σε όλους τους άγνωστους παράγοντες
- **Σφάλμα:** Το ποσοστό της συνολικής διακύμανσης των τιμών μιας ομάδας που προκύπτει κατά τη μέτρηση μιας συμπεριφοράς από μια συνισταμένη πολλαπλών παραγόντων που παρεμβαίνουν κατά τη στιγμή της εκδήλωσης της συμπεριφοράς
- Επιδίωξη κάθε ερευνητή είναι το μεγαλύτερο ποσοστό της συνολικής διακύμανσης των τιμών να οφείλεται στο χειρισμό της ανεξάρτητης μεταβλητής, ενώ το ποσοστό της διακύμανσης που οφείλεται στους άγνωστους παράγοντες (σφάλμα) να είναι ελάχιστο. Τα ποσοστά αυτά μπορούν να εκφραστούν με ένα κλάσμα (πηλίκο):

$$\frac{\text{αναμενόμενη διακύμανση ως αποτέλεσμα των ανεξάρτητων μεταβλητών}}{\text{διακύμανση που οφείλεται σε όλες τις άλλες μεταβλητές (σφάλμα)}}$$

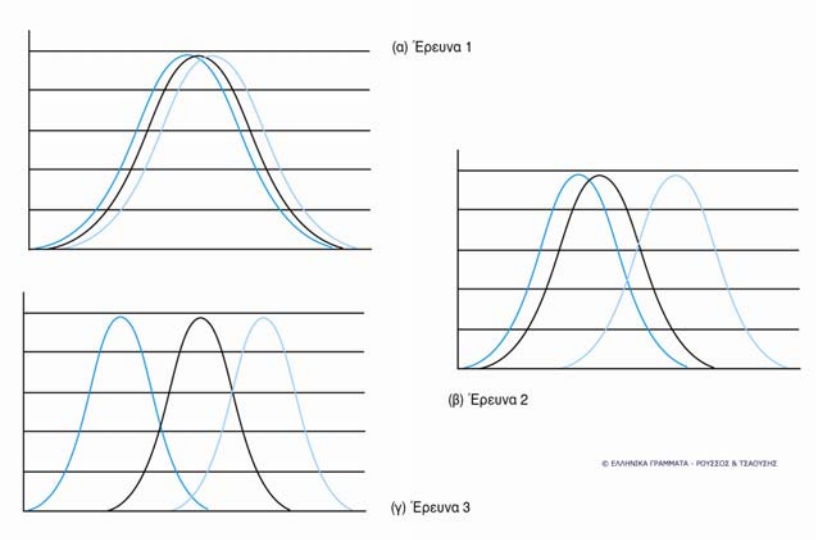
5

## Συστηματικές διαφορές και τυχαία σφάλματα

| | Έρευνα 1 | | | Έρευνα 2 | | | Έρευνα 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | α | β | γ | α | β | γ | α | β | γ |
| | 17 | 16 | 19 | 18 | 18 | 40 | 20 | 30 | 40 |
| | 16 | 18 | 25 | 21 | 18 | 44 | 19 | 30 | 41 |
| | 22 | 21 | 19 | 16 | 20 | 38 | 21 | 31 | 39 |
| | 16 | 18 | 25 | 21 | 18 | 42 | 20 | 29 | 41 |
| | 23 | 24 | 18 | 18 | 23 | 37 | 21 | 29 | 40 |
| | 20 | 23 | 20 | 20 | 23 | 39 | 19 | 31 | 39 |
| $\overline{X}$ | **19** | **20** | **21** | **19** | **20** | **40** | **20** | **30** | **40** |

6

## Συστηματικές διαφορές και τυχαία σφάλματα




7

## Ο υπολογισμός της διακύμανσης των τιμών

- Η έννοια της διακύμανσης
- Ο τύπος υπολογισμού της (όπου d η απόκλιση καθεμίας τιμής από το μέσο όρο):

$$s^2 = \frac{\Sigma d^2}{\mathrm{N}}$$

Επομένως, ο τύπος μπορεί να γραφεί και ως εξής:

$$s^2 = \frac{\Sigma(\mathrm{X} - \overline{\mathrm{X}})^2}{\mathrm{N}}$$

Το σημαντικότερο τμήμα στον υπολογισμό της διακύμανσης είναι το **άθροισμα των τετραγώνων** (Sum of Squares - SS), δηλαδή το άθροισμα των τετραγώνων των αποκλίσεων μιας ομάδας τιμών από το μέσο όρο τους:

$$\Sigma(\mathrm{X} - \overline{\mathrm{X}})^2$$

8

# Η διαδικασία της ανάλυσης διακύμανσης

- Ένα απλό παράδειγμα:

| Συνθήκη α | Συνθήκη β | Συνθήκη γ |
|---|---|---|
| 1 | 3 | 8 |
| 2 | 4 | 9 |
| 3 | 5 | 10 |
| 4 | 6 | 11 |
| 5 | 7 | 12 |

- Ο μέσος όρος του συνόλου των τιμών (λαμβάνοντας υπόψη όλες τις μετρήσεις της έρευνας) είναι **6** και το άθροισμα των τετραγώνων του συνόλου των τιμών ($SS_{total}$) είναι **160**

9

# Η διαδικασία της ανάλυσης διακύμανσης

| Συνθήκη α | | | Συνθήκη β | | | Συνθήκη γ | | |
|---|---|---|---|---|---|---|---|---|
| $X$ | $X - \overline{X}$ | $(X - \overline{X})^2$ | $X$ | $X - \overline{X}$ | $(X - \overline{X})^2$ | $X$ | $X - \overline{X}$ | $(X - \overline{X})^2$ |
| 1 | -2 | 4 | 3 | -2 | 4 | 8 | -2 | 4 |
| 2 | -1 | 1 | 4 | -1 | 1 | 9 | -1 | 1 |
| 3 | 0 | 0 | 5 | 0 | 0 | 10 | 0 | 0 |
| 4 | 1 | 1 | 6 | 1 | 1 | 11 | 1 | 1 |
| 5 | 2 | 4 | 7 | 2 | 4 | 12 | 2 | 4 |
| $\overline{X}$ = 3 | | | $\overline{X}$ = 5 | | | $\overline{X}$ = 10 | | |
| $\Sigma(X - \overline{X})^2 = 10$ | | | $\Sigma(X - \overline{X})^2 = 10$ | | | $\Sigma(X - \overline{X})^2 = 10$ | | |

10

## Η διαδικασία της ανάλυσης διακύμανσης

- Βασισμένοι στους υπολογισμούς του πίνακα της προηγούμενης διαφάνειας έχουμε:
- Το άθροισμα των τετραγώνων εντός των ομάδων ($SS_{within}$) είναι 30 (10+10+10)
- Το άθροισμα των τετραγώνων μεταξύ των συνθηκών ($SS_{between}$) είναι 130:

  *Αυτό προκύπτει αν πάρουμε μόνο τους τρεις μέσους όρους (3, 5 και 10), όπου διαπιστώνουμε ότι έχουν μέσο όρο 6 και άθροισμα τετραγώνων 26. Αυτές βεβαίως δεν είναι αρχικές τιμές αλλά μέσοι όροι και, εφόσον ο κάθε μέσος όρος υπολογίζεται από 5 τιμές, χρειάζεται να πολλαπλασιάσουμε το 26 επί 5 προκειμένου να πάρουμε ένα δείκτη της διακύμανσης των τιμών (και όχι των μέσων όρων) μεταξύ των συνθηκών.*
- Προσέξτε ότι:

$$SS_{total} = SS_{between} + SS_{within} \qquad \textbf{(160=130+30)}$$

11

## Η διαδικασία της ανάλυσης διακύμανσης

- Το επόμενο βήμα είναι ο υπολογισμός των βαθμών ελευθερίας για το σύνολο του δείγματος ($df_{total}$), εντός των ομάδων ($df_{within}$), και μεταξύ των συνθηκών ($df_{between}$).
- Ακριβώς όπως επιμερίσαμε τα αθροίσματα των τετραγώνων και τους βαθμούς ελευθερίας, μπορούμε να υπολογίσουμε και τη διακύμανση μεταξύ και εντός των συνθηκών της έρευνας:
  - **Διακύμανση μεταξύ των ομάδων**: Το ποσοστό της συνολικής διακύμανσης που οφείλεται στη διακύμανση των τιμών μέσα σε κάθε ομάδα (συνθήκη)
  - **Διακύμανση εντός των ομάδων**: Το ποσοστό της συνολικής διακύμανσης που οφείλεται στη διακύμανση των τιμών μέσα σε κάθε ομάδα (συνθήκη)

12

## Η διαδικασία της ανάλυσης διακύμανσης

- Τέλος, αν συγκρίνουμε τη διακύμανση μεταξύ των συνθηκών με τη διακύμανση εντός των ομάδων, θα πάρουμε ένα στατιστικό δείκτη που αποκαλύπτει τις συστηματικές διαφορές μεταξύ των συνθηκών (αν υπάρχουν). Ο στατιστικός αυτός δείκτης ονομάζεται **πηλίκο της διακύμανσης** ή **F-τιμή** (variance ratio ή F):

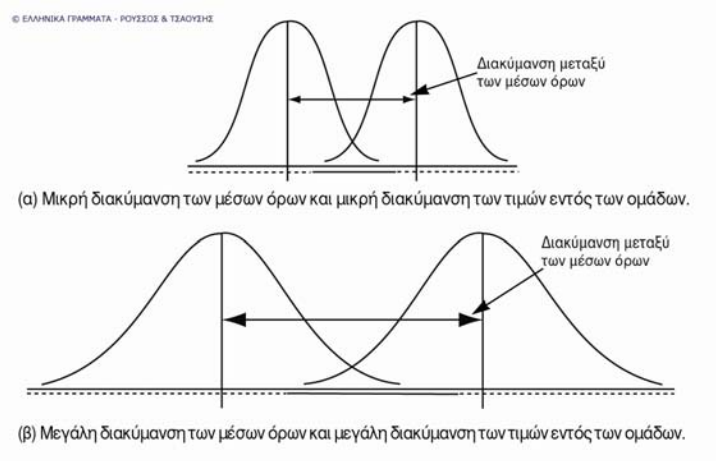
$$F = \frac{\text{Διακύμανση μεταξύ των συνθηκών}}{\text{Σφάλμα}}$$

Αυτό το πηλίκο θα μπορούσε επίσης να εκφραστεί ως εξής:

$$F = \frac{\text{Συστηματικές διαφορές + Σφάλμα}}{\text{Σφάλμα}}$$

13

## Οι δύο ανεξάρτητες εκτιμήσεις της διακύμανσης




Η σχέση μεταξύ του μεγέθους της διακύμανσης και του μεγέθους της διαφοράς των μέσων όρων και η επίδρασή τους στη στατιστική σημαντικότητα του F

14