κεφάλαιο 17

ΑΠΛΗ ΑΝΑΛΥΣΗ ΔΙΑΣΠΟΡΑΣ (ANOVA) ΓΙΑ ΑΝΕΞΑΡΤΗΤΑ ΔΕΙΓΜΑΤΑ
(SIMPLE ANALYSIS OF VARIANCE FOR INDEPENDENT SAMPLES)

17.1 Βασικές Έννοιες.

Στο παρόν κεφάλαιο περιγράφεται αναλυτικά η **Απλή Ανάλυση Διασποράς για «Ανεξάρτητα Δείγματα» (Simple Analysis of Variance, ANOVA, for Independent Samples)** ή αλλιώς «Μεταξύ Υποκειμένων» (Between Subjects) ή «Μεταξύ Ομάδων» (Between Groups). Αποτελεί γενίκευση του ελέγχου t για ανεξάρτητα δείγματα (t-tes for independent samples) και χρησιμοποιείται για τη σύγκριση πολλών (k) ανεξάρτητων δειγμάτων μιας ποσοτικής μεταβλητής. Η ANOVA καθιερώθηκε ως μέθοδος ανάλυσης πειραμάτων από τον Sir Ronald Fisher (1925).

Στην **απλή ANOVA (simple ANOVA, one-way ANOVA)** εξετάζεται η σχέση (επίδραση) της ανεξάρτητης μεταβλητής με την (στην) εξαρτημένη μεταβλητή. Η **ανεξάρτητη μεταβλητή (independent variable, factor)** είναι κατηγορική (categorical) και ταξινομεί τους συμμετέχοντες (participants, subjects) σε δείγματα (ομάδες, επίπεδα, κατηγορίες). Η **εξαρτημένη μεταβλητή (dependent variable)** είναι η ποσοτική ιδιότητα που μετράμε σε κάθε μέλος του δείγματος.

Με τον **έλεγχο t για ανεξάρτητα δείγματα** (Κεφ. 14.3) συγκρίνουμε τους μέσους 2 ανεξάρτητων δειγμάτων μιας ποσοτικής μεταβλητής. **Παράδειγμα**, η σύγκριση αγοριών - κοριτσιών στην ισορροπία. Ανεξάρτητη μεταβλητή εδώ είναι η «ομάδα» με 2 επίπεδα και εξαρτημένη η μετρήσιμη ιδιότητα (measure) «ισορροπία».

Με την **απλή ANOVA για ανεξάρτητα δείγματα** συγκρίνουμε τους μέσους 3+ ανεξάρτητων δειγμάτων μιας ποσοτικής μεταβλητής. **Παράδειγμα**, η σύγκριση δρομέων - αλτών - ριπτών στη δύναμη. Ανεξάρτητη μεταβλητή εδώ είναι το «άθλημα» με 3 επίπεδα (ομάδες) και εξαρτημένη μεταβλητή η μετρήσιμη ιδιότητα (measure) «δύναμη».

Με την ANOVA εξετάζουμε **ερευνητικά προβλήματα (research problems)** όπως η επίδραση διδακτικών μεθόδων στη μάθηση, η αποτελεσματικότητα τεχνικών προπόνησης στην αθλητική απόδοση, η βελτίωση της υγείας μέσω προγραμμάτων άσκησης.

Το ζητούμενο σε ερευνητικούς σχεδιασμούς ανεξάρτητων δειγμάτων είναι να εκτιμηθεί κατά πόσο *οι **διαφορές** μεταξύ των δειγμάτων (ομάδων) είναι **πραγματικές*** (ισχύουν σε πληθυσμιακό επίπεδο*) ή **τυχαίες*** (οφείλονται σε δειγματοληπτικό σφάλμα).

Με την ANOVA ελέγχουμε αν η **διαφορά** μεταξύ των k πληθυσμιακών μέσων (μ) είναι:

(i) στατιστικώς σημαντική (statistical significance), δηλαδή **πιθανή** και

(ii) ουσιαστικώς σημαντική (substantive significance), δηλαδή **αξιόλογη**.

Στον ερευνητικό σχεδιασμό περιλαμβάνεται **η μηδενική υπόθεση (null hypothesis)**

H_0: $\mu_1 = \mu_2 = ... = \mu_k$ «δεν υπάρχει διαφορά μεταξύ των k πληθυσμιακών μέσων»
με **εναλλακτική υπόθεση (alternative hypothesis)**

H_A: $\mu_1 \neq \mu_2 \neq ... \neq \mu_k$ «υπάρχει διαφορά μεταξύ των k πληθυσμιακών μέσων».

Ο έλεγχος της υπόθεσης H_0 γίνεται με **α ολική πιθανότητα σφάλματος τύπου Ι** (π.χ. 5% ή 1%) για όλες τις δυνατές συγκρίσεις μεταξύ των k πληθυσμιακών μέσων.

Παραδοχές της ANOVA για Ανεξάρτητα Δείγματα (αναλυτικά Κεφ. 17.10)

Η ANOVA για ανεξάρτητα δείγματα εδράζεται στην θεμελιώδη προϋπόθεση ότι τα υπό σύγκριση **δείγματα** είναι **τυχαία (random samples).** Αυτή η προϋπόθεση σπάνια τηρείται, καθότι η επιλογή τυχαίων δειγμάτων είναι πρακτικά πολύ δύσκολη αν όχι ανέφικτη (Κεφ. 1.3). Δεδομένης αυτής της σοβαρής «αδυναμίας» για «αναγκαστική» χρήση μη τυχαίων δειγμάτων, η ANOVA για ανεξάρτητα δείγματα βασίζεται σε 3 **παραδοχές (assumptions)**.

1) Ανεξαρτησία Παρατηρήσεων (independence of observations): μη συσχέτιση μεταξύ των παρατηρήσεων σε κάθε δείγμα και μεταξύ των δειγμάτων. Αυτό σημαίνει ότι κάθε αποτέλεσμα μέτρησης θα πρέπει να είναι ανεξάρτητο από κάθε άλλο στη μετρητική διαδικασία. Αν ο ερευνητικός σχεδιασμός είναι πλήρης και εφαρμοσθεί ορθά, η παραδοχή ισχύει. Αν, ο ερευνητικός σχεδιασμός είναι ελλιπής ή δεν εφαρμοσθεί ορθά, τότε «άδηλες» πειραματικές «συνθήκες» μπορεί να τη «νοθεύσουν». Όταν τα δεδομένα συσχετίζονται, η ANOVA καθίσταται μη έγκυρη. Αξιολογείται με γραφήματα των υπολοίπων (residual plots) και για τυχόν σειριακή συσχέτιση (serial correlation) με Durbin-Watson.

2) Κανονικότητα Κατανομών (normal distributions): οι συγκρινόμενοι πληθυσμοί της υπό ανάλυση μεταβλητής έχουν κατανομή κανονική. Αν παραβιαστεί έντονα το κριτήριο F καθίσταται μεροληπτικό. Βέβαια, η ANOVA στηρίζεται στη δειγματική κατανομή του μέσου ή των διαφορών που με βάση το θεώρημα των κεντρικών ορίων είναι συμμετρική. Ο έλεγχος F είναι γενικά ισχυρός (robust) σε μέτριες λοξότητες σε σχετικά ίσα δείγματα. Αξιολογείται με γραφήματα (π.χ. ιστόγραμμα) και έλεγχο κανονικότητας (π.χ. Shapiro-Wilk) σε επαρκή δείγματα. Σε έντονη λοξότητα μετασχηματίζουμε τη μεταβλητή (π.χ. logX) και, αν η λοξότητα παραμένει, επιλέγουμε μη παραμετρική ANOVA Kruskal-Wallis (Κεφ. 16.3).

3) **Ομοιογένεια Διασπορών (homogeneity of variance)**: οι συγκρινόμενοι πληθυσμοί της υπό ανάλυση μεταβλητής έχουν ίδια διασπορά (ομοιογένεια). Αξιολογείται με γραφήματα και τον έλεγχο Levene. Αν ο Levene δείξει σημαντική ανομοιογένεια, τρέχουμε ANOVA Welch ή Brown-Forsythe. Ο έλεγχος F είναι ευαίσθητος σε ουσιαστικές ανομοιογένειες ακόμα και σε ίσα δείγματα, με αποτέλεσμα να οδηγεί σε θετική μεροληψία, δηλαδή σε πολύ συχνή στατιστική σημαντικότητα (απόρριψη της μηδενικής υπόθεσης). Τα πολύ μικρά και πολύ άνισα δείγματα επιδεινώνουν τη μεροληψία αυτή, αλλά, γενικά, μικρές ετερογένειες δεν αποτελούν πρόβλημα στην ANOVA, που ελέγχεται και ως ομοιογένεια υπολοίπων.

Οι επιπτώσεις της παραβίασης των παραδοχών αυτών είναι θέμα βαθμού. Τα μικρά και πολύ άνισα δείγματα αυξάνουν τη σοβαρότητα των επιπτώσεων αυτών. Γενικά, όμως, ο έλεγχος F της ANOVA είναι σταθερός σε μικρές έως μέτριες αποκλίσεις από την κανονικότητα των κατανομών και την ομοιογένεια των διασπορών (Cohen, 1988, σελ. 19-20, 273-274).

17.2 Ανάλυση της Διασποράς και Έλεγχος F.

Η **Ανάλυση Διασποράς** (ANOVA) βασίζεται στα εξής **θεμελιώδη στατιστικά μεγέθη**:

- **Απόκλιση (Deviation) =** Τιμή – Μέσος,
- **Άθροισμα Τετραγώνων (Sum of Squares)** = SS = Σ(Τιμή – Μέσος)^2
- **Διασπορά (Variance, Mean Squares) = MS = s^2 =** SS / df = SS / (N – 1).

Για **παράδειγμα** σε ένα μικρό δείγμα με τιμές 2, 4, 6 και μέσο 4 έχουμε

- **Αποκλίσεις:** 2 – 4 = –2, 4 – 4 = 0, 6 – 4 = 2
- **Άθροισμα Τετραγώνων: SS** = (–2)^2 + 0^2 + 2^2 = 4 + 0 + 4 = **8**
- **Διασπορά (αμερόληπτη): s^2** = MS = SS / df = 8 / (3 – 1) = 8 / 2 = **4**.

Στην **ANOVA** έχουμε το **Άθροισμα Τετραγώνων Ολικό (total)** που επιμερίζεται σε «**Μεταξύ των Ομάδων**» και σε «**Εντός των Ομάδων**».

Το **άθροισμα τετραγώνων μεταξύ των ομάδων (between groups)** ή εναλλακτικά **μεταξύ των υποκειμένων (between subjects)** αφορά τις διαφορές μεταξύ των υπό σύγκριση δειγμάτων, δηλαδή τις διαφορές των αντίστοιχων πληθυσμιακών μέσων (μ) και ενίοτε λέγεται και **άθροισμα τετραγώνων υπόθεση (hypothesis)**. Αποτελεί το προς εξήγηση (συστηματικό) τμήμα της διασποράς της εξαρτημένης μεταβλητής.

Το **άθροισμα τετραγώνων εντός των ομάδων (within groups)** ή εναλλακτικά **εντός των υποκειμένων (within subjects)** αφορά το δειγματοληπτικό (τυχαίο) σφάλμα, δηλαδή τις διαφορές των τιμών εντός κάθε ομάδας και για το λόγο αυτό λέγεται και **άθροισμα τετραγώνων σφάλμα (error)**. Αποτελεί το μη εξηγήσιμο (υπόλοιπο) τμήμα της διασποράς της εξαρτημένης μεταβλητής.

Οι **αποκλίσεις** για τα αθροίσματα αυτά είναι:

- **Απόκλιση** για **Ολικό Άθροισμα** (SS_{TOTAL}): Τιμή – Μέσος Ολικός
- **Απόκλιση** για **Μεταξύ Ομάδων** (SS_{BETWEEN}): Μέσος Ομάδας – Μέσος Ολικός
- **Απόκλιση** για **Εντός Ομάδων** (SS_{WITHIN}): Τιμή – Μέσος Ομάδας.

Σε σύγκριση j = 1, 2, 3, . . . , k ανεξάρτητων δειγμάτων μιας μεταβλητής ιδιότητας X

με πλήθη δειγμάτων $N_1 \quad N_2 \quad N_3 \ldots \quad N_k$

με μέσους δειγμάτων $\bar{X}_1 \quad \bar{X}_2 \quad \bar{X}_3 \ldots \quad \bar{X}_k$

και ολικό μέσο $\bar{X}. = (X_1 + X_2 + \ldots + X_N) / N = \Sigma X / N$

ισχύει

$$(X - \bar{X}.) = (\bar{X}_j - \bar{X}.) + (X - \bar{X}_j) \qquad (17.1)$$

- **[Τιμή – Μέσος Ολικός] =**
  [Μέσος Ομάδας – Μέσος Ολικός] + [Τιμή – Μέσος Ομάδας]
- **[Απόκλιση Ολική] =**
  [Απόκλιση Μεταξύ των Μέσων] + [Απόκλιση εντός των Ομάδων]

Αν πάρουμε το τετράγωνο της 17.1 και αθροίσουμε για τις k ομάδες και τις N τιμές έχουμε

$$\mathbf{SS}_{\textbf{ΟΛΙΚΟ}} = \mathbf{SS}_{\textbf{ΜΕΤΑΞΥ ΟΜΑΔΩΝ}} + \mathbf{SS}_{\textbf{ΕΝΤΟΣ ΟΜΑΔΩΝ}}$$

$$SS_{\text{TOTAL}} = SS_{\text{BETWEEN GROUPS}} + SS_{\text{WITHIN GROUPS}}$$

$$\sum (X - \bar{X}.)^2 = \sum N_j (\bar{X}_j - \bar{X}.)^2 + \sum (X - \bar{X}_j)^2 \qquad (17.2)$$

Ο όρος στο αριστερό σκέλος της 17.2 δίνει το άθροισμα των τετραγώνων των αποκλίσεων όλων των N τιμών X από τον ολικό μέσο και συμβολίζεται ως $\mathbf{SS_T}$ (total, ολικό).

Ο 1ος όρος στο δεξί τμήμα της 17.2 δίνει το άθροισμα των τετραγώνων των αποκλίσεων των k μέσων από τον ολικό μέσο και συμβολίζεται ως $\mathbf{SS_B}$ (between, μεταξύ ομάδων).

Ο 2ος όρος στο δεξί τμήμα της 17.2 δίνει το άθροισμα των τετραγώνων των αποκλίσεων των τιμών X από τον μέσο κάθε ομάδας και συμβολίζεται ως $\mathbf{SS_W}$ (within, εντός ομάδων).

Για κάθε όρο στη ισότητα 17.2 αντιστοιχούν οι **Βαθμοί ελευθερίας (degrees of freedom)**:

Ο αριθμός των ανεξάρτητων «μονάδων πληροφορίας» σε ένα δείγμα στην εκτίμηση της παραμέτρου ή του υπολογισμού ενός στατιστικού. Πιο απλά είναι ο αριθμός των μονάδων πληροφορίας (π.χ. ατόμων, μέσων) μείον τον αριθμό των (υπό εκτίμηση πληθυσμιακών) παραμέτρων τού μοντέλου τής ανάλυσης (Walker, 1940):

- **Βαθμοί Ελευθερίας (df)** =
  [Αριθμός Ανεξάρτητων «Μονάδων»] – [Αριθμός Παραμέτρων Μοντέλου].
- **Άθροισμα Τετραγώνων (SS)** και **Βαθμοί Ελευθερίας (df):**
  Ολικό (Total)

$$SS_T = \sum (X - \bar{X}.)^2 \qquad df_T = N - 1 \tag{17.3α}$$

Μεταξύ Ομάδων (Between Groups)

$$SS_B = \sum N_j (\bar{X}_j - \bar{X}.)^2 \qquad df_B = k - 1 \tag{17.3β}$$

Εντός Ομάδων (Within Groups)

$$SS_W = \sum (X - \bar{X}_j)^2 \qquad df_W = N - k \tag{17.3γ}$$

- **Διασπορά (s^2, Mean Squares, MS = SS / df):**

  Μεταξύ των k ομάδων (between groups)

$$MS_B = \frac{SS_B}{df_B} = \frac{\sum N_j (\bar{X}_j - \bar{X}.)^2}{k-1} \tag{17.4α}$$

  Εντός των k ομάδων (within groups)

$$MS_W = \frac{SS_W}{df_W} = \frac{\sum (X - \bar{X}.)^2}{N-k} \tag{17.4β}$$

- **Κριτήριο (πηλίκο) F**

$$F = \frac{s_B^2}{s_W^2} = \frac{MS_B}{MS_W} \tag{17.5}$$

Ακολουθούν η στατιστική (statistical) και ουσιαστική (substantive) σημαντικότητα.

- **Στατιστική Σημαντικότητα** (statistical significance) **του F (έλεγχος):**

  Με **επίπεδο πιθανότητας** (σφάλματος τύπου Ι) **α (π.χ. 0.05, 0.01)** και **βαθμούς ελευθερίας $df_1 = df_{Between} = k - 1$** και **$df_2 = df_{Within} = N - k$** η **κρίσιμη τιμή F** (Πίν. 21.Η) είναι **$F_{critical} = F_{(\alpha,\, dfB,\, dfW)}$** οπότε

- **αν F ≥ F_c απορρίπτεται η υπόθεση H_0 και συμπεραίνουμε ότι οι k πληθυσμιακοί μέσοι (μ) διαφέρουν σημαντικά** (τουλάχιστον 2 εξ' αυτών),
- **αν F < F_c γίνεται αποδεκτή η υπόθεση H_0 και συμπεραίνουμε** ότι **οι k πληθυσμιακοί μέσοι (μ) δε διαφέρουν σημαντικά**.

Τι δείχνει η στατιστική σημαντικότητα του F στην ANOVA;

Ο έλεγχος της στατιστικής σημαντικότητας του F μας δίνει απλά την πιθανότητα τυχαίας εμφάνισης ενός F ίσου ή μεγαλύτερου από αυτό που μας έδωσαν τα δεδομένα του δείγματος δοσμένης της μηδενικής υπόθεσης. Πιο απλά, θεωρώντας ότι η μηδενική υπόθεση ισχύει (ότι δηλαδή δεν υπάρχουν διαφορές μεταξύ των συγκρινόμενων k πληθυσμιακών μέσων), η παραχθείσα τιμή σημαντικότητας (significance, observed p-value) μας δίνει την πιθανότητα να έχουμε ένα F τουλάχιστον τόσο ακραίο όσο αυτό που προέκυψε (observed F). Δεν μας λέει πόσο μεγάλο ή πρακτικά αξιόλογο είναι το τελικό αποτέλεσμα της ANOVA, δηλαδή πόσο ουσιαστικό είναι το μέγεθος της επίδρασης, της διαφοράς (effect size). Η **ορθή πρακτική** προβλέπει καλή γνώση των δεδομένων μας, έγκυρη εκτίμηση του μεγέθους της επίδρασης της ανεξάρτητης μεταβλητής στην εξαρτημένη και δόκιμη ερμηνεία του στο πλαίσιο του ερευνητικού πεδίου (Nuzzo, 2014). Με τη στατιστική σημαντικότητα, που έχει αμφισβητηθεί ως κριτήριο της μηδενικής υπόθεσης (Carver, 1978), θα πρέπει να δίνουμε και τη ουσιαστική σημαντικότητα, δηλαδή το μέγεθος της επίδρασης / διαφοράς (Sullivan & Feinn, 2012).

- **Ουσιαστική Σημαντικότητα** (substantive significance): **Μέγεθος Επίδρασης** (effect size).

Η ουσιαστική σημαντικότητα αξιολογείται με το **μέγεθος επίδρασης (effect size)** της ανεξάρτητης μεταβλητής (independent variable, π.χ. πειραματική αγωγή, ομάδες) πάνω στην εξαρτημένη μεταβλητή (dependent variable). Στην ουσία γίνεται εκτίμηση της σχέσης μεταξύ ανεξάρτητης και εξαρτημένης μεταβλητής με το **πηλίκο συσχέτισης (correlation ratio, R, η)** (Peters & Van Voorhis, 1940, σε Friedman, 1968, σελ. 246).

Στην ANOVA το μέγεθος της επίδρασης αξιολογείται (i) για το **δείγμα** με το «τετράγωνο του πηλίκου συσχέτισης» (squared correlation ratio, R^2, η^2), και (ii) για τον **πληθυσμό** με το «ωμέγα τετράγωνο» (omega squared, ω^2) (Hays, 1963, σελ. 382). Και οι 2 δείκτες δίνουν το ποσοστό (%) της διασποράς της εξαρτημένης που εξηγήθηκε από την ανεξάρτητη.

Με **k** τον αριθμό των δειγμάτων (ομάδων), **N** το μέγεθος του δείγματος, **SS** τα αθροίσματα τετραγώνων και **F** το κριτήριο της ANOVA έχουμε:

- **Μέγεθος Επίδραση για το δείγμα (R squared, eta squared)**

$$R^2 = \eta^2 = \frac{SS_{between}}{SS_{total}} = \frac{(k-1)*F}{(k-1)*F+(N-k)} \qquad (17.6\alpha)$$

¤ Μέγεθος Επίδρασης για τον πληθυσμό (omega squared)

$$\omega^2 = \frac{SS_{between} - (k-1) * MS_{within}}{SS_{total} + MS_{within}} = \frac{(k-1)*(F-1)}{(k-1)*(F-1) + \mathrm{N}} \qquad (17.6\beta)$$

¤ Αξιολόγηση του Μεγέθους Επίδρασης (effect size) η^2 & ω^2 - Κλίμακα Cohen

Η **κλίμακα Cohen** (1988, σελ. 285-288) δίνει το μέγεθος της επίδρασης στην ANOVA ως ποσοστό (%) της διασποράς της εξαρτημένης που εξηγήθηκε από την ανεξάρτητη (ομάδες)

με **κατώτερα όρια** **0.01 = μικρή** **0.06 = μέτρια** **0.14 = μεγάλη επίδραση**

και **διαστήματα** **1 % – 5.9 %** **6 % – 13.9 %** **14% – & άνω**.

Παράδειγμα 17.1 - Ανάλυση Διασποράς (ANOVA): Η Κύρια Διαδικασία.

Σε έρευνα μετρήθηκαν 3 ομάδες ατόμων ($N_1 = 3$, $N_2 = 4$, $N_3 = 5$) σε μια «Κινητική Ικανότητα» και πήραν τις τιμές X που δίνονται στον Πίν. 17.1α μαζί με τα ενδιάμεσα στατιστικά μεγέθη. Ζητείται να ελεγχθεί στο α = 0.05 η υπόθεση ότι «δεν υπάρχει διαφορά μεταξύ των 3 πληθυσμιακών μέσων» (H_0: $\mu_1 = \mu_2 = \mu_3$).

Πίνακας 17.1α - Ανάλυση Διασποράς (ANOVA): αθροίσματα τετραγώνων (SS)

Ομάδα (Δείγμα) j	Άτομο (Subject) i	Μεταβλητή Ιδιότητα X	ΟΛΙΚΟ (Total) $(X-\bar{X}.)^2$	ΕΝΤΟΣ (Within) $(X-\bar{X}_j)^2$	ΜΕΤΑΞΥ (Between) $N_j\,(\bar{X}_j-\bar{X}.)^2$
	1	2	$(2-6)^2 = 16$	$(2-4)^2 = 4$	
1	2	4	$(4-6)^2 = 4$	$(4-4)^2 = 0$	$3(4-6)^2 =$ **12**
$N_1 = 3$	3	6	$(6-6)^2 = 0$	$(6-4)^2 = 4$	
	1	3	$(3-6)^2 = 9$	$(3-5)^2 = 4$	
	2	5	$(5-6)^2 = 1$	$(5-5)^2 = 0$	
2	3	5	$(5-6)^2 = 1$	$(5-5)^2 = 0$	$4(5-6)^2 =$ **4**
$N_2 = 4$	4	7	$(7-6)^2 = 1$	$(7-5)^2 = 4$	
	1	7	$(7-6)^2 = 1$	$(7-8)^2 = 1$	
	2	7	$(7-6)^2 = 1$	$(7-8)^2 = 1$	
3	3	8	$(8-6)^2 = 4$	$(8-8)^2 = 0$	$5(8-6)^2 =$ **20**
	4	9	$(9-6)^2 = 9$	$(9-8)^2 = 1$	
$N_3 = 5$	5	9	$(9-6)^2 = 9$	$(9-8)^2 = 1$	
N = 12	Άθρ. Σ =	**72**	**SS_T = 56**	**SS_W = 20**	**SS_B = 36**

Στο Κεφ. 17.11 η ανάλυση αυτή γίνεται μέσω πολλαπλής γραμμικής παλινδρόμησης.

Υπολογιστικά Βήματα (Κύρια Ανάλυση)

1. Μέσοι των 3 ομάδων και ολικός μέσος

$\overline{X}_1 = 12 / 3 = \mathbf{4}$ $\overline{X}_2 = 20 / 4 = \mathbf{5}$ $\overline{X}_3 = 40 / 5 = \mathbf{8}$ $\overline{X}. = 72 / 12 = \mathbf{6}$

2. Άθροισμα Τετραγώνων *Μεταξύ Ομάδων* (Sum of Squares Between)

$$SS_B = \sum N_j(\overline{X}_j - \overline{X}.)^2 = \sum [\, N_j \,(\text{ Μέσος Ομάδας} - \text{Μέσος Ολικός }]^2 =$$

$$= 3(4-6)^2 + 4(5-6)^2 + 5(8-6)^2 = 12 + 4 + 20 = \mathbf{36}$$

3. Άθροισμα τετραγώνων *Εντός Ομάδων* (Sum of Squares Within)

$$SS_W = \sum X - \overline{X}_j \;^2 = \text{ άθροισμα (τιμή X − μέσος ομάδας)}^2 =$$

$= [\,(2-4)^2 + (4-4)^2 + (6-4)^2\,]$ (ομάδα 1)

$+ \;[\,(3-5)^2 + (5-5)^2 + (5-5)^2 + (7-5)^2\,]$ (ομάδα 2)

$+ \;[\,(7-8)^2 + (7-8)^2 + (8-8)^2 + (9-8)^2 + (9-8)^2]$ (ομάδα 3)

$= \;(4+0+4) + (4+0+0+4) + (1+1+0+1+1) = \mathbf{20}$

4. Βαθμοί Ελευθερίας (Degrees of Freedom)

Ολικοί (Total)	$df_T = N - 1 = 12 - 1$	$= 11$
Μεταξύ Ομάδων (Between)	$df_B = k - 1 = 3 - 1$	$= 2$
Εντός Ομάδων (Within)	$df_W = N - k = 12 - 3$	$= 9$

5. Μέσα Τετράγωνα (Mean squares, Variance, Διασπορά)

Μεταξύ Ομάδων (Between) $MS_B = SS_B / df_B = 36 / 2 = \mathbf{18}$ (τύπος 17.4α)

Εντός Ομάδων (Within) $MS_W = SS_W / df_W = 20 / 9 = \mathbf{2.2222}$ (τύπος 17.4β)

6. Κριτήριο (πηλίκο) F

$$F = MS_B / MS_W = 18 / 2.2222 = 8.10 \qquad \text{(τύπος 17.5)}$$

7. Στατιστική Σημαντικότητα (statistical significance) **του F (έλεγχος)**

Με α = 0.05 και βαθμούς ελευθερίας $df_B = 2$ & $df_W = 9$ (Πίν. 21.Η)
η κρίσιμη τιμή F είναι $F_c = F_{(0.05,\,2,\,9)} = 4.26$ οπότε,
επειδή $F = 8.10 > F_c = 4.26$, με πιθανότητα σφάλματος τύπου Ι ίση με 5%, απορρίπτουμε τη μηδενική υπόθεση (H_0: μ1 = μ2 = μ3) και **συμπεραίνουμε** ότι οι 3 πληθυσμιακοί μέσοι της κινητικής ιδιότητας Χ διαφέρουν στατιστικώς σημαντικά.

Τα αποτελέσματα της ANOVA δίνονται περιληπτικά στον Πίν. 17.1β στη μορφή που δίνονται σε στατιστικά προγράμματα, όπως το SPSS, το SAS κ.ά.

Πόσο μεγάλη είναι, όμως, αυτή η στατιστικώς σημαντική διαφορά ως μέγεθος;

Πίνακας 17.1β – Περιληπτικά Στατιστικά τής ANOVA

Πηγή (source)	Άθροισμα Τετραγώνων (SS)	Βαθμοί Ελευθερίας (df)	Διασπορά (MS)	Κριτήριο F	Σημαντ. P*	EffectSize η^2	ω^2
Μεταξύ (B)	36	2	18	8.10	< 0.01	0.64	0.54
Εντός (W)	20	9	2.22				
Ολικό(T)	56	11					

* Πιθανότητα (p-value) = 0.0097. η^2 (για το δείγμα), ω^2 (για τον πληθυσμό).

8. Μέγεθος Επίδρασης, Διαφοράς (Effect Size): Ουσιαστική Σημαντικότητα

Με βάση τα στατιστικά του Πίν. 17.1β υπολογίζουμε τα μεγέθη επίδρασης.

Για το **δείγμα** (τύπος 17.6α): η^2 = SS between / SS within = 26 / 56 = 0.64

ή εναλλακτικά $\eta^2 = \dfrac{(k-1)*F}{(k-1)*F+(N-k)} = \dfrac{(3-1)*8.10}{(3-1)*8.10+(12-3)} = 0.643$

Για τον **πληθυσμό** (τύπος 17.6β):

$$\omega^2 = \frac{SS_B-(k-1)*MS_W}{SS_T+MS_W} = \frac{36-(3-1)*2.22}{56+2.22} = 0.542$$

ή εναλλακτικά

$$\omega^2 = \frac{(k-1)*(F-1)}{(k-1)*(F-1)+\mathrm{N}} = \frac{(3-1)*(8.10-1)}{(3-1)*(8.10-1)+12} = 0.542$$

Ερμηνεία – Αξιολόγηση του μεγέθους επίδρασης:

Η ανεξάρτητη μεταβλητή «ομάδα» ανέδειξε το 64% της διασποράς της εξαρτημένης μεταβλητής «Κινητική Ικανότητα Χ» στο δείγμα και το 54% στον πληθυσμό. Πιο απλά η διαφορά μεταξύ των 3 ομάδων εξήγησε το 64% της διασποράς της ιδιότητας Χ (εκτίμηση για τον πληθυσμό 54%). Τα ποσοστά αυτά είναι πολύ μεγάλα, αφού η κλίμακα Cohen (1988, σελ. 283) δίνει ως μεγάλη επίδραση από 14% και πάνω. Η **πρακτική αξία** των μεγεθών αυτών αξιολογείται πάντα με βάση το ουσιαστικό περιεχόμενο της έρευνας.

9. Συγκρίσεις Κατά Ζεύγη ή / και Συνδυαστικά

Στη συνέχεια θα εξετάσουμε ποιες από τι 3 ομάδες διαφέρουν μεταξύ τους κατά ζεύγη. Αυτή η διαδικασία γίνεται μέσω πολλαπλών συγκρίσεων (multiple comparisons). Εδώ έχουμε κυρίως 2 διακριτές επιλογές: (i) να είχαμε προσχεδιάσει (pre-planned) μόνο τη σύγκριση της $1^{ης}$ ομάδας (π.χ. ελέγχου) με τη $2^{η}$ (π.χ. πειραματική 1) και την $3^{η}$ (π.χ. πειραματική 2) ξεχωριστά (οπότε αγνοούμε τη στατιστική σημαντικότητα του F), ή (ii) να αφήναμε μετά την ANOVA (post hoc) τη διερεύνηση των διαφορών κατά ζεύγη ή και συνδυαστικά. Επίσης (iii) θα μπορούσαμε να εξετάσουμε αν υπάρχουν τάσεις (trends) μεταξύ των 3 διαδοχικών μέσων (μ).

17.3 Πολλαπλές Συγκρίσεις (Multiple Comparisons).

Οι πολλαπλές συγκρίσεις (multiple comparisons) αποτελούν καταληκτικό μέρος της ANOVA στο πλαίσιο της διερεύνησης επί μέρους πτυχών της έρευνας, δηλαδή συγκρίσεων μεταξύ των πληθυσμιακών μέσων. Στόχος τής ανάλυσης αυτής είναι να διαπιστωθεί ποιος μέσος διαφέρει σημαντικά από ποιον, κατά ζεύγη ή και συνδυαστικά. Σε έρευνες σύγκρισης δειγμάτων ελέγχουμε (i) είτε λίγες προσχεδιασμένες συγκρίσεις επιβεβαιωτικά, είτε (ii) αρκετές μετά-ANOVA συγκρίσεις διερευνητικά (Kerlinger, 1992, σελ. 218-219).

1) Προσχεδιασμένες Συγκρίσεις (planned comparisons)

Οι συγκρίσεις αυτές προσχεδιάζονται (planned) στο στάδιο του ερευνητικού σχεδιασμού (research design) και στοχεύουν στη διερεύνηση συγκεκριμένων μόνο συγκρίσεων (comparisons, contrasts), δηλαδή διαφορών μεταξύ 2 μέσων ή συνδυασμών τους με άλλους, σε αντιστοιχία με συγκεκριμένες ερευνητικές υποθέσεις. Οι συγκρίσεις αυτές είναι συνήθως **λίγες** και ή δυνατόν **ορθογώνιες** (**orthogonal**), δηλαδή **ανεξάρτητες μεταξύ τους**, και διερευνώνται **ανεξάρτητα από την στατιστική σημαντικότητα του F της ANOVA**. Με τις προσχεδιασμένες συγκρίσεις «επιβεβαιώνουμε» μερικές ουσιαστικές υποθέσεις (πλευρές) σχετικές με το υπό εξέταση ερευνητικό πρόβλημα (confirmatory analysis).

Οι προσχεδιασμένες συγκρίσεις γίνονται είτε με έλεγχο F είτε με έλεγχο t. Τα αποτελέσματα είναι ίδια καθότι $F = t^2$. Για καλύτερη κατανόηση της διαδικασίας επιλέχθηκε ο έλεγχος t. Ουσιαστικά τρέχουμε έναν **ειδικό έλεγχο t** για κάθε **Ψ σύγκριση** (contrast) 2 μέσων ή συνδυασμών μέσων με τυπικό σφάλμα βασισμένο στη διασπορά «εντός» της ANOVA:

$$t_\psi = \Psi / S_\psi = [\text{σύγκριση}] / [\text{τυπικό σφάλμα}] = [\text{contrast}] / [\text{std. error}]$$

2) Μετά-ANOVA Συγκρίσεις (post-hoc comparisons)

Οι συγκρίσεις αυτές αφορούν τη διερεύνηση όλων ή των περισσότερων δυνατών συγκρίσεων μεταξύ των k πληθυσμιακών μέσων με στόχο να διερευνηθούν διάφορες (εν μέρει ή εν πολλοίς άγνωστες) «πτυχές» ενός ερευνητικού προβλήματος. Στον ερευνητικό σχεδιασμό διατυπώνουμε κάποιες **γενικές ερευνητικές υποθέσεις** και **αν το F** της ANOVA βγει **σημαντικό**, διερευνούμε αναλυτικά (post-hoc) ποιοι μέσοι διαφέρουν μεταξύ τους κατά ζεύγη ή / και συνδυαστικά. Με τις μετά-

ANOVA συγκρίσεις συνήθως διερευνούμε αναλυτικά ένα «άγνωστο» ερευνητικό πρόβλημα (exploratory research).

Οι **κύριοι μετά-ANOVA έλεγχοι** είναι ο έλεγχος **Scheffe** για όλες τις δυνατές συγκρίσεις κατά ζεύγη και συνδυαστικά, ο έλεγχος **Tukey** για πολλαπλές συγκρίσεις κατά ζεύγη, ο έλεγχος **Dunnett** για σύγκριση της **ομάδας ελέγχου** (control group) με κάθε **πειραματική ομάδα** (experimental group), και οι έλεγχοι **Bonferroni** και **Sidak** για συγκρίσεις κατά ζεύγη με διόρθωση του α κάθε σύγκρισης ως προς τον αριθμός τους. Οι μετά-ANOVA έλεγχοι παρέχουν επαρκή προστασία από τη διόγκωση της ολικής πιθανότητας σφάλματος τύπου Ι μιας οικογένειας συγκρίσεων όταν κάθε μία σύγκριση γίνεται π.χ. στο 0.05.

Τι επίπτωση έχουν οι πολλαπλές συγκρίσεις και πως διορθώνεται;

Οι πολλές συγκρίσεις (c) αυξάνουν την **ολική πιθανότητα σφάλματος τύπου Ι** (ολικό α_o). Έτσι κάνουμε **διόρθωση στο α_c κάθε σύγκρισης**, ώστε το **ολικό** να είναι πάντα $\boldsymbol{\alpha_o}$ <5%. Η διόρθωση (adjustment) είναι είτε ταυτόχρονη για όλες τις συγκρίσεις (simultaneous) είτε διαδοχική από σύγκριση σε σύγκριση (sequential) ιεραρχικά (Vagenas, Palaiothodorou, & Knudson, 2017). Η **ταυτόχρονη διόρθωση** είναι η κρατούσα και γίνεται με 2 μεθόδους:

Bonferroni ως $\alpha_{comp} = \alpha_o / m$ ή **Sidak** ως $\alpha_{comp} = 1 - (1 - \alpha_o)^{1/m}$.

Το **ολικό $\boldsymbol{\alpha_o}$** σε **m** συγκρίσεις με κάθε μία στο $\boldsymbol{\alpha_{comp}}$ γίνεται $\alpha_o = 1 - (1 - \alpha_c)^m$.

- Σε **3 συγκρίσεις** με κάθε μία στο $\boldsymbol{\alpha_{comp}}$ **= 0.05** το ολικό α γίνεται $\alpha_o = 1 - (1 - 0.05)^3 = 0.14$.
- Σε **10 συγκρίσεις** με κάθε μία στο $\boldsymbol{\alpha_{comp}}$**= 0.05** το ολικό α γίνεται $\alpha_o = 1 - (1 - 0.05)^{10} = 0.40$.

Το ολικό α γίνεται με 3 συγκρίσεις 14% και 10 συγκρίσεις 40%: (μη αποδεκτό).

- Σε **3 συγκρίσεις** με κάθε μία στο $\boldsymbol{\alpha_{comp}}$**= 0.01** το ολικό α γίνεται $\alpha_o = 1 - (1 - 0.01)^3 = 0.03$.
- Σε **10 συγκρίσεις** με κάθε μία στο $\boldsymbol{\alpha_{comp}}$ **= 0.01** το ολικό α γίνεται $\alpha_o = 1 - (1 - 0.01)^{10} = 0.10$.

Το ολικό α γίνεται με 3 συγκρίσεις 3% (αποδεκτό) και 10 συγκρίσεις 10% (μη αποδεκτό).

Ορθογώνιες Συγκρίσεις (Orthogonal Comparisons)

Δύο συγκρίσεις είναι ορθογώνιες αν (i) το άθροισμα των συντελεστών (coefficients) της κάθε σύγκρισης και (ii) το άθροισμα των γινομένων των συντελεστών τους είναι μηδέν.

Για **παράδειγμα**, σε ανάλυση 3 ομάδων μπορούν να επιλεγούν οι εξής 3 συγκρίσεις (Ψ):

Σύγκριση	Ομάδα 1	Ομάδα 2	Ομάδα 3
Ψ_1: $2\mu_1 - (\mu_2 + \mu_3)$	+2	−1	−1
Ψ_2: $\mu_2 - \mu_3$	0	+1	−1
Ψ_3: $\mu_1 - \mu_2$	+1	−1	0

Οι τιμές +1, −1 και 0 συμβολίζουν τη συμμετοχή (1), την αντίθεση (−1), την απουσία (0):

- Η Ψ_1 αφορά τη σύγκριση της ομάδας 1 με τη 2 και 3 μαζί.
- Η Ψ_2 αφορά τη σύγκριση της ομάδας 2 με την ομάδα 3.
- Η Ψ_3 αφορά τη σύγκριση της ομάδας 1 με την ομάδα 2.

Από τις 3 αυτές συγκρίσεις **ορθογώνιες μεταξύ τους** είναι μόνο η Ψ_1 με τη Ψ_2:

επειδή $(2) * (0) + (-1) * (1) + (-1) * (-1) = 0 + -1 + 1 = 0$.

Τα άλλα 2 ζεύγη συγκρίσεων δίνουν αθροίσματα γινομένων των συντελεστών τους > 0. Αν οι ερευνητικές μας υποθέσεις συμπίπτουν με τις συγκρίσεις Ψ_1 & Ψ_2, τότε έχουμε την ιδανική περίπτωση: 2 συγκρίσεις ανεξάρτητες μεταξύ τους στατιστικά και ερμηνευτικά.

Σε πειράματα με 3 ομάδες μία ομάδα είναι η «ελέγχου» (control group) και οι άλλες 2 «πειραματικές» (experimental groups). Στις περιπτώσεις αυτές 2 υποθέσεις μπορεί να είναι: (i) «ότι οι 2 πειραματικές ομάδες μαζί υπερέχουν από την ομάδα ελέγχου» (π.χ. η πειραματική αγωγή συνολικά θα έχει σημαντική επίδραση), και (ii) «ότι οι 2 πειραματικές ομάδες διαφέρουν μεταξύ τους» (π.χ. κάποια από αυτές να υπερέχει από την άλλη ως πιο αποτελεσματική). Αν, όμως, ελεγχθεί και κάποια από τις άλλες 2 συγκρίσεις, επειδή προβλέπεται σε αντίστοιχες υποθέσεις, τότε έχουμε μια βέλτιστη περίπτωση με 2 ανεξάρτητες (ορθογώνιες) και 1 μη ανεξάρτητη από τις άλλες σύγκριση.

17.4 Προσχεδιασμένες Συγκρίσεις (Planned Comparisons).

Ακολουθεί ο αλγόριθμος υπολογισμού, ελέγχου και αξιολόγησης κάθε προσχεδιασμένης σύγκρισης (Ψ) σε k ανεξάρτητα δείγματα (ομάδες). Ο αλγόριθμος βασίζεται στη διασπορά (MS_w) και στους βαθμούς ελευθερίας (df_w) «εντός των ομάδων» από την ολική ANOVA, στους μέσους $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$ στα πλήθη $N_1, N_2, \ldots, N_k$, και στους συντελεστές $c_1, c_2, \ldots, c_k$.

1) Σύγκριση, Αντίθεση (Contrast, Ψ)

$$\psi = \sum c_j \bar{X}_j = c_1 \bar{X}_1 + c_2 \bar{X}_2 + \ldots + c_k \bar{X}_k \qquad (17.6.\alpha)$$

2) Άθροισμα Πηλίκων c_j^2/N_j

$$\sum \left(c_j^2 / N_j \right) = c_1^2 / \mathrm{N}_1 + c_2^2 / \mathrm{N}_2 + \ldots + c_k^2 / \mathrm{N}_k \qquad (17.6.\beta)$$

3) Τυπικό Σφάλμα (Std. Error)

$$SE_\Psi = \sqrt{MS_w * \sum \left(c_j^2 / N_j \right)} \qquad (17.6.\gamma)$$

4) Κριτήριο* $t_\psi = \Psi / SE_\psi$ = [Σύγκριση] / [Τυπικό Σφάλμα]

5) Στατιστική Σημαντικότητα του t_ψ της σύγκρισης

Με πιθανότητα σφάλματος τύπου Ι α (π.χ. 0.05 ή 0.01) και με δίπλευρο έλεγχο και βαθμούς ελευθερίας για τη διασπορά «εντός των k ομάδων» $df_w = k - 1$
το κρίσιμο t είναι t_c (Πίν. 21.Δ), οπότε

- αν $t_\psi \geq t_c$ η σύγκριση Ψ (διαφορά) είναι **στατιστικώς σημαντική**,
- αν $t_\psi < t_c$ η σύγκριση Ψ (διαφορά) είναι **στατιστικώς μη σημαντική**.

6) Μέγεθος Επίδρασης, Διαφοράς (effect size): Ουσιαστική σημαντικότητα

Κάθε σύγκριση Ψ αξιολογείται ως μέγεθος επίδρασης (διαφοράς μεταξύ των υπό σύγκριση ομάδων) σύμφωνα με τον Steiger (2004) με το μερικό η^2 (partial eta squared)

$$\eta_p^2 = F_\psi / (F_\psi + df_w) = \mathbf{t^2_\psi / (t^2_\psi + df_w)}, \qquad \text{καθότι } F = t^2.$$

Η αξιολόγηση του η^2 γίνεται με βάση την **κλίμακα Cohen** (1988, σελ. 283) ως

μικρό = 0.01 – 0.059, μέτριο = 0.06 – 0.139, μεγάλο = 0.14 και πάνω,
1% – 5.9% 6% – 13.9% 14% + ..

* Ο **έλεγχος t_ψ** διαφέρει από τον απλό **έλεγχο t** που γνωρίσαμε στο Κεφ. 14 (σύγκριση 2 μέσων): το **τυπικό του σφάλμα (SE_ψ)** βασίζεται στη «διασπορά εντός» (MS_w) για τα k δείγματα της ANOVA και όχι στις τυπικές αποκλίσεις μόνο των συγκρινόμενων δειγμάτων (S_1, S_2, …). Έχει δηλαδή **κοινό όρο σφάλμα** (common error term) για κάθε προσχεδιασμένη σύγκριση. Για το λόγο αυτό, **ο απλός έλεγχος t είναι αδόκιμος για την ανάλυση πολλαπλών συγκρίσεων σε ανεξάρτητα δείγματα**.

Παράδειγμα 17.2 - Προσχεδιασμένες Συγκρίσεις (Planned Comparisons).

Στον Πίν. 17.1α έχουμε τα στατιστικά της σύγκρισης 3 ομάδων με μέσους και πλήθη

ομάδα 1: $\bar{X}_1 = 4$, $N_1 = 3$, ομάδα 2: $\bar{X}_2 = 5$, $N_2 = 4$, ομάδα 3: $\bar{X}_3 = 8$, $N_3 = 5$, και στον Πίν. 17.1β «εντός ομάδων» διασπορά $MS_w = 2.2222$ & βαθμοί ελευθερίας $df_w = 9$.

Υποθέστε ότι προσχεδιάσαμε τη σύγκριση Ψ «**ομάδα 1 με ομάδα 2 & 3 μαζί**» (Πίν. 17.1).

Οι συντελεστές c της σύγκρισης Ψ μπορεί να είναι $c_1 = 2$, $c_2 = -1$, $c_3 = -1$.

1) Σύγκριση (Contrast, Ψ)

$$=(2)(4)+(-1)(5)+(-1)(8)=-5$$

2) Άθροισμα Πηλίκων c_j^2/N_j

$$\sum\left(c_j^2 / N_j\right) = (2)^2 / 3 + (-1)^2 / 4 + (-1)^2 / 5 = 1.7833$$

3) Τυπικό Σφάλμα (Std. Error)

$$SE_\Psi = \sqrt{MS_w * \sum\left(c_j^2 / N_j\right)} = \sqrt{2.2222 * 1.7833} = 1.9906$$

4) Κριτήριο $t_\psi = \Psi / SE_\psi = (-5) / 1.9906 = \mathbf{-2.5118}$

5) Στατιστική Σημαντικότητα (statistical significance) του t_ψ (έλεγχος)

Με πιθανότητα σφάλματος τύπου Ι $\alpha = 0.05$ και $df_w = 9$ (Πίν. 21.Δ)
το **κρίσιμο t** είναι $t_c = 2.262$ και επειδή $t_\psi = 2.51 > t_c = 2.262$ συμπεραίνουμε ότι η **σύγκριση Ψ είναι στατιστικώς σημαντική:** οι ομάδες 2 & 3 υπερέχουν της ομάδας 1 στην προαγωγή της «Κινητής Ικανότητας».

6) Μέγεθος Επίδρασης, Διαφοράς (effect size): Ουσιαστική σημαντικότητα
Το μέγεθος διαφοράς (Steiger, 2004) μεταξύ της ομάδας 1 και των ομάδων 2 & 3 μαζί είναι

$$\eta_p^2 = t^2_\psi / (t^2_\psi + df_w) = (2.51)^2 / (2.51)^2 + 9) = \mathbf{0.41}$$

και δείχνει ότι η διαφορά αυτή εξήγαγε 41% της διασποράς της εξαρτημένης μεταβλητής Χ.

Το 41% είναι αρκετά μεγάλη επίδραση (Cohen, 1988, σελ. 283: μεγάλη επίδραση από 14% και πάνω). Όμως, **η πρακτική αξία** της διαφοράς αυτής θα μπορούσε να εκτιμηθεί μόνο με βάση το ουσιαστικό περιεχόμενο του ερευνητικού προβλήματος.

17.5 Μετα-ANOVA Έλεγχος Scheffe: Όλες οι Δυνατές Συγκρίσεις!

Ο έλεγχος **Scheffe** είναι κατάλληλος **για πολλαπλές συγκρίσεις κατά ζεύγη και συνδυαστικά μόνο μετά από στατιστικώς σημαντική ANOVA** (Scheffe, 1953). Παρέχει τη μέγιστη δυνατή προστασία από τη **διόγκωση (inflation)** της **ολικής πιθανότητας σφάλματος τύπου Ι**. Η διόγκωση αυτή προκύπτει όταν κάνουμε πολλές συγκρίσεις με κάθε μία π.χ. στο $\alpha = 0.05$. Αυτό αυξάνει την πιθανότητα λαθεμένης απόρριψης μιας «αληθούς» μηδενικής υπόθεσης (π.χ. $\mu_1 = \mu_2 = \mu_3$) και «αποδοχής» μιας «μη υπαρκτής» διαφοράς ως «στατιστικώς σημαντικής». Είναι κατάλληλος για την ανάλυση πειραμάτων με σύνθετες υποθέσεις και για το λόγο αυτό θεωρείται ως ο **πλέον αυστηρός μετά-ANOVA έλεγχος**.

Ο έλεγχος **Scheffe** δίνει το **κρίσιμο F Scheffe (critical F)**

$$F_{Scheffe} = (k-1) F_{(\alpha,\, df_b, df_w)} \tag{17.7α}$$

που ως **κρίσιμο t** (critical t) είναι

$$t_{Scheffe} = \sqrt{F_{Scheffe}} \tag{17.7β}$$

ή, εναλλακτικά, την **κρίσιμη διαφορά D Scheffe (critical difference)**

$$D_{Scheffe} = t_{Scheffe}\sqrt{\frac{MS_w}{\tilde{N}}} \qquad (17.7\gamma)$$

όπου
k ο αριθμός των ομάδων,
MS_w η διασπορά εντός των ομάδων (σφάλμα),
df_b οι βαθμοί ελευθερίας μεταξύ των ομάδων,
df_w οι βαθμοί ελευθερίας εντός των ομάδων,
$\tilde{N}$ το αρμονικό μέσο πλήθος των ομάδων (Κεφ. 5.2) και
F(α, df_b, df_w) το **κρίσιμο F** (Πίν. 21.Η).

- **Στατιστική Σημαντικότητα (Statistical Significance) κάθε Διαφοράς**
  - αν $D \geq D_{Scheffe}$ η διαφορά είναι **στατιστικώς σημαντική**,
  - αν $D < D_{Scheffe}$ η διαφορά είναι **στατιστικώς μη σημαντική**.
- **Μέγεθος Επίδρασης, Διαφοράς (Effect Size)**

  Το **μέγεθος της διαφοράς** σε κάθε σύγκριση είναι **d = D / S = Διαφορά / Τυπική Απόκλιση**. Ως τυπική απόκλιση **S** επιλέγουμε γενικά τη **συνδυασμένη τυπική απόκλιση** (pooled standard deviation, Κεφ. 14.3.2)

$$S_p = \sqrt{\left[\left(N_1 - 1\right)S_1^2 + \left(N_2 - 1\right)S_2^2\right) / \left(N_1 + N_2 - 2\right)\right]},$$

  η οποία σε **σχεδόν ίσες ομάδες** (i) με **άνισες τυπικές αποκλίσεις** απλουστεύεται σε $\sqrt{\left[S_1^2 + S_2^2\right] / 2}$ και (ii) με **σχεδόν ίσες τυπικές αποκλίσεις** απλουστεύεται πλήρως σε $S = (S_1 + S_2) / 2$.

Η **αξιολόγηση της διαφοράς d** γίνεται με την **κλίμακα Cohen** για τυποποιημένες διαφορές (standardized differences): μικρή = 0.20 – 0.49, μέτρια = 0.50 – 0.79, μεγάλη = 0.80 & άνω (Cohen, 1988, σελ. 20-27). Η **πρακτική αξία** της εκτιμάται στο πλαίσιο του ερευνητικού πεδίου, με βάση δηλαδή το ουσιαστικό περιεχόμενο του ερευνητικού προβλήματος.

Παράδειγμα 17.3 – Μετα-ANOVA Έλεγχος Scheffe: Συγκρίσεις μόνο κατά Ζεύγη.

Στο παράδειγμα 17.1 είδαμε ότι οι 3 ομάδες διέφεραν σημαντικά στην κινητική ιδιότητα Χ ($F = 8.10$, $p < 0.01$, $\eta^2 = 0.64$). Τα κύρια στατιστικά της ανάλυσης δίνονται στον Πίν. 17.1β:

«εντός των ομάδων» διασπορά **MS_w = 2.2222** και βαθμοί ελευθερίας **df_w = 9**, μέσοι, πλήθη, τυπικές αποκλίσεις:

ομάδα 1:	$\bar{X}_1 = 4$	$N_1 = 3$	$S_1 = 2$
ομάδα 2:	$\bar{X}_2 = 5$	$N_2 = 4$	$S_2 = 1.63$
ομάδα 3:	$\bar{X}_3 = 8$	$N_3 = 5$	$S_2 = 1.$

Με τον **έλεγχο Scheffe** θα μπορούσαμε να ελέγξουμε όλες τις δυνατές διαφορές των 3 μέσων κατά ζεύγη και συνδυαστικά. **Ας διερευνήσουμε μόνο τις διαφορές κατά ζεύγη**.

Το **αρμονικό μέσο πλήθος** των 3 ομάδων είναι $\bar{N}$ = 3 / [1/3 + 1/4 + 1/5] = **3.8298**. Με α = 0.05, $df_1 = df_b = 2$ και $df_2 = df_w = 9$ (Πίν. 21.Η) το **κρίσιμο F** είναι $F_c = 4.26$,

το **κρίσιμο F Scheffe** = (k – 1) F_c = (3 – 1)(4.26) = **8.52**,
το **κρίσιμο t Scheffe** = √(F Scheffé) = √(8.52) = **2.92** και
η **κρίσιμη διαφορά (critical difference D) Scheffe**

$$D_{Scheffe} = 2.92 * \sqrt{2.22/3.83} = 2.92 * 0.76 = 2.219.$$

Στατιστική Σημαντικότητα κάθε Διαφοράς (D)

¤ **1η Σύγκριση - Ομάδα 1 με Ομάδα 2:**

$\bar{X}_1 - \bar{X}_2 = 4 - 5 = -1$ και επειδή 1 < 2.219 η διαφορά είναι στατιστικώς μη σημαντική,

¤ **2η Σύγκριση - Ομάδα 1 με Ομάδα 3:**

$\bar{X}_1 - \bar{X}_3 = 4 - 8 = -4$ και επειδή 4 > 2.219 η διαφορά είναι **στατιστικώς σημαντική**,

¤ **3η Σύγκριση - Ομάδα 2 με Ομάδα 3:**

$\bar{X}_2 - \bar{X}_3 = 5 - 8 = -3$ και επειδή 3 > 2.219 η διαφορά είναι **στατιστικώς σημαντική**.

Μέγεθος Επίδρασης, Διαφοράς (effect size): Ουσιαστική Σημαντικότητα
Η τυποποιημένη διαφορά σε κάθε σύγκριση είναι d = διαφορά / τυπική απόκλιση = D / S. Οι 3 ομάδες έχουν σχεδόν ίσα πλήθη (N) και άνισες τυπικές αποκλίσεις (S). Άρα, η εκτίμηση του μεγέθους της διαφοράς μπορεί να γίνει επαρκώς ως

$$d = \mathrm{D} / \sqrt{[S_1^2 + S_2^2]/2}\ .$$

Η τυπική απόκλιση είναι για την 1η σύγκριση $\sqrt{\left(2^2 + 1.63^2\right)/2}$ = **1.82**, για τη 2η σύγκριση $\sqrt{\left(1.63^2 + 1^2\right)/2}$ = **1.35** και για την 3η σύγκριση $\sqrt{\left(2^2 + 1^2\right)/2}$ = **1.58**.

Τα **μεγέθη διαφοράς** είναι $d_{1.2}$ = 1 / 1.82 = **0.55**, $d_{1.3}$ = 4 / 1.35 = **2.96**, $d_{2.3}$ = 3 / 1.58 = **1.88**. Κατά τον Cohen (1988, σελ. 20-27): διαφορά 0.20 – 0.49 = μικρή, 0.50 – 0.79 = μέτρια, 0.80 και άνω = μεγάλη). Άρα η 0.56 είναι μέτρια, η 2.86 πολύ μεγάλη και η 1.88 μεγάλη.

Συμπέρασμα: Η στατιστική σημαντικότητα της ANOVA (F = 8.10, p < 0.01, η^2 = 0.64) οφείλεται στη διαφορά (i) μεταξύ της ομάδας 1 και 3, που ήταν στατιστικώς σημαντική (p < 0.05) και πολύ μεγάλη (Cohen d = 2.96), και (ii) μεταξύ της ομάδας 2 και 3, που ήταν επίσης στατιστικώς σημαντική (p < 0.05) και μεγάλη (Cohen d = 1.88). Η **πρακτική αξία** των μεγεθών επίδρασης αξιολογείται με βάση το ουσιαστικό περιεχόμενο της έρευνας.

17.6 Μετα-ANOVA Έλεγχος Tukey: Όλες οι Συγκρίσεις Κατά Ζεύγη.

Ο έλεγχος **Tukey** χρησιμοποιείται **για πολλαπλές συγκρίσεις κατά ζεύγη** μεταξύ των k μέσων μόνο μετά από στατιστικώς σημαντική ANOVA (Tukey, 1949). Θεωρείται λιγότερο «αυστηρός» από τον έλεγχο Scheffe, αλλά παρέχει επαρκή προστασία από τη διόγκωση του **ολικού (experiment-wise) σφάλματος τύπου I**. Η διόγκωση αυτή προκύπτει όταν κάνουμε πολλές συγκρίσεις, με κάθε μία π.χ. στο α = 0.05. Αυτό αυξάνει την πιθανότητα λαθεμένης απόρριψης της μηδενικής υπόθεσης (π.χ. $\mu_1 = \mu_2$) και «αποδοχής» μιας «μη υπαρκτής» διαφοράς ως «στατιστικώς σημαντικής». Είναι κατάλληλος για πειράματα με λίγες ομάδες και όχι σύνθετες συγκρίσεις και γι' αυτό θεωρείται ως ο **βέλτιστος μετα-ANOVA έλεγχος**.

Ο έλεγχος **Tukey** δίνει την **κρίσιμη διαφορά D Tukey (critical difference)**, τη δίκαια σημαντική διαφορά (honestly significant difference, HSD) για κάθε σύγκριση 2 μέσων και με **αρμονικό πλήθος** των 2 υπό σύγκριση ομάδων $\tilde{N}$ **= 2 / (1/N_1 + 1/N_2)**.

¤ **Κρίσιμη Διαφορά (critical difference) Tukey**

$$D_{Tukey} = q_{(\alpha,\, k,\, df_w)} \sqrt{\frac{MS_w}{\tilde{N}}} \qquad (17.8)$$

όπου **q** το στατιστικό **Studentized** (Πίν. 21.I) που ορίζεται με βάση το επίπεδο πιθανότητας **α**, τον αριθμό **k** των ομάδων και τους βαθμούς ελευθερίας «εντός των ομάδων» $\mathbf{df_w}$, και $\mathbf{MS_w}$ η διασπορά «εντός των ομάδων».

¤ **Στατιστική Σημαντικότητα κάθε Διαφοράς (D)**:

- αν $D \geq D_{Tukey}$ η διαφορά είναι **στατιστικώς σημαντική**,
- αν $D < D_{Tukey}$ η διαφορά είναι **στατιστικώς μη σημαντική**.

¤ **Μέγεθος Επίδρασης, Διαφοράς (Effect Size): Ουσιαστική Σημαντικότητα**

Το **μέγεθος της διαφοράς** σε κάθε σύγκριση είναι **d = D / S = Διαφορά / Τυπική Απόκλιση**. Ως τυπική απόκλιση **S** επιλέγουμε γενικά τη **συνδυασμένη τυπική απόκλιση** (pooled standard deviation, Κεφ. 14.3.2)

$$S_p = \sqrt{\left[(N_1-1)S_1^2 + (N_2-1)S_2^2\right) \,/\, (N_1+N_2-2)\right]},$$

η οποία σε **σχεδόν ίσες ομάδες** (i) με **άνισες τυπικές αποκλίσεις** απλουστεύεται σε $\sqrt{\left[S_1^2+S_2^2\right]/2}$ και (ii) με **σχεδόν ίσες τυπικές αποκλίσεις** απλουστεύεται πλήρως σε $S = (S_1 + S_2) / 2$.

Η **αξιολόγηση της διαφοράς d** γίνεται με την **κλίμακα Cohen** για τυποποιημένες διαφορές (standardized differences): μικρή = 0.20 – 0.49, μέτρια = 0.50 –

0.79, μεγάλη = 0.80 & άνω (Cohen, 1988, σελ. 20-27). Η **πρακτική αξία** της εκτιμάται στο πλαίσιο του ερευνητικού πεδίου, με βάση δηλαδή το ουσιαστικό περιεχόμενο του ερευνητικού προβλήματος.

Παράδειγμα 17.4 – Μετα-ANOVA Έλεγχος Tukey: Συγκρίσεις κατά Ζεύγη.

Στο παράδειγμα 17.1 είδαμε ότι οι 3 ομάδες διέφεραν σημαντικά στην κινητική ιδιότητα Χ ($F = 8.10$, $p < 0.01$, $\eta^2 = 0.64$). Τα κύρια στατιστικά της ανάλυσης δίνονται στον Πίν. 17.1β:

«εντός των ομάδων» διασπορά $\mathbf{MS_w = 2.2222}$ και βαθμοί ελευθερίας $\mathbf{df_w = 9}$, μέσοι, πλήθη, τυπικές αποκλίσεις:

ομάδα 1:	$\bar{X}_1 = 4$	$N_1 = 3$	$S_1 = 2$
ομάδα 2:	$\bar{X}_2 = 5$	$N_2 = 4$	$S_2 = 1.63$
ομάδα 3:	$\bar{X}_3 = 8$	$N_3 = 5$	$S_2 = 1.$

Ας διερευνήσουμε τις 3 δυνατές συγκρίσεις κατά ζεύγη: $\mu_1 - \mu_2$, $\mu_1 - \mu_3$, $\mu_2 - \mu_3$. Το αρμονικό πλήθος των ομάδων ανά 2 είναι για κάθε σύγκριση αντίστοιχα:

$$\tilde{N}_{1,2} = 2 / [1/3 + 1/4] = \mathbf{3.43},\ \tilde{N}_{1,3} = 2 / [1/3 + 1/5] = \mathbf{3.75},\ \tilde{N}_{2,3} = 2 / [1/4 + 1/5] = \mathbf{4.47}$$

Με $\alpha = 0.05$, $k = 3$ ομάδες και $df_w = 9$ (Πίν. 21.I)

το **κρίσιμο q Tukey** είναι $\mathbf{q_T = 3.95}$ και η **κρίσιμη διαφορά Tukey** είναι

για τη σύγκριση $\mu_1 - \mu_2$ $\quad D_{Tukey\ .1.2} = 3.95\sqrt{2.22 / 3.43} = 3.18,$

για τη σύγκριση $\mu_1 - \mu_3$ $\quad D_{Tukey\ .1.3} = 3.95\sqrt{2.22 / 3.75} = 3.04,$

για τη σύγκριση $\mu_1 - \mu_3$ $\quad D_{Tukey\ .2.3} = 3.95\sqrt{2.22 / 4.44} = 2.78$

Στατιστική Σημαντικότητα κάθε διαφοράς (D)

- **1η Σύγκριση - Ομάδα 1 με Ομάδα 2:**

  $\bar{X}_1 - \bar{X}_2 = 4 - 5 = -1$ και επειδή $1 < 3.18$ η διαφορά είναι στατιστικώς μη σημαντική,

- **2η Σύγκριση - Ομάδα 1 με Ομάδα 3:**

  $\bar{X}_1 - \bar{X}_3 = 4 - 8 = -4$ και επειδή $4 > 3.04$ η διαφορά είναι **στατιστικώς σημαντική**,

- **3η Σύγκριση - Ομάδα 2 με Ομάδα 3:**

  $\bar{X}_2 - \bar{X}_3 = 5 - 8 = -3$ και επειδή $3 > 2.78$ η διαφορά είναι **στατιστικώς σημαντική**.

Μέγεθος Επίδρασης, Διαφοράς (effect size)

Η τυποποιημένη διαφορά σε κάθε σύγκριση είναι d = διαφορά / τυπική απόκλιση = D / S.