One-way Analysis of Variance (ANOVA)

Between-subjects designs

Konstantinos I. Bougioukas, MSc, PhD

School of Psychology
Faculty of Philosophy
Aristotle University of Thessaloniki

Έλεγχος υποθέσεων-Βήματα

1. Καθορίζεται η **μηδενική υπόθεση Η0 (=)** και **εναλλακτική υπόθεση Η1 (≠) .**

2. Ορίζεται το **επίπεδο σημαντικότητας** α (συνήθως α=0.05).

3. Επιλέγεται μια κατάλληλη **στατιστική δοκιμασία** και υπολογίζεται η τιμή του στατιστικού με βάση τα δεδομένα του δείγματος.

4. Σύγκριση της **πιθανότητας p** να έχουμε την συγκεκριμένη τιμή του στατιστικού (ή κάτι πιο ακραίο) θεωρώντας ότι ισχύει η Ho, με το **επίπεδο σημαντικότητας α** (0.05). Στατιστικά σημαντικό αποτέλεσμα ($p < 0.05$).

5. **Ερμηνεία** αποτελεσμάτων.

One-way between-subjects ANOVA

ANOVA: Y = numeric variable → **Dependent**

X = categorical (factor) variable (>2 levels) → **Independent**

NOTE: If categorical variable has only 2 levels: 2-sample t-test

EXAMPLE

The SAT is used by a wide range of colleges and universities as part of the application process for college admission. Assume we are interested in the **effect of preparation time on SAT performance.**

Y: SAT score (math section: 200-800)
X: preparation time (five student groups: zero, two, four, six, or eight months prior to taking SAT)

One-way between-subjects ANOVA: hypotheses

H_0: The means of all the groups are equal ($\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$)

H_1: At least one mean differs from the other

- Doesn't say how or which ones differ.
- Can follow up with "post-hoc tests".

Assumptions of one-way between-subjects ANOVA

- **Between-subjects design** (participants are represented by one level of the independent variable)

- **Normality:** The outcome variable (SAT score in our example) must be normally distributed in each group (here preparation time 0, 2, 4, 6, or 8 months).

  Check this by looking at histograms, Q-Q plots, and Shapiro-Wilk test:

  If $P>0.05$ then normal distribution

  If $P<0.05$ then non-normal distribution

- **Homogeneity of variances:** the variance of each group are approximately equal.

  Check for homogeneity of variances (Levene's test)

  If $P>0.05$ then ANOVA

  If $P<0.05$ then ANOVA with corrected df (Welch's ANOVA test)

DATA

Factor levels	Preparation time					Total
	0 months	2 months	4 months	6 months	8 months	
	370	410	530	560	550	
	380	430	530	580	610	
	420	[illegible]	[illegible]	[illegible]	620	
	420	[illegible]	[illegible]	[illegible]	630	
	430	510	550	640	640	
	430	520	570	650	650	
	440	530	610	660	660	
Sum	2890	3320	3870	4300	4360	18740
Mean	412.9	474.3	552.9	614.3	623.9	**535.4**

SAT scores

How it works

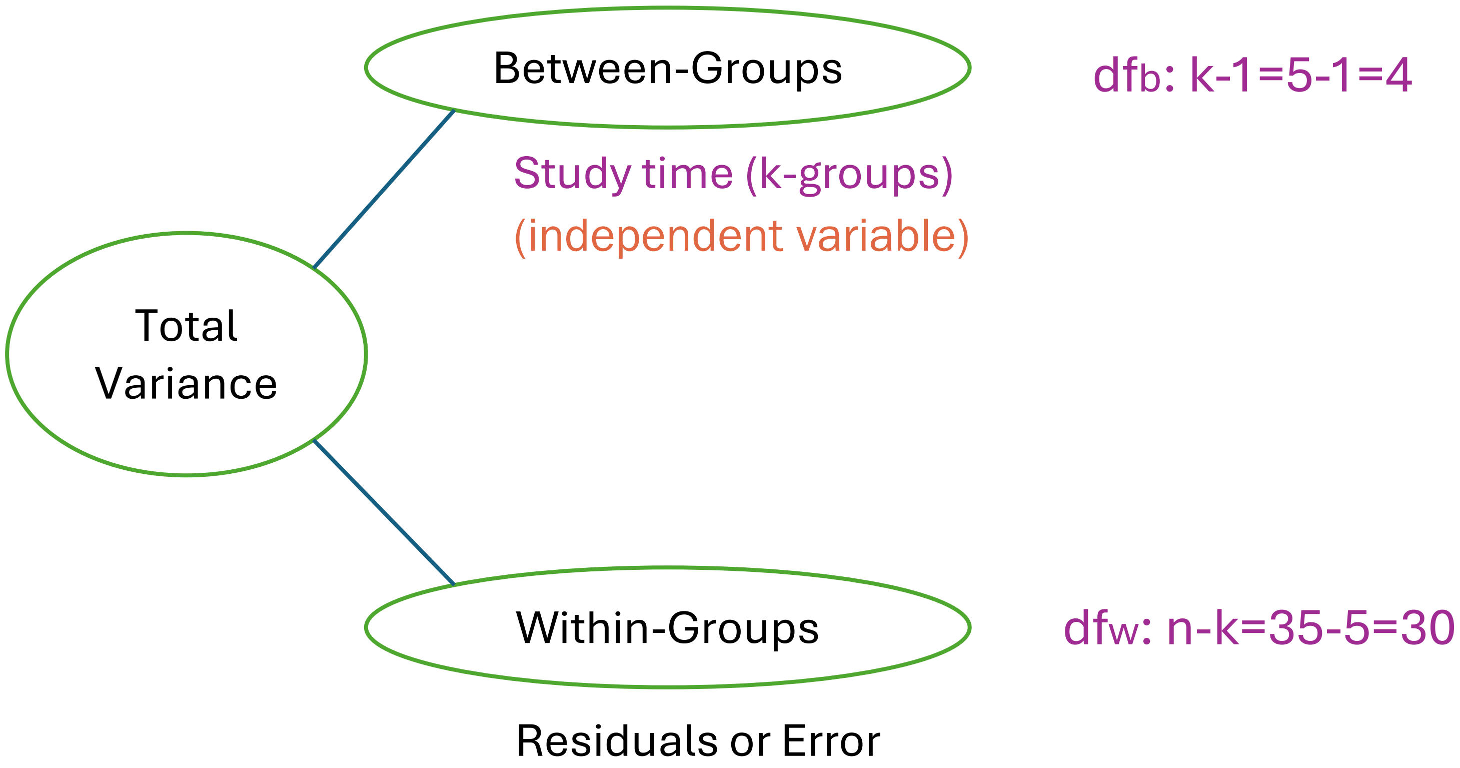
Factor levels	0 months	2 months	4 months	6 months	8 months	Total
	370	410	530	560	550	
	380	430	530	580	610	
	420	440	540	590	620	
	420	480	540	620	630	
	430	510	550	640	640	
	430	520	570	650	650	
	440	530	610	660	660	
Mean	**412.9**	474.3	552.9	614.3	623.9	**535.4**

Group means

Grand mean

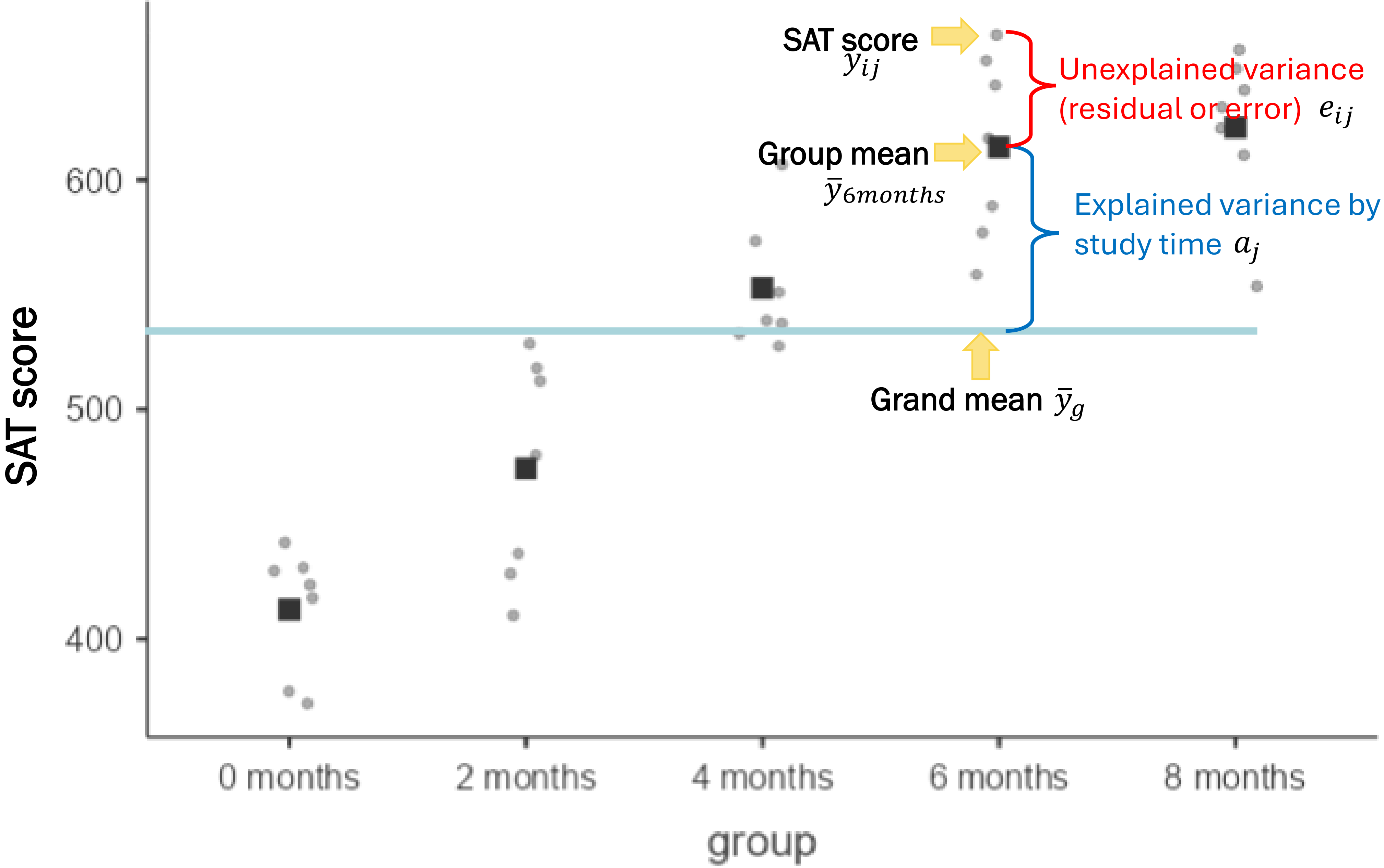
One-way between-subjects ANOVA

Partitioning the total variance into its sources



Explained and unexplained variation in data

$$y_{ij} = \bar{y}_g + a_j + e_{ij}$$



The ANOVA F-statistic is a ratio of the Between Group Variation divided by the Within Group Variation:

$$F = \frac{variation\ Between\ group\ means}{variation\ \ Within\ groups}$$

A large F is evidence *against* H_0, since it indicates that there is more difference between groups than within groups.

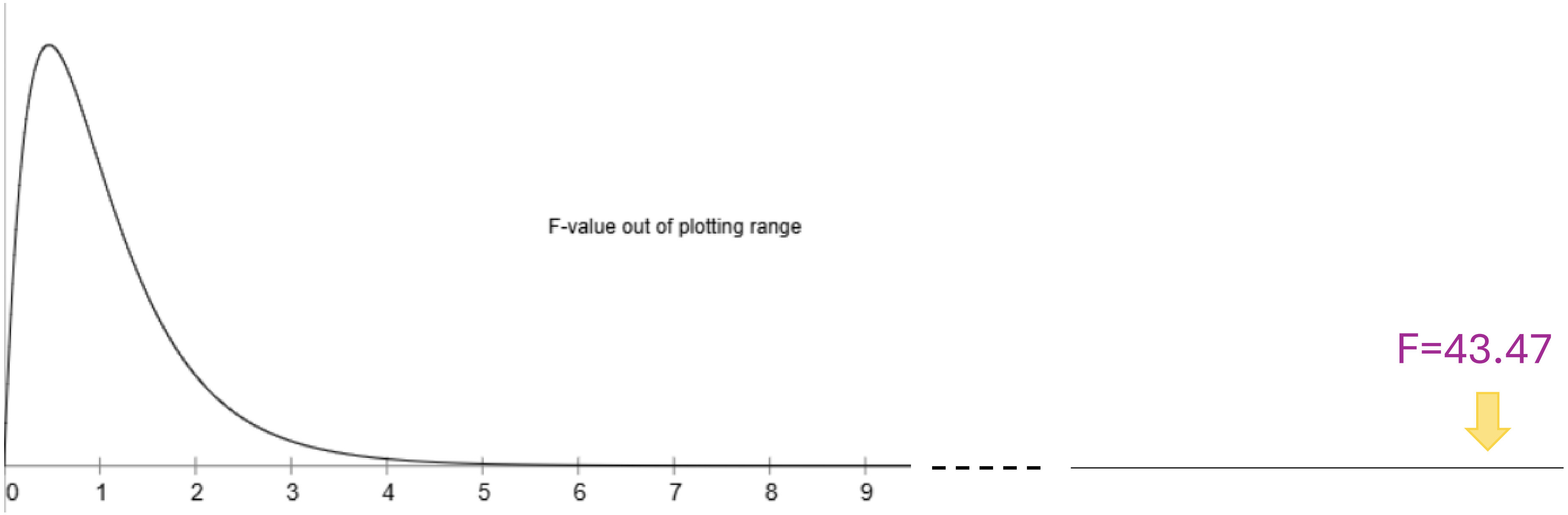
Rejection of the H_0 ($p<0.05$) will not tell us if one differs or which one it is: this is an overall test that we do!

ANOVA TABLE

ANOVA - statscore

	Sum of Squares	df	Mean Square	F	p	η^2
group	230497.14	4	57624.29	43.47	<.001	0.85
Residuals	39771.43	30	1325.71			

F-DISTRIBUTION



Multiple comparison problem (alpha inflation)

If we perform all pair-wise comparisons, type I error will not be equal to the originally decided α

FWER=1-(1-α)^c, c = no of pair-wise comparisons

- For α=0.05 and c= 3 pair-wise comparisons, the probability of type I error becomes 0.143
- For α=0.05 and c= 10 pair-wise comparisons, the probability of type I error becomes 0.401

Post hoc Tests

- Post hoc tests are tests of the statistical significance of differences between group means calculated after ("post") having done an analysis of variance (ANOVA) that shows an overall difference.

- Post hoc tests are designed to make all pairwise comparisons while maintaining the error rate at the pre-established α level.

- Most common post-hoc tests that adjust our p-value are:
  - Bonferroni Procedure
  - Tukey's Test
  - Scheffe
  - Holm