ARISTOTLE UNIVERSITY
OF THESSALONIKI

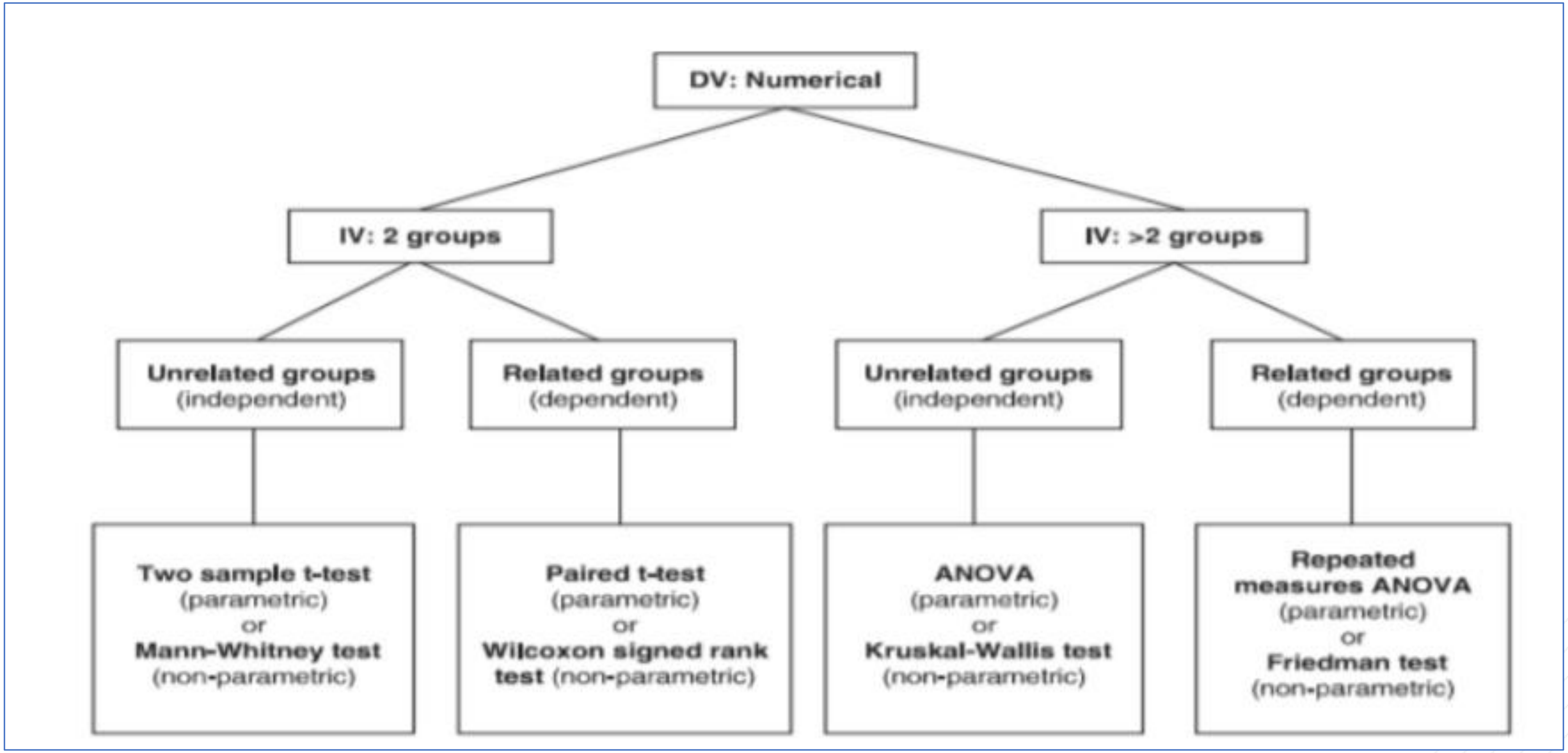
FACULTY OF HEALTH SCIENCES - SCHOOL OF MEDICINE
MSc Health Statistics and Data Analytics

Hypothesis testing – tests for more than two samples: One-Way ANOVA and the Kruskal-Wallis H test

Eirini Pagkalidou, PhDc
pagalidou@auth.gr

THESSALONIKI 2021-22

Statistical tests



Objectives

- ❑ Choose appropriate statistical procedure to compare a continuous outcome between more than two independent groups
- ❑ Know the assumptions of ANOVA
- ❑ Recognize when a Kruskal-Wallis test should be used instead
- ❑ Conclude on your hypothesis based on the test results

Framework

- Suppose we measure the LDL cholesterol levels in a group of individuals
- We classify them in 3 group according to their BMI (normal, overweighted, obese)
- We want to test whether the LDL levels in the three groups differ significantly
- What test should we perform?

Continuous outcome-Three or more independent groups

Example: LDL cholesterol levels, weight, height, Hematocrit, INR

	Independent (e.g BMI categories)
Parametric	**One-Way ANOVA (Analysis of Variance):** compares means between more than two independent groups
Non-parametric	**Kruskal-Wallis test:** non-parametric alternative to ANOVA

Assumptions of One-Way ANOVA (Analysis of Variance)

- Normality
  - The outcome variable (LDL in our example) must be normally distributed in each group
- Homogeneity of variances
  - 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 test)
- Independence
  - Each observation in the data must be a distinct and independent entity.

Inflation of type I error

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

- **$\alpha^*=1-(1-\alpha)^c$, 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= 6 pair-wise comparisons, the probability of type I error becomes 0.265

One-Way ANOVA

- One-Way ANOVA uses F-test to statistically test the equality of means.

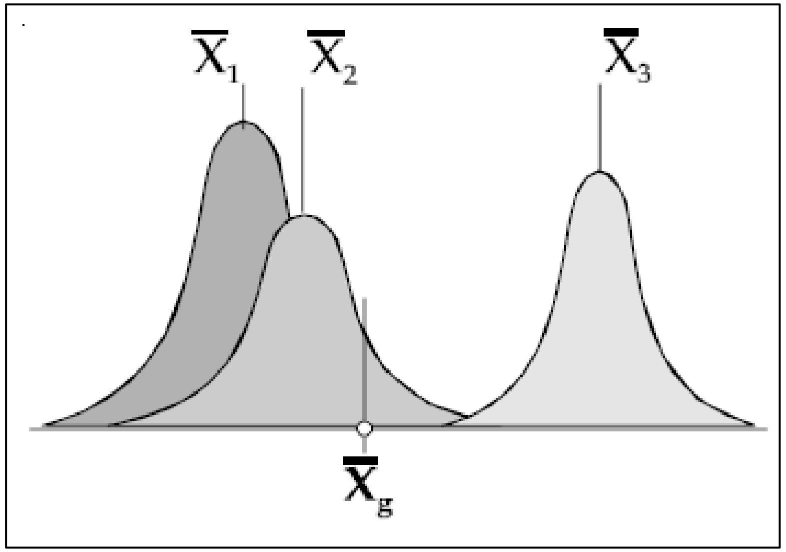
F = variation between group means / variation within the groups

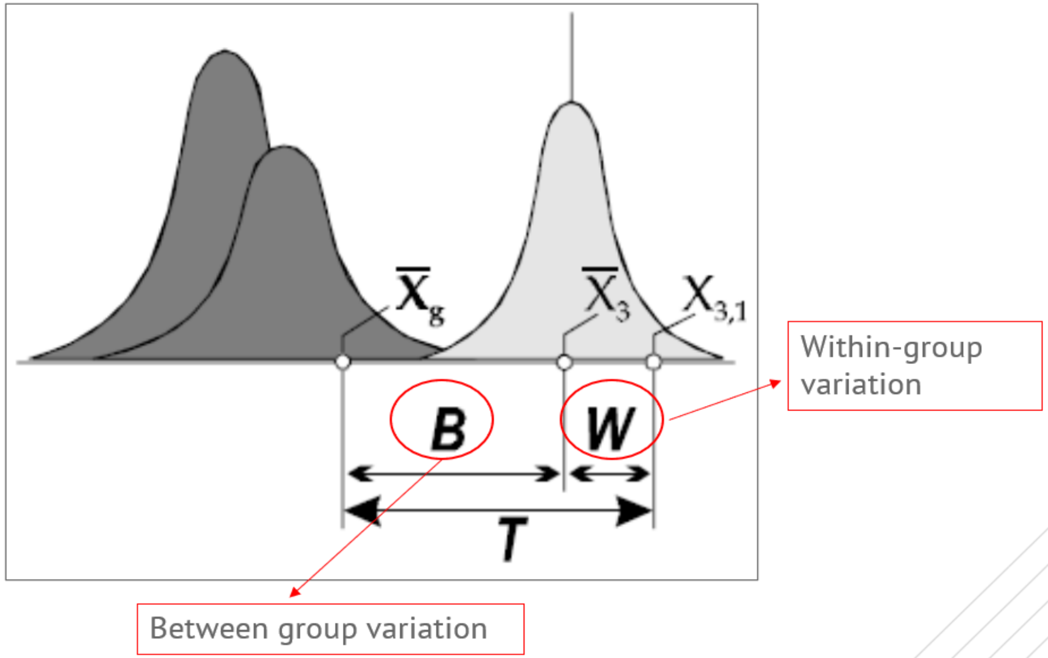
- If only two means are compared, then F test = independent samples t-test.

ANOVA

ANOVA separates the **total variability** seen in the data to:

- **Between-group variation**, i.e. one component that can be attributed to differences between groups
- **Within-group variation**, i.e. one component that is the random variation between the observations within each group.





One-Way ANOVA Hypotheses (example from the transfusion dataset)

- We assume that the LDL levels are normally distributed in each group

- The null hypothesis is that the mean LDL levels (μ) in the 3 BMI groups are equal

  $H_0: \mu_1=\mu_2=\mu_3$

- The alternative is that at least one of them is different

  $H_1: \mu_i \neq \mu_j$, i,j=1, 2, 3

- Rejection of the H_0 will not tell us if one differs or which one it is: this is an overall test that we do!

One-Way ANOVA hypotheses (cont.)

- One-way ANOVA test will only tell us whether at least two groups are different from each other
- BUT it will **not** tell us which groups are different
- For our example:

```
> summary(aov(LDL~BMI.cat))
                       Df Sum Sq Mean Sq F value Pr(>F)
as.factor(BMI.cat)      2  24858   12429    1577 <2e-16 ***
Residuals              37    292       8
```

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 are:
  - Bonferroni Procedure
  - Tukey's Test
  - Fisher's Least Significant Difference (LSD)

Retuning to our example:

```
Pairwise comparisons using t tests with pooled SD

data:  LDL and BMI.cat

                   Normal  Overweight
Overweight        <2e-16      -
Obese             <2e-16   <2e-16

P value adjustment method: bonferroni
```

The Kruskal-Wallis Rank Sum Test

- It is the non-parametric alternative to ANOVA
- It is used when normality assumptions of ANOVA have been violated
- It is also used when the outcome variable is ordinal or scale
- Independence of the observations is still a requirement for this test.

Kruskal-Wallis Test

- It is a non-parametric alternative to one-way ANOVA

- Like all non-parametric tests, the focus is on ranks, counting and the medians.

- The hypotheses statements are written as:

  H_0: All k populations have the same median

  H_1: Not all of the k population medians are the same

- As the ANOVA is a conceptual extension of the two sample t-test, so the K-W test is a conceptual extension of the Wilcoxon Rank Sum (Mann Whitney) test

Example

- Assume that we want to test whether persons in different BMI groups have different INR levels
- INR is not normally distributed between the BMI groups so the Kruskal Wallis test will be used
- R results:

```
Kruskal-Wallis rank sum test
data:  INR by BMI
Kruskal-Wallis chi-squared = 4.2101, df = 2, p-value = 0.122 > 0.05
```