ARISTOTLE UNIVERSITY
OF THESSALONIKI

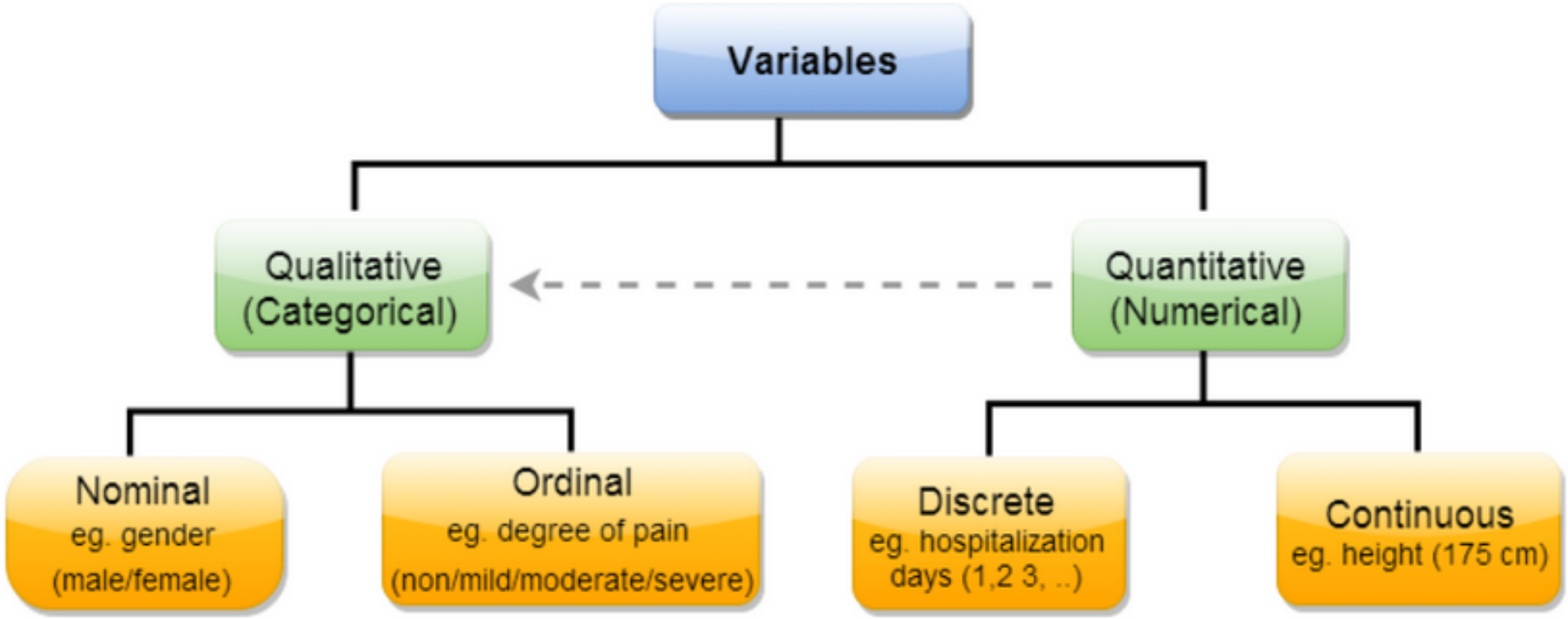
Basic Principles of Statistics

Anna-Bettina Haidich
Associate Professor of Medical Statistics –Epidemiology
haidich@auth.gr

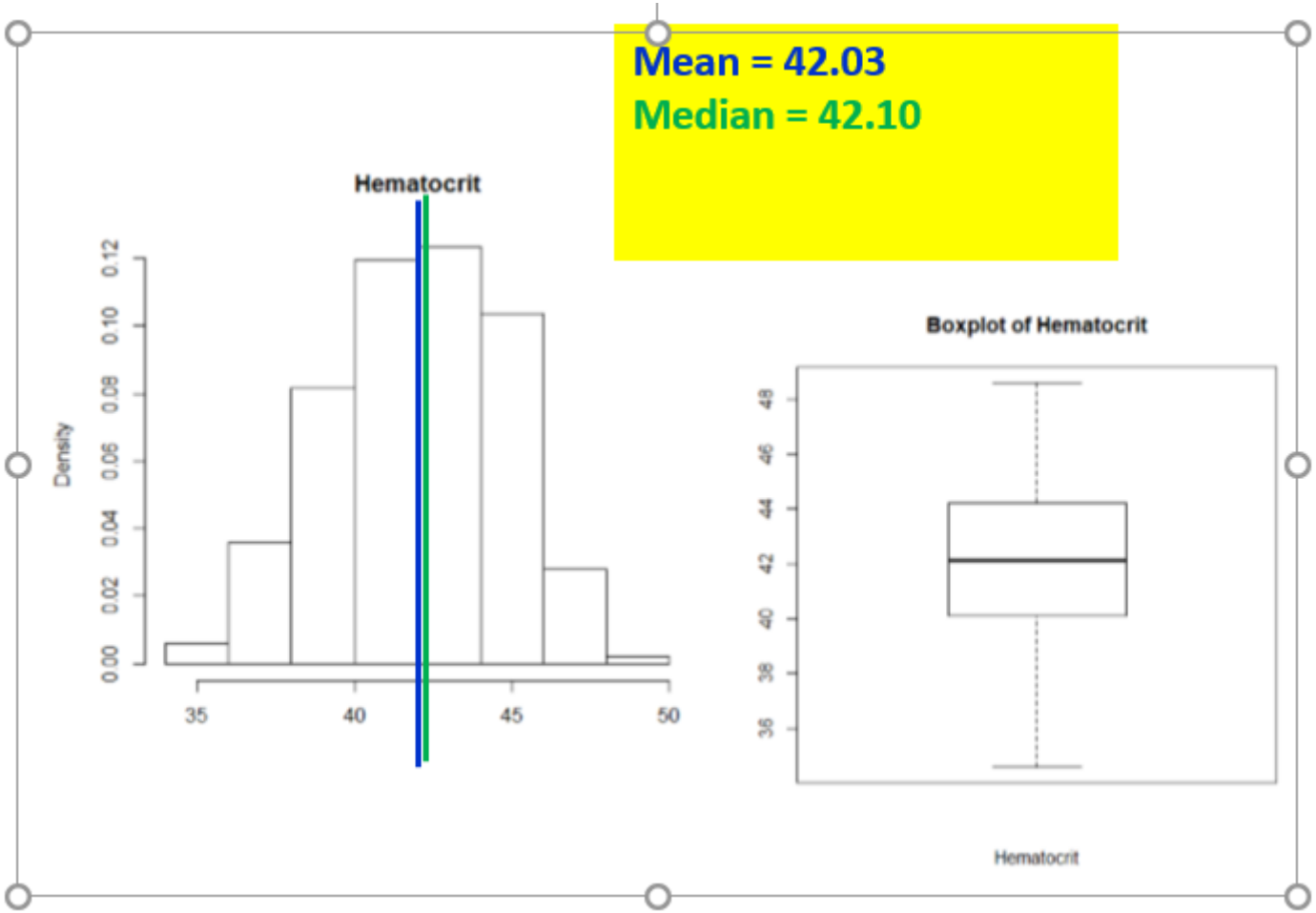
THESSALONIKI 2021-22

MSc Health Statistics & Data Analytics

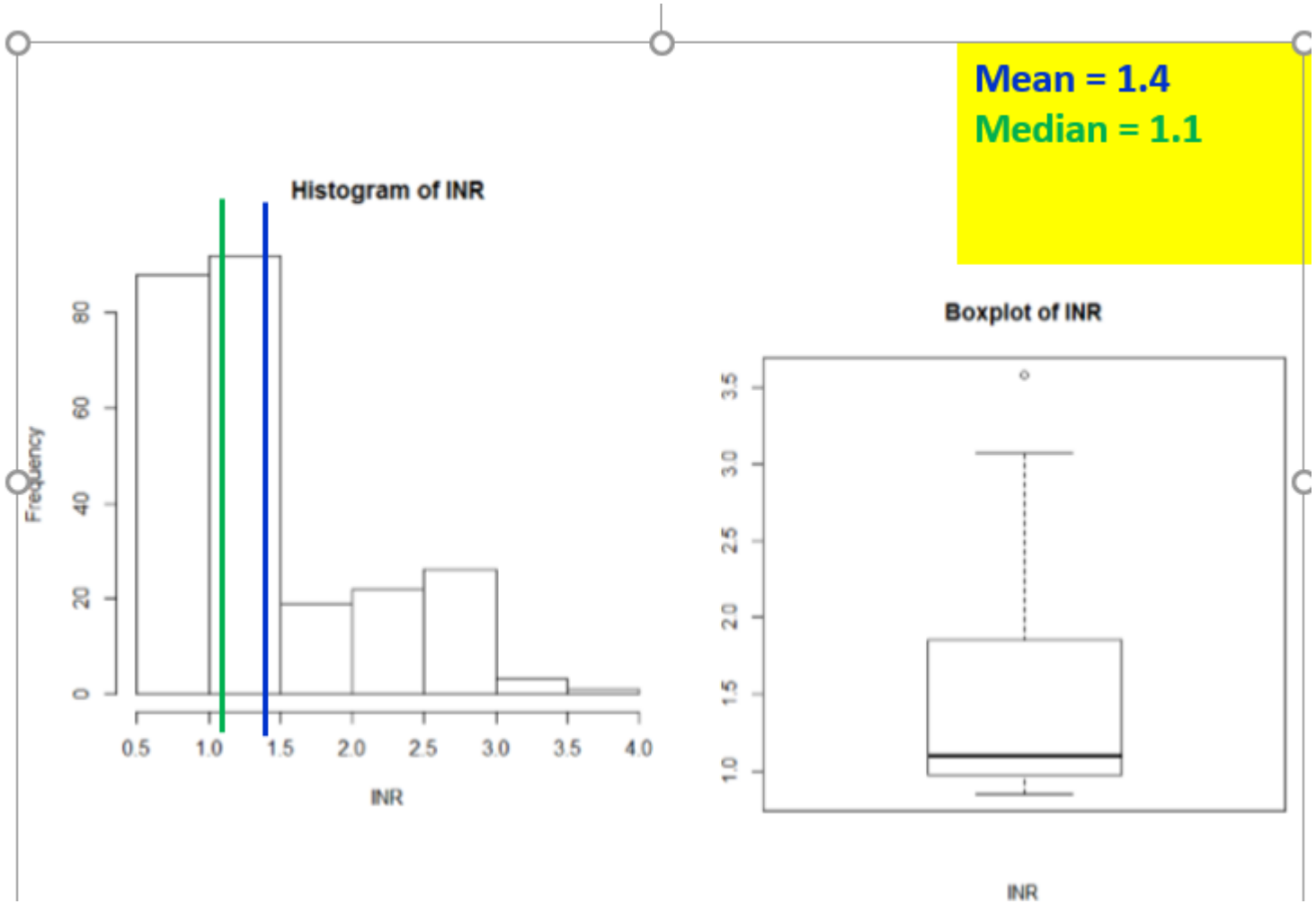
Types of variables



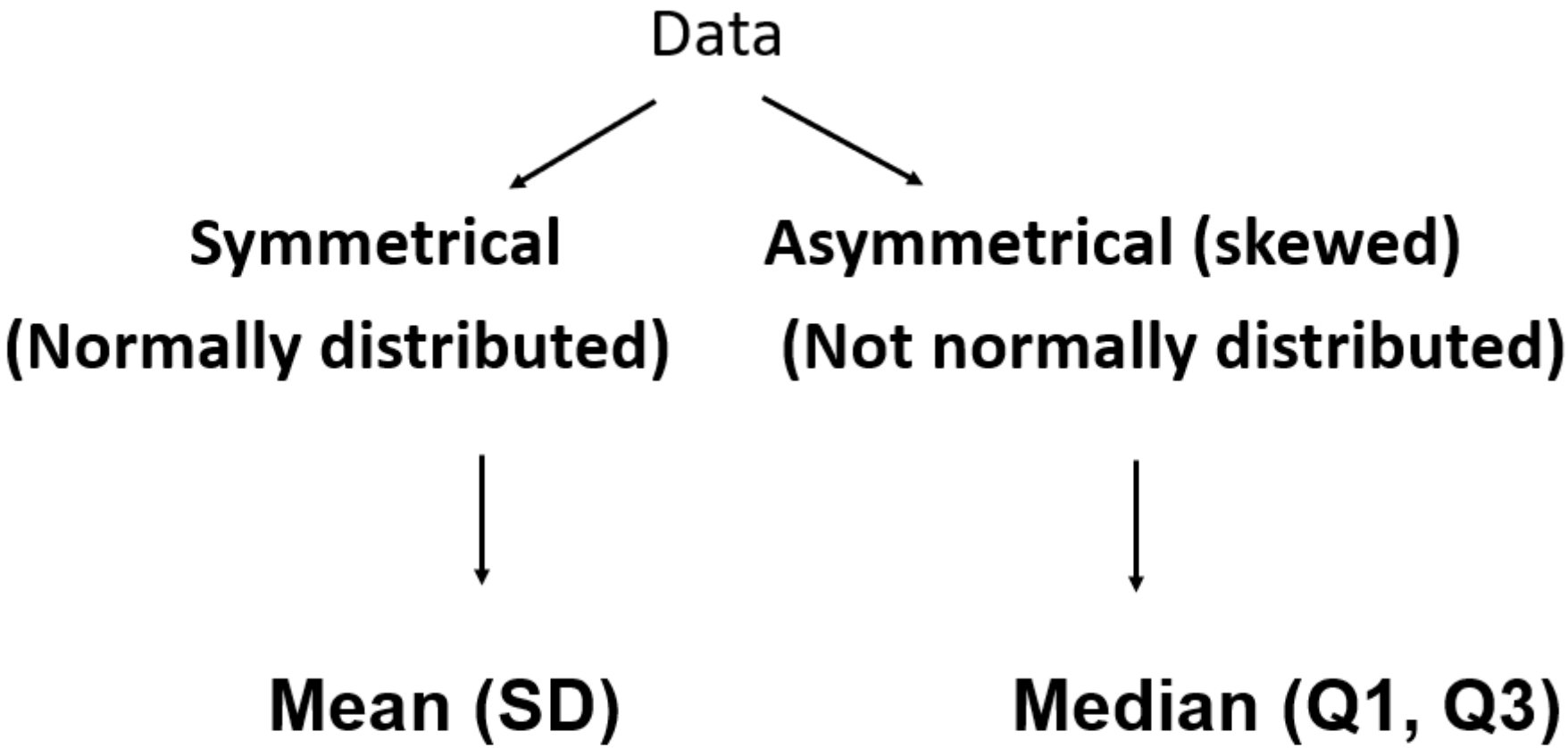
Symmetric (Normal) Distribution



Asymmetric (Not Normal) Distribution



Summary measures



Hypothesis testing

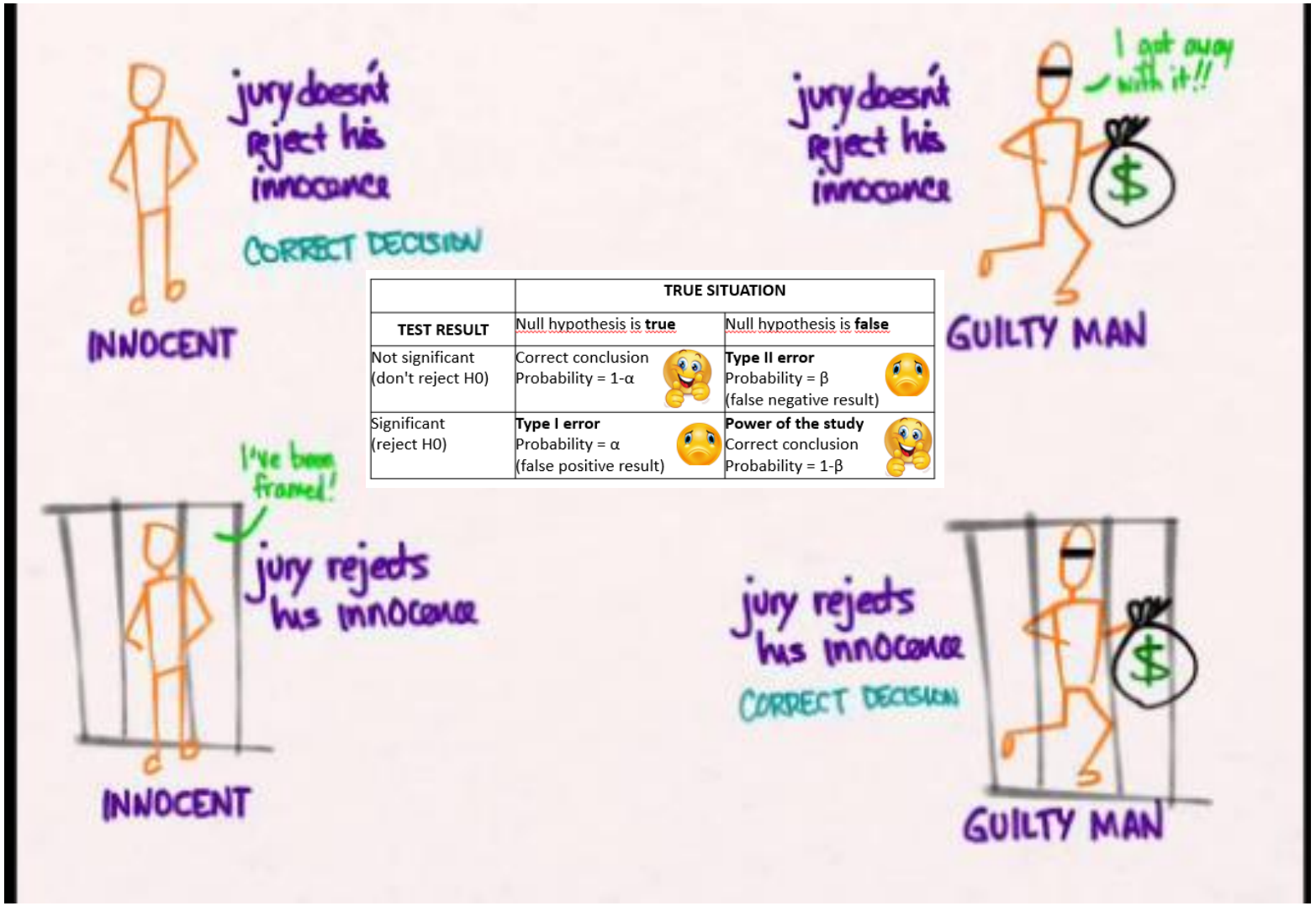
1. From the research question, determine the appropriate null hypothesis, H_0, and the alternative, H_1.
2. Set the level of significance, α (α=0.05)
3. Identify the appropriate test statistic and check the assumptions
4. Decide whether or not the result is statistically significant.
   - The ***p-value*** <0.05, this result ***is statistically significant***. **Reject the H0**
   - The ***p-value*** > 0.05, this result ***is NOT statistically significant***. **We cannot reject the H0** .
5. Interpret the results

Type I and Type II errors

	TRUE SITUATION	
TEST RESULT	Null hypothesis is **true**	Null hypothesis is **false**
Not significant (don't reject H0)	Correct conclusion Probability = 1-α 	**Type II error** Probability = β (false negative result) 
Significant (reject H0)	**Type I error** Probability = α (false positive result) 	**Power of the study** Correct conclusion Probability = 1-β 

- The power of the study is the probability of getting a statistically significant result with the selected sample if a true difference exists.
- The power is equal to 1-β - the larger the power of the study, the smaller the Type II error.

Court system and hypothesis testing



	TRUE SITUATION	
TEST RESULT	Null hypothesis is **true**	Null hypothesis is **false**
Not significant (don't reject H0)	Correct conclusion Probability = 1-α	**Type II error** Probability = β (false negative result)
Significant (reject H0)	**Type I error** Probability = α (false positive result)	**Power of the study** Correct conclusion Probability = 1-β

Two Independent samples test

1. State the null and alternative hypothesis
   - **H_0: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$ H_1: $\mu_1 \neq \mu_2$ or $\mu_1 - \mu_2 \neq 0$**

2. Check for normality in the continuous variable within the two groups

   - If normal (P>0.05) in both groups then t-test
     - Check for equality of variances (Levene's test)
       - If P>0.05 then unpaired t-test with pooled variance
       - If P<0.05 then unpaired t-test with unpooled variance (Welch's test)

   - If not normal (P<0.05) in at least one group then Mann-Whitney U test
     - H_0: $md_1 = md_2$ H_1: $md_1 \neq md_2$

> 2 Independent samples test

1. State the null and alternative hypothesis
   - $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ $\quad\quad$ $H_1: \mu_i \neq \mu_j$

2. Check for normality in the continuous variable within each group

   - If normal (P>0.05) in all groups then ANOVA
     - 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)

   - If not normal (P<0.05) in at least one group then Kruskal-Wallis test
     - $H_0: md_1 = md_2 \ldots = md_k$ $\quad\quad$ $H_1: md_i \neq md_j$

> 2 Independent samples test

- If ANOVA test significant (P<0.05) then find out which means differs with multiple pairwise t-tests and an alpha correction e.g Bonferroni

- If Kruskal-Wallis test significant (P<0.05) then find out which medians differs with multiple pairwise Mann-Whitney's U test and an alpha correction e.g Bonferroni

Paired test

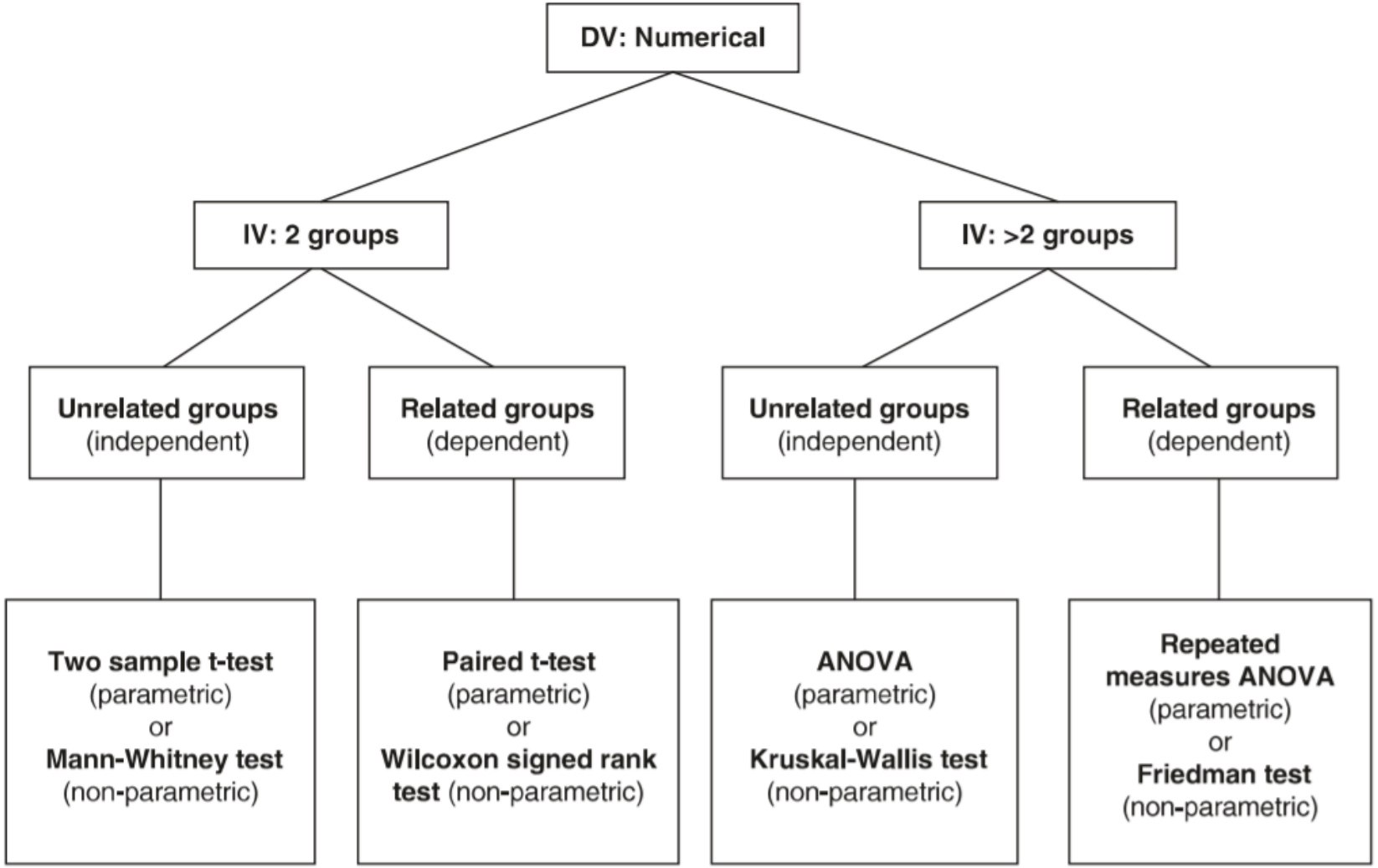
- State the null and alternative hypothesis
  - **H0: δ = 0 H1: δ ≠ 0**

- Calculate the difference di and check for normality
  - If normal (P>0.05) then **Paired t-test**
  - If not normal (P<0.05) the **Signed-Ranks Wilcoxon test**

Frequently used statistical tests

Parametric test	Respective non-parametric test	Use	Example
***t* test** for 2 independent groups	Mann-Whitney U test	Comparison of 2 independent groups	Compare height between boys and girls
***t* test** for 2 dependent groups	Wilcoxon signed ranks test	Compare two states of a variable	Compare birth weight and weight on discharge from the hospital
Variance analysis **ANOVA**	Kruskall-Wallis test	Compare 3 or more independent groups	Compare glucose levels on 4 different BMI groups: Underweight, Normal-weight, Overweight, Obese
Pearson's test	Spearman's test	Correlation between two numerical variables	Compare glucose levels and BMI in continuous scale

Flowchart of statistical tests



This weekend

Date	Hours	Topics	Tutor
19 November 2021	17:00-18:00	Repetition/queries	Anna-Bettina Haidich
	18:00 – 19:00	Repeated measures analysis	Eirini Pagkalidou
	19:30-21:00	Association of categorical variables	Anna-Bettina Haidich
20 November 2021	09:00-10:30	Measures of association and tests for categorical variables	Persefoni Talimtzi Konstantinos Bougioukas
	11:00- 12:30	Survival analysis: Log-rank test and Kaplan-Meier plots	Eirini Pagkalidou Persefoni Talimtzi
	13:00-15:00	Power and sample size calculation	Anna-Bettina Haidich Persefoni Talimtzi
21 November 2021	09:00-10:30	Statistical presentation+Repetition	Konstantinos Bougioukas
	11:00-12:30	Random sample selection	Eirini Pagkalidou
	13:00-15:00	Practice in datasets	All