WORKSHOP SCHEDULE

- 4 days
  - 1. Intro to R and data analysis
    - Descriptive statistics and visual data exploration
  - 2. Statistical inference & hypothesis testing
    - probability, CI, comparing 1 sample to population, comparing 2 samples, tests
  - 3. From association to prediction
    - Supervised ML (regression)
  - 4. Intro to ML, elements of power analysis
    - Unsupervised ML algorithm (PCA

- Each day will include:
  - Frontal class (MORNING)
  - Practical training with R about the topics discussed in the morning. (AFTERNOON)

DAY 1 – LECTURE OUTLINE

- Introduction to R and R-studio
  - Why R?
  - Principles of reproducible analysis with R + RStudio
- R objects, functions, packages
- Understanding different types of variables
  - Principles of "tidy data"
  - Data cleaning and manipulation
- Descriptive statistics
  - measures of central tendency, measures of variability (or spread), and frequency distribution
- Visual data exploration
  - {ggplot2}

LAB 1

- fatto !
- dataset autism
- REFERENCE
- + Benjamin Soltoff [Linear regression with a single predictor](https://info2950.infosci.cornell.edu/slides/16-models-single-predictor.html#/goals)
- Benjamin Soltoff [Linear regression with multiple predictors](https://info2950.infosci.cornell.edu/slides/17-models-multiple-predictors.html#/title-slide)
- + [biostats](https://biostats.w.uib.no/up-in-the-r-2/basic-statistics-in-r/)
- Modelling vocabulary
    - Predictor/feature/explanatory variable/independent variable
    - Outcome/dependent variable/response variable
    - Correlation
    - Regression line (for linear models)
        - Slope
        - Intercept

- +

DAY 2 – LECTURE OUTLINE

- Purpose and foundations of inferential statistics
  - Population and samples
  - Probability and random variables &
  - Meaningful probability distributions
  - Sampling distributions and Central Limit Theorem

- Getting to know the “language” of hypothesis testing
  - The null and alternative hypothesis
  - The probability of error? (α or “significance level")
  - The p-value probability and tests interpretation
  - Types of errors (Type 1 and Type 2)
  - Confidence Intervals (spostato dopo Z score)
  - Effective vs statistical significance (mettere altrove?)

- Hypothesis tests **examples**
  - Comparing sample mean to a hypothesized population mean (Z test & t test)
  - Comparing two independent sample means (t test)
  - Comparing sample means from 3 or more samples (ANOVA)

- A closer look at testing assumptions (with **examples**)
  - Testing two groups that are NOT independent
  - Testing if the data are not normally distributed: non-parametric tests
  - Testing samples without homogeneous variance of observations

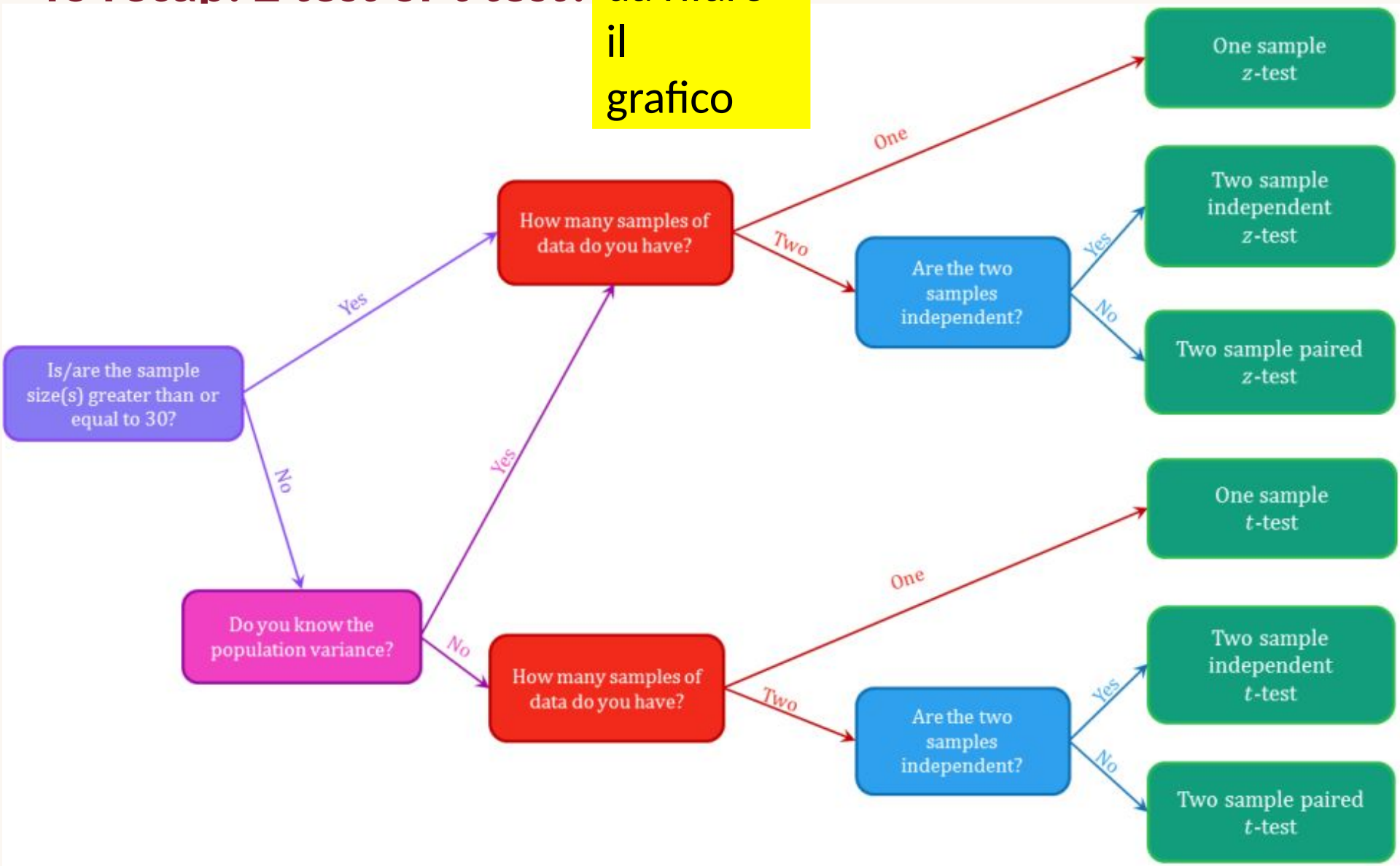
LAB 2

- vedi Esercitazione 2
  - dataset genes.xlsx
  - data/gait_clean.csv

- spostare esempi da Lecture a Practice?
- nuovi esempi?

To recap: Z test or t test?

da rifare il grafico



Source image: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/business/hypothesis-tests/

DAY 3 – LECTURE OUTLINE

- Testing for a correlation hypothesis (relationship of variables)
  - Pearson rho analysis (param)
  - Spearman test (no param)
- Measures of association
  - Fisher's Exact Test
  - Chi-Square Test of Independence

- From association to prediction -> Machine learning algorithms
  - Supervised
    - Example: Linear regression models
    - Multiple Linear Regression
    - logistic regression?

LAB 3

- vedi Esercitazione 3
  - ~~dataset metabolomica catanzaro ?~~
  - https://new.metaboanalyst.ca/MetaboAnalyst/upload/StatUploadView.xhtml

DAY 4 – LECTURE OUTLINE

- INQUADRAMENTO SCENARI DI RICERCA
  - causal inference v. prediction (ML)
  - observational - Longitudinal analysis ?
  - (vedere LEZIONE 3 cocca PDF)
  - ~~CLINICAL TRIAL ANALYSS ? Survival analisi ?~~

- From association to prediction -> Machine learning algorithms
  - from causal concern to prediction
  - Unsupervised
    - Example: PCA
  - PLS- DA
- Mostrare https://new.metaboanalyst.ca/MetaboAnalyst/upload/StatUploadView.xhtml

- Elements of Statistical Power analysis
  - Underlying principles of statistical power
  - Power calculations for basic study designs
  - Use power and sample size calculations as the basis of argument in support of study design, feasibility, and testing

- Mostrare (PACCHETTO metsize)

LAB 4

- vedi Esercitazione 4
  - dataset metabolomica catanzaro

- TEST + correzione