Choosing Appropriate Statistical Analysis

In the framework of the course:
"Applied Statistics"

How To Choose?

Prof. Dr. Mohamed Samer
Engineering in Biosystems, Energy and Environment
Department of Agricultural Engineering
Faculty of Agriculture, Cairo University
E-Mails: msamer@agr.cu.edu.eg; samer@cu.edu.eg
Website: http://scholar.cu.edu.eg/samer/biocv

Introduction

Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. It is basically a collection of quantitative data.

Descriptive statistics is a term given to the analysis of data that helps to describe, show and summarize data in a meaningful way.

Descriptive statistics is important to present raw data in effective and meaningful way using numerical calculations, graphs or tables.

This type of statistics is applied to already known data. It is a simple way to describe data.

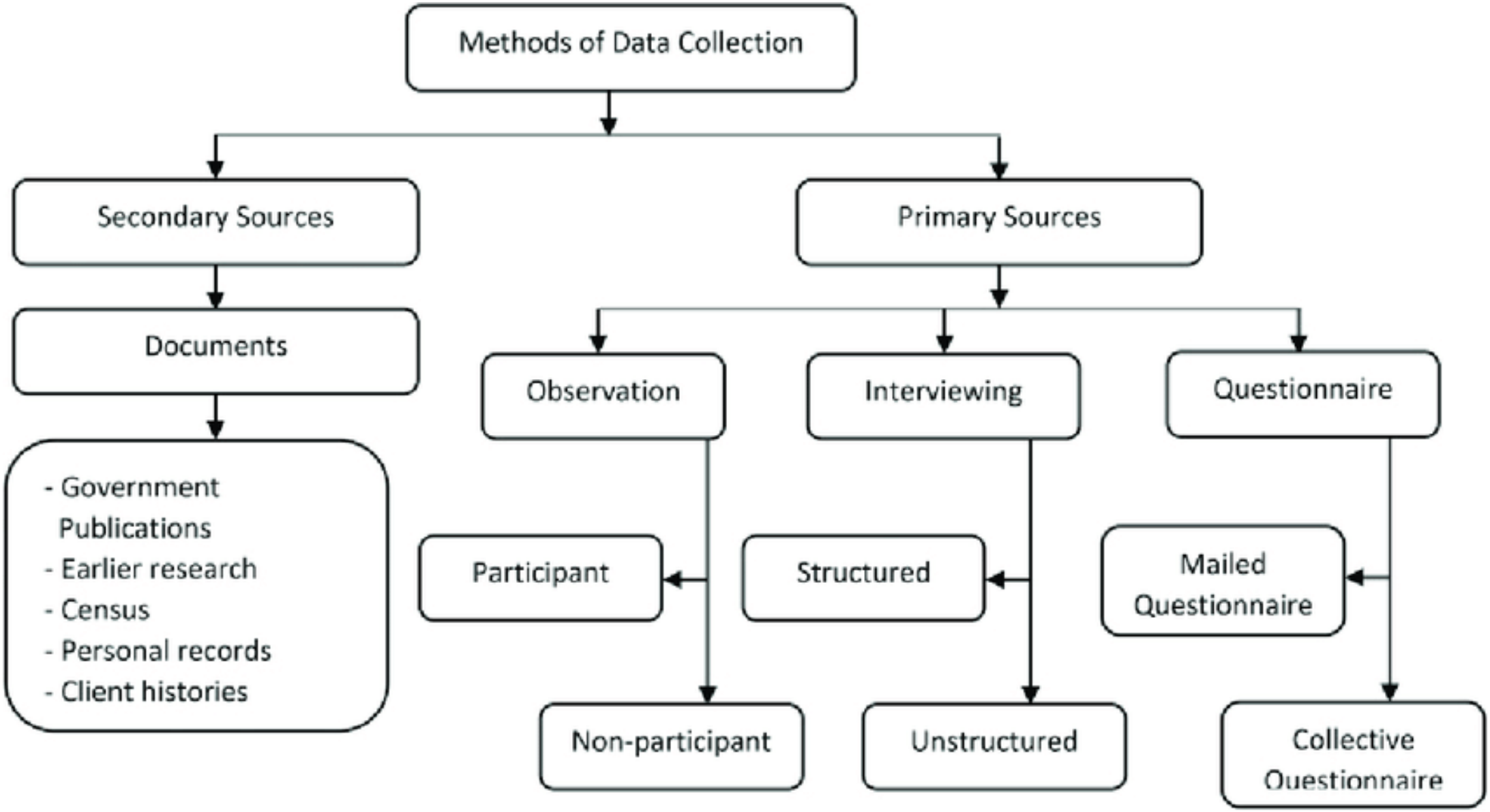
In inferential statistics, predictions are made by taking any group of data in which researchers are interested.

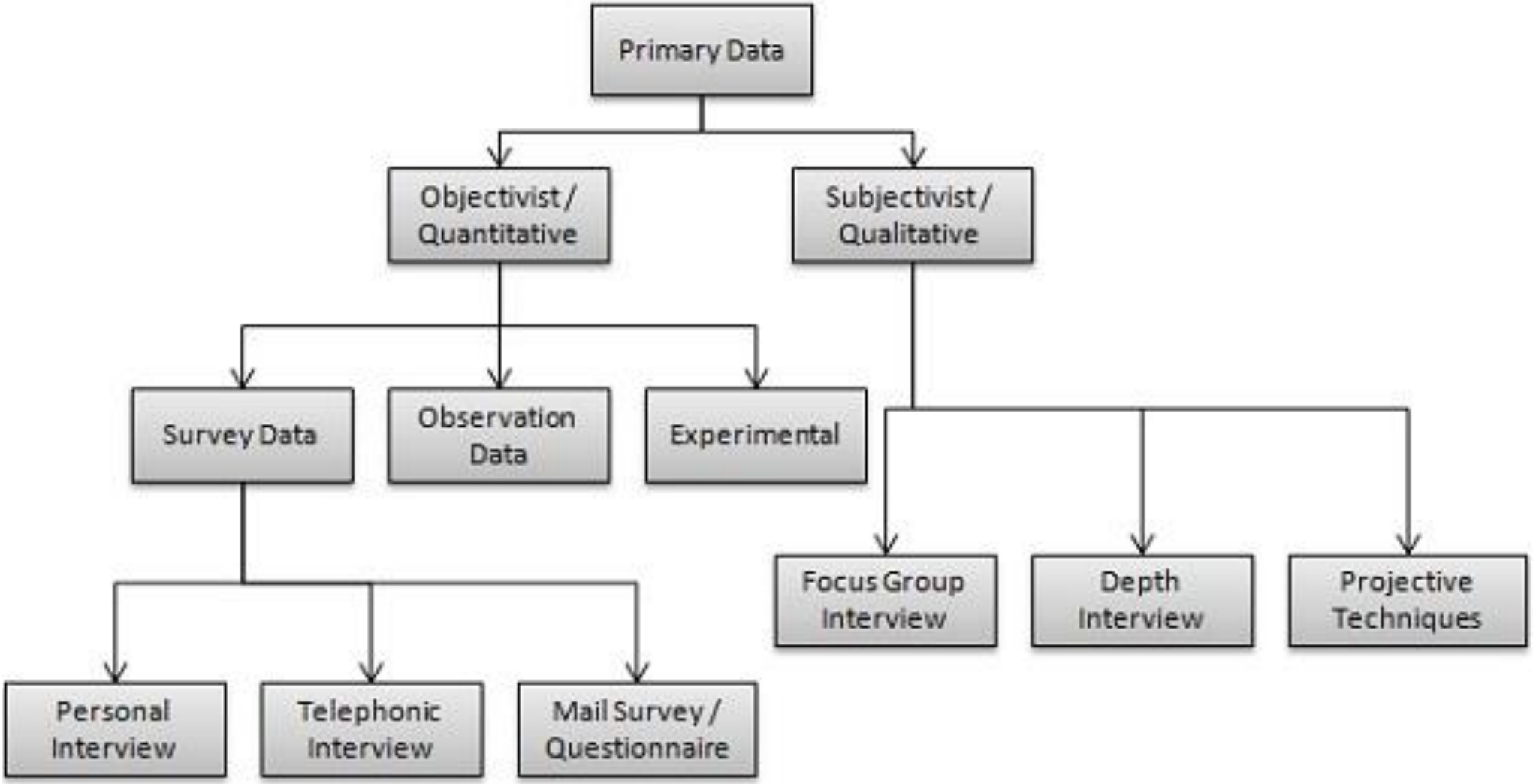
It can be defined as a random sample of data taken from a population to describe and make inferences about the population.

Any group of data that includes all the data researchers are interested in is known as population.

It basically allows researchers to make predictions by taking a small sample instead of working on the whole population.

Data Acquisition


Methods of Data Collection
Secondary Sources
Primary Sources
Documents
- Government Publications
- Earlier research
- Census
- Personal records
- Client histories
Observation
Interviewing
Questionnaire
Participant
Structured
Mailed Questionnaire
Non-participant
Unstructured
Collective Questionnaire

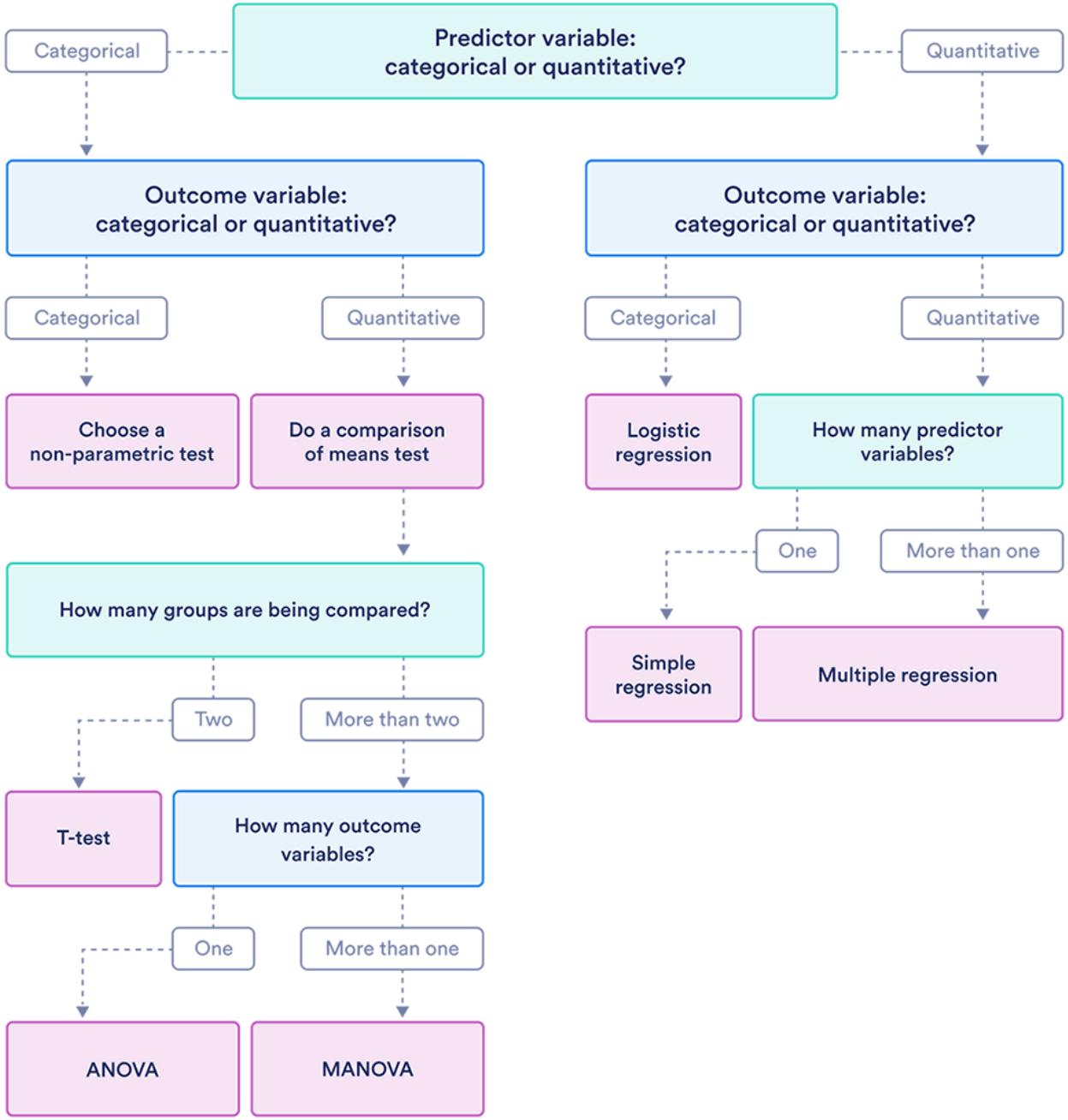

Primary Data
Objectivist / Quantitative
Subjectivist / Qualitative
Survey Data
Observation Data
Experimental
Focus Group Interview
Depth Interview
Projective Techniques
Personal Interview
Telephonic Interview
Mail Survey / Questionnaire

Choosing the Correct Statistical Test

Selection of Appropriate Statistical Method for Data Analysis

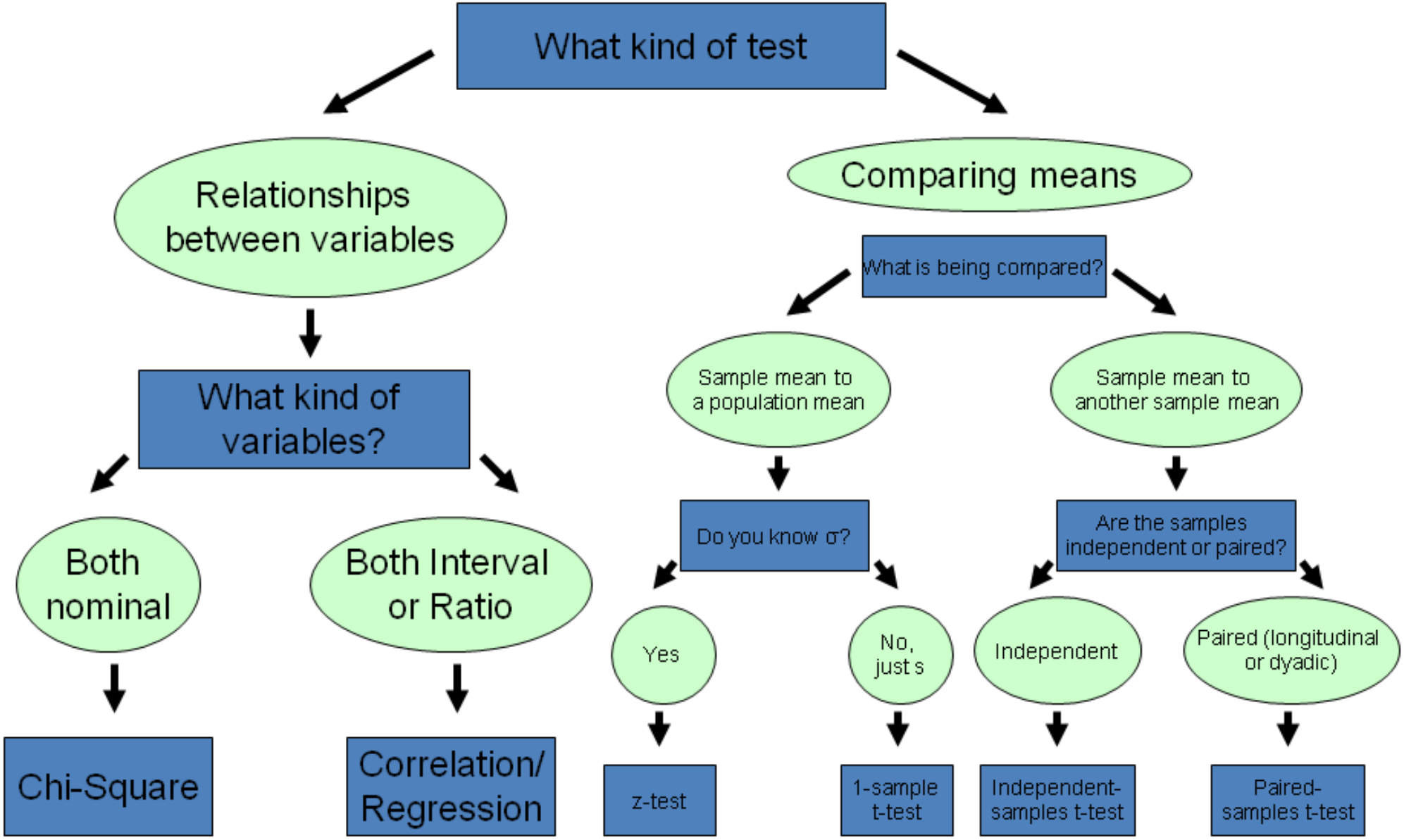
Choosing a statistical test

This flowchart helps you choose among parametric tests



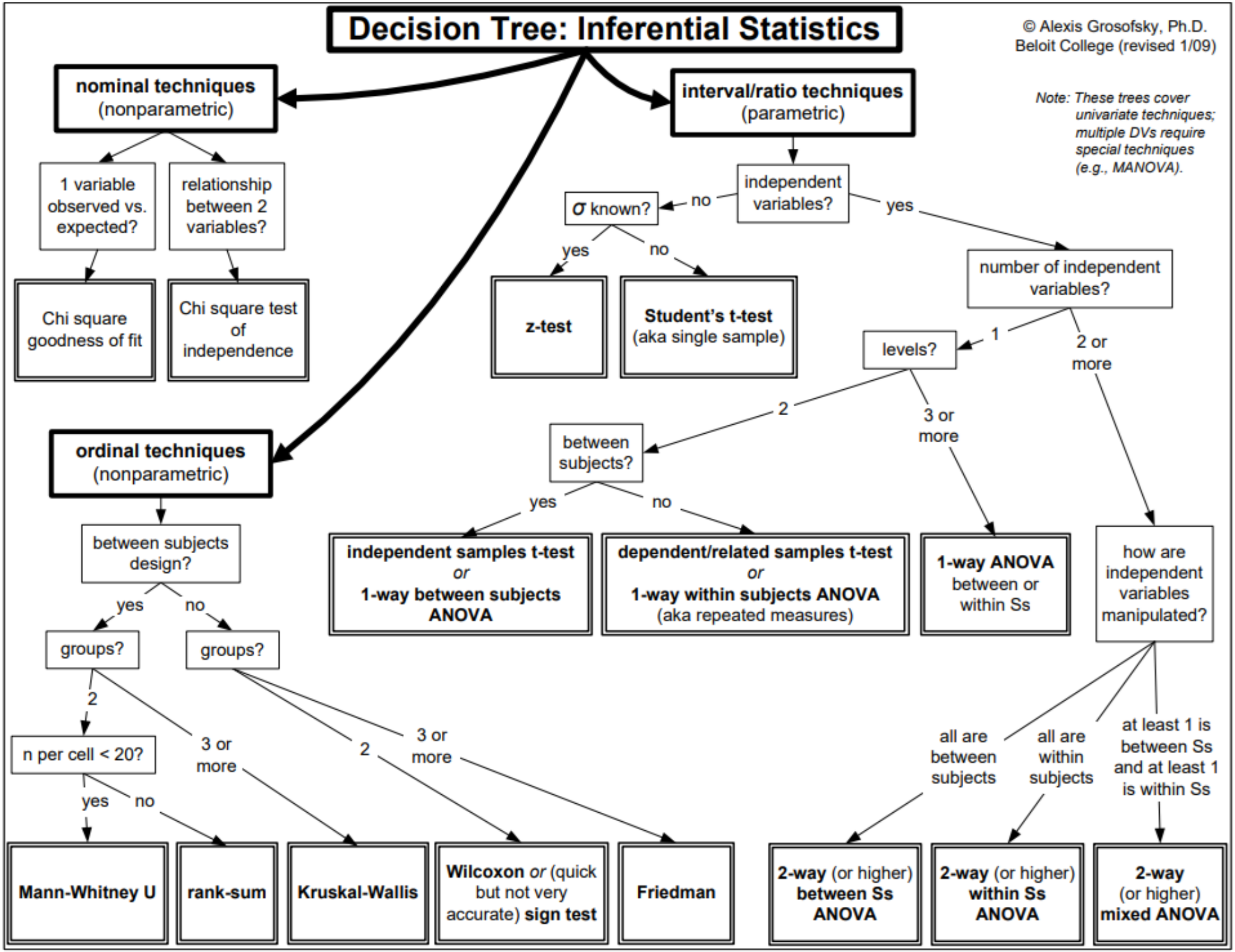

What Type of Research Question?
Are There Many (5+) Ordered Levels of the Independent Variable?
NO
↓
Difference Question (Compares Groups)
YES
↓
Associational Question (Relates Variables)
Is There More Than One Independent Variable?
Is There More Than One Independent Variable?
NO
↓
Basic Single Factor Difference Statistics
Between Groups:
e.g., t tests, Single Factor ANOVA Mann-Whitney U
OR
Within Subjects:
e.g., paired t
YES
↓
Multifactor Difference Statistics
e.g., Between, Within, or Mixed Factorial ANOVAs
NO
↓
Bivariate Associational Statistics
e.g., Pearson Correlation, Spearman Rho
YES
↓
Multivariate Associational Statistics
e.g., Multiple Regression

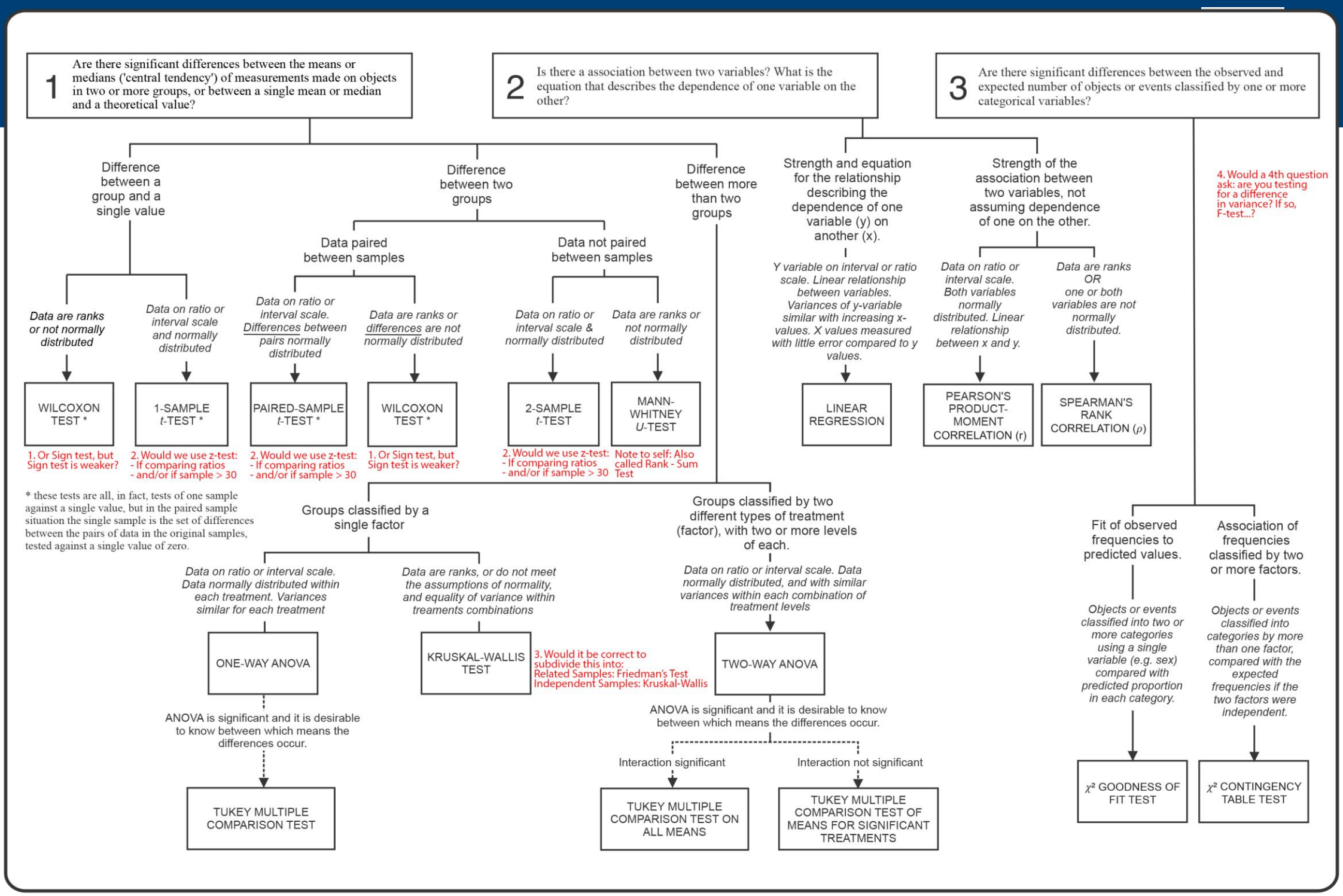
Decision Tree



Decision Tree: Inferential Statistics

Note: These trees cover univariate techniques; multiple DVs require special techniques (e.g., MANOVA).





Hypothesis Testing

Type of question:

Do data match an expected ratio?

Type of data: discrete, categorical (counts,frequencies)

a priori expectation? yes / no

one variable, one sample; # categories: two / more than two

- two: χ^2-goodness of fit test with Yates's corr.; Fischer's exact
- more than two: χ^2-goodness of fit test

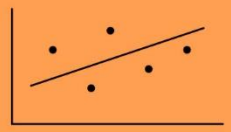
one variable, two or more samples, two or more categories

- χ^2-test for heterogeneity

two variables, both variables have two categories / at least one variable has more than two categories

- both variables have two categories: χ^2-test for association with Yates's corr.; Fischer's exact
- at least one variable has more than two categories: χ^2-test for association

Is there an association between two variables?



Type of data: non-parametric (nominal, ordinal, interval) / parametric (interval)

non-parametric (nominal, ordinal, interval):

two variables, data can be ranked (ordinal, interval)

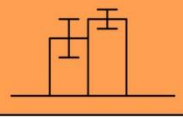
- Spearman's rank correlation

parametric (interval):

two variables: expl. var. under control or with smaller error than resp. var. / neither one under control and with similar error

- expl. var. under control or with smaller error than resp. var.: simple linear regression (**) / principal axis regression (**)
- neither one under control and with similar error: Pearson's product moment correlation

Do samples come from the same or different populations?



Type of data: non-parametric / parametric

non-parametric:

treatment variables: one / two

- one: # categories: two / more than two
  - two: data: unpaired / paired
    - unpaired: # observ. per sample: less than 30 / at least 30
      - less than 30: Wilcoxon's rank paired test (***)
      - at least 30: *z*-test for paired data (*)
    - paired: one-way anova, Kruskal-Wallis
  - more than two: two-way non-parametric anova

parametric:

treatment variables: one / two or more

- one: # categories: two / more than two
  - two: data: unpaired / paired
    - unpaired: variances homogeneous: yes / no
      - yes: # observ. per sample: less than 30 / at least 30
        - less than 30: *t*-test for unpaired data (*)
        - at least 30: *z*-test for unpaired data (*)
      - no: Welch *t*-test
    - paired: # observ. per sample: less than 30 / at least 30
      - less than 30: *t*-test for paired data (*)
      - at least 30: *z*-test for paired data (*)
  - more than two: one-way parametric anova (*), (**)
- two or more: multi-way parametric anova (possibly nested) (*), (**)

Always check the literature for the details!

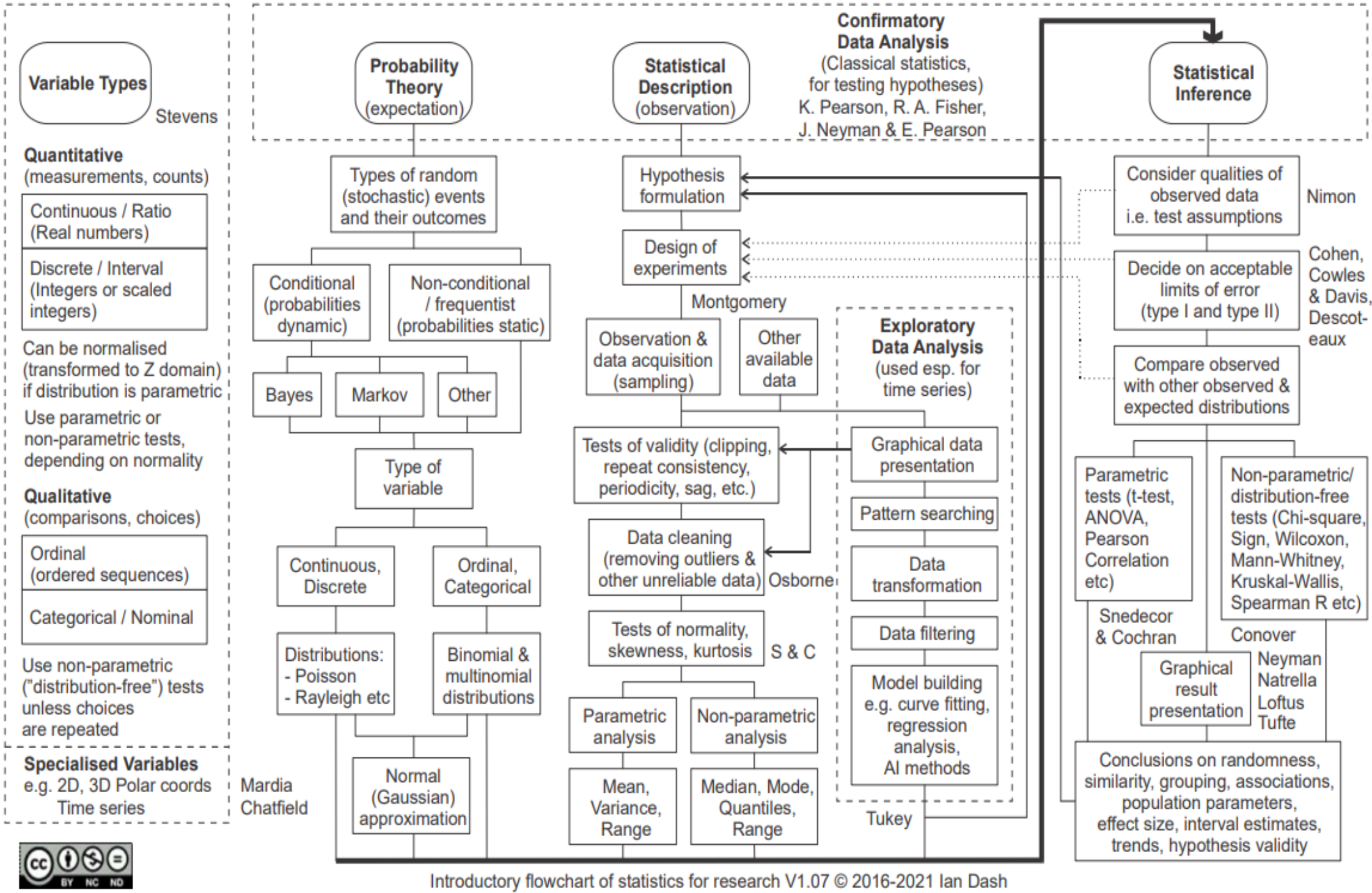
(*): requires homogeneous variances (*F* test)
(**): check for normally distributed residuals
(***): requires similarly shaped distributions

If you have data for which no test seems available, try to transform your data.

Statistics

"A method for extracting meaning from numerical data"

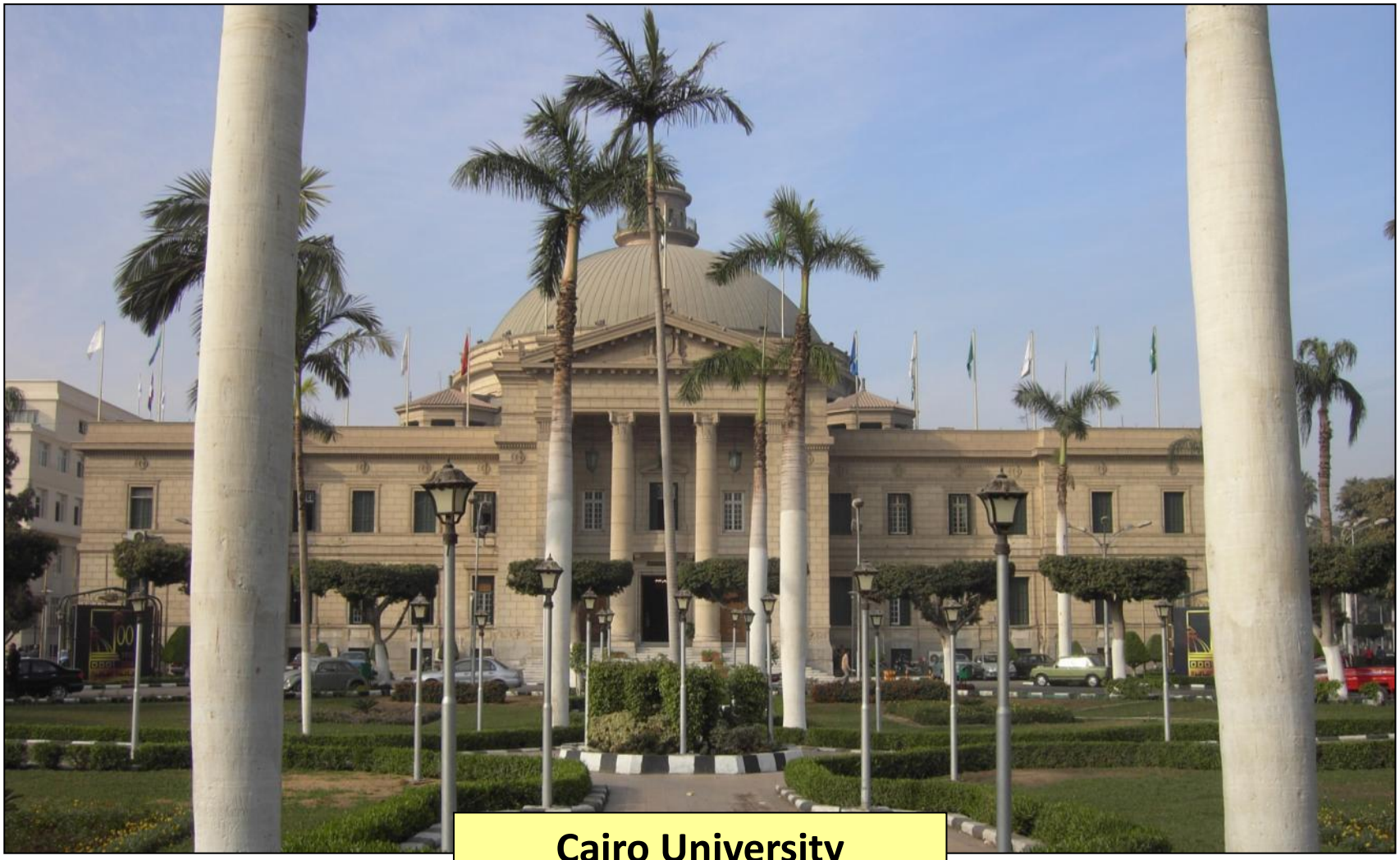
"The science of uncertainty"



Examples

Exercises

Thank You!



Cairo University