Statistical Methodology for Software Engineering

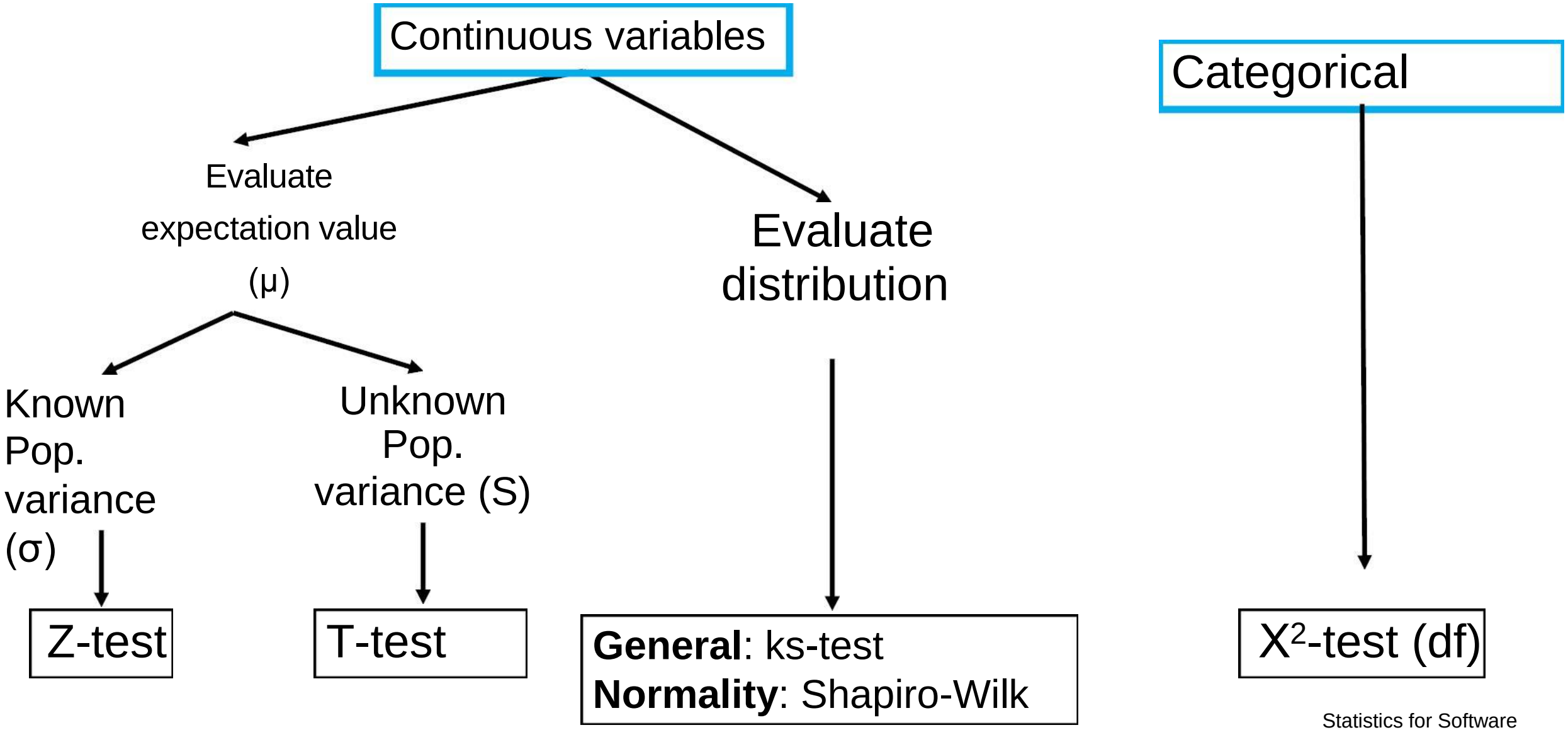
Hadas Lapid, PhD

Contents

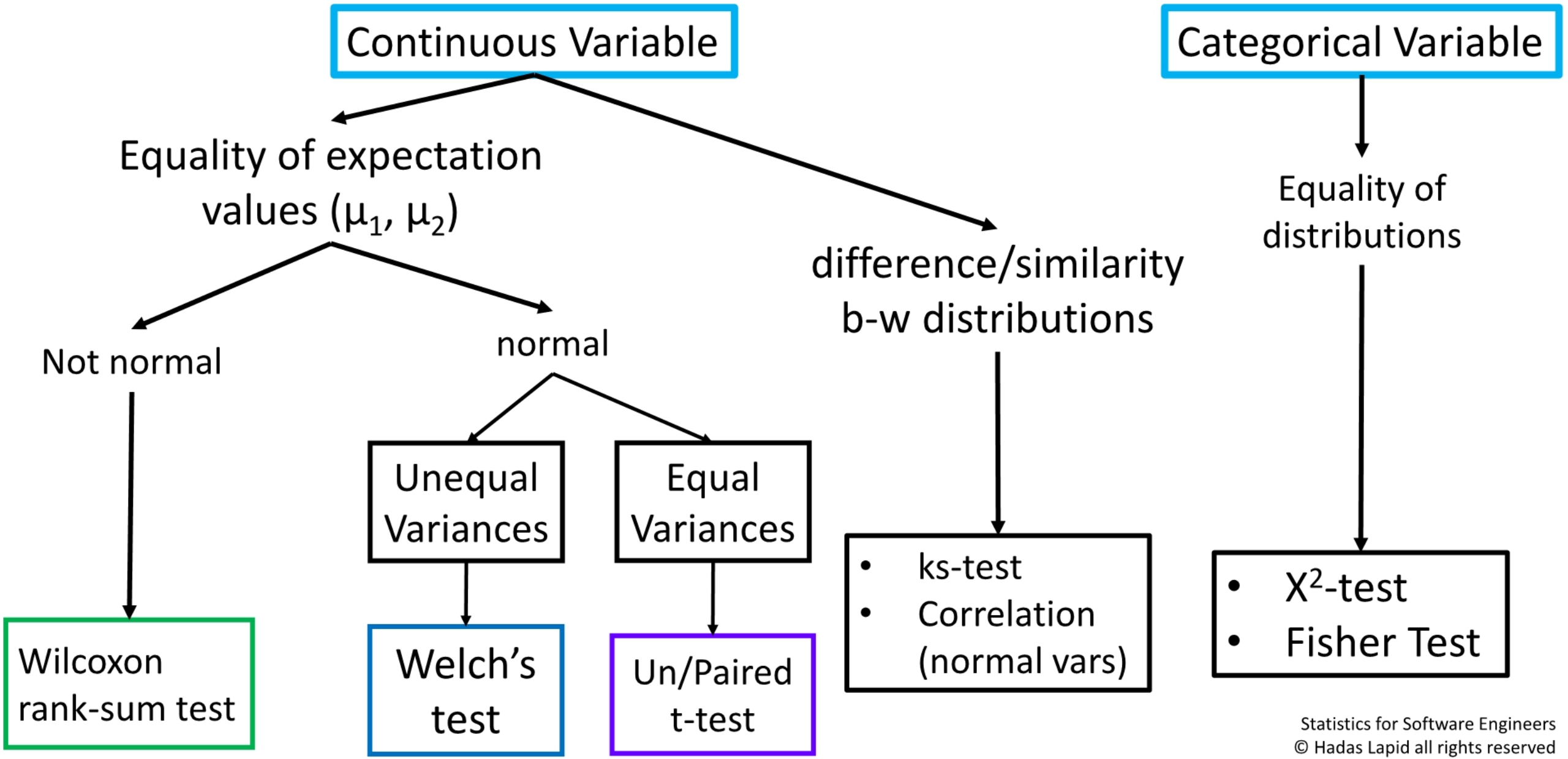
How to choose correct statistical analysis

- Single sample
- Two samples
- Multiple parameters

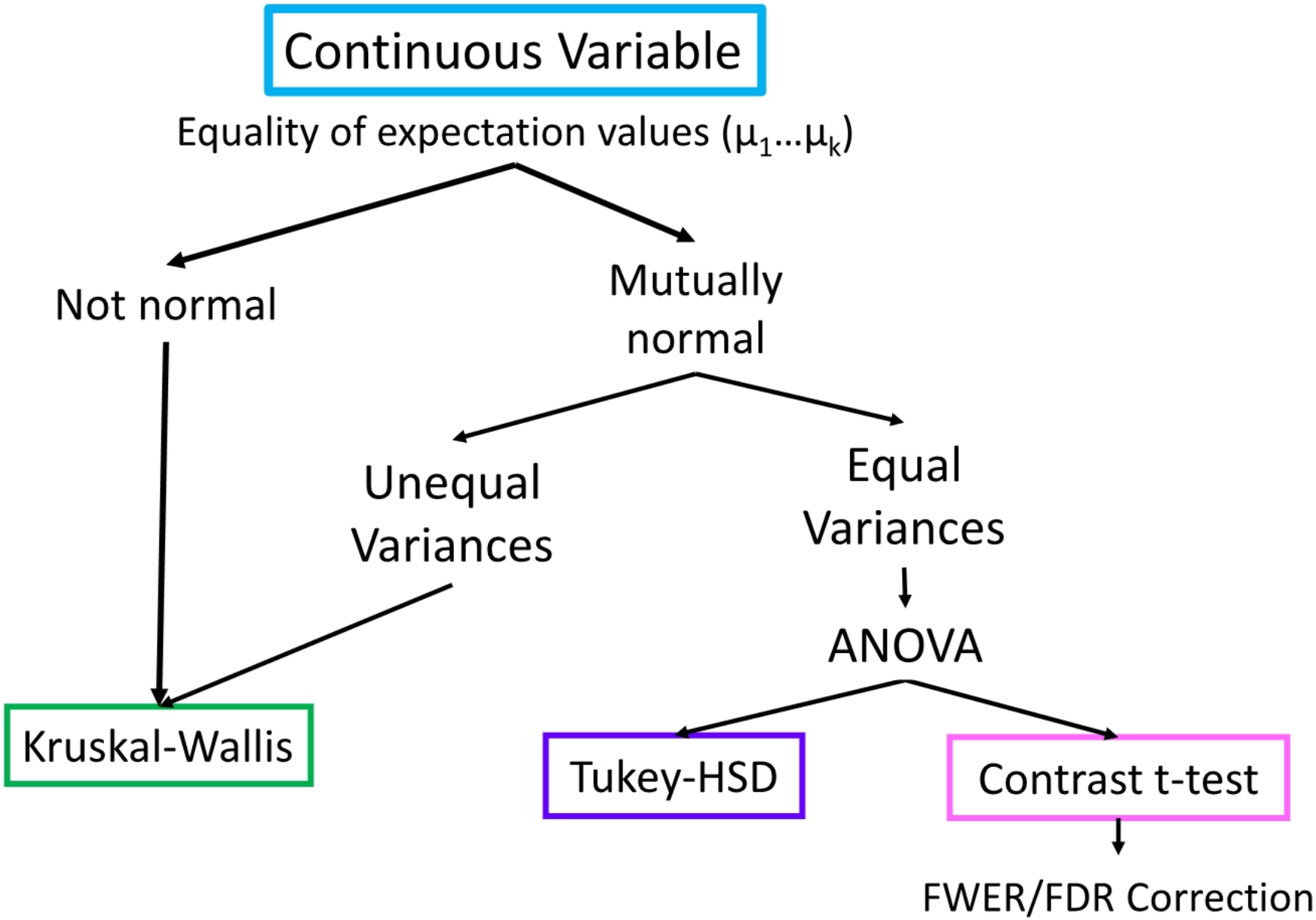
Single sample against theoretical distribution

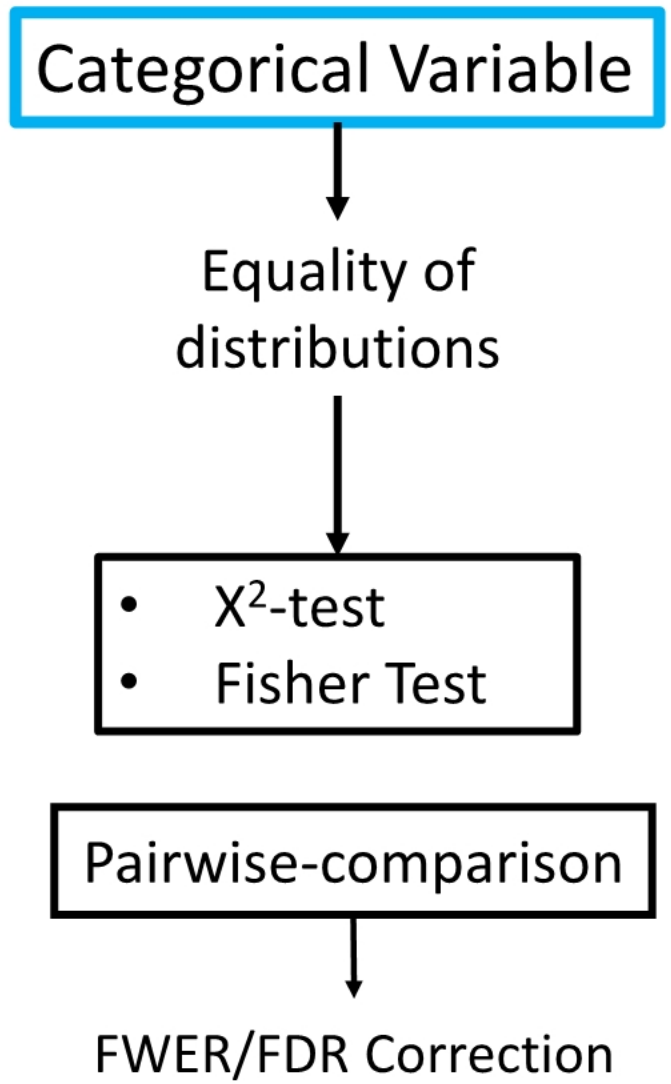


Comparison of two samples



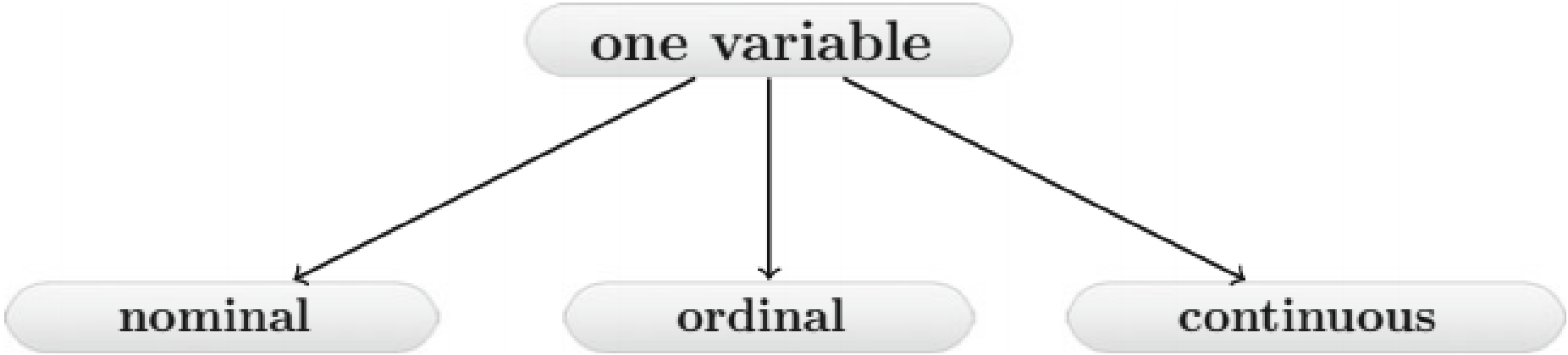
Multiple Comparisons



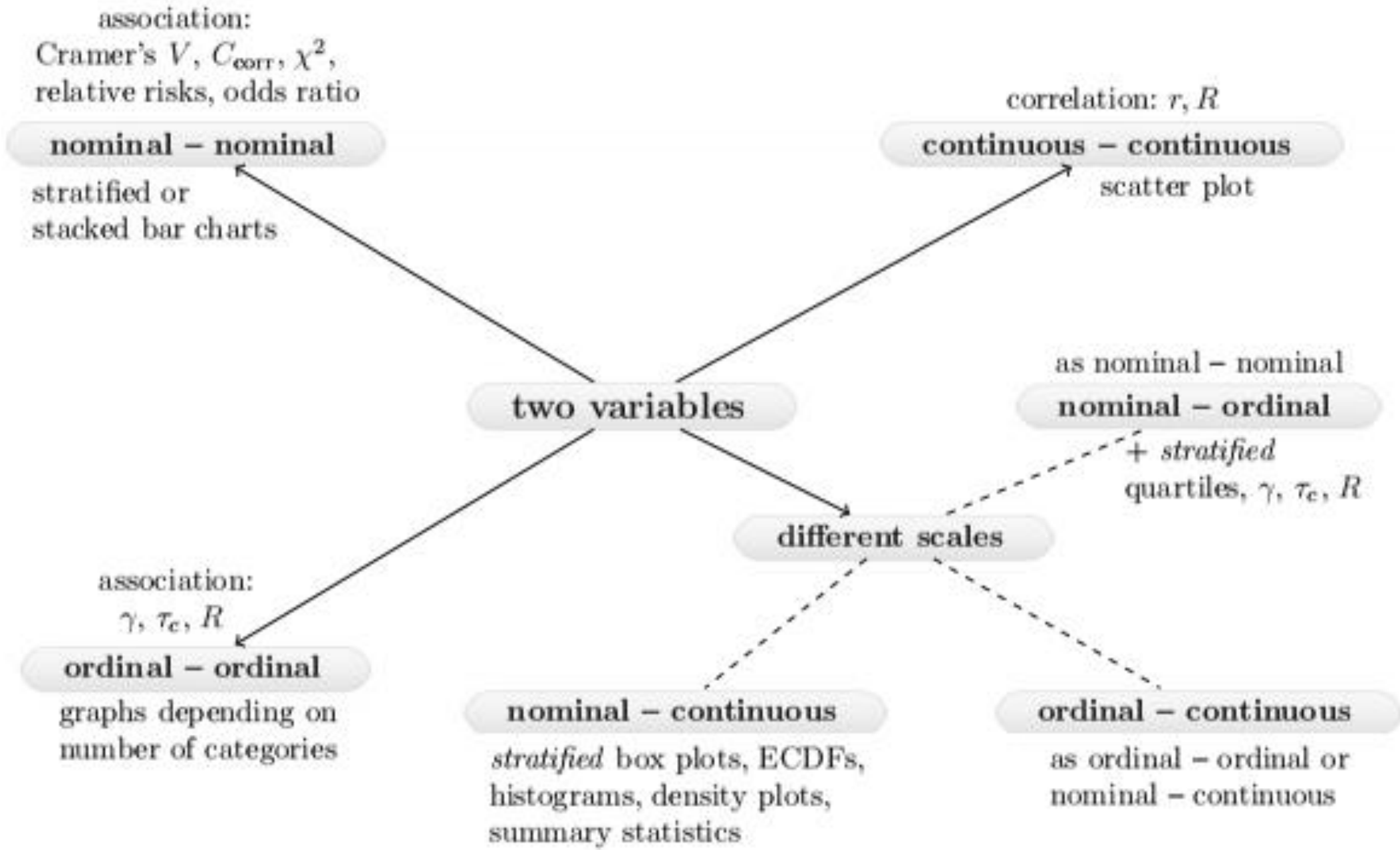


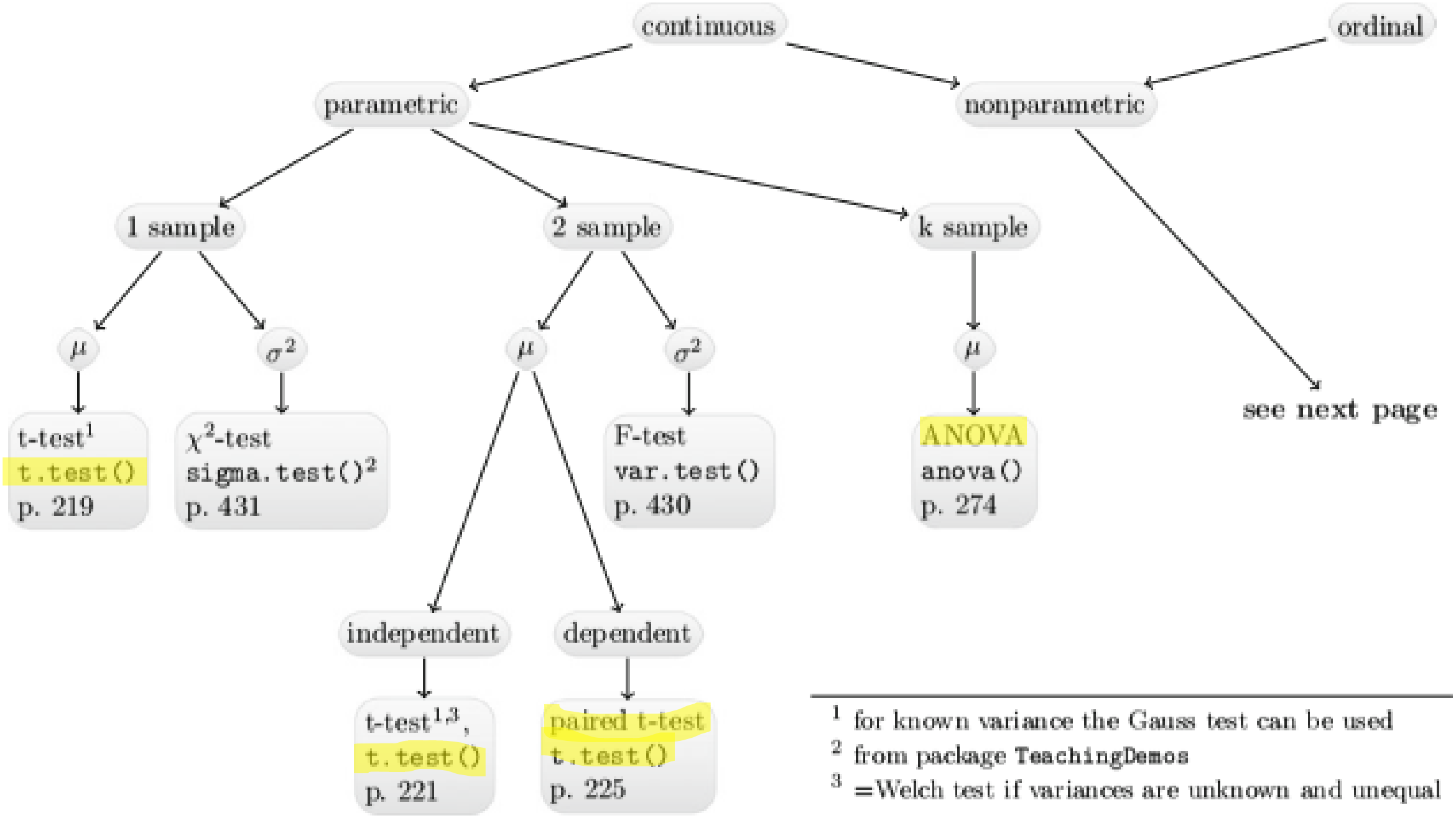
Auxilliary Summaries

Descriptive Data Analysis



nominal	ordinal	continuous
bar chart pie chart freq. table	bar chart pie chart freq. table box plot ECDF	box plot histogram kernel density plot ECDF
mode	median, mode, quartiles, range, IQR	median, mean, quartiles, variance, absolute median deviation, range, IQR

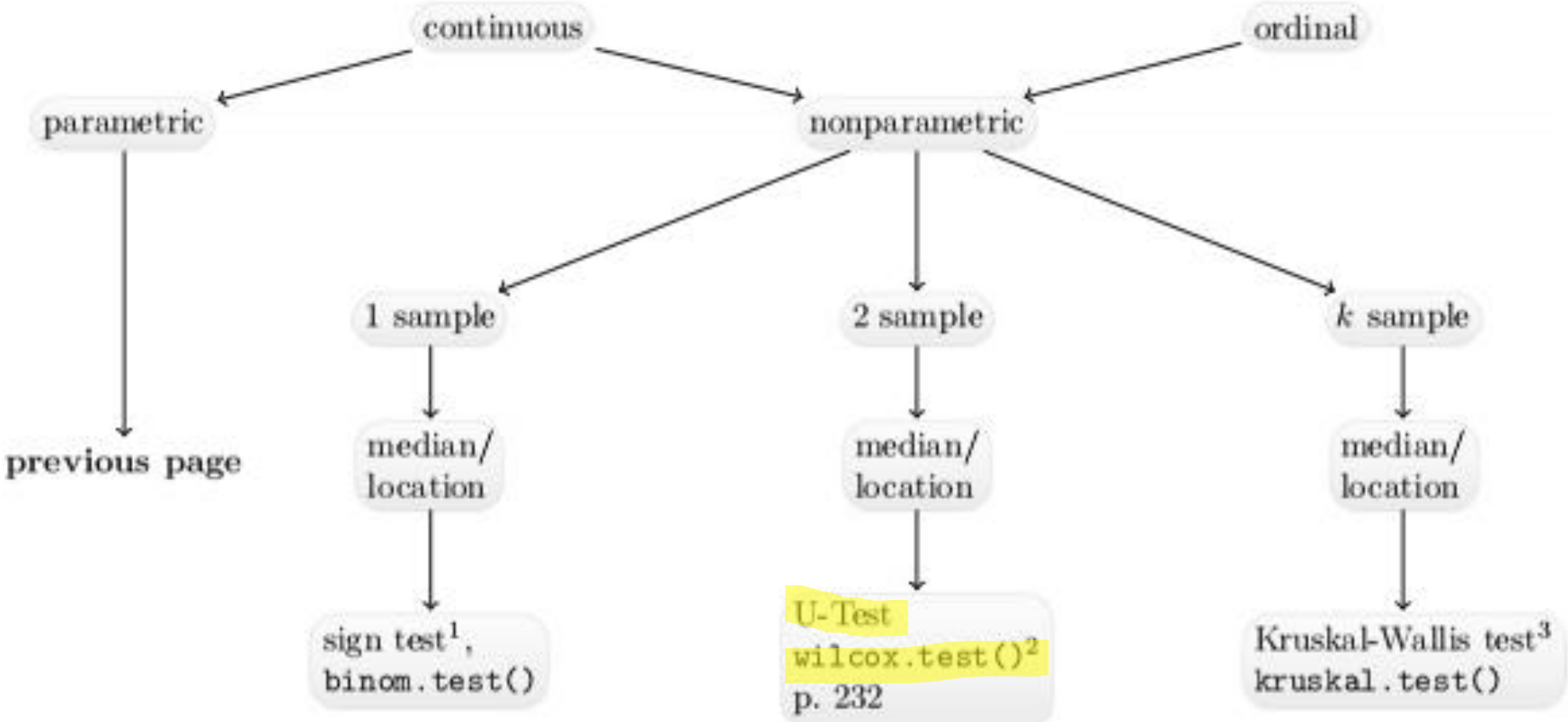

association:
Cramer's V, C_{corr}, χ^2,
relative risks, odds ratio
nominal – nominal
stratified or
stacked bar charts
correlation: r, R
continuous – continuous
scatter plot
two variables
as nominal – nominal
nominal – ordinal
+ stratified
quartiles, γ, τ_c, R
different scales
association:
γ, τ_c, R
ordinal – ordinal
graphs depending on
number of categories
nominal – continuous
stratified box plots, ECDFs,
histograms, density plots,
summary statistics
ordinal – continuous
as ordinal – ordinal or
nominal – continuous



[1] for known variance the Gauss test can be used

[2] from package `TeachingDemos`

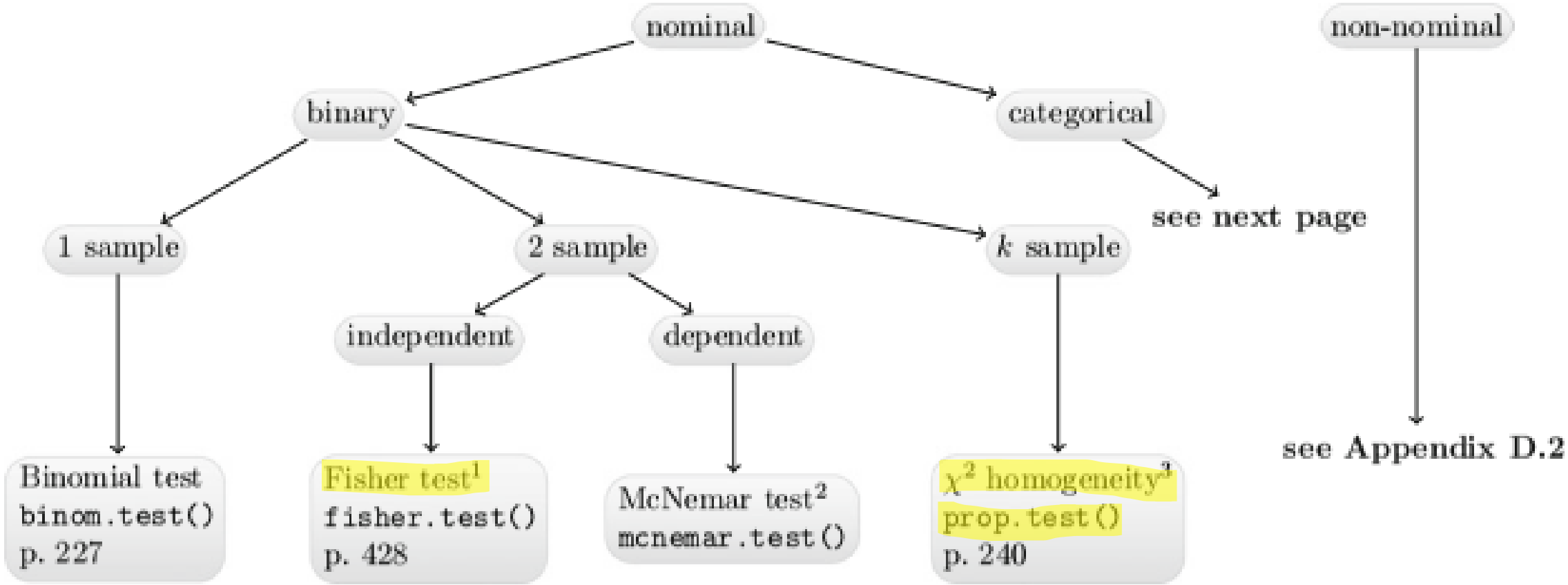
[3] =Welch test if variances are unknown and unequal



[1] not explained in this book; alternative: Mood's median test

[2] use option `paired=TRUE` for dependent data

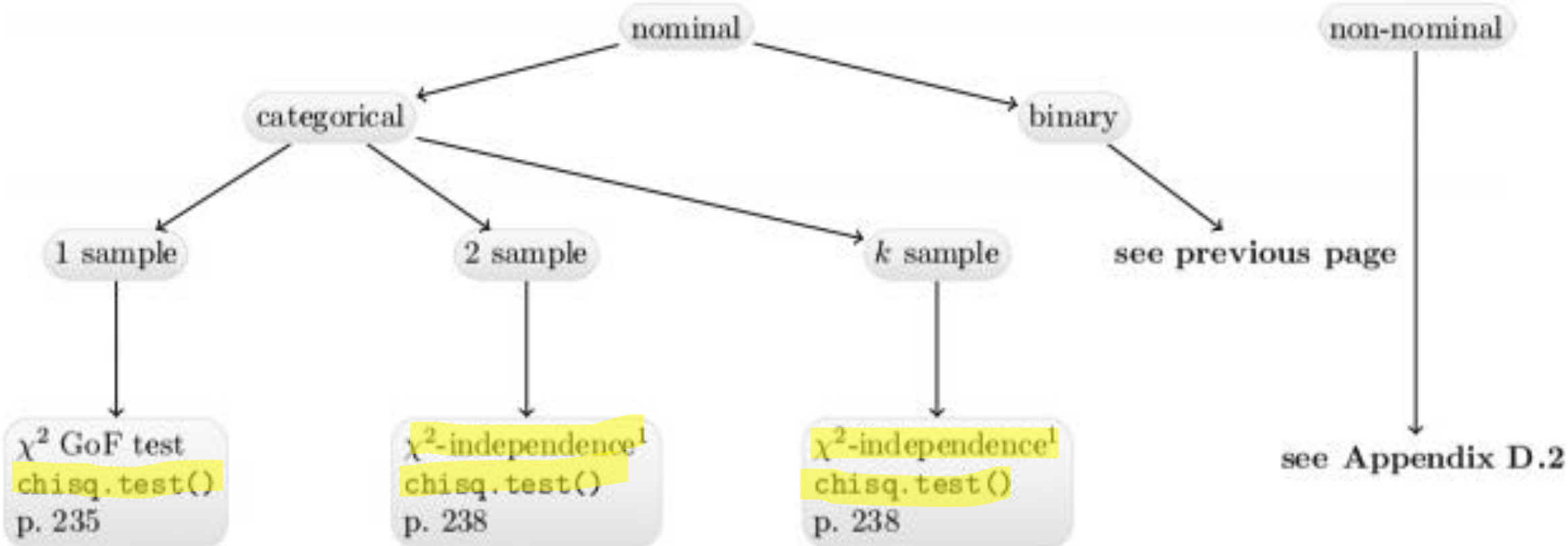
[3] not explained in this book; use Friedman test for dependent data



[1] alternative: χ^2-independence test (`chisq.test()`)
(test decision equivalent to χ^2-homogeneity test)

[2] not explained in this book

[3] test decision equivalent to χ^2-independence test



[1] test decision equivalent to χ^2-homogeneity test (prop.test)