CAS Applied Data Science – Module 2 – Day 3

Statistical Inference for Data Science

Dr. Anja Mühlemann

28. August 2025

UNIVERSITÄT BERN

Questions from Day 2

Day 3

Hypothesis Testing

Today's Topics

- Hypotheses
- p-values
- Error types
- Frequently used tests

Inferential Statistics

Inferential Statistics

With a certain degree of certainty, one would like to draw conclusions from empirical data, even if the data are subject to error or incomplete.

3 main techniques

- **Parameter estimates:** Calculation estimate for unknown parameter of underlying probability distribution
- **Confidence intervals:** Calculation of a region within which unknown parameter should lie with certain degree of certainty
- **Tests:** Tests are intended to prove that a certain effect, e.g. the effect of a vaccine, is indeed present.

Tests

- Method for deciding on the correctness of hypotheses under uncertainty

  e.g., new medication is better than the old one

2 Hypotheses:

- Working hypothesis (H1): Motivation of the study

  e.g., the new medication is better than the old one
- Null hypothesis (H0): Opposite of H1

  e.g. the new medication is not better than the old one

Goal: reject the null hypothesis with some degree of certainty

Tests

- Statistical test rely on a test statistic, for which distribution under the test assumptions and H0 is known.
- We calculate the value of the test statistic for the sample at hand ($\hat{T}$)
- And check whether this value is probable for the distribution under H0.
- To this end the p-value is calculated
- If the p-value < 1 - desired degree of certainty, we reject H0
- Otherwise, we cannot reject H0, which does not necessarily imply that H1 holds

p-value

- *To illustrate what a p-value is, I'll illustrate the **one sample t-test** in a little more detail*
- **Assumptions:** independent observations, approx. normal
- **Possible hypotheses:**

$$H_0\colon \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0,$$
$$H_0\colon \mu \leq \mu_0, \qquad H_1 : \mu > \mu_0,$$
$$H_0\colon \mu \geq \mu_0, \qquad H_1\colon \mu < \mu_0,$$

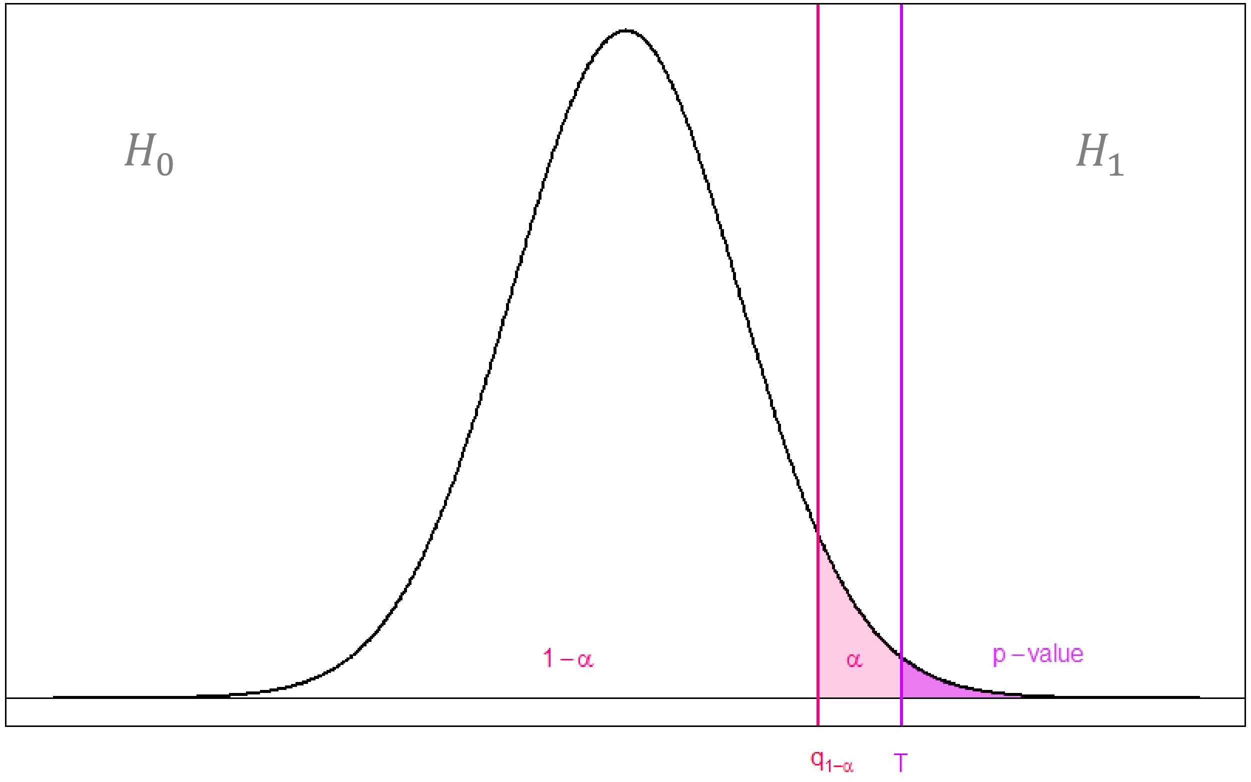
- **Test statistic:**

$$T = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}}\sqrt{n}$$

- Thus, T is large (≥ 0) if H_1 holds, and T is small (< 0) if H_0 holds

p-value

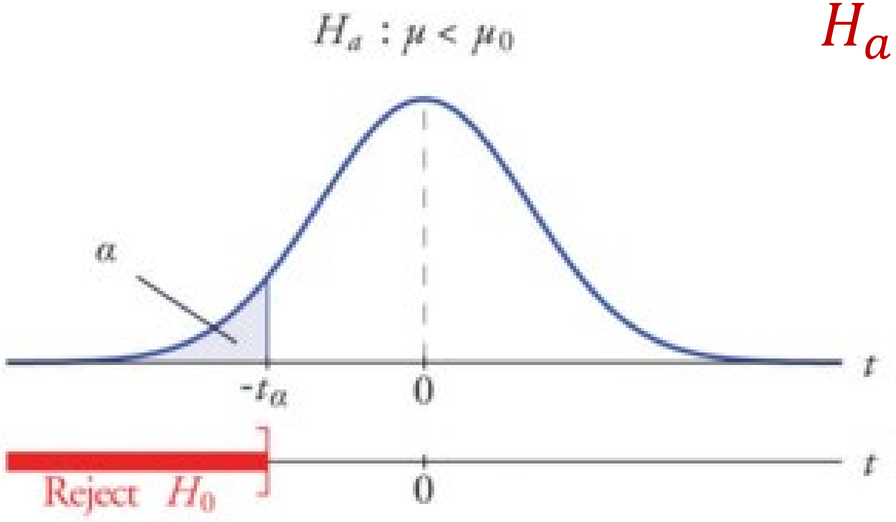
- **Distribution:** One can show that if H_0 holds, then T follows a t-distribution with $n-1$ degrees of freedom

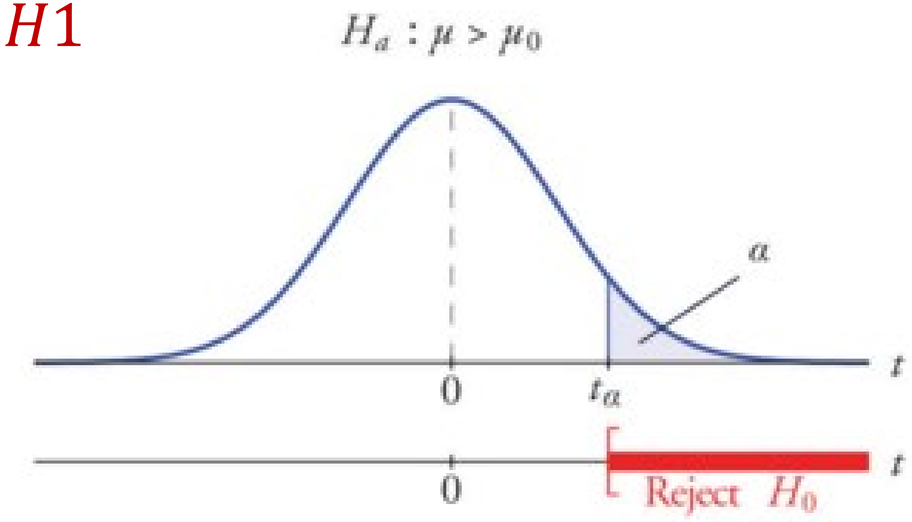


- Thus, the p-value is «*the probability to observe a even more extreme value in terms of H_0 than the one at hand*»

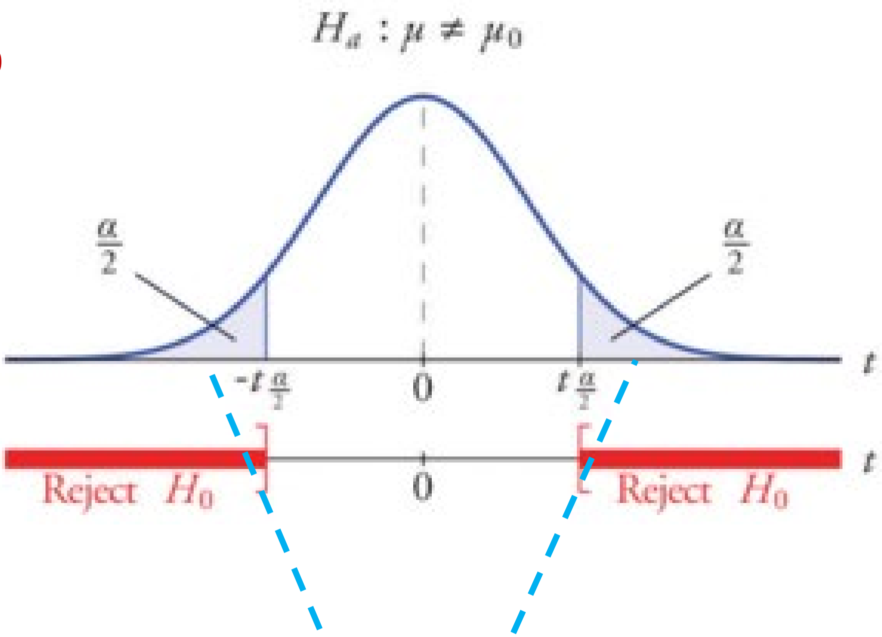
Tests

$H_a = H1$





Typically $\alpha = 5\%$



Errors

- Type 1: Wrongly reject the null hypothesis due to a fluctuation (false positive)
- Type 2: Wrongly keep the null hypothesis by interpreting a real effect as a fluctuation (false negative)

		Reality	
		True	False
Measured or Perceived	True	Correct ☺	Type 1 error False Positive
	False	Type 2 error False Negative	Correct ☺

Prison example

Innocent person set free	Innocent person jailed
Guilty person set free	Guilty person jailed

Types of Tests

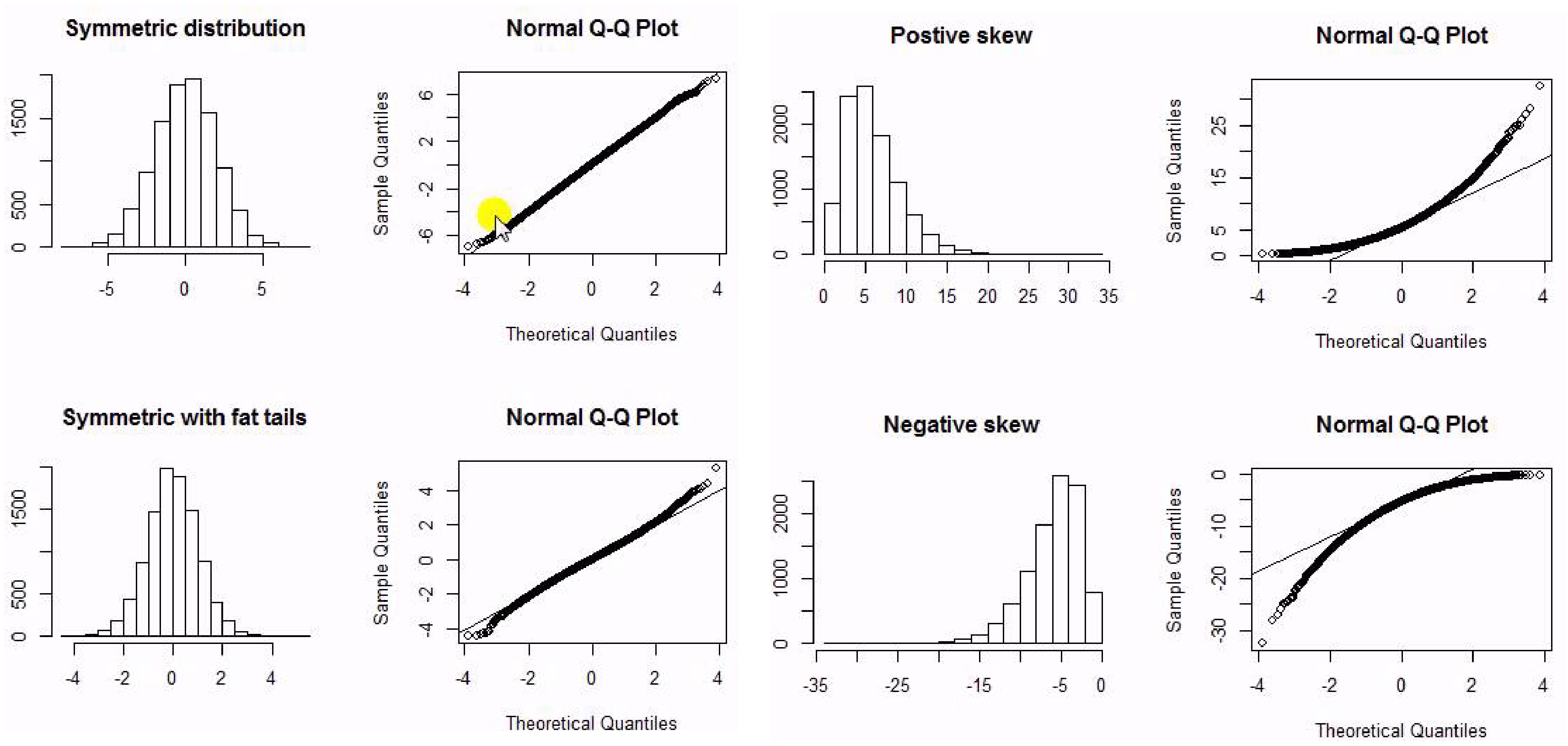
- One group: the mean monthly income is larger than 5000.-
- Two groups: the mean income of men is larger than that of women
- ≥ Three groups: effect of tea on weight loss (green, black, none)

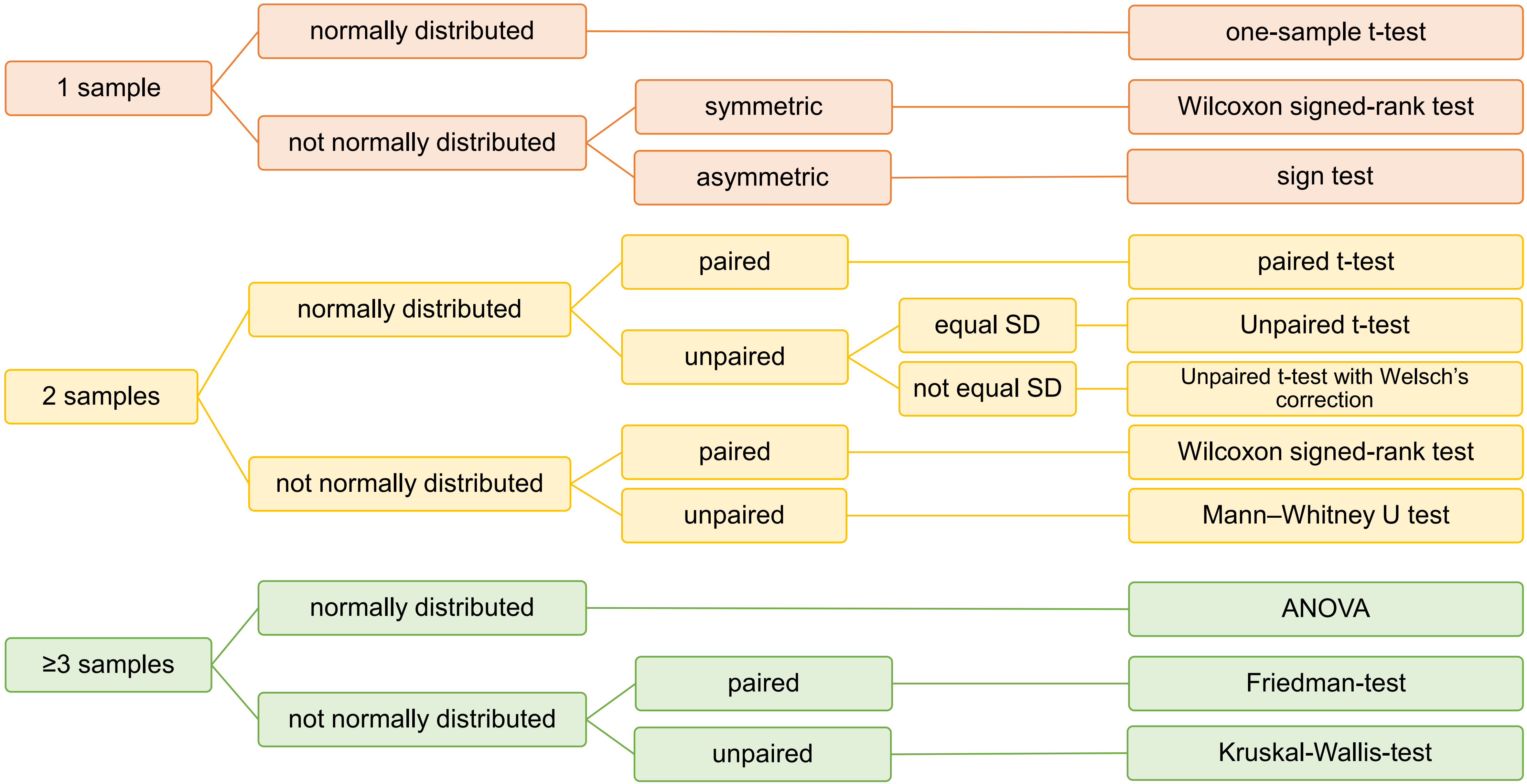
≥ Two groups:

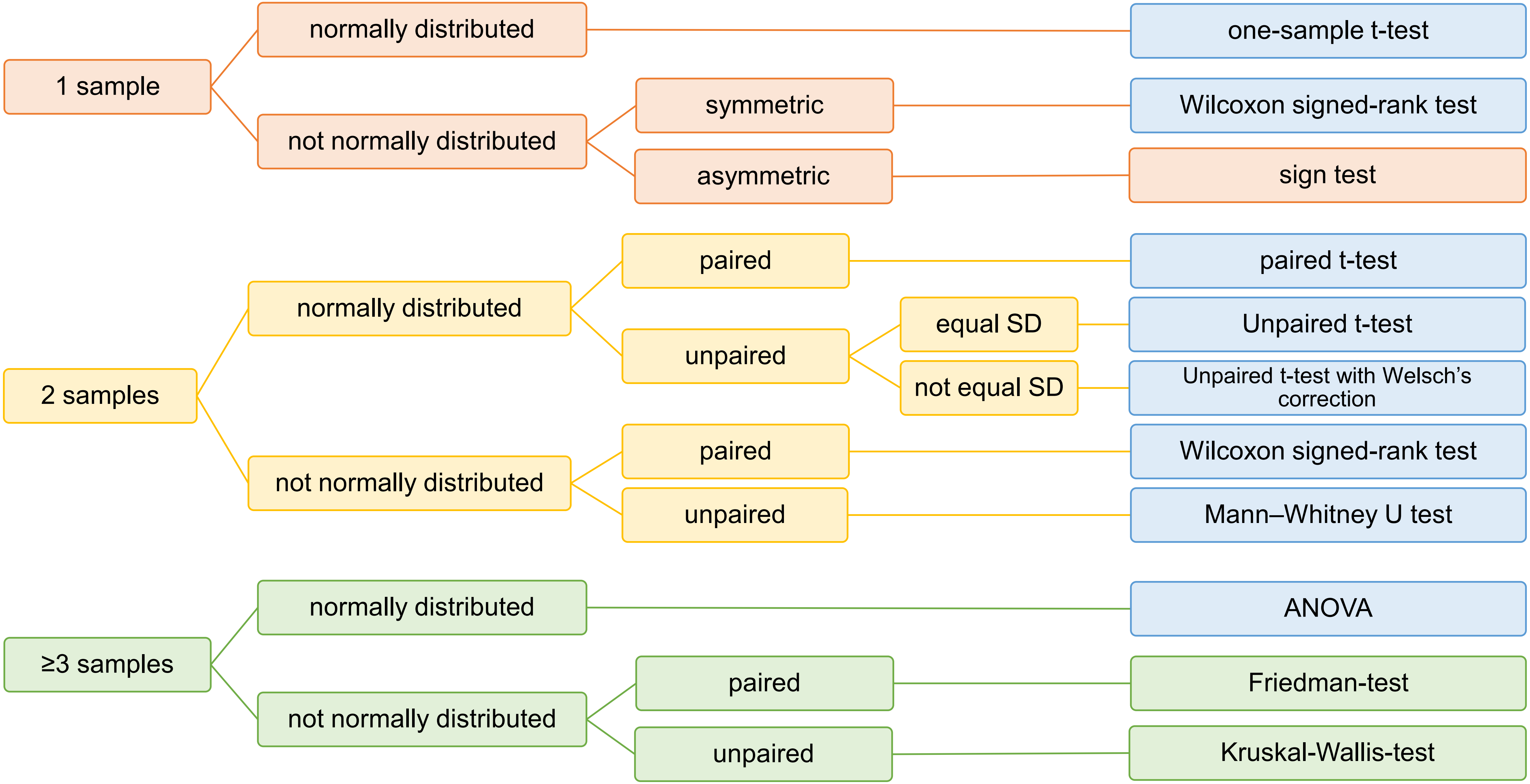
- Paired: dependent, repeated measurements on same individual, e.g. blood pressure before and after surgery
- Unpaired: independent, from separate individuals, e.g. blood pressure after medication 1 vs. blood pressure after medication 2

Normality

- Many test assume that the sample comes from a normal distribution
- Thus, we need to check whether this is fulfilled before performing such a test
- Shapiro-Wilk test, Shapiro-Francia test, Q-Q-Plot, …




1 sample
normally distributed
one-sample t-test
not normally distributed
symmetric
Wilcoxon signed-rank test
asymmetric
sign test
2 samples
normally distributed
paired
paired t-test
unpaired
equal SD
Unpaired t-test
not equal SD
Unpaired t-test with Welsch's correction
not normally distributed
paired
Wilcoxon signed-rank test
unpaired
Mann–Whitney U test
≥3 samples
normally distributed
ANOVA
not normally distributed
paired
Friedman-test
unpaired
Kruskal-Wallis-test


1 sample
normally distributed
one-sample t-test
not normally distributed
symmetric
Wilcoxon signed-rank test
asymmetric
sign test
2 samples
normally distributed
paired
paired t-test
unpaired
equal SD
Unpaired t-test
not equal SD
Unpaired t-test with Welsch's correction
not normally distributed
paired
Wilcoxon signed-rank test
unpaired
Mann–Whitney U test
≥3 samples
normally distributed
ANOVA
not normally distributed
paired
Friedman-test
unpaired
Kruskal-Wallis-test

Exercise

- 3 Slides to be uploaded to ILIAS today
  - 1 slide: Question that the test tries to answer, assumptions on data, other details
  - 1 slide: example from “real live” (if possible)
  - 1 slide: your conclusion from the Notebook on this test
- Will be presented at tomorrow’s discussion session

Nr	Test
1	One-sample t-test
2	One-sample Wilcoxon SR test
3	Paired t-test
4	Paired Wilcoxon SR test
5	Unpaired t-test
6	Unpaired t-test with Welsch’s correction
7	Mann-Withney U test
8	One-way ANOVA