



# Is my sample size large enough? Parametric vs nonparametric tests

With a demo using XLSTAT

---

Jean-Paul Maalouf  
jpmaalouf@xlstat.com

**Addinsoft**





## Webinar plan

- **Principles of statistical testing**
- **Parametric vs nonparametric tests (Power vs robustness)**
- **Why does sample size matter?**
  - > Large sample size = better representativeness
  - > Large sample size allows the verification of parametric conditions
- **XLSTAT: ANOVA vs Kruskal-Wallis**
- **Take home message**
- **Survey**



## Principles of statistical testing

- Two **competing** hypotheses
- **Null Hypothesis  $H_0$** : treatments have **no effect** the studied population
- **Alternative hypothesis  $H_a$** : treatments induce **change** in the studied population
- A test allows to **reject/not to reject  $H_0$**
- Tests involve an **extrapolation** from an observed sample to the (unobserved) population

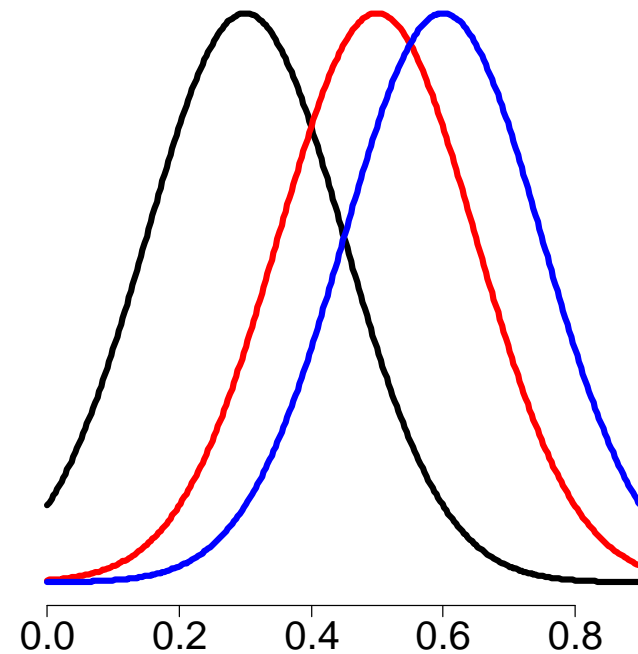
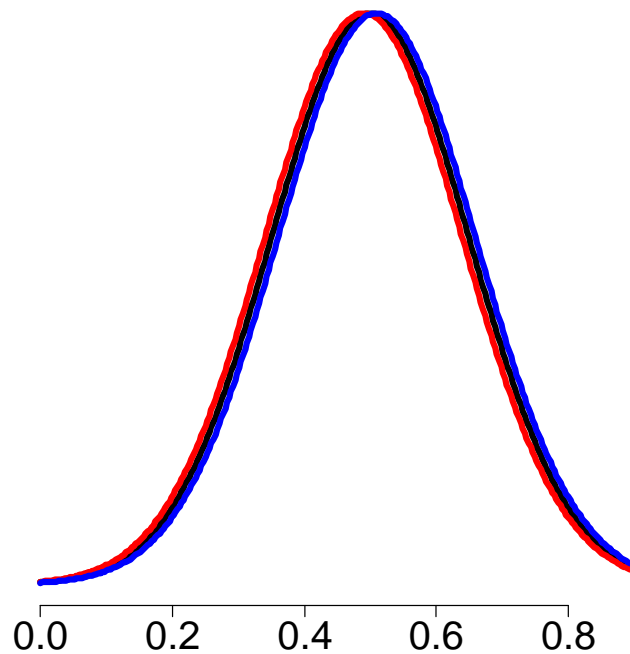


## Principles of statistical testing

- Example: do medications A or B induce a change in the insulin rate of mice?

**$H_0$ :** *insulin(Control mice)*  
= *insulin(medic. A)*  
= *insulin(medic. B)*

**$H_a$ :** *insulin(Control mice)*  
 $\neq$  *insulin(medic. A)*  
 $\neq$  *insulin(medic. B)*



Insulin rate

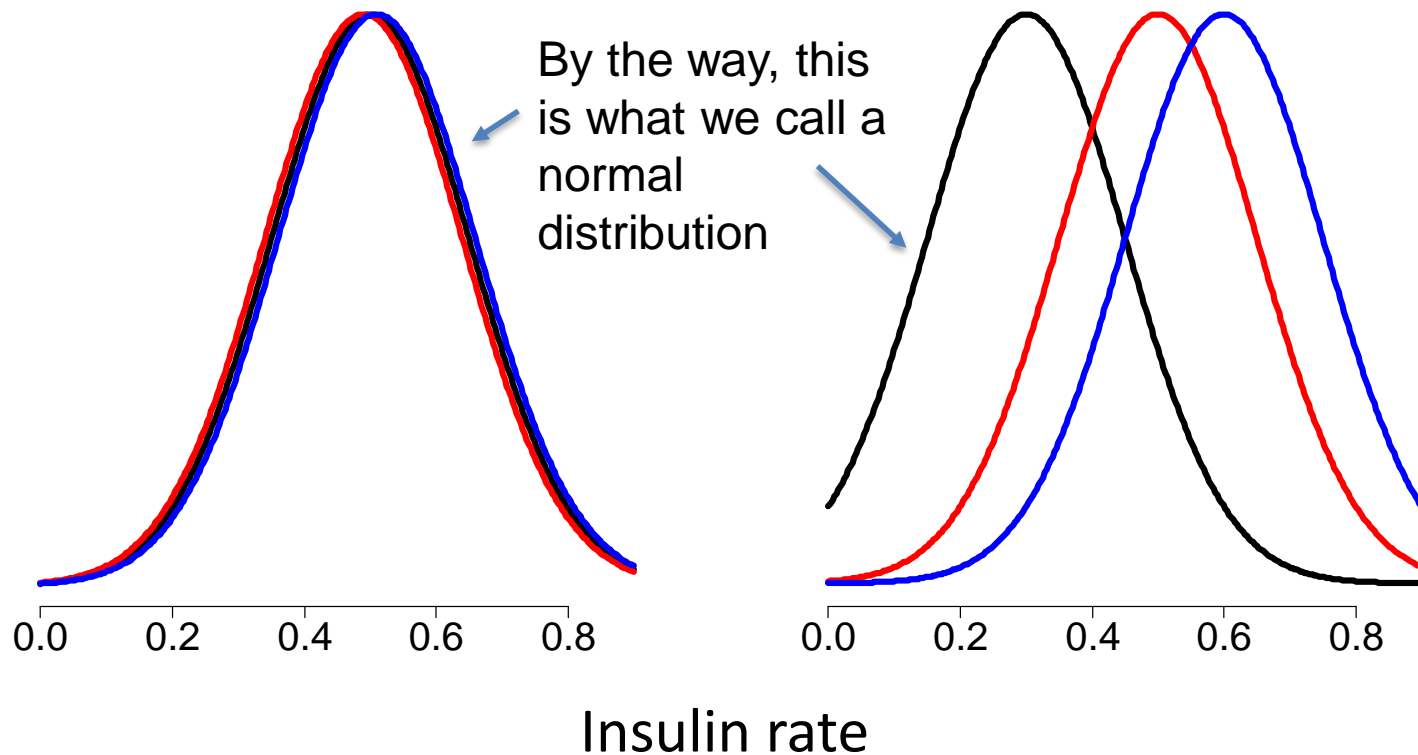


## Principles of statistical testing

- Example: do medications A or B induce a change in the insulin rate of mice?

**$H_0$ :**  $insulin(\text{Control mice})$   
 $= insulin(\text{medic. A})$   
 $= insulin(\text{medic. B})$

**$H_a$ :**  $insulin(\text{Control}) \neq insulin(A)$   
and/or  $insulin(A) \neq insulin(B)$   
and/or  $insulin(\text{Control}) \neq insulin(B)$





## Principles of statistical testing

**$H_0$** :  $insulin(\text{Control mice}) = insulin(\text{medic. } A) = insulin(\text{medic. } B)$

**$H_a$** :  $insulin(\text{Control}) \neq insulin(A)$  and/or  $insulin(A) \neq insulin(B)$   
and/or  $insulin(\text{Control}) \neq insulin(B)$

We define a risk threshold **alpha** (= 0.05, commonly)

We run a statistical test (ANOVA for example) and obtain a **p-value**

- If **p-value** < 0.05: We reject  $H_0$  and accept  $H_a$ , with a risk proportional to p-value of being wrong
- If **p-value** > 0.05: We do not reject  $H_0$



## Parametric vs nonparametric

- A test may be **parametric** or **nonparametric**
- **Parametric tests** assume the populations are normally distributed (→ **conditions** must be verified).
- **Nonparametric tests** assume no specific distribution in the population.



## Parametric vs nonparametric

- A test may be **parametric** or **nonparametric**
- **Parametric tests** assume the populations are normally distributed (→ **conditions** must be verified).
- **Nonparametric tests** assume no specific distribution in the population.

	Parametric tests	Nonparametric tests
Assume underlying distribution?	Yes	No
Populations should be normal?	Yes	Not necessarily
Variances should be equal?	Yes	Not necessarily





## Parametric vs nonparametric

- A test may be **parametric** or **nonparametric**
- **Parametric tests** assume the population is normally distributed.
- **Nonparametric tests** assume no specific distribution in the population.
- **Robustness**: reliability of the result

	Parametric tests	Nonparametric tests
Assume underlying distribution?	Yes	No
Populations should be normal?	Yes	Not necessarily
Variances should be equal?	Yes	Not necessarily
Robust?	If the conditions are verified	In all situations



## Parametric vs nonparametric

So why do we still use  
parametric tests?



## Parametric vs nonparametric

- A test may be **parametric** or **nonparametric**
- **Parametric tests** assume the population is normally distributed.
- **Nonparametric tests** assume no specific distribution in the population.
- **Robustness**: reliability of the result
- **Power**: ability to reject  $H_0$  if it is false

	Parametric tests	Nonparametric tests
Assume underlying distribution?	Yes	No
Populations should be normal?	Yes	Not necessarily
Variances should be equal?	Yes	Not necessarily
Robust?	If the conditions are verified	In all situations
Powerful?	More	Less



## Parametric vs nonparametric

- A test may be **parametric** or **nonparametric**
- **Parametric tests** assume the population is normally distributed.
- **Nonparametric tests** assume no specific distribution in the population.
- **Robustness**: reliability of the result
- **Power**: ability to reject  $H_0$  if it is false

	Parametric tests	Nonparametric tests
Assume underlying distribution?	Yes	No
Populations should be normal?	Yes	Not necessarily
Variances should be equal?	Yes	Not necessarily
Robust?	If the conditions are verified	In all situations
Powerful?	More	Less

***When valid, parametric tests should be preferred***



## Parametric vs nonparametric

- Many parametric tests have nonparametric equivalents

### Parametric test

Independent samples t-test

Paired samples t-test

1-way ANOVA

1-way Within-factor ANOVA

...

### Nonparametric equivalent

Mann-Whitney test

Wilcoxon test

Kruskal-Wallis test

Friedman test

...



## Why does sample size matter?

- Large sample: better representativeness
- Large sample: allows a better verification of the validity of parametric tests



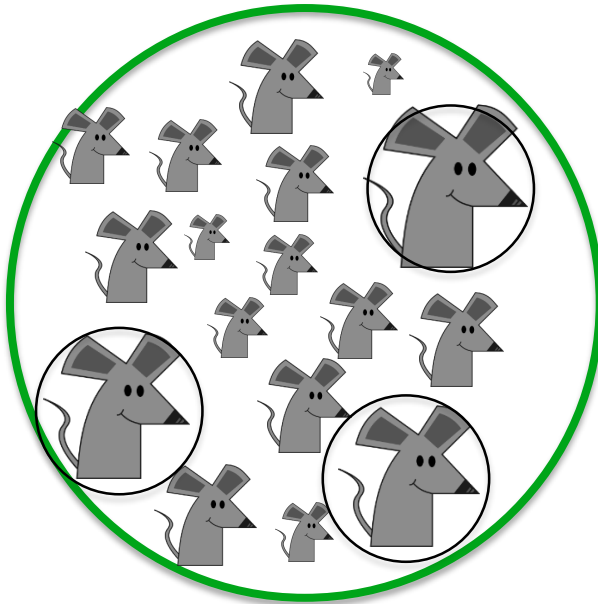
## Why does sample size matter? representativeness

- A test allows to **extrapolate a conclusion** from a sample to the population
- Samples should thus be **representative** of the population

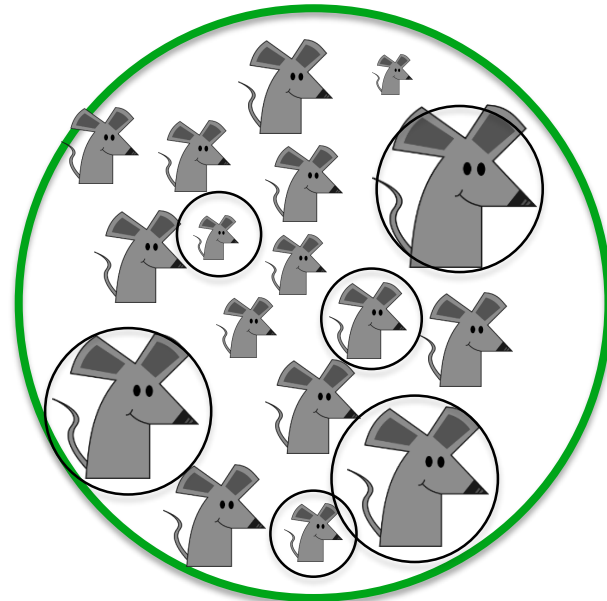


## Why does sample size matter? representativeness

- A test allows to **extrapolate a conclusion** from a sample to the population
- Samples should thus be **representative** of the population
- Larger random samples means better representativeness, in general



Bad sampling



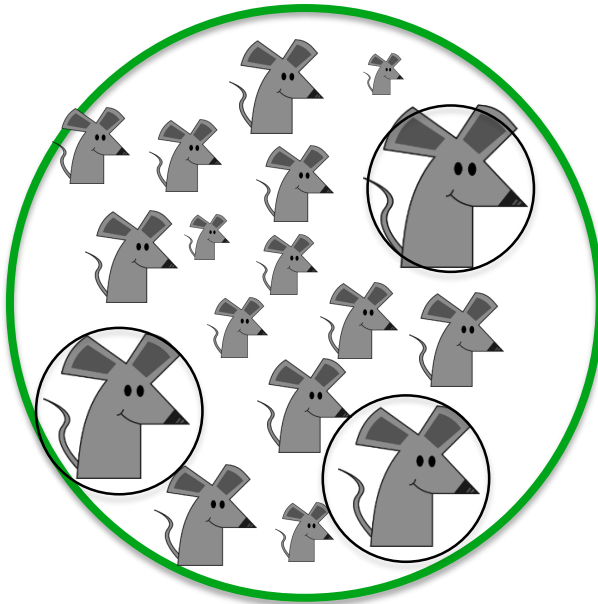
Better sampling



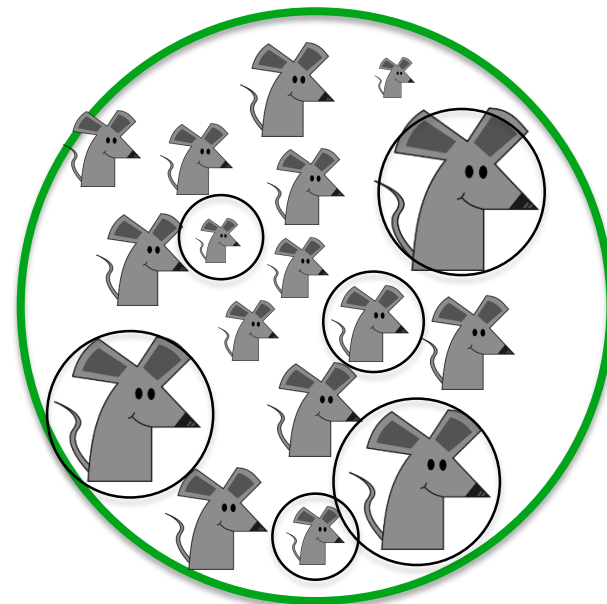


## Why does sample size matter? representativeness

- It is not recommended to run a test if samples are not representative of the population
- Hereafter, we consider that samples are representative of the population in all of the investigated cases



Bad sampling



Better sampling



## Why does sample size matter? Parametric tests: conditions

### Some diagnosis tools require minimum sample sizes

In the following slides, sample size recommendations we propose are rules of thumb. There are no precise rules in literature. We strongly advise to refer to your fields' specific recommendations.

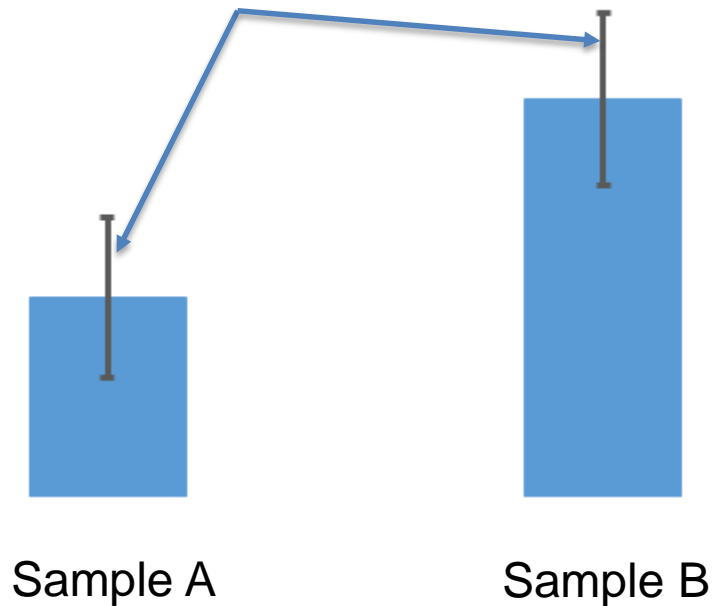


## Why does sample size matter? Parametric tests: conditions

Some diagnosis tools require minimum sample sizes

*t-test on 2 independent samples*

- **Condition 1: equal variances**
  - > **variance tests** (at least 3 observations/sample)



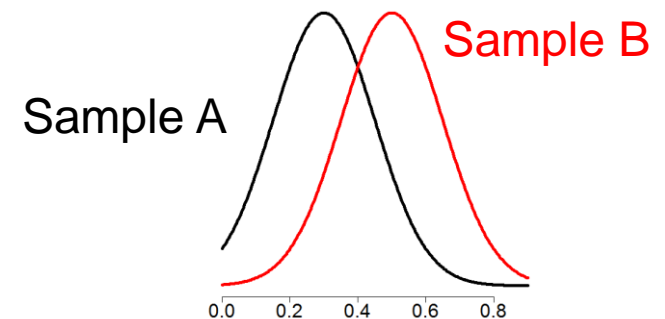


## Why does sample size matter? Parametric tests: conditions

Some diagnosis tools require minimum sample sizes

*t-test on 2 independent samples*

- **Condition 1: Homogeneous variances**
  - > **variance tests** (at least 3 observations/sample)
- **Condition 2: Normality of each population**
  - > If sample size < 20, normality of each population **should be assumed** based on prior knowledge (not recommended). Otherwise, the test cannot be used.
  - > If sample size > 20, normality **should be verified** in each sample → **graphics** or **normality tests** (at least 20 observations/sample)





## Why does sample size matter? Parametric tests: conditions

Some diagnosis tools require minimum sample sizes

*t-test on 2 paired samples*

- **Condition 1: Normality of the differences between pairs**
  - > If number of pairs < 20, normality **should be assumed** based on prior knowledge. Otherwise, the test cannot be used.
  - > If number of pairs > 20, normality should be verified → **normality tests** (at least 20 observations/sample)



## Why does sample size matter? Parametric tests: conditions

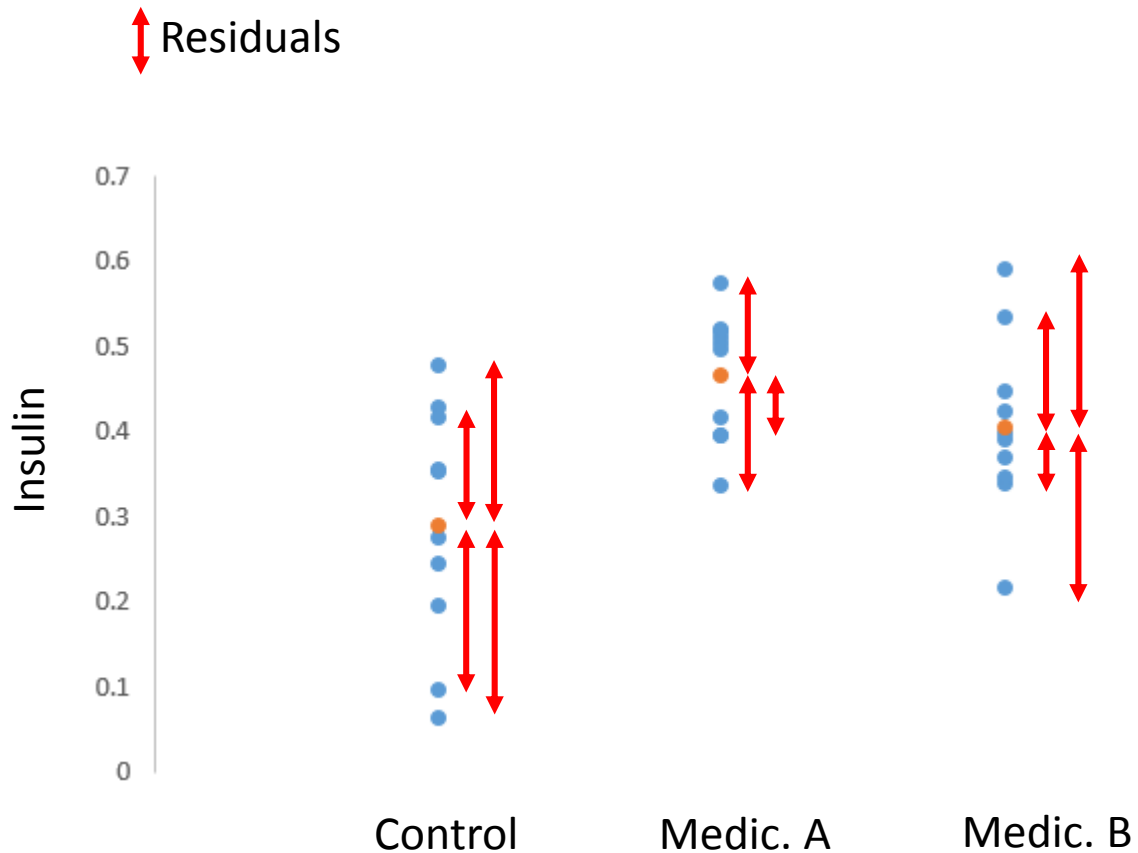
Some diagnosis tools require minimum sample sizes

*ANOVA on independent samples*

- **Condition 1: Homogeneous variances**
  - > **variance tests** (at least 3 observations/sample)
- **Condition 2: Normality of residuals**
  - > If *the whole experiment* includes < 20 observations, normality of residuals **should be assumed** based on prior knowledge (not recommended). Otherwise, the test cannot be used.
  - > If the whole experiment includes > 20 observations, normality of residuals should be verified → **normality tests**



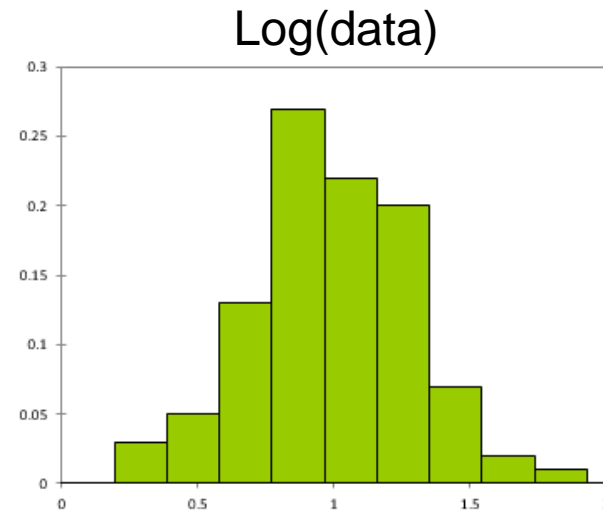
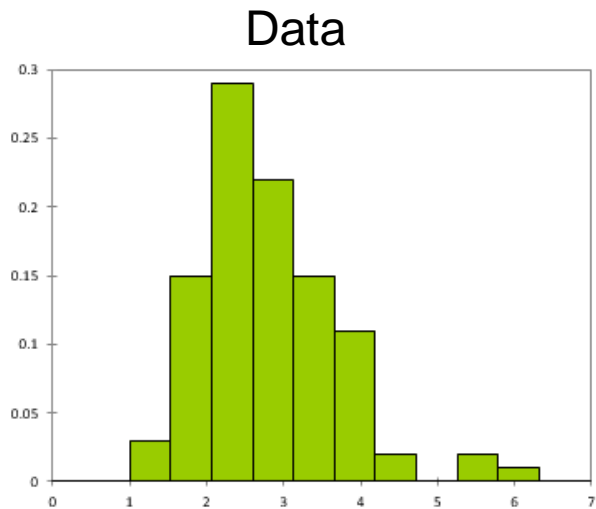
# What are residuals in ANOVA?





## What to do if conditions are not met?

- Transform your data (e.g. log)
- Remove outliers (if justified)
- Variance heterogeneity: use the Welch correction
- Run nonparametric tests







Demo time

# 1-way ANOVA *vs* Kruskal-Wallis test: XLSTAT demo



## Take home message

- Parametric tests are more powerful. We recommend to prefer them when they are valid
- A sample should be representative of the population (more likely if larger)
- Parametric conditions could be reached by transforming the data / removing outliers if justified
- Rules of thumb. If  $S$  is Sample size (t-test) or number of observations (ANOVA):
  - > If variances are heterogeneous → **nonparametric** (or **parametric** with Welch correction)
  - > If  $S < 20$  AND normality is assumed → **parametric**
  - > If  $S < 20$  AND normality is not assumed → **nonparametric**
  - > If  $S > 20$  AND normality is validated through verification → **parametric**
  - > If  $S > 20$  AND if normality is not validated through verification → **nonparametric**
- You should also refer to your field's specific recommendations



**Thank you for your attention!**

