Descriptive [Bio-]statistics

Robert Haase

With material from

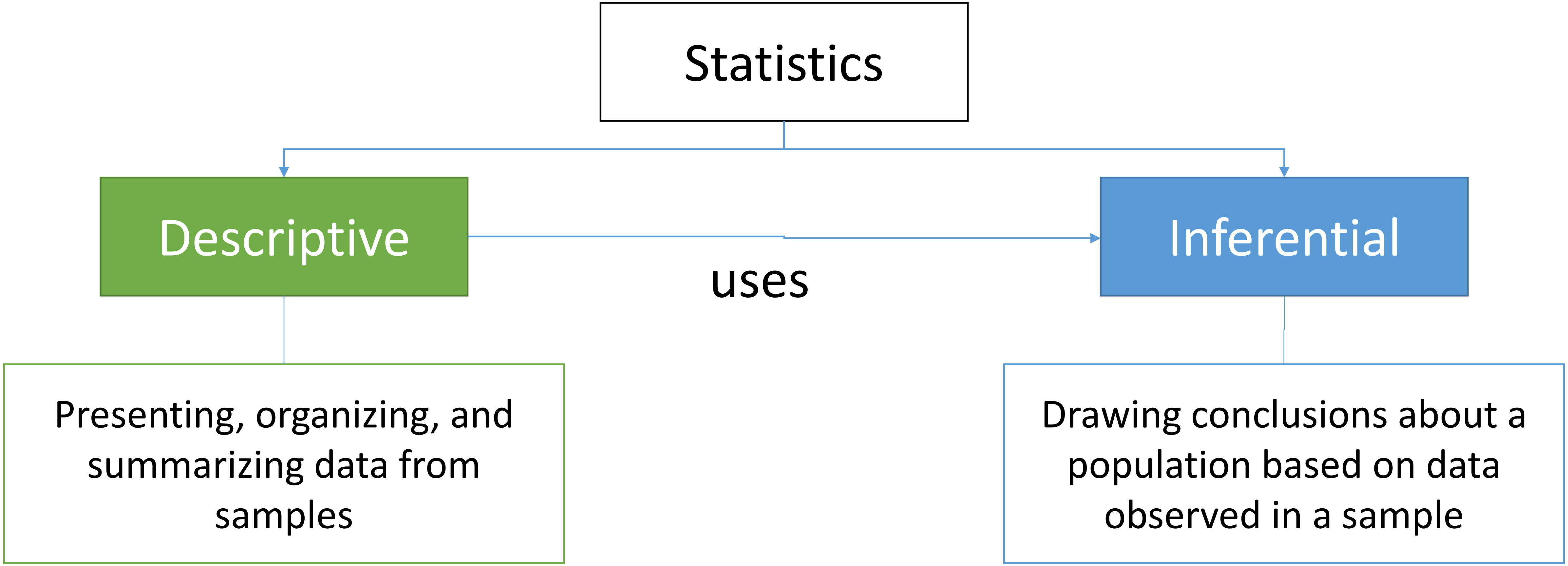
Aldo Acevedo Toledo, Biotec, TU Dresden

Martin J. Bland and Douglas G. Altman

May 2022

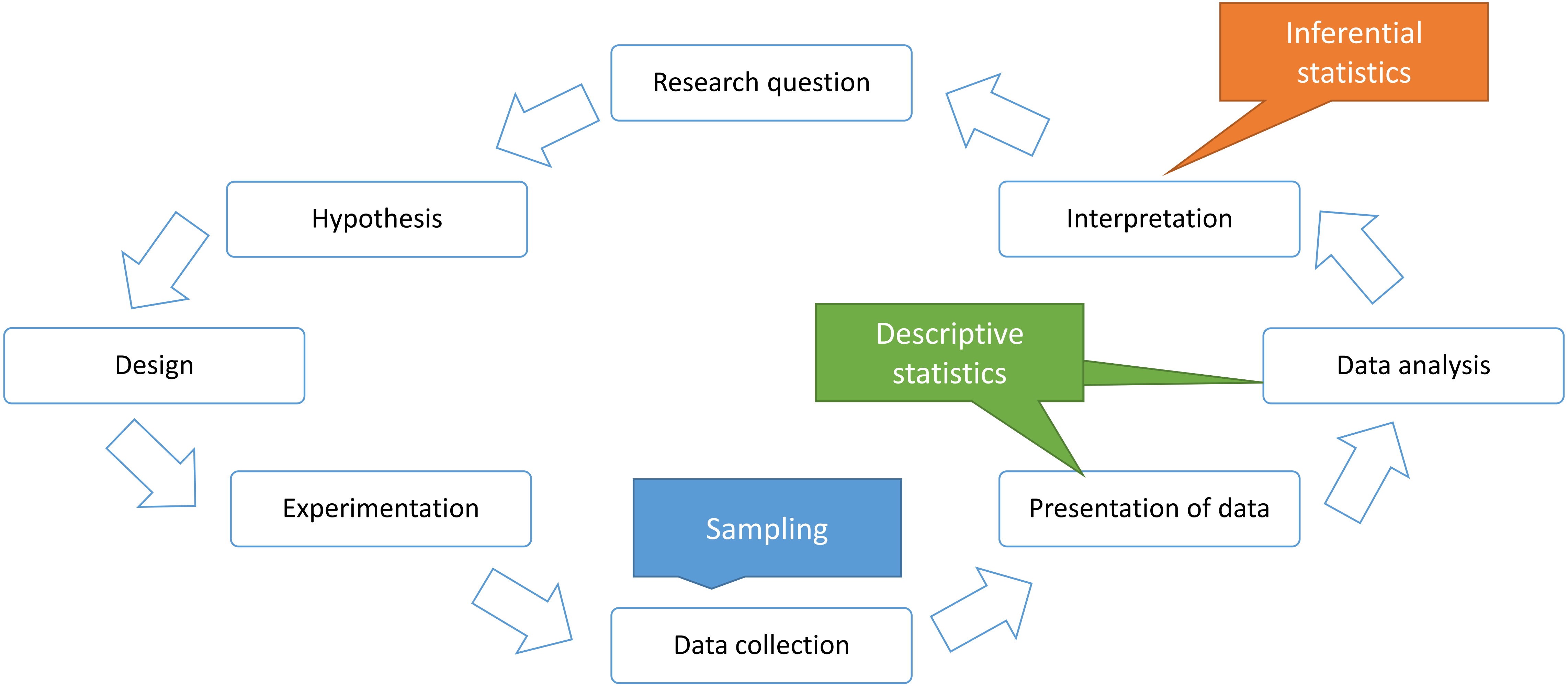
@haesleinhuepf

Descriptive statistics



Slide adapted from
Aldo Acevedo Toledo, Biotec, TU Dresden

Research workflow



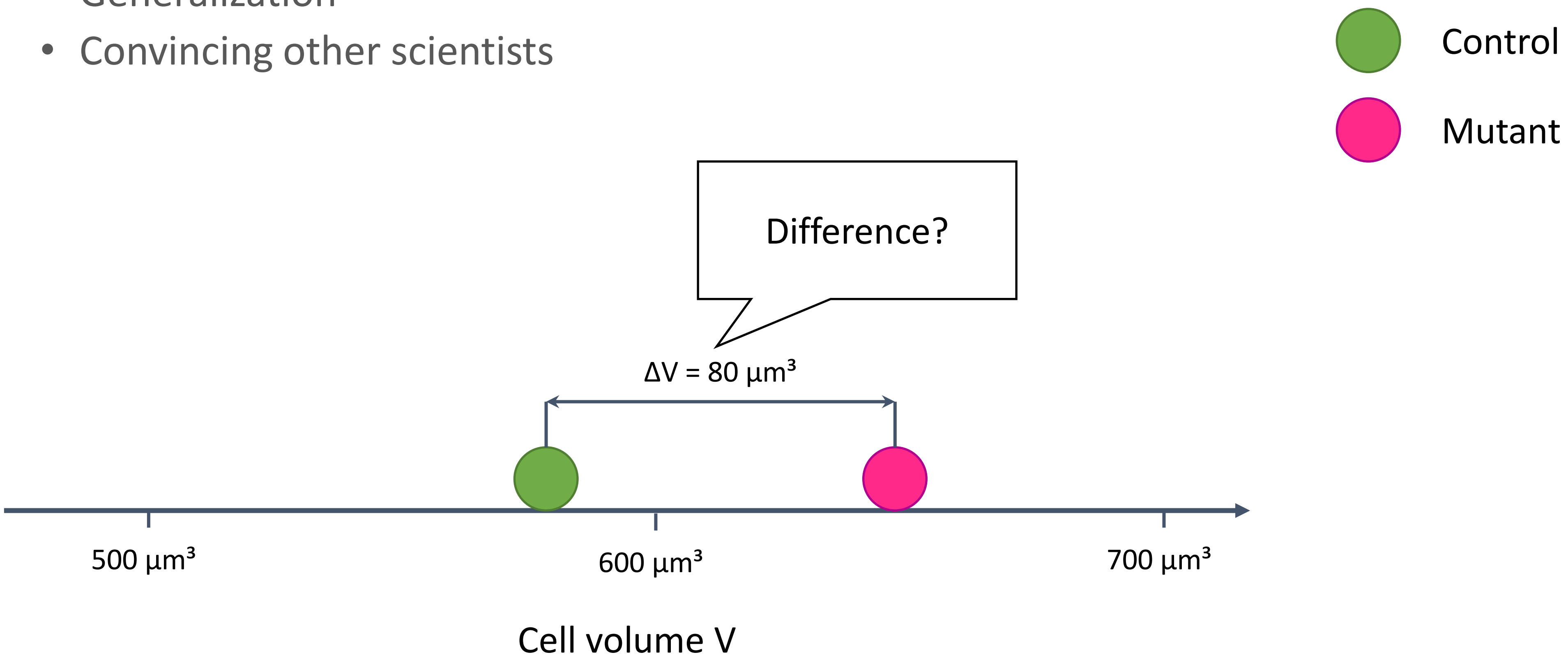
Sampling

Convenience Sampling	• Select the most accessible and available subjects in target population • Inexpensive, less time consuming, but sample is nearly always non-representative of target population
Random Sampling	• Select subjects at random from the target population • Need to identify all in target population first • Provides representative sample frequently
Systematic Sampling	• Identify all in target population • Select every n^{th} item as a subject
Stratified Sampling	• Identify important sub-groups in your target population. Sample from these groups randomly or by convenience • Ensures that important sub-groups are included in sample • May not be representative

Slide adapted from
Aldo Acevedo Toledo, Biotec, TU Dresden

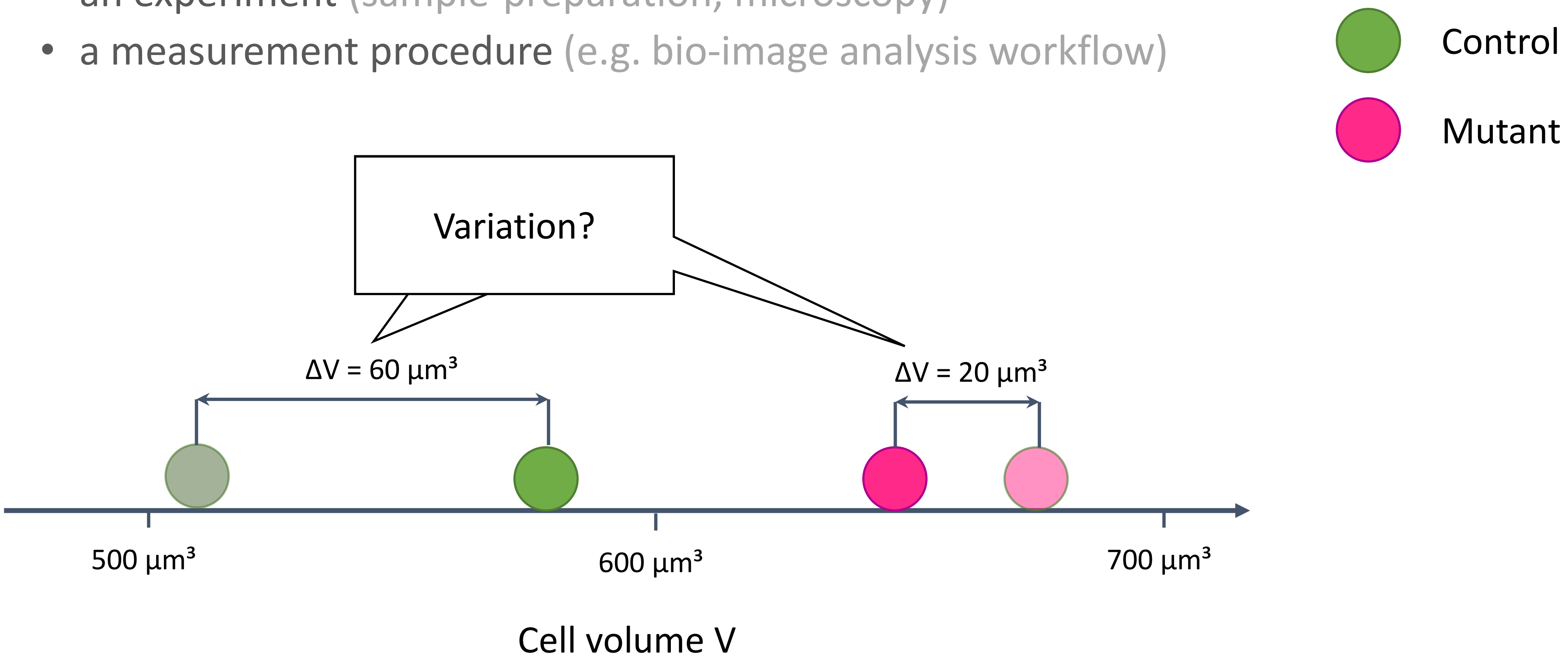
Sampling

- Raw individual measurements are limited regarding
  - Interpretability
  - Generalization
  - Convincing other scientists



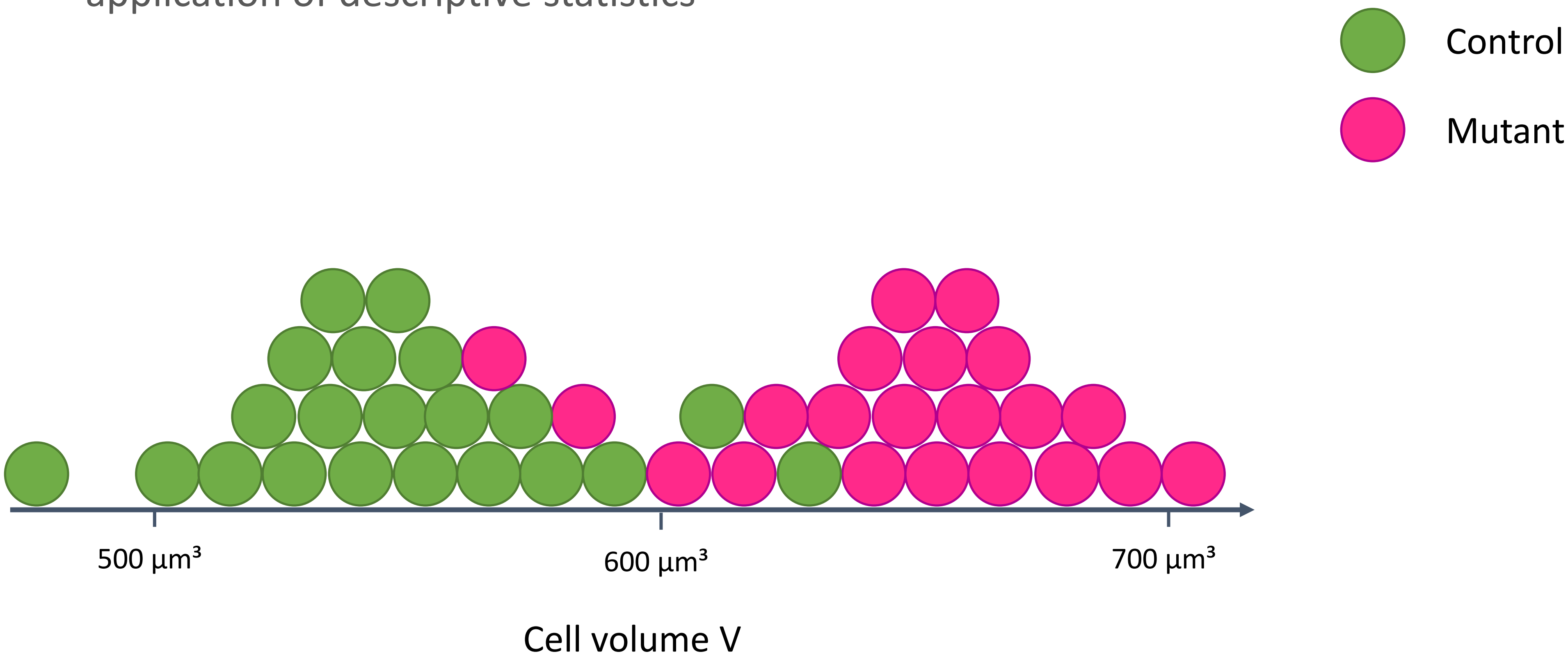
Sampling

- Raw individual measurements are result of
  - a stochastic process (cell life-cycle, biology)
  - an experiment (sample-preparation, microscopy)
  - a measurement procedure (e.g. bio-image analysis workflow)



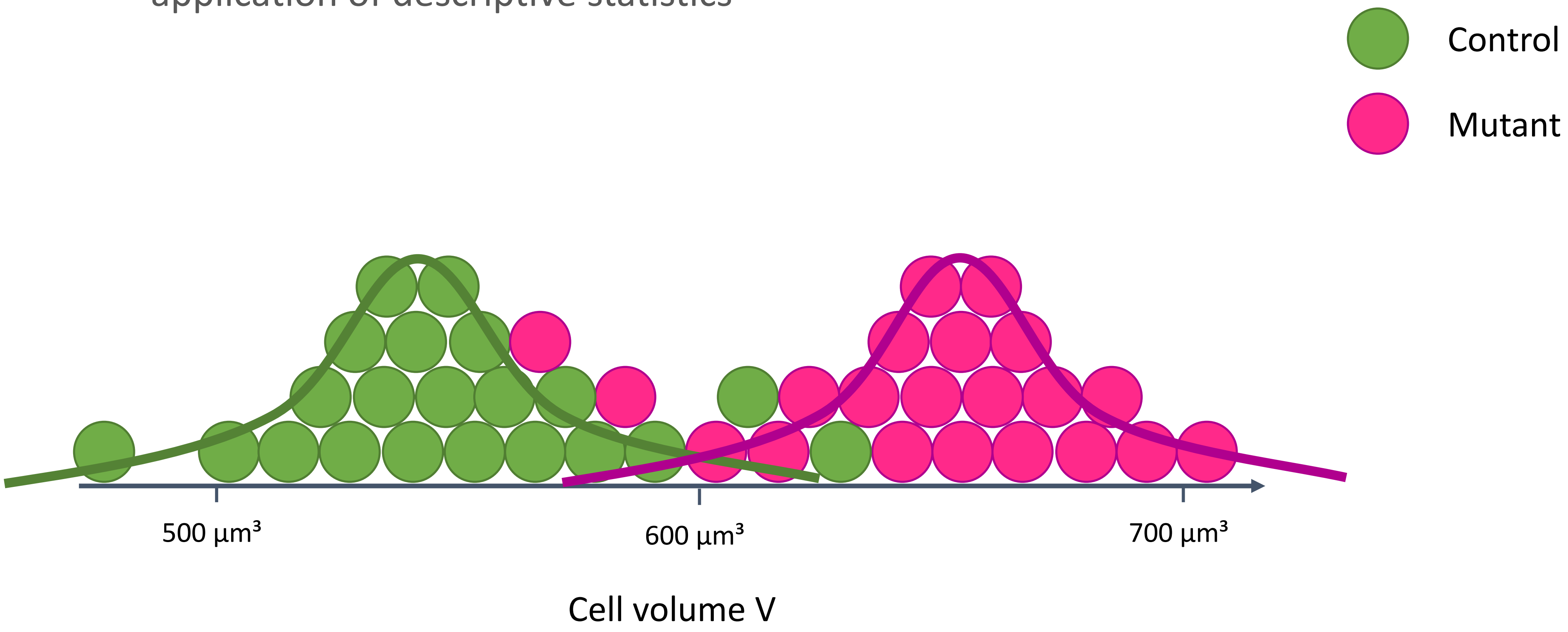
Sampling

- Repetitive experiments allow
  - a closer view
  - application of descriptive statistics



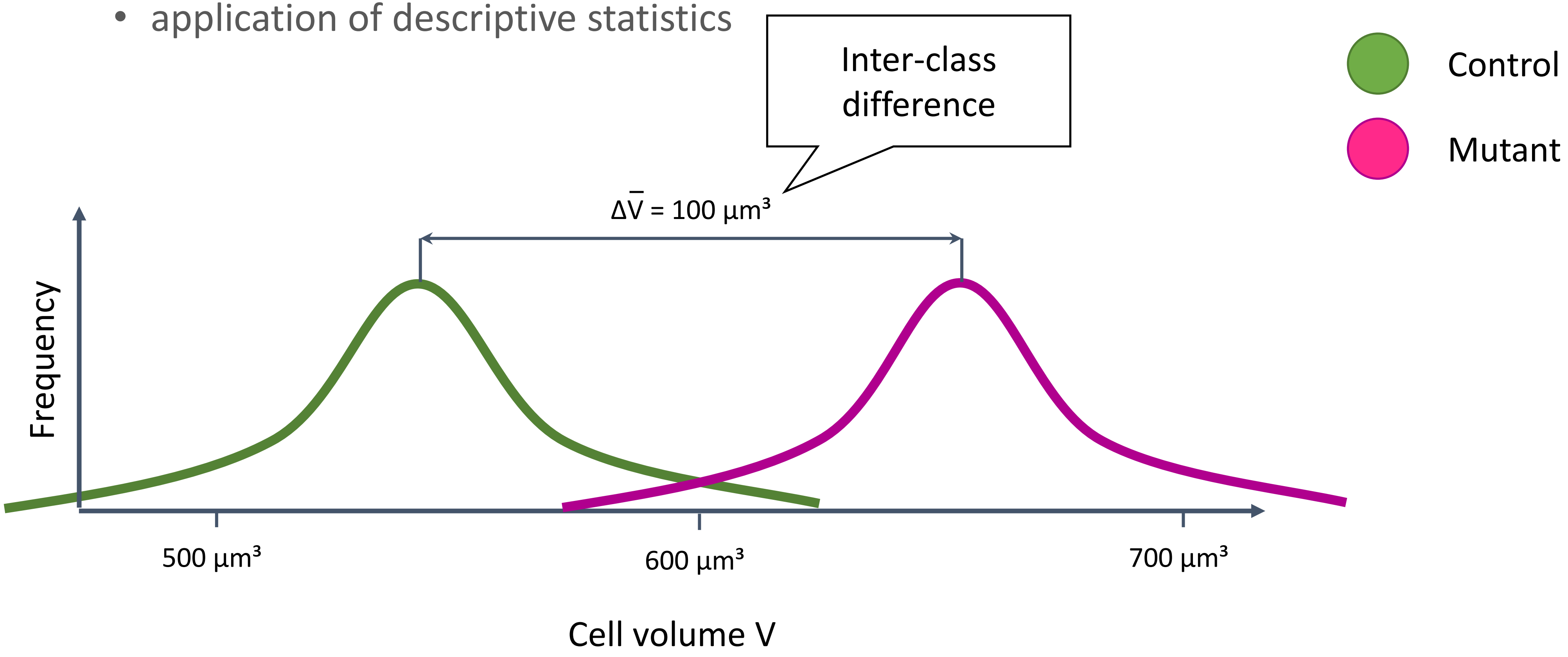
Sampling

- Repetitive experiments allow
  - a closer view
  - application of descriptive statistics



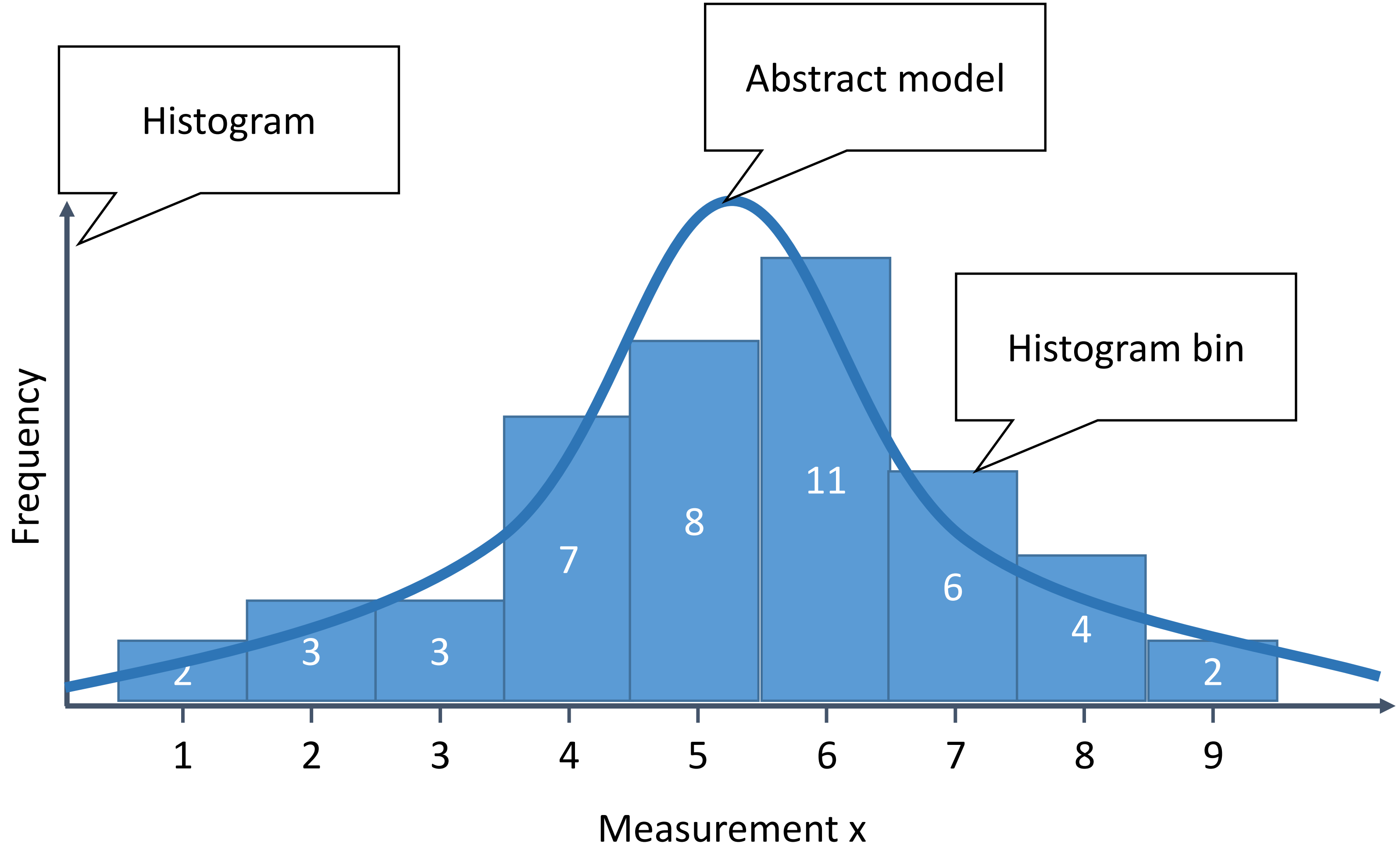
Descriptive statistics

- Repetitive experiments allow
  - a closer view
  - application of descriptive statistics



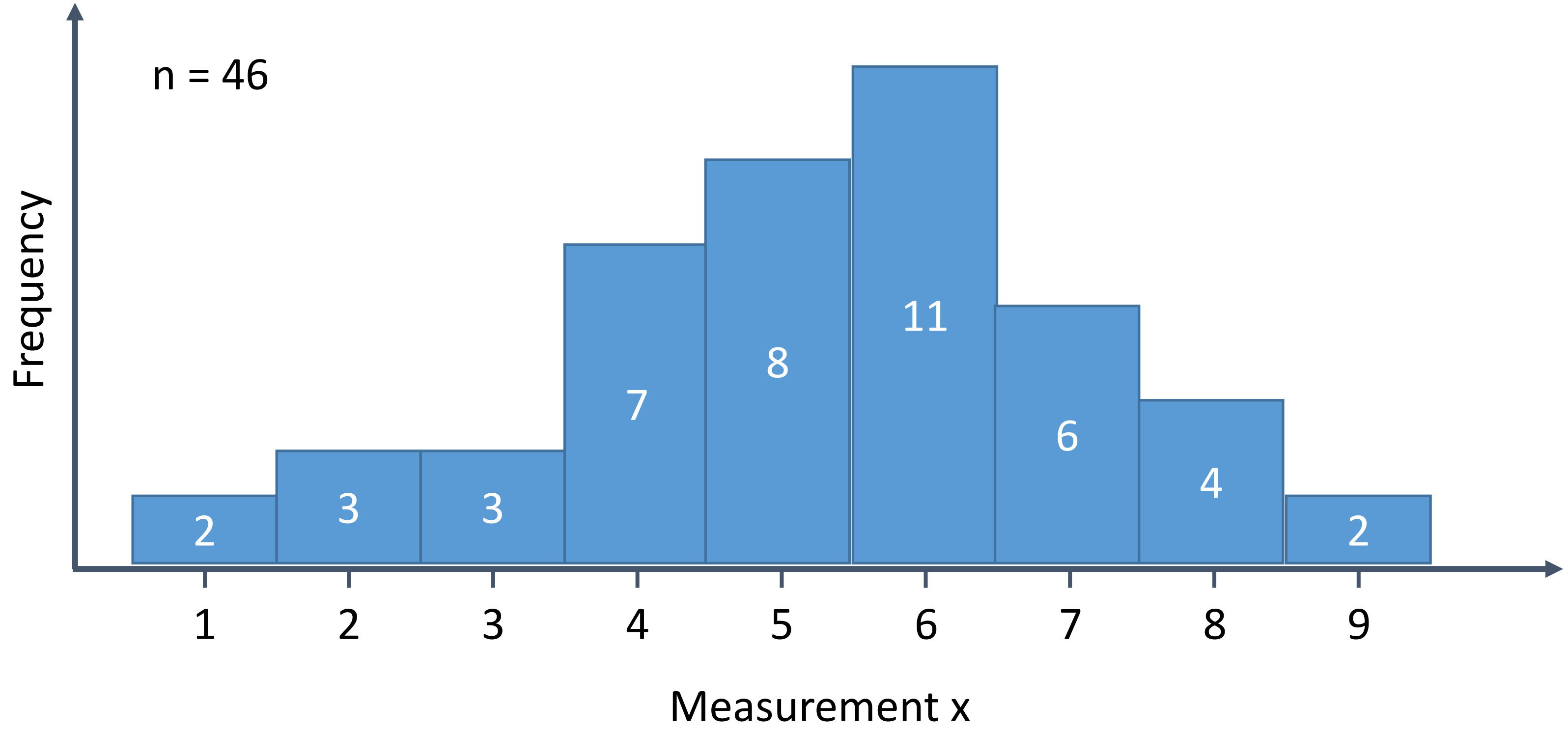
Descriptive statistics

- Descriptive statistics enables summarizing data
  - Abstract models
  - Histograms



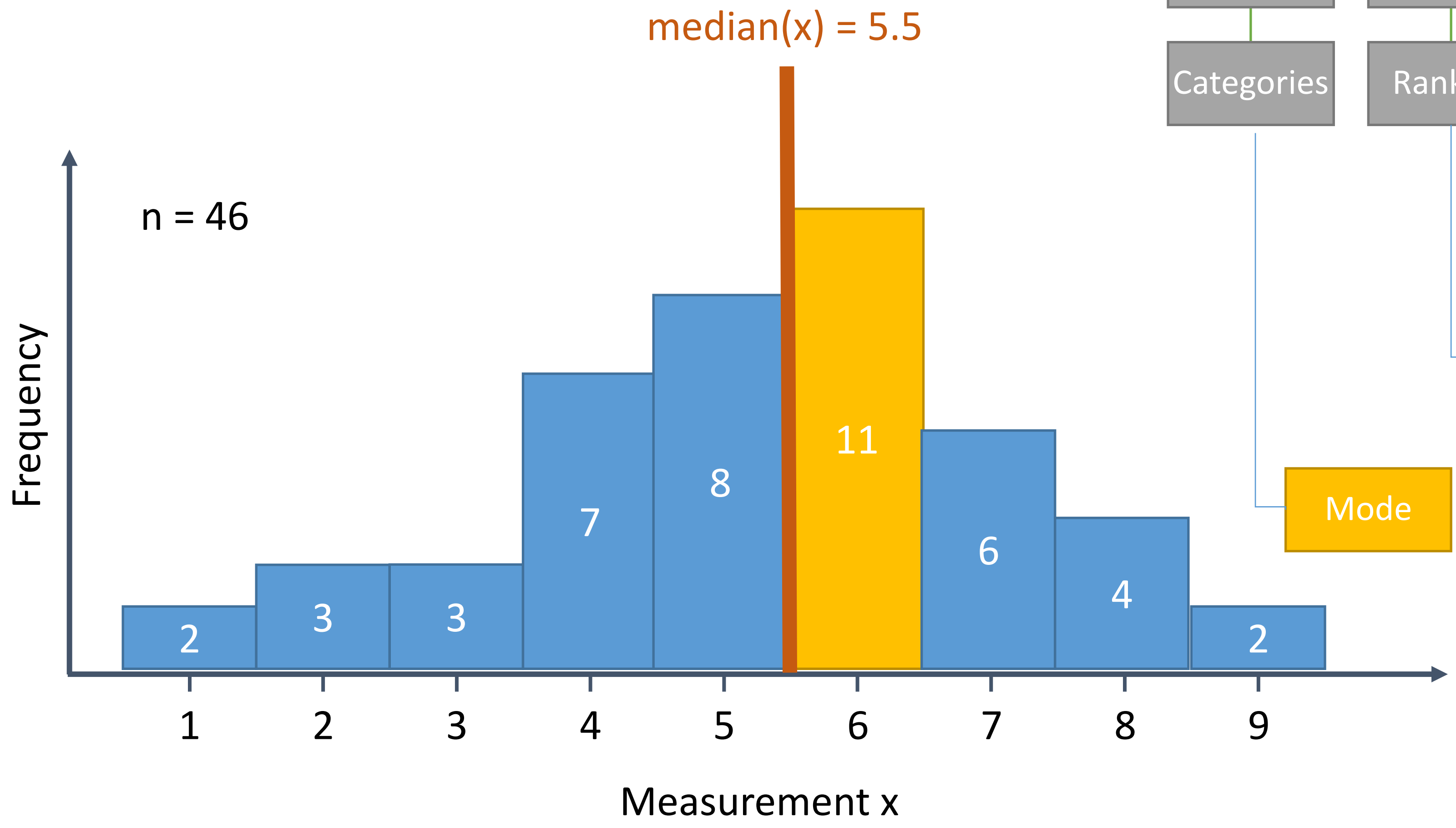
Measures of central tendency

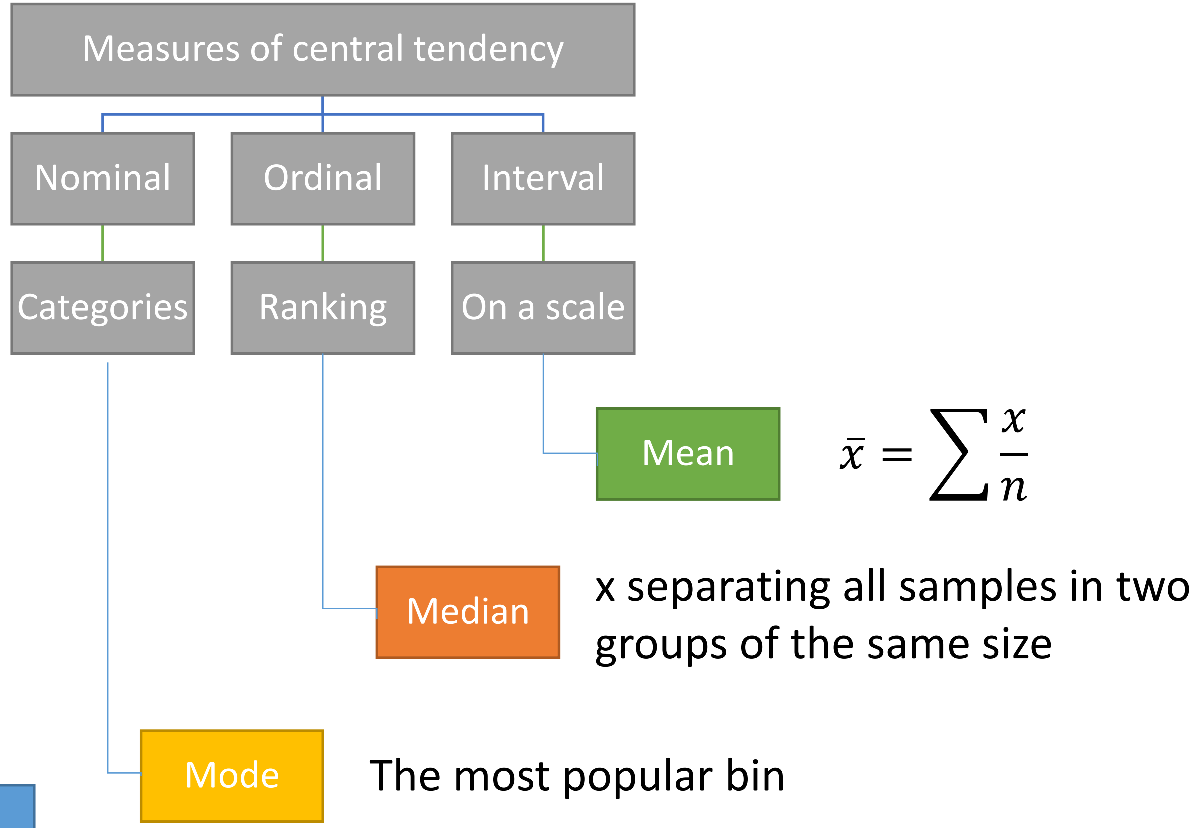
- "Where" in parameter space are my samples located?



Measures of central tendency

- "Where" in parameter space are my samples located?

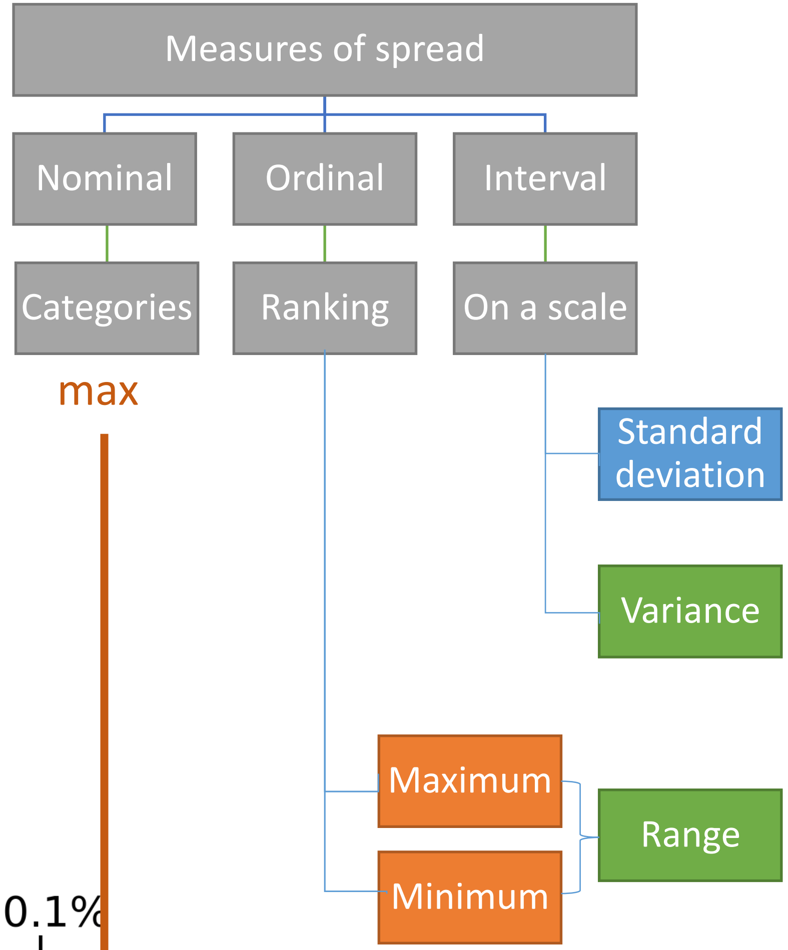




Slide adapted from
Aldo Acevedo Toledo, Biotec, TU Dresden

Measures of spread

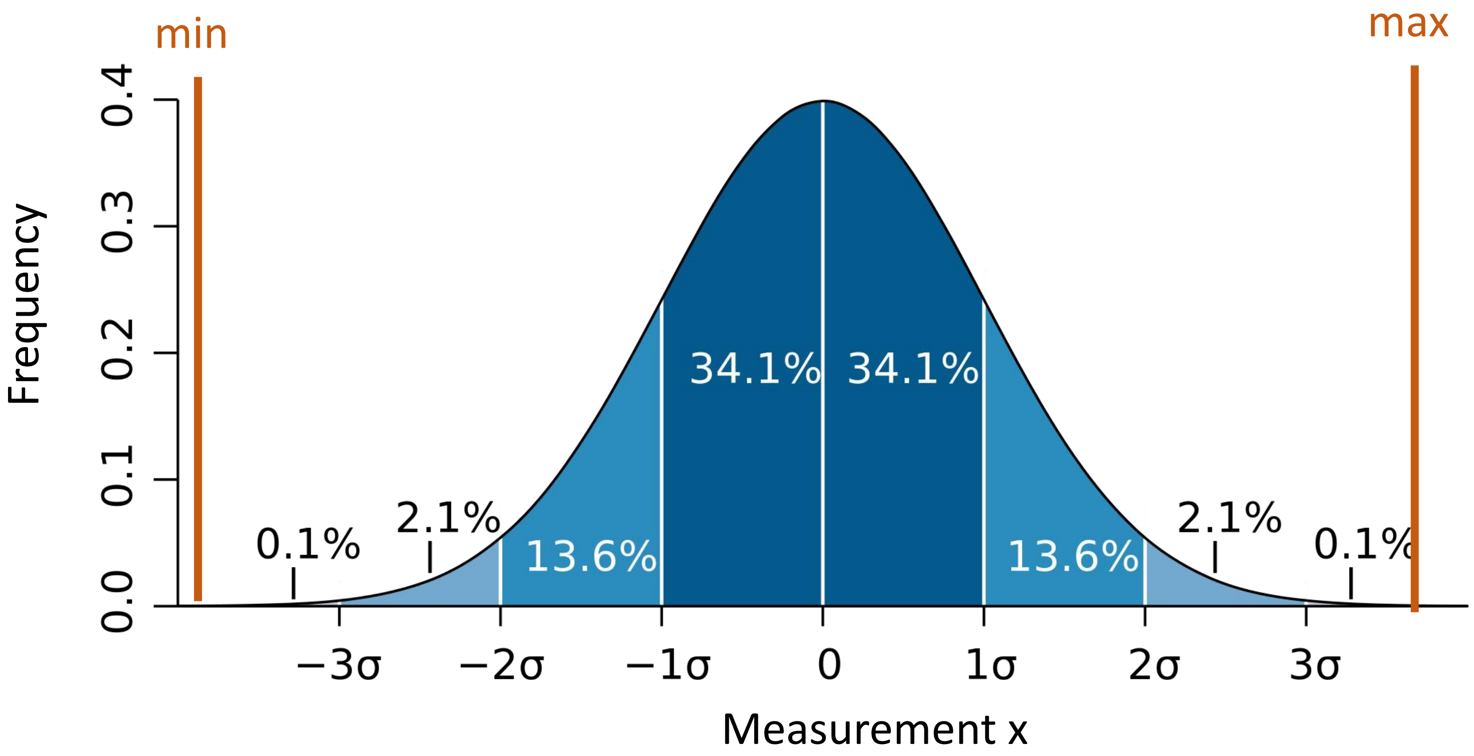
- How do my samples vary in parameters space?



$$\bar{x} = \sum \frac{x}{n}$$

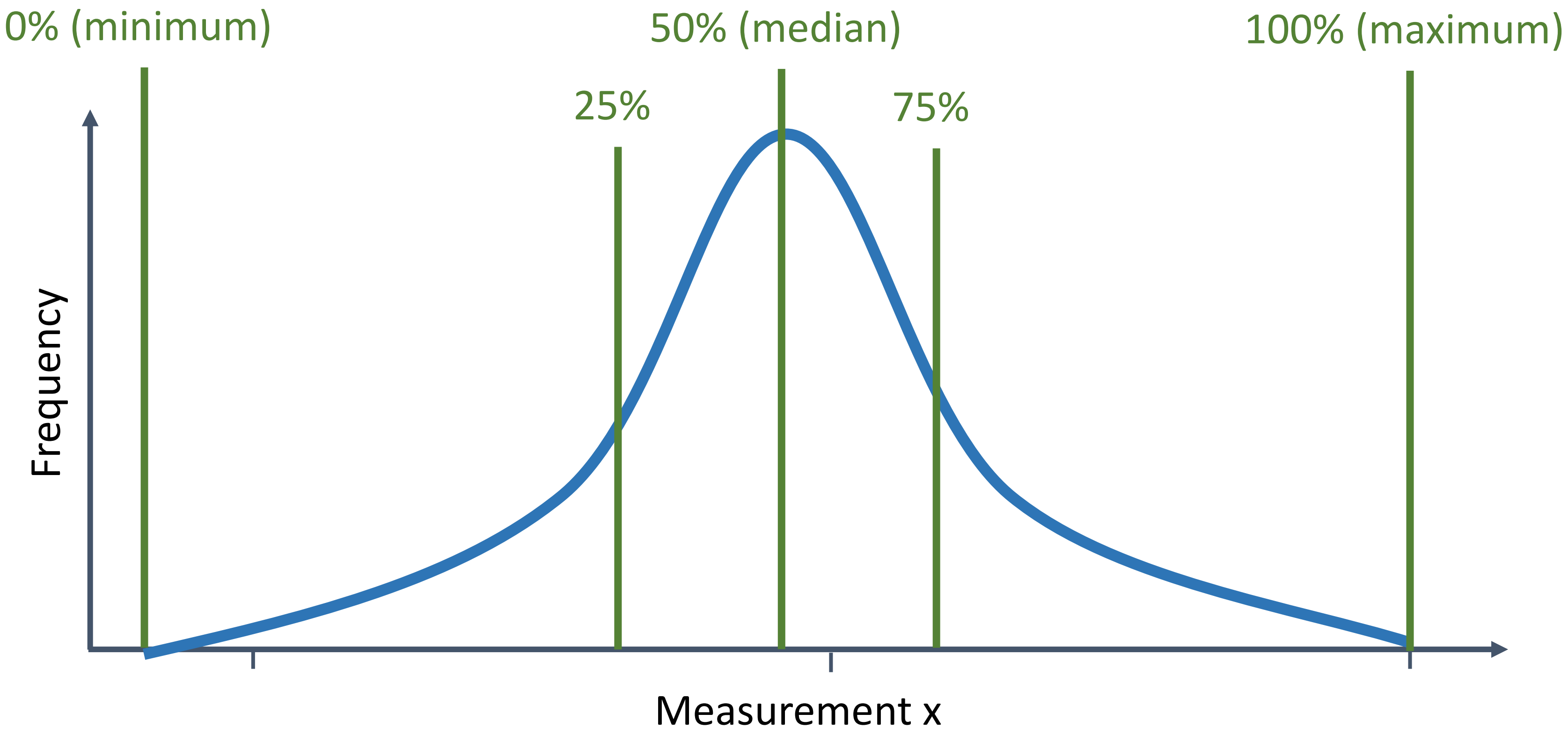
$$\sigma = \sqrt{\sum \frac{(x - \bar{x})^2}{n}}$$

$$Var(x) = \sigma^2$$



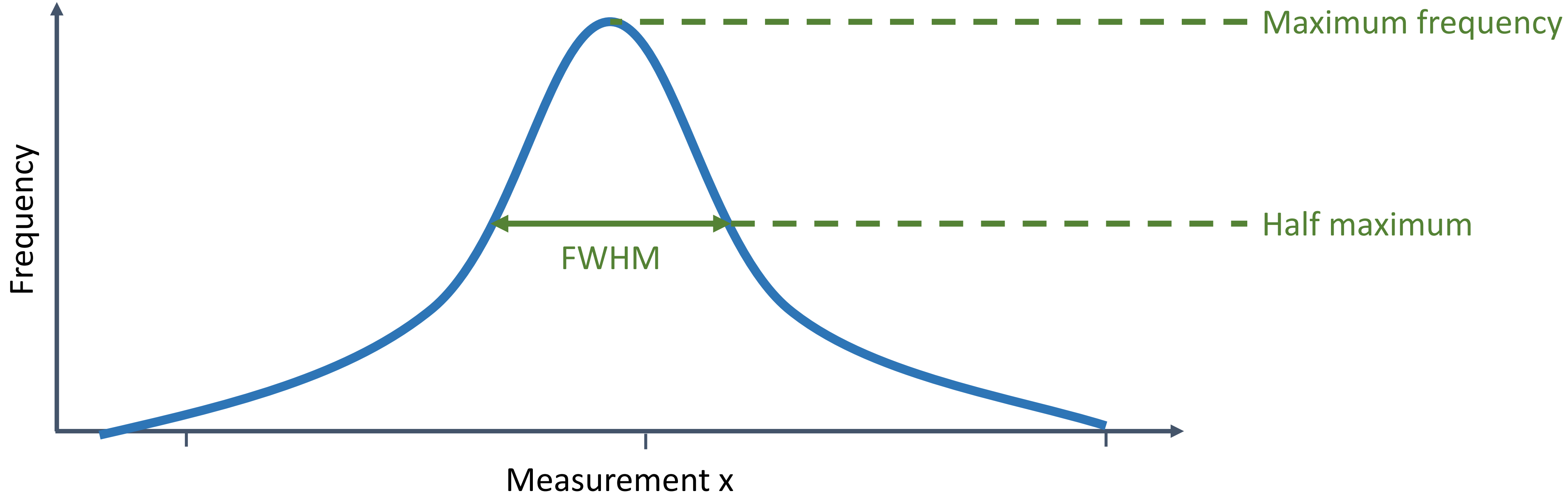
Measures of spread

- Percentiles
  - The value under which a given percentage of our samples lie



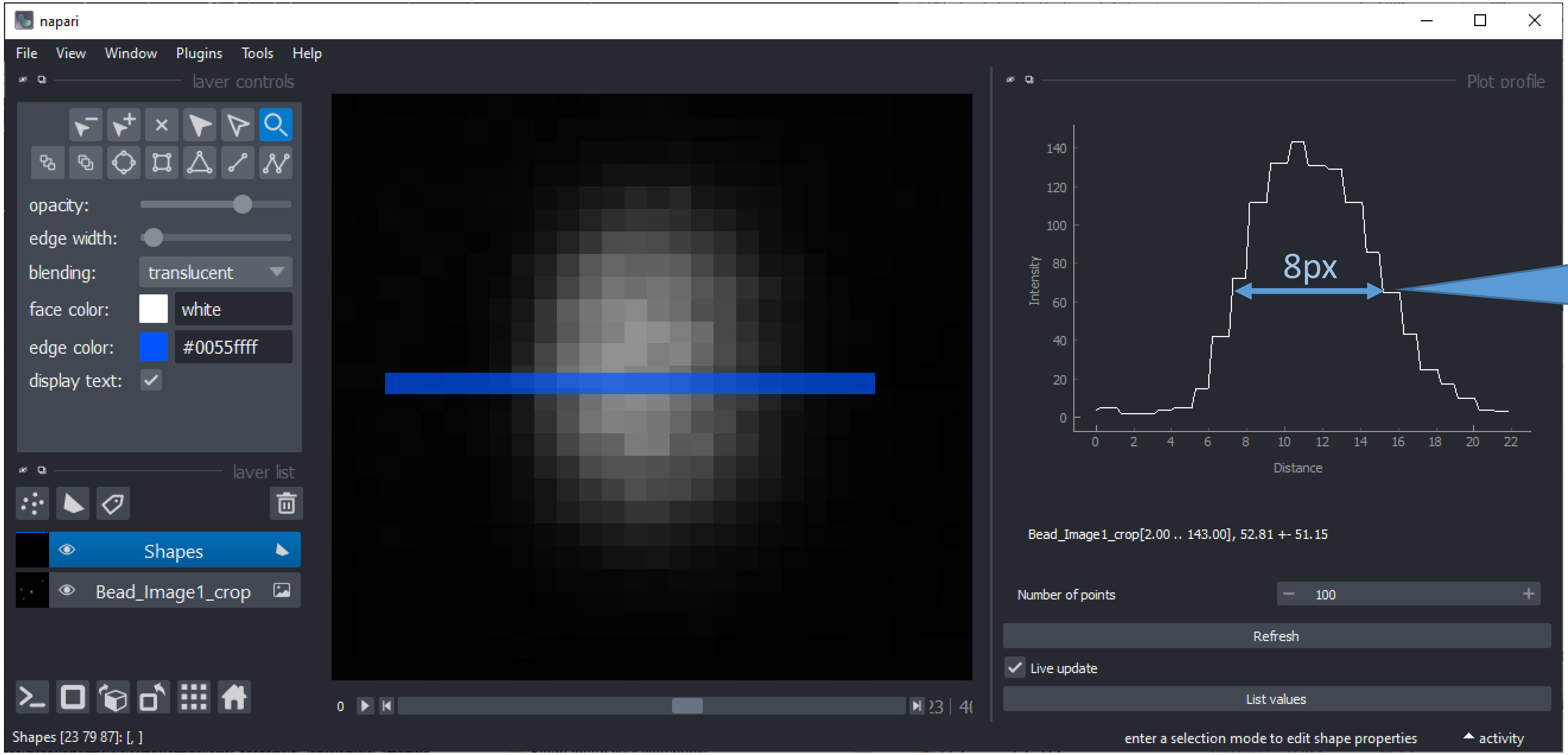
Measures of spread

- Full width at half maximum (FWHM)



Measures of spread

- Full width at half maximum (FWHM)
  - Imaging sub-resolution fluorescent beads
  - A profile plot through the image of a bead allows us to estimate the spatial resolution of a microscope



8 px
times 0.022 µm

-> 1.8 µm FWHM

Bland-Altman analysis

Robert Haase

Based on material by Martin J. Bland
and Douglas G. Altman

May 2022

@haesleinhuepf

Method comparison studies

- Scenario
  - You work in a lab and try to improve procedures
    - Chemical protocols
    - Sample preparation
    - Analysis protocols
      - Physical measurements
      - Image analysis

Unpaired data

- Analyze independent sample sets
- Conclude about their similarity or relationship

Inferential statistics

Paired data

- The same dataset analyzed twice with different methods
- The same dataset analyzed twice with the same method

Direct method comparison –descriptive statistics

Method comparison studies

Martin Bland and Douglas Altman work on Method Comparison (excerpt)

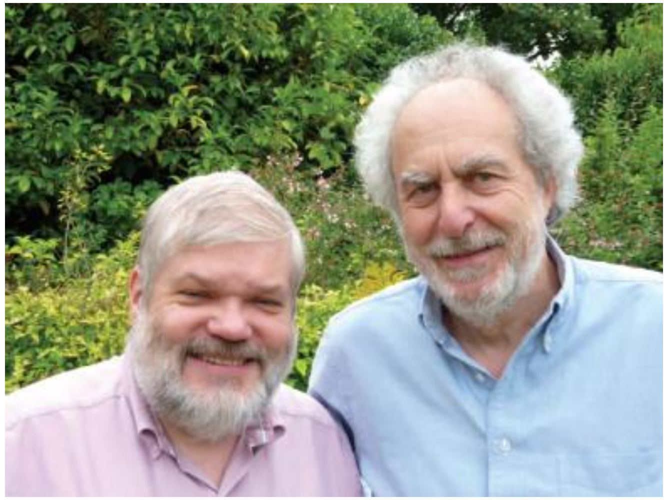
- https://www-users.york.ac.uk/~mb55/meas/ab83.pdf (Open Access)
- https://www.sciencedirect.com/science/article/pii/S0140673686908378

The Statistician 32
© 1983 Institute of

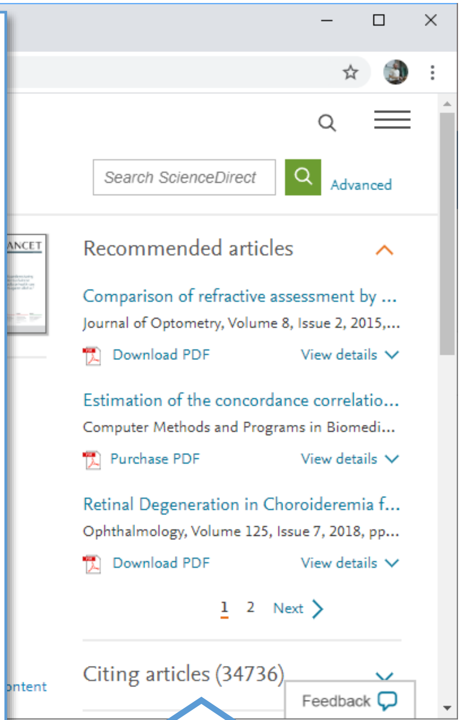
Measuren
Comparis

D. G. ALTMA

Division of Co
Research Cent
‡ Department of Clinical Epidemiology and Social Medicine,
St George's Hospital Medical School, Cranmer Terrace, London SW17



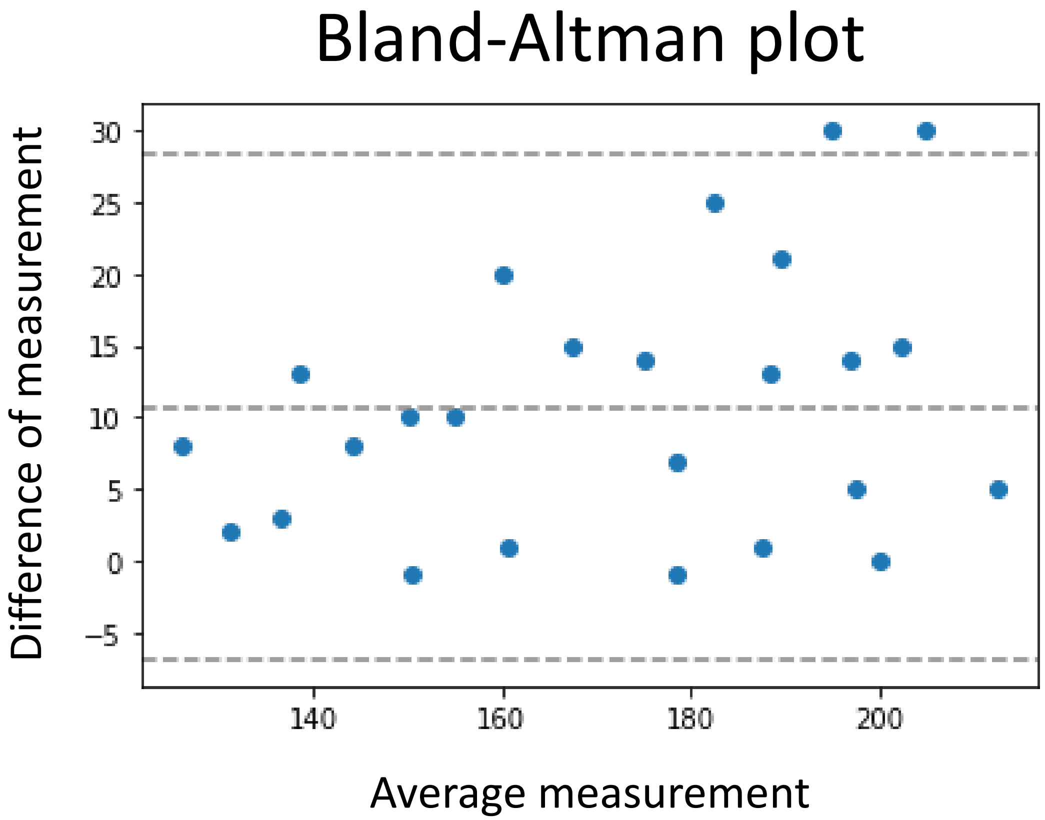
Copyright J. Martin Bland and Douglas G. Altman.



Citing articles (34736)

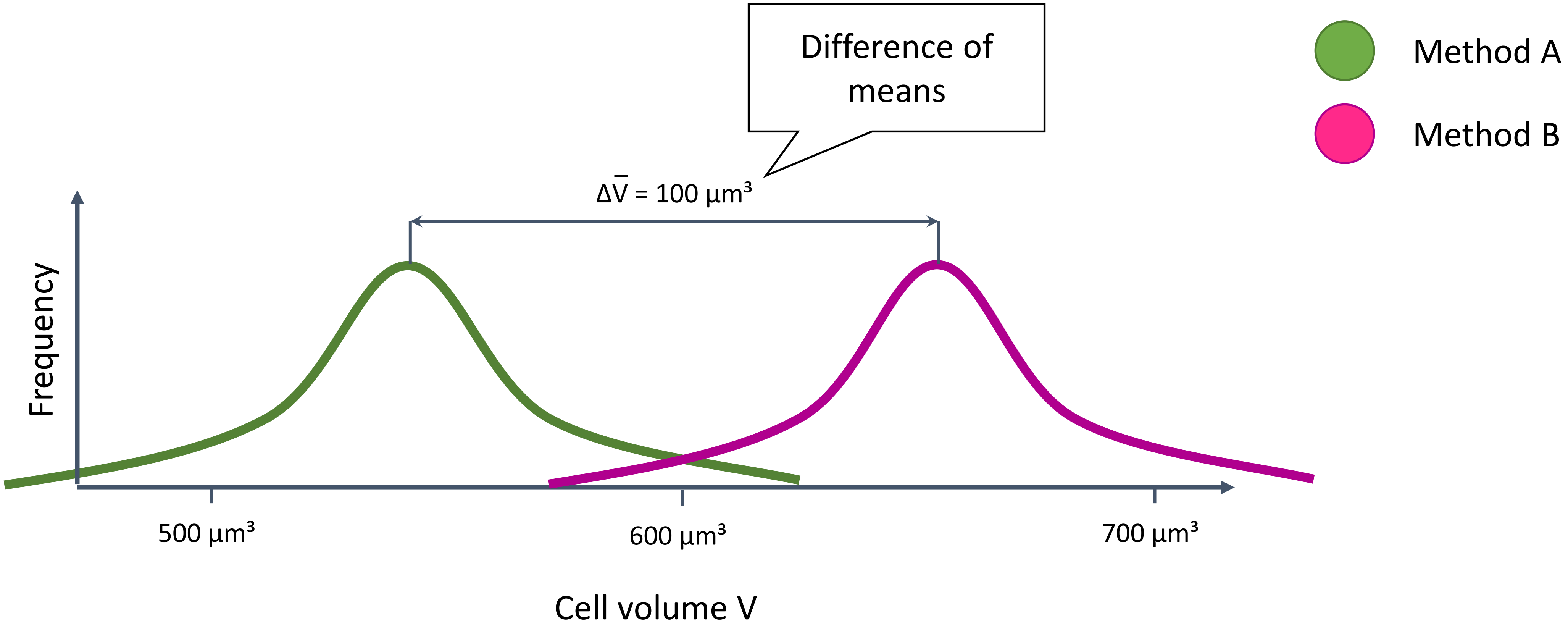
Bland-Altman plots

- Visualized method comparison



Comparison of means

- Comparing mean measurements appears reasonable on the first view.



Comparison of means

- Are two methods doing the same if their mean measurement is similar?

A	B
1	4
9	5
7	5
1	7
2	4
8	5
9	4
2	6
1	6
7	5
8	4

```
Mean(A) = 5.0
Mean(B) = 5.0
```

What if means were “very” different?

Yes

Method B cannot replace method A

No

Similar means is a necessary condition, but is it sufficient?

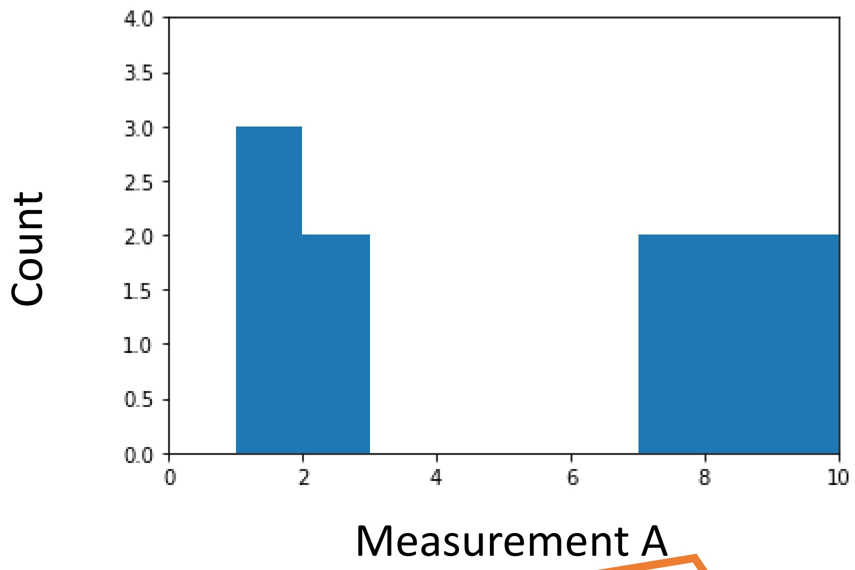
Comparison of means

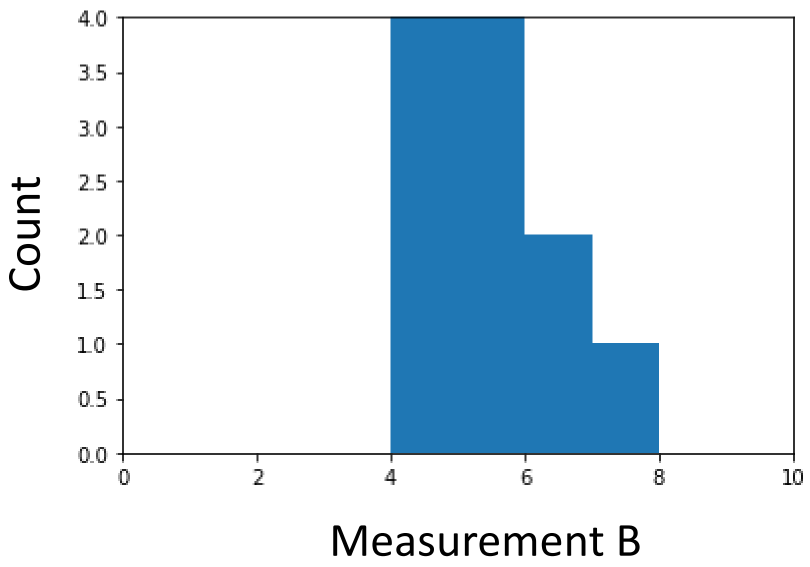
- Are two methods doing the same if their mean measurement is similar?

A	B
1	4
9	5
7	5
1	7
2	4
8	5
9	4
2	6
1	6
7	5
8	4

```
Mean(A) = 5.0
Mean(B) = 5.0
```

- Draw histograms! How can two methods do the same if histograms from their measurements are different?





Similar means is a necessary condition, but it is NOT sufficient!

The scientific method: Show that a method doesn’t work with *just one* example. And you have *proven* that the method doesn’t work in general.

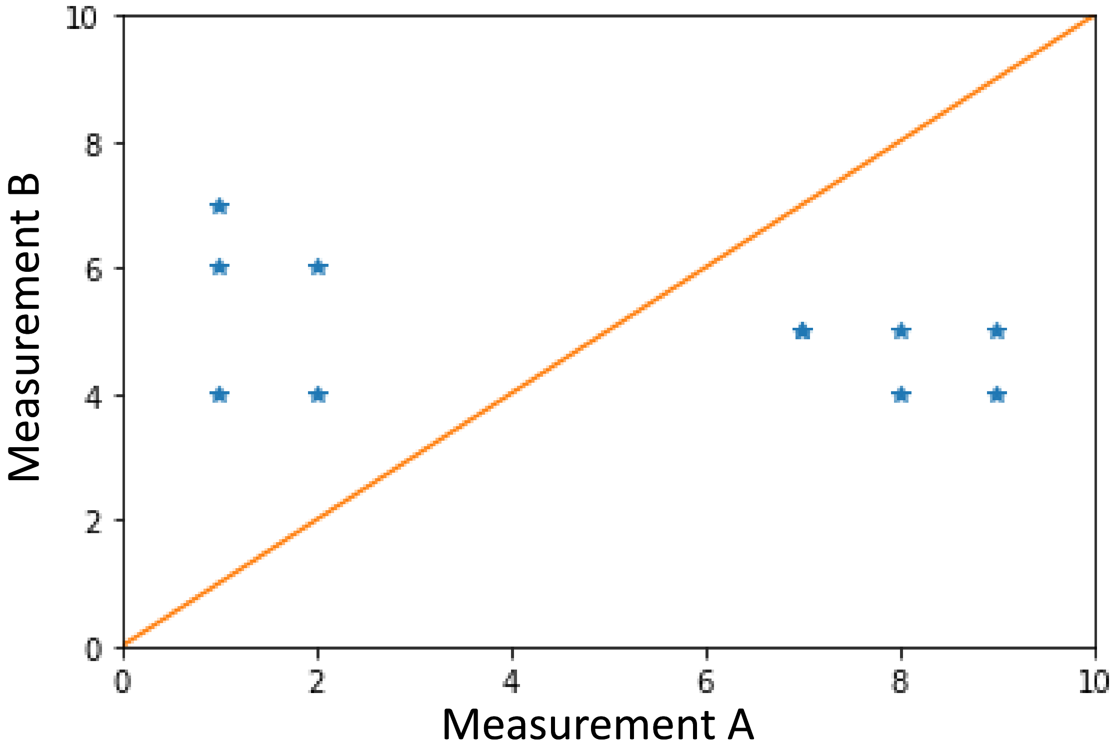
Comparison of means

- Are two methods doing the same if their mean measurement is similar?

A	B
1	4
9	5
7	5
1	7
2	4
8	5
9	4
2	6
1	6
7	5
8	4

```
Mean(A) = 5.0
Mean(B) = 5.0
```

- Plot the measurements against each other. What does it mean if they lie on a straight line? What if not?



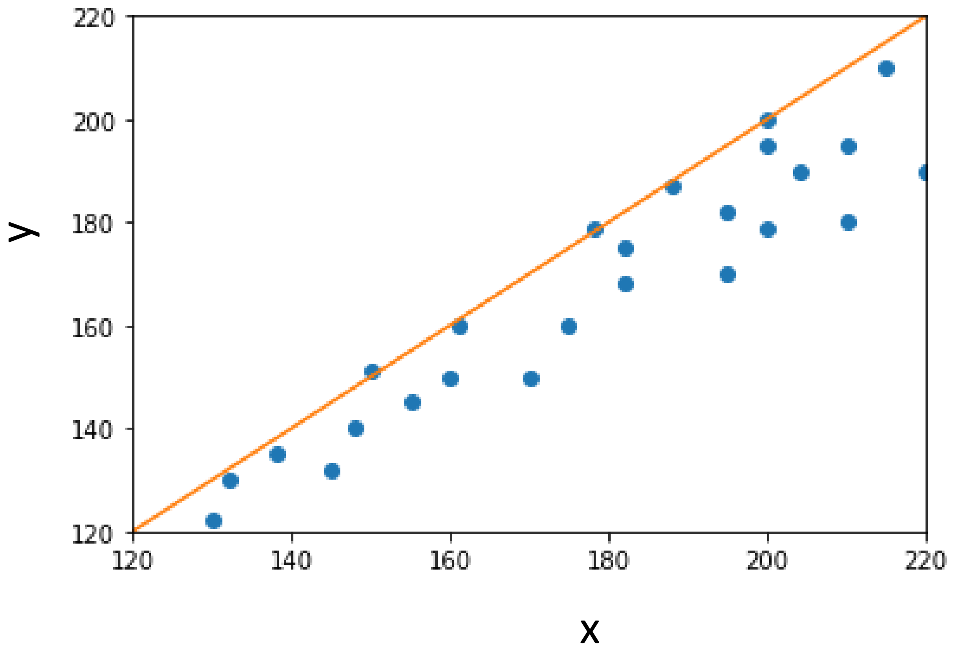
The “criterion of agreement was that the two methods gave the same mean measurement; ‘the same’ appears to stand for ‘not significantly different’. Clearly, this approach tells us very little about the accuracy of the methods.”[1]

[1] Altman & Bland, The Statistician 32, 1983

Correlation

- Are two methods doing the same if they correlate?
  - Correlation: Any kind of relationship.
  - Measurable; e.g. using Pearson's Correlation Coefficient *r* enumerated linear correlation.

Comparison of two methods of measuring systolic blood pressure (Data take from [1])



Expectation E

Mean average μ

$$r(X,Y) = \frac{E(X - \mu_X)(Y - \mu_y)}{\sigma_X \sigma_Y}$$

Standard deviation σ

Disclosure: Mean and standard deviation must be obtained from the whole population or from a sample set which is sufficiently large.

In practice *E* is the weighted sum:

$$r(X,Y) = \frac{\sum_{x \in X, y \in Y} \frac{(x - \mu_X)(y - \mu_Y)}{n}}{\sigma_X \sigma_Y}$$

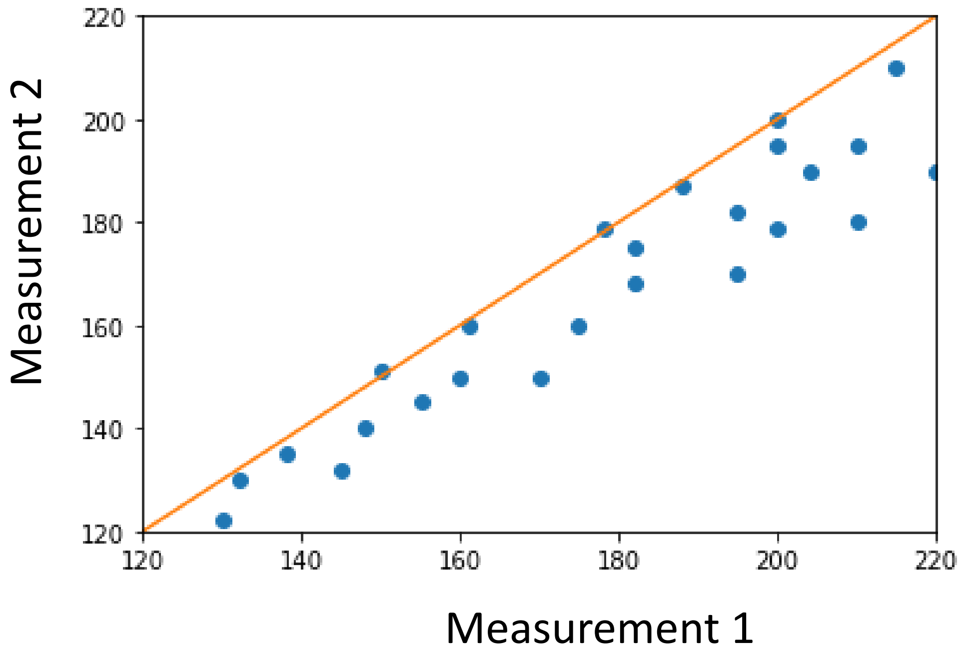
Number of measurements *n*

[1] Altman & Bland, The Statistician 32, 1983

Correlation

- Are two methods doing the same if they correlate?
  - Correlation: Any kind of relationship.
  - Measurable; e.g. using Pearson's Correlation Coefficient *r* enumerated linear correlation.

Comparison of two methods of measuring systolic blood pressure (Data take from [1])



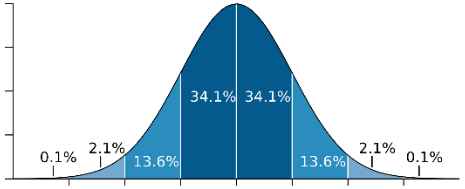
$$r(X,Y) = \frac{\sum_{x\in X, y\in Y}\frac{(x-\mu_X)(y-\mu_Y)}{n}}{\sigma_X\sigma_Y} = 0.94$$

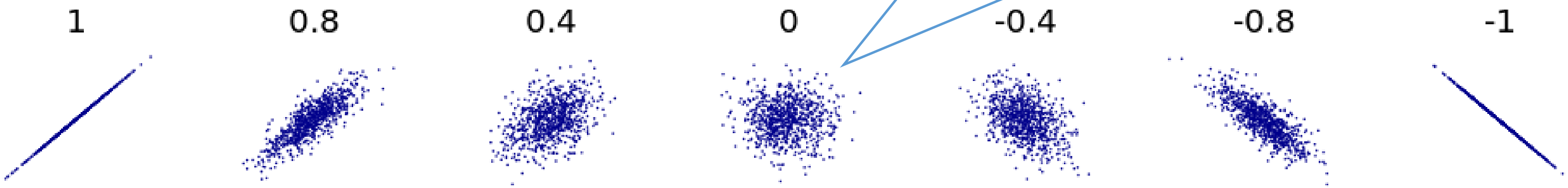
[1] Altman & Bland, The Statistician 32, 1983

Correlation: Pearson's *r*

- Pearson's *r* lies between -1 and 1
  - 1: Positive linear correlation
  - 0: No linear correlation
  - -1: Negative linear correlation

2-dimensional normal distribution



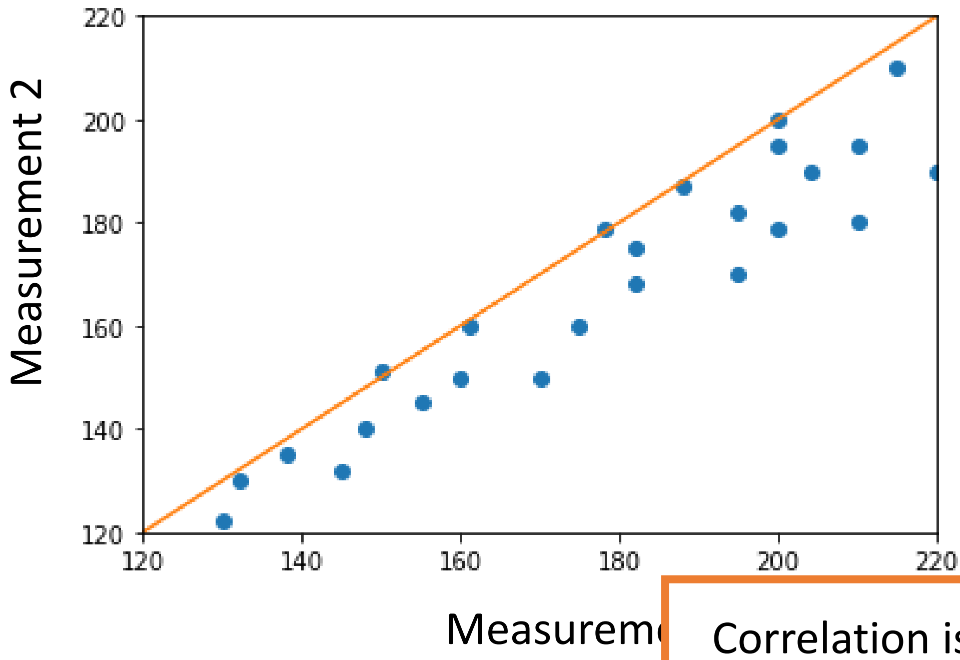


https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#/media/File:Correlation_examples2.svg

Correlation

- Are two methods doing the same if they correlate?
  - Correlation: Any kind of relationship.
  - Measurable; e.g. using Pearson's Correlation Coefficient *r* enumerated linear correlation.

Comparison of two methods of measuring systolic blood pressure (Data take from [1])



$$r(X,Y) = \frac{\sum_{x \in X, y \in Y} \frac{(x - \mu_X)(y - \mu_Y)}{n}}{\sigma_X \sigma_Y} = 0.94$$

"Positive linear correlation"

Measurement 1 is almost always larger than measurement 2

Correlation is necessary, but it is NOT sufficient!

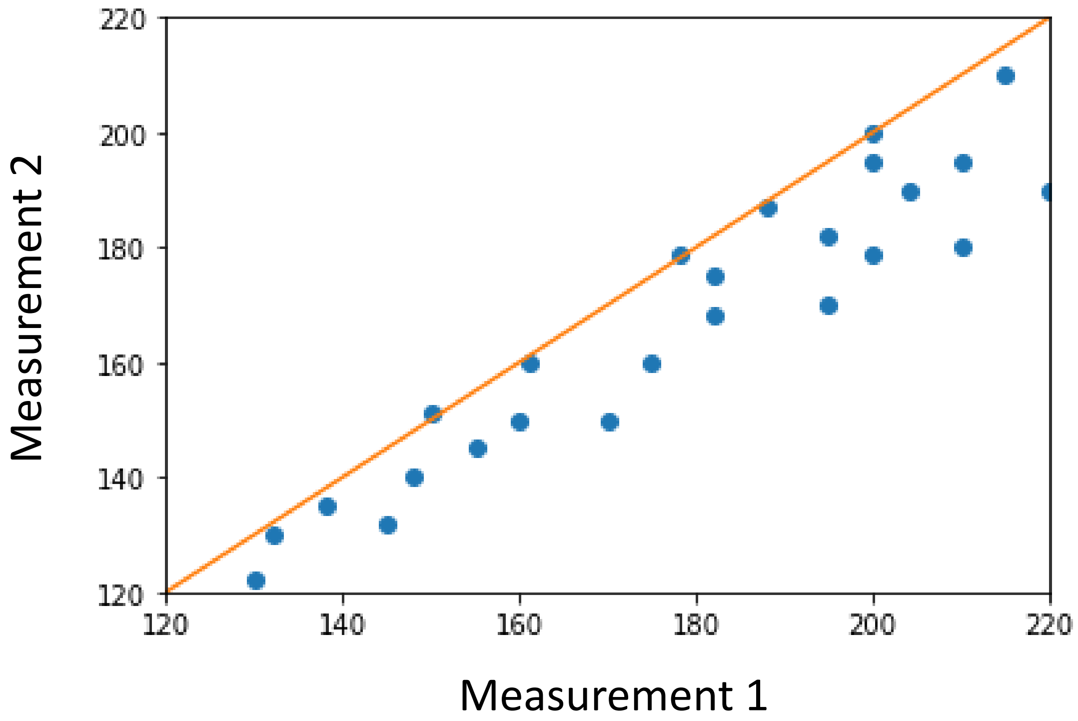
The scientific method: Show that a method doesn't work with *just one* example. And you have *proven* that the method doesn't work in general.

[1] Altman & Bland, The Statistician 32, 1983

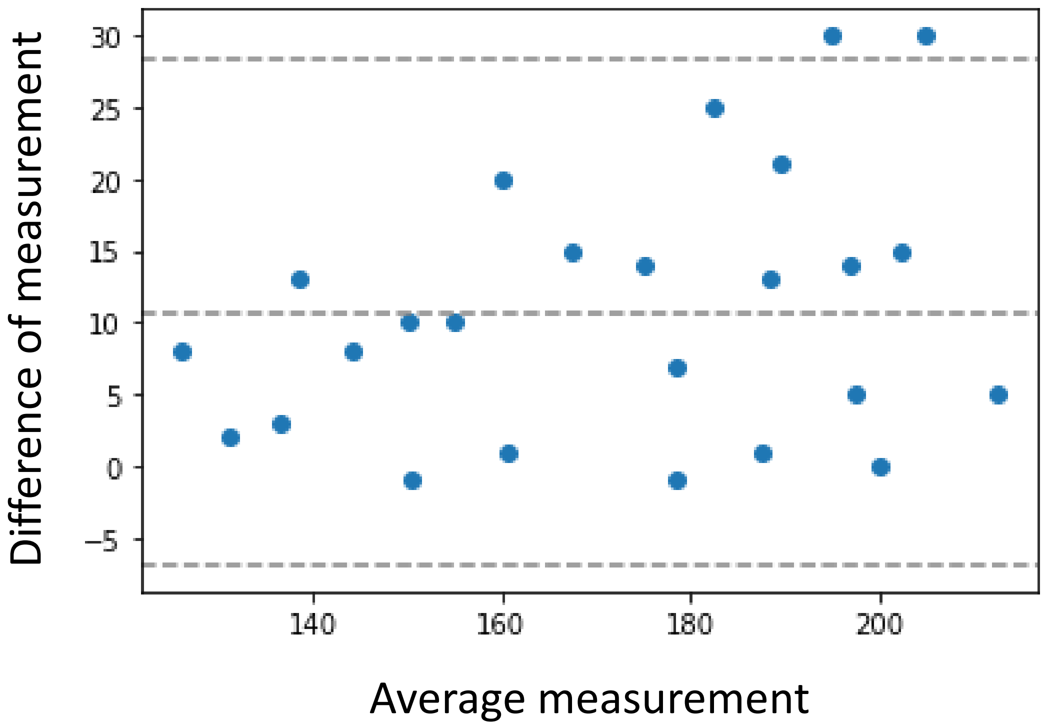
Bland-Altman plots

- In order to evaluate the difference between two methods, you should visualize them first.
- *"The purpose of computing is insight, not numbers."*, Richard Hamming

Scatter plot



Bland-Altman plot



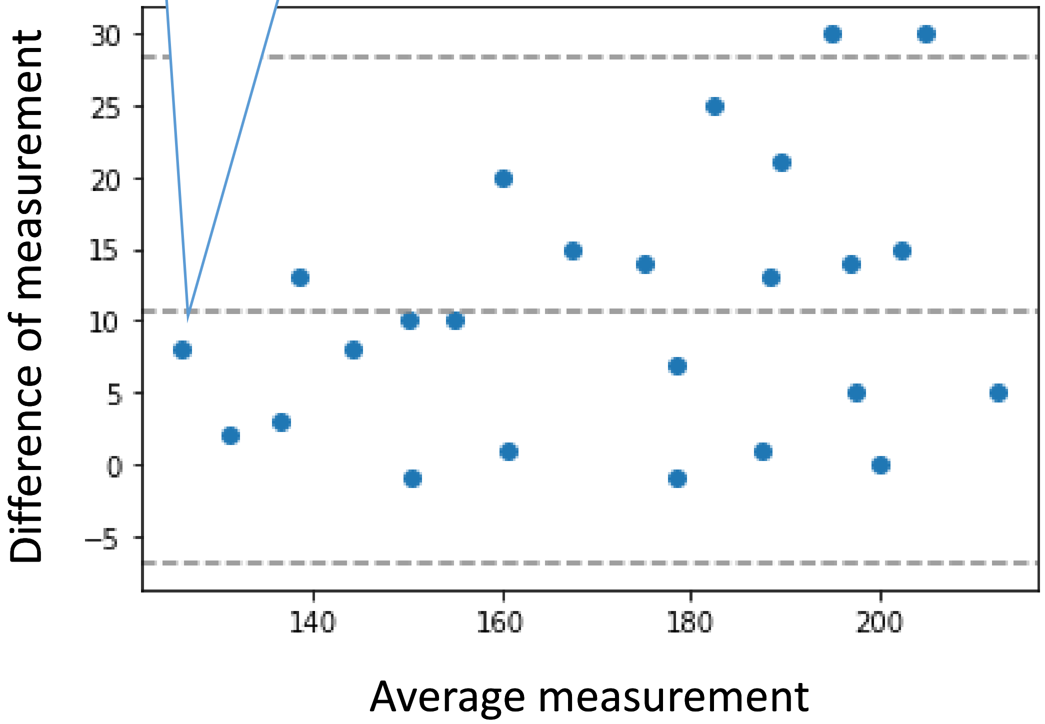
The confidence interval

PoL
Physics of Life
TU Dresden

- “The British Standards Institution (1979) define a coefficient of repeatability as ‘the value below which the difference between two single test results ... may be expected to lie with a specified probability; in the absence of other indications, the probability is 95 per cent’.”[1]

The mean difference between the methods

Bland-Altman plot



The confidence interval *CI* of agreement

$$CI = (\mu - 2\sigma, \mu + 2\sigma)$$

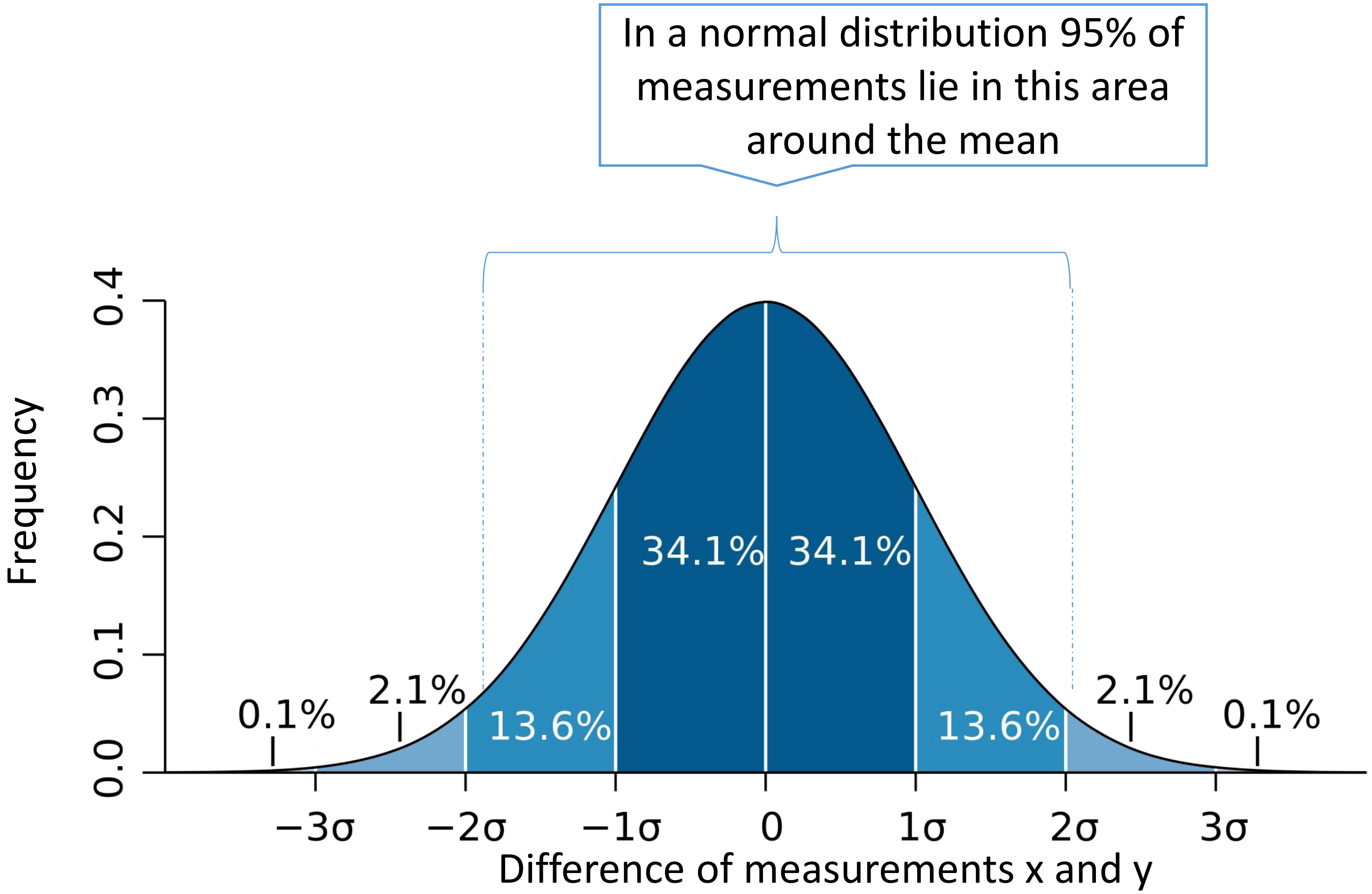
Mean difference

Standard deviation of differences

[1] Altman & Bland, The Statistician 32, 1983

The confidence interval & the coefficient of repeatability.

- "The British Standards Institution (1979) define a coefficient of repeatability as 'the value below which the difference between two single test results ... may be expected to lie with a specified probability; in the absence of other indications, the probability is 95 per cent'."[1]

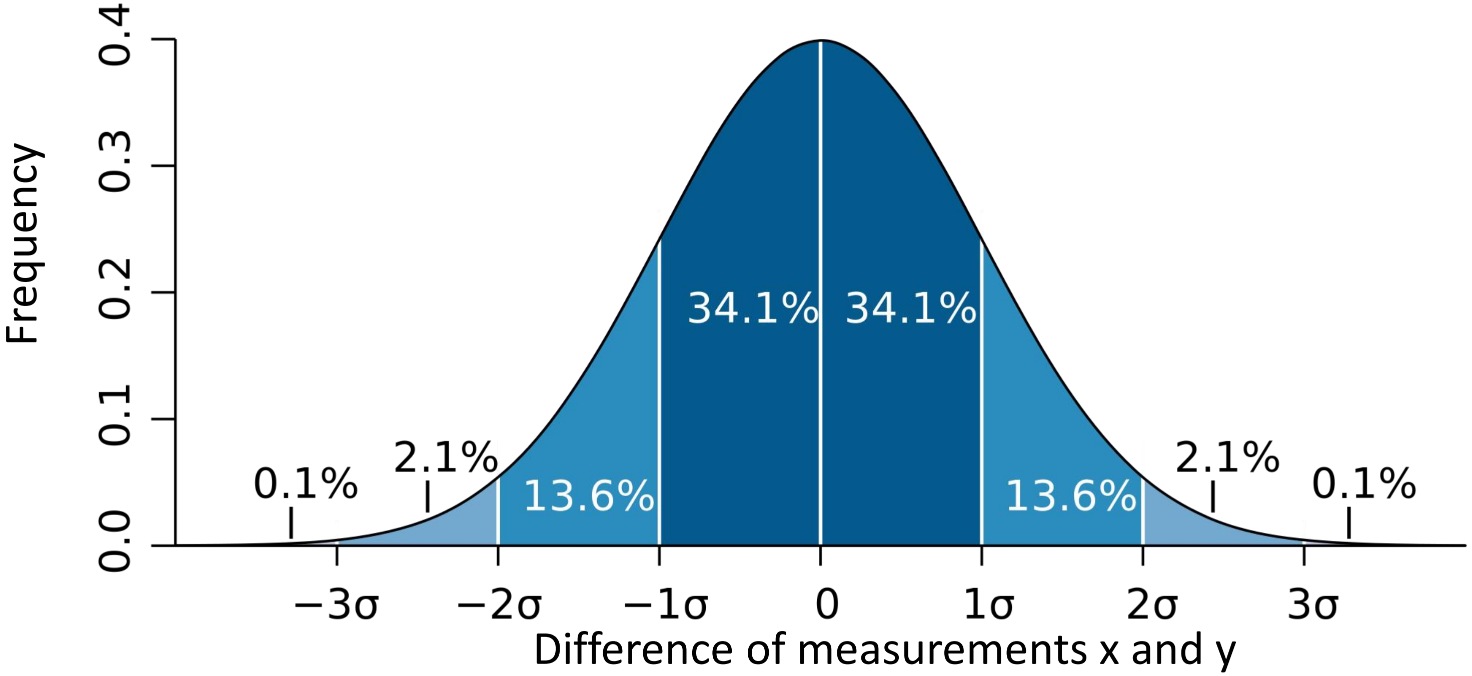


Graph adapted from: M. W. Toews - Own work, based (in concept) on figure by Jeremy Kemp, on 2005-02-09, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=1903871

[1] Altman & Bland, The Statistician 32, 1983

The confidence interval & the coefficient of repeatability.

- "The British Standards Institution (1979) define a coefficient of repeatability as 'the value below which the difference between two single test results ... may be expected to lie with a specified probability; in the absence of other indications, the probability is 95 per cent'."[1]
- If the two measurements come from the same method which just repeated twice, we can assume that the main difference is zero. The coefficient of repeatability *CR* can then be estimated: It's the standard deviation of differences.[2]



$$\mathrm{CR}(X,Y) = \sqrt{\sum_{x\in X, y\in Y} \frac{(x-y)^2}{n}}$$

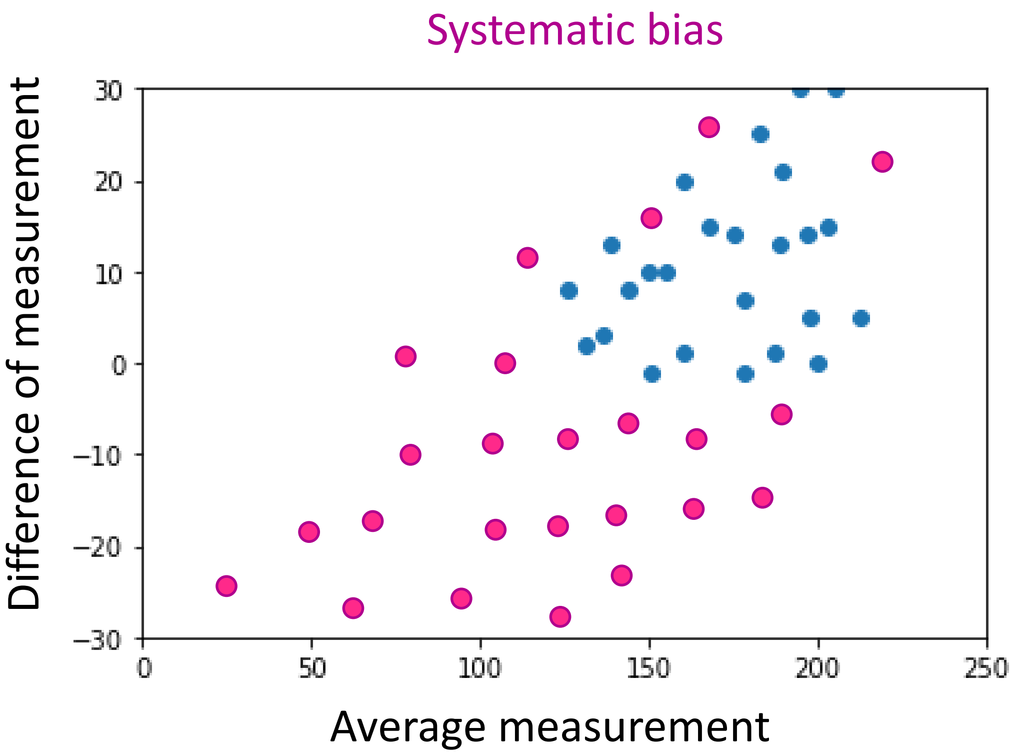
Graph adapted from: M. W. Toews - Own work, based (in concept) on figure by Jeremy Kemp, on 2005-02-09, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=1903871

[1] Altman & Bland, The Statistician 32, 1983

2 Bland & Altman, Lancet , 1986

In practice:

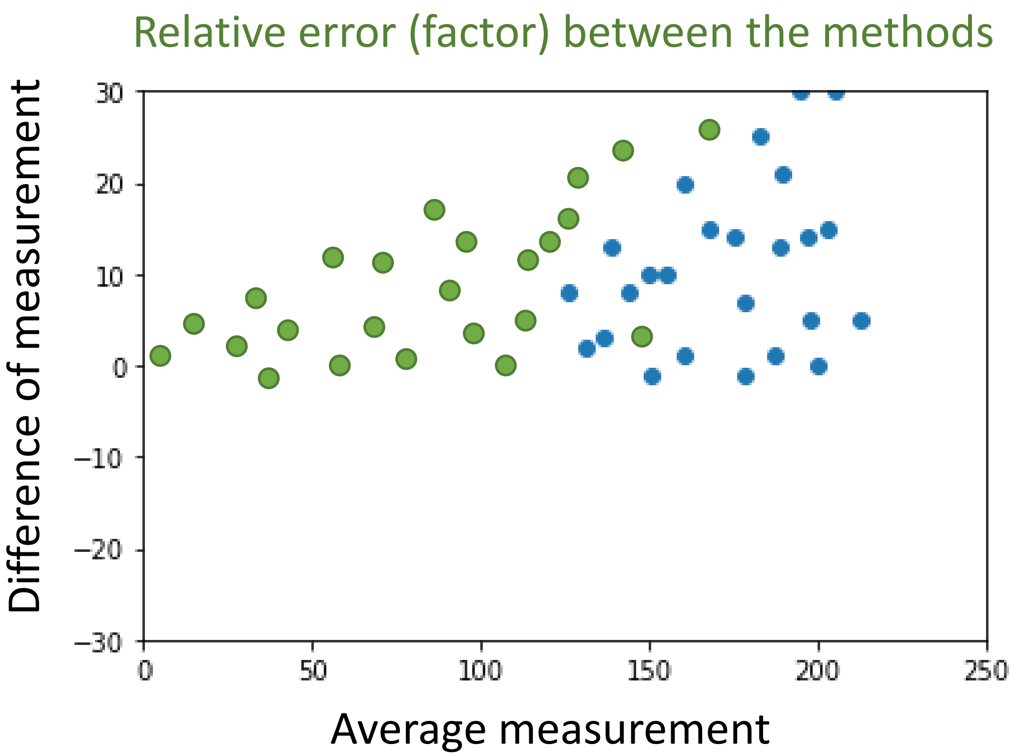
- Bland-Altman plots allow us to differentiate various kinds of bias.

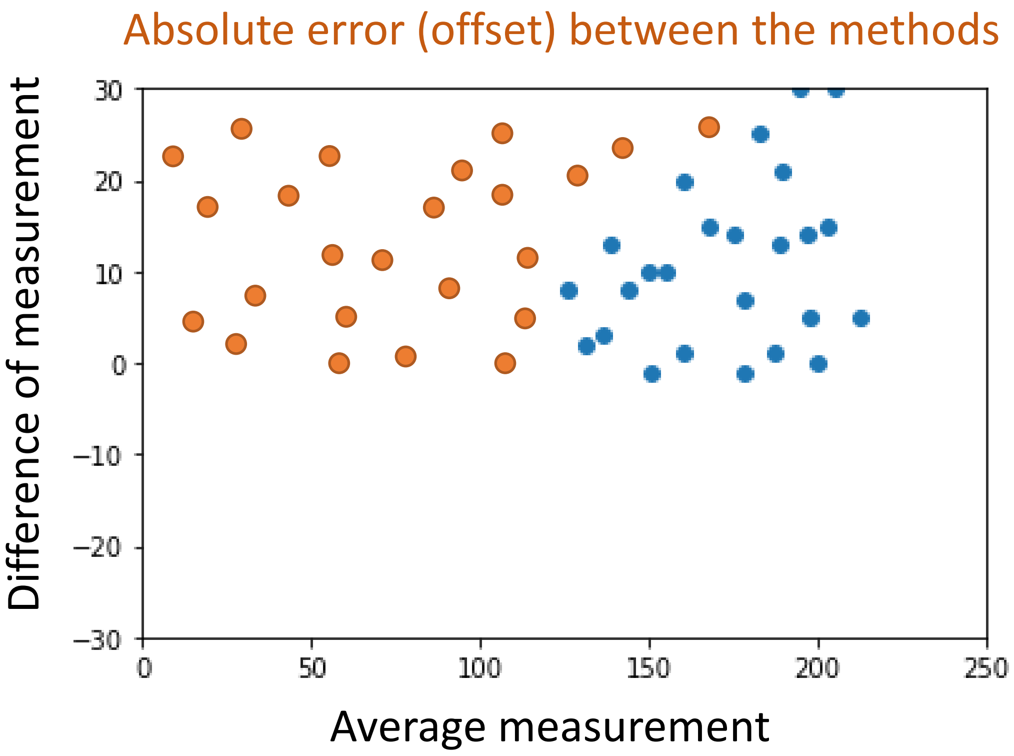


- Both distributions could have the same mean difference and confidence interval.

In practice:

- Bland-Altman plots allow us to differentiate various kinds of bias.





- Both effects can be corrected by calibration.

Summary: descriptive statistics

- Summarizing measurements
  - To communicate with others
  - To draw conclusions from them (-> inferential statistics)
- Measures of central tendency
  - Mean
  - Median
  - Mode
- Measures of spread
  - Standard deviation
  - Variance
  - Percentiles
  - Full width at half maximum
- Method comparison
  - Comparison of means (not sufficient!)
  - Pearson’s correlation coefficient
  - Scatter plots
  - Bland-Altman plots
  - Confidence interval
  - Coefficient of repeatability.

- Coming up next:
  - Machine learning for pixel and object classification