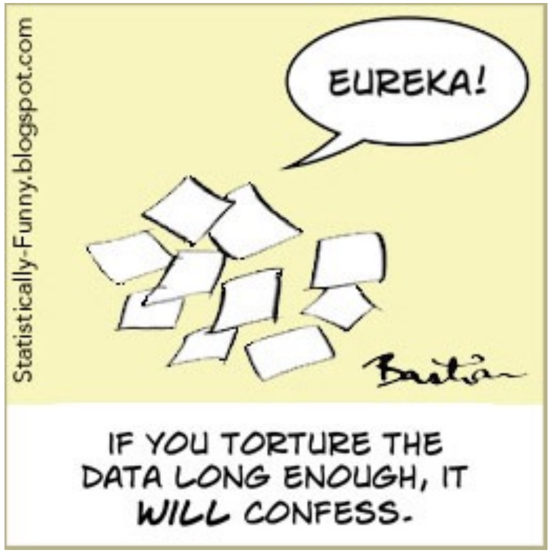
# Introduction to biostatistics

Dr. Hristo Todorov,
AG Prof. Dr. Susanne Gerber

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

## What statistics is (not) all about


"Data don't make any sense, we will have to resort to statistics."




## Branches of statistics

- **Descriptive statistics**
  - Describe, summarize, order or graphically represent empirical data
- **Exploratory data analysis**
  - Identify patterns or structures in the data
- **Inferential statistics**
  - Predict, estimate and generalize about populations based on data derived from samples

# Descriptive statistics

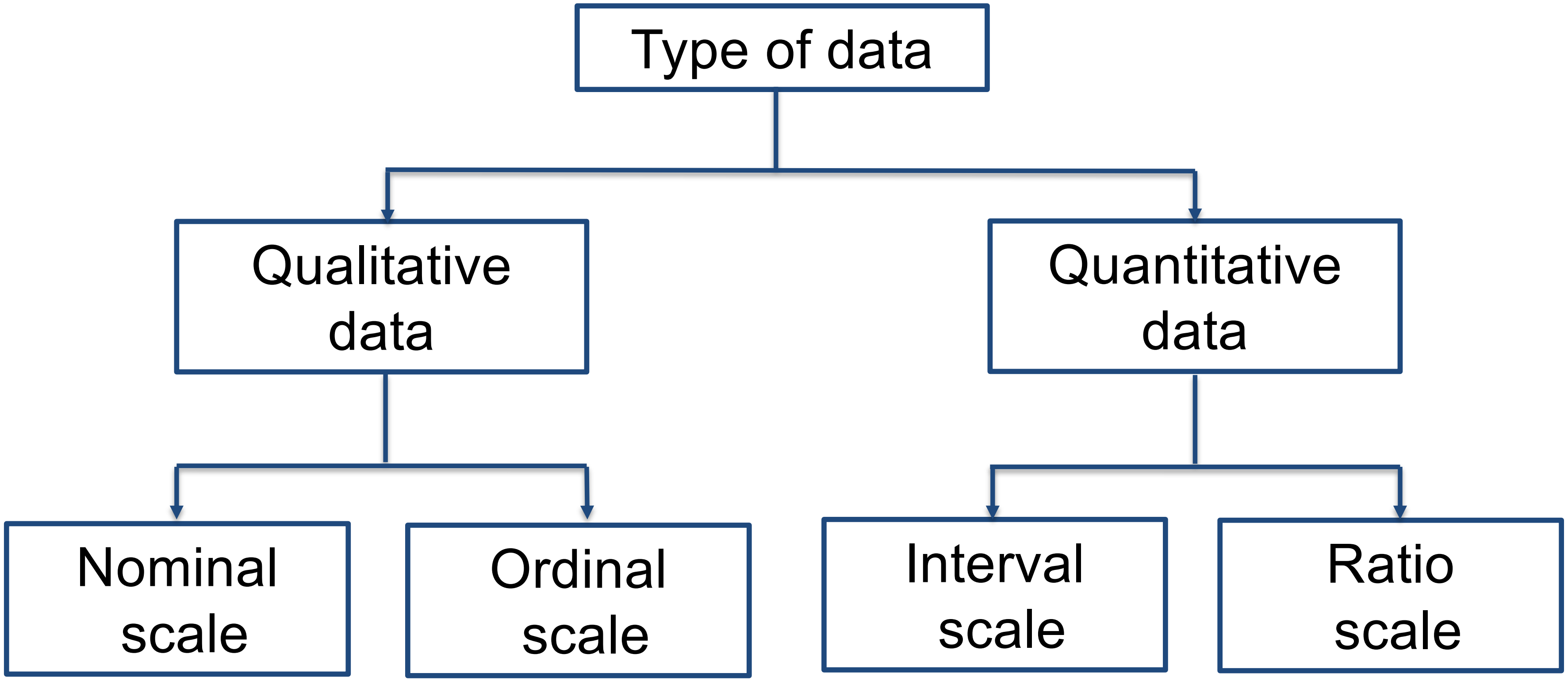
# Descriptive statistics

## Introduction to descriptive statistics

- Descriptive statistics provide the basis of quantitative data analysis
- The focus is to describe, summarize and graphically represent the data as it is without making any assumptions or generalizations
- Descriptive statistics provide measures to:
  - Describe the central tendency of data
    - Mean, median, modus
  - Describe the dispersion of the data
    - Variance, standard deviation, range
  - Describe how different data are related to each other
    - Correlation coefficient

# Descriptive statistics

## Measurement scales



# Descriptive statistics

## Measurement scales

- Nominal scale
  - The lowest level of measurement
  - Data belong to mutually exclusive categories
  - The only possible comparison between elements is “equal to” or “not equal to”
  - Examples include:
    - Blood type
    - Color of your eyes
    - Nucleotides in the DNA sequence
    - University field of studies

# Descriptive statistics

## Measurement scales

- Ordinal scale
  - Data belong to mutually exclusive categories which can be ordered (<,> comparisons allowed)
  - Differences between categories are not allowed
  - Examples include:
    - Exam grades (A, B, C, D)
    - Questionnaire options (“strongly agree”, “agree”, disagree”, strongly disagree”)
    - Levels of happiness, satisfaction, etc.
    - Some clinical scores

# Descriptive statistics

## Measurement scales

- Interval scale
  - A metric scale which allows building differences between values but not ratios
  - Examples include:
    - Celsius temperature
    - IQ score
    - Time on a clock
  - No “true” zero value is defined on an interval scale
    - 0 degrees Celsius does not mean there is no temperature
    - 0:00 does not mean time does not exist

# Descriptive statistics

## Measurement scales

- Ratio scale
  - A metric scale which allows building differences and ratios between values
  - A value of zero indicates non-existence
  - Examples include:
    - Age
    - Height
    - Weight
    - Number of children
    - Distance
    - Blood pressure

- Note: higher levels of measurements can be reduced to lower scales but not the other way around

# Descriptive statistics

## Measures of central tendency

- Central tendency, center or location of the data is a central or typical value

- Mode
  - The element if the data which appears most often
  - The mode can be determined for variables measured on any scale

  *Out of 40 participants in a statistics course, 10 are studying bioinformatics, 25 are studying biomedicine and 5 are studying biochemistry. What is the mode for the the variable field of studies of the course takers?*

# Descriptive statistics

## Measures of central tendency

- Median
  - The “middle” value in an ordered data set, half of the data are smaller and half of the data are larger than the median
  - Can be calculated for data measured at least on an ordinal scale

- Arithmetic mean
  - The arithmetic mean only makes sense for variables on a metric scale

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

# Descriptive statistics

## Measures of central tendency

- Mean or median?
- Consider a small sample containing the following values

$$\{3.5; 4; 2.5; 6; 7; 5.5\}$$

$$\bar{x} = \frac{3.5 + 4 + 2.5 + 6 + 7 + 5.5}{6} = 4.75$$

$$Median = 4.75$$

- Consider adding the value 500 to the set

$$\{3.5; 4; 2.5; 6; 7; 5.5; 500\}$$

$$\bar{x} = \frac{3.5 + 4 + 2.5 + 6 + 7 + 5.5 + 500}{7} = 75.5$$

$$Median = 5.5 - \ a\ robust\ statistic$$

# Descriptive statistics

## Measures of dispersion

- Empirical variance

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

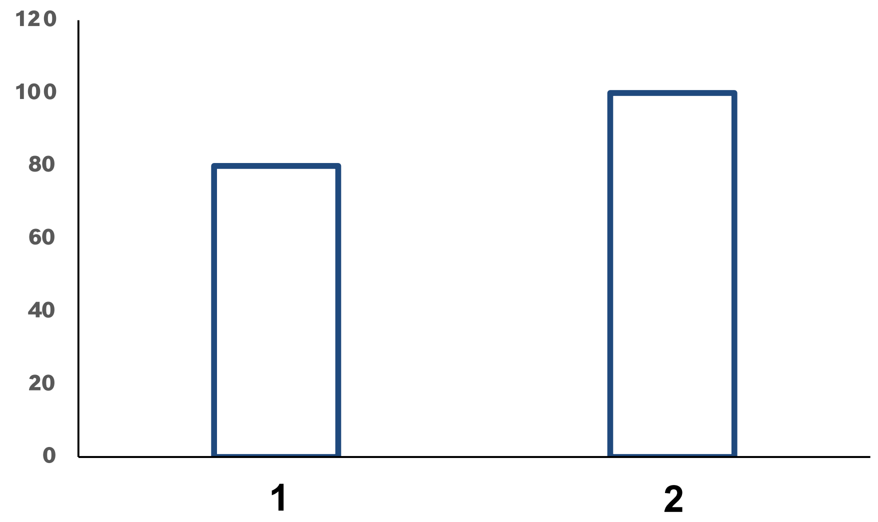
- Standard deviation

$$s = \sqrt{s^2}$$

# Descriptive statistics

## Graphical representation of location and dispersion

- *Suppose that blood pressure was measured in two samples of 30 patients each. Mean diastolic blood pressure for the first sample turned  out to be 80, whereas mean pressure in the second sample was 100 mmHg.*

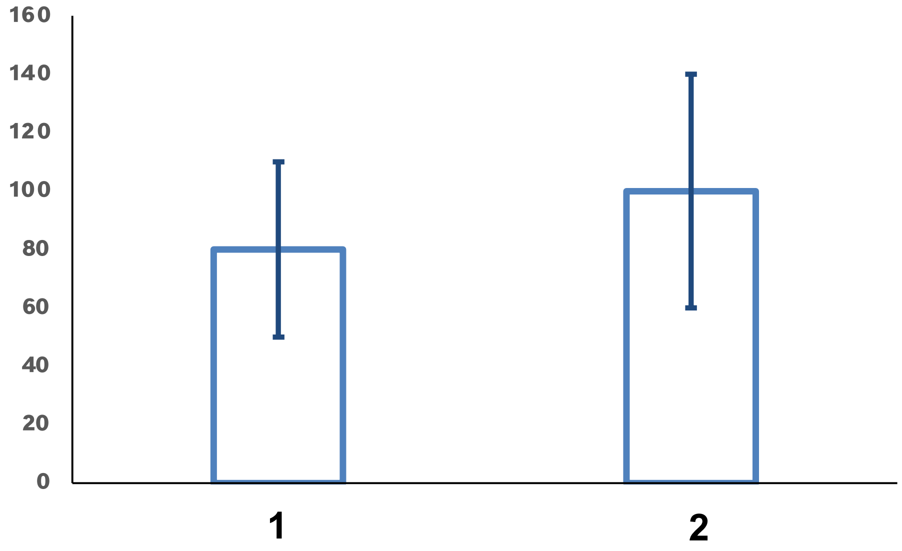



- *Do people in the second sample have higher diastolic blood pressure?*

# Descriptive statistics

## Graphical representation of location and dispersion

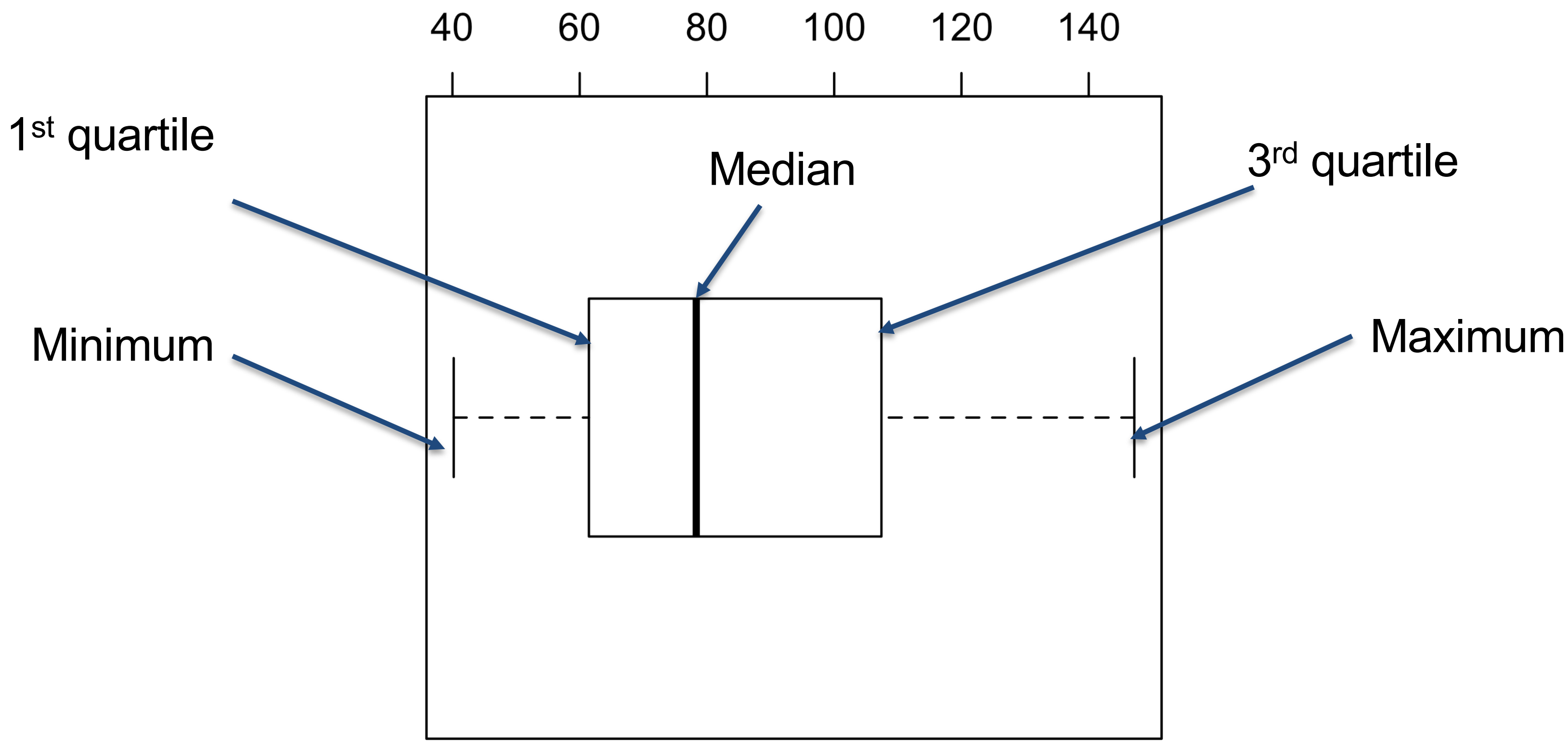
- *Suppose that blood pressure was measured in two samples of 30 patients each. Mean diastolic blood pressure for the first sample turned  out to be 80* ± 30*, whereas mean pressure in the second sample was 100* ± 40 *mmHg.*



# Descriptive statistics

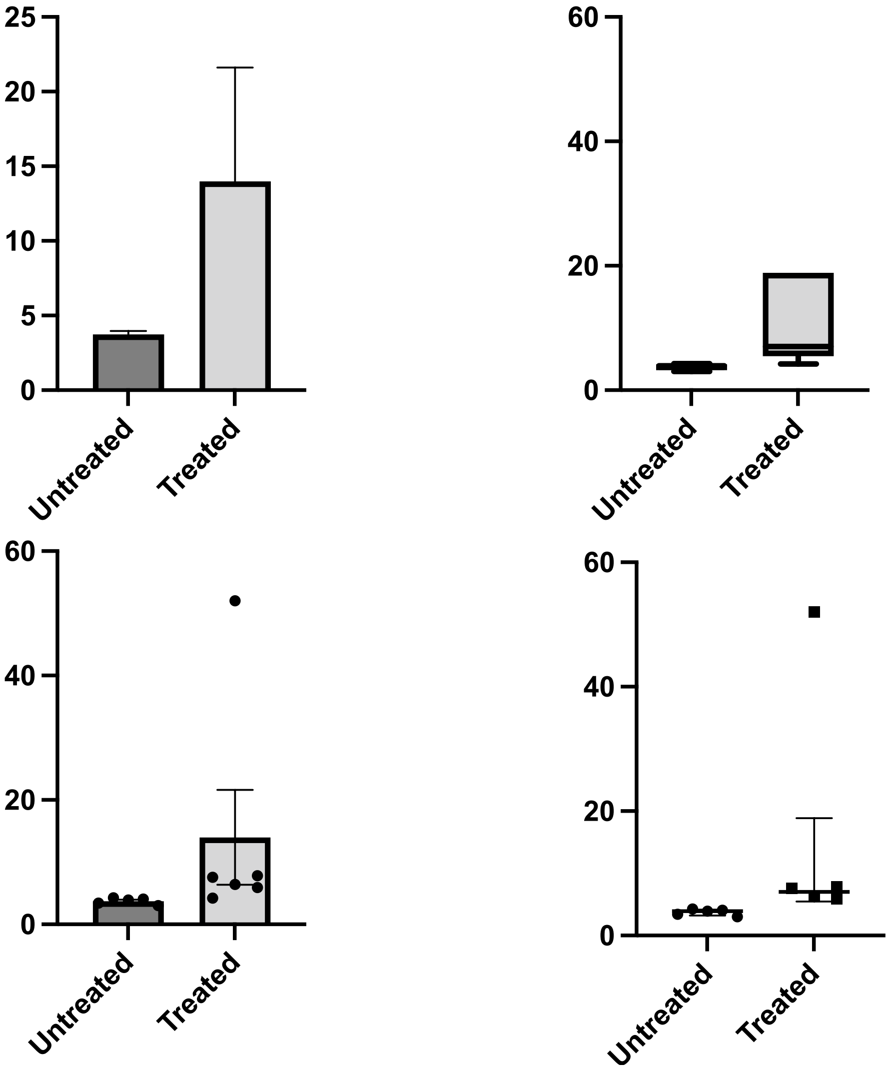
## Boxplot

- Consider the first sample of 30 patients with mean diastolic blood pressure of 80, with a standard deviation of 30



# Descriptive statistics

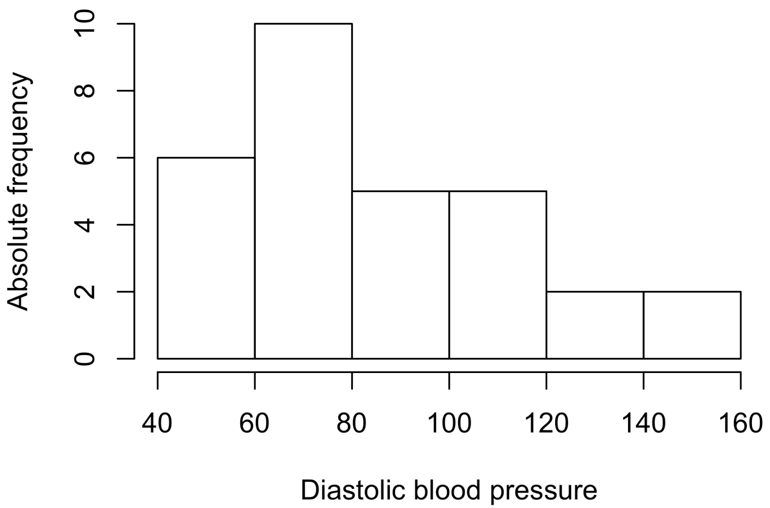
Which is the "best" graphical representation?

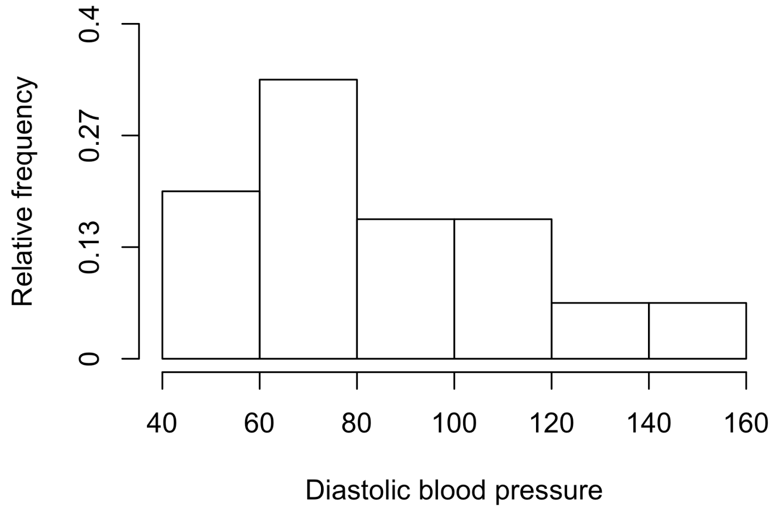


# Descriptive statistics

## Histograms

- Consider the first sample of 30 patients with mean diastolic blood pressure of 80, with a standard deviation of 30

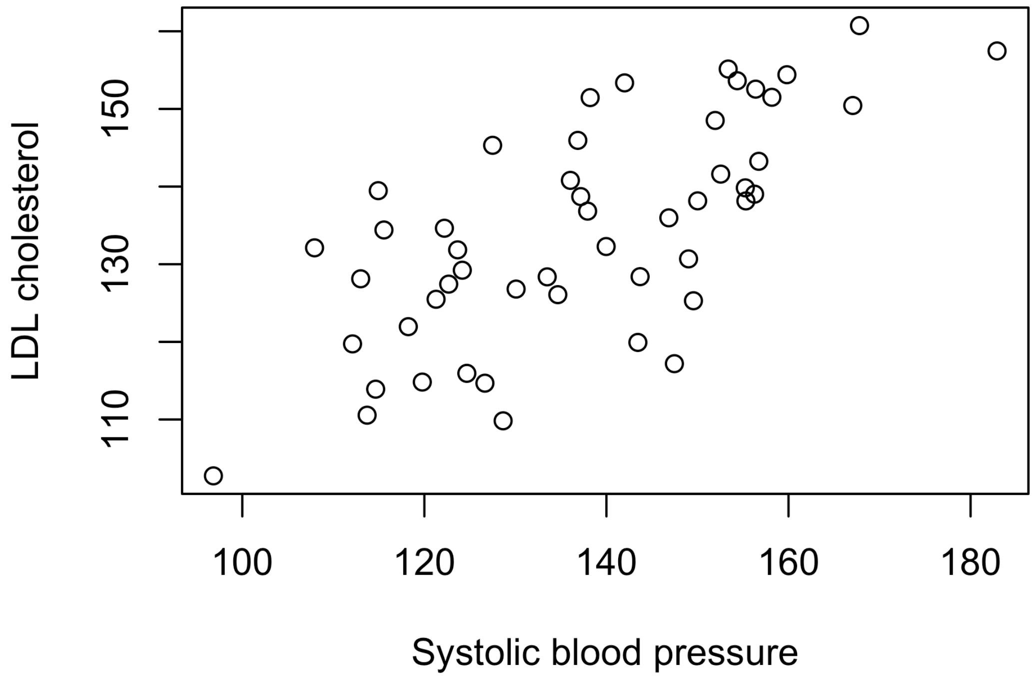


# Descriptive statistics

## Association between metric variables

- *In a study of 50 patients, systolic blood pressure and LDL cholesterol were measured. Mean blood pressure was 137.5 mmHg with a standard deviation of 18.27. Average LDL cholesterol was 134 mg/dL with a standard deviation of 14.14.*
  - *Is higher blood pressure associated with higher LDL cholesterol levels?*



# Descriptive statistics

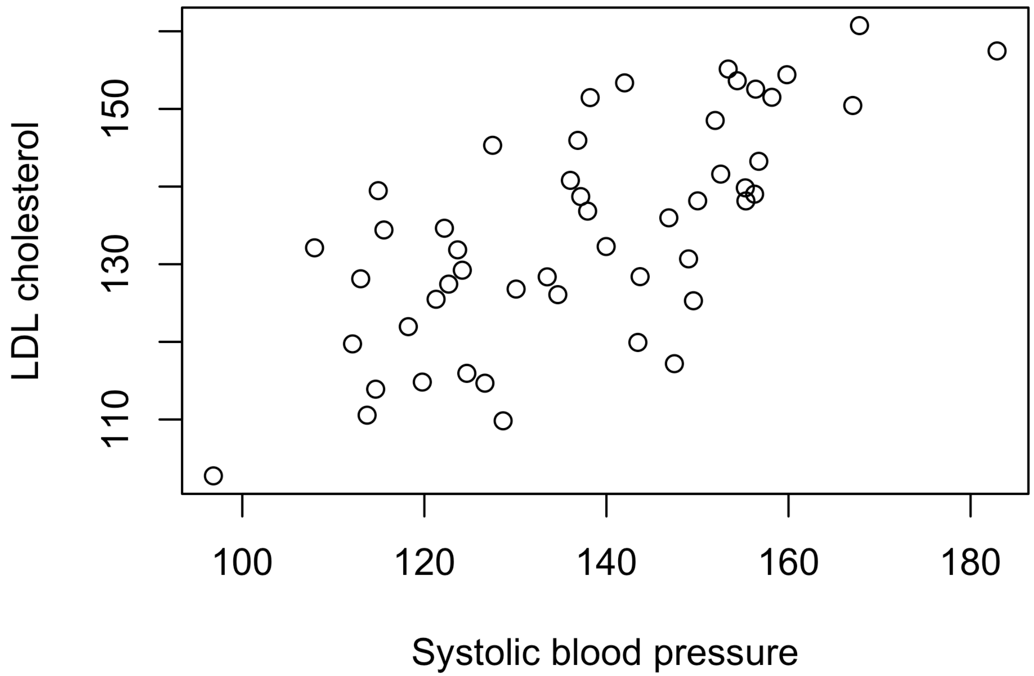
## Association between metric variables

- Pearson correlation coefficient
  - Measures the linear relationship between metric variables
  - Values range from -1 to 1
    - -1 - perfect negative correlation
    - 0 – no correlation
    - +1 – perfect positive correlation

# Descriptive statistics

## Association between metric variables

- *In a study of 50 patients, systolic blood pressure and LDL cholesterol were measured. Mean blood pressure was 137.5 mmHg with a standard deviation of 18.27. Average LDL cholesterol was 134 mg/dL with a standard deviation of 14.14.*
  - *Is higher blood pressure associated with higher LDL cholesterol levels?*

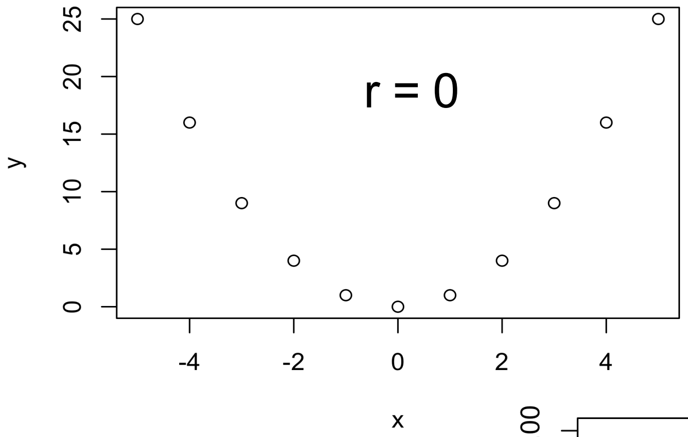


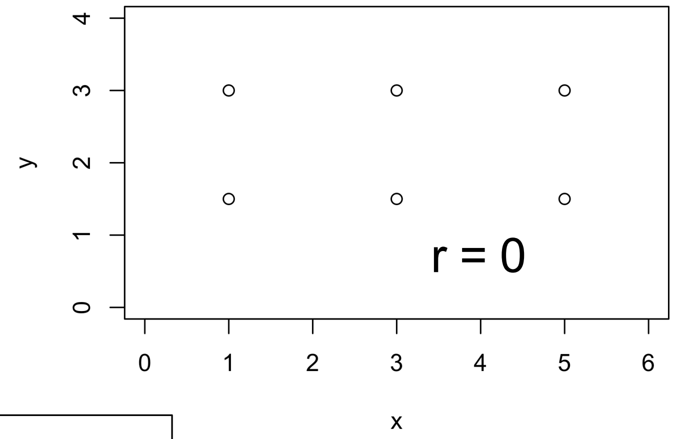

Pearson r = 0.71

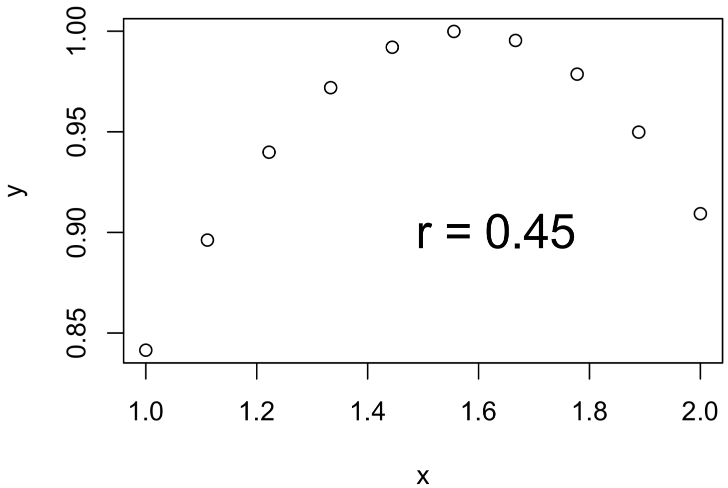
# Descriptive statistics

## Pearson correlation

- Pearson correlation coefficient is only appropriate if the relationship between the variables is linear



# Descriptive statistics

## Spearman correlation

- The Spearman correlation coefficient is an alternative to the Pearson correlation coefficient appropriate when:
  - The relationship between the variables is monotonically increasing or decreasing but not linear
  - Data are measured on an ordinal scale
- Values range from -1 to 1, 0 corresponds to no correlation
- Data are arranged in an increasing order and every data point is assigned a rank
- Spearman correlation belongs to the category of robust statistics because it is robust in the presence of outliers

# Pearson vs. Spearman correlation

- Influence of outliers

| x | y |
|---|---|
| 1 | 5 |
| 2 | 6 |
| 3 | 7 |
| 4 | 8 |
| Pearson r = 1 | |
| Spearman r = 1 | |

| x | y |
|---|---|
| 1 | 5 |
| 2 | 6 |
| 3 | 7 |
| 4 | 8 |
| 100 | 9 |
| Pearson r = 0.72 | |
| Spearman r = 1 | |

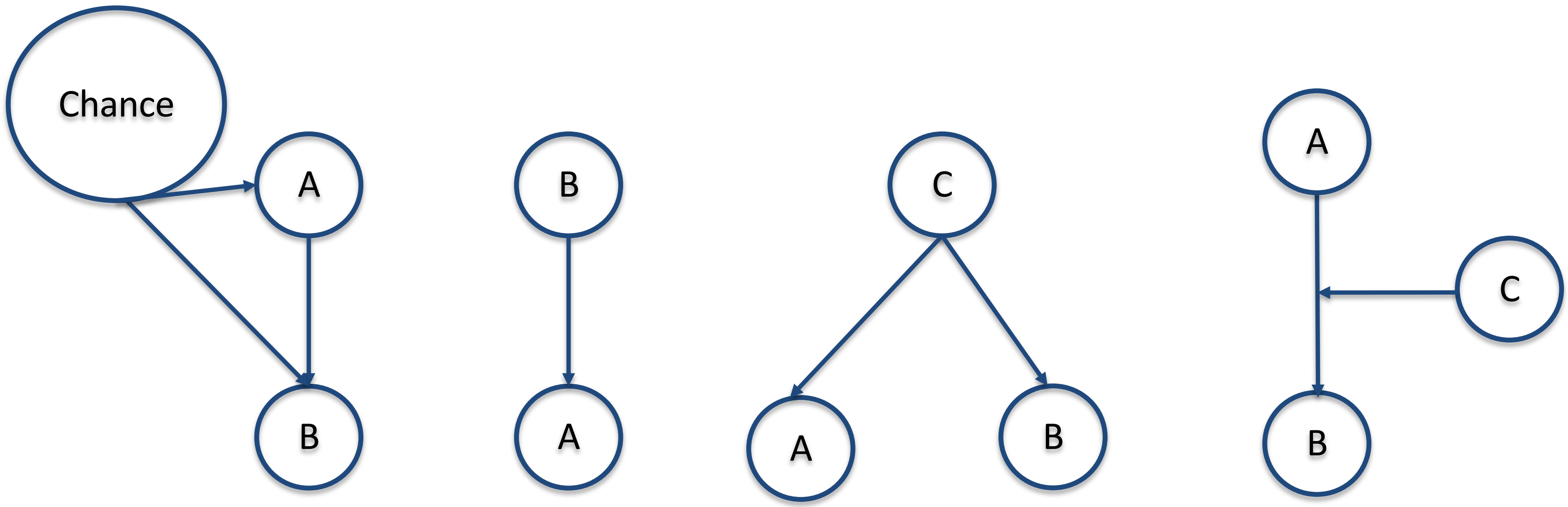
## Correlation versus causality

- Causality is often the focus of research
  - ➢ We try to explain why phenomena occur and what causes them
- Caution: statistics only provides measures of association. The question of what is cause and what is consequence remains open.
- Causality, therefore, requires careful theoretical consideration and appropriate experimental designs
  - Causality can often be masked by random confounding factors or latent factors

# Descriptive statistics

## Correlation versus causality

- Consider two variables A and B which are highly correlated with each other.
- Different causal relationships are possible



# Probability theory

# Probability theory

## Random processes

- Random process – a phenomenon with an uncertain outcome. Examples include:
  - A coin toss or throwing a dice
  - The sex of an unborn child
  - Result from an exam
  - A scientific experiment
- The set of all possible outcomes of a random process is called a **sample space Ω.**
  - E.g., the colour of a gummy bear we take out of a bag with blue, red, yellow and green gummy bears is a random process with the following sample space:

$\Omega$ = {„blue“, „red“, „yellow“, „green“}

# Probability theory

## Radom events

- The outcome of a **random experiment** is called a **random event**
  - **Simple event** – an event which contains only a single outcome

    *E.g. A green gummy bear*
  - An event is a unification of simple events

    *E.g. A green gummy bear or a yellow gummy bear;*

    *Not a blue gummy bear*
- Although the single outcome of a random process is uncertain, outcomes follow certain distributions if the random experiment is repeated many times
- The chance of a random outcome occurring is described by **the probability**

## Frequency versus probability

- Consider the following random experiment:
  - *A pot contains 50 white and 50 black balls. What is the probability of drawing a white ball if we draw with replacement (we put the ball back in the pot after drawing it)*
  - *An R simulation was performed investigating relative frequency of white balls after repeating the experiment 10 000 times by drawing 1 to 10 000 balls.*

# Probability theory

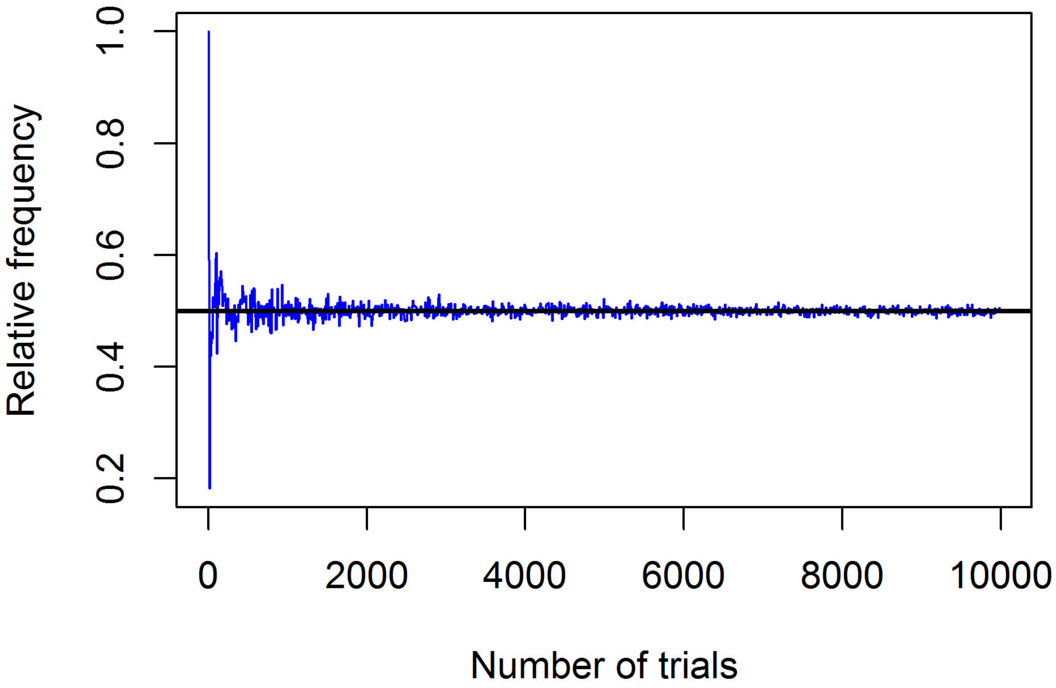
## Frequency versus probability

| Number of draws | Number of white balls | Relative frequency of white balls |
|---|---|---|
| 1 | 1 | 1 |
| 11 | 2 | 0.182 |
| 101 | 61 | 0.604 |
| 501 | 238 | 0.475 |
| 1001 | 494 | 0.494 |
| 9991 | 4985 | 0.499 |

- The relative frequency approaches 0.5 as the number of trials increases

# Probability theory

## Frequency versus probability




- The **relative frequency** converges against a limit value as the number of trials **n** increases. The limit is called **probability**

# Random variables and probability distributions

# Random variables

## Definition of a random variable

- A **random variable** is a numerical representation of a random phenomenon. The values of the random variable correspond to the outcomes of the random process
- The values of the random variable are called realizations
  - For example, flipping a coin is a random variable X with values 0 (in case of heads) and 1 (in case of tails)
  - The color of a ball we draw from a pot with blue, red and green balls is a random variable Y with values 0,1 and 2, respectively
  - The outcome of throwing a dice is a random variable Z with values {1,2,3,4,5,6}
- Notation:
  - Random variables – X,Y,Z
  - Realizations – $x_i$, $y_i$, $z_i$

# Random variables

## Discrete random variables

- Random variables with a finite, countable set of possible values are called **discrete random variables**
  - The number of heads after flipping a coin n times
  - The number of children in a family
  - The field of studies from a randomly selected student on the campus at a given time point

## Continuous random variables

- A random variable which can take an infinite number of values in an interval [a; b] is called a **continuous random variable**
  - Height and weight of a sample of citizens
  - Blood pressure of patients in a study to test anti-hypertension therapy
  - The time needed to solve an exercise during an exam
  - The daily revenue of a supermarket

- Random variables in statistics follow certain distributions called **probability distributions**

# Normal distrubution

## Motivation for the normal distribution

- Many **empirical distributions** can be approximated by a normal distribution (e.g. height of adults, duration of pregnancy)
- Represents the **limit distribution** of many other probability distributions (e.g. binomial, poisson, t-distribution)
- **Sampling distributions** asymptotically approximate the normal distribution for large sample sizes
- The normal distribtution provides the theoretical basis for numerous models in statistics

# Normal distribution

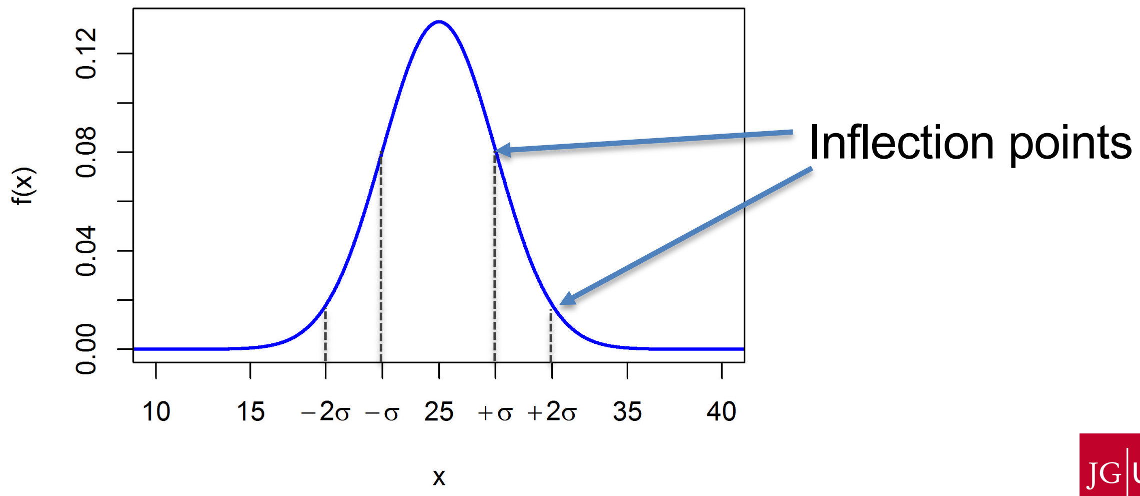
## Features of the normal distribution

- Completely defined by its two paramers expected value $\mu$ and variance $\sigma^2$:

$$X \sim N\,(\mu;\,\sigma^2)$$

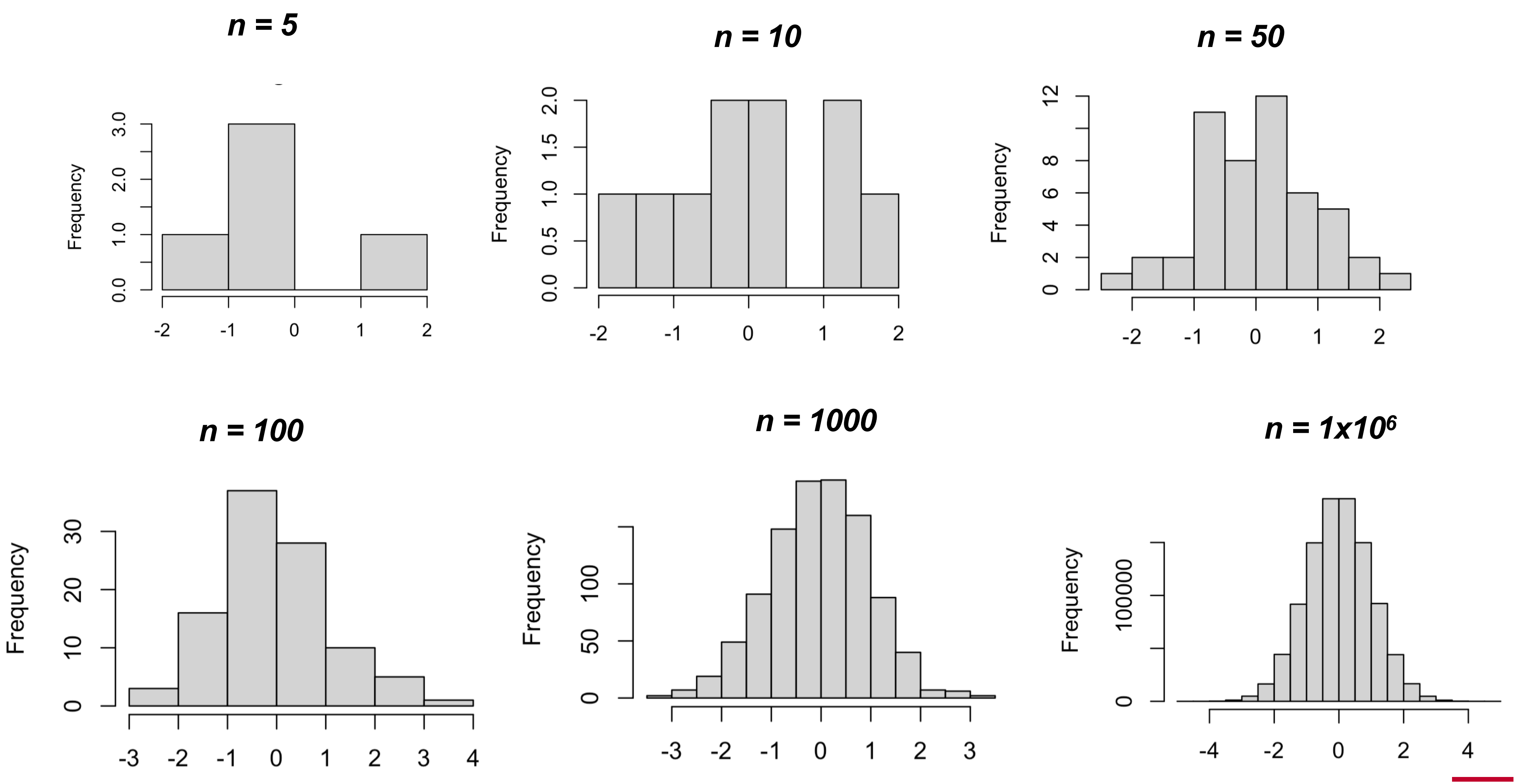
- Density function is unimodal with characteristic „bell“ shape
- Density is maximal at $X = \mu$ and symmetrical around $\mu$ with inflection points $+\sigma$ and $-\sigma$

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$



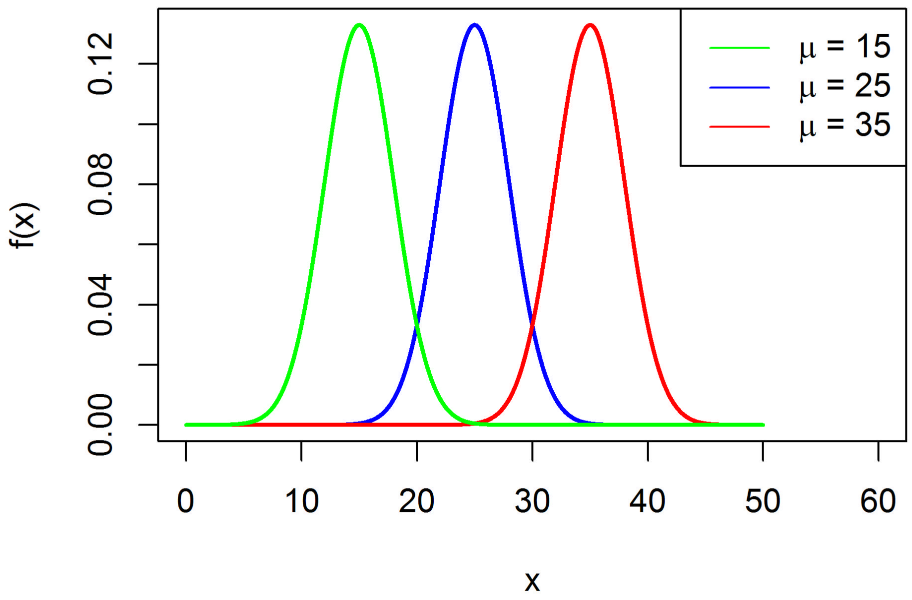
# Normal distribution

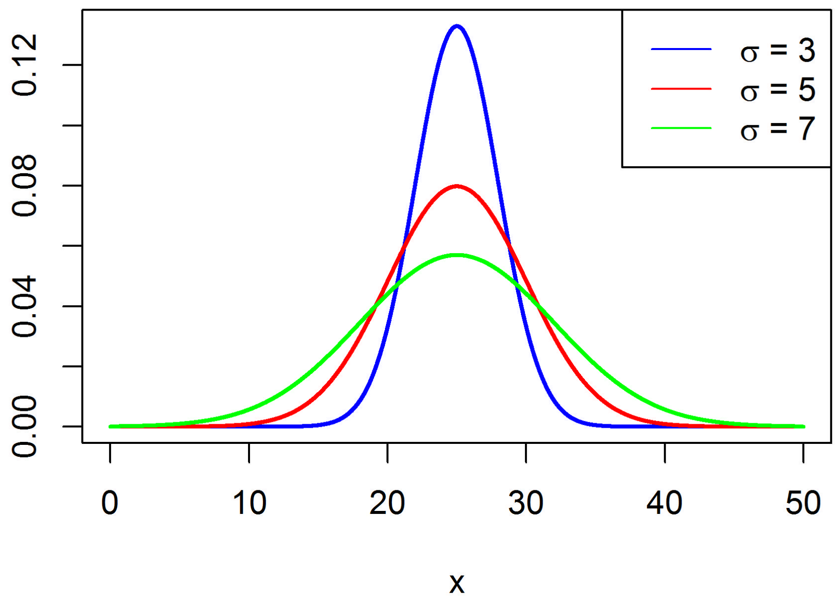
## From frequency to density



# Normal distribution

## Normal distributions for different parameters



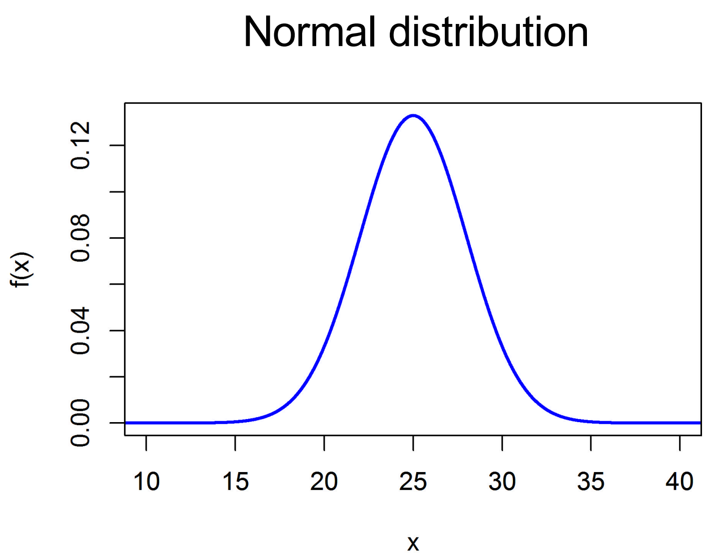



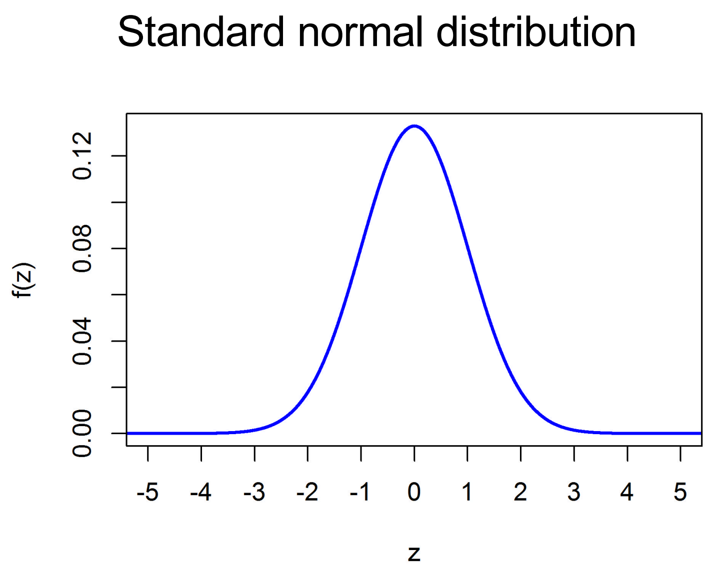
# Normal distribution

## Standard normal distribution

- Expected value E(*Z*) = μ = 0
- Variance Var(X) = $\sigma^2 = 1$
- Every normal distribution can be transformed in the standard normal distribution using the following transformation

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$



# Normal distribution

## Quantiles of the standard normal distribution

- Quantiles of the standard normal distribution with the corresponding cumulative distribution function values are summarized in statistical tables
- Since there are infinitely many normal distributions, they are usually transformed into a standard normal distribution

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |

# Inferential statistics and hypothesis testing

# Inferential statistics

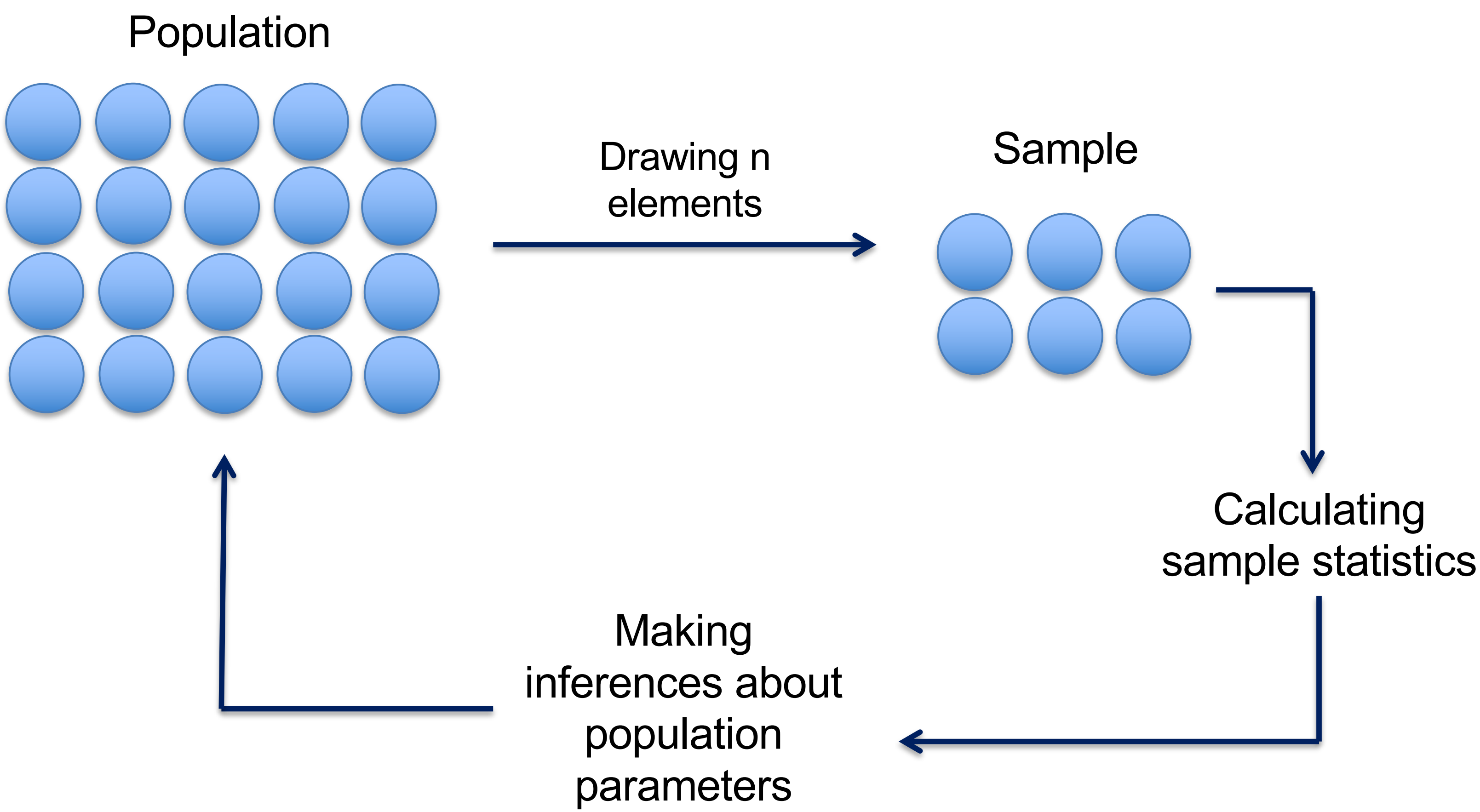
## Population vs. sample

- **Statistical population –** a set of similar elements of interest in a reasearch question or experiment
  - *E.g. All patients undergoing cardiac surgery over the age of 65*
  - Usually it is not possible to investigate a whole population due to time or cost reasons

- **Statistical sample –** a subset of a statistical population chosen by following a specific stragegy or criterion
  - *E.g. All patients undergoing heart valve surgery at the University Medicine Mainz in the period between January 2021 and December 2021*

# Inferential statistics

## Population vs. sample



## Population parameters and sample statistics

- Population parameters are unknown because the entire population is not available
- The measures which characterize samples are called statistics – directly calculated from the data
- Sample statistics are used to infer population parameters

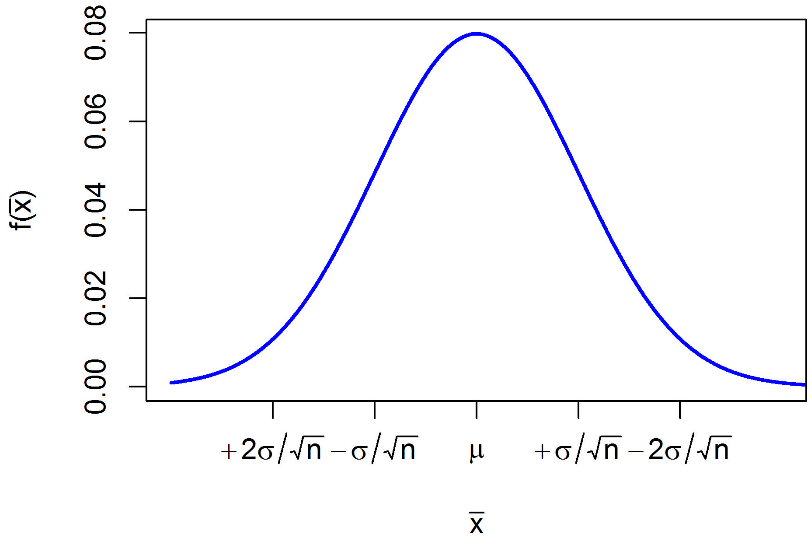
| Population parameter | Sample statistic |
|---|---|
| Expected value $\mu$ | Sample mean $\bar{x}$ |
| Variance $\sigma^2$ | Sample variance $s^2$ |
| Standard deviation $\sigma$ | Sample standard deviation $s$ |
| Correlation $\rho$ | Correlation coefficient $r$ |

# Inferential statistics

## Sampling distributions

- Sample statistics are random variables because they are based on random samples
- Therefore, sample statistics have their own probability distributions called **sampling distributions**
  - *E.g. Let the height of Germans in cm be a normally distributed random variable **X** with mean **μ** and variance **σ²**. Let $\bar{X}$ be the mean height in a sample of size **n**.*
  - $\bar{X}$ is a random variable with the following distribution:

$$\bar{X} \sim \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right)$$



# Inferential statistics

## Hypothesis tests

- Researchers are often interested in yes/no questions regarding population parameters
- E.g. A new medicine to prevent cold was tested against a control treatment. In a small pilot experiment, 20 subjects were randomized to the control and 20 to the new treatment. 11 subjects in the control group got a cold compared to 7 subjects in the treatment group. *Does the new treatment reduce the chance of getting a cold?*

| | **Control** | **Treatment** |
|---|---|---|
| Got a cold | 11 | 7 |
| Remained healthy | 9 | 13 |

JG|U

# Inferential statistics

## Hypothesis tests

- Results based on sample statistics initially only apply to the investigated sample
- Hypothesis tests are needed to decide if the observed results are due to population differences or simply the result of chance.
- The basis of hypothesis test is establishing the **hypothesis pair:**
  - **Null hypothesis:** Represents the status quo, absence of a difference – usually the hypothesis we want to disprove

    *$H_0$: The new therapy does not reduce the risk of getting a cold.*

  - **Alternative hypothesis:** Represents the alternative to $H_0$ – usually represents the desired outcome

    *$H_1$: The new therapy  reduced the risk of getting a cold.*

# Inferential statistics

## Hypothesis tests

- The null and alternative hypothesis make assumptions about an **unknown** population parameter **Θ** (e.g. population mean **μ** or population proportion **p**)
- Types of hypothesis pairs:
  - $H_0$: $\Theta = \Theta_0$ vs. $H_1$: $\Theta \neq \Theta_0$: **two-tailed** test
    - *E.g. Expression levels of gene X under condition A are different compared to condition B*
  - $H_0$: $\Theta = \Theta_0$ vs. $H_1$: $\Theta < \Theta_0$: **left-tailed** test
    - *E.g. Therapy A reduces blood pressure compared to therapy B*
  - $H_0$: $\Theta = \Theta_0$ vs. $H_1$: $\Theta > \Theta_0$: **right-tailed** test
    - *E.g. Therapy A increases survival time in cancer patients compared to control.*

## Hypothesis tests

- At the end of a specific test we always retain one of the hypotheses and reject the other. However, we never know if this decision is correct.
- What we do instead is try to reduce the probability of a false decision
- The following possibilities exist:

| | Decision for a test | |
|---|---|---|
| **Reality** | Reject $H_0$ | Reject $H_1$ |
| $H_0$ is true | Type I error ($\alpha$) | Correct decision |
| $H_1$ is true | Correct decision | Type II error ($\beta$) |

# Inferential statistics

## Hypothesis tests

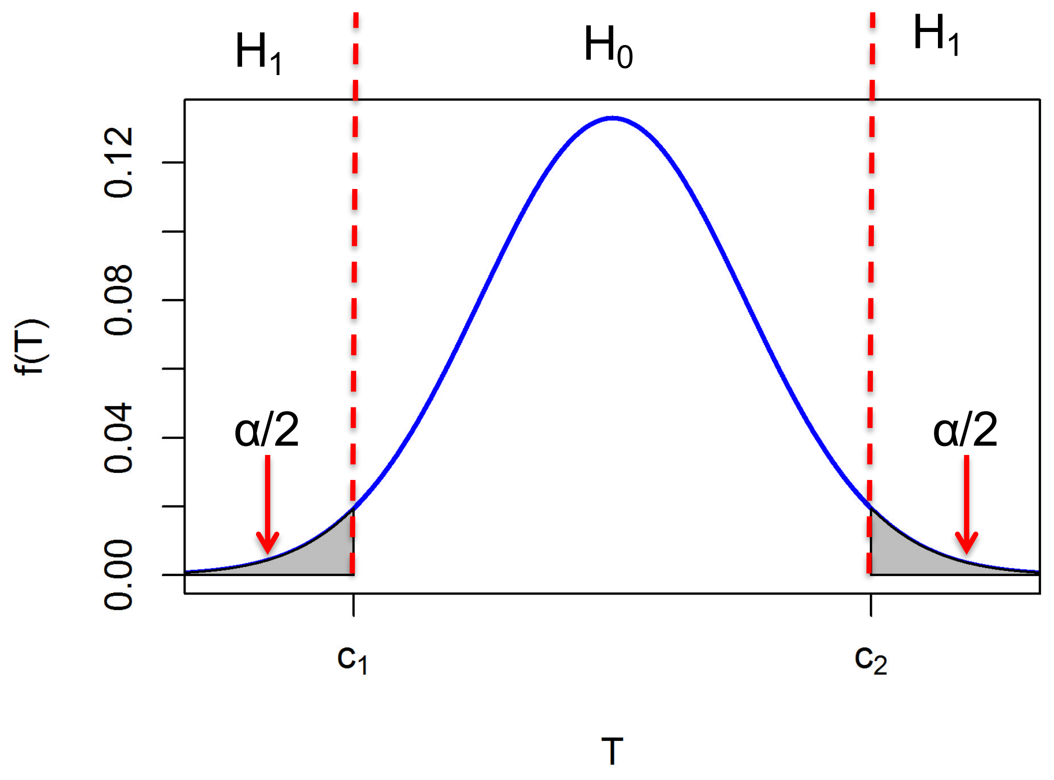
- Researchers typically control the type I error probability, each statistical test is performed at a predefined **significance level α.**
- Typical values include α = 0.05, α = 0.01, α = 0.001
- In order to decide which hypothesis to retain, a **test statistic T** is calculated.
- **T** is a function of the sample statistic, e.g. the sample mean $\bar{X}$
  - Remember that hypothesis are defined for the unknown population parameters
- **T** is therefore a random variable with a probability distribution
- In order to decide which hypothesis to retain/reject, the distribution of **T** under the **null hypothesis** is evaluated.
- The **null hypothesis** is rejected if **T** exceeds critical values which depend on **α.**

# Inferential statistics

## Hypothesis tests

Distribution of the test statistic **T,** two-tailed test

- Reject $H_0$ if $T < c_1$ or if $T > c_2$
- $c_1$ and $c_2$ are the so called critical values

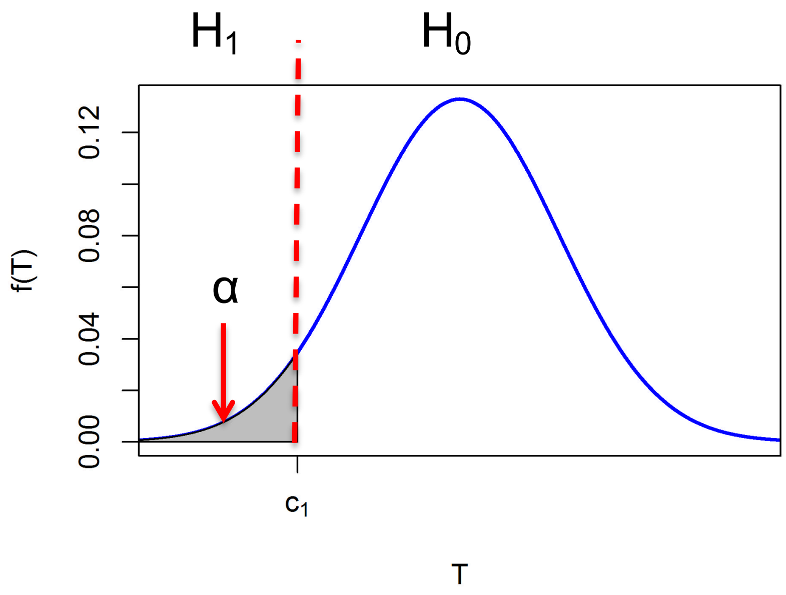


# Inferential statistics

## Hypothesis tests

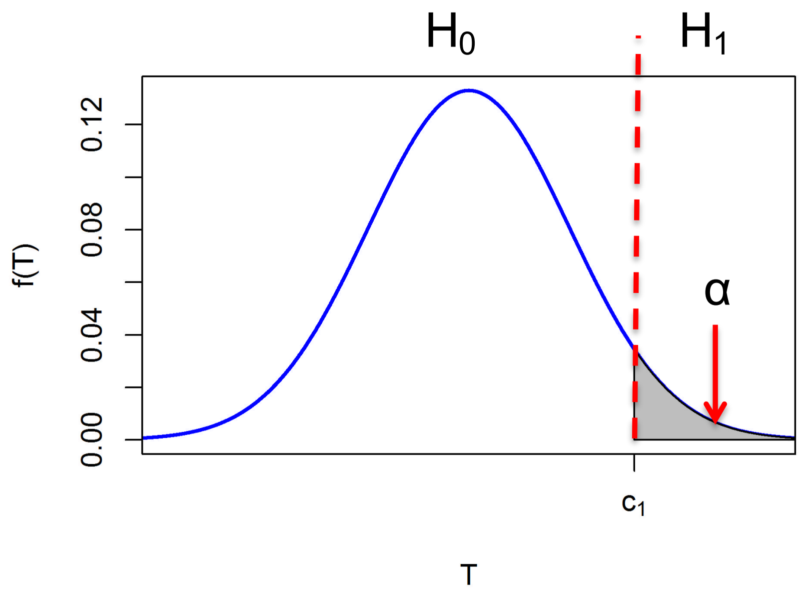
Distribution of the test statistic **T,** left- and right- tailed tests

Left-tailed test




Reject $H_0$ if $T < c_1$

Right-tailed test




Reject $H_0$ if $T > c_1$

# Hypothesis tests

## Steps in hypothesis testing

1. Evaluation of test assumptions
2. Definition of the null and alternative hypothesis
3. Definition of significance level
4. Calculation of the test statistic
5. Definition of the rejection region based on the probability distribution of the test statistic, critical values specification
6. Comparison of the test statistic with critical values
7. Decision to retain/reject null hypothesis and interpretation

# Hypothesis tests

## One sample z test

- The one sample z-test investigates if a population mean $\mu$ significantly deviates from a given value $\mu_0$ when the population variance is known.

*E.g. The average note in the final biostatistics exam of biomedicine students at the University Mainz from the last 5 years was 2.2 with a known variance of 0.9. The average note of the 49 students who took the exam this year was 1.8. Does the result for this year's exam represent a systematic change or only random fluctuation?*

$\longrightarrow$ Investigate hypothesis with the help of the one sample z-test using the 7-point scheme outlined on the previous slide

# Hypothesis tests

## One sample z test example

*The average note in the final biostatistics exam of biomedicine students at the University Mainz from the last 5 years was 2.2 with a known variance of 0.9. The average note of the 49 students who took the exam this year was 1.8. Does the result for this year's exam represent a systematic change or only random fluctuation?*

1. Evaluation of test assumptions
   - Normally distributed random variable and known population variance
2. Definition of the null and alternative hypothesis

- Null hypothesis: The average note of biomedicine students in statistics is 2.2

  $H_0$: $\mu = 2.2$

- Alternative: The average note of biomedicine students in statistics is not 2.2.

  $H_1$: $\mu \neq 2.2$ → two-sided test

# Hypothesis tests

## One sample z test example

*The average note in the final biostatistics exam of biomedicine students at the University Mainz from the last 5 years was 2.2 with a known variance of 0.9. The average note of the 49 students who took the exam this year was 1.8. Does the result for this year's exam represent a systematic change or only random fluctuation?*

3. Definition of significance level
   - α = 0.05 (type I error = 5 %)
4. Calculation of the test statistic

$$T = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}$$

$\bar{X}$ = 1.8, $\mu_0$ = 2.2, $\sigma$ = $\sqrt{0.9}$ , n = 49

$$T = \frac{1.8 - 2.2}{\sqrt{0.9}}\sqrt{49} = -2.951$$

# Hypothesis tests

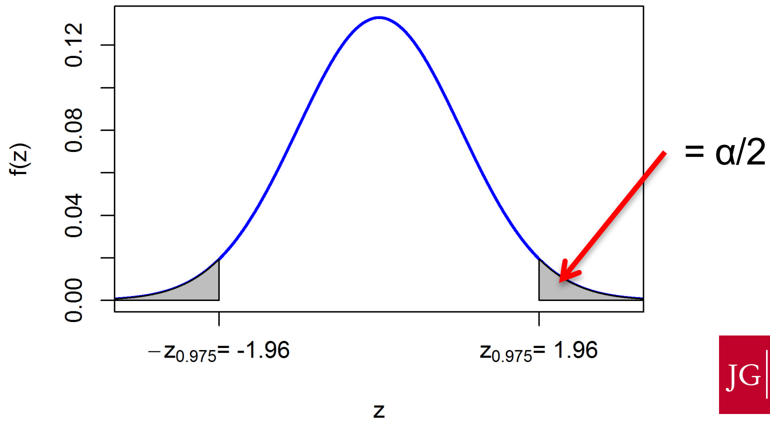
## One sample z test example

*The average note in the final biostatistics exam of biomedicine students at the University Mainz from the last 5 years was 2.2 with a known variance of 0.9. The average note of the 49 students who took the exam this year was 1.8. Does the result for this year's exam represent a systematic change or only random fluctuation?*

5. Definition of the rejection region based on the probability distribution of the test statistic, critical values specification

- When assumptions of the test are met, the test statistic follows a standard normal distribution, $T \sim \mathcal{N}(0,1)$.
- Critical values for a two sided test:

$$\pm z_{1-\alpha/2} = \pm z_{1-0.025} = \pm z_{0.975} = \pm 1.96$$



# Hypothesis tests

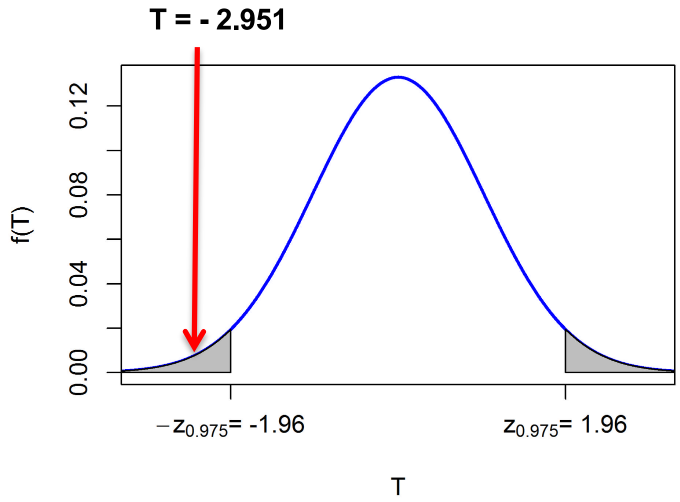
## One sample z test example

*The average note in the final biostatistics exam of biomedicine students at the University Mainz from the last 5 years was 2.2 with a known variance of 0.9. The average note of the 49 students who took the exam this year was 1.8. Does the result for this year's exam represent a systematic change or only random fluctuation?*

6. Comparison of the test statistic with critical values

$T = -2.951 < -z_{0.975} = -1.96$

- Test statistic exceeds the critical value



# Hypothesis tests

## One sample z test example

*The average note in the final biostatistics exam of biomedicine students at the University Mainz from the last 5 years was 2.2 with a known variance of 0.9. The average note of the 49 students who took the exam this year was 1.8. Does the result for this year‘s exam represent a systematic change or only random fluctuation?*

7. Decision to retain/reject null hypothesis and interpretation

- Null hypothesis is rejected
- Interpretation:

  The average note of biomedicine students in statistics this year is significantly different from the 5 year-average.

# Hypothesis tests

## Power of a test

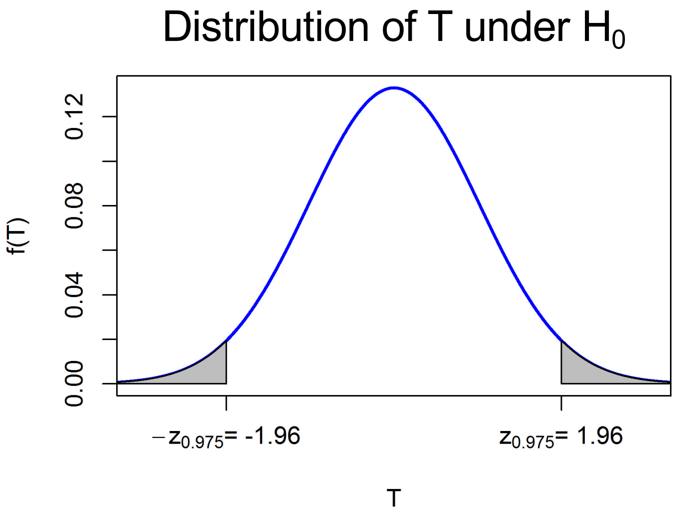
- Usually, the desired outcome of a hypothesis test is to find sufficient empirical evidence to reject the null hypothesis.
- Even if $H_1$ is true, we still might fail to reject $H_0$ (Type II error: false negative)
- Type II errors can have serious consequences especially in the field of clinical research
- Power = $1 - \beta$, probability to reject a false $H_0$
- Ways to increase power of a test:
  - Increase the significance level
  - Increase sample size

# Hypothesis tests

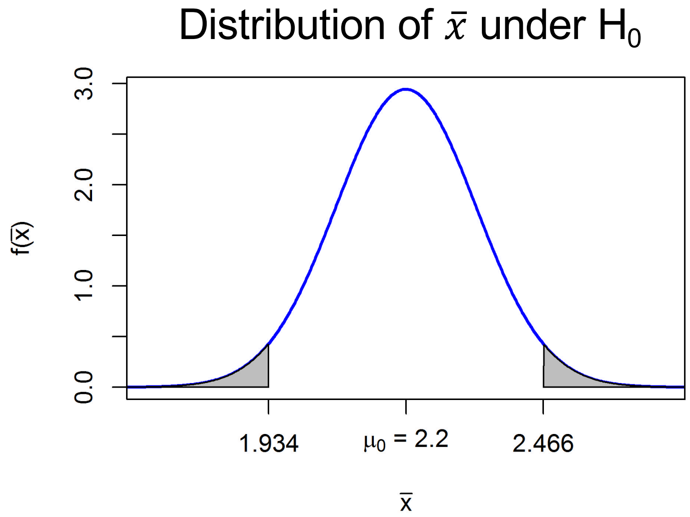
## Power of a test

- Power can only be calculated for specific values of the parameters under the alternative hypothesis.
- *The average note in the final biostatistics exam of biomedicine students at the University Mainz from the last 5 years was 2.2 with a known variance of 0.9. The average note of the 49 students who took the exam this year was 1.8. What is the power of a two-sided z-test given the alternative hypothesis that $\mu_1$ = 1.9?*

$$\text{Power} = P(T < -z_{0.975} \text{ or } T > z_{0.975} \mid \mu_1 = 1.9) = ?$$



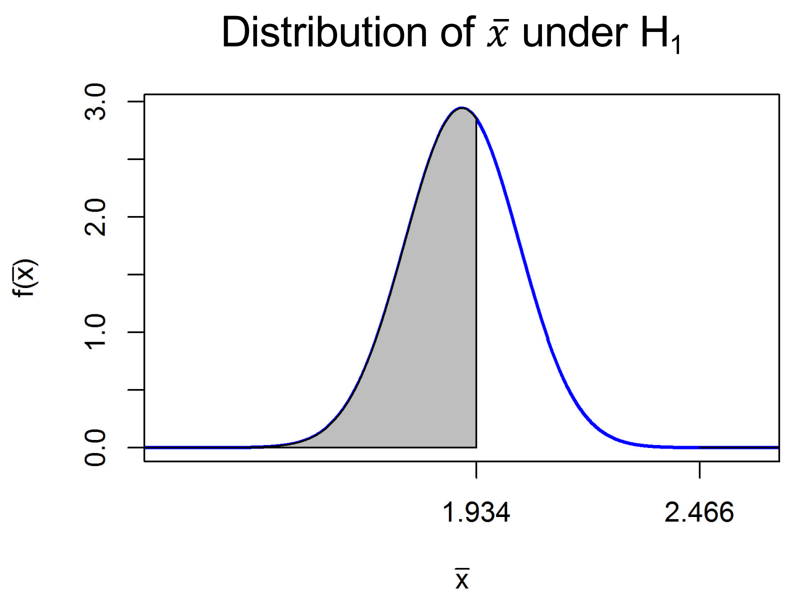

$$\bar{x}_{crit} = \pm \frac{z_{0.975}\sigma}{\sqrt{n}} + \mu_0$$



# Hypothesis tests

## Power of a test

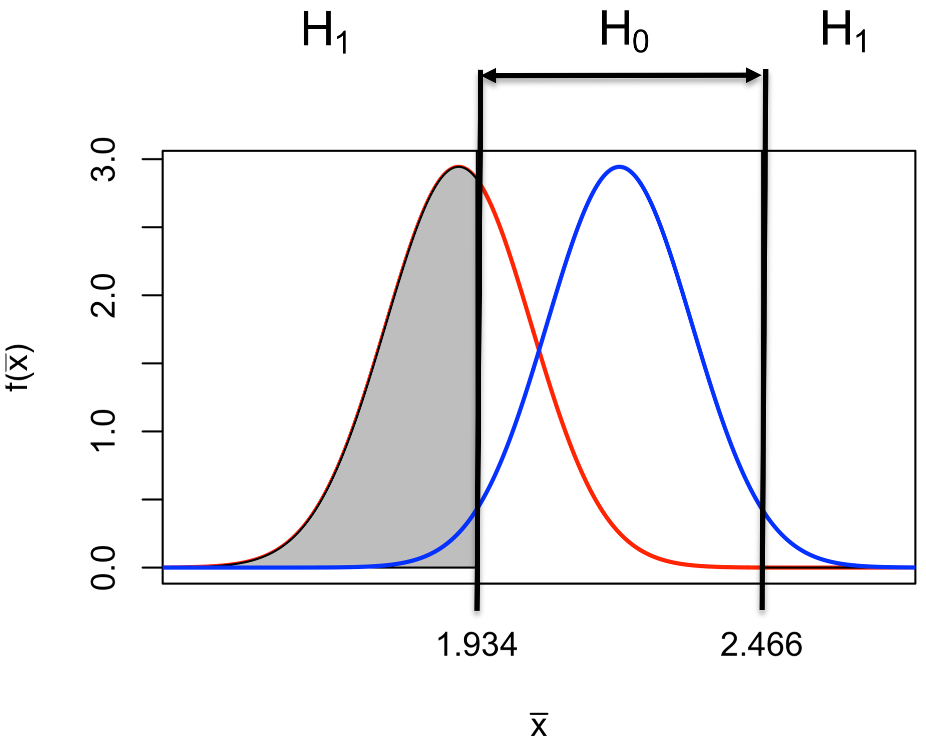
- Reject $H_0$ if $\bar{x} < 1.934$ or $\bar{x} > 2.466$
- Power = P($\bar{x} < 1.934$ or $\bar{x} > 2.466$ | $\mu_1 = 1.9$) = 0.6



# Hypothesis tests

## Power of a test

- Reject $H_0$ if $\bar{x} < 1.934$ or $\bar{x} > 2.466$
- Power = $P(\bar{x} < 1.934 \text{ or } \bar{x} > 2.466 \mid \mu_1 = 1.9) = 0.599$



# Hypothesis tests

## Effect size

- Rejecting a null hypothesis does not automatically imply that results have practical relevance
- In large samples, even trivial differences might be statistically significant
- This calls for standard measures to evaluate effect size, e.g.:
  - Standardized mean difference
  - Percentage of variance explained by the model
  - Odds ratio of disease in the presence of a risk factor
  - Fold change of expression levels of a gene
- Effect sizes should be reported with results from significance tests

# Hypothesis tests

## Effect size example

- Cohen‘s *d* is a measure of effect size evaluating standardized mean difference
  - $$Cohen's\ d = \frac{|\mu_1 - \mu_2|}{\sigma}$$
- Reference values:
  - d = 0.2 – small effect
  - d = 0.5 – medium effect
  - d > 0.8 – large effect

- *E.g. The average note of biomedicine students from this year‘s exam* $\bar{x} = 1.8$ *was shown to be significantly different from the 5-year average of 2.2. What is the effect size?*

$$d = \frac{|1.8 - 2.2|}{\sqrt{0.9}} = 0.422$$

# Hypothesis tests

## Power and effect size in controlled experiments

- Controlled clinical experiments are associated with significant costs and time
- Failing to detect a true alternative hypothesis could be detrimental
- From a statistical point of view, studies are planned to maximize the chances of proving a study hypothesis which is also of practical relevance:
  - Sample size is based on the level of power desired to be achieved, typically 80%
  - Effect size is based on theoretical knowledge or pilot investigations

JG|U