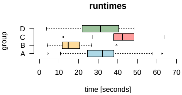
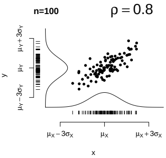
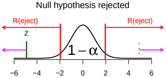
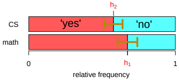
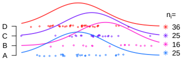
Revision

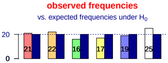










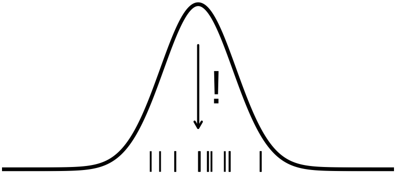


For comments please email: M. Messer, michael.messer@tuwien.ac.at

Overview

We differentiate:

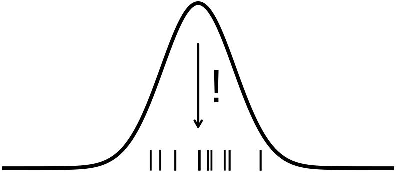
Probability theory
(Stochastics)
$=$
Theory of randomness



Overview

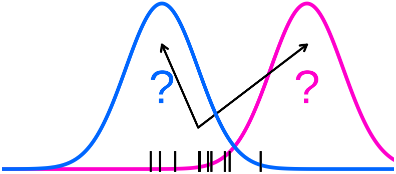
We differentiate:

Probability theory
(Stochastics)
=
Theory of randomness



and

Statistics
=
Description of data $\longrightarrow$
(using stochastic models)



Data collection

How long is the runtime of an algorithm that you implemented?

Data collection

How long is the runtime of an algorithm that you implemented?

$n = 121$ students requested (same technical setup)

Data collection

How long is the runtime of an algorithm that you implemented?

$n = 121$ students requested (same technical setup)

Results (in seconds):

24.6, 24, 31.4, 29.9, 37.8, 19.9, 46.1, 32.8, 30.3, 29, 47.1, 27.8, 33.8, 30.1, 53.3, 23.8, 32.1, 4.2, 42.8, 25.2, 52.3, 35, 30.1, 43.2, 25.4, 62.5, 35.4, 25.2, 37.6, 37.1, 22.9, 29.5, 44.5, 34.8, 33.3, 21.9, 37.2, 24, 37, 34, 24.1, 10.8, 24.9, 37.2, 52, 30.8, 22, 18.6, 22, 26.8, 52.3, 27, 23.6, 33.5, 30.8, 20.9, 35.6, 37.2, 57.5, 46.2, 36.1, 19.8, 38.1, 36.9, 26.5, 23.6, 30.3, 49.9, 39, 50.2, 35.7, 11.4, 24.1, 27.5, 36.4, 29.8, 49, 42.6, 22.5, 32.7, 34.3, 21.4, 34.7, 47.3, 20.3, 35.4, 41.8, 24.9, 15.2, 42.2, 29.1, 25.1, 22.7, 41, 28.2, 30.3, 25.6, 41.8, 16.6, 38, 43.1, 29.5, 40.3, 20.5, 39.9, 24.5, 33.7, 14.6, 23.3, 36.7, 34.7, 34.9, 39.1, 32.2, 43, 12.1, 19.8, 27.4, 39.3, 35, 46.3

Data collection

How long is the runtime of an algorithm that you implemented?

$n = 121$ students requested (same technical setup)

Results (in seconds):

24.6, 24, 31.4, 29.9, 37.8, 19.9, 46.1, 32.8, 30.3, 29, 47.1, 27.8, 33.8, 30.1, 53.3, 23.8, 32.1, 4.2, 42.8, 25.2, 52.3, 35, 30.1, 43.2, 25.4, 62.5, 35.4, 25.2, 37.6, 37.1, 22.9, 29.5, 44.5, 34.8, 33.3, 21.9, 37.2, 24, 37, 34, 24.1, 10.8, 24.9, 37.2, 52, 30.8, 22, 18.6, 22, 26.8, 52.3, 27, 23.6, 33.5, 30.8, 20.9, 35.6, 37.2, 57.5, 46.2, 36.1, 19.8, 38.1, 36.9, 26.5, 23.6, 30.3, 49.9, 39, 50.2, 35.7, 11.4, 24.1, 27.5, 36.4, 29.8, 49, 42.6, 22.5, 32.7, 34.3, 21.4, 34.7, 47.3, 20.3, 35.4, 41.8, 24.9, 15.2, 42.2, 29.1, 25.1, 22.7, 41, 28.2, 30.3, 25.6, 41.8, 16.6, 38, 43.1, 29.5, 40.3, 20.5, 39.9, 24.5, 33.7, 14.6, 23.3, 36.7, 34.7, 34.9, 39.1, 32.2, 43, 12.1, 19.8, 27.4, 39.3, 35, 46.3

We see: n data: $x_1 = 24.6, x_2 = 24.0, \ldots, x_n = 46.3$

Data collection

How long is the runtime of an algorithm that you implemented?

$n = 121$ students requested (same technical setup)

Results (in seconds):

24.6, 24, 31.4, 29.9, 37.8, 19.9, 46.1, 32.8, 30.3, 29, 47.1, 27.8, 33.8, 30.1, 53.3, 23.8, 32.1, 4.2, 42.8, 25.2, 52.3, 35, 30.1, 43.2, 25.4, 62.5, 35.4, 25.2, 37.6, 37.1, 22.9, 29.5, 44.5, 34.8, 33.3, 21.9, 37.2, 24, 37, 34, 24.1, 10.8, 24.9, 37.2, 52, 30.8, 22, 18.6, 22, 26.8, 52.3, 27, 23.6, 33.5, 30.8, 20.9, 35.6, 37.2, 57.5, 46.2, 36.1, 19.8, 38.1, 36.9, 26.5, 23.6, 30.3, 49.9, 39, 50.2, 35.7, 11.4, 24.1, 27.5, 36.4, 29.8, 49, 42.6, 22.5, 32.7, 34.3, 21.4, 34.7, 47.3, 20.3, 35.4, 41.8, 24.9, 15.2, 42.2, 29.1, 25.1, 22.7, 41, 28.2, 30.3, 25.6, 41.8, 16.6, 38, 43.1, 29.5, 40.3, 20.5, 39.9, 24.5, 33.7, 14.6, 23.3, 36.7, 34.7, 34.9, 39.1, 32.2, 43, 12.1, 19.8, 27.4, 39.3, 35, 46.3

We see: n data: $x_1 = 24.6$, $x_2 = 24.0, \ldots, x_n = 46.3$

We understand: nothing?

Data collection

How long is the runtime of an algorithm that you implemented?

$n = 121$ students requested (same technical setup)

Results (in seconds):

24.6, 24, 31.4, 29.9, 37.8, 19.9, 46.1, 32.8, 30.3, 29, 47.1, 27.8, 33.8, 30.1, 53.3, 23.8, 32.1, 4.2, 42.8, 25.2, 52.3, 35, 30.1, 43.2, 25.4, 62.5, 35.4, 25.2, 37.6, 37.1, 22.9, 29.5, 44.5, 34.8, 33.3, 21.9, 37.2, 24, 37, 34, 24.1, 10.8, 24.9, 37.2, 52, 30.8, 22, 18.6, 22, 26.8, 52.3, 27, 23.6, 33.5, 30.8, 20.9, 35.6, 37.2, 57.5, 46.2, 36.1, 19.8, 38.1, 36.9, 26.5, 23.6, 30.3, 49.9, 39, 50.2, 35.7, 11.4, 24.1, 27.5, 36.4, 29.8, 49, 42.6, 22.5, 32.7, 34.3, 21.4, 34.7, 47.3, 20.3, 35.4, 41.8, 24.9, 15.2, 42.2, 29.1, 25.1, 22.7, 41, 28.2, 30.3, 25.6, 41.8, 16.6, 38, 43.1, 29.5, 40.3, 20.5, 39.9, 24.5, 33.7, 14.6, 23.3, 36.7, 34.7, 34.9, 39.1, 32.2, 43, 12.1, 19.8, 27.4, 39.3, 35, 46.3

We see: n data: $x_1 = 24.6, x_2 = 24.0, \ldots, x_n = 46.3$

We understand: nothing?

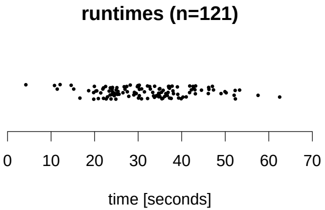
Thus: descriptive Statistics $\rightarrow$ graphical representation and summary of data

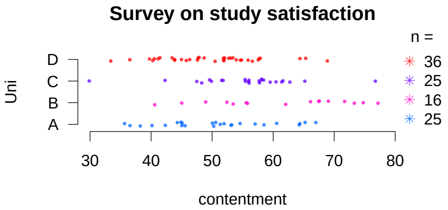
Descriptive statistics

- Graphical representations
- Summarizing statistics

Graphical representations

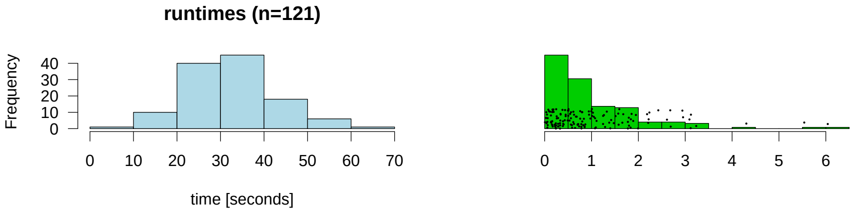
- stripchart





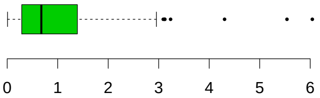
Graphical representations

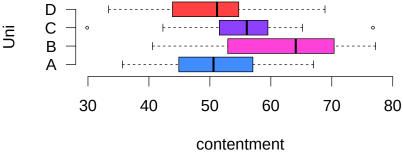
- stripchart
- histogram



Graphical representations

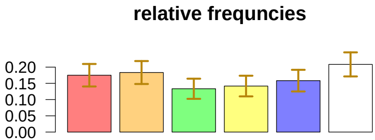
- stripchart
- histogram
- boxplot

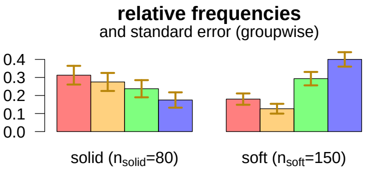




Graphical representations

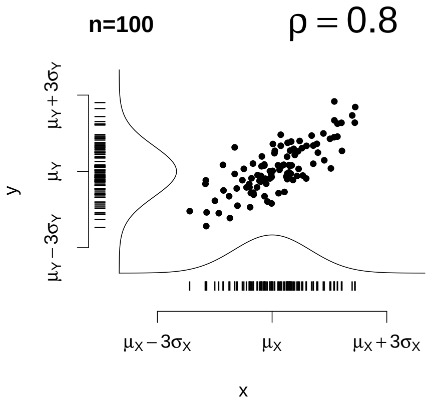
- stripchart
- histogram
- boxplot
- barplot

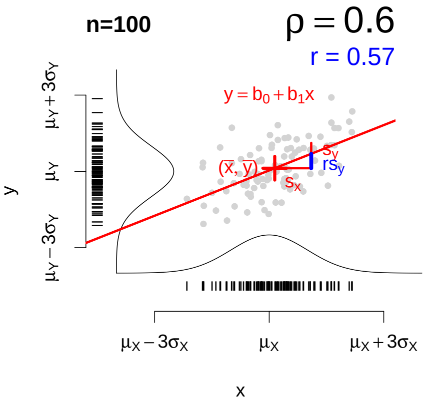




Graphical representations

- stripchart
- histogram
- boxplot
- barplot
- plot



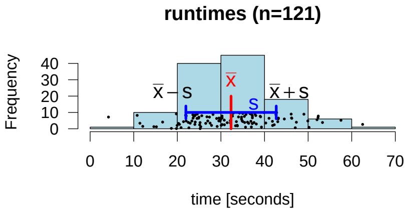


Graphical representations

- stripchart
- histogram
- boxplot
- barplot
- plot
- ...?

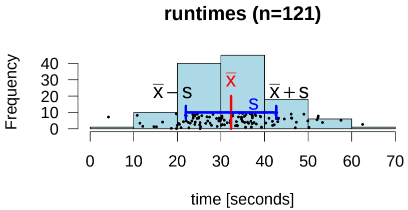
Summary statistics

- mean $\bar{x}$



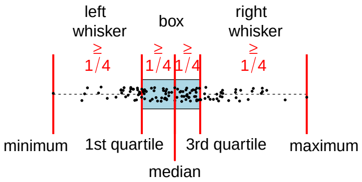
Summary statistics

- mean $\bar{x}$
- empirical standard deviation s (and emp. variance s^2)



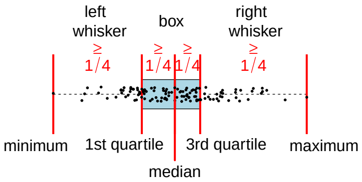
Summary statistics

- mean $\bar{x}$
- empirical standard deviation s (and emp. variance s^2)
- minimum $x_{(1)}$ and maximum $x_{(n)}$



Summary statistics

- mean $\bar{x}$
- empirical standard deviation s (and emp. variance s^2)
- minimum $x_{(1)}$ and maximum $x_{(n)}$
- empirical quantiles q_p (e.g., median, quartiles)



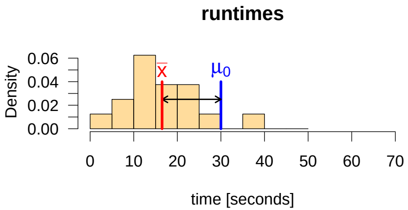
Summary statistics

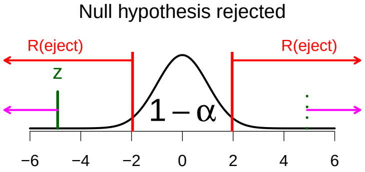
- mean $\bar{x}$
- empirical standard deviation s (and emp. variance s^2)
- minimum $x_{(1)}$ and maximum $x_{(n)}$
- empirical quantiles q_p (e.g., median, quartiles)
- ...?

For details see the lecture: 'Descriptive Statistics'

Statistical hypothesis test

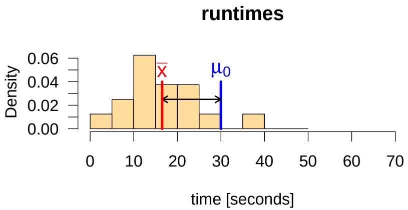
- How compatible are the data with an assertion?

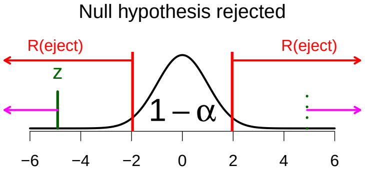




Statistical hypothesis test

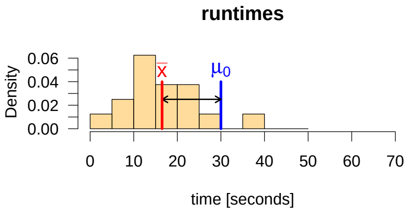
- How compatible are the data with an assertion?
- Notion of incompatibility via probability statements within statistical models ('data interpreted as realizations of random variables')

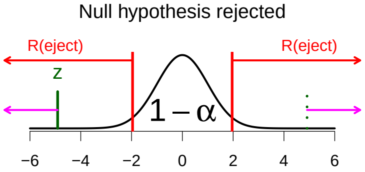




Statistical hypothesis test

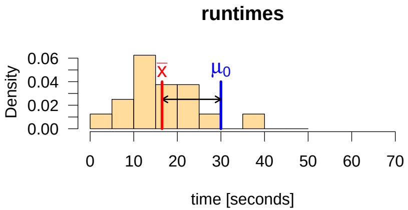
- How compatible are the data with an assertion?
- Notion of incompatibility via probability statements within statistical models ('data interpreted as realizations of random variables')
- Formalized in hypothesis tests (or confidence intervals)

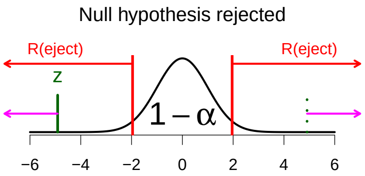




Statistical hypothesis test

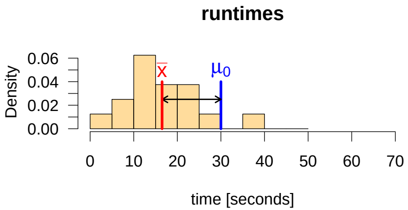
- How compatible are the data with an assertion?
- Notion of incompatibility via probability statements within statistical models ('data interpreted as realizations of random variables')
- Formalized in hypothesis tests (or confidence intervals)
- Need additional statistics
  - test statistics (t-statistic, f-statistic, χ^2-statistic etc.)
  - auxiliary statistics (e.g., standard errors)
  - p-value

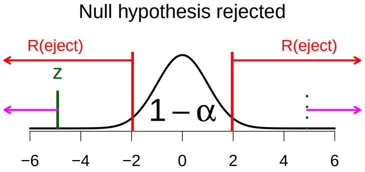




Statistical hypothesis test

- How compatible are the data with an assertion?
- Notion of incompatibility via probability statements within statistical models ('data interpreted as realizations of random variables')
- Formalized in hypothesis tests (or confidence intervals)
- Need additional statistics
  - test statistics (t-statistic, f-statistic, χ^2-statistic etc.)
  - auxiliary statistics (e.g., standard errors)
  - p-value





For general concepts of modeling and testing see the lecture: 'Basic ideas of hypothesis testing'

Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \ldots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$) (The data $x_1, \ldots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ <u>and unknown variance σ^2</u>)

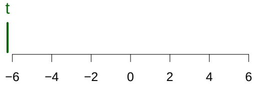
Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \ldots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \ldots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ <u>and unknown variance σ^2</u>)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)

Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \ldots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \ldots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ and unknown variance σ^2)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)
- *Test statistic* for the evaluation of the data (measures discrepancy). Now t-statistic

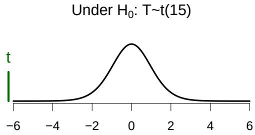
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$$



Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \ldots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \ldots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ and unknown variance σ^2)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)
- *Test statistic* for the evaluation of the data (measures discrepancy). Now t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$$



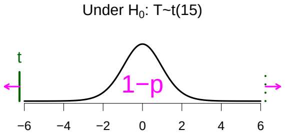
- Distribution of the test statistic if H_0 holds true

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(15)$$

Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \ldots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \ldots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ <u>and unknown variance σ^2</u>)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)
- *Test statistic* for the evaluation of the data (measures discrepancy). Now t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$$



- Distribution of the test statistic if H_0 holds true

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(15)$$

- *p-value:* quantifies discrepancy (judge t according to the distribution of T)

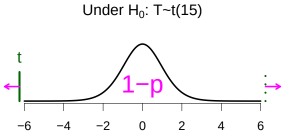
$$p = \mathbb{P}_{H_0}(|T| \geqslant |t|) \approx 1.5 \cdot 10^{-5}$$

  Probability to make an observation which is at least as extreme as in the data, if H_0 holds true (here: two-sided test)

Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \dots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \dots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ <u>and unknown variance σ^2</u>)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)
- *Test statistic* for the evaluation of the data (measures discrepancy). Now t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$$



- Distribution of the test statistic if H_0 holds true

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(15)$$

- *p-value:* quantifies discrepancy (judge t according to the distribution of T)

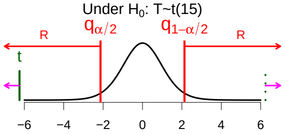
$$p = \mathbb{P}_{H_0}(|T| \geqslant |t|) \approx 1.5 \cdot 10^{-5}$$

  Probability to make an observation which is at least as extreme as in the data, if H_0 holds true (here: two-sided test)
- *Decision:* Reject the null hypothesis, because $p \leqslant \alpha$
  Say: the observed discrepancy was *significant* ($p < 10^{-4}$)

Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \ldots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \ldots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ <u>and unknown variance σ^2</u>)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)
- *Test statistic* for the evaluation of the data (measures discrepancy). Now t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$$



- Distribution of the test statistic if H_0 holds true

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(15)$$

- *p-value:* quantifies discrepancy (judge t according to the distribution of T)

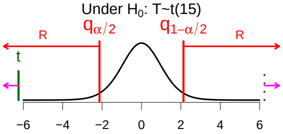
$$p = \mathbb{P}_{H_0}(|T| \geqslant |t|) \approx 1.5 \cdot 10^{-5}$$

  Probability to make an observation which is at least as extreme as in the data, if H_0 holds true (here: two-sided test)
- *Decision:* Reject the null hypothesis, because $p \leqslant \alpha \Leftrightarrow t \in R$
  Say: the observed discrepancy was *significant* ($p < 10^{-4}$)

Example: the (one-sample) t-Test

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \dots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \dots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ <u>and unknown variance σ^2</u>)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)
- *Test statistic* for the evaluation of the data (measures discrepancy). Now t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$$



- Distribution of the test statistic if H_0 holds true

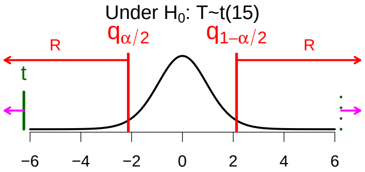
$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(15)$$

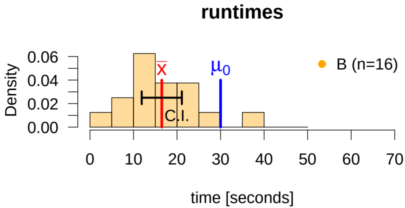
- *p-value:* quantifies discrepancy (judge t according to the distribution of T)

$$p = \mathbb{P}_{H_0}(|T| \geqslant |t|) \approx 1.5 \cdot 10^{-5}$$

  Probability to make an observation which is at least as extreme as in the data, if H_0 holds true (here: two-sided test)
- *Decision:* Reject the null hypothesis, because $p \leqslant \alpha \Leftrightarrow t \in R$
  Say: the observed discrepancy was *significant* ($p < 10^{-4}$)
- *Interpretation*: If H_0 holds true, then something very unlikely was observed. In that sense, the data are hardly compatible with H_0.

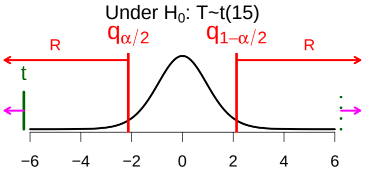
Equivalence of test and confidence interval

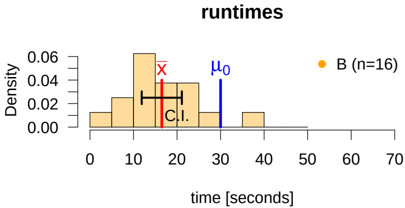




Reject H_0 if and only if μ_0 is not overlapped by the confidence interval
$i := (\bar{x} - q_{1-\alpha/2} \cdot sem, \bar{x} + q_{1-\alpha/2} \cdot sem)$ with $sem := s/\sqrt{n}$

Equivalence of test and confidence interval





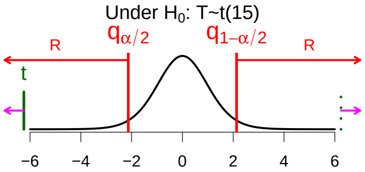
Reject H_0 if and only if μ_0 is not overlapped by the confidence interval
$i := (\bar{x} - q_{1-\alpha/2} \cdot sem, \bar{x} + q_{1-\alpha/2} \cdot sem)$ with $sem := s/\sqrt{n}$

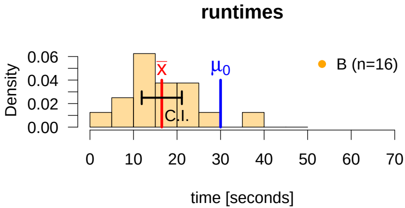
Recall that the upper test and confidence interval are referred to as 'Student's' versions (one-sample situation).
But also remember that the structure of the test statistic and the confidence interval is inherited to other situations, i.e., 'more general'

$$T = \frac{\spadesuit - \spadesuit}{\spadesuit} \qquad \text{and} \qquad I = (\spadesuit - q \cdot \spadesuit, \spadesuit + q \cdot \spadesuit)$$

while ♠ denotes some summary statistic, ♠ denotes the null-parameter, and ♠ a standard-error of the statistic, see e.g., the two-sample situation.

Equivalence of test and confidence interval





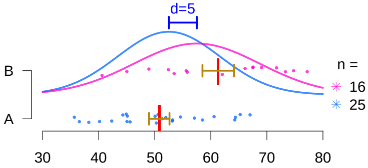
Reject H_0 if and only if μ_0 is not overlapped by the confidence interval
$i := (\bar{x} - q_{1-\alpha/2} \cdot sem, \bar{x} + q_{1-\alpha/2} \cdot sem)$ with $sem := s/\sqrt{n}$

Recall that the upper test and confidence interval are referred to as 'Student's' versions (one-sample situation).
But also remember that the structure of the test statistic and the confidence interval is inherited to other situations, i.e., 'more general'

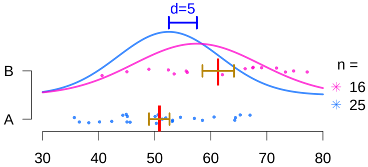
$$T = \frac{\spadesuit - \spadesuit}{\spadesuit} \qquad \text{and} \qquad I = (\spadesuit - q \cdot \spadesuit, \spadesuit + q \cdot \spadesuit)$$

while ♠ denotes some summary statistic, ♠ denotes the null-parameter, and ♠ a standard-error of the statistic, see e.g., the two-sample situation.
For details see the lecture: 'Surrounding the one-sample t-test'

Two-sample *t*-test

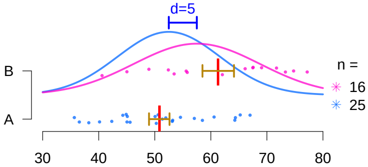


Two-sample t-test



- Let $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$ be independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \dots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \dots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
- $d = \mu_2 - \mu_1$
- Let $q_{1-\alpha/2}$ be the $(1 - \alpha/2)$-quantile of the $t(\nu)$-distribution (R knows ν)

Two-sample t-test

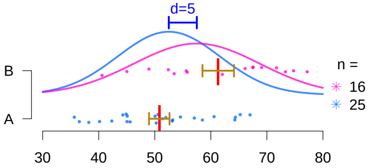


- Let $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ be independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
- $d = \mu_2 - \mu_1$
- Let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $t(\nu)$-distribution (R knows ν)

Under $H_0 : d = d_0$ it holds (approx)

$$T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$$

Two-sample t-test



- Let $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ be independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
- $d = \mu_2 - \mu_1$
- Let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $t(\nu)$-distribution (R knows ν)

Under $H_0 : d = d_0$ it holds (approx)

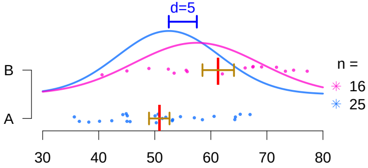
$$T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$$

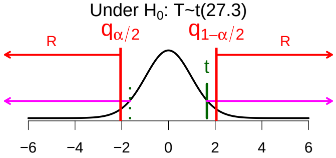
and equivalently: The confidence interval

$$I := \left((\bar{Y} - \bar{X}) - q_{1-\alpha/2} \cdot \sqrt{SEM_y^2 + SEM_x^2}, \ (\bar{Y} - \bar{X}) + q_{1-\alpha/2} \cdot \sqrt{SEM_y^2 + SEM_x^2} \right)$$

overlaps the parameter d_0 with probability (approx) $1 - \alpha$

Two-sample t-test





- Let $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ be independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
- $d = \mu_2 - \mu_1$
- Let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $t(\nu)$-distribution (R knows ν)

Under $H_0 : d = d_0$ it holds (approx)

$$T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$$

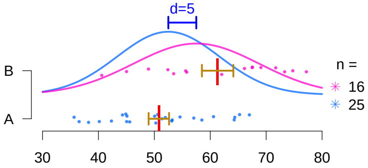
and equivalently: The confidence interval

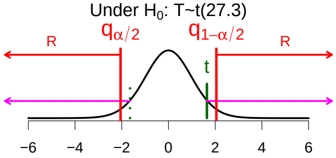
$$I := \left((\bar{Y} - \bar{X}) - q_{1-\alpha/2} \cdot \sqrt{SEM_y^2 + SEM_x^2},\ (\bar{Y} - \bar{X}) + q_{1-\alpha/2} \cdot \sqrt{SEM_y^2 + SEM_x^2} \right)$$

overlaps the parameter d_0 with probability (approx) $1 - \alpha$

- Under $H_0 : d = 5$ ($\Leftrightarrow \mu_2 = \mu_1 + 5$) $\rightarrow$ here: can <u>not</u> reject H_0 ($\rightarrow$ Welch-test)

Two-sample t-test





- Let $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ be independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
- $d = \mu_2 - \mu_1$
- Let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $t(\nu)$-distribution (R knows ν)

Under $H_0 : d = d_0$ it holds (approx)

$$T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$$

and equivalently: The confidence interval

$$I := \left((\bar{Y} - \bar{X}) - q_{1-\alpha/2} \cdot \sqrt{SEM_y^2 + SEM_x^2},\ (\bar{Y} - \bar{X}) + q_{1-\alpha/2} \cdot \sqrt{SEM_y^2 + SEM_x^2} \right)$$

overlaps the parameter d_0 with probability (approx) $1 - \alpha$

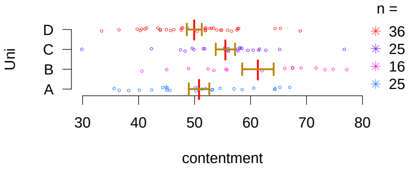
- Under $H_0 : d = 5$ ($\Leftrightarrow \mu_2 = \mu_1 + 5$) $\rightarrow$ here: can not reject H_0 ($\rightarrow$ Welch-test)

For details see the lecture: 'Surrounding the two-sample t-test'

ANOVA

Model:

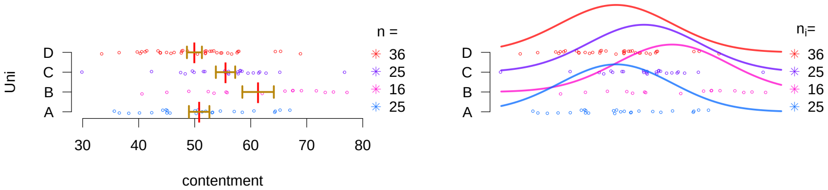
- Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$
- Thus, k groups and $n = \sum_{i=1}^{k} n_i$ observations in total



ANOVA

Model:

- Let $X_{1,1}, \ldots, X_{1,n_1},\ \ X_{2,1}, \ldots, X_{2,n_2},\ \ \ldots,\ \ X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$
- Thus, k groups and $n = \sum_{i=1}^{k} n_i$ observations in total

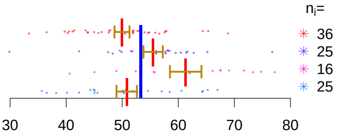


ANOVA

Model:

- Let $X_{1,1},\ldots,X_{1,n_1},\ \ X_{2,1},\ldots,X_{2,n_2},\ \ \ldots,\ \ X_{k,1},\ldots,X_{k,n_k}$ be independent RVs and for $i=1,\ldots,k$ let $X_{i,j}\sim N(\mu_i,\sigma^2)$ for $j=1,\ldots,n_i$, with $(\mu_1,\ldots,\mu_k,\sigma^2)\in\mathbb{R}^k\times\mathbb{R}^+$
- Thus, k groups and $n=\sum_{i=1}^k n_i$ observations in total

$$F := \frac{\frac{1}{k-1}\sum_{i=1}^k n_i(\bar{X}_{i,\cdot}-\bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j}-\bar{X}_{i,\cdot})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}$$



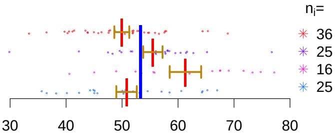
ANOVA

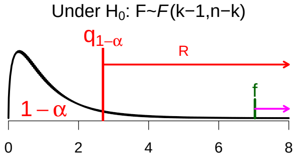
Model:

- Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$
- Thus, k groups and $n = \sum_{i=1}^{k} n_i$ observations in total

Under $H_0 : \mu_1 = \cdots = \mu_k$ it holds

$$F := \frac{\frac{1}{k-1} \sum_{i=1}^{k} n_i (\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k} \sum_{i,j} (X_{i,j} - \bar{X}_{i,\cdot})^2} \sim \mathscr{F}(k-1, n-k)$$





- Here: reject H_0.

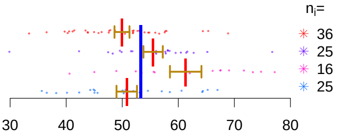
ANOVA

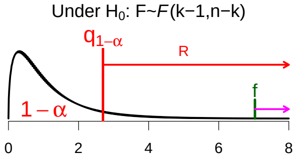
Model:

- Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$
- Thus, k groups and $n = \sum_{i=1}^k n_i$ observations in total

Under $H_0 : \mu_1 = \cdots = \mu_k$ it holds

$$F := \frac{\frac{1}{k-1} \sum_{i=1}^k n_i (\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k} \sum_{i,j} (X_{i,j} - \bar{X}_{i,\cdot})^2} \sim \mathscr{F}(k-1, n-k)$$



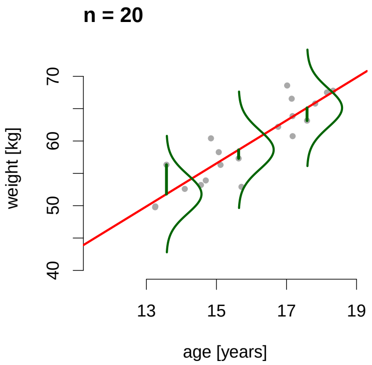


- Here: reject H_0.

For details on comparisons on multiple groups see the lecture: 'Analysis of variance and multiple testing'

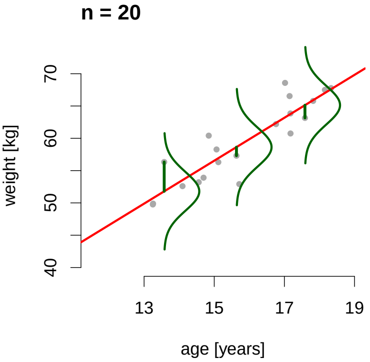
Linear regression

- For $i = 1, \ldots, n$ let
$$Y_i = \beta_0 + \beta_1 \cdot x_i + \sigma Z_i,$$
with $Z_1, \ldots, Z_n$ i.i.d. RVs and $Z_1 \sim N(0,1)$, and $(\beta_0, \beta_1, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$
- $q_{1-\alpha/2}$ the $(1-\alpha/2)$-quantile of the $t(n-2)$-distribution



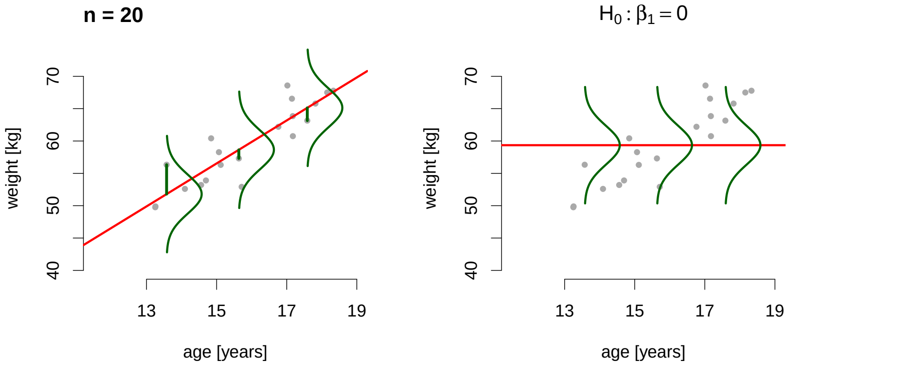
Linear regression

- For $i = 1, \ldots, n$ let
$$Y_i = \beta_0 + \beta_1 \cdot x_i + \sigma Z_i,$$
with $Z_1, \ldots, Z_n$ i.i.d. RVs and $Z_1 \sim N(0,1)$, and $(\beta_0, \beta_1, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$
- $q_{1-\alpha/2}$ the $(1-\alpha/2)$-quantile of the $t(n-2)$-distribution
- estimators B_0, B_1 and S_r via 'least-squares', $SE_{B_1} = S_r/(s_x \cdot \sqrt{n-1})$



Linear regression

- For $i = 1, \ldots, n$ let
$$Y_i = \beta_0 + \beta_1 \cdot x_i + \sigma Z_i,$$
with $Z_1, \ldots, Z_n$ i.i.d. RVs and $Z_1 \sim N(0,1)$, and $(\beta_0, \beta_1, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$
- $q_{1-\alpha/2}$ the $(1-\alpha/2)$-quantile of the $t(n-2)$-distribution
- estimators B_0, B_1 and S_r via 'least-squares', $SE_{B_1} = S_r/(s_x \cdot \sqrt{n-1})$



Linear regression

- For $i = 1, \ldots, n$ let
$$Y_i = \beta_0 + \beta_1 \cdot x_i + \sigma Z_i,$$
with $Z_1, \ldots, Z_n$ i.i.d. RVs and $Z_1 \sim N(0,1)$, and $(\beta_0, \beta_1, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$
- $q_{1-\alpha/2}$ the $(1-\alpha/2)$-quantile of the $t(n-2)$-distribution
- estimators B_0, B_1 and S_r via 'least-squares', $SE_{B_1} = S_r/(s_x \cdot \sqrt{n-1})$

Under $H_0 : \beta_1 = \beta_1^{(0)}$ it holds

$$T := \frac{B_1 - \beta_1^{(0)}}{SE_{B_1}} \sim t(n-2)$$

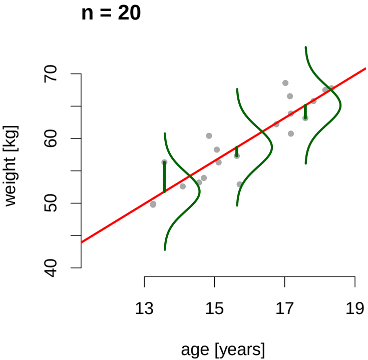
and equivalently: the confidence interval

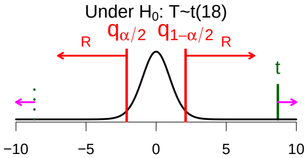
$$I := \left(B_1 - q_{1-\alpha/2} \cdot SE_{B_1},\ B_1 + q_{1-\alpha/2} \cdot SE_{B_1}\right)$$

overlaps $\beta_1^{(0)}$ with probability $1-\alpha$

Linear regression

- For $i = 1, \ldots, n$ let
$$Y_i = \beta_0 + \beta_1 \cdot x_i + \sigma Z_i,$$
with $Z_1, \ldots, Z_n$ i.i.d. RVs and $Z_1 \sim N(0,1)$, and $(\beta_0, \beta_1, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$
- $q_{1-\alpha/2}$ the $(1-\alpha/2)$-quantile of the $t(n-2)$-distribution
- estimators B_0, B_1 and S_r via 'least-squares', $SE_{B_1} = S_r/(s_x \cdot \sqrt{n-1})$

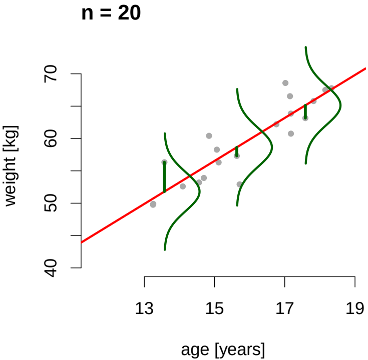


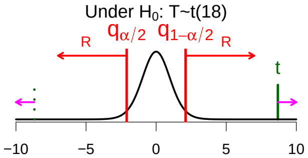


- Here: reject $H_0 : \beta_1 = 0$

Linear regression

- For $i = 1, \dots, n$ let
$$Y_i = \beta_0 + \beta_1 \cdot x_i + \sigma Z_i,$$
with $Z_1, \dots, Z_n$ i.i.d. RVs and $Z_1 \sim N(0,1)$, and $(\beta_0, \beta_1, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$
- $q_{1-\alpha/2}$ the $(1-\alpha/2)$-quantile of the $t(n-2)$-distribution
- estimators B_0, B_1 and S_r via 'least-squares', $SE_{B_1} = S_r/(s_x \cdot \sqrt{n-1})$



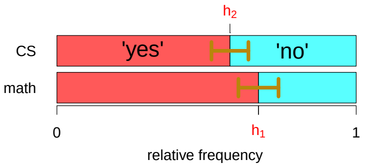


- Here: reject $H_0 : \beta_1 = 0$

For details on correlation and linear regression see the lecture: 'Linear Regression'

Frequencies

- Let $Y_{1,1},\dots,Y_{1,n_1},\ Y_{2,1},\dots,Y_{2,n_2}$ be independent RVs with $Y_{1,i}\sim ber(p_1)$ for $i=1,\dots,n_1$ and $Y_{2,j}\sim ber(p_2)$ for $j=1,\dots,n_2$, and $(p_1,p_2)\in(0,1)^2$
- let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $N(0,1)$-distribution



$$H_j := \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{j,i} \qquad \text{and} \qquad SE_{H_j} := \sqrt{\frac{H_j(1-H_j)}{n_j}}$$

Frequencies

- Let $Y_{1,1}, \ldots, Y_{1,n_1},\ Y_{2,1}, \ldots, Y_{2,n_2}$ be independent RVs with $Y_{1,i} \sim ber(p_1)$ for $i = 1, \ldots, n_1$ and $Y_{2,j} \sim ber(p_2)$ for $j = 1, \ldots, n_2$,

  and $(p_1, p_2) \in (0,1)^2$
- let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $N(0,1)$-distribution

Under $H_0 : p_2 - p_1 = 0$ it holds (approximately)

$$Z := \frac{(H_2 - H_1) - 0}{\sqrt{SE_{H_2}^2 + SE_{H_1}^2}} \overset{d}{\approx} N(0,1)$$

and equivalently: the confidence interval

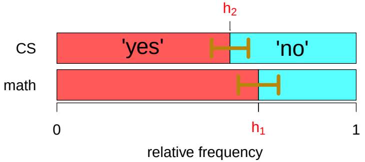
$$I := \left((H_2 - H_1) - q_{1-\alpha/2} \cdot \sqrt{SE_{H_2}^2 + SE_{H_1}^2},\ (H_2 - H_1) + q_{1-\alpha/2} \cdot \sqrt{SE_{H_2}^2 + SE_{H_1}^2} \right)$$

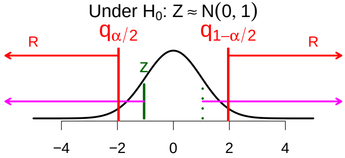
overlaps 0 with probability about $1 - \alpha$

$$H_j := \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{j,i} \qquad \text{and} \qquad SE_{H_j} := \sqrt{\frac{H_j(1-H_j)}{n_j}}$$

Frequencies

- Let $Y_{1,1}, \ldots, Y_{1,n_1},\ Y_{2,1}, \ldots, Y_{2,n_2}$ be independent RVs with $Y_{1,i} \sim ber(p_1)$ for $i = 1, \ldots, n_1$ and $Y_{2,j} \sim ber(p_2)$ for $j = 1, \ldots, n_2$,
  and $(p_1, p_2) \in (0,1)^2$
- let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $N(0,1)$-distribution

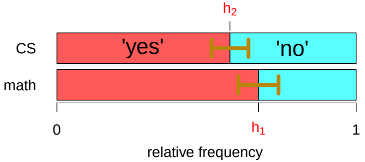


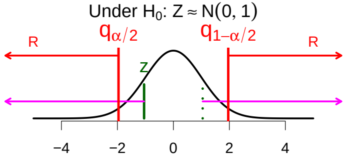


$$\color{red}{H_j} := \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{j,i} \qquad \text{and} \qquad SE_{H_j} := \sqrt{\frac{H_j(1-H_j)}{n_j}}$$

Frequencies

- Let $Y_{1,1}, \ldots, Y_{1,n_1}$, $Y_{2,1}, \ldots, Y_{2,n_2}$ be independent RVs with $Y_{1,i} \sim ber(p_1)$ for $i = 1, \ldots, n_1$ and $Y_{2,j} \sim ber(p_2)$ for $j = 1, \ldots, n_2$,
  and $(p_1, p_2) \in (0,1)^2$
- let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $N(0,1)$-distribution



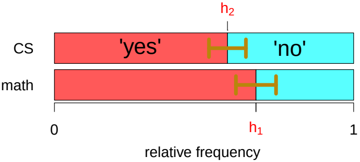


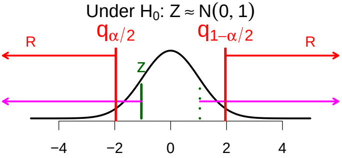
$$H_j := \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{j,i} \qquad \text{and} \qquad SE_{H_j} := \sqrt{\frac{H_j(1-H_j)}{n_j}}$$

- Here: can <u>not</u> reject $H_0 : p_2 = p_1$

Frequencies

- Let $Y_{1,1}, \ldots, Y_{1,n_1}, \ Y_{2,1}, \ldots, Y_{2,n_2}$ be independent RVs with $Y_{1,i} \sim ber(p_1)$ for $i = 1, \ldots, n_1$ and $Y_{2,j} \sim ber(p_2)$ for $j = 1, \ldots, n_2$,
and $(p_1, p_2) \in (0,1)^2$
- let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $N(0,1)$-distribution





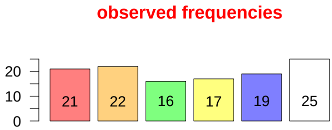
$$H_j := \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{j,i} \qquad \text{and} \qquad SE_{H_j} := \sqrt{\frac{H_j(1-H_j)}{n_j}}$$

- Here: can <u>not</u> reject $H_0 : p_2 = p_1$

For details on frequencies see the lecture: 'Proportions'

The χ^2-test (goodness of fit)

- Let $\mathfrak{X} = (X_1, \ldots, X_d)^t \sim mult(n, p)$, with $p \in (0,1)^d$ and $\sum_{k=1}^d p_k = 1$



The χ^2-test (goodness of fit)

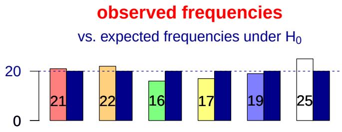
- Let $\mathfrak{X} = (X_1, \ldots, X_d)^t \sim mult(n, p)$, with $p \in (0,1)^d$ and $\sum_{k=1}^d p_k = 1$

Under $H_0 : p = (p_{0,1}, \ldots, p_{0,d})^t$ it holds (approximately)

$$X^2 := \sum_{k=1}^{d} \frac{(X_k - \mathbb{E}_{H_0}[X_k])^2}{\mathbb{E}_{H_0}[X_k]} \overset{d}{\approx} \chi^2(d-1)$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2(d-1)$ as $n \to \infty$

- here: $n = 120$ and $d = 6$, as well as $p_0 = (1/6, \ldots, 1/6)^t$



The χ^2-test (goodness of fit)

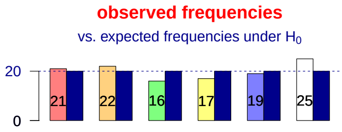
- Let $\mathfrak{X} = (X_1, \dots, X_d)^t \sim mult(n, p)$, with $p \in (0,1)^d$ and $\sum_{k=1}^d p_k = 1$

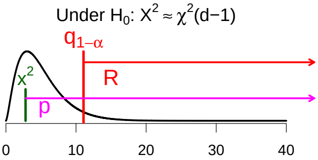
Under $H_0 : p = (p_{0,1}, \dots, p_{0,d})^t$ it holds (approximately)

$$X^2 := \sum_{k=1}^{d} \frac{(X_k - \mathbb{E}_{H_0}[X_k])^2}{\mathbb{E}_{H_0}[X_k]} \overset{d}{\approx} \chi^2(d-1)$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2(d-1)$ as $n \to \infty$

- here: $n = 120$ and $d = 6$, as well as $p_0 = (1/6, \dots, 1/6)^t$
- Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the $\chi^2(d-1)$-distribution





The χ^2-test (goodness of fit)

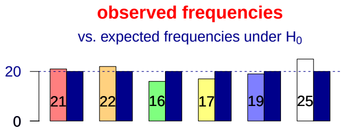
- Let $\mathfrak{X} = (X_1, \ldots, X_d)^t \sim mult(n, p)$, with $p \in (0,1)^d$ and $\sum_{k=1}^{d} p_k = 1$

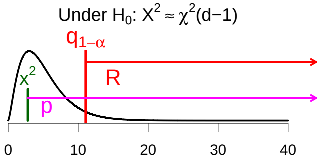
Under $H_0 : p = (p_{0,1}, \ldots, p_{0,d})^t$ it holds (approximately)

$$X^2 := \sum_{k=1}^{d} \frac{(X_k - \mathbb{E}_{H_0}[X_k])^2}{\mathbb{E}_{H_0}[X_k]} \overset{d}{\approx} \chi^2(d-1)$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2(d-1)$ as $n \to \infty$

- here: $n = 120$ and $d = 6$, as well as $p_0 = (1/6, \ldots, 1/6)^t$
- Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the $\chi^2(d-1)$-distribution
- Here: can <u>not</u> reject H_0





The χ^2-test (goodness of fit)

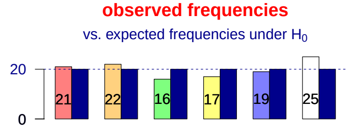
- Let $\mathfrak{X} = (X_1, \dots, X_d)^t \sim mult(n,p)$, with $p \in (0,1)^d$ and $\sum_{k=1}^d p_k = 1$

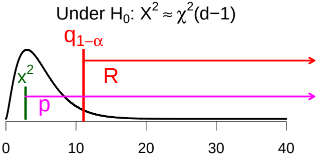
Under $H_0 : p = (p_{0,1}, \dots, p_{0,d})^t$ it holds (approximately)

$$X^2 := \sum_{k=1}^{d} \frac{(X_k - \mathbb{E}_{H_0}[X_k])^2}{\mathbb{E}_{H_0}[X_k]} \overset{d}{\approx} \chi^2(d-1)$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2(d-1)$ as $n \to \infty$

- here: $n = 120$ and $d = 6$, as well as $p_0 = (1/6, \dots, 1/6)^t$
- Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the $\chi^2(d-1)$-distribution
- Here: can not reject H_0

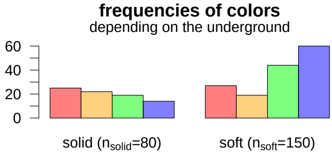




For details see the lecture: 'The χ^2-test (goodness of fit)'

The χ^2-test for independence

- Model: Let $\mathfrak{X} = (X_{1,1}, \ldots, X_{d_1,d_2})^t \sim mult(n, p)$ with $p = (p_{1,1}, \ldots, p_{d_1,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j,k} p_{j,k} = 1$



The χ^2-test for independence

- Model: Let $\mathfrak{X} = (X_{1,1}, \ldots, X_{d_1,d_2})^t \sim mult(n,p)$ with $p = (p_{1,1}, \ldots, p_{d_1,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j,k} p_{j,k} = 1$

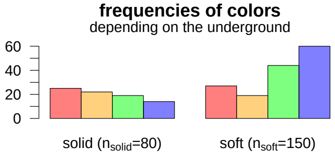
Under $H_0 : p = p_0 := (p_{1,\cdot} \cdot p_{\cdot,1}, \ldots, p_{d_1,\cdot} \cdot p_{\cdot,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j=1}^{d_1} p_{j,\cdot} = \sum_{k=1}^{d_2} p_{\cdot,k} = 1$

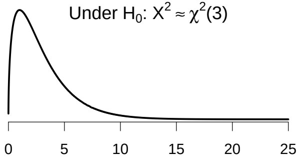
it holds (approximately)

$$X^2 := \sum_{j,k} \frac{\left(X_{j,k} - \frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}\right)^2}{\frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}} \overset{d}{\approx} \chi^2((d_1 - 1) \cdot (d_2 - 1))$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2((d_1 - 1)(d_2 - 1))$ as $n \to \infty$

- Here: $d_1 = 2$, $d_2 = 4$, i.e., $X^2 \overset{H_0}{\sim} \chi^2(3)$ (approx)
- $X_{j,\cdot}$, $X_{\cdot,k}$, $p_{j,\cdot}$ and $p_{\cdot,k}$ are the 'marginal frequencies / probabilites'





The χ^2-test for independence

- Model: Let $\mathfrak{X} = (X_{1,1}, \ldots, X_{d_1,d_2})^t \sim mult(n,p)$
  with $p = (p_{1,1}, \ldots, p_{d_1,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j,k} p_{j,k} = 1$

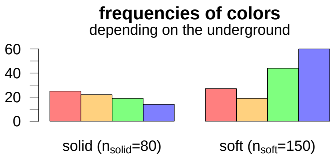
Under $H_0 : p = p_0 := (p_{1,\cdot} \cdot p_{\cdot,1}, \ldots, p_{d_1,\cdot} \cdot p_{\cdot,d_2})^t \in (0,1)^{d_1 \cdot d_2}$
and $\sum_{j=1}^{d_1} p_{j,\cdot} = \sum_{k=1}^{d_2} p_{\cdot,k} = 1$

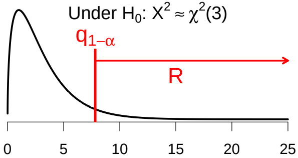
it holds (approximately)

$$X^2 := \sum_{j,k} \frac{\left(X_{j,k} - \frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}\right)^2}{\frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}} \overset{d}{\approx} \chi^2((d_1 - 1) \cdot (d_2 - 1))$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2((d_1 - 1)(d_2 - 1))$ as $n \to \infty$

- Here: $d_1 = 2$, $d_2 = 4$, i.e., $X^2 \overset{H_0}{\sim} \chi^2(3)$ (approx)
- $X_{j,\cdot}$, $X_{\cdot,k}$, $p_{j,\cdot}$ and $p_{\cdot,k}$ are the 'marginal frequencies / probabilites'
- Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the $\chi^2([d_1 - 1] \cdot [d_2 - 1])$-distr.





The χ^2-test for independence

- Model: Let $\mathfrak{X} = (X_{1,1}, \ldots, X_{d_1,d_2})^t \sim mult(n,p)$ with $p = (p_{1,1}, \ldots, p_{d_1,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j,k} p_{j,k} = 1$

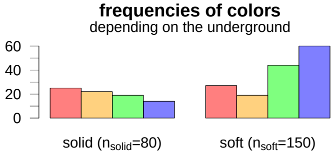
Under $H_0 : p = p_0 := (p_{1,\cdot} \cdot p_{\cdot,1}, \ldots, p_{d_1,\cdot} \cdot p_{\cdot,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j=1}^{d_1} p_{j,\cdot} = \sum_{k=1}^{d_2} p_{\cdot,k} = 1$

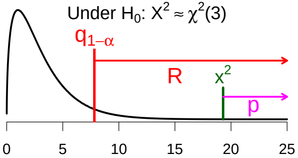
it holds (approximately)

$$X^2 := \sum_{j,k} \frac{\left(X_{j,k} - \frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}\right)^2}{\frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}} \overset{d}{\approx} \chi^2((d_1 - 1) \cdot (d_2 - 1))$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2((d_1 - 1)(d_2 - 1))$ as $n \to \infty$

- Here: $d_1 = 2, d_2 = 4$, i.e., $X^2 \overset{H_0}{\sim} \chi^2(3)$ (approx)
- $X_{j,\cdot}$, $X_{\cdot,k}$, $p_{j,\cdot}$ and $p_{\cdot,k}$ are the 'marginal frequencies / probabilites'
- Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the $\chi^2([d_1 - 1] \cdot [d_2 - 1])$-distr.
- Here: reject H_0





The χ^2-test for independence

- Model: Let $\mathfrak{X} = (X_{1,1}, \ldots, X_{d_1,d_2})^t \sim mult(n,p)$ with $p = (p_{1,1}, \ldots, p_{d_1,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j,k} p_{j,k} = 1$

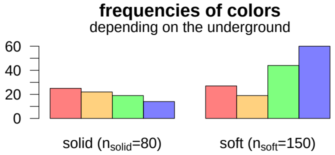
Under $H_0 : p = p_0 := (p_{1,\cdot} \cdot p_{\cdot,1}, \ldots, p_{d_1,\cdot} \cdot p_{\cdot,d_2})^t \in (0,1)^{d_1 \cdot d_2}$ and $\sum_{j=1}^{d_1} p_{j,\cdot} = \sum_{k=1}^{d_2} p_{\cdot,k} = 1$

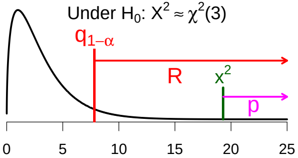
it holds (approximately)

$$X^2 := \sum_{j,k} \frac{\left(X_{j,k} - \frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}\right)^2}{\frac{X_{j,\cdot} \cdot X_{\cdot,k}}{n}} \overset{d}{\approx} \chi^2((d_1 - 1) \cdot (d_2 - 1))$$

in fact, it holds that $X^2 \xrightarrow{d} \chi^2((d_1 - 1)(d_2 - 1))$ as $n \to \infty$

- Here: $d_1 = 2, d_2 = 4$, i.e., $X^2 \overset{H_0}{\sim} \chi^2(3)$ (approx)
- $X_{j,\cdot}$, $X_{\cdot,k}$, $p_{j,\cdot}$ and $p_{\cdot,k}$ are the 'marginal frequencies / probabilites'
- Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the $\chi^2([d_1 - 1] \cdot [d_2 - 1])$-distr.
- Here: reject H_0





For details see the lecture: 'The χ^2-test for independence'

Thank you!