Part III: Estimation

April 3, 2009

Example: Quality control

Determine the rate of defect parts

- data: 200 items are checked,
  G good D defect

  $$\cdots G\ G\ G\ G\ D\ G\ G\ G\ D\ G\ G\ G \cdots\ D\ G\ G\ D\ D\ G\ G\ G\ G\ G\ D\ G \cdots$$

  suppose: 30 times D and 170 times G
- intuitive **estimate for rate of defect parts**
  number of defect parts divided by all checked parts

  $$\hat{p} = \frac{30}{200} = 0.15$$

  relative frequency of defect parts

Example: Quality control

- **mathematical model : Binomial distribution**

$$\underline{z} = (z_1, \ldots, z_n), \qquad z_i = \begin{cases} 1 & \text{if } D \\ 0 & \text{if } G \end{cases}$$

- $$\cdots 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0 \cdots\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 0 \cdots$$

- z_i realizations of $\mathbf{B(1, p)}$

$$\mathbf{P}(Z_i = 1) = p \quad \mathbf{P}(Z_i = 0) = 1 - p$$

parameter $\boldsymbol{p}$ rate of defect parts
(probability of success)

Example: Quality control

- in other words
  The parameter p is estimated by the relative frequency

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} z_i$$

which is the arithmetic mean of the indicator variables z_i

Example: Quality control

- estimation: use the data to determine a plausible value of an unknown parameter

Questions

1. How precise is this estimate?
2. Are there better estimators? Do we lose information by using only the counts of 'defect' and 'good parts' and not the whole data vector of 0's and 1's?
3. How confident can we be in statements based on this estimator? (queries)
4. Which role does the sample size $n = 200$ play?

Medical Investigations: Birth weight

- data ($n = 1239$): age of parents, weight of mother, weight father, height...., age of parents,....
- column bwt (in once)

$$\cdots 120 \;\; 113 \;\; 128 \;\; 123 \;\; 98 \;\; 108 \;\; \cdots\cdots \;\; 138 \;\; 124 \;\; 130 \;\; 119 \;\; \cdots$$

- **mathematical model:**
  data $z_1, \ldots, z_n$ are values of random variables $\boldsymbol{Z_i}$ varying around a constant, i.e. an **expectation** μ, their **dispersion** is characterized by a parameter σ^2

$$\mathbf{E}Z_i = \mu \qquad \mathbf{Var}\ Z_i = \sigma^2 \qquad i = 1, \ldots, n$$

Example: Birth weight

- estimates for these parameters: empirical analogs
- for $\mathrm{E}Z_i = \mu$ **arithmetic mean**

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \qquad \text{here} \quad \bar{z} = 119.7$$

for Var $Z_i = \sigma^2$ **empirical variance**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 \qquad \text{here} \quad s^2 = 376,9$$

Example: Birth weight

Questions:

1. How precise are these estimates?
2. Is a newborn child with a weight of 85 once serious underweight? Are there "norm values"? (not fulfilled $\rightarrow$ medical treatment)
3. Why has the empirical variance the factor $(n-1)^{-1}$ and not n^{-1}?

Example: Life time

- **Reliability investigations:** Failure time of a certain item of equipment
  Life time for 10 items

$$2,\ 72,\ 51,\ 60,\ 33,\ 27,\ 14,\ 24,\ 4,\ 21$$

- parameter of interest: expected life time $\mathrm{E}Z$
- intuitive estimate: arithmetic mean $\quad \overline{z} = 30.8$

Example: Life time

- **Reliability investigations:** The 10 pieces were installed at different times. At the end of the study three items had still not failed, and their failure times are censored; we know for these items only that their failure time exceed the observed time: Life time for 10 items

$$2,\ \geq 72,\ 51,\ \geq 60,\ 33,\ 27,\ 14,\ 24,\ 4,\ \geq 21$$

- parameter of interest: expected life time
  intuitive estimate:? arithmetic mean $\overline{z} = 30.8$
  or arithmetic mean of the uncensored observations? or....

Example: Braking distance/ speed

- data

$$(x_1, y_1), \ldots, (x_n, y_n), \quad (x_i, y_i) \in \mathbb{R}_+ \times \mathbb{R}_+$$

x_i speed $\qquad$ y_i braking distance

- dependence described by a polynomial of degree 2

$$y_i = a + bx_i + cx_i^2 + \varepsilon_i$$

ε_i random error, random noise

Example: Braking distance/ speed

$$y_i = a + bx_i + cx_i^2 + \varepsilon_i$$

- parameter of interest:

  coefficients (a, b, c)

- Estimation of the coefficients by **Methode der kleinsten Quadrate/Method of least squares**

What is an estimator

- problem: estimate a parameter γ given data $z_1, \ldots, z_n$

Definition

A (point) ***estimator*** *(Schätzer, Schätzfunktion) is a function from the data space in in the parameter space:*

$$\widehat{\gamma} : (z_1, \ldots, z_n) \to \widehat{\gamma}(z_1, \ldots, z_n) \in \Gamma.$$

The value $\widehat{\gamma}(z_1, \ldots, z_n)$ *is the* ***estimate*** *(Schätzung, Schätzwert).*

often no difference between the function $\hat{\gamma}$ and the value of the function $\hat{\gamma}(z_1, \ldots, z_n)$

Properties of an estimator

- Properties of the **procedure**, of the function - not of the estimate, that is **not of the value**!!
- in other words
  Criterion: How **good** is the estimation procedure **for all possible data** which can be drawn from the underlying phenomenon!

Requirements

1. **”in average”** the **right value** of the parameter γ
2. **”in average ”** **not highly variable** about the parameter γ

- **”in average”** means ”average over all possible data” mathematically: consider data as values of r.v.’s and take the **expectation**;

  !!!! expectation corresponds to the average over all possible data

Properties of an estimator

1 ”in average” the **right (unknown) value** of the parameter γ
mathematically: unbiased (erwartungstreu, unverzerrt)

$$\mathbf{E}\hat{\gamma}(Z_1, \ldots, Z_n) = \gamma$$

2 ”in average ” **not highly variable** about the parameter γ
mathematically: small variance

$$\mathbf{Var}\ \hat{\gamma}(Z_1, \ldots, Z_n) \quad \textbf{minimal!}$$

Unbiased estimators

Given an estimator $\hat{\gamma}$ for the parameter γ

$$\hat{\gamma} = \hat{\gamma}(Z_1, \ldots, Z_n)$$

Definition

An estimator $\widehat{\gamma} \in \Gamma$ is called unbiased (erwartungstreu) if

$$\mathbf{E}\widehat{\gamma} = \gamma \qquad \textit{for all } \gamma$$

Unbiasedness of the arithmetic mean

- parameter of interest: **expectation of a distribution**,
  success probability $\mathrm{B}(1, p)$: $\gamma = p$,
  normal distribution $\mathrm{N}(\mu, \sigma^2)$: $\gamma = \mu$,
  exponential distribution $\mathrm{Exp}(\lambda)$: $\gamma = 1/\lambda$,
  unspecified: expectation (e. g. expected survival time)
- $Z_1, \ldots, Z_n \quad \mathrm{E}Z_i = \gamma$

$$\hat{\gamma} = \overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

$$\mathrm{E}\,\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}Z_i = \frac{1}{n}\sum_{i=1}^{n} \gamma = \gamma$$

$\Rightarrow$ $\overline{Z}$ is unbiased for γ.

Unbiasedness of the empirical variance

- parameter of interest: variance of a distribution

$$Z_1, \ldots, Z_n \qquad \text{Var } Z_i = \gamma$$

$$\hat{\gamma} = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \overline{Z})^2$$

One can show $\qquad \mathrm{E} \sum_{i=1}^{n} (Z_i - \overline{Z})^2 = (n-1)\gamma$

$$\Rightarrow \quad \text{factor} \frac{1}{n-1}$$

$$\Rightarrow \quad \mathrm{E}\,\widehat{\gamma} = \gamma$$

$\Rightarrow \quad s^2$ is unbiased for $\gamma = \text{Var } Z_i$.

Unbiasedness

- Let $z_1, \ldots, z_n$ be values of independent r.v. 's Z_i with the same distribution (not necessarily specified); then we have

The arithmetic mean is an unbiased estimator for the expectation.
The empirical variance is an unbiased estimator for the variance.

Unbiasedness - Bernoulli - Relative frequency

- Example: **Quality control**
  data are values of $\mathrm{B}(1, p)$-distributed r.v.'s Z_i;
  p estimated by **relative frequency** which is the arithmetic mean of the Z_i's
  $\Rightarrow$ it is **unbiased**: $\mathrm{E}\hat{p} = p$
- parameter of interest: variance, i.e. $\gamma = p(1-p)$
  "**plug-in**" estimator: $\hat{\gamma} = \hat{p}(1-\hat{p})$ is not unbiased, because

$$\mathrm{E}\,\hat{\gamma} = \mathrm{E}\hat{p} - \mathrm{E}\hat{p}^2 = p(1-p)(1-\frac{1}{n})$$

The **bias** is:

$$\mathrm{E}\hat{\gamma} - \gamma = -\frac{1}{n}p(1-p)$$

Estimator with smallest variance

- **An estimator should have minimal variance!**

Definition

Let $\mathcal{E}$ be the set of all unbiased estimators for γ. If

$$\textbf{Var } \hat{\gamma}^* \leq \textbf{Var } \hat{\gamma}$$

*for all $\hat{\gamma} \in \mathcal{E}$ and for all $\vartheta \in \Theta$ then $\hat{\gamma}^*$ is called **best estimator** in the set $\mathcal{E}$.*

(γ is part of the parameter ϑ which characterizes the underlying distribution.)

Estimators with minimal variance

- Variance of the arithmetic mean
- $Z_1, \ldots, Z_n \quad \text{with} \quad \mathrm{E}Z_i = \mu \quad \text{and} \quad \mathrm{Var}\, Z_i = \sigma^2$

$$\mathbf{Var}\ \overline{Z} = n^{-1}\sigma^2$$

  For many distributions (binomial, normal exponential,...) it can be shown that the **arithmetic mean is the best estimator** for the expectation (under all unbiased estimators).
- If the data are normal distributed ($\mathbf{N}(\mu, \sigma^2)$) one can show that the **empirical variance s^2** is the **best estimator for σ^2** under all unbiased estimators.

Standard error

- **variance of the arithmetic mean**

$$\mathrm{Var}\,\overline{Z} = n^{-1}\sigma^2$$

- The variance of the arithmetic mean can be estimated (and is estimated automatically in standard software packages) by

$$n^{-1}\,s^2 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \overline{z})^2$$

- The quantity

$$\sqrt{n^{-1}s^2} = \boldsymbol{s/\sqrt{n}}$$

is called **standard error**. (Standardfehler)

Mathematical Background *

- Notation of **sufficiency**:

$$\text{data} \xrightarrow{\text{Reduction}} \text{estimate}$$

example: $(z_1, \ldots, z_n)$ (n-dim.) $\longrightarrow$ $(\overline{z}, s^2)$ 2-dim.

- If this reduction from the data to the estimator does not mean a loss of information about the underlying parameter then the estimator is called ”sufficient”.
- Roughly speaking: If an unbiased estimator is sufficient then it also optimal.

Mathematical Background - Sufficiency

Example: **Normal distribution** $\mathrm{N}(\mu, \sigma^2)$

- parameter of interest: $\boldsymbol{\mu}$
  It is only necessary to know $\overline{Z}$; the whole vector $(Z_1, \ldots, Z_n)$ or the order of the Z_i's do not provide more information
- If you know only $\overline{Z}$ you have lost information about σ^2.
- $(\sum_{i=1}^n Z_i, \sum_{i=1}^n Z_i^2)$ contains the complete information about (μ, σ^2).
- $(\sum_{i=1}^n Z_i, \sum_{i=1}^n Z_i^2)$ is sufficient for (μ, σ^2).

Simulation

to show what means:

- An estimator is a **random variable**.
- **Unbiasedness** of an estimator
- **Var** $\widehat{\gamma} = \cdot$

Method

- **Generation of a sample** of size n of $\mathrm{B}(1, p)$-distributed random numbers with p
  (here $n = 100$ and $p = 0.2$)
- **Forget the value** $p = 0.2$; in other words, we have the problem of estimating the unknown parameter of a $B(1, p)$-distribution
- Application of the **estimation** procedure
- **Repeat** this simulation M times $\Rightarrow$ M estimates
  (here $M = 500$)

Simulation procedure

- For $m = 1, \ldots, M$
  - generate values
    $$z_1^{(m)}, z_2^{(m)}, \ldots, z_n^{(m)}$$
    of $\mathbf{B}(1, p)$-distributed r.v.'s,
  - compute the sum $S_n^{(m)} = \sum_{i=1}^n z_i^{(m)}$ and the estimate
    $$\hat{p}^{(m)} = \frac{1}{n} \sum_{i=1}^{n} z_i^{(m)}$$

Simulation

```
> #sample number 1
Y[1,]
  [1] 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
 [38] 0 0 1 0 0 0 0 0 1 0 0 0 1 0 1 0 0 1 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
 [75] 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 1 1 1 0 0 0 0 0
> #sample number 40
> Y[40,]
  [1] 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0
 [38] 1 0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0
 [75] 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 0 0
> p.hat
  [1] 0.22 0.30 0.17 0.18 0.19 0.20 0.13 0.21 0.17 0.13 0.21 0.20 0.24 0.24 0.17
 [16] 0.15 0.26 0.27 0.25 0.25 0.15 0.16 0.19 0.18 0.17 0.16 0.18 0.26 0.18 0.26
 [31] 0.15 0.22 0.22 0.21 0.19 0.12 0.25 0.18 0.21 0.18 0.20 0.23 0.18 0.19 0.23
 [46] 0.19 0.26 0.24 0.25 0.14 0.20 0.19 0.23 0.19 0.20 0.21 0.24 0.21 0.22 0.13
 [61] 0.21 0.20 0.26 0.14 0.27 0.28 0.15 0.17 0.19 0.21 0.27 0.15 0.23 0.16 0.18
 [76] 0.24 0.18 0.26 0.22 0.20 0.20 0.19 0.16 0.24 0.22 0.17 0.22 0.17 0.14 0.20
 [91] 0.21 0.18 0.25 0.16 0.22 0.26 0.21 0.25 0.22 0.20 0.21 0.21 0.25 0.18 0.23
[106] 0.27 0.17 0.13 0.16 0.12 0.13 0.20 0.13 0.26 0.22 0.20 0.23 0.20 0.20 0.29
[121] 0.12 0.17 0.21 0.20 0.27 0.28 0.21 0.20 0.19 0.30 0.10 0.18 0.21 0.17 0.20
[136] 0.15 0.24 0.18 0.20 0.26 0.29 0.14 0.27 0.27 0.21 0.18 0.23 0.19 0.25 0.19
[151] 0.22 0.18 0.15 0.27 0.20 0.21 0.25 0.18 0.19 0.18 0.21 0.24 0.26 0.23 0.18
[166] 0.16 0.20 0.22 0.24 0.20 0.21 0.23 0.17 0.22 0.27 0.15 0.21 0.21 0.24 0.19
[181] 0.21 0.22 0.19 0.16 0.18 0.15 0.15 0.26 0.24 0.13 0.16 0.26 0.23 0.17 0.23
[196] 0.23 0.19 0.21 0.20 0.29 0.21 0.17 0.14 0.20 0.21 0.21 0.20 0.31 0.20 0.15
[211] 0.20 0.16 0.20 0.32 0.13 0.19 0.20 0.15 0.27 0.17 0.16 0.23 0.23 0.21 0.16
```

Result 1

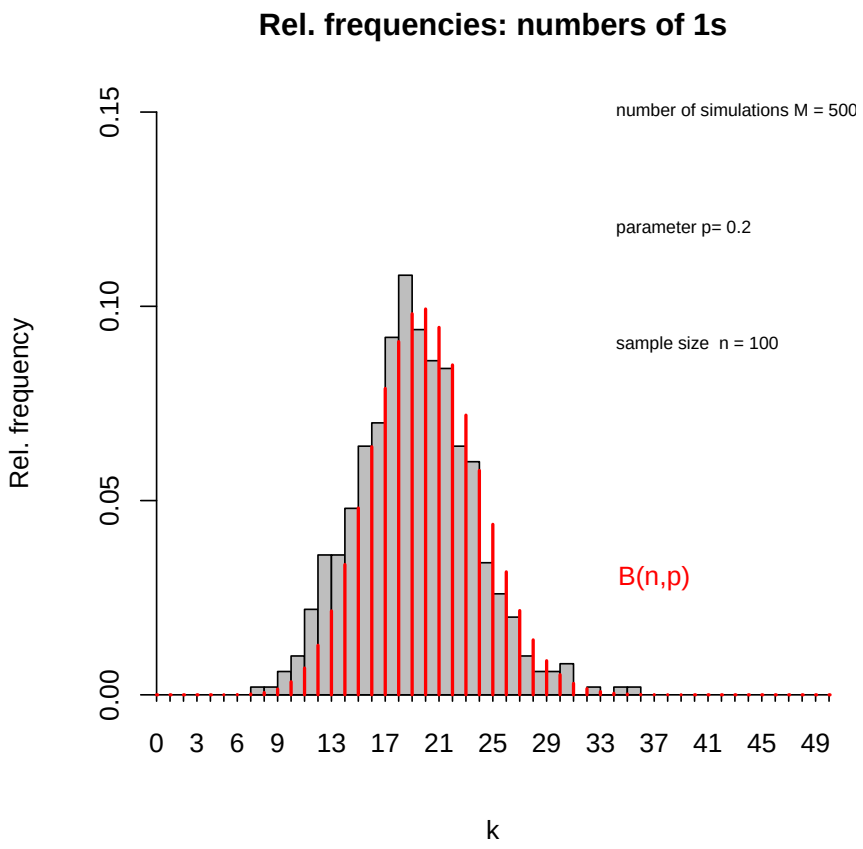
- From probability theory: For each m

$$\sum_{i=1}^{n} Z_i^{(m)} \sim \mathrm{B}(n, p)$$

- the empirical analog of this sum of r.v.'s is $S_n^{(m)}$

$\Rightarrow$ the **histogram** of the $S_n^{(m)}$'s must be similar to the **barplot** of $\mathrm{B}(n, p)$

- Plot the histogram of the sums $S_n^{(m)}$'s and compare with the theoretic distribution $\mathrm{B}(n, p)$.



Result 2

- The estimator is unbiased (erwartungstreu), i.e. ”expectation of the estimator = parameter p”
- we expect that
  the mean of the estimates

$$\frac{1}{M}\sum_{m=1}^{M}\hat{p}^{(m)}$$

  is close to p
- We estimate the **bias** (theoretically =0) by

$$\frac{1}{M}\sum_{m=1}^{M}\hat{p}^{(m)} - p$$

Result 3

- Interpretation of the variance
  theoretically we have for the variance of the estimator:

$$\mathrm{Var}\ \hat{p} = n^{-1} p(1-p)$$

  (here $10^{-2} \cdot 0.2 \cdot 0.8 = 0.16 \cdot 10^{-2}$)
- we can approximate this by the empirical variance of our simulations

$$\frac{1}{M-1} \sum_{m=1}^{M} (\hat{p}^{(m)} - \overline{\hat{p}^{(\cdot)}})^2$$

Results

- Compute the mean of the estimates and the simulation-estimator of the bias

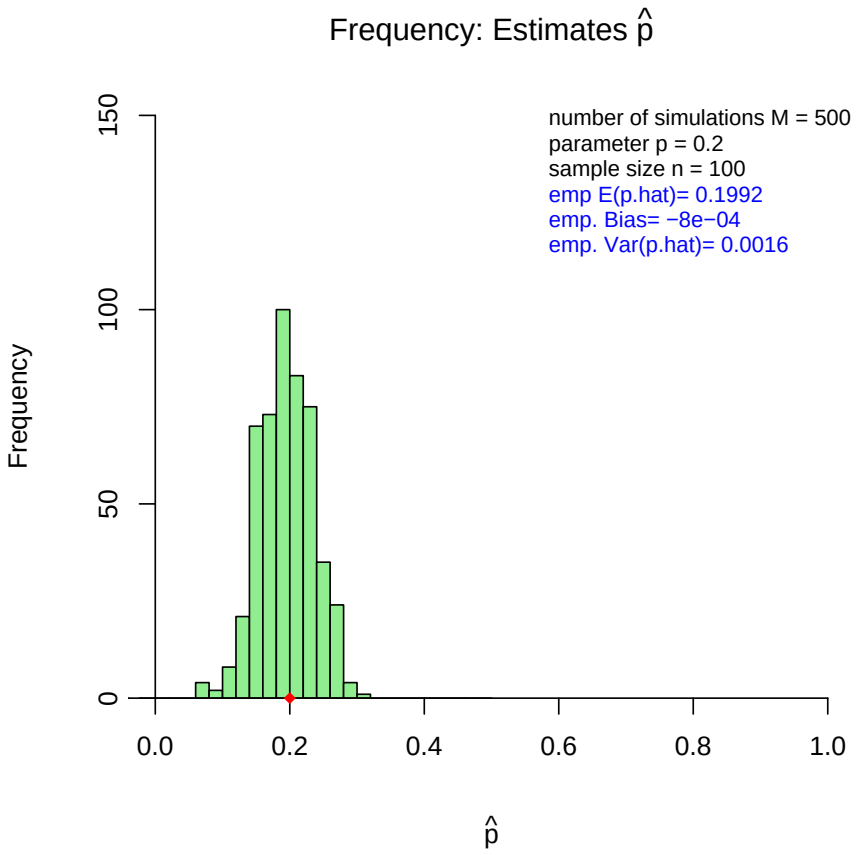
$$emp.\mathrm{E}(\hat{p}) = \frac{1}{M}\sum_{m=1}^{M}\hat{p}^{(m)} \qquad emp.Bias(\hat{p}) = emp.\mathrm{E}(\hat{p}) - p$$

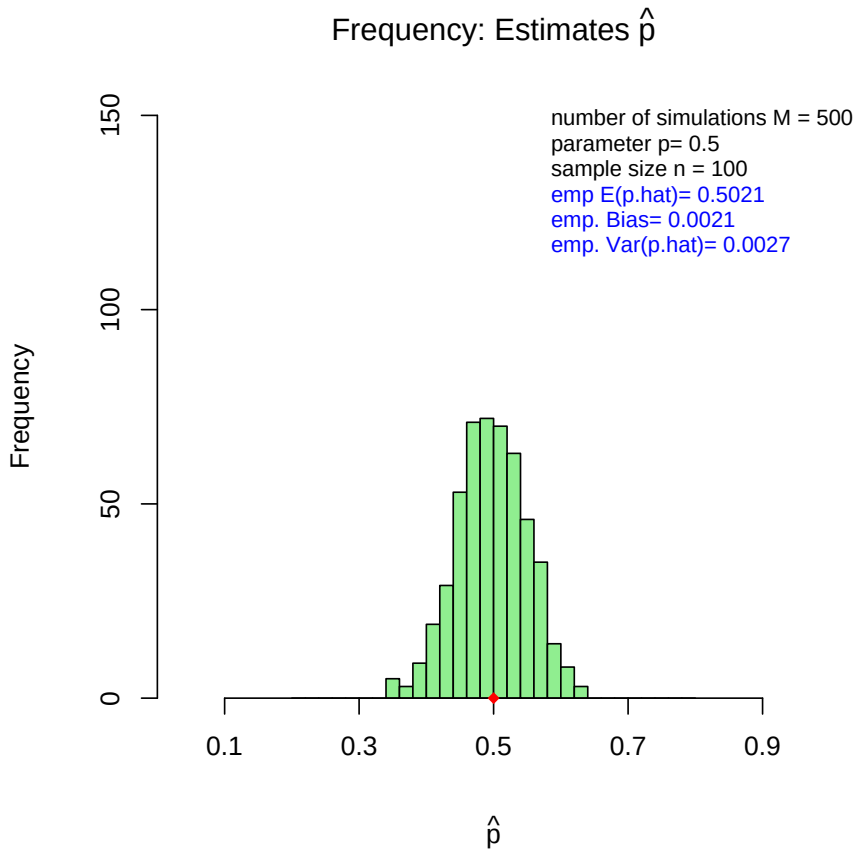
- Compute the empirical variance of the estimates

$$emp.\mathrm{Var}\,(\hat{p}) = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{p}^{(m)} - emp.\mathrm{E}(\hat{p}))^2$$

$emp.\mathrm{E}(\hat{p})$	$emp.Bias(\hat{p})$	$emp.\mathrm{Var}\,(\hat{p})$
0.19916	-0.00084	0.001622940

  $(p = 0.2)$


Frequency: Estimates p̂
number of simulations M = 500
parameter p = 0.2
sample size n = 100
emp E(p.hat)= 0.1992
emp. Bias= −8e−04
emp. Var(p.hat)= 0.0016
Frequency
0
50
100
150
0.0
0.2
0.4
0.6
0.8
1.0
p̂


Frequency: Estimates p̂
number of simulations M = 500
parameter p= 0.5
sample size n = 100
emp E(p.hat)= 0.5021
emp. Bias= 0.0021
emp. Var(p.hat)= 0.0027
Frequency
0
50
100
150
0.1
0.3
0.5
0.7
0.9
p̂

The mean square error

- **MSE** **Mittlerer quadratischer Fehler** (MQF)
- unbiasedness of an estimator, i.e. no systematic error suitable requirement
  but useless if the variance is very large
- combination of both

$$\mathbf{MSE}(\hat{\gamma}) = \mathbf{E}(\hat{\gamma} - \gamma)^2$$

Mean Squared Error

$$
\begin{aligned}
\mathrm{MSE}(\hat{\gamma}) &= \mathrm{E}(\hat{\gamma} - \gamma)^2 \\
&= \mathrm{E}(\hat{\gamma} - \mathrm{E}\hat{\gamma} + \mathrm{E}\hat{\gamma} - \gamma)^2 \\
&= \mathrm{E}(\hat{\gamma} - \mathrm{E}\hat{\gamma})^2 + 2\mathrm{E}(\hat{\gamma} - \mathrm{E}\hat{\gamma})(\mathrm{E}\hat{\gamma} - \gamma) + \mathrm{E}(\mathrm{E}\hat{\gamma} - \gamma)^2 \\
&= \mathrm{E}(\hat{\gamma} - \mathrm{E}\hat{\gamma})^2 + \qquad 0 \qquad\qquad + (\mathrm{E}\hat{\gamma} - \gamma)^2 \\
&= \mathrm{Var}\ \hat{\gamma} + [\mathrm{Bias}(\hat{\gamma})]^2
\end{aligned}
$$

Goal: estimation procedure with small MSE (biased ?)

Example for an useful biased estimator

Estimation of the success probability in $\mathrm{B}(1, p)$

- $Z_1, \ldots, Z_n$ independent $B(1, p)$-distributed r.v.'s
- we know: unbiased estimator for p (best unbiased estimator):

$$\hat{p}_1 = \frac{1}{n} \sum_{i=1}^{n} Z_i$$

- consider

$$\hat{p}_2 = \frac{\sum_{i=1}^{n} Z_i + 1}{n + 2} = \frac{n\hat{p}_1 + 1}{n + 2}$$

Example for an useful biased estimator

We have for $\hat{p}_2 = \frac{n\hat{p}_1+1}{n+2}$

$$\hat{p}_2 \geq \hat{p}_1 \qquad \Leftrightarrow \qquad \hat{p}_1 \leq 1/2$$

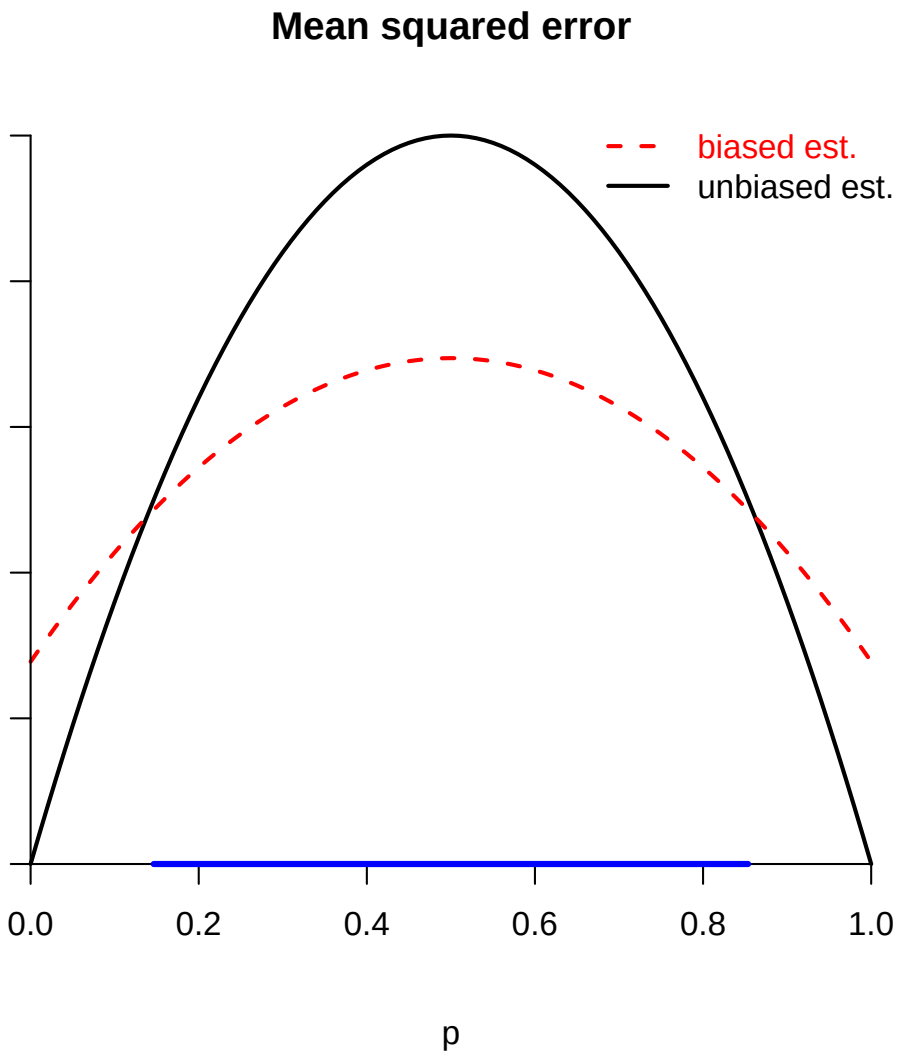
- **MSE of $\hat{p}_2$**

$$B(\hat{p}_2) = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}$$

$$\mathrm{MSE}(\hat{p}_2) = \frac{np(1-p)+(1-2p)^2}{(n+2)^2}$$

- **MSE of $\hat{p}_1$**

$$B(\hat{p}_1) = 0 \quad \mathrm{MSE}(\hat{p}_1) = \frac{p(1-p)}{n}$$


Mean squared error
biased est.
unbiased est.
0.0
0.2
0.4
0.6
0.8
1.0
p

Example for an useful biased estimator

$$\mathrm{MSE}(\hat{p}_2) \leq \mathrm{MSE}(\hat{p}_1)$$

if and only if

$$\frac{|p - \frac{1}{2}|^2}{p(1-p)} \leq 1 + \frac{1}{n}$$

this is satisfied for arbitrary sample size n if

$$\boldsymbol{|p - 1/2| \leq 1/\sqrt{8} \approx 0.35}$$

- If we know by some **prior information** that p takes only values around 0.5 then it is better to use the estimator $\hat{p}_2$ to estimate p than $\hat{p}_1$. (↷ Bayesian estimation theory)

A word about the sample size

- Investigation of the estimator for large n
  idea: more data $\longrightarrow$ more information
  $\longrightarrow$ estimate $\hat{\gamma}_n$ more close to γ
- mathematically: **limit**

$$\boldsymbol{n} \longrightarrow \infty \qquad \hat{\gamma}_{\boldsymbol{n}} \longrightarrow \gamma$$

- strong consistency

$$\mathrm{P}(\lim_{n\to\infty} \hat{\gamma}_n = \gamma) = 1$$

Weak consistency of an estimator

- limit of a (nonrandom) sequence a_n: $\lim_{n\to\infty} a_n = 0$:
  For all $\varepsilon > 0$ there exists a number $n_0(\varepsilon)$ such that for all $n > n_0$ $|a_n| \leq \varepsilon$.
- For estimators: **Weak consistency**
  For all $\delta > 0$
  $$\lim_{n\to\infty} \mathbf{P}\left(|\hat{\gamma}_n - \gamma| > \delta\right) = 0$$
- The probability that the estimate differs from the parameter more than δ (small) tends to zero as $n \to \infty$.
  Roughly speaking: If n is large the probability that $\hat{\gamma}_n$ differs from γ more than δ is small.
- Do not use estimators which are not consistent!
- Proof of consistency $\to$ mathematicians!

Until now

- Until now: $\overline{Z}$ is an unbiased consistent estimator for the expectation of a distribution, the empirical variance S^2 is an unbiased and consistent estimator for the variance of a distribution.
- The empirical distribution function $\hat{F}_n$

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1(Z_i \leq t)$$

is an unbiased and consistent estimator for the distribution function F defined by

$$F(t) = \mathrm{P}(Z_i \leq t).$$

Other quantities of a distribution: Median

- The **median** of a distribution is defined by the value z_{med} with

$$F(z_{med}) = 1/2$$

("halbiert die Wahrscheinlichkeitsmasse")

- The median is estimated by the **empirical or sample median** defined by:

$$\hat{F}_n(\hat{z}_{med}) = 1/2$$

Definition

$z_{1:n} \leq \cdots \leq z_{n:n}$ *ordered data*

$$\hat{z}_{med} = \begin{cases} z_{(n+1)/2:n} & \textit{for odd } n \\ \frac{1}{2}(z_{n/2:n} + z_{n/2+1:n}) & \textit{for even } n \end{cases}$$

Quartiles and percentiles

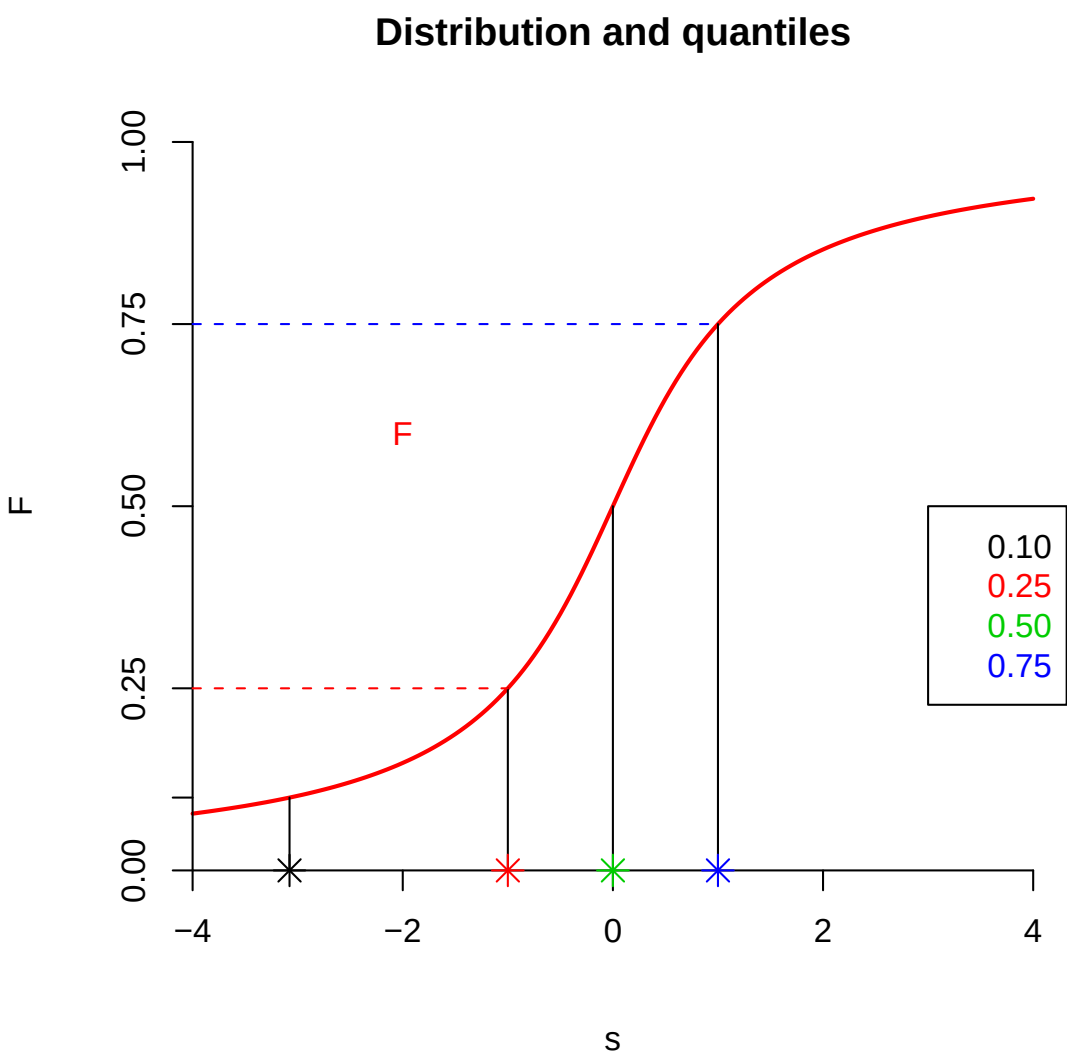
- (empirical) median divides the (data) probability mass in half; the **three sample quartiles** of a set of numerical data are numbers that divide the ordered data into four segments of about equal size
- estimators of z_q with $z_q = F^{inv}(q)$ for $q = 1/4,\ 1/2,\ 3/4$
- quartiles are special quantiles:

Definition

z_q is q-th quantile of a distribution F: $\quad z_q = F^{inv}(q)$

$\hat{z}_q$ is q-th empirical quantile of a sample: $\quad \hat{z}_q = \hat{F}_n^{inv}(q)$

Quartiles and percentiles



Mean, empirical variance and empirical quantiles in R

- data $z = (z_1, \ldots, z_n)$

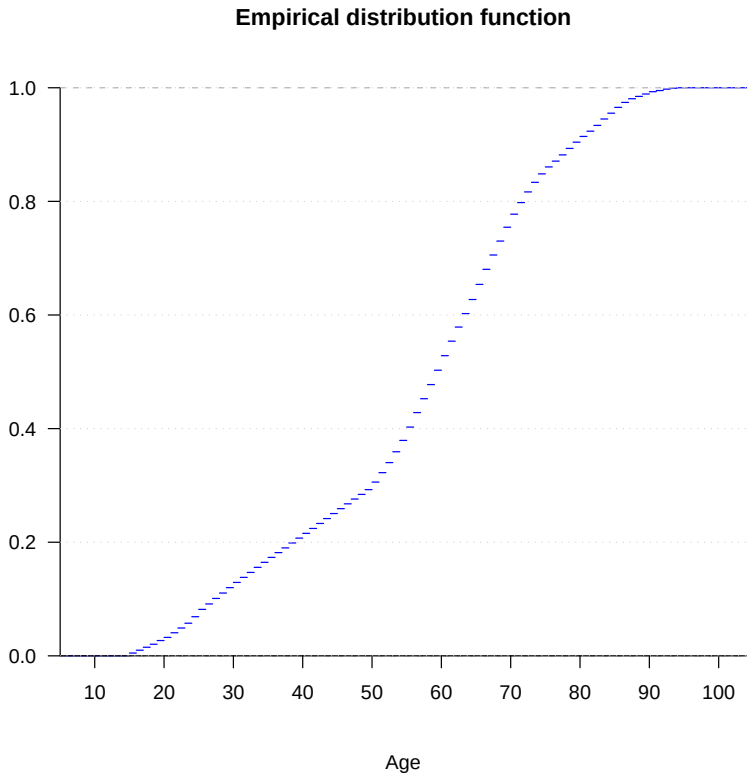
mean	mean(z)
median	median(z)
Emp. variance	var(z)
standard deviation	sd(z)

- sample quantile: quantile (z,0.1)
  sample quartile: quantile(z,seq(0,1,0.25))
- distribution quantile: q*name of distribution*

Quartiles and percentiles: Data Example

- data age of consumer
- **summary**(z) yields

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.00	45.00	59.00	56.72	69.00	94.00

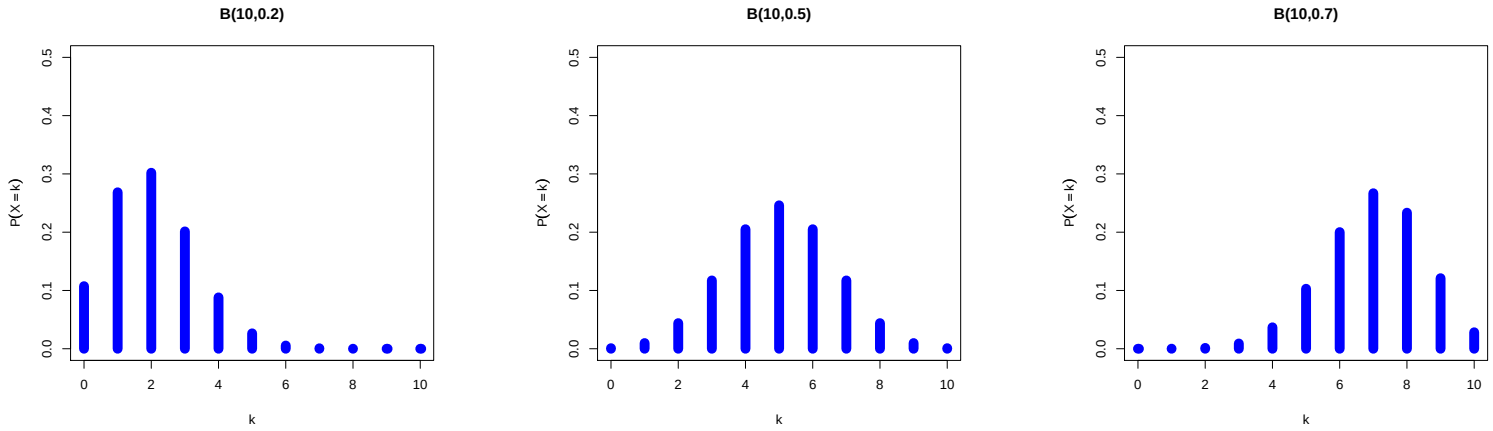


Skewness (Schiefe)

- Skewness of a distribution characterized by normalized central third moment

$$\frac{\mathrm{E}(Z - \mathrm{E}Z)^3}{(\mathrm{Var}\ Z)^{3/2}}$$

- example $\mathrm{B}(n, p)$ skewness: $\frac{1-2p}{\sqrt{np(1-p)}}$



(here: 0.474, 0, -0.276)

Skewness

- symmetric distributions: skewness = 0
- **empirical skewness**

$$\frac{\frac{1}{n}\sum(z_i - \overline{z})^3}{(\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})^2)^{3/2}}$$

- not unbiased, but consistent estimator

Linear dependence of two r.v.'s

- random variables X and Y with **numerical values**, not categories

$$\rho = \mathbf{Cor}(X, Y) = \frac{\mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)]}{\sqrt{\mathbf{Var}\ (X)\ \mathbf{Var}\ (Y)}}$$
$$= \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}\ (X)\ \mathbf{Var}\ (Y)}}$$

Correlation (coefficient)

- measure for **linear** dependence; suppose
  $Y = aX + b$ $\quad a, b$ constant
  $\Rightarrow \rho = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$
- We have $|\rho| \leq 1$

Empirical correlation: Estimator for ρ

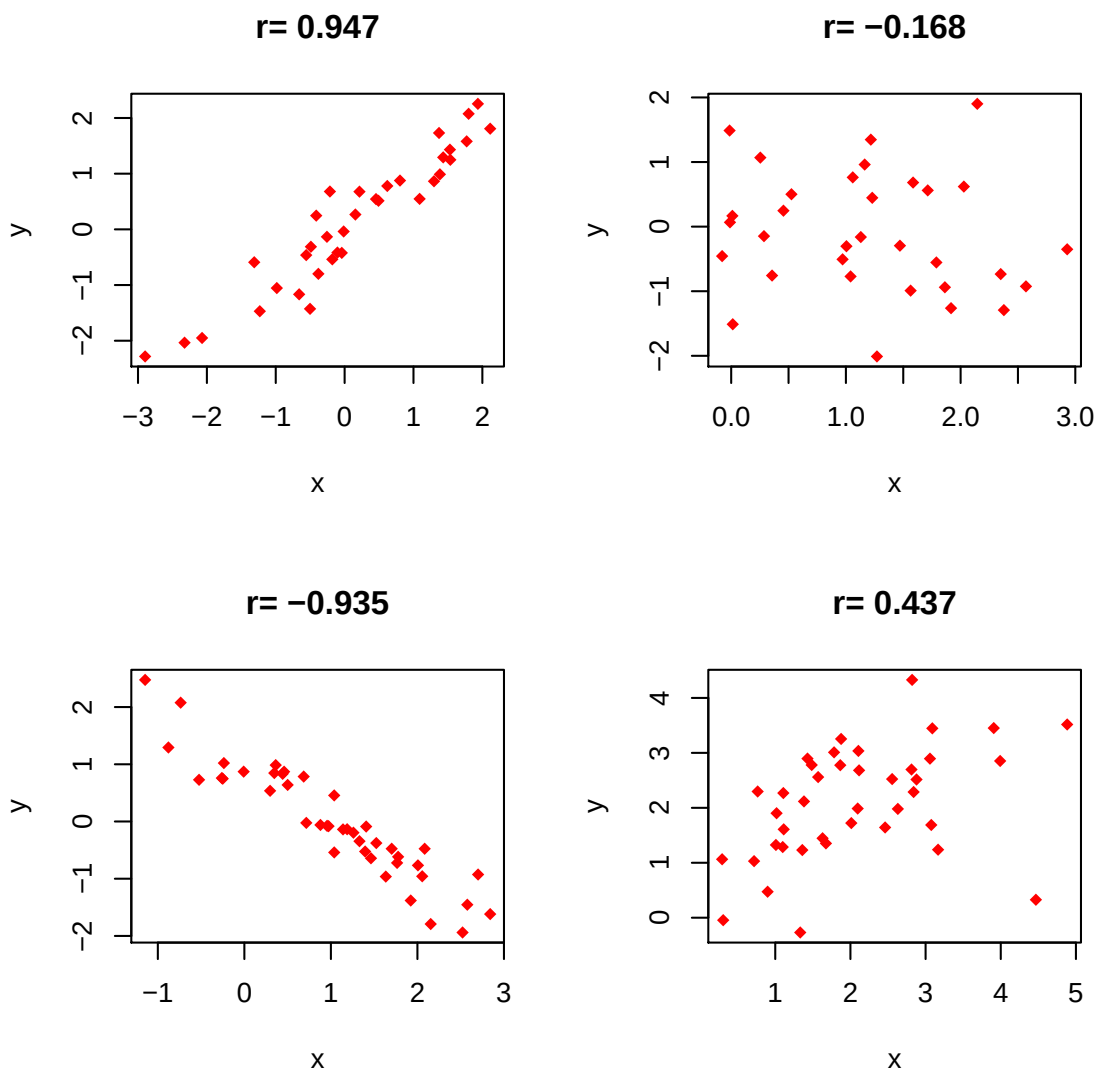
- data $(x_i, y_i) \quad i = 1, \ldots, n$

Definition (Empirical correlation)

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{s_x\, s_y}$$

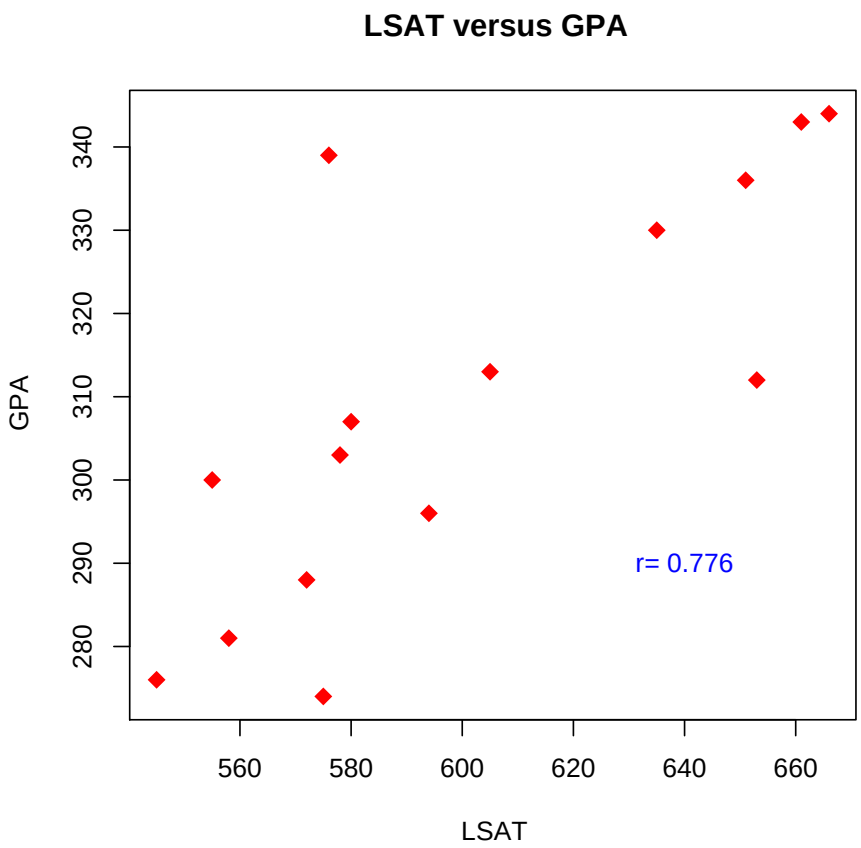
$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$

- Properties:
  If $y_i = ax_i + b$ for all $i \quad \Rightarrow \quad r = \pm 1$
  r is a consistent estimator ($r \to \rho$ as $n \to \infty$.)
  (Numerator is an unbiased estimator for the covariance of X and Y)


r= 0.947
r= −0.168
r= −0.935
r= 0.437
x
y

Example: Law school data

A random sample of size n=15 from the universe of 82 USA law schools. Two measurements: LSAT (average score on a national law test) and GPA (average undergraduate grade-point average).



Dependence of ordinal data - Ranks

- for definition: consider one-dimensional data $z_1, \ldots, z_n$

$$z_{1:n} \leq z_{2:n} \leq \ldots \leq z_{n:n}$$

rank $\boldsymbol{R_i = R(z_i)}$ (Rang)

i	1	2	3	4	5	6	7	8
z_i	8	5	10	3	9	2	7	1
$z_{i:n}$	1	2	3	5	7	8	9	10
R_i	6	4	8	3	7	2	5	1

Ranks

- If there are equal values in the data set....(ties, Bindungen)

i	1	2	3	4	5	6	7	8
z_i	8	8	10	3	9	2	7	1
$z_{i:n}$	1	2	3	7	8	8	9	10
R_i	5.5	5.5	8	3	7	2	4	1

(means of the corresponding ranks)

- in R: `rank(x,ties.method="average")`

Dependence of ordinal data

- back to two-dimensional data (x_i, y_i)
  determine the ranks for each component

$$R(x_i) \text{ und } R(y_i)$$

- We have

$$\sum_{i=1}^{n} R(x_i) = \sum_{i=1}^{n} R(y_i) = \frac{n(n+1)}{2}$$

  and for the means of the ranks

$$\bar{R}(x) = \frac{1}{n}\sum_{i=1}^{n} R(x_i) = \bar{R}(y) = \frac{n+1}{2}$$

Dependence of ordinal data: Spearman's rho

- Compute the (usual) correlation of the ranks, i.e.

Definition (Spearman rank correlation)

$$
\begin{aligned}
r_S &= r_{R(x)R(y)} \\
&= \frac{\sum_{i=1}^{n}(R(x_i) - \bar{R}(x))(R(y_i) - \bar{R}(y))}{\sqrt{\sum_{i=1}^{n}(R(x_i) - \bar{R}(x))^2 \sum_{i=1}^{n}(R(y_i) - \bar{R}(y))^2}}
\end{aligned}
$$

- transformations $\Rightarrow$

Spearman's rho

$$r_S = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - \frac{n}{4}(n+1)^2}{\sqrt{\left(\sum_{i=1}^{n} R^2(x_i) - \frac{n}{4}(n+1)^2\right)\left(\sum_{i=1}^{n} R^2(y_i) - \frac{n}{4}(n+1)^2\right)}}$$

- We have:

$$-1 \leq r_S \leq 1.$$

Spearman's rho

- $r_S = 1$ if and only if the ranks are equal
  $R(x_i) = R(y_i)$ for $i = 1, \ldots, n$
- $r_S = -1$ – the ranks are completely inverse (increase of the x-values means decrease of the y-values and vice versa)
- r_S close to 1 – strong positive rank correlation
  If x_i has a high (small) rank then also y_i has a high (small) rank (mostly).
- r_s close to -1 – strong negative rank correlation
  High (small) ranks of x-values correspond to small (high) ranks of y-values.

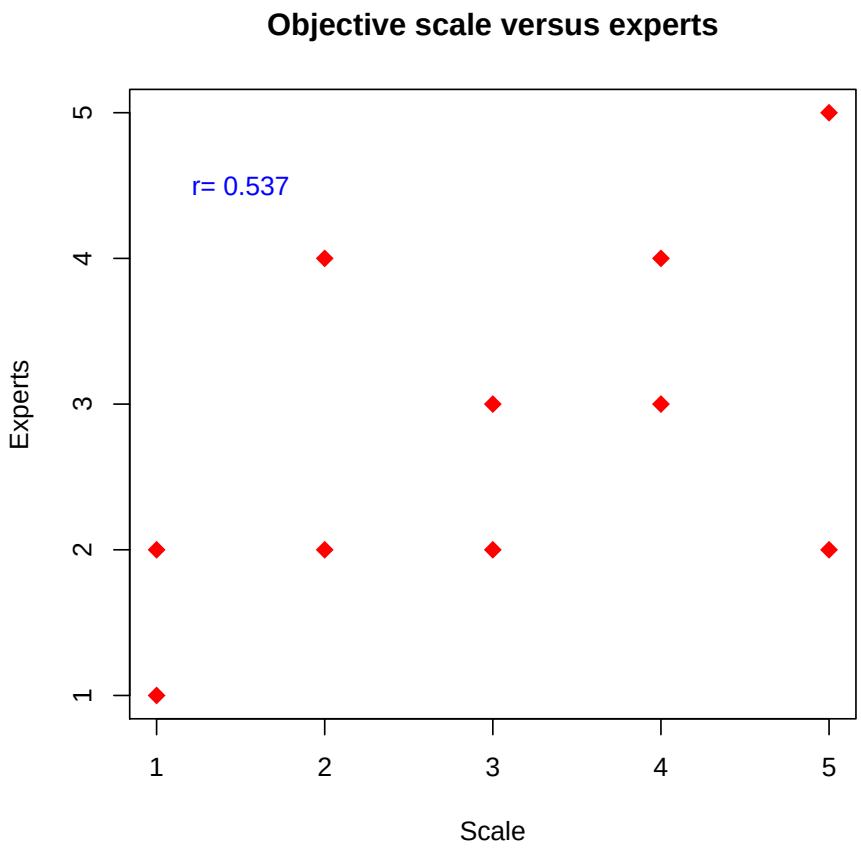
Example: Spearman's rho

- According to a given scale the quality of certain products is evaluated (x). The same products are evaluated by a group of experts (y).
- The following points were given:

Product	1	2	3	4	5	6	7	8	9	10
x_i	1	3	4	2	2	5	5	4	3	1
y_i	1	2	3	2	4	2	5	4	3	2

- The following figure shows little concordance:

Example


Objective scale versus experts
r= 0.537
Experts
Scale
1
2
3
4
5
1
2
3
4
5

Example

```
x<-c(1,3,4,2,2,5,5,4,3,1)
y<-c(1,2,3,2,4,2,5,4,3,2)
x.r<-rank(x);y.r<-rank(y)
r.s<-cor(x.r,y.r)
plot(x,y,main="Objective scale versus experts",
     pch=18,col=2,cex=1.5,
     xlab="Scale",ylab="Experts")
text(1.5,4.5,paste("r=",round(r.s,3)),col=4)
```

Maximum Likelihood Estimator

- Assumption: special type of distribution
- in general: estimators with good properties

Toy example

- Box with 10 balls, red and black;
- number of red balls ϑ (unknown)
- 3 balls are drawn: 2 red and 1 black $\Rightarrow$
  know: possible values for ϑ

$$2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9$$

- no value can be excluded, however some are more likely than others
- **Computation of the probability of the observed event**
  Z number of red balls in (first) drawing (of 3 balls) **$\mathbf{P}(Z = 2)$ for different values of ϑ**:

Toy example

- $$\mathsf{P}_\vartheta(Z=2) = \frac{\binom{\vartheta}{2}\binom{10-\vartheta}{1}}{\binom{10}{3}} = \frac{\vartheta(\vartheta-1)(10-\vartheta)}{240}$$
- We have for $p_\vartheta(2) = \mathsf{P}_\vartheta(Z=2)$

ϑ	0	1	2	3	4	5	6	**7**	8	9	10
$p_\vartheta(2)$	0	0	.067	.175	.3	.417	.5	**.0.525**	.467	.3	0

- Suppose $\vartheta = 2$, then the observed event has a relative small probability;
  for the value $\vartheta = 7$ the probability for a observation is maximal; it is plausible to choose as estimate: $\boldsymbol{\hat{\vartheta} = 7}$

Toy example

- three balls back, and repeat the experiment;
  result: 3 black balls
- Computation of the probability of both observations:
- $$L_\vartheta := \mathrm{P}_\vartheta(Z_1 = 2, Z_2 = 0) \;=\; \mathrm{P}_\vartheta(Z_1 = 2) \cdot \mathrm{P}_\vartheta(Z_2 = 0)$$
- $$\mathrm{P}_\vartheta(Z_2 = 0) \;=\; \frac{\binom{\vartheta}{0}\binom{10-\vartheta}{3}}{\binom{10}{3}} \;=\; \frac{(10-\vartheta)(9-\vartheta)(8-\vartheta)}{720}$$
-

ϑ	0	1	2	**3**	4	5	6	7	8	9	10
L_ϑ	0	0	.031	**.051**	.05	.035	.017	.004	0	0	0

Toy example

- most likely value: $\hat{\vartheta} = \mathbf{3}$
- idea is also applicable in the case that the unknown parameter takes infinite many value

General idea:
A maximum likelihood estimate is a parameter value under which the data (observation) actually obtained have the highest probability.

- likelihood = probability

MLE - General method

Data:

$$z_1, \ldots, z_n$$

- **Likelihood function for discrete r.v.'s** with single probabilities

$$\mathrm{P}(Z = \cdot) = p(\cdot; \vartheta), \qquad j = 1, \ldots, k$$

$$L(\vartheta) = L(\vartheta; z_1, \ldots, z_n) \; = \; \prod_{i=1}^{n} p(z_i; \vartheta)$$

- **Likelihood function for continuous r.v.'s** with density $f(\cdot; \vartheta)$

$$L(\vartheta) = L(\vartheta; z_1, \ldots, z_n) \; = \; \prod_{i=1}^{n} f(z_i; \vartheta)$$

m.l.e. - General method

Definition

A maximum likelihood estimate of ϑ is any value $\hat{\vartheta}$ that maximizes the likelihood function ***L****, i.e.*

$$\boldsymbol{L}(\hat{\vartheta}) \geq \boldsymbol{L}(\vartheta) \quad \textit{for all } \vartheta$$

m.l.e. (MLS)

- how to compute?
  finite number of values of ϑ directly
  infinite many: take log L (log likelihood); often: maximum by differentiating

Example: Normal distribution

- $Z_1, \ldots, Z_n$ independent $\mathrm{N}(\mu, \sigma^2)$-distr.; $\vartheta = (\mu, \sigma^2)$
- **likelihood function:** product of densities

$$
\begin{aligned}
L(\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_i - \mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^{n} \frac{(z_i - \mu)^2}{2\sigma^2}\right)
\end{aligned}
$$

- **log likelihood function**

$$
l(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = -\frac{n}{2}(\ln(2\pi) + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \mu)^2
$$

m.l.e. Normal distribution

- **Maximize**

$$l(\mu, \sigma^2) = -\frac{n}{2}(\ln(2\pi) + \ln \sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i - \mu)^2$$

- **system of equations**

$$\frac{\partial l(\mu, \sigma^2)}{\partial \mu} \quad : \quad \frac{2}{2\sigma^2}\sum_{i=1}^{n}(z_i - \mu) = 0$$

$$\frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} \quad : \quad -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(z_i - \mu)^2 = 0$$

- **solution:** $\hat{\mu}_{m.l.e} = \frac{1}{n}\sum_{i=1}^{n} z_i \quad \hat{\sigma}_{m.l.e} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2$

m.l.e. Binomial distribution

- $Z_1, \ldots, Z_n$ independent $\mathrm{B}(1, p)$-distributed; values 0 and 1
- **likelihood function:** product of the single probabilities

$$\begin{aligned} L(p) &= \prod_{i=1}^{n} p^{z_i}(1-p)^{1-z_i} \\ &= p^{\sum_i z_i}(1-p)^{n-\sum_i^n z_i} \end{aligned}$$

- **log likelihood function**

$$l(p) = \ln L(p) = \sum_i z_i \ln p + (n - \sum_i z_i) \ln(1-p).$$

m.l.e. Binomial distribution

- **Maximize**

$$l(p) = \ln L(p) \;=\; \sum_i z_i \ln p \;+\; (n - \sum_i z_i) \ln(1-p)$$

- **Equation**

$$\frac{\partial l(p)}{\partial p} \quad : \quad \frac{\sum_i^n z_i}{p} - \frac{n - \sum_i^n z_i}{1-p} \;=\; 0$$

- Solution: $\hat{p}_{m.l.e} = \frac{1}{n} \sum_{i=1}^{n} z_i$ **relative frequency**

m.l.e. Continuous uniform distribution

- $Z_1, \ldots, Z_n$ independent **uniformly distributed** over the interval $[0, \vartheta]$[1]

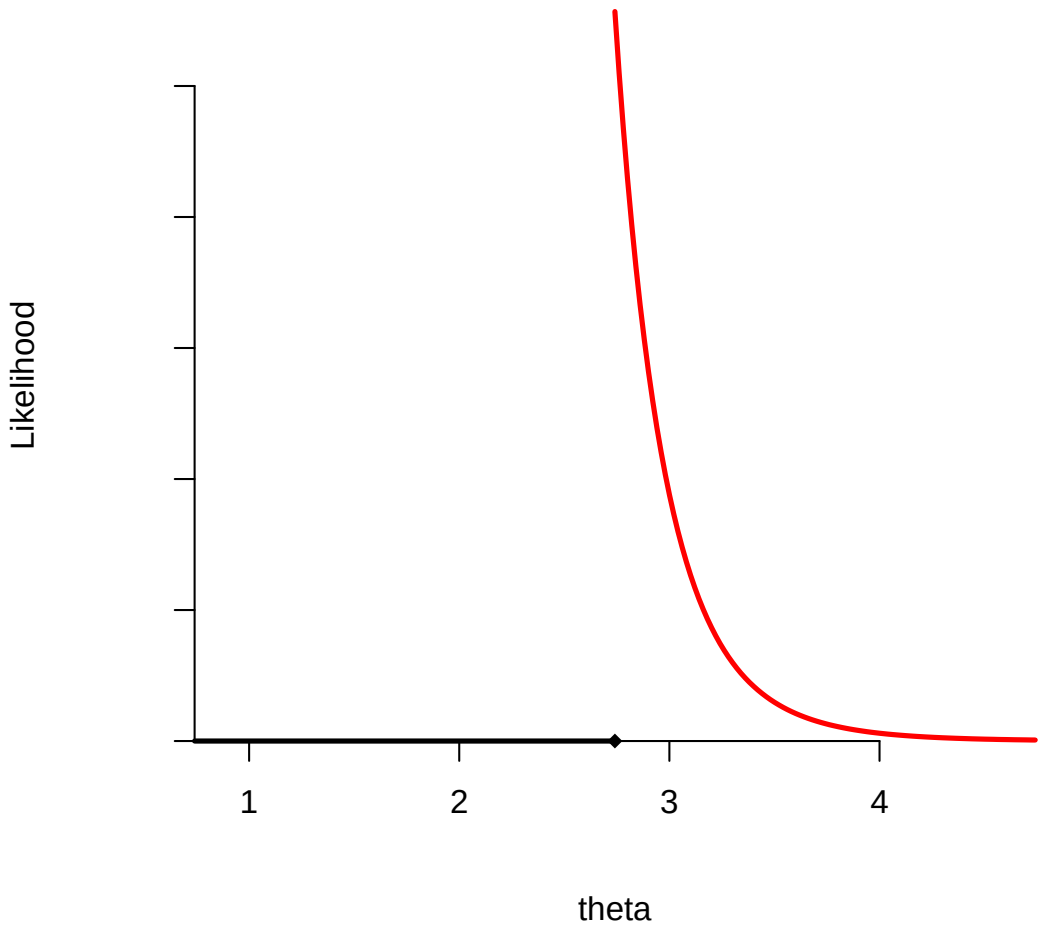
$$f(z_i) = \begin{cases} \frac{1}{\vartheta} & \text{if } z_i \in [0, \vartheta] \\ 0 & \text{otherwise} \end{cases}$$

- likelihood function: product of the densities

$$L(\vartheta) = \begin{cases} \frac{1}{\vartheta^n} & \text{if } z_i \in [0, \vartheta] \text{ for } \textbf{all } i \\ 0 & \text{otherwise} \end{cases}$$

- **Solution not by differentiating!**

[1] Important for generation of random numbers U[0,1].


Likelihood
1
2
3
4
theta

m.l.e. Continuous uniform distribution

- The figure shows the likelihood function for 12 data; the maximal value is 1.6. $\Rightarrow \quad \vartheta_{m.l.e.} = \mathbf{2.74} = z_{\max}$
- One can prove:

$$\mathbf{E}Z_{\max} = \frac{n}{n+1}\vartheta$$

The corrected unbiased estimator is

$$\frac{n+1}{n} Z_{\max}$$

Estimate (example): 2.97

m.l.e. Discrete uniform distribution

- **discrete** analog: **Uniform distribution** over the natural numbers $\{1, \ldots, M\}$ with unknown M. (Taxi problem)
  asymptotical unbiased estimator, resulting from the m.l.e.:

$$\hat{M} = \frac{n+1}{n} Z_{\max},$$

  where n is the sample size and $Z_{\max}$ is the largest observed number.
- We have

$$\mathrm{E} Z_{\max} = M - \frac{1}{M^n} \sum_{k=1}^{M} k^n.$$

m.l.e. Exponential distribution

- $Z_1, \ldots, Z_n$ independent $\mathrm{Exp}(\lambda)$-distributed
- density $f(z) = \lambda \exp(-\lambda z)$ $\mathsf{E}Z = 1/\lambda$
- **likelihood function:** product of densities

$$L(\lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} z_i)$$

- **log likelihood function**

$$l(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} z_i$$

m.l.e. Exponential distribution

$$\frac{\partial l(\lambda)}{\partial \lambda} : \quad \frac{n}{\lambda} - \sum_{i=1}^{n} z_i = 0$$

- Solution:

$$\hat{\lambda}_{m.l.e} = \frac{n}{\sum_{i=1}^{n} z_i} = \frac{1}{\overline{z}}$$

- <u>Remark!!</u> If the parameter of interest is $\gamma = \mathbf{P}(Z > t_0)$ i.e.

$$\gamma = 1 - F(t_0) = \exp(-\lambda t_0)$$

- m.l.e is given by

$$\hat{\gamma}_{m.l.e} = \exp(-t_0/\overline{z})$$

m.l.e. Exponential distribution

- another estimator for $\gamma = 1 - F(t_0)$ is given by **empirical distribution function** $\hat{F}_n$

$$\tilde{\gamma} = 1 - \hat{F}_n(t_0)$$

  where $\hat{F}_n$ is the empirical df
- **better than** $\hat{\gamma} = \exp(-t_0/\overline{Z})$**?**
  If assumption "Exponential distribution" is justified then $\hat{\gamma}$ is better!
  If there are some doubts whether the data are values of an exponential distribution then it is better to apply $\tilde{\gamma}$.

  This estimator is good for all distributions; (does not use Exp.-assumption)

Computation in R

- package: **Stats4**

1. procedure: ***mle*** (minuslogl)
   (negative log likelihood function
   for example ***d...***)
2. or directly define as function negative log likelihood function and apply procedure ***nlm*** (minimization of a function with Newton Raphson)

Definition of a confidence interval

- instead of a plausible value for an unknown parameter consider an **interval of plausible values**
- this interval **covers** the unknown parameter with a chosen **degree of confidence**; with a given probability
  confidence interval
- the interval is random; depends on the data

Definition

An interval $\mathcal{I} = [\hat{\gamma}_L, \hat{\gamma}_U]$ is called a confidence interval for a parameter γ with confidence level $1 - \alpha$ (coverage probability) if

$$\mathbf{P}(\hat{\gamma}_L \leq \gamma \leq \hat{\gamma}_U) \geq 1 - \alpha$$

Konfidenzintervall zum Niveau $1 - \alpha$

Construction of confidence intervals

- $\hat{\gamma}_L, \hat{\gamma}_U$ are lower and upper interval bounds
- $1 - \alpha$ confidence level, coverage probability;
  usual values for α: 0.05, 0.1, 0.01
- for the construction of a confidence interval one has to assume
  - a special distribution
  - large sample size *n* (approximation by the central limit theorem)

Confidence interval: Normal distribution

- $Z_1, \ldots, Z_n$ independent $\mathrm{N}(\mu, \sigma_0^2)$-distr.; σ_0^2 known
- $\overline{Z}$ good estimator; plausible values are around $\overline{Z}$;
- $\overline{Z} \sim \mathrm{N}(\mu, n^{-1}\sigma_0^2)$ equivalent:

$$\frac{\sqrt{n}(\overline{Z} - \mu)}{\sigma_0} \sim \mathrm{N}(0, 1) \quad \mathrm{P}\left(\frac{\sqrt{n}|\overline{Z} - \mu|}{\sigma_0} \leq u\right) = 2\Phi(u) - 1$$

- Choose u such that $\Phi(u) = 0.975$, d.h. $u = 1.96$ (0.975-quantile of $\mathrm{N}(0, 1)$) then

$$\mathbf{P}\left(\frac{\sqrt{n}|\overline{Z} - \mu|}{\sigma_0} \leq u\right) = 0.95$$

Confidence interval: Normal distribution

- Transformation of

$$\mathrm{P}\left(\frac{\sqrt{n}|\overline{Z}-\mu|}{\sigma_0} \leq u\right) = 0.95$$

leads to

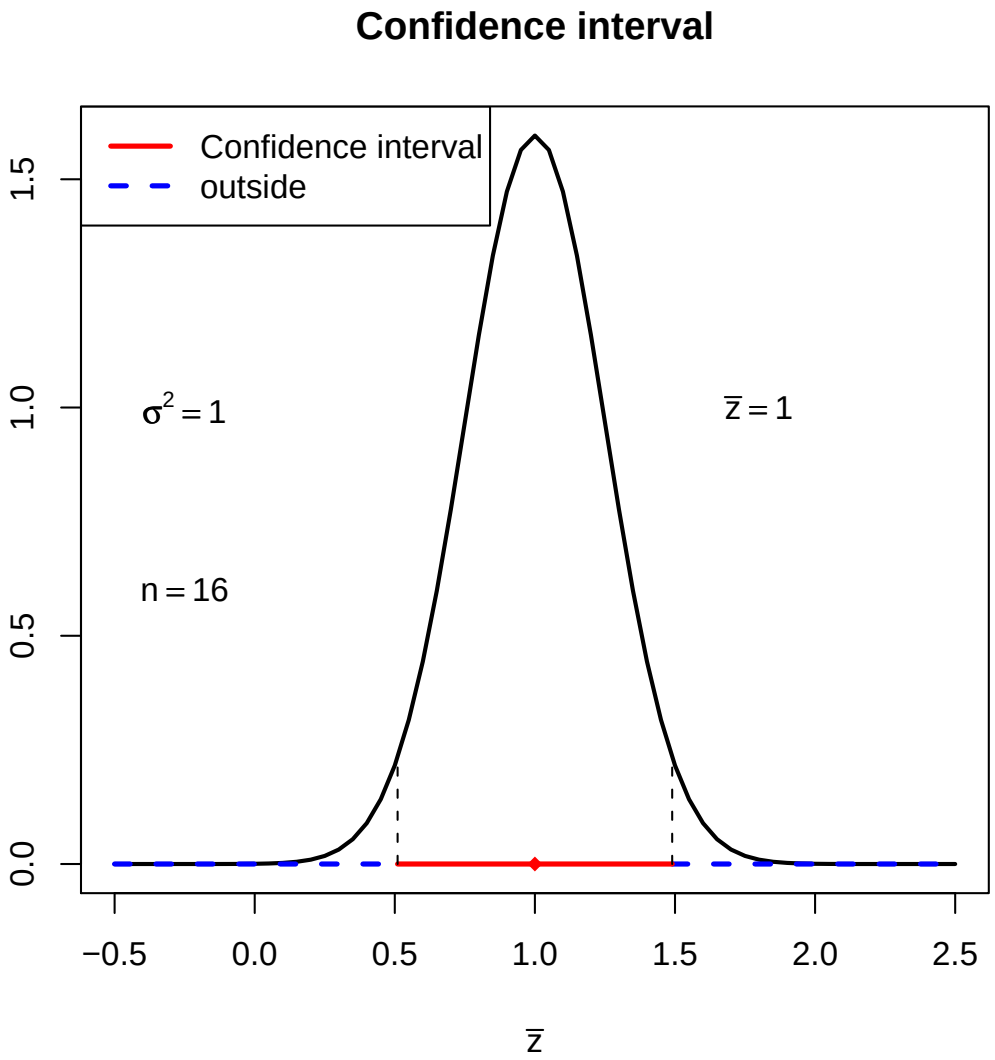
$$\mathrm{P}\left(\overline{Z} - 1.96\frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \overline{Z} + 1.96\frac{\sigma_0}{\sqrt{n}}\right) = 0.95$$

- confidence interval:

$$\boldsymbol{\mathcal{I}_n = [\overline{Z} - 1.96\frac{\sigma_0}{\sqrt{n}}, \overline{Z} + 1.96\frac{\sigma_0}{\sqrt{n}}]}$$

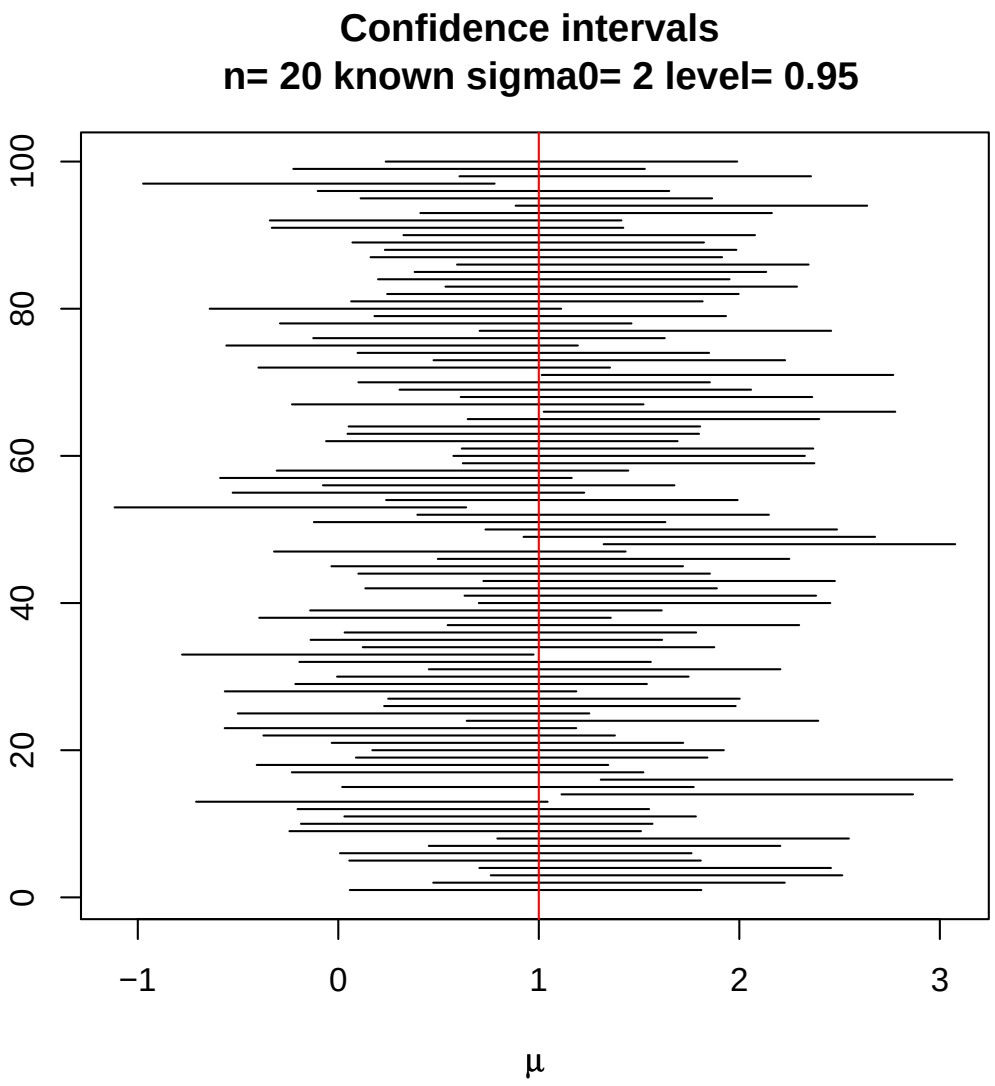
- for general α: instead $u_{0.975} = 1.96$

$$u_{1-\frac{\alpha}{2}} \quad \text{mit} \quad \Phi(u_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$


Confidence interval
Confidence interval
outside
$\sigma^2 = 1$
$\bar{z} = 1$
$n = 16$
1.5
1.0
0.5
0.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
$\bar{z}$

Confidence intervals interpretation

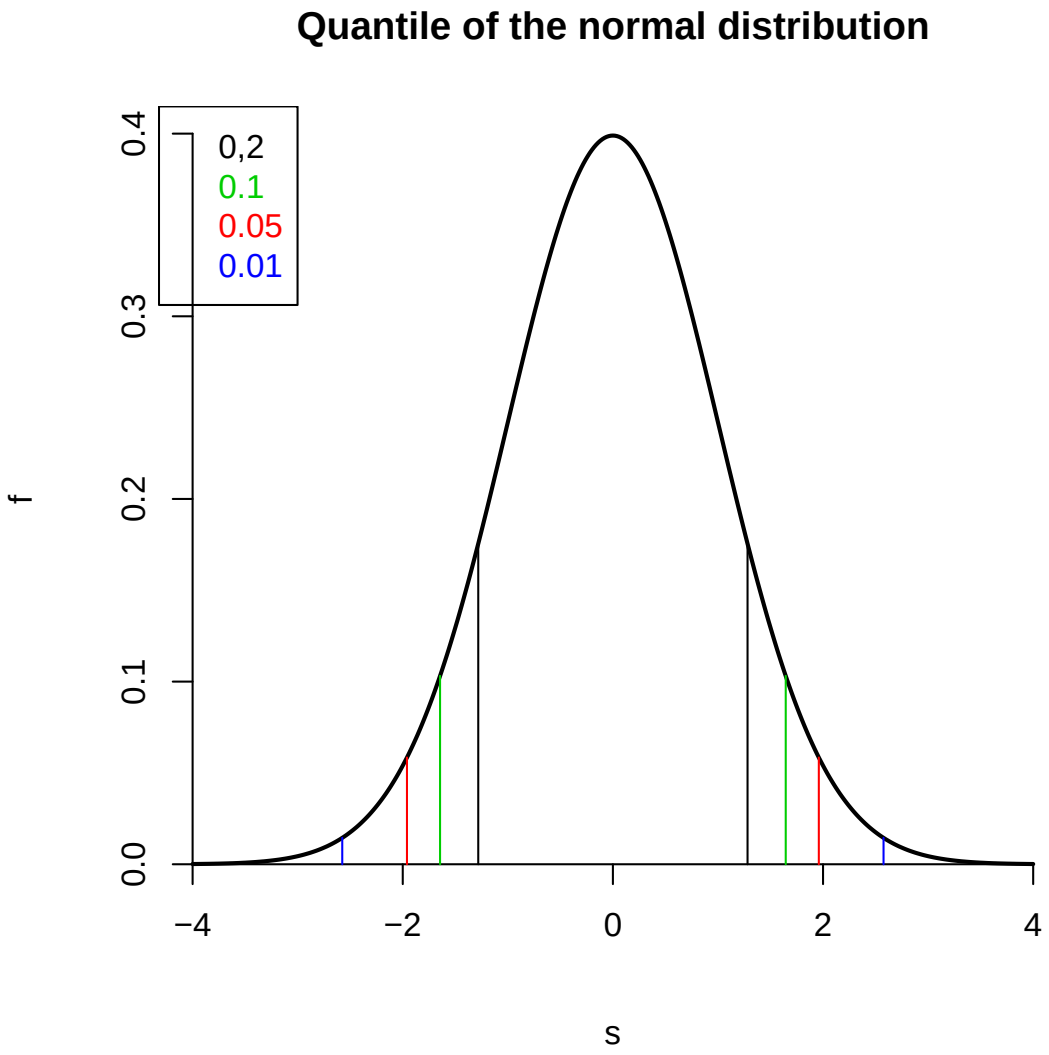
- **! NOT:** "μ lies with probability 0.95 in $\mathcal{I}_n$"
- means: If one applies the proposed procedure to "all possible samples" of $\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma_0^2})$-distributed data to construct a confidence interval then 95% of the resulting confidence intervals would cover the parameter $\boldsymbol{\mu}$
- *The confidence level refers to the method used to construct the interval rather than to any particular interval, such as the one we obtained.*
- α is also called error probability (Irrtumswahrscheinlichkeit)


Confidence intervals
n= 20 known sigma0= 2 level= 0.95
100
80
60
40
20
0
−1
0
1
2
3
μ

Confidence interval: Example

Suppose that we have data with mean 0.2 (resulting from samples of size 10, 50, 200)

	α		
n	0.05	0.01	0.001
10	$[0.138, 0.261]$	$[0.118, 0.281]$	$[0.10, 0.304]$
50	$[0.172, 0.236]$	$[0.163, 0.236]$	$[0.153, 0.246]$
200	$[0.186, 0.213]$	$[0.181, 0.218]$	$[0.176, 0.223]$


Quantile of the normal distribution
0,2
0.1
0.05
0.01
f
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
s

Confidence interval: Discussion

Length of the interval

$$\text{length} = 2u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}$$

1. α fixed, σ_0^2 fixed n larger $\implies$ shorter
2. α fixed, n fixed σ_0^2 larger $\implies$ larger
3. n fixed, σ_0^2 fixed α smaller $\implies$ larger

- To divide the length in half (fixed α and σ_0^2) you have to quadruple the sample size!

Confidence intervals: Normal distribution

- $Z_1, \ldots, Z_n$ independent $\mathrm{N}(\mu, \sigma^2)$-distr.
- σ^2unknown replaced by **estimator S^2**
- result from probability theory:

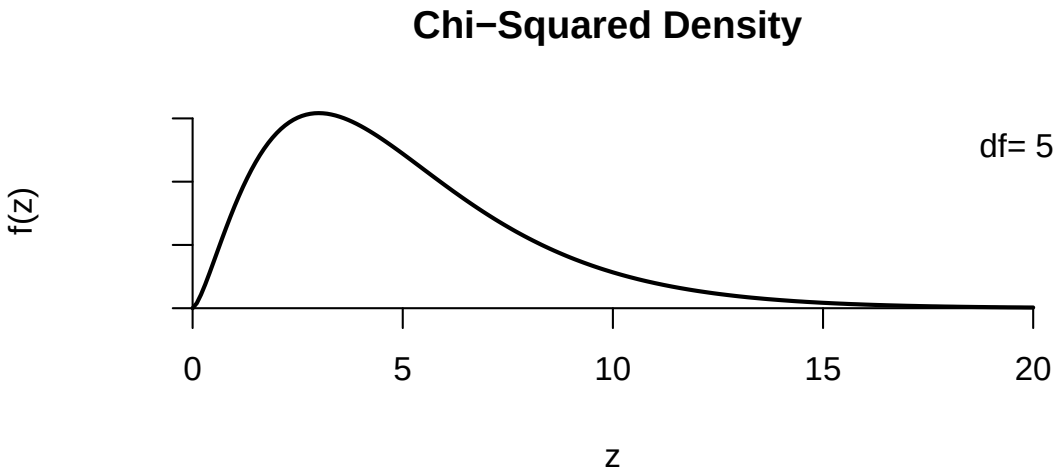
$$\overline{Z} \sim \mathrm{N}(\mu, n^{-1}\sigma^2) \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

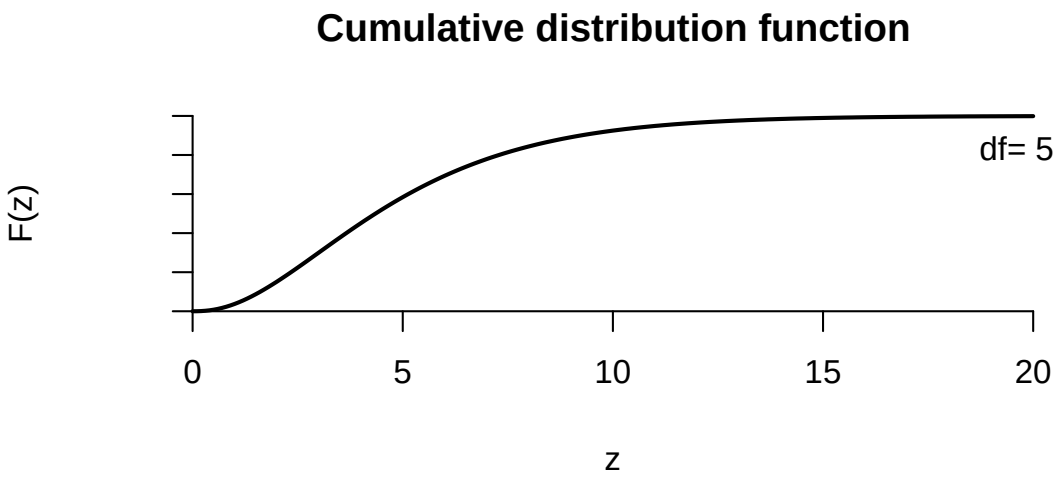
$\overline{Z}$ und S^2 stochastically independent

$$\frac{\sqrt{n}(\overline{Z} - \mu)}{\sqrt{S^2}} \sim t_{n-1}$$

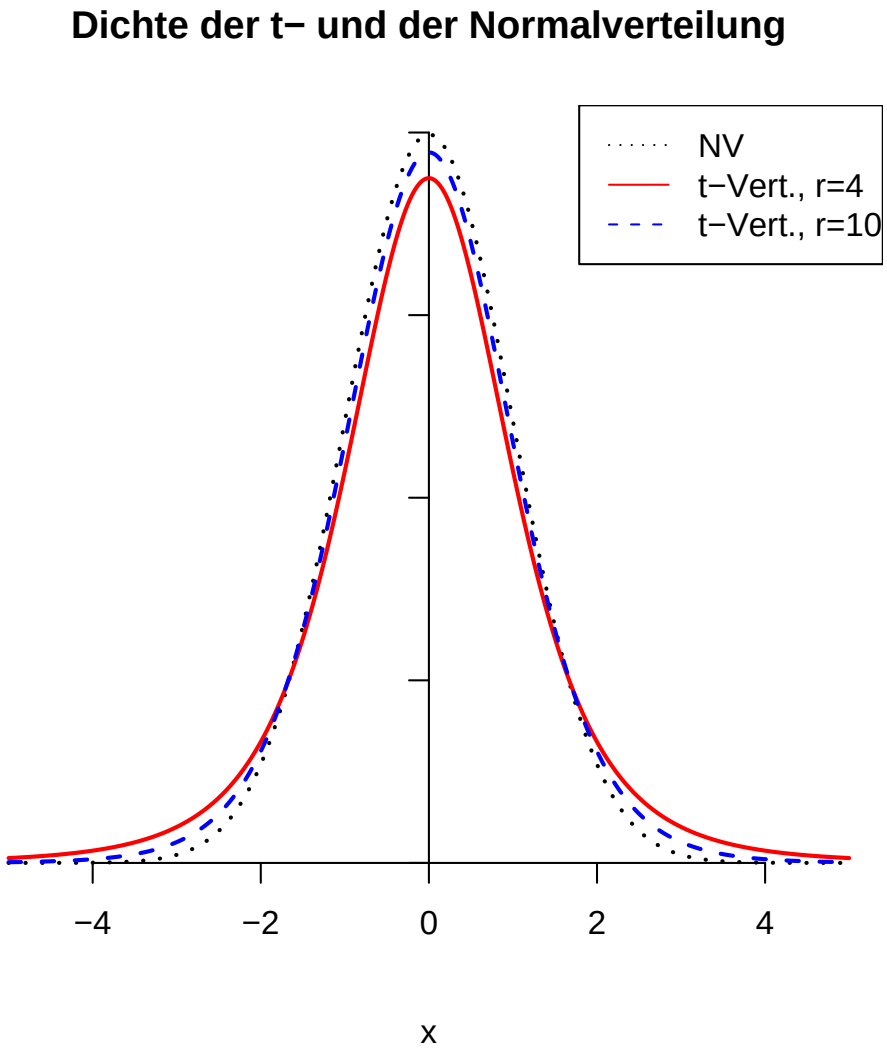
t-distributed with $n - 1$ degrees of freedom (df

t-distribution Student distribution

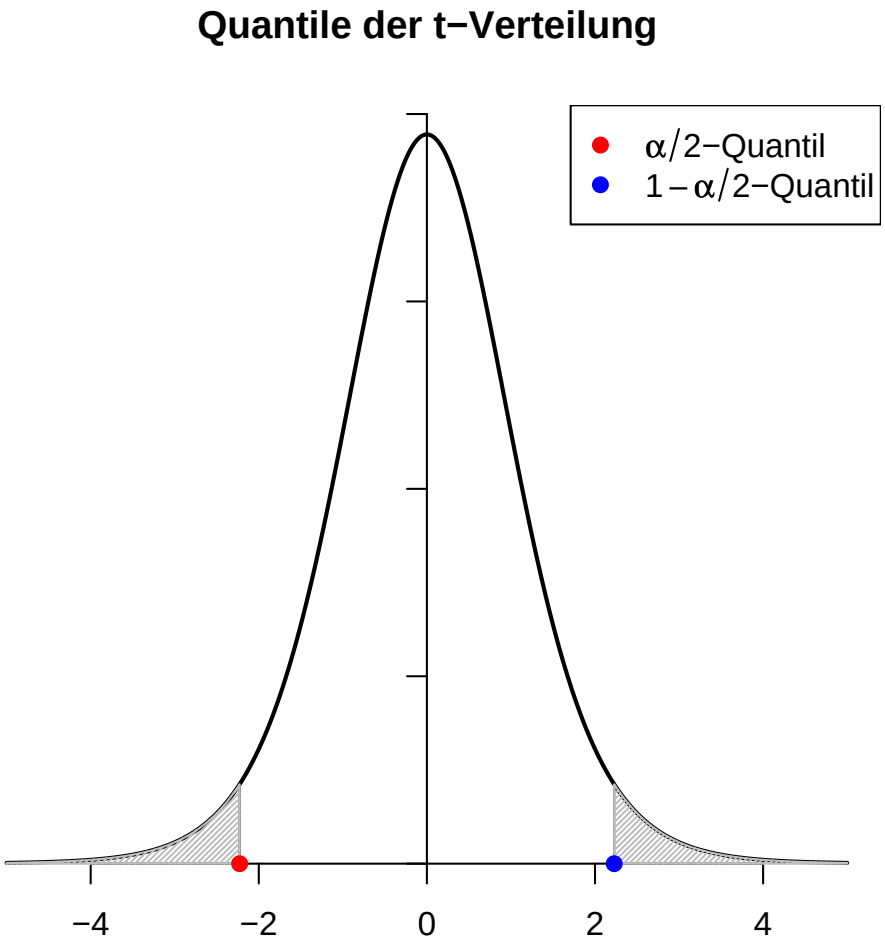




t-distribution Student distribution



t-distribution Student distribution



Flächeninhalt der schraffierten Fläche α

Confidence intervals: Normal distribution

- Confidence interval for μ

$$\mathrm{P}\left(\overline{Z} - t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{Z} + t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$$\mathcal{I}_n = [\overline{Z} - t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \overline{Z} + t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}]$$

- in comparison: interval with known σ^2

$$\mathcal{I}_n = [\overline{Z} - u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}}, \overline{Z} + u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}}]$$

Normal and t- distribution

- Quantile:

β	0.95	0.975	0.99	0.995
u_β	1.64	1.96	2.33	2.58
$t_{3;\beta}$	2.35	3.18	4.54	5.84
$t_{10;\beta}$	1.81	2.23	2.76	3.17
$t_{25;\beta}$	1.71	2.06	2.49	2.79
$t_{100;\beta}$	1.66	1.98	2.36	2.63

- Consequence: Intervals with **estimated variance** are longer; intuitively : **more uncertainty** than by known variance
  with **increasing sample size *n*** : difference becomes smaller; estimate more precise, **uncertainty disappears**

Remark: degree of freedom

- consider $\sum_{i=1}^{n}(Z_i - \mu)^2 = \sum_{i=1}^{n} Y_i^2$
  vector $Y = (Y_1, \ldots, Y_n)$ can take values in $\mathbb{R}^n$
- consider $\sum_{i=1}^{n}(Z_i - \overline{Z})^2 = \sum_{i=1}^{n} \tilde{Y}_i^2$
  Since
  $$\sum_{i=1}^{n} \tilde{Y}_i = \sum_{i=1}^{n} Z_i - n\overline{Z} = 0$$
  the vector $\tilde{Y} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)$ can take values only in $\mathbb{R}^{n-1}$.
  "A degree of freedom is lost."

Confidence interval: Normal distribution

- S^2 good estimator for σ^2
- $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$\Longrightarrow$ confidence interval for σ^2

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Approximative confidence intervals

- Basis: **Central limit theorem**: Let $Z_1, \ldots, Z_n$ be i.i.d. r.v.'s with $\mathrm{E}Z_i = \mu$ and Var $Z_i = \sigma^2$ (**no specification of the type of the distribution!**), then for large sample size n the distribution of $\overline{Z}$ can be approximated by the **normal distribution**.
- S^2 is a consistent estimator, $S^2 \to \sigma^2$.

$$\frac{\sqrt{n}(\overline{Z} - \mu)}{S} \overset{\text{approx}}{\sim} \mathbf{N(0,1)}$$

- **approximative confidence interval** $\mathcal{I}_n = [\overline{Z} - u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{Z} + u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}]$
- !!! Statistical software packages compute this interval automatically (without justification!)

Approximative confidence interval for a probability

- $Z_1, \ldots, Z_n$ B$(1, p)$ distr. r.v.'s
  $\overline{Z} = \hat{\vartheta}$ (relative frequency) $S^2 = \frac{n}{n-1}\hat{p}(1-\hat{p}) \approx \hat{p}(1-\hat{p})$
- **approximative confidence interval**

$$\mathcal{I}_n = \left[\hat{p} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

- Example: Public opinion poll: 250 persons 42 answer M $\Rightarrow$ rel. frequency. $\hat{p} = 0.168 \Rightarrow$ interval for $\alpha = 0.05$:
  [0.12 , 0.21];
  1000 persons; 168 answer M
  $\hat{p} = 0.168 \Rightarrow$ interval [0.14 , 0.19]

Approximative confidence interval for a probability

- Different question:**How many persons** have to be asked such that the confidence interval is not larger than $2 \cdot 0.025$ (5%)?
- p **completely unknown**
  $2 \cdot u_{1-\frac{\alpha}{2}} \cdot \sqrt{p(1-p)/n} \leq 2 \cdot 0.025 \quad u_{1-\frac{\alpha}{2}} = 1.96$
  worst case: $p = 0.5$
  $2 \cdot u_{1-\frac{\alpha}{2}} \cdot \sqrt{1/4n} \leq 2 \cdot 0.025 \quad \Rightarrow \quad n \geq \frac{1}{4}\left(\frac{1.96}{0.025}\right)^2 = \mathbf{1536.6}$
- known: **less than** 10% answer M $\Rightarrow$ $p \leq 0.1$
  $p(1-p) \leq 0.09 \quad \Rightarrow \quad n \geq 0.09\left(\frac{1.96}{0.025}\right)^2 = \mathbf{553.2}$

Confidence interval for the correlation ρ

- Assumption (X_i, Y_i) normally distributed, for $n \geq 50$

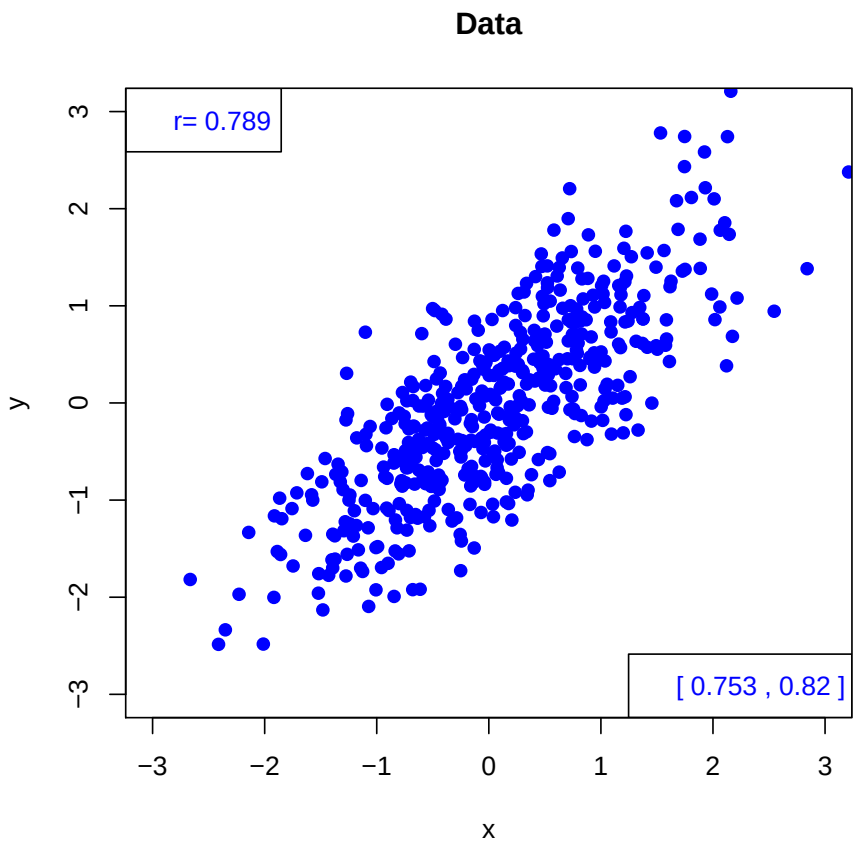
$$Z = \frac{\sqrt{n-3}}{2}\left(\ln\frac{1+r}{2} - \ln\frac{1+\rho}{1-\rho}\right) \sim_{asympt.} \mathrm{N}(0,1)$$

- With

$$A = \ln\frac{1+r}{1-r} - \frac{2u_{1-\alpha/2}}{\sqrt{n-3}} \quad B = \ln\frac{1+r}{1-r} + \frac{2u_{1-\alpha/2}}{\sqrt{n-3}}$$

it follows that an approximative confidence interval is given by

$$\mathrm{P}\left(\frac{\exp(A)-1}{\exp(A)+1} \leq \rho \leq \frac{\exp(B)-1}{\exp(B)+1}\right) \approx 1-\alpha$$


Data
r= 0.789
[0.753 , 0.82]
y
x
−3
−2
−1
0
1
2
3