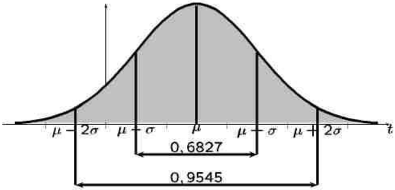
Statistics

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



Bachelor Studiengang Informatik

Prof. Dr. Egbert Falkenberg

Fachbereich Informatik & Ingenieurwissenschaften

Wintersemester 23/24

Section 1

Inferential Statistics

Introduction I

Example: Average height of all adults (over 18 years old) in the U.S.

- Population: all adults over 18 years of age in the U.S.
- Census: measure every adult and then compute the average $\rightarrow$ time-consuming and cost-intensive

Introduction II

- Using statistics:
  - Take a random sample and measure the heights
  - Conjecture that the average (**point estimation**) of the total population is “close to” the average of our sample
  - Calculate an interval containing the true value in for example 95% of all cases (**confidence interval**)
  - According to the Centers for Disease Control and Prevention Trusted Source [1], the average is 175.4 centimeters. Based on the sample value can we state that this value has been changed? (**hypothesis testing**)
- Goal of **inferential statistics**: use sample statistics to make inference about population parameters
- **Estimation** and **Hypothesis Testing** will be discussed in the following

[1] https://www.cdc.gov/nchs/data/nhsr/nhsr122-508.pdf, published in December 2018 based on data collected between 1999 and 2016

Characteristics of Estimators I

compare: Heumann, Schomaker 9.2

Notations: Let $x = \{x_1, x_2, ..., x_n\}$ be observations of a random sample from a population.

- Random sample $x = \{x_1, x_2, ..., x_n\}$ = realized values of a random variable X.
- More formally: x_i are realisations of independent and identically distributed (i.i.d) random variables X_i.
- Statistic = function of random variables. For example: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- A statistic $T(X)$ is used to estimate a parameter ϑ.
- $T(X)$ is called an estimator of ϑ.
- $\hat{\vartheta} = T(X)$ denote the estimate of ϑ using $T(X)$.
- $T(X)$ is a random variable but $T(x)$ is its observed value calculated from the sample $x = (x_1, x_2, ..., x_n)$.

Characteristics of Estimators II

Important characteristics of estimators:

bias and sampling variability

- Bias refers to whether an estimator tends to either over or underestimate the parameter.
- Sampling variability refers to how much the estimate varies from sample to sample.
- An estimator is biased if the long-term average value of the statistic is not equal to the parameter being estimated.
- More technically: biased if the expected value is not equal to the parameter to be estimated.

Example: A stopwatch that is a little bit fast gives biased estimates of elapsed time.

Characteristics of Estimators III

Definition: An estimator T(X) is unbiased if

$$E_\vartheta(T(X)) = \vartheta$$

The bias of an estimator $T(X)$ is defined as

$$Bias_\vartheta(T(X)) = E_\vartheta(T(X)) - \vartheta.$$

Remark: The index ϑ denotes that the expectated value is calculated with respect to the distribution whose parameter is ϑ.

Characteristics of Estimators IV

Remark: An unbiased estimator is not necessarily an accurate statistic.

- If a statistic is sometimes much too high and sometimes much too low, it can still be unbiased. It would be very imprecise, however.
- A slightly biased statistic that systematically results in very small overestimates of a parameter could be quite efficient.

Measure for the quality of an estimator: $E((T(X) - \vartheta)^2)$

Remark:

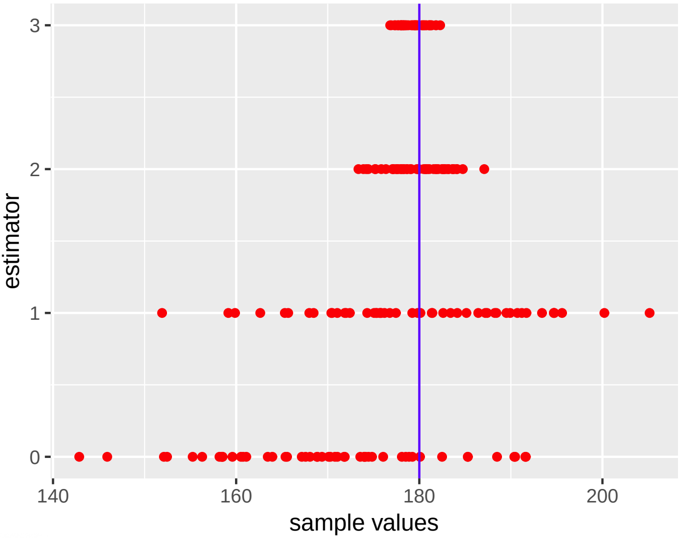
$$
\begin{aligned}
E_\vartheta((T(X) - \vartheta)^2) &= Var_\vartheta(T(X)) + (E_\vartheta(T(X)) - \vartheta)^2 \\
&= Var_\vartheta(T(X)) + Bias_\vartheta(T(X))^2
\end{aligned}
$$

$Var_\vartheta(T(X)) =$ Variance of the Estimator; a measure for the sampling variability

Characteristics of Estimators V

Example: different estimators for the paramter $\theta = 180$

estimator with different properties



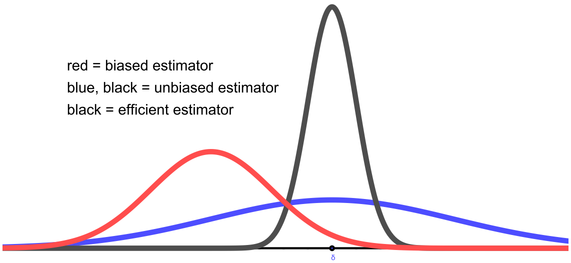
Characteristics of Estimators VI

Remark:

- Since an estimator is a random variable usually a new sample leads to a new estimate of θ.
- The estimators 0 and 3 are biased. The others are unbiased.
- The variability of estimator is 2 is much lower than the variability of estimator 1.
- Since the variability of estimator 3 is very low and the bias is not too high, it is the most accurate estimator of these 4 estimators.

Characteristics of Estimators VII

Objective: Find an unbiased estimator with smallest variance! Such an estimator is called efficient.



- Efficiency of a statistic decribes the precision of the estimate.
- The precision of a parameter estimator increases with the efficiency of a statistic.

Characteristics of Estimators VIII

Consistency of Estimators: If the estimator's values approach the estimated parameter as the sample size increases, we consider the estimator to be consistent.

Definition: Let $T_i = T_i(X_1, X_2, ..., X_i), i \in \mathbb{N}$ a sequence of estimators for the parameter ϑ. The sequence is a consistent sequence of estimators for ϑ if for every $\varepsilon > 0$

$$\lim_{n \to \infty} P(|T_n - \vartheta| < \varepsilon) = 1$$

As the sample size increases, the probability that T_n is getting closer to ϑ is approaching 1.

Characteristics of Estimators IX

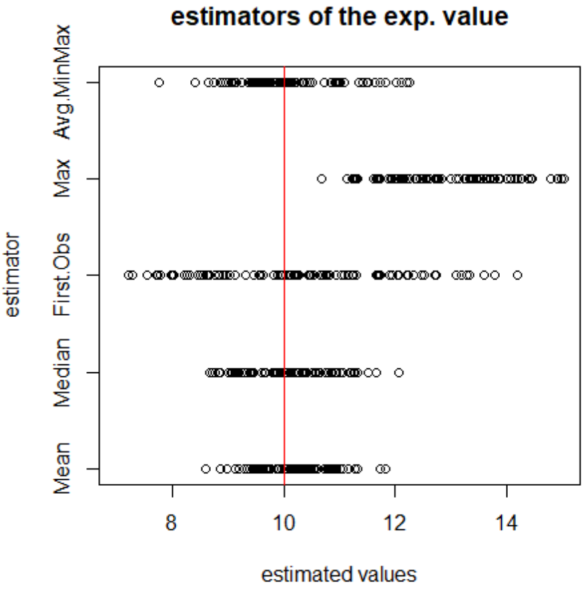
Example: A Bookseller operates a large number of stores. The expected monthly profit in 1000 Euro of a store is to be estimated. To do this, the monthly profit of 10 randomly selected stores is taken into account.

Observation of different estimators for expected monthly profit over the time: mean, median, first store, max value, average of the min and max value

The first 6 samples:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Mean	Median	First.Obs	Max	Avg.MinMax
1	13.51	9.92	15.15	10.27	13.94	10.73	10.73	12.54	13.14	10.42	12.03	11.64	13.51	15.15	12.53
2	14.05	14.00	12.61	12.43	9.10	11.78	10.35	11.73	8.37	10.23	11.47	11.75	14.05	14.05	11.21
3	10.38	9.67	11.91	9.74	9.60	9.01	10.74	11.42	8.39	7.43	9.83	9.70	10.38	11.91	9.67
4	9.04	9.67	11.44	14.03	8.21	11.00	10.77	12.19	8.02	11.87	10.62	10.89	9.04	14.03	11.03
5	11.90	11.64	10.76	10.75	11.13	11.70	9.77	11.13	12.31	9.03	11.01	11.13	11.90	12.31	10.67
6	7.84	8.40	10.13	10.49	8.98	8.63	7.96	7.02	9.48	11.33	9.03	8.81	7.84	11.33	9.18

Characteristics of Estimators IX



Questions

Which of the following statements are true or false?

t	f	
☐	☐	If you have estimated an unknown parameter in a sample it will not change in a new sample.
☐	☐	If an estimator deviates on average from the true value to be estimated, he has a bias.
☐	☐	The mean square deviation of an estimator from its true value is the bias of the estimator.
☐	☐	An efficient estimator is an unbiased estimator with smallest variance.
☐	☐	In the case of a consistent estimator, the estimates become with probability 1 more accurate as the sample size increases.

Maximum Likelihood Method I

Let $f(x;\theta)$ be the density function of $X_i, i = 1, 2, ..., n$.

$$L(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

density function of the random sample $X_1, X_2, ..., X_n$ is called **Likelihood-Function.**

Idea: Choose as an estimator $\hat{\theta}$ for θ the value which maximizes the Likelihood-Function, i.e.

$$L(x_1, x_2, ..., x_n; \theta) \leq L(x_1, x_2, ..., x_n; \hat{\theta}) \quad \text{for all } \theta$$

Remark: A maximum likelihood estimator is not necessary unbiased!

Maximum Likelihood Method II

Example: Urn with 10 black and white marbles. The number θ of black marbles is unknown.
1: 3 marbles are randomly drawn without replacement, X=2 marbles are black

θ	$P_\theta(X = 2)$
0	0
1	0
2	0.067
3	0.175
4	0.4
5	0.417
6	0.5
7	**0.525**
8	0.467
9	0.3
10	0

$$L(x_1;\theta) = P_\theta(X = 2) = \frac{\binom{\theta}{2}\binom{10-\theta}{1}}{\binom{10}{3}}$$

$$\hat{\theta} = 7$$

Maximum Likelihood Method III

2: The 3 drawn marbles are replaced and again 3 marbles are randomly drawn without replacement, X=0 marbles are black

θ	$P_\theta(X_2 = 0)$	$P_\theta(X_1 = 2) \cdot P_\theta(X_2 = 0)$
0	1	0
1	0.7	0
2	0.0467	0.031
3	0.292	**0.051**
4	0.167	0.050
5	0.083	0.035
6	0.033	0.017
7	0.008	0.004
8	0	0
9	0	0
10	0	0

$$\begin{aligned} L(x_1, x_2; \theta) &= P_\theta(X_1 = 2, X_2 = 0) = P_\theta(X_1 = 2) \cdot P_\theta(X_2 = 0) \\ &= \frac{\binom{\theta}{2}\binom{10-\theta}{1}}{\binom{10}{3}} \cdot \frac{\binom{\theta}{0}\binom{10-\theta}{3}}{\binom{10}{3}} \\ \hat{\theta} &= 3 \end{aligned}$$

Maximum Likelihood Method IV

Example: A bus will arrive in exactly θ minutes, but the time is unknown. You are arriving at the bus stop, randomly. Let T denote the the waiting time for the bus. T follows a uniform distribution ranging $[0, \theta]$, i.e. the density of T is $f(t; \theta) = \begin{cases} \frac{1}{\theta} & 0 \leq t \leq \theta \\ 0 & \text{else} \end{cases}$

If $T_1, T_2, ..., T_n$ are the waiting times at n days [2] the Likelihood-Function is

$$L(t_1, t_2, ..., t_n; \theta) = \prod_{i=1}^{n} f(t_i, \theta) = \begin{cases} 0 & \text{if one } t_i > \theta \\ \frac{1}{\theta^n} & \text{all } t_i \leq \theta \end{cases}$$

$$= \begin{cases} 0 & \max_i(t_i) > \theta \\ \frac{1}{\theta^n} & \text{else} \end{cases}$$

[2] we can assume that these are independent of each other

Maximum Likelihood Method V

- $L(t_1, t_2, ..., t_n; \theta)$ maximal if $\theta = \max_i(t_i)$, i.e. $T_{(n)} = \max(T_1, ..., T_n)$ is a Maximum Likelihood Estimator.
- Since $E(T_{(n)}) = \frac{n}{n+1}\theta$ and $\mathrm{Var}(T_{(n)}) = \frac{n}{(n+2)(n+1)^2}\theta^2$ the estimator $T_{(n)} = \max(T_1, ..., T_n)$ is biased but consistent.
- $\hat{T}_{(n)} = \frac{n+1}{n} T_{(n)}$ is an unbiased and consistent estimator.

Maximum Likelihood Method VI

Example: Simulation experiment

- Experiment: Measure the waiting times $T_1, T_2, ..., T_{10}$ at 10 days and $\theta = 5$
- estimator of θ: $E_i = \max(T_1, ..., T_i)$
- Sample: waiting times and esimators at one day

i	T_i	E_i
1	1.12	1.12
2	0.0964	1.12
3	4.18	4.18
4	3.73	4.18
5	0.894	4.18
6	4.61	4.61
7	4.90	4.90
8	2.87	4.90
9	1.81	4.90
10	3.54	4.90

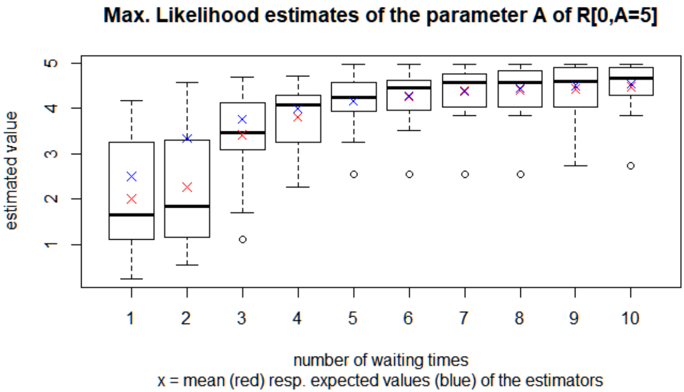
- Repeat this experiment 20 times:
  values of E_{10}: 4.90, 4.55, 4.70, 4.78, 4.28, 3.91, 4.63, 4.91, 4.93, 4.72, 4.30, 4.91, 4.89, 4.43, 4.97, 3.97, 2.74, 4.47, 3.84, 4.76

Maximum Likelihood Method VII

- some characteristic numbers of E_i

i	Min	q1	q2	Mean	q3	Max
1	0.247	1.12	1.66	2.00	3.26	4.18
3	1.10	3.14	3.46	3.42	4.11	4.68
5	2.55	3.94	4.25	4.15	4.56	4.97
7	2.55	4.06	4.57	4.40	4.77	4.97
10	2.74	4.29	4.66	4.48	4.89	4.97

- Boxplots over i=1,2,...,10



- Mention the biased but consistent estimator E_i

Estimating the Mean I

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from the distribution of a real-valued random variable X that has mean μ and standard deviation σ.

Estimator of μ: sample mean, defined by

$$\bar{X}_{(n)}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- $E(\bar{X}_{(n)}(X)) = \mu$, i.e. $\bar{X}_{(n)}$ is an unbiased estimator of μ.
- $\mathrm{Var}(\bar{X}_{(n)}(X)) = \frac{\sigma^2}{n}$, i.e. the estimator tends to get closer towards the parameter being estimated as the sample size increases. Therefore the estimator is a **consistent** estimator.
- $T_n = \sum_{i=1}^{n} \alpha_i X_i$ with $\sum_{i=1}^{n} \alpha_i = 1$ is an unbiased estimator for E(X). $\mathrm{Var}(T_n) = \mathrm{Var}(X)\sum_{i=1}^{n} \alpha_i^2$ is minimal if $\alpha_i = \frac{1}{n}$, i.e. $\bar{X}_{(n)}$ is an efficient estimator of E(X) in the set of linear estimators.

Estimating the Mean II

Some important special cases:

- $X(\omega) = I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{else} \end{cases}$ indicator variable for an event A with probability $P(A)$.

  $\bar{X}_{(n)}(X)$: relative frequency $P_n(A)$ of A

  $\Rightarrow P_n(A)$ unbiased and consistent estimator of $P(A)$.
- F distribution function of a real-valued random variable X

  For fixed x, the value $F_n(x)$ of the empirical distribution function is the sample mean for a random sample of size n from the distribution of the indicator variable $I_{X \le x}$.

  $\Rightarrow F_n(x)$ is a unbiased and consistent estimator of $F(x)$.

Estimating the Mean III

- X random variable with a discrete distribution on a countable set S

  f probability density function of X

  For fixed $x \in S$, the empirical probability density function $f_n(x)$ is the sample mean for a random sample of size n from the distribution of the indicator variable $I_{X=x}$.

  $\Rightarrow f_n(x)$ unbiased and consistent estimator of $f(x)$.

Estimating the Variance I

Let $X = (X_1, X_2, ..., X_n)$ be a random sample of size n from the distribution of a real-valued random variable X that has mean μ and standard deviation σ.

- μ is known (usually an artificial assumption)

  $$W_n^2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$ is an unbiased and consistent estimator of σ^2.

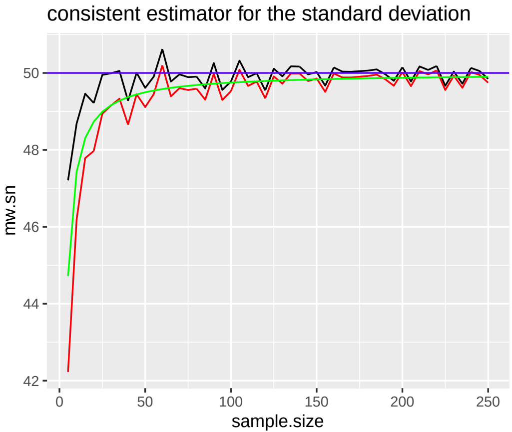
- μ is unknown (the more realistic assumption)

  $$S_n^2(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)}(X))^2$$ is an unbiased and consistent estimator of σ^2.

Remark: $\hat{W}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)}(X))^2$ is a biased maximum likelihood estimator of σ^2, it tends to underestimate σ^2.

Estimating the Variance II

- S_n unbiased and consistent estimator for σ
- $W_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ is a biased but consistent estimator for σ with $E(W_n^2) = \sigma^2 \cdot (1 - \frac{1}{n})$



- blue line: true value of σ
- black line: mean value of S_n
- green line: expected value of W_n
- red line: mean value of W_n

Questions

Which of the following statements are true or false?

t	f	
☐	☐	The arithmetic mean of a random sample $(X_1, ..., X_n)$ with i.i.d random variable X_i is an unbiased and consistent estimator for $E(X)$.
☐	☐	The arithmetic mean of a random sample $(X_1, ..., X_n)$ with i.i.d random variable X_i is an efficient estimator for $E(X)$.
☐	☐	The mean square deviation of the observed values from the sample mean in the random sample $(X_1, ..., X_n)$ with i.i.d random variable X_i is an unbiased estimator of the variance.
☐	☐	The empirical distribution is an unbiased estimator of the distribution function.

Introduction to Confidence Intervals I

- A confidence interval provides an estimated range of values that probably includes an unknown population parameter.
- It is being calculated from a given set of sample data.

Example: Mean heigt of male students at FHF

- It is impractical to weigh all male students.
- Sample of 25: mean height 177,52 cm
  177,52 is a point estimate of the population mean.
- A point estimate does not reveal the uncertainty associated with the estimate.
- Can you be confident that the population mean is within 5 cm of 177,52?

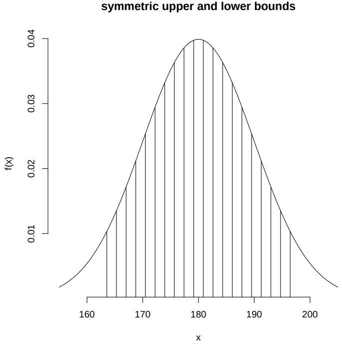
Introduction to Confidence Intervals II

Confidence intervals provide more information than point estimates:

- They are constructed using a procedure that will contain the unknown population parameter a specified proportion of the time, typically either 95% or 99% of the time.
- In case of i.i.d. samples a 95% confidence interval calculated for each sample 95% of the intervals will include the unknown population parameter.
- Width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter.

Confidence Intervals for Normal Distributions: parameter μ I

Example: Assume the height of male students follows a normal distribution with mean of $\mu = 180$ and a standard deviation of $\sigma = 10$.



$$\bar{X}_{(n)} \sim N(\mu, \frac{\sigma^2}{n})$$

We compute for a given α an upper bound o and lower bound u for the possible values of $\bar{X}_{(n)}$ thuch that:

$$P(u \leq \bar{X}_{(n)} \leq o) = 1 - \alpha$$

Confidence Intervals for Normal Distributions: parameter μ II

$$P(u \leq \bar{X}_{(n)} \leq o) = 1 - \alpha \Rightarrow \Phi(\frac{o-\mu}{\sigma/\sqrt{n}}) - \Phi(\frac{u-\mu}{\sigma/\sqrt{n}}) = 1 - \alpha$$

Let: $[u, o] = [\mu - \delta, \mu + \delta]$; u resp. o is the 2.5% resp. 97.5% quantil of $\bar{X}_{(n)}$.

$$\Phi(\frac{\delta}{\sigma/\sqrt{n}}) - \Phi(\frac{-\delta}{\sigma/\sqrt{n}}) = \Phi(\frac{\delta}{\sigma/\sqrt{n}}) - (1 - \Phi(\frac{\delta}{\sigma/\sqrt{n}})) = 1 - \alpha \Rightarrow$$

$$\Phi(\frac{\delta}{\sigma/\sqrt{n}}) = 1 - \alpha/2 \Rightarrow \delta = \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$$

$$\text{i.e. } u = \mu - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}, o = \mu + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$$

With $\mu = 180, \sigma = 10, n = 25, \alpha = 0,05$ we get $u_{1-\alpha/2} = u_{0.975} = 1.96$ and $u = 176.08, o = 183.92$

Confidence Intervals for Normal Distributions: parameter μ III

$$P(\mu - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} \leq \bar{X}_{(n)} \leq \mu + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}) = 1 - \alpha \Rightarrow$$

$$P(\bar{X}_{(n)} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} \leq \mu \leq \bar{X}_{(n)} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}) = 1 - \alpha$$

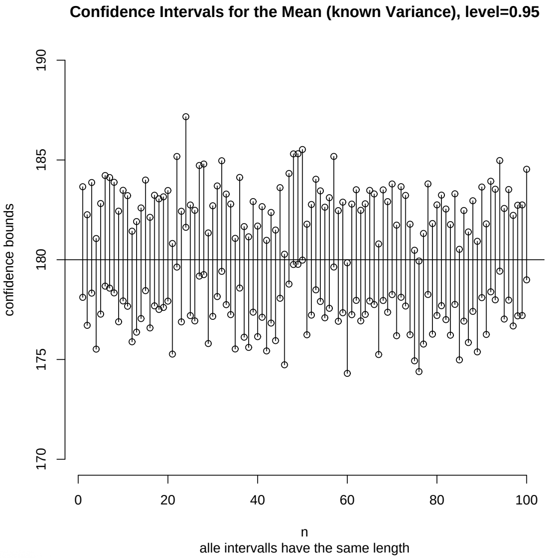
Thus we have the following confidence interval:

$$\left[\bar{X}_{(n)} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}, \bar{X}_{(n)} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right] = [173.61, 181.44]$$

Remark:

- $1 - \alpha$ is called the confidence level.
- The lower and upper bounds are random variables.

Confidence Intervals for Normal Distributions: parameter μ IV



If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the intervals would contain the population mean. Naturally, 5% of the intervals would not contain the population mean.

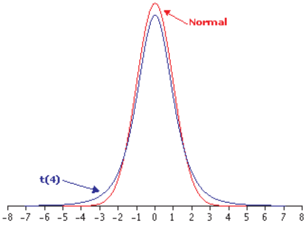
Confidence Intervals for Normal Distributions: parameter μ V

Remark:

- The formula of the confidence interval is derived using
$$\frac{\bar{X}_{(n)} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$
in the case of i.i.d. $N(\mu, \sigma^2)-$ distributed random variables $X_1, ..., X_n$.
- In the case of unknown standard deviation σ, σ must be estimated.
$$\frac{\bar{X}_{(n)} - \mu}{S_{(n)} / \sqrt{n}} \sim t_{n-1}$$
i.e. t-distribution with $n-1$ degrees of freedom instead of the standard normal distribution.

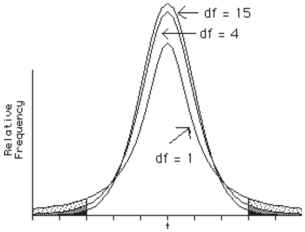
Confidence Intervals for Normal Distributions: parameter μ VI

t-distribution:



HyperStat Online Chapter 8

- The shape depends on the degrees of freedom (df) that went into the estimate of the standard deviation.
- The distribution has a greater number of scores in its tails than the normal distribution.
- As the degrees of freedom increases, the t distribution approaches the standard normal distribution.



Stat Online Chapter 8

t distributions with 1, 4, and 15 degrees of freedom: Areas greater than +2 and less than -2 are shaded. This figure shows that the t distribution with 1 df has the least area in the middle of the distribution and the greatest area in the tails.

Confidence Intervals for Normal Distributions: parameter μ VII

Remark: Usually the variance is unknown. Therefore, the construction of a confidence interval involves the estimation of both μ and σ and the t-distribution is to be used instead of the normal distribution.

Confidence interval for μ and unknown σ:

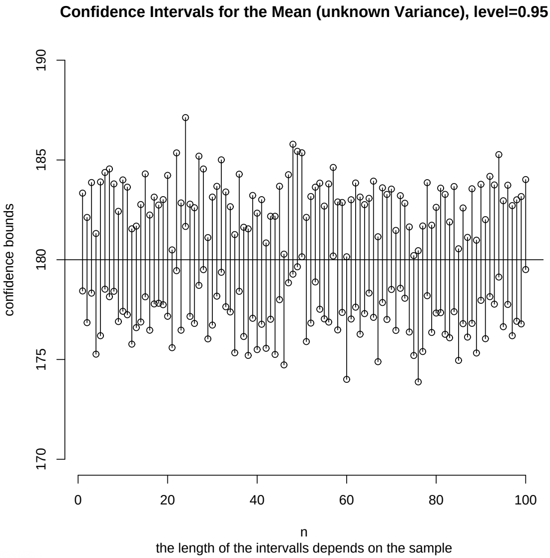
$$\left[\bar{X}_{(n)} - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}, \bar{X}_{(n)} + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}\right]$$

Example: Sample of 25 male students (i.i.d. normally distributed)

- $\bar{x}_{(25)} = 177.52, s_{(25)} = 8.227$
- confidence level $1 - \alpha = 0.95 \Rightarrow t_{24,0.975} = 2.064$

$\Rightarrow$ confidence interval for μ: $[174.12, 180.92]$

Confidence Intervals for Normal Distributions: parameter μ VIII



Note that the location and the length of the confidence interval are varying from sample to sample.

Confidence Intervals for Normal Distributions: parameter σ^2 I

In the case of i.i.d. $N(\mu, \sigma^2)-$distributed random variable $X_1, ..., X_n$ it can be shown that

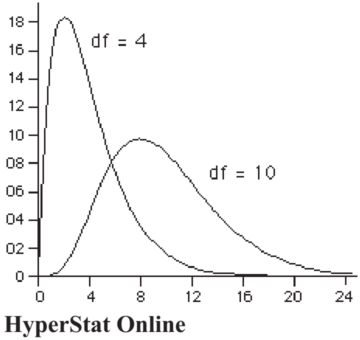
$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_n \text{ and } \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}_{(n)}}{\sigma}\right)^2 \sim \chi^2_{n-1}$$

Using these results the following formulas for the confidence interval for σ^2 can be derived.

1. mean μ_0 known:
$$\left[\frac{Q_{(n)}}{\chi^2_{n;1-\frac{\alpha}{2}}}, \frac{Q_{(n)}}{\chi^2_{n;\frac{\alpha}{2}}}\right] \quad \text{with} \quad Q_{(n)} = \sum_{i=1}^{n}(X_i-\mu_0)^2$$
2. mean unknown:
$$\left[\frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Confidence Intervals for Normal Distributions: parameter σ^2 II

Chi Square Distribution:



- One parameter: degrees of freedom (df)
- A positive skewness; the skewness is less with more degrees of freedom
- As the df increase, the chi square distribution approaches a normal distribution.
- The mean of a chi square distribution is its df. The mode is df - 2 and the median is approximately df - 0.7.

Tail Areas und Chi-Squared Distributions

Source: "Tail Areas under Chi-Squared Distributions"from the Wolfram Demonstrations Project

http://demonstrations.wolfram.com/TailAreasUnderChiSquaredDistributions/

Confidence Intervals for Normal Distributions: parameter σ^2 III

Example: Sample of 25 male students (i.i.d. normally distributed)

- $\mu_0 = 180$ known
  - $Q_{(n)} = 1778$
  - confidence level $1 - \alpha = 0.95 \Rightarrow$ $\chi^2_{25,0.975} = 40.65, \chi^2_{25,0.025} = 13.12$

  $\Rightarrow$ confidence interval for σ^2: $[43.75, 135.53]$
- μ unknown
  - $S^2_{(n)} = 67.68$
  - confidence level $1 - \alpha = 0.95 \Rightarrow$ $\chi^2_{24,0.975} = 39.36, \chi^2_{24,0.025} = 12.40$

  $\Rightarrow$ confidence interval for σ^2: $[41.27, 130.98]$

Questions

Which of the following statements are true or false?

t	f	
☐	☐	The confidence level $1-\alpha$ indicates the likelihood that the unknown parameter is within the confidence interval.
☐	☐	Increasing the confidence level consecutively shortens the length of the confidence interval.
☐	☐	The length of confidence intervals depends on the confidence level and the variability of the sample values.
☐	☐	Increasing the sample sizes will (on the average) enhance the precision of the intervall estimation.

Confidence Intervals for Normal Distributions: Summary I

Let $\bar{X}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2_{(n)} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)})^2$ and $1-\alpha$ be the level of confidence.

Assumptions: Normal distribution $N(\mu, \sigma^2)$, scores are sampled randomly and are independent

1. Confidence interval for μ, standard deviation known:
$$\left[\bar{X}_{(n)} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X}_{(n)} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$
2. Confidence interval for μ, standard deviation unknown:
$$\left[\bar{X}_{(n)} - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}, \bar{X}_{(n)} + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S_{(n)}}{\sqrt{n}}\right]$$

Confidence Intervals for Normal Distributions: Summary II

3. Confidence interval for σ^2, mean μ_0 known:
$$\left[\frac{Q_{(n)}}{\chi^2_{n;1-\frac{\alpha}{2}}}, \frac{Q_{(n)}}{\chi^2_{n;\frac{\alpha}{2}}}\right] \quad \text{with} \quad Q_{(n)} = \sum_{i=1}^{n}(X_i - \mu_0)^2$$
4. Confidence interval for σ^2, mean unknown:
$$\left[\frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Remark: You get a lower resp. an upper confidence bound if you change in the corresponding bound $\alpha/2$ by α.

Confidence Intervals on the Proportion I

Example: Proportion p of X-ray machines that malfunction and produce excess radiation

- A random sample of 40 machines is taken and 12 of the machines are of the machines malfunction.
- Although the point estimate of the proportion $\hat{p} = \frac{12}{40}$ is informative, it is important to also compute a confidence interval.

Confidence Intervals on the Proportion II

Assumption: Observations are sampled randomly and independently.

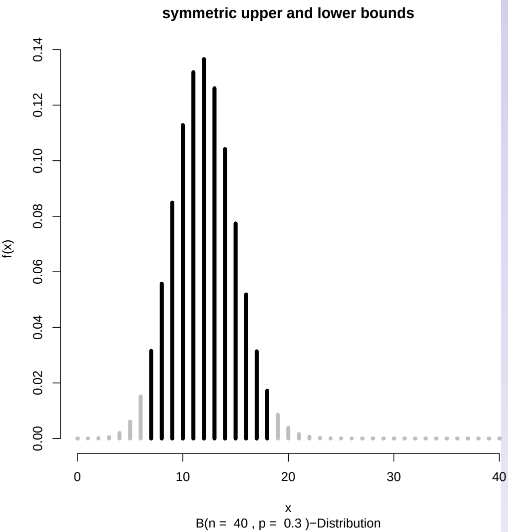
Let $X_i = \begin{cases} 1 & \text{machine i malfunctions} \\ 0 & \text{else} \end{cases}$ then $X_1, \ldots, X_n$ are independent identically $B(1,p)$-distributed with unknown p

$$\Rightarrow X = \sum_{i=1}^{n} X_i \sim B(n,p)$$

$$\Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{with} \quad E(\hat{p}) = p, \text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

Confidence Intervals on the Proportion III

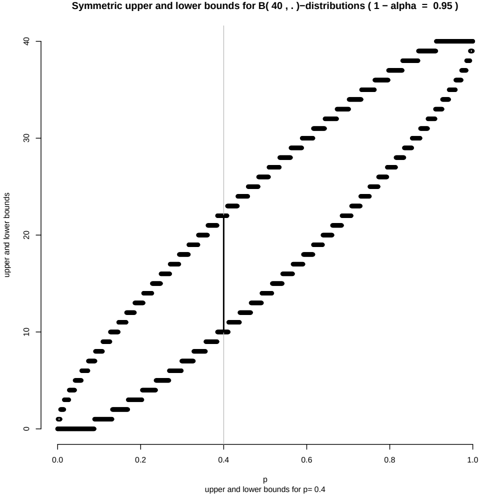
We first discuss symmetric $1-\alpha$ upper $u_{n,\alpha}()$ and lower bounds $l_{n,\alpha}(p)$ of $X \sim B(n,p)$.



Confidence Intervals on the Proportion IV

n and α fix; $p \in [0, 1]$

- $l_{n,\alpha}(p)$, $u_{n,\alpha}(p)$ are uniquely defined for p
- both functions are monotonously increasing in p

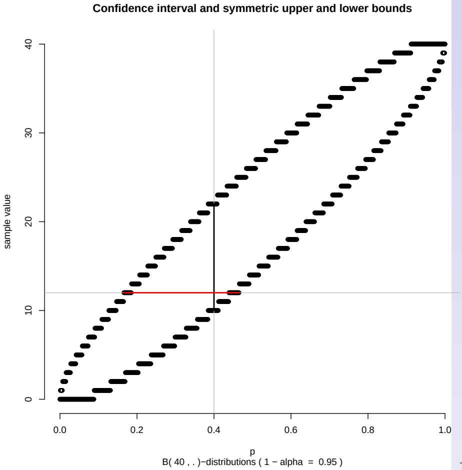


Confidence Intervals on the Proportion V

consider the sets $u_{n,\alpha}^{-1}(x) = \{p \mid u_{n,\alpha}(p) = x\}$, $l_{n,\alpha}^{-1}(x) = \{p \mid l_{n,\alpha}(p) = x\}$ and take the min and max values of these sets

sample value $x \in [0, n]$:

- $pu_{n,\alpha}(x) = \min_p u_{n,\alpha}^{-1}(x)$
- $pl_{n,\alpha}(x) = \max_p l_{n,\alpha}^{-1}(x)$.
- $p \in [pu_{n,\alpha}(x), pl_{n,\alpha}(x)]$ $(*) \Rightarrow x \in [l_{n,\alpha}(x), u_{n,\alpha}(x)]$
- $(*)$ is a $1-\alpha$ confidence interval for p.



Confidence Intervals on the Proportion VI

From the Moivre-Laplace Theorem we get from $P(l \leq X \leq u) = 1 - \alpha$ approximate confidence bounds for p for the level $1 - \alpha$ with $c = u_{1-\alpha/2}$:

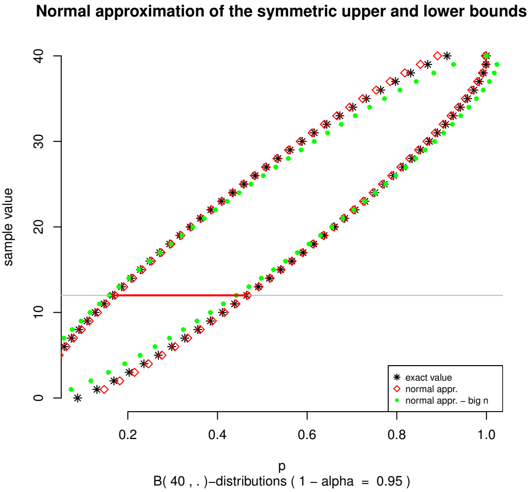
$$\left[\frac{X - 0.5 + \frac{c^2}{2} - c\sqrt{X - 0.5 + \frac{c^2}{4} - \frac{(x-0.5)^2}{n}}}{c^2 + n}, \right.$$

$$\left. \frac{X + 0.5 + \frac{c^2}{2} + c\sqrt{X + 0.5 + \frac{c^2}{4} - \frac{(X+0.5)^2}{n}}}{c^2 + n} \right]$$

For big n we get approximately with $\hat{p} = \frac{X}{n}$

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

Confidence Intervals on the Proportion VII



Confidence Intervals on the Proportion VIII

Example: $X = 12, n = 40$: $\hat{p} = 12/40 = 0.30$. The estimated value of $s = \sqrt{\frac{0.3 \cdot 0.7}{40}}$. For $1 - \alpha = 0.95$ we get the following bounds:

bound	exact value	normal appr.	normal appr. - big n
lower	0.1657	0.1709	0.1580
upper	0.4653	0.4671	0.4420

Confidence Intervals on the Proportion IX

Remark: The adequacy of the normal approximation relies on the sample size n and p. Although there are no strict guidelines, the subsequent provides a reference for the necessary sample size:

- If p is between 0.4 and 0.6 then an n of 10 is adequate. If p is as low as 0.2 or as high as 0.8 then n should be at least 25. For p as low as 0.1 or as high as 0.9, n should be at least 30.
- A more conservative rule of thumb often recommended is that both np and $n(1-p)$ should be at least 10.

Hypothesis Testing I

compare: Online Statistics IX

Hypothesis testing: statistical procedure for testing whether chance is a plausible explanation of an experimental finding.

Logic of Hypothesis Testing I

compare: Online Statistics IX

Example: Experiment to ascertain if Mr. Bond has a higher success rate than chance in determining if a martini is shaken or stirred.

- 16 tests:
  - A fair coin is flipped to determine whether to stir or shake the martini.
  - Mr. Bond decides whether it was shaken or stirred.
- Correct 13 / 16 times
- Is this proof Mr. Bond can detect a stirred martini?

Logic of Hypothesis Testing II

How plausible is the explanation that Mr. Bond was just lucky?

- Probability of getting 13 or more if just guessing?

  $X =$ number of right decisions,
  π probability of a right decision.
  $\Rightarrow X \sim B(16, \pi)$.
- $\pi \leq 0.5$: Mr. Bond is just guessing or tends to make a wrong decision
- $P_{\pi=0.5}(X > 12) = 0,0106$ and $P_{\pi=0.5}(X > x) \geq P_{\pi\leq 0.5}(X > x)$
- Strong evidence for not just guessing

x	$P_{\pi=0.5}(X > x)$
0	0.99998
1	0.99974
2	0.99791
3	0.98936
4	0.96159
5	0.89494
6	0.77275
7	0.59819
8	0.40181
9	0.22725
10	0.10506
11	0.03841
12	0.01064
13	0.00209
14	0.00026
15	0.00002
16	0.00000

Logic of Hypothesis Testing III

Probability Value:

- In the James Bond example, the computed probability of 0.0106 is the probability he would be correct on 13 or more taste tests (out of 16) if he were just guessing or tends to make a wrong decision.
- The probability of 0.016 is the probability of a certain outcome (13 or more out of 16) assuming a certain state of the world.
- If the probability of recognizing a stirred martini is less equal than 0,5 we get a probability value less than 0,0106.
- Thus the probability value is
  - not the probability he cannot tell the difference,
  - it is the probability of a certain outcome assuming a state of the world

Logic of Hypothesis Testing IV

- **Null Hypothesis** hypothesis about a state of the world (here about a population parameter)
  Typically a hypothesis of no differences, i.e. that an apparent effect is due to chance.
- Purpose of hypothesis testing: test the viability of the null hypothesis based on experimental data.
- Depending on the data, the null hypothesis will either be rejected or fail to be rejected.

James Bond Example: Is Mr. Bond better at chance at tasting a stirred martini?
Null Hypothesis: $\pi \leq 0.5$

Logic of Hypothesis Testing V

compare: Hedderich, Sachs: Angewandte Statistik, Auflage 2020, S. 456

- Statistical hypothesis = assertion about properties of one or more random variables
- Hypotheses are usually only indirectly testable.
- Examples increase the reliability of research results through empirical evidence.
- Since a hypothesis (H_1) can never be confirmed directly a counter hypothesis (H_0) is made and attempted to disprove it.
- This allows the hypothesis H_1 to be indirectly confirmed.
- If the data sample can only be explained with a low probability, assuming the hypothesis of H_0, it is considered evidence towards the hypothesis of H_1.

Logic of Hypothesis Testing VI

Significance Testing:

- Statistical tests are performed with a test statistic. [3]
- Probability-value = Probability of the test statistic or more extreme values occurring if H0 is true.
- Low probability values cast doubt on the null hypothesis
- The probablility value below which the null hypothesis is rejected is called significance level or simply α level
- Conventional significance levels are 0,05 and 0,01
- A result is called statistically significant if the null hypothesis is rejected

[3] Rule according to which a number is calculated from the sample data. Depending on the value of the number, a decision is made for or against H0.

Logic of Hypothesis Testing VII

Type I and II Errors: Two kinds of errors:

- A true null hypothesis can be incorrectly rejected (type I error).
- A false null hypothesis can fail to be rejected (type II error).

Statistical decision	True state of the Null Hypothesis H_0	
	H_0 True	H_0 False
Reject H_0	Type I Error	Correct
Do not reject H_0	Correct	Type II Error

α probability of rejecting H_0 given that H_0 is true

β probability of not rejecting H_0 given that H_0 is false

- Probability of recognizing the working hypothesis H_1 as such, i.e., $P(H_1|H_1)$ is called power. It is equal to 1-beta

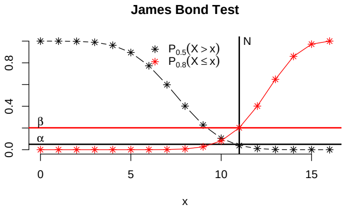
Logic of Hypothesis Testing VIII

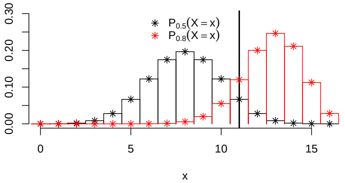
Example James Bond: $H_0 : \pi \leq 0.5 \qquad H_1 : \pi > 0.5$

X number of right decisions $\sim B(n, \pi)$

significance level $\alpha = 0.05$

Decision Rule: Reject H_0 if $X > N$





- $P_\pi(X > x)$ is a monotonously increasing function in π. Thus N is the $(1 - \alpha)$-quantil of the $B(n, \pi = 0.5)$-distribution.

  The probability of a type I error is $P_{\pi \leq 0.5}(X > N) \leq \alpha$
- The probability β of a type II error is $\beta(\pi) = P_\pi(X \leq N)$ for $\pi > 0.5$. Since $P_\pi(X < N)$ is a monotonuously decreasing function in π, β increases if π decreases to 0.5.

Logic of Hypothesis Testing IX

- **A Type II error is not really an error.**
  When a statistical test is not significant, it indicates that the data do not provide convincing evidence that the null hypothesis is incorrect. A lack of significance does not support the conclusion that the null hypothesis is correct.
  As in a court of law: “In doubt for the accused”
- **A Type I error is really an error.**
  If a Type I error occurs, the researcher incorrectly believes that the null hypothesis is false when it is actually true.

Therefore, Type I errors are generally considered more serious than Type II errors. The probability of a Type I error is set by the experimenter.

Logic of Hypothesis Testing X

Tradeoff between Type I and Type II errors:

- The more an experimenter protects himself or herself against a Type I error by choosing a low level, the greater the chance of a Type II error.
- Requiring very strong evidence to reject the null hypothesis makes it very unlikely that a true null hypothesis will be rejected.
- However, it increases the chance that a false null hypothesis will not be rejected.

$\rightarrow$ For any given set of data, type I and type II errors are inversely related:

the smaller the risk of one, the higher the risk of the other

Logic of Hypothesis Testing XI

One and Two Tailed Tests:

- Depending on the form of the rejection area we distinguish one and two tailed tests.
- In the James Bond Example our question is whether Mr. Bond is better than chance at determining whether a martini is stirred or not, i.e.

$$H_0 : \pi \leq 0.5 \qquad H_1 : \pi > 0.5$$

Rejection, if Mr. Bond does very well, i.e. X is bigger a certain value $\rightarrow$ **one tailed test**

Logic of Hypothesis Testing XII

two tailed test:

- If we are asking whether Mr. Bond can tell the difference between shaken or stirred martinis, then we would conclude he could if he performed either much better than chance or much worse than chance, i.e.

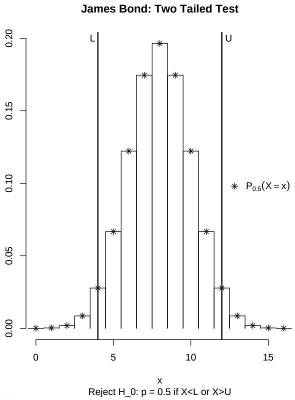
$$H_0 : \pi = 0.5 \qquad H_1 : \pi \neq 0.5$$

- If he performed much worse than chance, we would conclude that he can tell the difference, but he does not know which is which. So, since we want to know whether Mr. Bond does either very well or very bad, i.e. rejection if X is greater than some or X is less than some value.

Logic of Hypothesis Testing XIII

Example: James Bond

- $H_0 : \pi = 0.5 \quad H_1 : \pi \neq 0.5$ with level $\alpha = 0.05$
- Decision Rule: Reject H_0 if $X > U$ or $X < L$
- U is the $(1 - \alpha/2)$-quantil and L is the $\alpha/2$-quantil of the $B(n, 0.5)$-distribution.



x	$P_{0.5}(X \leq x)$
0	0.00002
1	0.00026
2	0.00209
3	0.01064
4	0.03841
5	0.10506
6	0.22725
7	0.40181
8	0.59819
9	0.77275
10	0.89494
11	0.96159
12	0.98936
13	0.99791
14	0.99974
15	0.99998
16	1.00000

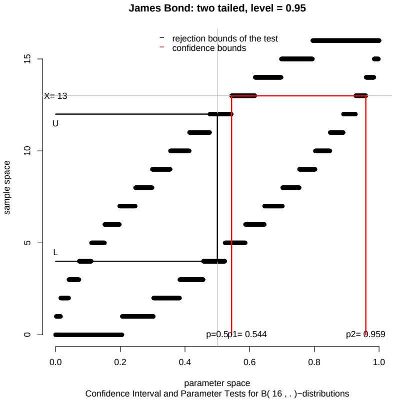
Logic of Hypothesis Testing XIV

Relationship between confidence intervals and hypothesis testing:

- A 95% confidence interval is constructed.
- Values in the interval are considered as plausible values for the parameter being estimated.
- Values outside the interval are rejected as relatively implausible.
- If the value of the parameter specified by the null hypothesis is contained in the 95% interval then the null hypothesis cannot be rejected at the 0.05 level.
- If the value specified by the null hypothesis is not in the interval then the null hypothesis can be rejected at the 0.05 level.
- Be careful if the value is close to the bounds of the confidence interval.

Logic of Hypothesis Testing XV

Example: James Bond



- A two tailed test with level $\alpha = 0.05$

  $H_0 : \pi = 0.5 \quad H_1 : \pi \neq 0.5$

  will be rejected if $X \notin [L, U]$.
- For the sample value X we get the confidence interval $[p1, p2]$ for π with level $1 - \alpha = 0.95$
- If $\pi = 0.5 \in [p1, p2]$ the null hypothesis can not be rejected at the 0.05 level.

Questions

Which of the following statements are true or false?

t	f	
☐	☐	By conducting a statistical test, one can verify the accuracy of a hypothesis.
☐	☐	If one decides for the correctness of the hypothesis, although it is wrong, then one makes an error of 1st kind.
☐	☐	The significance level alpha of a test is the max. probability of an error 1st kind.
☐	☐	The probability of a 1st kind error and the probability of a 2nd kind error are always the same.
☐	☐	A test is used to check whether a value of interest can be brought into conformity with the data by taking into account a certain probability of error. If the value of interest lies within a confidence interval, this is confirmed and a test is obtained from a confidence interval. This also works the other way round.

Basic Model I

Starting Point:

- Random experiment: outcome is a sequence of n observable random variables taking values in a sample space S:

$$X = (X_1, X_2, ..., X_n).$$

- A particular outcome $x = (x_1, x_2, ..., x_n)$ of the experiment forms our data.
- Most important special: a random sample of size n from the distribution of X, i.e. $X_1, X_2, ..., X_n$ are n independent, identicaly distributed variables

Statistical Hypothesis: A statement about the distribution of the random variable X

Basic Model II

Question: Is there sufficient statistical evidence to reject a presumed null hypothesis H_0 in favor of a conjectured alternative hypothesis H_1.

Hypothesis Test $=$ Statistical Decision:

- Conclusion: reject H_0 in favor of H_1, or fail to reject H_0
- Decision is based on the data vector X

$$R \subset S : \text{reject } H_0 \text{ if and only if } X \in R$$

- Usually, the critical region R is defined in terms of a statistic $W(X)$ (test statistic).
  Asymmetry between H_0 and H_1: We assume H_0 and then see if there is sufficient evidence in X to overturn this assumption in favor of the alternative.

Basic Model III

Types of errors:

1. Type 1 error: reject the null hypothesis when it is true.
2. Type 2 error: fail to reject the null hypothesis when it is false.

- H_0 is true: $P(X \in R)$ is the probability of a type 1 error
  - Maximum probability of a type 1 error: significance level α of the test
  - R is constructed so that the significance level is a prescribed, small value (typically 0.1, 0.05, 0.01).
- H_1 is true: $P(X \notin R)$ is the probability of a type 2 error

Basic Model IV

- **Tradeoff between the type 1 and type 2 error probabilities:** If we decrease the likelihood of a type 1 error, by shrinking the rejection area R smaller, we inherently increase the likelihood a type 2 error because the complementary region $S \setminus R$ enlarges.
- **p-value:** The p-value of the data variable X, denoted $p(X)$ is defined to be the smallest α for which $X \in R_\alpha$; that is, the smallest significance level for which H_0 is rejected, given X.
  - If $p(X) \leq \alpha$ then we would reject H_0 at significance level α.
  - If $p(X) > \alpha$ then we fail to reject H_0 at significance level α.

Basic Model V.a

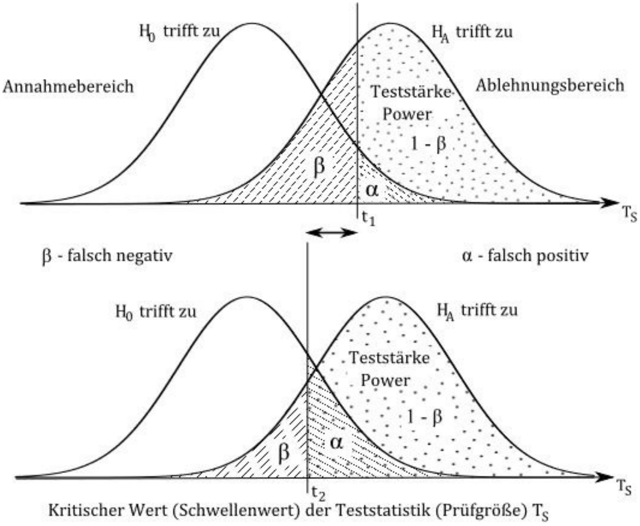
Hedderich, Sachs, Kapitel 7.1.5 Powerfunktion und Operationscharakteristik

- In many cases, there are different test procedures for testing a null hypothesis.
- Assessment of the (quality) of a test by the power function.
- Power function: probability of rejection as a function of the depending to be estimated parameter δ

$$G(\delta) = P(T \in R_\alpha | \delta)$$

- The function characterizes the probability for a wrong decision (error of 1st kind, α), if $\delta \in \Omega_0$ (H_0) and for a correct decision (power, $1 - \beta$) if $\delta \in \Omega_1$ (H_1): $\sup_{\delta \in \Omega_0} G(\delta) = \alpha$

Basic Model V.b



Source: Sachs, Hedderich Angewandte Statistik, Auflage 17, Kapitel 7.1.7

Basic Model VI.a

- A small possible β error of maintaining a false null hypothesis depends on:
  - size of the sample: If the sample size increases, it will be more likely that a difference between two populations will be detected for a given probability of error α.
  - the degree of difference between the hypothetical and the true state of the effect to be detected, that is the amount by which the null hypothesis is false.
  - the power of the test, i.e. $1 - \beta$
    - higher information content of the output data (frequencies, rankings and measured values) increases the power
    - if more preconditions about the distribution of the values are made we have a higher power: A test that requires normal distribution and variance homogeneity is generally much stronger than one that makes no assumptions at all.
    - depends on the directionality of the test (two- or one-sided test)

Basic Model VII

One-sided or two-sided hypothesis formulation?

- When switching from the one-sided to the two-sided hypothesis the power decreases.
- With the same sample size, a one-tailed test is always more discriminating than the two-tailed test, provided that it can be justified!
- Two-sided hypotheses should in principle be used as long as there is no good factual justification for a one-sided hypothesis.
- The one-sided questioning is preferred, if the two-sided questioning is obviously meaningless.

Basic Model VIII

How many observations are required?

- Sample sizes that are too small are not even capable of detecting large differences between two parameters.
- Sample sizes that are too large detect tiny differences that are practically meaningless.
- Therefore, one must first consider what difference (effect), if any, is important to find.
- Then it has to be determined with which probability or power at least this difference/effect should be found.

Basic Model IX

- Only in the case of large n or in the case of a large difference (effect) statistical significance will result if a very small α is given.
- For the two-sided test, as the distance $\mu - \mu_0$ increases the probability to reject the null hypothesis increases and if the significance level or/and the sample size become smaller, it become more difficult to accept a true alternative hypothesis.

Parameter Tests in the Normal Model I

Suppose that

$$X = (X_1, X_2, ..., X_n)$$

is a random sample of size n from the normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in (0, \infty)$.

Objective: *Hypothesis tests for μ and σ*

Parameter Tests in the Normal Model II

Example: The length of a certain part is supposed to be 70 centimeters.

- Due to imperfections in the manufacturing process, the actual length is a random variable.
- The standard deviation remains relatively stable over time due to inherent factors in the process. From historical data, the standard deviation is known to be 4.
- The mean may be set by adjusting various parameters in the process and hence may change to an unknown value fairly frequently.
- Sample of n parts: X_i measured length of part i $(1 \leq i \leq n)$

$$X = (X_1, X_2, ..., X_n) \quad \text{with } X_i \text{ i.i.d. } N(\mu, 16)$$

Parameter Tests in the Normal Model III

Test Statistic: $\bar{X}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$.

$H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ (significance level α)

Decision Rule: Reject H_0 if $\bar{x}_{(n)} \notin [\mu_0 - c, \mu_0 + c]$ where

$$P_{\mu=\mu_0}(\mu_0 - c \leq \bar{X}_{(n)} \leq \mu_0 + c) = 1 - \alpha$$

Therefore we get $c = u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ and the rejection region

$$R_\alpha = \left[\mu_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]^c$$

Using the test statistic $\frac{\bar{X}_{(n)} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ we get the decision rule:

Rejection of $H_0 \Leftrightarrow \frac{\bar{x}_{(n)} - \mu_0}{\sigma/\sqrt{n}} \notin [-u_{1-\alpha/2}, u_{1-\alpha/2}]$

Parameter Tests in the Normal Model IV

- H_0 will be rejected if

$$\bar{X}_{(n)} < \mu_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{X}_{(n)} > \mu_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\iff \mu_0 \notin \left[\bar{X}_{(n)} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_{(n)} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

  i.e. if μ_0 is outside the corresponding confidence interval.

- p-value: with $c = |\bar{x} - \mu_0|$, $\bar{x}$ = sample mean

$$p(\bar{x}) = 1 - P_{\mu=\mu_0}(\mu_0 - c \leq \bar{X}_{(n)} \leq \mu_0 + c) = 2(1 - \Phi(\frac{c}{\sigma/\sqrt{n}}))$$

- probability of type 2 error:

$$\begin{aligned}\beta(\mu) &= P_\mu(\bar{X}_{(n)} \notin R_\alpha) \\ &= P_\mu(\mu_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_{(n)} \leq \mu_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})\end{aligned}$$

  If α decreases then $\beta(\mu)$ increases.

Parameter Tests in the Normal Model V

If only deviations $|\mu - \mu_0| \geq e$ are relevant, these deviations should be detected with a probability of at least $1 - \beta_{max}$, what is the mininmal sample size?

- Let for example $\mu_1 = \mu_0 + e$ then the nullhypothesis $H_0 : \mu \geq \mu_1, H_1 : \mu < \mu_1$ should have a significane level of at most β_{max}.
- $$\mu_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = \mu_1 + u_{1-\beta_{max}}\frac{\sigma}{\sqrt{n}} \Rightarrow$$
  $$n \geq \frac{(u_{1-\alpha/2} + u_{1-\beta_{max}})^2}{(\mu_1 - \mu_0)^2}\sigma^2$$

Parameter Tests in the Normal Model VI

Example: Sample of 100 parts with mean 71: $\alpha = 0,05$ and $\sigma = 4$

Decision Rule: Reject H_0 if $\bar{X}_{(n)} \notin [69,216; 70,784]$

- Since $\bar{X}_{(n)} = 71$ H_0 should be rejected.
- p-value $= 1 - P_{\mu=70}(69 \leq \bar{X}_{(n)} \leq 71) = 0,0124$
- probability of type 2 error:

$$\begin{aligned} \beta(\mu) &= P_\mu(69,216 \leq \bar{X}_{(n)} \leq 70,784) \Rightarrow \\ \beta(71) &= \Phi\left(\frac{70,784 - 71}{0,4}\right) - \Phi\left(\frac{69,216 - 71}{0,4}\right) = 0,295 \end{aligned}$$

- minimal sample size for $|\mu - 70| \geq 0.5$ and $\beta(\mu) \leq 0.1$ for all $\mu \leq 69.5$ or $\mu \geq 70.5$ is $\frac{(u_{0.975}+u_{0.9})^2}{(70.5-70)^2} 4^2 \approx 673$

Example: Shiny App about Gauß-Test

Parameter Tests in the Normal Model VII

Summary parameter tests in the normal model:
Analogously we get using the relationship between confidence intervals and hypothesis testing the following results for the significance level α.

I) **Gauß-Test:** $N(\mu, \sigma_0^2)$ with μ unknown and σ_0 known

Teststatistic: $\dfrac{\bar{X}_{(n)} - \mu}{\sigma_0}\sqrt{n} \sim N(0,1)$

Decision Rule: $\dfrac{\bar{x} - \mu_0}{\sigma_0}\sqrt{n} \in R \;\Rightarrow$ reject H_0

H_0	rejection region R
$\mu = \mu_0$	$(-\infty, -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty)$
$\mu \leq \mu_0$	$(u_{1-\alpha}, \infty)$
$\mu \geq \mu_0$	$(-\infty, -u_{1-\alpha})$

Parameter Tests in the Normal Model VIII

II) **t-Test:** $N(\mu, \sigma^2)$ with μ and σ_0 unknown

Teststatistic: $\frac{\bar{X}_{(n)} - \mu}{s_{(n)}}\sqrt{n} \sim t_{n-1}$ with

$S^2_{(n)} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)})^2$

Decision Rule: $\frac{\bar{x} - \mu_0}{s_{(n)}}\sqrt{n} \in R \Rightarrow$ reject H_0

H_0	rejection region R
$\mu = \mu_0$	$(-\infty, -t_{n-1,1-\frac{\alpha}{2}}) \cup (t_{n-1,1-\frac{\alpha}{2}}, \infty)$
$\mu \leq \mu_0$	$(t_{n-1,1-\alpha}, \infty)$
$\mu \geq \mu_0$	$(-\infty, -t_{n-1,1-\alpha})$

Parameter Tests in the Normal Model IX

III) $N(\mu, \sigma^2)$ with μ and σ unknown

Teststatistic: $\frac{(n-1)S^2_{(n)}}{\sigma^2} \sim \chi^2_{n-1}$ with

$S^2_{(n)} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_{(n)})^2$

Decision Rule: $\frac{(n-1)s^2_{(n)}}{\sigma_0^2} \in R \Rightarrow$ reject H_0

H_0	rejection region R
$\sigma^2 = \sigma_0^2$	$(0, \chi^2_{n-1,\frac{\alpha}{2}}) \cup (\chi^2_{n-1,1-\frac{\alpha}{2}}, \infty)$
$\sigma^2 \leq \sigma_0^2$	$(\chi^2_{n-1,1-\alpha}, \infty)$
$\sigma^2 \geq \sigma_0^2$	$(0, \chi^2_{n-1,\alpha})$

Parameter Tests in the Normal Model X

Remark: Calculation of the p-value in these cases:

- Exchange the quantile in the rejection region R by the value of the teststatistic for the given data. $\rightarrow \tilde{R}$
- p-value$= P_{H_0}$(teststatistic $\in \tilde{R}$)

Parameter Tests in the Normal Model XI

R functions for one sample tests and confidence intervals:

- Gauß test: z.test(), package TeachingDemos
- t-test: t.test(), package stats
- variance test: sigma.test(), package TeachingDemos

Remark: Beside conducting the tests the functions calculate the corresponding confidence intervals and the p-values, too.

Example: R script: one_sample_test_conf_intervals.R

Questions

A sample of the size n=25 from a normal distribution resulted in $\bar{x} = 9$ and $s = 2$. You want to conduct a hypothesis on μ at the $\alpha = 0.05$-level.
Which of the following statements are true or false?

t	f	
☐	☐	The value of the test statistic T is 2.5.
☐	☐	$H_0 : \mu = 10$ can be rejected if the absolute value of the test statistic T is bigger $u_{0.975}$.
☐	☐	$H_0 : \mu > 10$ can be rejected if the value of the test statistic $T < -t_{24,0.95}$.
☐	☐	The p-value for $H_0 : \mu = 10$ is ca. 0.02.
☐	☐	The rejection region for $H_0 : \sigma^2 \leq 1.5^2$ is $\frac{24 * s^2}{1.5^2} > \chi^2_{24,0.95}$.

Tests in the Bernoulli Model I

- $X_i, 1 \leq i \leq n$ are independent random variables taking the values 1 and 0 with probabilities p and $1 - p$ respectively.
- $X = (X_1, X_2, ..., X_n)$ is a random sample from the Bernoulli distribution with unknown success parameter $p \in (0, 1)$.

Tests in the Bernoulli Model II

Applications:

- Event of interest in a basic experiment with unknown probability p.
- Replicate the experiment n times and define $X_i = 1$ if and only if the event occurred on run i.

Tests in the Bernoulli Model III

Example:

- Population of objects of several different types.
- p unknown proportion of objects of a particular type of interest.
- Select n objects at random from the population and let $X_i = 1$ if and only if object i is of the type of interest.
- When the sampling is with replacement, these variables really do form a random sample from the Bernoulli distribution.
- When the sampling is without replacement, the variables are dependent, but the Bernoulli model may still be approximately valid.

Tests in the Bernoulli Model IV

- Number of successes $X = \sum_{i=1}^{n} X_i \sim B(n, p)$, i.e. $E(X) = np$ and Var(X)$= np(1-p)$.
- In case of large n the distribution of X is approximately normal, by the central limit theorem. An approximate normal test can be constructed using the test statistic $\dfrac{r_n - p}{\sqrt{p(1-p)/n}}$ with $r_n = \dfrac{X}{n}$. [4]
- **Decision Rule:** $\dfrac{r_n - p_0}{\sqrt{p_0(1-p_0)/n}} \in R \Rightarrow$ reject H_0

H_0	rejection area R
$p = p_0$	$(-\infty, -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty)$
$p \leq p_0$	$(u_{1-\alpha}, \infty)$
$p \geq p_0$	$(-\infty, -u_{1-\alpha})$

[4] The adequacy of the normal approximation depends on n and p. A rule of thumb is that np and n(1 - p) should both be at least 10.

Tests in the Bernoulli Model V

Example: (compare Heumann, Schomaker, p 228) A party wants to know whether the proportion of votes will exceed 30%. In a representative sample of size n=2000 of eligible voters 700 have voted the party.

- test statistic
  $$T = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.35 - 03}{\sqrt{0.3(1-0.3)}}\sqrt{2000} = 4.8795$$
- If $\alpha = 0.05$, $T = 4.8795 > u_{1-\alpha} = 1.64$, $H_0 = p \leq 0.3$ can be rejected
- p-value: $P(T \geq 4.8795) = 5.318e - 07$

Tests in the Bernoulli Model VI

Exact binomial test: above example with $\alpha = 0.05$

- Teststatistic $T = X \sim B(n = 2000, p_0)$
- Critical region: find c with $P_{p=0.3}(T \geq c) \leq 0.05$, i.e. $P_{p=0.3}(T < c) \geq 0.95$
  From R (qbinom(p=0.95, size=2000, prob=0.3)) we get c = 634. Since $T = 700 > 634$ we reject $H_0 : p \leq 0.3$.
- p-value: $P_{p=0.3}(T \geq 700) = 1 - P_{p=0.3}(T \leq 699) = 8.395e-07$ (pbinom(699,size=2000,prob=0.3))
- The result can be easily get by the R function binom.test(): binom.test(700,2000, p=0.3, alternative="greater"). The function determines a corresponding confidence interval and the -value, too.
  compare: one_sample_test_conf_intervals.R

Questions

In a big city, 20% of all households have subscribed to a certain magazine so far. In a random sample of 100 households, 16 households have subscribed to the magazine. Let $H_0 : \pi >= 0.2$ and $\alpha = 0.05$.
Which of the following statements are true or false?

t	f	
☐	☐	Based on a normal approximation the value of the test statistic T is -0.875.
☐	☐	The rejection region of the approximation of the binomial test is $T < -u_{0.95}$ if T is the test statistic.
☐	☐	The p-value of the approximation of the binomial test is $\Phi(t)$, if t is value of the test statistic.
☐	☐	The p-value of the exact binomial test is $P(X <= 16)$ if $X \sim B(100, 0.2)$.

Section 2

Two Sample Tests

Types of Tests I

Tests can be classified in Parametric and Nonparametric Tests

- Many statistical test are based upon the assumption that the data are sampled from a Gaussian distribution. These tests are referred to as parametric tests.
- Tests that do not make assumptions about the population distribution are referred to as nonparametric-tests.

Types of Tests II

- Another classification refers to the number of samples.
- Until now before we discus only tests from one sample.
- In many situations we want to compare two or more groups resp. samples.
- When comparing two or more groups, the sample can be independent or not.
  1. Independent samples: the individual values are not paired or matched with one another.
  2. Dependent samples:
     - the individual values represent repeated measurements on one subject (before and after an intervention) or
     - measurements on matched subjects, i.e. values in one group are more closely correlated with a specific value in the other group than with random values in the other group. The subjects were matched or paired before the data were collected.

Types of Tests III

- One sample tests are used to test if the population parameter is different from a specified value.
- Two sample tests are used to detect the difference between the parameters of two populations.
- In the following we will introduce two sample tests in the normal model and the binomial model.
- At the end of this chapter we will introduce some nonparametric tests to compare two samples.
- The following chapters are mainly based on Heumann, Schoemaker, Shalabh, chapter 10.3, 10.4, 10.5

Normal Model I

Testing two means in a normal model
Assumptions: two sample of normally distributed random variables

- $X_1, ..., X_{n_1}$: sample size n_1 with i.i.d random variables $X_i \sim N(\mu_1, \sigma_1^2)$
- $Y_1, ..., Y_{n_2}$: sample size n_2 with i.i.d random variables $Y_i \sim N(\mu_2, \sigma_2^2)$

The following cases must be distinguished.
compare: 2sample_tests_mean.R

Normal Model II

Case I - Variances are known: 2-sample Gauss test

- **Assumptions:** μ_1, μ_2 are unknown and σ_1, σ_2 are known
- **Hypothesis:**
  a) $H_0 : \mu_1 = \mu_2$, b) $H_0 : \mu_1 \geq \mu_2$ or c) $H_0 : \mu_1 \leq \mu_2$
- **Teststatistic:** $\sim N(0, 1)$

$$T(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}) = \frac{\bar{X}_{(n_1)} - \bar{Y}_{(n_2)} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- **Decision rule:** reject H_0 if
  a) $|T| > u_{1-\alpha/2}$, b) $T < u_\alpha$ or c) $T > u_{1-\alpha}$

Normal Model III

Example: Two machine produce bolts. The length of the bolts from machine 1 are $N(\mu_1, 0.25)$-distributed and from machine 2 are $N(\mu_2, 0.36)$-distributed. We have two i.i.d. sample from both machines:

x_i :	5.46	5.34	4.34	4.82	4.40	5.12	5.69	5.53
	4.77	5.82						
y_i :	5.45	5.31	4.11	4.69	4.18	5.05	5.72	5.54
	4.62	5.89	5.60	5.19	3.31	4.43	5.30	4.09

- Is the length of a bolt from machine 2 significantly ($\alpha = 0.05$) less than the length of a bolt from machine 1?
- $H_0 : \mu_1 \leq \mu_2$
- Since $T(x_1, ..., x_{10}, y_1, ...y_{15}) = 1.03 < u_{0.95} = 1.645$, H_0 can not be rejected.

Normal Model IV

Case II - Variances are identical but unknwon:

2-sample t-test

- **Assumptions:** μ_1, μ_2 are unknwon and $\sigma_1 = \sigma_2$ but unknown
- $S^2_{X,n_1} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x}_{(n_1)})^2$ and

  S^2_{Y,n_2} - analogously defined - are plausible estimators of σ^2, but they use only the informations given by the samples separately.
- The pooled sample variance

$$S_p^2 = \frac{(n_1 - 1)S^2_{X,n_1} + (n_2 - 1)S^2_{Y,n_2}}{n_1 + n_2 - 2}$$

  uses all informations given by the two samples and is a better unbiased estimator of σ^2.

Normal Model V

Hypothesis:
a) $H_0 : \mu_1 = \mu_2$, b) $H_0 : \mu_1 \geq \mu_2$ or c) $H_0 : \mu_1 \leq \mu_2$
Teststatistic:

$$T(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}) = \frac{\bar{X}_{(n_1)} - \bar{Y}_{(n_1)} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}}$$
$$\sim t_{n_1+n_2-2}$$

Decision rule: reject H_0 if
a) $|T| > t_{n_1+n_2-2,1-\alpha/2}$, b) $T < t_{n_1+n_2-2,\alpha}$ or
c) $T > t_{n_1+n_2-2,1-\alpha}$

Normal Model VI

Example: In a study the influence of a certain drug on the growth is examined. A sample of 15 animals are treated with this drug and a control group of 20 animals do not get this drug. From the observerd values of the growth $x_1, ..., x_{15}$ in the first group one get $\bar{x} = 72$ cm and $s_x = 13$ cm while the values of the control group $y_1, ..., y_{20}$ lead to $\bar{y} = 75$ cm and $s_y = 12$ cm.
Does the drug has any influence on the growth of the animals?

Assumptions: The growth in samples are indepedently identically $N(\mu_1, \sigma^2)$ resp. $N(\mu_2, \sigma^2)$ distributed with identical but unknown σ.

Test $H_0 : \mu_1 = \mu_2$ with $\alpha = 0.05$

Since $|T(x_1, ..., x_{15}, y_1, ..., y_{20})| = 0.7064$ and $t_{33,0.975} = 2.0345$, H_0 can not be rejected.

Normal Model VII

Case III - Variances are unequal and unknwon:
Welsh test

- **Assumptions:** μ_1, μ_2 are unknwon and $\sigma_1 \neq \sigma_2$ both unknown
- A plausible estimator of $\mathrm{Var}(\bar{X}_{(n_1)} - \bar{Y}_{(n_2)}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ is $\frac{S_{X,n_1}^2}{n_1} + \frac{S_{Y,n_2}^2}{n_2}$
- Exchanging the nominator in the test statistic of the two sample t-test by this estimator we get

$$T(X,Y) = \frac{\bar{X}_{(n_1)} - \bar{Y}_{(n_1)} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_{X,n_1}^2}{n_1} + \frac{S_{Y,n_2}^2}{n_2}}}$$

the test statistic of the Welsh test.

Normal Model VIII

Hypothesis:

a) $H_0 : \mu_1 = \mu_2$, b) $H_0 : \mu_1 \geq \mu_2$ or c) $H_0 : \mu_1 \leq \mu_2$

Teststatistic:

$$T(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}) = \frac{\bar{X}_{(n_1)} - \bar{Y}_{(n_1)} - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2_{X,n_1}}{n_1} + \frac{S^2_{Y,n_2}}{n_2}}} \sim t_\nu$$

with $\nu = \dfrac{\left(\frac{S^2_{X,n_1}}{n_1} + \frac{S^2_{Y,n_2}}{n_2}\right)^2}{\frac{(S^2_{X,n_1}/n_1)^2}{n_1-1} + \frac{(S^2_{Y,n_2}/n_2)^2}{n_2-1}}$

Decision rule: reject H_0 if

a) $|T| > t_{\nu,1-\alpha/2}$, b) $T < t_{\nu,\alpha}$ or c) $T > u_{\nu,1-\alpha}$

Example - compare two sample Gauss test:

$n_1 = 10, n_2 = 16, \alpha = 0.05$ and $H_0 : \mu_1 \leq \mu_2$

Since $T(x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}) = 0.9086$, $\nu = 23.372 \approx 23$ and $t_{23,0.95} = 1.713872$ the hypothesis H_0 can not be rejected.

Normal Model IX

Case IV - Two Paired Sample t-Test:

- Until now we have assumed that the two samples we have compared are independent samples.
- In practice we often have two groups of observations based on the same sample of subjects who were tested twice (e.g., before and after a treatment).
- By computing the difference between first score from the second for each subject we test whether the mean difference is significantly different from 0.

Normal Model X

- **Assumptions:** The paired differences are independently normally distributed (σ unknown).
- **Hypothesis:** a) $H_0 : \mu = 0$, b) $H_0 : \mu \leq 0$, c) $H_0 : \mu \geq 0$
- $\Rightarrow$ **Teststatistic:** $T(X_1, \ldots, X_n) = \sqrt{n} \cdot \frac{\bar{X}_{(n)}}{\sqrt{s^2_{(n)}}}$ is t_{n-1}-distributed, if $\mu = 0$ with $s^2_{(n)} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_{(n)})^2$
- **Decision rule:** reject H_0, if a) $|T| > t_{n-1,1-\frac{\alpha}{2}}$, b) $T > t_{n-1,1-\alpha}$, c) $T < -t_{n-1,1-\alpha}$

Normal Model XII

Example: A study investigated the cognitive effects of stimulant medication in children with mental retardation and Attention Deficit / Hyperactivity Disorder (ADHD). A sample of 24 children with ADHD was tested under two dosage levels (placebo D0 and dosage D60). Here we have one group of subjects, each subject being tested in both the D0 and D60 conditions.

Normal Model XIII

subject	1	2	...	24
D0	57	27	...	33
D60	62	49	...	29
D60-D0	5	22	...	-4

The mean difference score is 4.96 which is significantly different ($\alpha = 0.05$) from 0, since $T = 3.22 > t_{23,0.975} = 2.0687$. If you had mistakenly used the method for an independent-groups t test with these data, you would have found that $T = 1.42 < t_{46,0.975} = 2.019$. That is, the difference between means would not have been found to be statistically significant.

Remark: This is a typical result: paired t tests almost always produces "better"results (i.e., it is always more sensitive).

Normal Model XIV

Testing two Variances - F Test:

- **Assumptions:** 2 independent samples from normally distributed population with unknown expectations and variances.
- **Objective:** Tests concerning the variances of the populations
- $X_1, \ldots, X_{n_1}$ identically $N(\mu_1; \sigma_1^2)$-distributed
- $Y_1, \ldots, Y_{n_2}$ identically $N(\mu_2; \sigma_2^2)$-distributed
- μ_1, μ_2 and σ_1^2, σ_2^2 unknown

$\Longrightarrow$

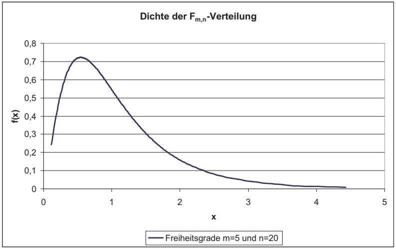
- $\frac{n_1-1}{\sigma_1^2} \cdot S^2_{(n_1)} = \frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} (X_i - \bar{X}_{(n_1)})^2 \sim \chi^2_{n_1-1}$
- $\frac{n_2-1}{\sigma_2^2} \cdot S^2_{(n_2)} = \frac{1}{\sigma_2^2} \sum_{i=1}^{n_2} (Y_i - \bar{Y}_{(n_2)})^2 \sim \chi^2_{n_2-1}$

$\Rightarrow \frac{S^2_{(n_1)}}{S^2_{(n_2)}} \sim F_{n_1-1, n_2-1}$, if $\sigma_1^2 = \sigma_2^2$

Normal Model XV

Definition: If Y is a χ^2_m-distributed random variable, Z is a χ^2_n-distributed random variable and both are independent, then is $\dfrac{Y/m}{Z/n} \sim F_{m,n}$.

$E(X) = \frac{n}{n-2}$ if $n > 2$,
$Var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ if $n > 4$



Remarks:

- Quantiles for $F_{m,n}$ distribution: $F_{m,n;1-p} = \dfrac{1}{F_{n,m;p}}$
- For big n the distribution of $m \cdot Z$ is approximately χ^2_m-distributed, if $Z \sim F_{m,n}$.

Normal Model XVI

Assumptions F-Test:

- two independent samples $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_2}$
- $X_i \sim N(\mu_1; \sigma_1^2)$ for $i = 1, ..., n_1$
- $Y_i \sim N(\mu_2; \sigma_2^2)$ for $i = 1, ..., n_2$
- μ_1, μ_2 and σ_1^2, σ_2^2 are unknown

a) $H_0 : \sigma_1 = \sigma_2$, b) $H_0 : \sigma_1 \leq \sigma_2$, c) $H_0 : \sigma_1 \geq \sigma_2$

$\Rightarrow$ **Teststatistic:**

$$T(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}) = \frac{S^2_{(n_1)}}{S^2_{(n_2)}} \sim F_{n_1-1, n_2-1},$$

if $\sigma_1^2 = \sigma_2^2$

Decision rule: reject H_0, if

a) $T < F_{n_1-1,n_2-1;\alpha/2}$ or $T > F_{n_1-1,n_2-1;1-\alpha/2}$,

b) $T > F_{n_1-1,n_2-1;1-\alpha}$, c) $T < F_{n_1-1,n_2-1;\alpha}$

Normal Model XVII

R-function: var.test(), package stats perfoms the F-test

Example: In a study two types of food for calves are examined. 9 of 22 calves get food of type A and the other 13 calves get food of type B. After a certain time period the gains in weight are observerd.

compare: 2sample_tests_var.R

group X	7.0	11.8	10.1	8.5	10.7	13.2	9.4	7.9	11.1				
group Y	13.4	14.6	10.4	11.9	12.7	16.1	10.7	8.3	13.2	10.3	11.3	12.9	9.7

Do both types of food have the same gains of weight, i.e. can we reject the hypothesis $H_0 : \mu_x = \mu_y$ *with* $\alpha = 0.05$*?*

We get from the sample values:

$$\bar{x} = 9.97, \bar{y} = 11.96, s_x^2 = 3.9, s_y^2 = 4.57$$

Normal Model XVIII

The basic assumptions of the unpaired t-test is variance homogeneity of both random variables X and Y. We first examine this assumption with the F-test.

F-test:

teststatistic F-test:	$\frac{s_x^2}{s_y^2} = 0.85$	$\alpha = 0.1$
quantiles:	$F_{8,12,0.05} = 0.30$	$F_{8,12,0.95} = 2.85$

$\Rightarrow$ The hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$ can not be rejected at level $\alpha = 0.1$. Thus we can apply the unpaired t-test.

Normal Model XIX

- teststatistic t-test:
$$\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1+n_2-2)}{n_1+n_2}} \cdot \frac{\bar{X}_{(n_2)}-\bar{Y}_{(n_2)}}{\sqrt{(n_1-1)s_x^2+(n_2-1)s_y^2}} = 2.21 \text{ with}$$
$n_1 = 8, n_2 = 12$
- $\alpha = 0.05$: $t_{20,0.95} = 2.09$
- $\Rightarrow$ The hypothesis $H_0 : \mu_x = \mu_y$ can be rejected at level $\alpha = 0.05$.

Remark: We use in the F-test a higher level as in the t-test, i.e. we are more critical towards the hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$. For $\alpha = 0.05$ H_0 we get a bigger rejection area

$$[F_{8,12,0.025}, F_{8,12,0.975}] = [0.24, 3.51]$$

as for $\alpha = 0.1$!

Binomial Model I

compare Heumann, Schomaker p. 230 ff

- two independent i.i.d. sample from Bernoulli distributions with parameters p_1, p_2

$$(X_1, X_2, ..., X_{n_1}), X_i \sim B(1, p_1)$$

$$(Y_1, Y_2, ..., Y_{n_2}), Y_i \sim B(1, p_2) \Rightarrow$$

$$X = \sum_{i=1}^{n_1} X_i \sim B(n_1, p_1), Y = \sum_{i=1}^{n_2} Y_i \sim B(n_2, p_2)$$

- As in the one-sample case both exact and approximate tests exists. The exact is called **exact test of Fisher**. Here we consider only the approximate test.
- Since $\frac{X}{n_1}$ resp. $\frac{Y}{n_2}$ are approximately $N(p_i, \frac{p_i(1-p_i)}{n_i})$, i=1,2 the difference is approximately normally distributed, too: $D = \frac{X}{n_1} - \frac{Y}{n_2} \sim N(0, p(1-p)(\frac{1}{n_1} + \frac{1}{n_2}))$ under $H_0 : p_1 = p_2$

Binomial Model II

- Under $H_0 : p = p_1 = p_2$ we can estimate p by $\hat{p} = \frac{X+Y}{n_1+n_2}$
- If n_1 and n_2 are sufficiently large and p is not close to 0 and 1, the test statistic T approximately follows a standard normal distribution

$$T = \frac{D}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} \sim N(0,1)$$

$\Rightarrow$ Test can be conducted for the one-sided and two-sided case as the Gauß-Test.

Binomial Model III

Example: Consider two lotteries

- A: 14 winning tickets in a sample of 63 tickets
- B: 13 winning tickets in a sample of 45 tickets

We want to test whether the probabilities of winning are different, i.e. $H_0 : p_1 = p_2, H_1 : p_1 \neq p_2$ at 5%-level.

- $\hat{p}_A = 14/63, \hat{p}_B = 13/45, \hat{d} = \hat{p}_A - \hat{p}_B = -1/15$
- Under H_0 an estimate of p is $\hat{p} = (14 + 13)/(65 + 45) = 0.25.$
- value of the test statistics: $t = \frac{-1/15}{\sqrt{0.25(1-0.25)(1/63+1/45)}} = -0.79$
- H_0 not rejected, since $= 0.79 < 1.96 = u_{0.975}$, i.e. no statistical evidence for different winning probabilities.
- **compare:** 2sample_tests_binom.R

χ^2- Goodness-of-Fit Test I

Objective: testing the goodness of fit for the observed data to a given distribution f_0.

Example: To check if a die is fair it has been thrown 600 times.

n	frequency
1	86
2	117
3	109
4	73
5	105
6	110
Σ	600

We have k=6 possible values of the random variable (rolling a die), n=600 observations and observed frequencies n_i given in the table.

H_0 : die is fair, i.e. $H_0 : f = f_0 = R[1, 2, ...6]$

χ^2- Goodness-of-Fit Test II

- test statistic: $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$, which measures the difference between the observerved frequencies n_i and under H_0 expected frequencies np_i
- Under the null hypothesis for $n \to \infty$ the distribution of χ^2 converges to the chi-square distribution with $k-1$ degrees of freedom χ^2_{k-1}.
- **Decision rule:** Reject $H_0 : f = f_0$ with significance level α, if $\chi^2 > \chi^2_{k-1,1-\alpha}$
- **Rule of thumb:** If $n \cdot p_i > 5$ for all i, the χ^2_{k-1} distribution can be used as an approximation of the distribution of χ^2.

χ^2- Goodness-of-Fit Test III

- **Example:** Can we conjecture that the die is a fair one with significance level $\alpha = 0.05$?
  Decision Rule: reject H_0, if $\chi^2 > \chi^2_{5,0.95} = 11.0705$
  Since $\chi^2 = 14.2$, H_0 can be rejected.
- A χ^2-goodness-of-fit test can be used to compare two distributions of nominal scaled random variables.

χ^2- Goodness-of-Fit Test IV

- A χ^2-test can be used to compare two distributions of nominal scaled random variables.
- **Example:** In two samples of students from the universities of Bochum and Muenchen of the same size we got:

study	Bochum	Muenchen
social science	125	145
natural science	100	85
medicine	50	40
else	60	65

Hypothesis: The choice of the studies does not differ in both universities, i.e. both sample are from the same population.

- These type of tests are called χ^2-tests of homogeneity.

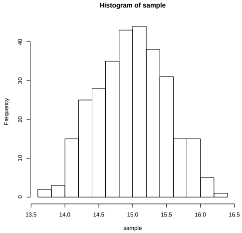
χ^2- Goodness-of-Fit Test V

- $\Omega = \{$social science, natural science, medicine, else$\}$
- $n = 335$, $k = 4$
- $np_i =$ number of students in study i from Muenchen.

H_0 : Both samples are from the same population; $\alpha = 0.1$
Decision rule: reject H_0 if $\chi^2 > \chi^2_{k-1,1-\alpha}$
Since
$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i}$ with
$\chi^2 = \frac{(125-145)^2}{145} + \frac{(100-85)^2}{85} + \frac{(50-40)^2}{40} + \frac{(60-65)^2}{65} = 8.290295$
and $\chi^2_{3,0.9} = 6.251389$, we reject H_0.

χ^2- Goodness-of-Fit Test VI

- In case of ordinal or continous variables, the number of different values can be large. Then it necessary to group the data into k intervals before applying the test.
- **Example:** 200 diameters of screws (in mm) Histogram and frequency table for equidistant classes



Frequency table:

from	to	counts
	13.8	2
13.8	14	3
14	14.2	15
14.2	14.4	25
14.4	14.6	28
14.6	14.8	35
14.8	15	43
15	15.2	44
15.2	15.4	38
15.4	15.6	31
15.6	15.8	15
15.8	16	15
16	16.2	5
16.2		1

χ^2- Goodness-of-Fit Test VII

- H_0 : The data follow a normal distribution.
  H_1 : The data do not follow a normal distribution.
- It can be shown, that the test statistic χ^2 is approximately χ^2_{k-r-1} distributed with r number of estimated parameters (here: $r = 2$) and k number of the classes.
- If the parameters are known, r is 0 and p_i is computed using the known parameters.
- **Decision rule** reject H_0, if $\chi^2 > \chi^2_{k-r-1,1-\alpha}$.

χ^2- Goodness-of-Fit Test VIII

from	to	counts	exp. counts
	14	5	8.345
14	14.2	15	10.667
14.2	14.4	25	19.117
14.4	14.6	28	29.554
14.6	14.8	35	39.411
14.8	15	43	45.335
15	15.2	44	44.985
15.2	15.4	38	38.505
15.4	15.6	31	28.430
15.6	15.8	15	18.107
15.8	16	15	9.948
16		6	7.596

- After merging some classes: $n = 300$ observations in $k = 15$ classes. The estimation of the parameters are $\bar{x}_{(n)} = 14.9895$ and $\hat{s} = \sqrt{\frac{n-1}{n} \cdot s^2} = 0.5169179$.
- Value of the test statistic: $\chi^2 = 9.30662$.
- For $\alpha = 0.05$ we get $\chi^2_{\hat{k}-3,1-\alpha} = 16.91898$. Since the value of χ^2 is far away from rejection, there is no reason to doubt on a normal distribution.

χ^2- Goodness-of-Fit Test IX

R function: chisq.test(x, p) performs chi-squared goodness-of-fit tests.

- x: numeric vector
- p: a vector of probabilities of the same length of x
- The hypothesis is tested whether the population probabilities equal those in p, or are all equal if p is not given.
- **compare:** chi2_tests.r

χ^2-Test: Associations of qualitative variables I

Example: Data from a Mediterranean Diet and Health case study

	Outcome				
Diet	Cancers	Fatal Heart Disease	Non-Fatal Heart Disease	Healthy	Total
AHA	15	24	25	239	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Is there a significant relationship between diet and outcome?

We compute: $\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ where O_{ij} resp. E_{ij} are the observed resp. expected frequencies for cell i, j.

χ^2-Test: Associations of qualitative variables II

Observed and Expected Frequencies for Diet and Health Study

	Outcome				
Diet	Cancers	Fatal Heart Disease	Non-Fatal Heart Disease	Healthy	Total
AHA	15 (11.02)	24 (19.03)	25 (16.53)	239 (256.42)	303
Mediterranean	7 (10.98)	14 (18.97)	8 (16.47)	273 (255.58)	302
Total	22	38	33	512	605

χ^2-Test: Associations of qualitative variables III

Generally: We want to test whether the components X, Y of the two dimensional random variable (X, Y) are independent. The sets of the possible values of X and Y are divided in r and k disjoint sets $I_1, ..., I_r$ and $J_1, ..., J_s$. Let

$p_{ij} = P(X \in I_i, Y \in J_j), p_{i\cdot} = \sum_{j=1}^{s} p_{ij}, p_{\cdot j} = \sum_{i=1}^{r} p_{ij}$

- If X and Y are independent, we have for all $i, j : p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$.
- $H_0 : p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ for all i, j versus $H_1 : p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j}$ for at least one pair i, j
- The teststatistic $\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ with $E_{ij} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{n}$ and $n =$ total sum is under H_0 χ^2 approximatly $\chi^2_{(r-1)(s-1)}$ distributed.
- **Decision rule:** reject H_0, if $\chi^2 > \chi^2_{(r-1)(s-1),1-\alpha}$

χ^2-Test: Associations of qualitative variables IV

- **Example:** The degrees of freedom is $(r-1)(c-1) = (2-1)(4-1) = 3$. For $\alpha = 0.01$ we get $\chi^2_{3,0.99} = 16.55445$. Since the value of teststatistic χ^2 is 11.34487 the null hypothesis of no relationship between diet and outcome can not be rejected.
- **Remark:** The formula for Chi Square yields a statistic that is only approximately a Chi Square distribution. In order for the approximation to be adequate, the total number of subjects should be at least 20.

χ^2-Test: Associations of qualitative variables V

R function: chisq.test(x, y=NULL) performs chi-squared contingency table tests

- x numeric matrix with at least two rows and columns, y not given: x is taken as a contingency table.
- Otherwise, x and y must be vectors of the same length and the contingency table is computed from these.
- **compare:** chi2_tests_qual.r

Wilcoxon-Mann-Whitney U-Test I

The two sample tests we already know require that data follow a normal distribution. Sometimes this is not the case.

- Samples are from highly-skewed distributions: Transforming data by taking logarithms or square roots somtimes make them follow a normal distribution.
- Sample size is so small that it is difficult to ascertain whether or not the data a normally distributed.

Non parametric tests do not require the data to follow a particular distribution. The Wilcoxon-Mann-Whitney U-test is an example of these tests.

Wilcoxon-Mann-Whitney U-Test II

- unpaired samples.
- **Assumptions:**
  - $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ are independent random variables.
  - $X_1, \ldots, X_{n_1}$ are identically distributed with a continous distribution F.
  - $Y_1, \ldots, Y_{n_2}$ are identically distributed with a continous distribution G.

  $$H_0 : F = G \quad \text{versus} \quad H_1 : F > G \text{ or } F < G$$

  We are comparing the entire distributions of X and Y. If there is location shift in the sense that one distribution is shifted left or right compared with the other distribution, H_0 will be rejected.

Wilcoxon-Mann-Whitney U-Test III

The teststatistic is computed from the ranks of the ordered sample values:

$$x_{i_1} < x_{i_2} < y_{j_1} < x_{i_3} < \ldots < x_{i_{n_1}} < y_{j_{n_2}}$$

1. Taking each observation in sample X and count the number of observations in sample Y that are smaller than it.
2. The total of these counts is U_Y the teststatistic of the U-test.
3. U_Y can be evaluated using R_Y, i.e. sum of the ranks for the observations which came from sample Y.

$$U_Y = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_Y$$
$$U_X = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_X$$

Wilcoxon-Mann-Whitney U-Test IV

Example: Let $x_1, ..., x_7$ and $y_1, ..., y_6$ be two samples with pairwise different values. If the form of the ranked observations is

$$x \quad x \quad y \quad y \quad y \quad x \quad x \quad y \quad x \quad x \quad y \quad y \quad x$$

we get:

$n_1 = 7$, $n_2 = 6$,
$U_Y = 3 \cdot 5 + 1 \cdot 3 + 2 \cdot 1 = 20$,
$U_X = 2 \cdot 3 + 4 \cdot 1 + 6 \cdot 2 = 22$,
$R_X = 1 + 2 + 6 + 7 + 9 + 10 + 13 = 48$,
$R_Y = 3 + 4 + 5 + 8 + 11 + 12 = 43$.

$$20 = U_Y = 6 \cdot 7 + 0.5 \cdot 6 \cdot 7 - 43$$

$$22 = U_X = 6 \cdot 7 + 0.5 \cdot 7 \cdot 8 - 43$$

Wilcoxon-Mann-Whitney U-Test V

- If H_0 is valid, U_Y is symmetrically distributed with $E(U_Y) = 0.5 \cdot n_1 \cdot n_2$.
- **Decision rule:** reject H_0, if $U_Y < k$ or $U_Y > n_1 \cdot n_2 - k$ with $k = w^{U_Y}_{n_1,n_2;\alpha/2}$ quantil of the U_Y-distribution, i.e.

$$P_{H_0}(U_Y < k) = P_{H_0}(U_Y > n_1 n_2 - k) \leq \alpha/2$$

Wilcoxon-Mann-Whitney U-Test VI

Example: In a genetic inheritance study samples of individuals from several ethnic groups were taken. Blood samples were collected from each individual and several variables measured. We compare the groups labeled “Native American” and “Caucasian” with respect to the variable MSCE (mean sister chromatid exchange). The data is as follows:

Native American	8.50	9.48	8.65	8.16	8.83	7.76	8.63		
Caucasian	8.27	8.20	8.25	8.14	9.00	8.10	7.20	8.32	7.70

Wilcoxon-Mann-Whitney U-Test IX

Ordering all sample values result in:

$C \quad C \quad A \quad C \quad C \quad A \quad C \quad C \quad C \quad C \quad A \quad A \quad A \quad A \quad C \quad A$

group	A	A	A	A	A	A	A	C	C	C	C	C	C	C	C	C
value	8.5	9.48	8.65	8.16	8.83	7.76	8.63	8.27	8.2	8.25	8.14	9	8.1	7.2	8.32	7.7
rank	11	16	13	6	14	3	12	9	7	8	5	15	4	1	10	2

$$n_A = 7, R_A = 75, U_A = 16, n_C = 9, R_C = 61, U_C = 47$$

Since $w^C_{7,9,0.025} = 13 \leq U_C \leq w^C_{7,9,0.975} = 50$ the null hypothesis, which says that the MSCE distribution for Native Americans is the same as that for Caucasians, can not be rejected.

Wilcoxon-Mann-Whitney U-Test X

Remark:

- The values of the quantiles of U_Y-distribution can be evaluated by the R-function qwilcox().
- For big n and m, i.e.

$$n_1 \geq 4, \quad n_2 \geq 4, \quad n_1 + n_2 \geq 20$$

  a normal approximation of U_Y is usefull. Since $E(U_Y) = \frac{n_1 n_2}{2}$ and $Var(U_Y) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$ we get

$$w^{U_Y}_{n_1,n_2,p} \approx \frac{1}{2} n_1 \cdot n_2 + u_p \cdot \sqrt{\frac{1}{12} n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}$$

- The above procedure assume that all values of the samples from X and Y do not have identical values. In the case of some identical values (ties) in both samples, a slightly different test-procedure is valid.

Wilcoxon-Mann-Whitney U-Test XI

- **Remark:** The test can be conducted by the R-function wilcox.test().
  wilcox.test(A, C, conf.level=0.05) computes U_C and conduct the test at level alpha:

  > *Wilcoxon rank sum exact test*
  > *data: A and C*
  > *W = 47, p-value = 0.1142*
  > *alternative hypothesis: true location shift is not equal to 0*

  wilcox.test(C, A, conf.level=0.05) computes U_A.
- **compare:** 2sample_tests_par_free.R

Content

Inferential Statistics
Introduction
Point Estimator
Characteristics of Estimators
Maximum Likelihood Method
Estimating the Mean
Estimating the Variance
Confidence Intervals
Introduction to Confidence Intervals
Confidence Intervals for Normal Distributions: parameter μ
Confidence Intervals for Normal Distributions: parameter σ^2
Confidence Intervals for Normal Distributions: Summary
Confidence Intervals on the Proportion
Hypothesis Testing
Logic of Hypothesis Testing
Basic Model
Parameter Tests in the Normal Model
Tests in the Bernoulli Model
Two Sample Tests
Types of Tests
Normal Model
Binomial Model