

1 R Tutorial

1.1 Loading Data

```

1 # loading csv files
2 data <- read.table("whatever.csv", sep=";", header=T)
3
4 # csv files can be stored with (almost) any kind of file ending, e.g.:
5 data <- read.table("whatever.dat", sep=";", header=T)
6 data <- read.table("whatever.txt", sep=";", header=T)

```

2 Probability And Statistics

2.1 Probability Models for Measurement Data

2.1.1 Random Variables

Random Variables												
Definition	$X : \Omega \longrightarrow W_x$											
Example	<p>A Coin is thrown three times, head and tails is observed:</p> $\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$ <p>Total number of heads $W_x = \{0, 1, 2, 3\}$</p> <p>Total number of tails $W_x = \{0, 1, 2, 3\}$</p> <p>Number of heads minus tails $W_x = \{-3, -1, 1, 3\}$</p>											
Probability Mass Function												
Definition	The probability distribution of a discrete random variable: $P(X = x)$											
Example	<table><tr><td>x</td><td>0</td><td>1</td><td>2</td><td>3</td></tr><tr><td>$P(X = x)$</td><td>$\frac{1}{8}$</td><td>$\frac{3}{8}$</td><td>$\frac{3}{8}$</td><td>$\frac{1}{8}$</td></tr></table>	x	0	1	2	3	$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	
x	0	1	2	3								
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$								

2.1.2 Probability Distributions

Cumulative Density Function (cdf)	
Definition	$F(x) = P(X \leq x)$
Properties	$P(a < X \leq b) = F(b) - F(a)$ $0 \leq F(x) \leq 1$ $P(X = a) = F(a) - F(a) = 0$

Probability Density Function (pdf)	
Definition	$f(x) = \frac{dF(x)}{dx}$
Properties	$f(x) \geq 0$ $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$ $\int_{-\infty}^{\infty} f(x)dx = 1$

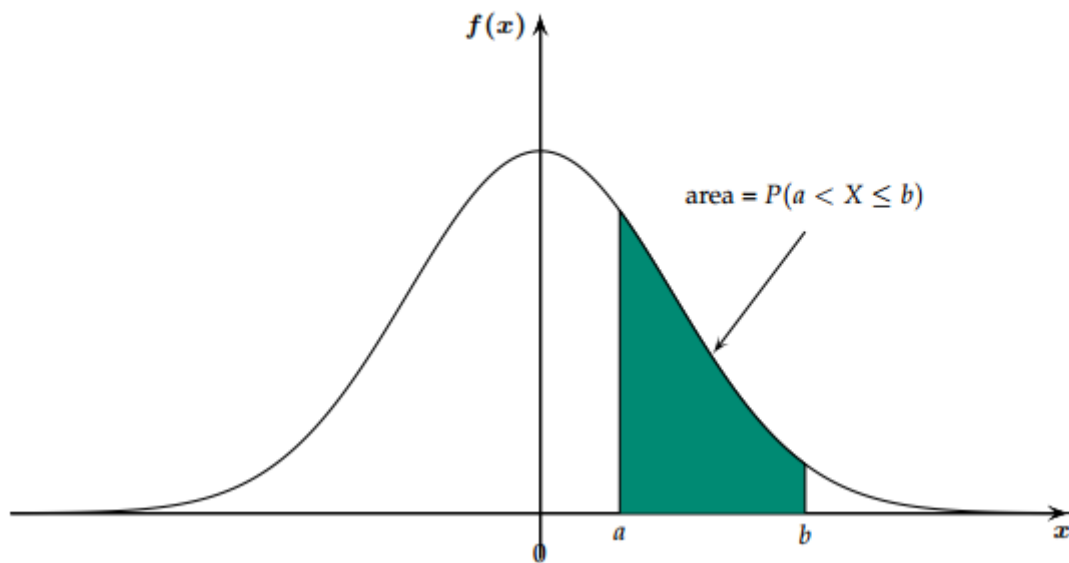


Figure 1: Probability density of a random variable and the probability of measuring a value from (a,b]

2.1.3 Summary Statistics of Continuous Distributions

Expected Value, Variance and Quantile	
Expected value	Discrete: $E(X) = \sum_i x_i P(X = x_i)$ Continuous: $E(X) = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x)dx$
Variance	$\text{Var}(X) = \sigma_x^2 = E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)dx$
Quantile	$P(X \leq q(\alpha)) = \alpha$ $F(q(\alpha)) = \alpha \Leftrightarrow q(\alpha) = F^{-1}(\alpha)$ <i>Note: When you're asked for the 50%-quantile, that means $\alpha = 50\%$, and you must find $q(0.5)$</i>
Example Body Length	If $\alpha=0.75$ and the corresponding quantile is $q(\alpha)=182.5\text{cm}$ then 75% of the persons is shorter or equal 182.5cm.

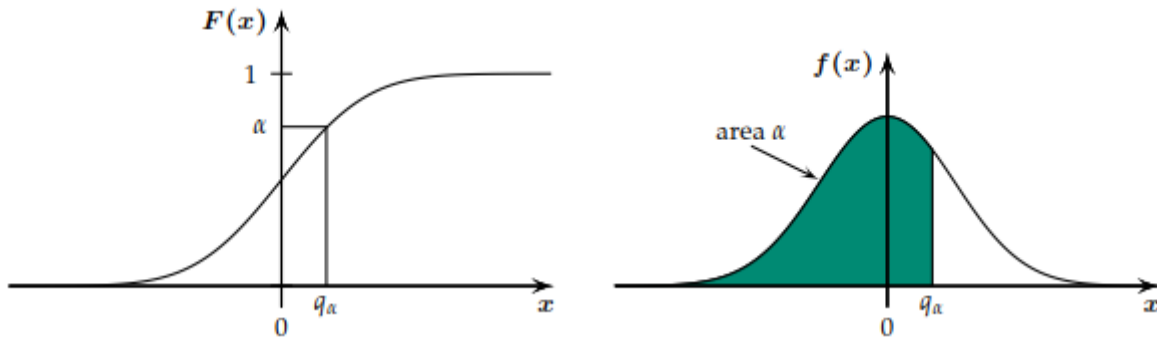


Figure 2: Quantiles

2.1.4 Important Distributions

2.1.4.1 Uniform Distribution

Theory	Code Example
$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$ $F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$ $E(x) = \frac{a+b}{2}$ $\text{Var}(x) = \frac{(b-a)^2}{12}$ $\sigma_x = \frac{b-a}{\sqrt{12}}$	<pre> 1 # value of the probability density function Uniform([1, 10]) at the position x = 5 2 dunif(x=5, min=1, max=10) 3 [1] 0.1111111 4 5 # P(X <= 5) 6 punif(q=5, min=1, max=10) 7 [1] 0.4444444 8 9 # P(1.2 < X <= 4.8) 10 punif(4.8, 1, 10) - punif(1.2, 1, 10) 11 [1] 0.4 12 13 # 5 uniformly distributed random values in Uniform([1, 10]) 14 runif(5,min=1,max=10) 15 [1] 1.061933 6.484813 5.928334 8.459887 8.852405 16 17 # TODO: ADD MORE HERE </pre>

2.1.4.2 Exponential Distribution

Theory	Code Example
$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$ $F(x) = \begin{cases} 1 - \lambda \cdot e^{-\lambda \cdot x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$ $E(x) = \frac{1}{\lambda}$ $\text{Var}(x) = \frac{1}{\lambda^2}$ $\sigma_x = \frac{1}{\lambda}$	<pre> 1 # P(0 <= X <= 4) of X ~ Exp(3) 2 pexp(4, rate=3) 3 [1] 0.9999939 4 5 # TODO: ADD MORE HERE </pre>

2.1.4.3 Normal Distribution

Theory	Code Example
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $F(x) = \int_{-\infty}^x f(y) dy$ $E(x) = \mu$ $\text{Var}(x) = \sigma^2$ $\sigma_x = \sigma$	<pre> 1 # X~N(u, sigma^2) --> X~N(100,15^2) 2 # In R we compute P(X>130) as 1 - P(X<=130) 3 1-pnorm(130, mean=100, sd=15) 4 [1] 0.02275013 5 6 #P(85<=X<=115) 7 pnorm(115, mean=100, sd=15)-pnorm(85, mean=100, 8 sd=15) 9 [1] 0.6826895 10 # TODO: ADD MORE HERE </pre>

2.1.4.4 Linear Transformation of Random Variables

Properties of Linear Transformation of a Random Variable	
Definition	<p>For $Y = a + bX$ the following apply</p> <p>(i) $E(Y) = a + bE(X)$</p> <p>(ii) $\text{Var}(Y) = b^2\text{Var}(X)$, $\sigma_Y = b \sigma_X$</p> <p>(iii) $\alpha - \text{Quantile of } Y = q_Y(\alpha) = a + bq_X(\alpha)$</p> <p>(iv) $f_Y(y) = \frac{1}{ b } f_X\left(\frac{y-a}{b}\right)$</p>
Summary Statistics of S_n and \bar{X}_n	
Summary Statistics of Sample Total S_n	$E(S_n) = E(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n E(X_i) = n\mu$ $\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) = n\text{Var}(X_i)$ $\sigma(S_n) = \sqrt{n}\sigma_X$
Summary Statistics of Sample Mean \bar{X}_n	$E(\bar{X}_n) = E\left(\frac{X_1+X_2+\dots+X_n}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} nE(X_i) = \mu$ $\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} n\sigma_X^2 = \frac{\sigma_X^2}{n}$ <p>Standard Error</p> $\sigma(\bar{X}_n) = \frac{\sigma_X}{\sqrt{n}}$

2.1.4.5 Distributions of S_n and \bar{X}_n

Theory	Code Example
<p>1. For $X_i \in \{0, 1\}$, we have</p> $S_n \sim \text{Bin}(n, \pi) \text{ with } \pi = P(X_i = 1)$ <p>2. For $X_i \sim \text{Pois}(\lambda)$, we have</p> $S_n \sim \text{Pois}(n\lambda)$ <p>3. For $X_i \sim N(\mu, \sigma^2)$</p> $S_n \sim N(n\mu, n\sigma^2) \text{ and } \bar{X}_n \sim N\left(\mu, \frac{\sigma_X^2}{n}\right)$	<pre> 1 What is the probability that among 10000 tosses 2 of a fair coin, heads would appear in 3 maximum 5100 cases? 4 #Approximated: X~N(5000,2500) 5 pnorm(5100,5000,sqrt(2500)) 6 [1] 0.9772499 7 8 # "True Result": X~Bin(10000,0.5) 9 pbinom(5100,10000,0.5) 10 [1] 0.9777871 </pre>

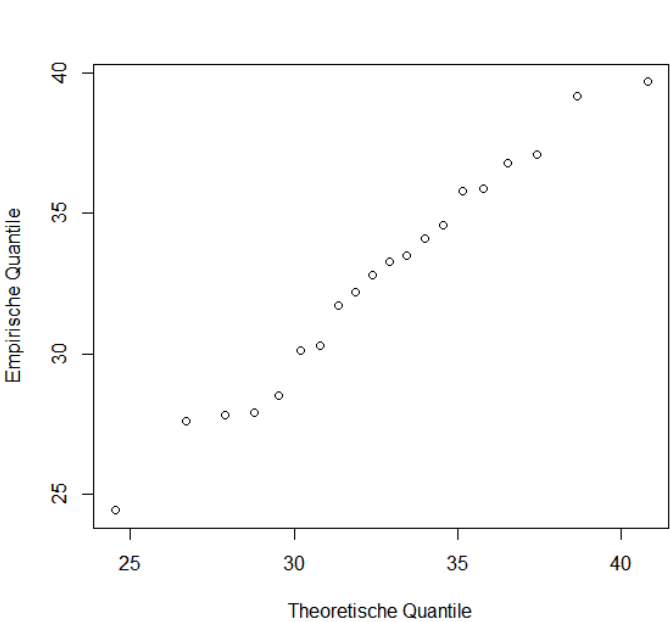
2.2 Statistics for Measurement Data

bgg

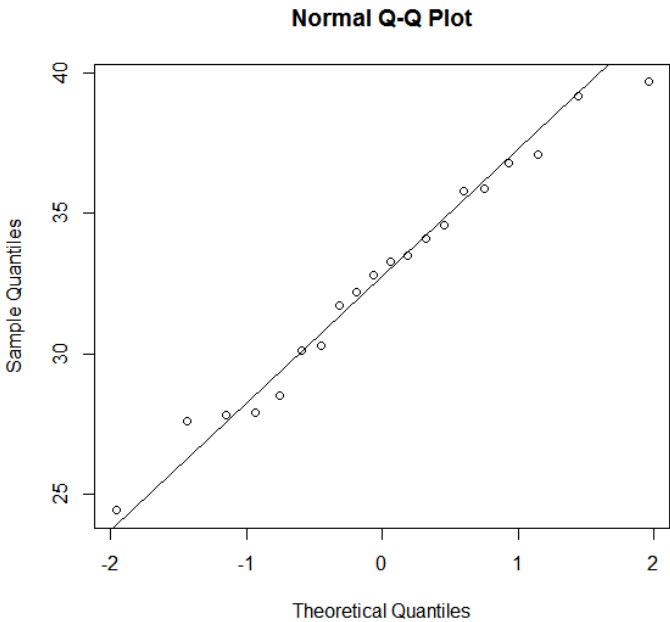
2.2.1 Assess the Normal Distribution Assumption

2.2.1.1 Q-Q Plot

Theory	Code Example
<div>1. For$\alpha_k = \frac{k-0.5}{n}$with $k = 1, \dots, n$calculate the corresponding theoretical quantiles of the model distribution$q(\alpha_k) = F^{-1}(\alpha_k)$</div> <div>2. Determine the empirical α_k-quantiles,$x_{(1)} < x_{(2)} < \dots < x_{(n)}$</div> <div>3. Plot the empirical quantiles x_k on the y-axis against the theoretical quantiles $q(\alpha_k)$ on the x-axis.</div>	<pre>1 x <- c(24.4, 27.6, 27.8, 27.9, 28.5, 30.1, 30.3, 2 31.7, 32.2, 32.8, 33.3, 33.5, 34.1, 34.6, 3 35.8, 35.9, 36.8, 37.1, 39.2, 39.7) 4 5 alpha_k <- (seq(1, length(x), by=1)-0.5)/length(x) 6 7 quantile_th <- qnorm(alpha_k, mean=mean(x), sd=sd(x)) 8 9 quantile_emp <- sort(x) 10 #image qqplot 11 qqplot(quantile_th, quantile_emp, xlab="Theoretische Quantile", ylab="Empirische Quantile") 12 #image qqnorm;qqline 13 qqnorm(x);qqline(x)</pre>



(a) qqplot()



(b) qqnorm();qqline()

k	$x_{(k)}$	$\alpha_k = (k - 0.5)/n$	q_{α_k} for $\mathcal{N}(32.7, 4.15^2)$	$\Phi^{-1}(\alpha_k)$
1	24.4	0.0250	24.5	-1.96
2	27.6	0.075	26.7	-1.44
3	27.8	0.125	27.9	-1.15
4	27.9	0.175	28.8	-0.935
5	28.5	0.225	29.5	-0.755
6	30.1	0.275	30.2	-0.600
7	30.3	0.325	30.8	-0.453
8	31.7	0.375	31.3	-0.319
9	32.2	0.425	31.9	-0.189
10	32.8	0.475	32.4	-0.0627
11	33.3	0.525	32.9	0.0627
12	33.5	0.575	33.4	0.189
13	34.1	0.625	34.0	0.319
14	34.6	0.675	34.5	0.454
15	35.8	0.725	35.1	0.598
16	35.9	0.775	36.0	0.755
17	36.8	0.825	36.5	0.935
18	37.1	0.875	37.4	1.15
19	39.2	0.925	38.6	1.44
20	39.7	0.975	40.8	1.96

```

1 #x(k) are the measured values N(u,sigma^2)
2 x <- c(24.4, 27.6, 27.8, 27.9, 28.5, 30.1, 30.3,
3       31.7, 32.2, 32.8, 33.3, 33.5, 34.1, 34.6, 35.8,
4       35.9, 36.8, 37.1, 39.2, 39.7)
5 mean(x)
6 [1] 32.665
7 sd(x)
8 [1] 4.149734
9 #N(32.7, 4.15)
10 #a_k = (k-0.5)/n = qnorm(q_ak, 32.7, 4.15)
11 pnorm(24.5, 32.7, 4.15)
12 [1] 0.02408285
13 pnorm(32.4, 32.7, 4.15)
14 [1] 0.4711859
15 pnorm(35.8, 32.7, 4.15)
16 [1] 0.7724646
17 pnorm(40.8, 32.7, 4.15)
18 [1] 0.9745195
19 #q_ak for N(32.7, 4.15) = qnorm(a_k, 32.7, 4.15)
20 qnorm(0.025, 32.7, 4.15)
21 [1] 24.56615
22 qnorm(0.475, 32.7, 4.15)
23 [1] 32.43977
24 qnorm(0.725, 32.7, 4.15)
25 [1] 35.1807
26 qnorm(0.975, 32.7, 4.15)
27 [1] 40.83385
28 #phi^{-1}(a_k)
29 qnorm(0.025)
30 [1] -1.959964
31 qnorm(0.475)
32 [1] -0.06270678
33 qnorm(0.725)
34 [1] 0.5977601
35 qnorm(0.975)
36 [1] 1.959964

```

2.2.2 Parameter Estimation for Continuous Probability Distributions

Method of Moments (not unbiased)

1. We consider our data measurements x_1, x_2, \dots, x_n as realization of random variables X_1, X_2, \dots, X_n originating from the same known distribution.

2. We calculate the expected value $E(X)$ and solve the equation for the unknown parameter that we intend to estimate.

3. We replace the expected value with its counterpart, the empirical mean value and obtain an estimate of the unknown parameter. A method of moments estimate of the standard deviation is the empirical standard deviation.

$$\mu = E(X) \Rightarrow \hat{\mu} = \bar{x}_n = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{653.3}{20} = 32.7$$

$$\sigma^2 = E(X^2) - E(X)^2 = E(X^2) - \mu^2$$

$$\hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$$

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

$$\hat{\sigma}^2 = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2} = \sqrt{\frac{1}{20} \sum_{i=1}^{20} (x_i - 32.7)^2} = 4.04$$

Method of Maximum Likelihood

We have n observations that are i.i.d.

For a discrete probability distribution: probability that these n observations (events) actually have occurred can be expressed as follows

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$$

$$P[(X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n)] = P[X_1 = x_1] \cdot P[X_2 = x_2] \cdot \dots \cdot P[X_n = x_n] = \prod_{i=1}^n P[X_i = x_i]$$

<p>Probability that the n independent random variables x_1, x_2, \dots, x_n are observed, depends on parameter θ, which we wish to estimate. Therefore the Likelihood function is given by $L(\theta)$ where $P[X_i = x_i \theta]$ denotes probability mass function that value x_i has been observed, given the parameter value θ.</p> <p>Idea of Maximum Likelihood : estimate the parameter θ in such a way that the likelihood is maximized, that is, that it makes the observed data most likely or most probable.</p> <p>Continuous probability distributions : with probability density function $f(x; \theta)$. Probability, that each observation x_i falls into its corresponding interval $[x_i, x_i + dx_i]$:</p> <p>Infinitesimal intervals dx_i do not depend on the parameter value θ : we omit them in the likelihood function</p> <p>If assumed probability density function $f(x_i; \theta)$ and parameter value of θ are correct, we expect a high probability for the actually observed data to occur : maximization of $L(\theta)$</p>	$L(\theta) = P[X_1 = x_1 \theta] \cdot P[X_2 = x_2 \theta] \cdot \dots \cdot P[X_n = x_n \theta] = \prod_{i=1}^n P[X_i = x_i \theta]$ $\prod_{i=1}^n f(x_i; \theta) dx_i$ $\prod_{i=1}^n f(x_i; \theta)$
Example: Maximum Likelihood for Exponential Distribution	
Let X_1, X_2, \dots, X_n i.i.d. $\sim \text{Exp}(\lambda)$, that is	$f(x_i; \lambda) = \lambda e^{-\lambda x_i}$
Likelihood function for a given data set x_1, x_2, \dots, x_n is given by	$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i}$
Log likelihood function is	$\log(L(\lambda)) = n \log(\lambda) - \lambda \sum_{i=1}^n x_i$
If we calculate the derivative of the log likelihood function with respect to λ and set it equal to 0, then we obtain	$\frac{d \log(L(\lambda))}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i \stackrel{!}{=} 0$
The maximum likelihood estimate $\hat{\lambda}$ thus corresponds to the solution of the previous equation	$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$

2.2.3 Statistical Tests and Confidence Interval for Normally Distributed Data

z-Test (σ_x known)	
1. Model:	X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma_X^2)$, σ_X known
2. Null hypothesis:	$H_0: \mu = \mu_0$
Alternative:	$H_A: \mu \neq \mu_0$ (or $<$ or $>$)
3. Test statistic:	$Z = \frac{(\bar{X}_n - \mu_0)}{\sigma_{\bar{X}_n}} = \frac{(\bar{X}_n - \mu_0)}{\sigma_{X_n}/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_{X_n}} = \frac{\text{observed} - \text{expected}}{\text{standard error}}$
Null distribution (assuming H_0 is true):	$Z \sim N(0, 1)$
4. Significance level:	α
5. Rejection region for the test statistic:	$K = (-\infty, z_{\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, \infty)$ with $H_A: \mu \neq \mu_0$, $K = (-\infty, z_{\alpha}]$ with $H_A: \mu < \mu_0$, $K = [z_{1-\alpha}, \infty)$ with $H_A: \mu > \mu_0$
where	$z_{\frac{\alpha}{2}} = \Phi^{-1}(\alpha/2)$

6. Test decision:	Check whether the observed value of the test statistic falls into the rejection region.
-------------------	---

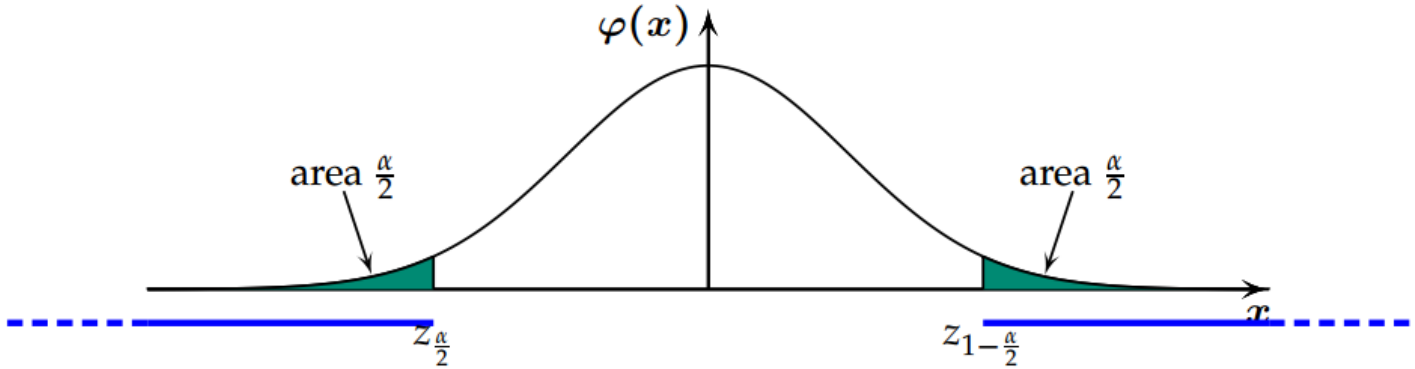


Figure 4: z-Test: Rejection Region

z-Test (σ_x known): Example	
Measurement of fusion heat:	The empirical mean value of $n = 13$ measurements is 80.02. From previous measurements the standard deviation is $\sigma_X = 0.01$. Is a fusion heat of exactly $80.00 \frac{\text{g}}{\text{cal}}$ plausible?
1. Model:	X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma_X^2)$, $\sigma_X = 0.01$ known, $n = 13$
2. Null hypothesis:	$H_0: \mu = \mu_0 = 80.00$
Alternative:	$H_A: \mu \neq \mu_0$
3. Test statistic:	$Z = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_{X_n}}$
Null distribution (assuming H_0 is true):	$Z \sim N(0, 1)$
4. Significance level:	$\alpha = 0.05$ (commonly used α -level)
5. Rejection region for the test statistic:	$K = (-\infty, z_{\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, \infty)$ with $H_A: \mu \neq \mu_0$
Given $\alpha = 0.05$, R yields the following 2.5% quantile of the standard normal distribution.	$\begin{array}{l} 1 \parallel \text{qnorm}(0.025) \\ 2 \parallel [1] \quad -1.959964 \end{array}$
The following rejection region for the test statistic results	$z_{\frac{\alpha}{2}} = \Phi^{-1}(\alpha/2) = \Phi^{-1}(0.025) = -1.96$ $K = (-\infty, -1.96] \cup [1.96, \infty)$
6. Test decision:	Hence the value for the statistics is
	$z = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_{X_n}} = \frac{\sqrt{13}(80.02 - 80.00)}{0.01} = 7.211$
Remarks: Standardizing is in principle unnecessary because of technical aid of computer software.	Therefore the observed value falls into the rejection region.
3. Test statistic: (not standardized)	The mean value of the measurements
	$T: \bar{X}_n$

Null distribution (assuming H_0 is true):	$T \sim N(\mu_0, \frac{\sigma_X^2}{n}) = N(80, \frac{0.01^2}{13})$
5. Rejection region for the test statistic: (not standardized) Given $\alpha = 0.05$, R yields the following 2.5% quantile of the standard normal distribution.	$K = (-\infty, c_u] \cup [c_o, \infty)$ with $H_A : \mu \neq \mu_0$ <pre> 1 qnorm(0.025, 80.0, 0.01/sqrt(13)) 2 [1] 79.99456 3 qnorm(0.975, 80.0, 0.01/sqrt(13)) 4 [1] 80.00544 </pre>
In this way, we obtain the rejection region for the test statistic:	$K = (-\infty, 79.99] \cup [80.01, \infty)$

t-Test (σ_x unknown)	
1. Model:	X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma_X^2)$, σ_X is estimated by $\hat{\sigma}_X$
2. Null hypothesis:	$H_0: \mu = \mu_0$
Alternative:	$H_A: \mu \neq \mu_0$ (or $<$ or $>$)
3. Test statistic:	$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} = \frac{\text{observed} - \text{expected}}{\text{estimated standard error}}$
Null distribution (assuming H_0 is true):	$T \sim t_{n-1}$
4. Significance level:	α
5. Rejection region for the test statistic:	$K = (-\infty, t_{n-1; \frac{\alpha}{2}}] \cup [t_{n-1; 1-\frac{\alpha}{2}}, \infty)$ with $H_A : \mu \neq \mu_0$, $K = (-\infty, t_{n-1; \alpha}]$ with $H_A : \mu < \mu_0$, $K = [t_{n-1; 1-\alpha}, \infty)$ with $H_A : \mu > \mu_0$
6. Test decision:	Check whether the observed value of the test statistic falls into the rejection region.
Example	
1. Model:	X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma_X^2)$, σ_X is estimated, $\hat{\sigma}_X = 0.024$
2. Null hypothesis:	$H_0: \mu = \mu_0 = 80.00$
Alternative:	$H_A: \mu \neq \mu_0$
3. Test statistic:	$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X}$
Null distribution (assuming H_0 is true):	$T \sim t_{n-1}$
4. Significance level:	$\alpha = 0.05$
5. Rejection region for the test statistic: We determine the value by means of R , where $\alpha = 0.05$ and $n = 13$.	$K = (-\infty, t_{n-1; \frac{\alpha}{2}}] \cup [t_{n-1; 1-\frac{\alpha}{2}}, \infty)$ with $H_A : \mu \neq \mu_0$, $t_{n-1; 1-\frac{\alpha}{2}} = t_{12; 0.975} = 2.179$ <pre> 1 qt(0.975, 12) 2 [1] 2.178813 </pre>
The rejection region of the test statistic thus is given by	$K = (-\infty, -2.179] \cup [2.179, \infty)$
6. Test decision:	On the basis of $n = 13$ measurements, we find

	$\bar{x} = 80.02$ and $\hat{\sigma}_X = 0.024$ Hence, the realized value of the test statistic is $t = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} = \frac{\sqrt{13}(80.02 - 80.00)}{0.024} = 3.00$ The observed value falls into the rejection region. Therefore, the null hypothesis is rejected at 5% level.
<p>The t-test directly performed in R using the function <code>t.test()</code></p> <p>Remarks:</p> <p>(i) The observed value of the test statistic is 3.12. Assuming the null hypothesis is true, then the test statistic follows a t-distribution with $df = 12$ degrees of freedom.</p> <p>(ii) The observed mean value of the data is 80.02. A 95% confidence interval for the true mean is [80.006, 80.035].</p> <p>(iii) The R functions <code>qt(p,df)</code> calculates the quantile from the probability density and the degrees of freedom and <code>pt(q,df)</code> calculates the probability density from the quantile and the degrees of freedom.</p> <p>(iv) The confidence interval for measurement data consists of the values μ, for which the corresponding statistical test does not reject the null hypothesis.</p>	<pre> 1 x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 2 80.03, 80.04, 79.97, 80.05, 80.03, 3 80.02, 80.00, 80.02) 4 5 t.test(x, alternative = "two.sided", 6 mu = 80.00, conf.level = 0.95) 7 8 ## 9 ## One Sample t-test 10 ## 11 ## data: x 12 ## t = 3.1246, df = 12, p-value = 0.008779 13 ## alternative hypothesis: true mean is not 14 ## equal to 80 15 ## 95 percent confidence interval: 16 ## 80.00629 80.03525 17 ## sample estimates: 18 ## mean of x 19 ## 80.02077 20 21 qt(0.975,12) 22 [1] 2.178813 23 pt(2.178813,12) 24 [1] 0.975 25 qt(0.5,12) 26 [1] 0.0 27 pt(0.0,12) 28 [1] 0.5 </pre>
P-Value	
<p>The p-value is the probability that the test statistic will take on a value that is at least as extreme (with respect to the alternative hypothesis) as the observed value of the statistic when the null hypothesis H_0 is true.</p> <p>In R we compute the one-sided and the two sided p-value as follows:</p> <p>These p-values are evidence against the null hypothesis at 5% level. Whereas the two-sided value is statistically significant at the 5% value.</p>	<p>For the one-sided alternative hypothesis $H_A: \mu > \mu_0$, the p-value can be calculated as follows - the observed value of the statistics is $t = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} = 3.1246$:</p> <p>$p\text{-value} = P(T > t) = P(T > 3.1246) = 0.00439$ For the two-sided alternative hypothesis $H_A: \mu \neq \mu_0$, the p-value can be calculated as follows (the observed value of the test statistics is $t = \frac{\sqrt{n} \bar{X}_n - \mu_0 }{\hat{\sigma}_X}$):</p> <p>$p\text{-value} = 2 \cdot P(T > t)$</p> <pre> 1 #one-sided p-value 2 1-pt(3.1246, df=12) 3 [1] 0.004389739 4 #two-sided p-value 5 2*(1-pt(3.1246, df=12)) 6 [1] 0.008779477 </pre>

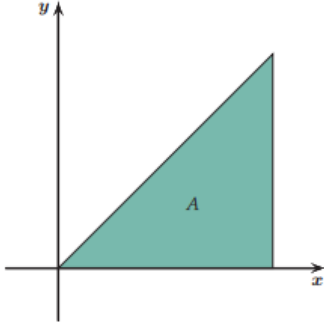
<p>p-value and Statistical Test</p> <ol style="list-style-type: none"> 1. Reject H_0 if $p\text{-value} \leq \alpha$ 2. Retain H_0 if $p\text{-value} > \alpha$ <p>The <i>p-value</i> is the <i>smallest level of significance</i> that would lead to rejection of the null hypothesis H_0 with the given data.</p>	<p>The <i>p-value</i> quantifies how significant an alternative is:</p> <p>$p\text{-value} \approx 0.05$: weakly significant, "."</p> <p>$p\text{-value} \approx 0.01$: weakly significant, "**"</p> <p>$p\text{-value} \approx 0.001$: weakly significant, "***"</p> <p>$p\text{-value} \leq 10^{-4}$: weakly significant, "****"</p>
--	--

2.3 Joint Distributions

2.3.1 Joint, Marginal and Conditional Distributions

Discrete Joint Probability Distribution																																									
The Joint Probability Distribution of X and Y is defined by the following distributions:				$P(X = x, Y = y), x \in W_x, y \in W_y$																																					
Marginal Distributions are single distributions $P(X = x)$ of X and $P(Y = y)$ of Y . They can be calculated based on their joint distribution: Joint distribution of (X, Y) starting from the marginal distribution of X and Y is only possible for independent X and Y . Then it holds:				$P(X = x) = \sum_{y \in W_y} P(X = x, Y = y), x \in W_x$ $P(X = x, Y = y) = P(X = x) \cdot P(Y = y), x \in W_x, y \in W_y$																																					
Conditional probability of X given $Y = y$ is defined as: The marginal distributions then can be expressed as follows:				$P(X = x Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}$ $P(X = x) = \sum_{y \in W_y} P(X = x Y = y)P(Y = y), x \in W_x$																																					
Conditional Expected Value of Y given $X = x$ is defined as:				$E[Y X = x] = \sum_{y \in W_y} y \cdot P(Y = y X = x)$																																					
Example																																									
<table><tr><th>X\ Y</th><th>1</th><th>2</th><th>3</th><th>4</th><th>Σ</th></tr><tr><th>1</th><td>0.080</td><td>0.015</td><td>0.003</td><td>0.002</td><td>0.100</td></tr><tr><th>2</th><td>0.050</td><td>0.350</td><td>0.050</td><td>0.050</td><td>0.500</td></tr><tr><th>3</th><td>0.030</td><td>0.060</td><td>0.180</td><td>0.030</td><td>0.300</td></tr><tr><th>4</th><td>0.001</td><td>0.002</td><td>0.007</td><td>0.090</td><td>0.100</td></tr><tr><th>Σ</th><td>0.161</td><td>0.427</td><td>0.240</td><td>0.172</td><td>1</td></tr></table>				X\ Y	1	2	3	4	Σ	1	0.080	0.015	0.003	0.002	0.100	2	0.050	0.350	0.050	0.050	0.500	3	0.030	0.060	0.180	0.030	0.300	4	0.001	0.002	0.007	0.090	0.100	Σ	0.161	0.427	0.240	0.172	1	$P(X = 3, Y = 4) = 0.030 \text{ or } P(X = 3 \cup Y = 4) = 0.030$ $P(X = 3) = P(X = 3, Y = 1) + P(X = 3, Y = 2) + P(X = 3, Y = 3) + P(X = 3, Y = 4) = 0.030 + 0.060 + 0.180 + 0.030 = 0.300$ $P(Y = 2 X = 4) = \frac{P(Y=2, X=4)}{P(X=4)} = \frac{0.002}{0.1} = 0.02$ $P(X = Y) = P(X = 1, Y = 1) + P(X = 2, Y = 2) + P(X = 3, Y = 3) + P(X = 4, Y = 4) = 0.700$ If random variables are independent it must hold that $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ From the marginal distribution follows $P(X = 1) \cdot P(Y = 2) = 0.100 \cdot 0.427 = 0.043$ and this is not equal to $P(X = 1, Y = 2) = 0.15$ X and Y are not independent	
X\ Y	1	2	3	4	Σ																																				
1	0.080	0.015	0.003	0.002	0.100																																				
2	0.050	0.350	0.050	0.050	0.500																																				
3	0.030	0.060	0.180	0.030	0.300																																				
4	0.001	0.002	0.007	0.090	0.100																																				
Σ	0.161	0.427	0.240	0.172	1																																				

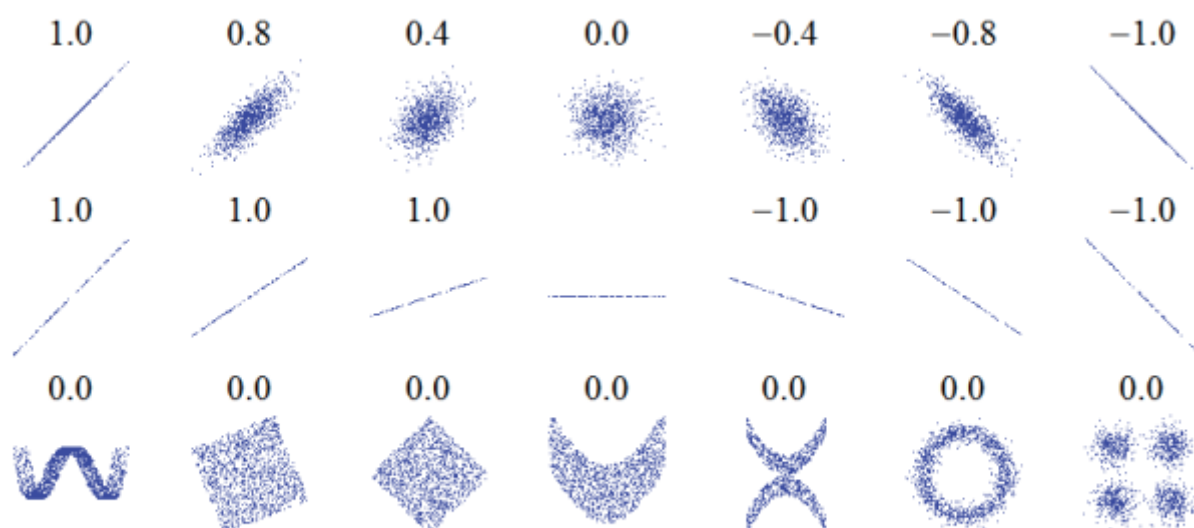
Joint Density Function	
The probability that the joint random variable (X, Y) lies in a two-dimensional region A , i.e., $A \subset \mathbb{R}^2$, is given by	$P((X, Y) \in A) = \iint_A f_{X,Y}(x, y) dx dy$

The (bivariate) joint density function needs to satisfy	$\iint_{\mathbb{R}} f_{X,Y}(x,y) dx dy = 1$
X and Y are only independent if	$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y), x, y \in \mathbb{R}$
Marginal Density	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy, f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$
Conditional Probability	$f_{Y X=x}(y) = f_Y(y X=x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
X and Y are only independent if the following apply:	$f_{Y X=x}(y) = f_Y(y)$ resp. $f_{X Y=y}(x) = f_X(x)$
Conditional Expected Value of a continuous random variable Y given $X = x$	$E[Y X=x] = \int_{-\infty}^{\infty} y \cdot f_{Y X=x}(y) dy$
Example	
<p>Two machines with exponentially distributed life expectancy $X \sim \text{Exp}(\lambda_1)$ and $Y \sim \text{Exp}(\lambda_2)$, where X and Y are independent. $f_X(x) = \lambda_1 e^{-\lambda_1 x}$ and $f_Y(y) = \lambda_2 e^{-\lambda_2 y}$</p> 	<p>Due to independence: $f_{X,Y}(x,y) = \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 y}$</p> <p>$P(Y < X) = \int_0^{\infty} \left(\int_0^x \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 y} dy \right) dx$</p> <p>$P(Y < X) = \int_0^{\infty} \lambda_1 e^{-\lambda_1 x} (1 - e^{-\lambda_2 x}) dx = \frac{\lambda_2}{\lambda_1 + \lambda_2}$</p>

2.3.2 Covariance and Correlation

Covariance and Correlation	
Covariance	$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$
X, Y independent	$E[XY] = E[X]E[Y]$
	$Cov(X,X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2] = Var(X)$
Sum of Variances	$Var\left(\sum_{i=1}^n X_i\right) = Cov\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j)$
2 Random Variables	$Var(X+Y) = Cov(X+Y, X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
If all X_i are independent	$Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$
Correlation	$Cor(X,Y) = \rho_{XY} = \frac{Cov(X,Y)}{\rho_X \rho_Y}$ where $-1 \leq Cor(X,Y) \leq 1$
Measure for strength and direction of the <i>linear dependency</i> between X and Y .	$Cor(X,Y) = +1$ if $Y = a + bX$ for $a \in \mathbb{R}$ and $b > 0$ $Cor(X,Y) = -1$ if $Y = a + bX$ for $a \in \mathbb{R}$ and $b < 0$

X and Y linear independent	$ Cor(X, Y) = 1$ means perfect linear relationship between X and Y . $Cor(X, Y) = 0$ means X and Y are uncorrelated. $Cor(X, Y) = 0$ (and thus $Cov(X, Y) = 0$)
---------------------------------------	---



If $Cor(X, Y) = 0$, then X and Y may still exhibit (non-linear) dependency.

Figure 5: Correlations

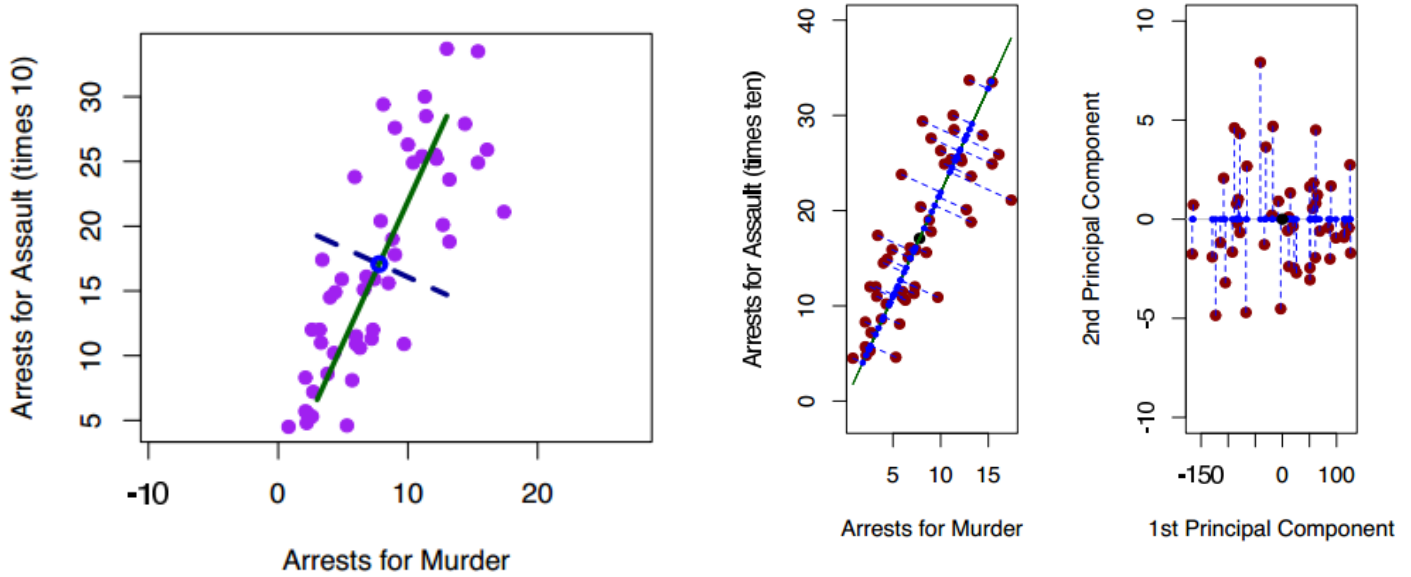
2.3.3 Bivariate Normal Distribution

Bivariate Normal Distribution	
Expected values and variances of the marginal distribution	μ_X, σ_X^2 and μ_Y, σ_Y^2
Covariance between X and Y	$Cov(X, Y) = \rho_{XY} \sigma_X \sigma_Y$
Joint Density	$f_{X,Y}(x, y) =$ $\frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu_X, y - \mu_Y) \Sigma^{-1} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix}\right)$
Covariance Matrix	$\Sigma = \begin{pmatrix} Cov(X, X) & Cov(X, Y) \\ Cov(Y, X) & Cov(Y, Y) \end{pmatrix} =$ $\begin{pmatrix} \sigma_X^2 & \rho_{XY} \sigma_X \sigma_Y \\ \rho_{XY} \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix}$

2.3.4 Principal Component Analysis (PCA)

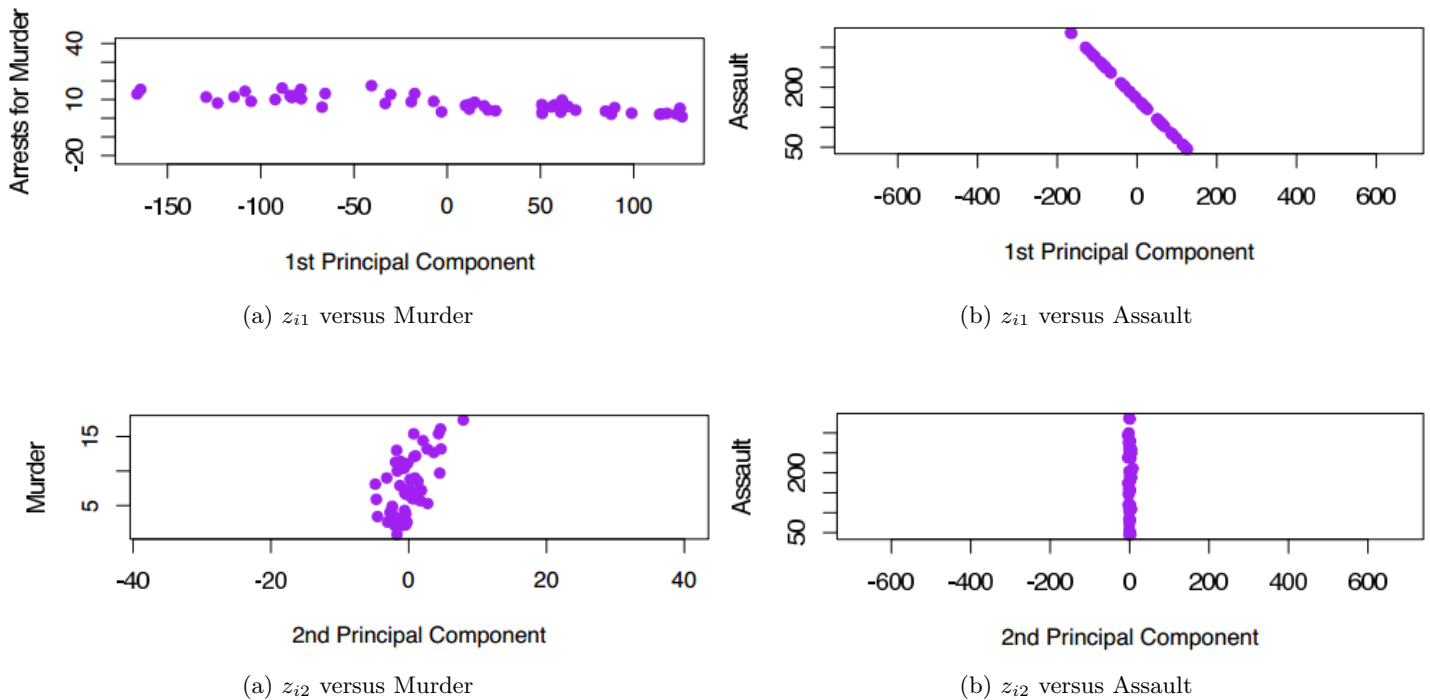
PCA is a popular approach for deriving a low-dimensional set of features from a large set of variables. PCA is a technique for reducing the dimension of a $n \times p$ data matrix X where n corresponds to the number of observations and p to the number of variables.

2.3.4.1 Example: USArrests

Assault versus Murder Rate, US, 1973

(a) The data vary the most along the first principal component (b) Counter Clockwise Rotation that 1. PC coincides with x-axis

Figure 6: 1st and 2nd Principal Component



(a) z_{i1} versus Murder

(b) z_{i1} versus Assault

(a) z_{i2} versus Murder

(b) z_{i2} versus Assault

Figure 8: The fact that the 2nd principal component scores are much closer to zero indicates that this component captures far less information as the 1st principal component.

Theory	Code Example
$Z_1 = -0.0419126(\text{Murder} - \overline{\text{Murder}}) - 0.9991213(\text{Assault} - \overline{\text{Assault}})$ $\phi_{11} = -0.0419126$ and $\phi_{21} = -0.9991213$ are the <i>principal component loadings</i> The idea is that every that out of every linear combination of Murder and Assault sucht that $\phi_{11}^2 + \phi_{21}^2 = 1$ and $\text{Var}(\phi_{11})(\text{Murder} - \overline{\text{Murder}}) + \phi_{21}(\text{Assault} - \overline{\text{Assault}})$ is maximized. <code>prcomp()</code> centers the variables to have mean zero. This corresponds to how the first principal component is defined. $z_{i1} = -0.0419126(\text{Murder} - \overline{\text{Murder}}) - 0.9991213(\text{Assault} - \overline{\text{Assault}})$ The values of z_{i1}, \dots, z_{n1} are known as <i>principal component scores</i> , seen in the right-hand panel in Figure 6. $z_{i1} > 0$ indicates a state with below-average arrests for murder and below average for assault. A negative score suggests the opposite. $Z_2 = 0.9991213(\text{Murder} - \overline{\text{Murder}}) - 0.0419126(\text{Assault} - \overline{\text{Assault}})$	<pre> 1 #First principal component 2 pr.out <- prcomp(USArrests[,c("Murder", "Assault")]) 3 pr.out\$rotation[,1] 4 5 ## Murder Assault 6 ## -0.0419126 -0.9991213 7 8 #principal component scores (z_i1 to z_in) 9 pr.out <- prcomp(USArrests[,c("Murder", "Assault")]) 10 head(pr.out\$x) 11 12 ## PC1(z_i1) PC2(z_i2) 13 ## Alabama -65.40950 2.6728663 14 ## Alaska -92.25166 -1.6559620 15 ## Arizona -123.14478 -4.8535831 16 ## Arkansas -19.26551 0.2047123 17 ## California -105.19832 -3.1999471 18 ## Colorado -33.21549 -1.2812733 19 20 #Second principal component 21 pr.out <- prcomp(USArrests[,c("Murder", "Assault")]) 22 pr.out\$rotation[,2] 23 24 ## Murder Assault 25 ## 0.9991213 -0.0419126 </pre>

With two-dimensional data, such as in our USArrests example, we can construct at most two principal components. However, if we had other variables, such as Rape, then additional components could be constructed.

2.3.4.2 PCA and Covariance Matrix

The covariance matrix of two random variables X and Y is defined as

$$\Sigma = \begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \sigma_Y^2 \end{pmatrix}$$

Since Z_1 and Z_2 are required to be uncorrelated, this implies for their covariance matrix Σ to have vanishing off diagonal elements. Therefore the covariance has to be diagonalized. This can be done with by a rotation matrix Φ so that

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = (X - \mu_X, Y - \mu_Y) \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} = \begin{pmatrix} \phi_{11}(X - \mu_X) & \phi_{12}(Y - \mu_Y) \\ \phi_{21}(X - \mu_X) & \phi_{22}(Y - \mu_Y) \end{pmatrix}$$

and

$$\begin{pmatrix} \text{Cov}(Z_1, Z_1) & \text{Cov}(Z_1, Z_2) \\ \text{Cov}(Z_2, Z_1) & \text{Cov}(Z_2, Z_2) \end{pmatrix} = \begin{pmatrix} \sigma_{Z_1}^2 & 0 \\ 0 & \sigma_{Z_2}^2 \end{pmatrix}$$

The rotation matrix Φ needs to satisfy the condition $\phi_{11}^2 + \phi_{21}^2 = 1$ and $\phi_{12}^2 + \phi_{22}^2 = 1$ It is straightforward to generalize the case of $p = 2$ to an arbitrary p .

2.3.4.3 Proportion of Variance Explained by Principal Components

Theory	Code Example
<p>There is an information loss of the given data by projecting the observations onto the first few principal components. Therefore we want to know the <i>proportion of variance explained (PVE)</i>. The <i>total variance</i> is defined as</p> $\sum_{j=1}^p \text{Var}(X_j) = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n x_{ij}^2$ <p>and the variance of the mth principal component is</p> $\frac{1}{n} \sum_{i=1}^n z_{im}^2 = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p \phi_{jm} x_{ij} \right)^2$ <p>Therefore the PVE by the mth principal component is given by</p> $\frac{\sum_{i=1}^n \left(\sum_{j=1}^p \phi_{jm} x_{ij} \right)^2}{\sum_{j=1}^p \sum_{i=1}^n x_{ij}^2}$	<pre> 1 pr.out <- prcomp(USArrests[,c("Murder", "Assault")], scale=FALSE) 2 pr.var <- pr.out\$sdev^2 3 pve <- pr.var/sum(pr.var) 4 pve 5 6 7 ## [1] 0.9990292327 0.0009707673 8 9 ## Most of the information of the data about the arrests for Murder and Assault is contained in the first principal component. </pre>

3 Regression Analysis

3.1 Simple Linear Regression

TO DO: Chapter 5

...

3.1.1 Estimating the Coefficients

Theory	Code Example
<p>Estimation of response variable Y based on a predictor variable X.</p> $Y \simeq \beta_0 + \beta_1 X$	<pre> 1 lm(Y ~ X, data=someData) </pre>

Source code:	Output:
<pre> 1 advertising <- read.csv("../Data/Advertising.csv") 2 model <- lm(sales ~ TV, data=advertising) 3 summary(model) </pre>	<pre> 1 ## 2 ## Call: 3 ## lm(formula = sales ~ TV, data = Advertising) 4 ## 5 ## Residuals: 6 ## Min 1Q Median 3Q Max 7 ## -8.3860 -1.9545 -0.1913 2.0671 7.2124 8 ## 9 ## Coefficients: 10 ## Estimate Std. Error t value Pr(> t) 11 ## (Intercept) 7.032594 0.457843 15.36 <2e-16 ** 12 ## TV 0.047537 0.002691 17.67 <2e-16 *** 13 ## --- 14 ## Signif. codes: 15 ## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 16 ## 17 ## Residual standard error: 3.259 on 198 degrees 18 ## of freedom 19 ## Multiple R-squared: 0.6119, Adjusted R-squared 20 ## : 0.6099 21 ## F-statistic: 312.1 on 1 and 198 DF, p-value: 22 ## < 2.2e-16 </pre>
Interpretation of output:	
TO DO: interpretation here	

3.2 Residual Analysis

TO DO: Chapter 6

3.3 Multiple Linear Regression

TO DO: Chapter 7

3.4 Linear Model Selection

TO DO: Chapter 8

4 Classification

4.1 Logistic Regression

TO DO: Chapter 10

4.2 Decision Trees

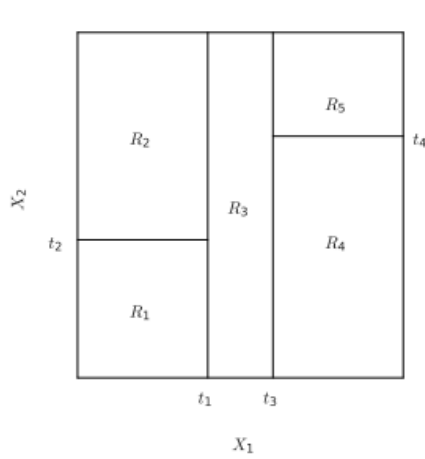
Decision trees are applied to both, classification and regression. TO DO: Chapter 11

4.2.1 Classification Trees

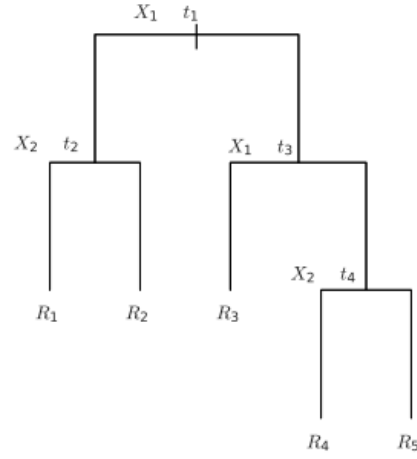
4.2.1.1 Binary Splitting

In binary splitting, a training set is used to split up the predictor domain into regions which contain data for which the response variable belongs to the same class. By **binary** it is meant that a region is split into **two** subregions (i.e. “is a predictor less or greater than a threshold value?” → yes/no).

Theory	Code Example
<p>Algorithm:</p> <ol style="list-style-type: none"> 1. Initialise the set of regions $\mathcal{R} = R$ by the predictor domain R 2. Choose the optimal region R in \mathcal{R} and the optimal predictor X_i such that a binary split of R with respect to X $R_1 = \{\vec{x} \in R x_i > t\} \quad \text{and} \quad R_2 = \{\vec{x} \in R x_i \leq t\}$ gives the highest gain in purity (for some threshold t). 3. Replace R in \mathcal{R} with R_1 and R_2 and return to 2. <p>The iteration is stopped if the current splitting fulfils a pre-defined stopping criterion.</p>	<pre> 1 require(tree) 2 #default controls 3 tc = tree.control(nobs = 303, mincut = 5, 4 minsize = 10, mindev = 0.01) 5 6 #grow tree 7 tree.model = tree(AHD~MaxHR+Age, data = heart, 8 control = tc) 9 10 #plot tree and label splits 11 plot(tree.model) 12 text(tree.model, cex=0.8) 13 14 #plot partition (only for two predictor case) 15 partition.tree(tree.model) 16 points(Age~MaxHR, data = heart, col = cols[label 17], pch=20) </pre>



(a) Example regions resulting from binary splitting



(b) Example decision tree resulting from binary splitting

4.2.1.2 Node Purity

Notation:

Variable	Description
Y	Response variable
K	Levels (categories) of the response variable
T	The decision tree
M	Amount of terminal nodes
\hat{p}_{mk}	proportion of the training data in region m from level k

Purity Measures:

Classification error rate	$E_m(T) = 1 - \max_k (\hat{p}_{mk})$
---------------------------	--------------------------------------

Gini index	$G_m(T) = \sum_{k=1}^K \hat{p}_{mk} \cdot (1 - \hat{p}_{mk})$
Cross-entropy	$D_m(T) = - \sum_{k=1}^K \hat{p}_{mk} \cdot \log(\hat{p}_{mk})$

Code example: Cross Entropy and Gini measures in R

```

1 require(tree)
2 # deviance or cross entropy
3 tree.model = tree(AHD~MaxHR+Age, data = heart, split = "deviance")
4 plot(tree.model)
5 text(tree.model, cex=0.8)
6 partition.tree(tree.model)
7 points(Age~MaxHR, data = heart, col = cols[label], pch=20)
8
9 # Gini index
10 tc = tree.control(303, mincut = 5, minsize = 60, mindev = 0.01)
11 tree.model = tree(AHD~MaxHR+Age, data = heart, split = "gini", control = tc)
12 plot(tree.model)
13 text(tree.model, cex=0.8)
14 partition.tree(tree.model)
15 points(Age~MaxHR, data = heart, col = cols[label], pch=20)

```

4.3 Random Forests

TO DO: Chapter 12

5 Time Series Analysis

5.1 Introduction to Time Series

Models are not always *independent of the order* of the training data. Many real life measuring and data recording processes result in data sets that are *serially correlated*. For example *machine monitoring, stock, environmental observations or federal statistics*. These kind of data is called *time series data*. Usually there are several goals that one wants to achieve in time series data.

- Descriptive Analysis
- Modelling and Interpretation
- Decomposition
- Prediction
- Regression

TO DO: Chapter 13

5.1.1 Time Series with R

Theory	Code Example
<p>All data in R are stored in objects, which provide a range of methods. The class of an object can be found using the <code>class</code> function. For example, we have already encountered the <code>data.frame</code> class. It has a series of methods, such as <code>names</code> or <code>nrow</code>:</p> <p>(The data set <code>iris</code> contains 50 samples of three types of Iris flowers, measured along four variables.)</p>	<pre> 1 class(iris); names(iris); nrow(iris) 2 3 ## [1] "data.frame" 4 ## [1] "Sepal.Length" "Sepal.Width" "Petal. Length" "Petal.Width" 5 ## [5] "Species" 6 ## [1] 150 </pre>

5.1.1.1 The `ts` Class

Theory	Code Example
<p>Basic properties: The <code>AirPassengers</code>-data is a built in set of class <code>ts</code>. Most important methods for <code>ts</code> class are:</p> <ol style="list-style-type: none"> 1. <code>start()</code> returns the start time of the series. 2. <code>end()</code> returns the end time of the series. 3. <code>frequency()</code> returns the number of samples per unit time. 4. <code>plot()</code> displays the time series as a function over the time axis. <code>plot</code> function calls <code>plot.ts</code> which is tailored for time series. <code>plot.ts</code> joins discrete time points automatically with lines. See Figure <code>AirPassengers</code>. 	<pre> 1 class(AirPassengers) 2 ## [1] "ts" 3 4 start(AirPassengers); end(AirPassengers); frequency(AirPassengers) 5 ## [1] 1949 1 6 ## [1] 1960 12 7 ## [1] 12 8 9 #1/frequency = 1/12 = 0.0833 10 deltat(AirPassengers) 11 ## [1] 0.0833 12 13 #output in figure AirPassengers. 14 plot(AirPassengers, main = "Passengers", ylab=" Number (in 1000s)") 15 grid() </pre>

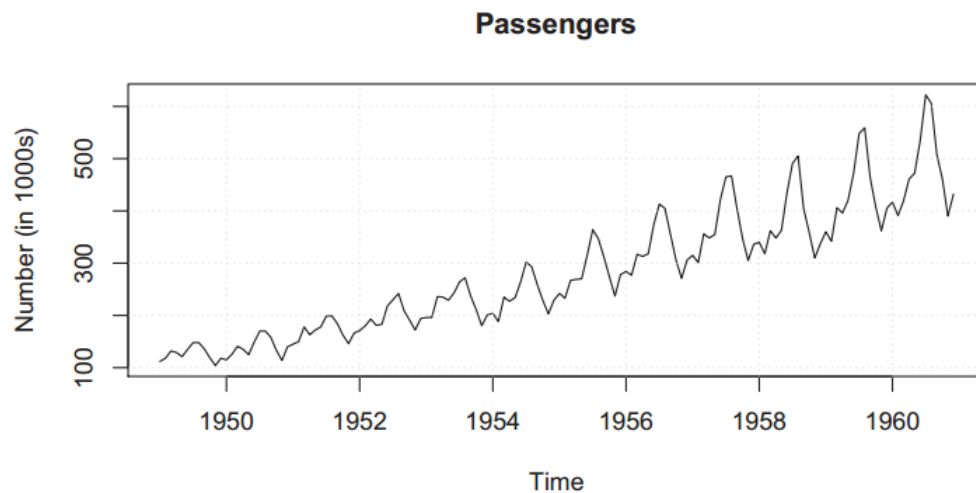


Figure 10: `AirPassengers`

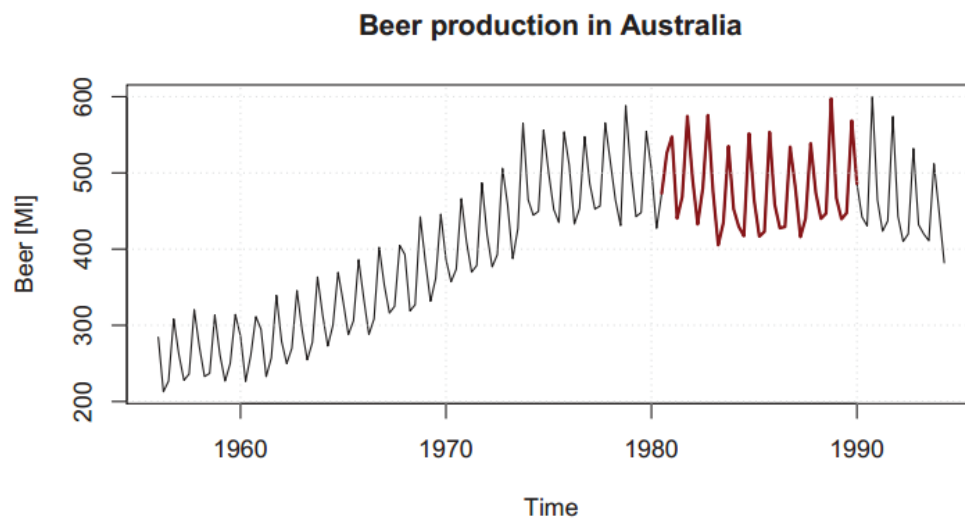
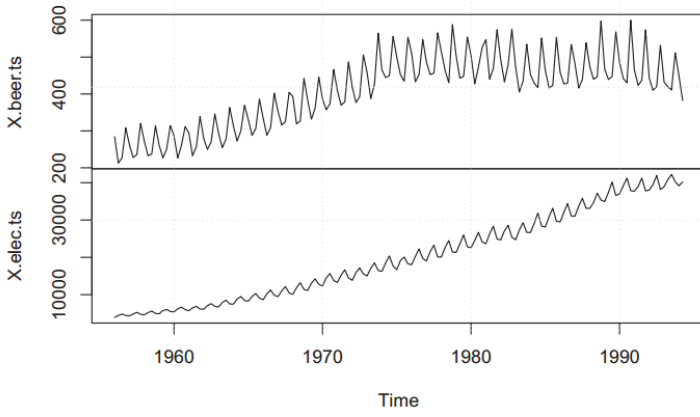


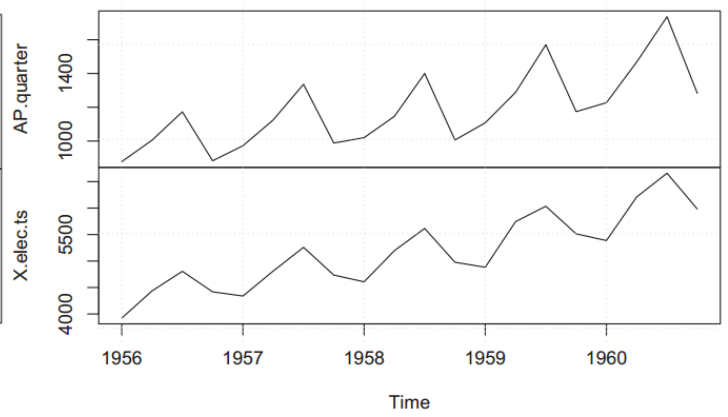
Figure 11: Subset of a Time Series (seasonal behaviour)

Theory	Code Example
<p>Defining a <code>ts</code> class</p> <p>If data is not in a time series form we can make a <code>ts</code> object by using the <code>ts</code> function. This is not necessary for AirPassengers, therefore the example AustralianBeer is used.</p> <ol style="list-style-type: none"> <code>summary()</code> gives the five-number summary as well as the mean of the time series. This function shows the minimum, the first quartile, the median, the second quartile and the maximum of the time series. This is called the <i>five-number-summary</i> of a data set. Additionally the mean is also computed. <code>window()</code> returns a subset of the time series defined by a start and an end time. 	<pre> 1 X.beer = read.table("../Daten/AustralianBeer.csv", sep=";", header = T) 2 3 X.beer.ts = ts(X.beer[,2], start = c(1956,1), end = c(1994, 2), frequency = 4) 4 summary(X.beer.ts) 5 ## Min. 1st Qu. Median Mean 3rd Qu. Max. 6 ## 213 325 427 408 467 600 7 8 #Figure Subset of Time Series 9 plot(X.beer.ts, ylab="Beer [Ml]", main="Beer production in Australia") 10 X.ts.w = window(X.beer.ts, start = c(1980,3), end = c(1990, 1)) 11 summary(X.ts.w) 12 ## Min. 1st Qu. Median Mean 3rd Qu. Max. 13 ## 405 437 467 478 530 598 14 lines(X.ts.w, col = "darkred", lwd=2) 15 grid() </pre>

Theory	Code Example
<p>Multivariate Time Series</p> <p>A few important ideas and concepts related to multivariate time series data illustrated with the following example:</p> <p>The quarterly supply of electricity in Australia compared to the quarterly beer production see Figure 12a.</p> <p>The plots show increasing trends in production for both goods, partly due to the rising population in Australia from about 10 million to about 18 million over the same period. But notice that electricity production has risen by a factor of 7 during which the population has not quite doubled.</p> <p>There are many functions in R for handling more than one series, including <code>ts.intersect</code> to obtain the intersection of two series that overlap in time. There are some pitfalls shown with AirPassenger and electricity in Figure 12b.</p> <p>The two series are correlated but there is of course no causal dependence of the two series. They are confounded by seasonal effects.</p> <p>Non-equidistant time series are not covered by the <code>ts</code> class. There are further packages:</p> <ul style="list-style-type: none"> zoo-package: It provides methods for regular and irregular spaced times series as well as arbitrary date formats. xts-package: It is an extension of the <code>zoo</code>-package which allows for further customization. 	<pre> 1 ##Example beer vs electricity (Australia) 2 3 ## First load the electricity data from file and create a time series out of it. 4 X.elec = read.table("../Daten/AustralianElectricity.csv", sep=";", header = T) 5 X.elec.ts = ts(X.elec[,2], start = c(1956,1), end = c(1994, 2), frequency 6 7 8 ##Bind the two separate series together by means of the cbind command and plot the series 9 X.ts = cbind(X.beer.ts, X.elec.ts) 10 plot(X.ts, main="Beer and electricity production in Australia") 11 grid() 12 13 ##aggregate the monthly data of the AirPassengers data to quarterly data 14 ##aggregate sums the data set to the desired frequency up 15 AP.quarter = aggregate(AirPassengers, nfrequency = 4) 16 17 18 #Extract common time points and combine the corresponding data values to a new, bivariate time series 19 AP.elec = ts.intersect(AP.quarter, X.elec.ts) 20 plot(AP.elec, main="Air Passenger bookings and electricity production") 21 grid() </pre>



(a) Beer and electricity production in Australia



(b) Air Passenger bookings and electricity production

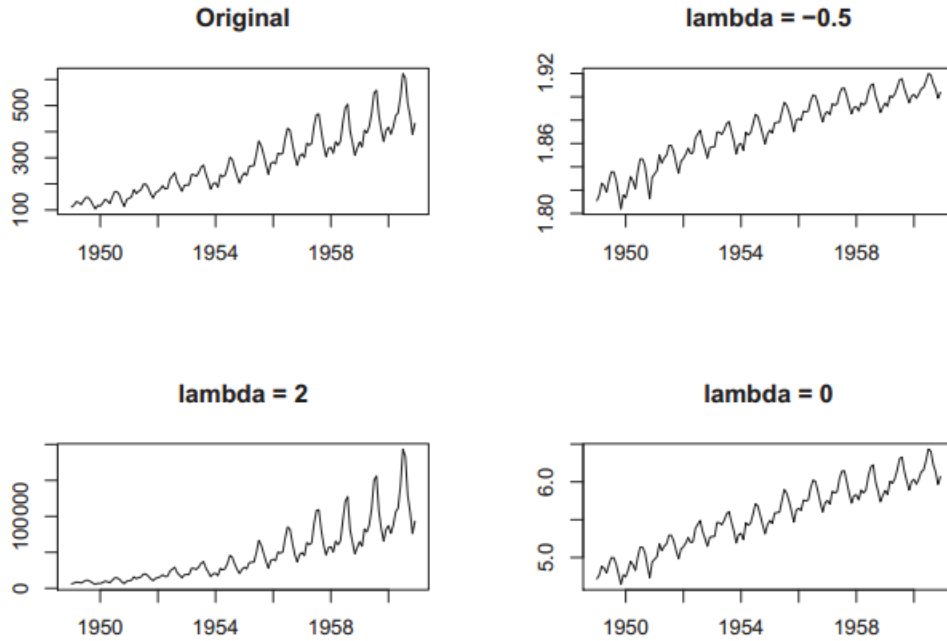
5.1.2 Basic transformation, visualization and decomposition of time series

5.1.2.1 Data transformation

In many situations it is desirable or necessary to transform a time series before the application of models and predictions. Many methods require

- **Gaussian** or **symmetric** distribution of the data.
- A **linear** trend relationship between time and data.
- A **constant variance** across time.

Theory	Code Example
<p>Box-Cox-transformation</p> <p>For highly skewed or heteroskedastic data - data whose variance is not constant across time - it is often better to use not the original series $\{x_1, x_2, \dots\}$ but a transformed series $\{g(x_1), g(x_2), \dots\}$. The Box-Cox-transformation is well suited for correcting skewness and variance.</p> <p>For a times series $\{x_1, x_2, \dots\}$ with positive values the Box-Cox transformations are defined as</p> $g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(x) & \lambda = 0 \end{cases}$ <p>As in Figure 13 to see the original data exhibits clear seasonal effects and an upward trend. The intensity of the seasonal influence, i.e. the variance over time, is also increasing. The parameter $\lambda = 0$, i.e. the log-transform of the data, yields a stabilized image: a seemingly linear trend with homogeneous seasonal effects.</p>	<pre> 1 # define the Box-Cox transformation 2 box.cox <- function(x, lambda) { 3 if (lambda==0) log(x) else (x^lambda - 1)/lambda 4 } 5 6 # plot the original and the transformed data 7 # --> see figure Box-Cox-transformation for 8 different values of lambda 9 layout(matrix(c(1,2,3,4), 2,2)) 10 plot(AirPassengers, main = "Original", ylab="", 11 xlab="") 12 plot(box.cox(AirPassengers, 2), main = "lambda = 13 2", 14 ylab="", xlab="") 15 plot(box.cox(AirPassengers, -0.5), main = " 16 lambda = -0.5", 17 ylab="", xlab="") 18 plot(box.cox(AirPassengers, 0), main = "lambda = 19 0", 20 ylab="", xlab="") </pre>

Figure 13: Box-Cox-transformation for different values of λ

Theory	Code Example
<p>Time-shift transformation</p> <p>Sometimes it is necessary to transform the time-axis as well. The most simple form version of time transforms is shifting.</p> <p>Let $\{x_1, x_2, \dots\}$ be a time series.</p> <ol style="list-style-type: none"> 1. The time-shift by a <i>lag</i> of $k \in \mathbb{Z}$ is defined by $g(x_i) = x_{i-k}$ 2. For the particular case where $k = 1$ the time-shift is called <i>backshift</i> $B(x_i) = x_{i-1}$ <p>In other words, applying a time-shift to a time series amounts to go back k steps (if $k > 0$) or go ahead $-k$ steps (if $k < 0$) in the series.</p> <p>In R the function lag is used to apply a time shift for various values of k.</p> <p>The back-shift operator is applied if <i>differences</i> of times series are computed, since $x_i - x_{i-1} = x_i - B(x_i)$. In particular, differencing is often combined with Box-Cox transformations. For example in the <i>log-returns</i> of a (financial) time series are defined as</p> $y_i = \log(x_i) - \log(x_{i-1}) = \log\left(\frac{x_i}{x_{i-1}}\right)$	<pre> 1 #lag function --> see figure time-shift transformations 2 AP = AirPassengers 3 AP.back = lag(AP, k = 4) 4 AP.ahead = lag(AP, k = -5) 5 plot(AP, lwd=2) 6 lines(AP.back, col="darkcyan") 7 lines(AP.ahead, col = "darkred") 8 grid() 9 legend("topleft", legend=c("original", "shifted back", "shifted ahead"), 10 lty=1, col = c("black", "darkcyan", "darkred")) </pre>

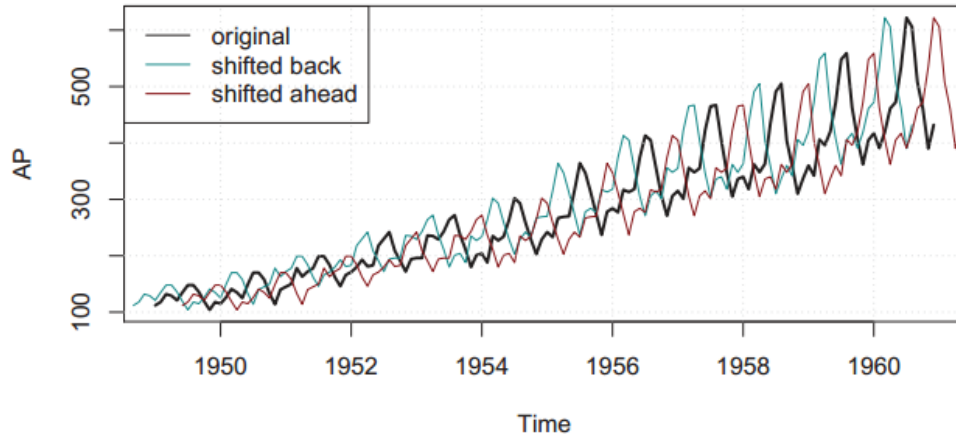
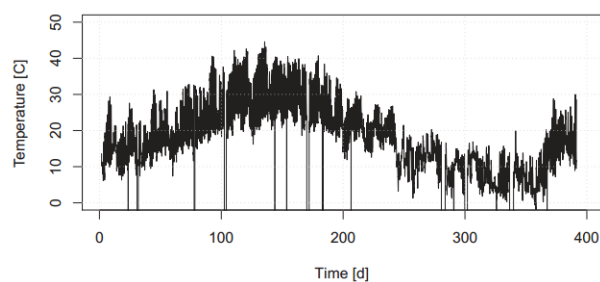


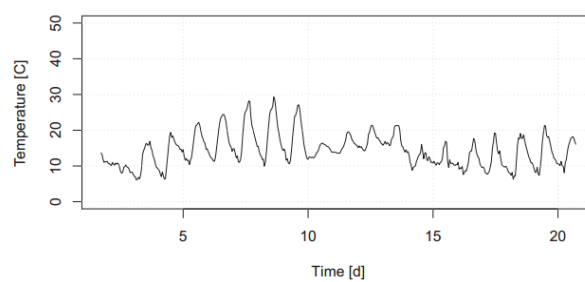
Figure 14: Time-shift transformation

5.1.2.2 Visualizations

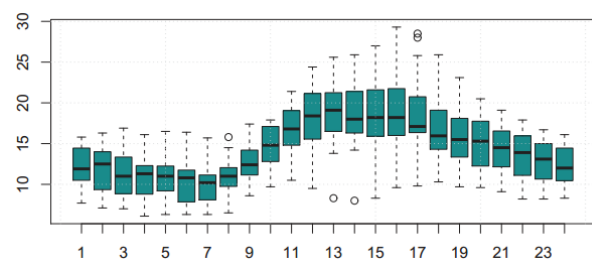
Theory	Code Example
<p>Visualization Example</p> <p>We have hourly measurements of several sensors, where we now focus on the air temperature. The time series is defined with a basic units of days starting at day 1 and frequency 24. The <code>end()</code>-command shows that the time series lasts 396 days and 8 hours, or in other words, has 9512 data points. The plot in Figure 15a shows the complete time series. The <code>ylim</code> option limits the temperature axis to non-negative values.</p> <p>We focus on a period of 20 days to analyse the temperature behaviour in more detail. Figure 15b shows the data.</p> <p>Figure 15c shows how data aggregation can be visualized with the <code>boxplot</code>-command. We are going to generate a boxplot for each full hour in one figure. To this end the <code>cycle()</code> function is very convenient: it returns for a given time series the positions in the cycle of each observation. In our example, a cycle is one day consisting of 24 hours. This means, that the first entry in the time series is at cycle position 18, i.e. measured at 6 p.m. and that the 875-th measurement is at cycle position 4, which corresponds to 4 a.m. The subset of observations that share a common cycle are called cycle-subseries and will be used later for time series decomposition.</p> <p>A useful graphical approach for visually inspecting correlations of consecutive observations are <i>lagged scatterplots</i>. They amount to produce scatterplots of the original time series values against a time-shifted version, i.e. plotting the data pairs (x_i, x_{i-k}). This can be done in R by the <code>lag.plot</code> command.</p>	<pre> 1 ##Figure Air Temperature measurement: 9512 points 2 AirData = read.table("../Daten/AirQualityUCI/ AirQualityUCI.csv", 3 sep=";", header=T, dec = ",") 4 AirTmp.ts = ts(AirData[,c(13)], start = c(1,18), frequency = 24) 5 end(AirTmp.ts) 6 ## [1] 396 8 7 8 plot(AirTmp.ts, main = "Air Temperature measurement: full data set", 9 ylab="Temperature [C]", xlab="Time [d]", ylim = c(0,50)) 10 grid() 11 12 ##Figure Air Temperature measurement: 480 points 13 AirTmpWin.ts = window(AirTmp.ts, start = c(1, 18), end=c(20, 18)) 14 plot(AirTmpWin.ts, main = "Air Temperature measurement: detail", 15 ylab="Temperature [C]", xlab="Time [d]", ylim = c(0,50)) 16 grid() 17 18 ##Figure Air temperature: Boxplot 19 cycle(AirTmp.ts)[1]; cycle(AirTmp.ts)[875] 20 ## [1] 18 21 ## [1] 4 22 23 boxplot(AirTmpWin.ts ~ cycle(AirTmpWin.ts), 24 col = "darkcyan", main = "Air temperature") 25 grid() 26 27 ##Figure lag.plot 28 lag.plot(AirTmpWin.ts, pch=20, main = "") 29 lag.plot(AirTmpWin.ts, pch=20, main = "", set. lags = 10) </pre>



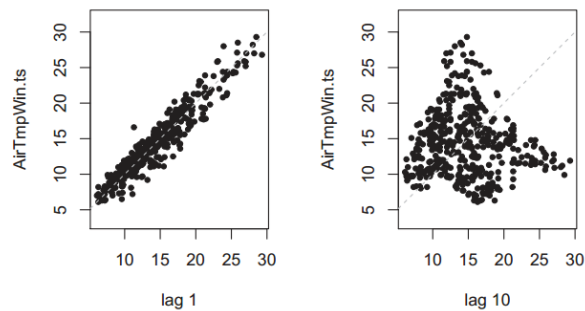
(a) Air temperature measurement: 9512 points



(b) Air temperature measurement: 480 points



(c) Air temperature: Boxplot



(d) lag.plot

Figure 15: Visualization

5.2 Mathematical Models for Time Series

TO DO: Chapter 14

5.3 Forecasting ime Series

TO DO: Chapter 15

6 Idiotenseite

6.1 Dreiecksformeln

Cosinussatz

$$c^2 = a^2 + b^2 - 2 \cdot a \cdot b \cdot \cos \gamma$$

Sinussatz

$$\frac{a}{\sin \alpha} = \frac{b}{\sin \beta} = \frac{c}{\sin \gamma} = 2r = \frac{u}{\pi}$$

Pythagoras beim Sinus

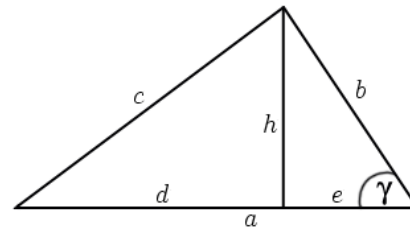
$$\sin^2(b) + \cos^2(b) = 1 \quad \tan(b) = \frac{\sin(b)}{\cos(b)}$$

$$\sin \beta = \frac{b}{a} = \frac{\text{Gegenkathete}}{\text{Hypotenuse}}$$

$$\cos \beta = \frac{c}{a} = \frac{\text{Ankathete}}{\text{Hypotenuse}}$$

$$\tan \beta = \frac{c}{b} = \frac{\text{Gegenkathete}}{\text{Ankathete}}$$

$$\cot \beta = \frac{c}{b} = \frac{\text{Ankathete}}{\text{Gegenkathete}}$$



6.2 Funktionswerte für Winkelargumente

deg	rad	sin	cos	tan	deg	rad	sin	cos	deg	rad	sin	cos	deg	rad	sin	cos
0°	0	0	1	0	90°	$\frac{\pi}{2}$	1	0	180°	π	0	-1	270°	$\frac{3\pi}{2}$	-1	0
30°	$\frac{\pi}{6}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{3}$	120°	$\frac{2\pi}{3}$	$\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	210°	$\frac{7\pi}{6}$	$-\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	300°	$\frac{5\pi}{3}$	$-\frac{\sqrt{3}}{2}$	$\frac{1}{2}$
45°	$\frac{\pi}{4}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	1	135°	$\frac{3\pi}{4}$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$	225°	$\frac{5\pi}{4}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$	315°	$\frac{7\pi}{4}$	$-\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$
60°	$\frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$	150°	$\frac{5\pi}{6}$	$\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	240°	$\frac{4\pi}{3}$	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	330°	$\frac{11\pi}{6}$	$-\frac{1}{2}$	$\frac{\sqrt{3}}{2}$

6.3 Periodizität

$$\cos(a + k \cdot 2\pi) = \cos(a) \quad \sin(a + k \cdot 2\pi) = \sin(a) \quad (k \in \mathbb{Z})$$

6.4 Quadrantenbeziehungen

$$\begin{aligned} \sin(-a) &= -\sin(a) & \cos(-a) &= \cos(a) \\ \sin(\pi - a) &= \sin(a) & \cos(\pi - a) &= -\cos(a) \\ \sin(\pi + a) &= -\sin(a) & \cos(\pi + a) &= -\cos(a) \\ \sin\left(\frac{\pi}{2} - a\right) &= \sin\left(\frac{\pi}{2} + a\right) = \cos(a) & \cos\left(\frac{\pi}{2} - a\right) &= -\cos\left(\frac{\pi}{2} + a\right) = \sin(a) \end{aligned}$$

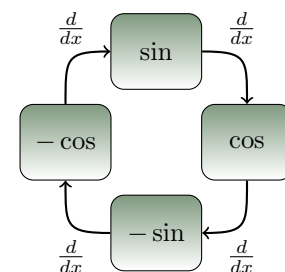
6.6 Additionstheoreme

$$\begin{aligned} \sin(a \pm b) &= \sin(a) \cdot \cos(b) \pm \cos(a) \cdot \sin(b) \\ \cos(a \pm b) &= \cos(a) \cdot \cos(b) \mp \sin(a) \cdot \sin(b) \\ \tan(a \pm b) &= \frac{\tan(a) \pm \tan(b)}{1 \mp \tan(a) \cdot \tan(b)} \end{aligned}$$

6.8 Produkte

$$\begin{aligned} \sin(a) \sin(b) &= \frac{1}{2}(\cos(a - b) - \cos(a + b)) \\ \cos(a) \cos(b) &= \frac{1}{2}(\cos(a - b) + \cos(a + b)) \\ \sin(a) \cos(b) &= \frac{1}{2}(\sin(a - b) + \sin(a + b)) \end{aligned}$$

6.5 Ableitungen



6.7 Doppel- und Halbwinkel

$$\begin{aligned} \sin(2a) &= 2 \sin(a) \cos(a) \\ \cos(2a) &= \cos^2(a) - \sin^2(a) = 2 \cos^2(a) - 1 = 1 - 2 \sin^2(a) \\ \cos^2\left(\frac{a}{2}\right) &= \frac{1 + \cos(a)}{2} \quad \sin^2\left(\frac{a}{2}\right) = \frac{1 - \cos(a)}{2} \end{aligned}$$

6.9 Euler-Formeln

$$\begin{aligned} \sin(x) &= \frac{1}{2j}(e^{jx} - e^{-jx}) & \cos(x) &= \frac{1}{2}(e^{jx} + e^{-jx}) \\ e^{x+jy} &= e^x \cdot e^{jy} = e^x \cdot (\cos(y) + j \sin(y)) \\ e^{j\pi} &= e^{-j\pi} = -1 \end{aligned}$$

6.10 Summe und Differenz

$$\begin{aligned}\sin(a) + \sin(b) &= 2 \cdot \sin\left(\frac{a+b}{2}\right) \cdot \cos\left(\frac{a-b}{2}\right) \\ \sin(a) - \sin(b) &= 2 \cdot \sin\left(\frac{a-b}{2}\right) \cdot \cos\left(\frac{a+b}{2}\right)\end{aligned}$$

$$\begin{aligned}\cos(a) + \cos(b) &= 2 \cdot \cos\left(\frac{a+b}{2}\right) \cdot \cos\left(\frac{a-b}{2}\right) \\ \cos(a) - \cos(b) &= -2 \cdot \sin\left(\frac{a+b}{2}\right) \cdot \sin\left(\frac{a-b}{2}\right) \\ \tan(a) \pm \tan(b) &= \frac{\sin(a \pm b)}{\cos(a) \cos(b)}\end{aligned}$$

6.12 Ableitungen elementarer Funktionen S436

Funktion	Ableitung	Funktion	Ableitung
C (Konstante)	0	$\sec x$	$\frac{\sin x}{\cos^2 x}$
x	1	$\sec^{-1} x$	$\frac{-\cos x}{\sin^2 x}$
x^n ($n \in \mathbb{R}$)	nx^{n-1}	$\arcsin x$ ($ x < 1$)	$\frac{1}{\sqrt{1-x^2}}$
$\frac{1}{x}$	$-\frac{1}{x^2}$	$\arccos x$ ($ x < 1$)	$-\frac{1}{\sqrt{1-x^2}}$
$\frac{1}{x^n}$	$-\frac{n}{x^{n+1}}$	$\arctan x$	$\frac{1}{1+x^2}$
\sqrt{x}	$\frac{1}{2\sqrt{x}}$	$\operatorname{arccot} x$	$-\frac{1}{1+x^2}$
$\sqrt[n]{x}$ ($n \in \mathbb{R}, n \neq 0, x > 0$)	$\frac{1}{n\sqrt[n]{x^{n-1}}}$	$\operatorname{arcsec} x$	$\frac{1}{x\sqrt{x^2-1}}$
e^x	e^x	$\operatorname{arccossec} x$	$-\frac{1}{x\sqrt{x^2-1}}$
e^{bx} ($b \in \mathbb{R}$)	be^{bx}	$\sinh x$	$\cosh x$
a^x ($a > 0$)	$a^x \ln a$	$\cosh x$	$\sinh x$
a^{bx} ($b \in \mathbb{R}, a > 0$)	$ba^{bx} \ln a$	$\tanh x$	$\frac{1}{\cosh^2 x}$
$\ln x$	$\frac{1}{x}$	$\coth x$ ($x \neq 0$)	$-\frac{1}{\sinh^2 x}$
$\log_a x$ ($a > 0, a \neq 1, x > 0$)	$\frac{1}{x} \log_a e = \frac{1}{x \ln a}$	$\operatorname{Arsinh} x$	$\frac{1}{\sqrt{1+x^2}}$
$\lg x$ ($x > 0$)	$\frac{1}{x} \lg e \approx \frac{0.4343}{x}$	$\operatorname{Arcosh} x$ ($x > 1$)	$\frac{1}{\sqrt{x^2-1}}$
$\sin x$	$\cos x$	$\operatorname{Artanh} x$ ($ x < 1$)	$\frac{1}{1-x^2}$
$\cos x$	$-\sin x$	$\operatorname{Arcoth} x$ ($ x > 1$)	$-\frac{1}{x^2-1}$
$\tan x$ ($x \neq (2k+1)\frac{\pi}{2}, k \in \mathbb{Z}$)	$\frac{1}{\cos^2 x} = \sec^2 x$	$[f(x)]^n$ ($n \in \mathbb{R}$)	$n[f(x)]^{n-1} f'(x)$
$\cot x$ ($x \neq k\pi, k \in \mathbb{Z}$)	$\frac{-1}{\sin^2 x} = -\operatorname{cosec}^2 x$	$\ln f(x)$ ($f(x) > 0$)	$\frac{f'(x)}{f(x)}$

6.11 Einige unbestimmte Integrale S1074

$\int dx = x + C$	$\int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1} + C, x \in \mathbb{R}^+, \alpha \in \mathbb{R} \setminus \{-1\}$
$\int \frac{1}{x} dx = \ln x + C, x \neq 0$	$\int e^x dx = e^x + C$
$\int a^x dx = \frac{a^x}{\ln a} + C, a \in \mathbb{R}^+ \setminus \{1\}$	$\int \sin x dx = -\cos x + C$
$\int \cos x dx = \sin x + C$	$\int \frac{dx}{\sin^2 x} = -\cot x + C, x \neq k\pi \text{ mit } k \in \mathbb{Z}$
$\int \frac{dx}{\cos^2 x} = \tan x + C, x \neq \frac{\pi}{2} + k\pi \text{ mit } k \in \mathbb{Z}$	$\int \sinh x dx = \cosh x + C$
$\int \cosh x dx = \sinh x + C$	$\int \frac{dx}{\sinh^2 x} = -\coth x + C, x \neq 0$
$\int \frac{dx}{\cosh^2 x} = \tanh x + C$	$\int \frac{dx}{ax+b} = \frac{1}{a} \ln ax+b + C, a \neq 0, x \neq -\frac{b}{a}$
$\int \frac{dx}{a^2x^2+b^2} = \frac{1}{ab} \arctan \frac{b}{a}x + C, a \neq 0, b \neq 0$	$\int \frac{dx}{a^2x^2-b^2} = \frac{1}{2ab} \ln \left \frac{ax-b}{ax+b} \right + C, a \neq 0, b \neq 0, x \neq \pm \frac{b}{a}$
$\int \sqrt{a^2x^2+b^2} dx = \frac{x}{2} \sqrt{a^2x^2+b^2} + \frac{b^2}{2a} \ln(ax + \sqrt{a^2x^2+b^2}) + C, a \neq 0, b \neq 0$	$\int \sqrt{a^2x^2-b^2} dx = \frac{x}{2} \sqrt{a^2x^2-b^2} - \frac{b^2}{2a} \ln ax + \sqrt{a^2x^2-b^2} + C, a \neq 0, b \neq 0, a^2x^2 \geq b^2$
$\int \sqrt{b^2-a^2x^2} dx = \frac{x}{2} \sqrt{b^2-a^2x^2} + \frac{b^2}{2a} \arcsin \frac{a}{b}x + C, a \neq 0, b \neq 0, a^2x^2 \leq b^2$	$\int \frac{dx}{\sqrt{a^2x^2-b^2}} = \frac{1}{a} \ln(ax + \sqrt{a^2x^2+b^2}) + C, a \neq 0, b \neq 0$
$\int \frac{dx}{\sqrt{a^2x^2-b^2}} = \frac{1}{a} \ln ax + \sqrt{a^2x^2-b^2} + C, a \neq 0, b \neq 0, a^2x^2 > b^2$	$\int \frac{dx}{\sqrt{b^2-a^2x^2}} = \frac{1}{a} \arcsin \frac{a}{b}x + C, a \neq 0, b \neq 0, a^2x^2 < b^2$
Die Integrale $\int \frac{dx}{X}, \int \sqrt{X} dx, \int \frac{dx}{\sqrt{X}}$ mit $X = ax^2 + 2bx + c, a \neq 0$ werden durch die Umformung $X = a(x + \frac{b}{a})^2 + (c - \frac{b^2}{a})$ und die Substitution $t = x + \frac{b}{a}$ in die oberen 4 Zeilen transformiert.	$\int \frac{x dx}{X} = \frac{1}{2a} \ln X - \frac{b}{a} \int \frac{dx}{X}, a \neq 0, X = ax^2 + 2bx + c$
$\int \sin^2 ax dx = \frac{x}{2} - \frac{1}{4a} \cdot \sin 2ax + C, a \neq 0$	$\int \cos^2 ax dx = \frac{x}{2} + \frac{1}{4a} \cdot \sin 2ax + C, a \neq 0$
$\int \sin^n ax dx = -\frac{\sin^{n-1} ax \cos ax}{na} + \frac{n-1}{n} \int \sin^{n-2} ax dx, n \in \mathbb{N}, a \neq 0$	$\int \cos^n ax dx = \frac{\cos^{n-1} ax \sin ax}{na} + \frac{n-1}{n} \int \cos^{n-2} ax dx, n \in \mathbb{N}, a \neq 0$
$\int \frac{dx}{\sin ax} = \frac{1}{a} \ln \left \tan \frac{ax}{2} \right + C, a \neq 0, x \neq k\frac{\pi}{a} \text{ mit } k \in \mathbb{Z}$	$\int \frac{dx}{\cos ax} = \frac{1}{a} \ln \left \tan \left(\frac{ax}{2} + \frac{\pi}{4} \right) \right + C, a \neq 0, x \neq \frac{\pi}{2a} + k\frac{\pi}{a} \text{ mit } k \in \mathbb{Z}$
$\int \tan ax dx = -\frac{1}{a} \ln \cos ax + C, a \neq 0, x \neq \frac{\pi}{2a} + k\frac{\pi}{a} \text{ mit } k \in \mathbb{Z}$	$\int \cot ax dx = \frac{1}{a} \ln \sin ax + C, a \neq 0, x \neq k\frac{\pi}{a} \text{ mit } k \in \mathbb{Z}$
$\int x^n \sin ax dx = -\frac{x^n}{a} \cos ax + \frac{n}{a} \int x^{n-1} \cos ax dx, n \in \mathbb{N}, a \neq 0$	$\int x^n \cos ax dx = \frac{x^n}{a} \sin ax - \frac{n}{a} \int x^{n-1} \sin ax dx, n \in \mathbb{N}, a \neq 0$
$\int x^n e^{ax} dx = \frac{1}{a} x^n e^{ax} - \frac{n}{a} \int x^{n-1} e^{ax} dx, n \in \mathbb{N}, a \neq 0$	$\int e^{ax} \sin bxdx = \frac{e^{ax}}{a^2+b^2} (a \sin bx - b \cos bx) + C, a \neq 0, b \neq 0$
$\int e^{ax} \cos bxdx = \frac{e^{ax}}{a^2+b^2} (a \cos bx + b \sin bx) + C, a \neq 0, b \neq 0$	$\int \ln x dx = x(\ln x - 1) + C, x \in \mathbb{R}^+$
$\int x^\alpha \cdot \ln x dx = \frac{x^{\alpha+1}}{(\alpha+1)^2} [(\alpha+1) \ln x - 1] + C, x \in \mathbb{R}^+, \alpha \in \mathbb{R} \setminus \{-1\}$	