

Schätzen und Testen

Andreas Handl

Torben Kuhlenkasper

8. Januar 2016

Grundlage des vorliegenden Skripts sind Aufzeichnungen von Andreas Handl, die er bis zum Jahr 2007 an der Universität Bielefeld verfasst und für seine Lehrveranstaltungen verwendet hat. Seit 2012 werden die Skripten von Torben Kuhlenkasper weitergeführt sowie fortlaufend aktualisiert und erweitert.

Anmerkungen und Vorschläge zur Verbesserung und Ergänzung sind jederzeit willkommen und können an statistik@kuhlenkasper.de gesendet werden. Weitere Skripten sind unter www.skripten.kuhlenkasper.de zu finden.

Inhaltsverzeichnis

1	Das Einstichprobenproblem	4
1.1	Die klassische Analyse eines Datensatzes	4
1.1.1	Die Daten und einige Annahmen	4
1.1.2	Schätzen	8
1.1.3	Testen	48
1.2	Die Normalverteilungsannahme	76
1.3	Robuste Schätzer	97
1.4	Nichtparametrische Tests	112
1.4.1	Der Vorzeichentest	112
1.4.2	Fishers Permutationsprinzip	119
1.4.3	Ränge	125
1.4.4	Der Vorzeichen-Rangtest	131
1.4.5	Vergleich der Test	140
2	Das Zweistichprobenproblem	147
2.1	Verbundene Stichproben	151
2.1.1	Stetige Variablen	151
2.1.2	Binäre Variablen	172
2.2	Unverbundene Stichproben	176
2.2.1	Kategoriale Variablen	176
2.2.2	Stetige Variablen	186
3	Das c-Stichprobenproblem	212
3.1	Unverbundene Stichproben	213
3.1.1	Einfaktorielle Varianzanalyse	213
3.1.2	Kruskal-Wallis-Test	226
3.1.3	Der Jonckheere-Test	236
3.2	Verbundene Stichproben	249

INHALTSVERZEICHNIS

3

3.2.1	Zweifaktorielle Varianzanalyse	249
3.2.2	Der Friedman-Test	274

Kapitel 1

Grundlegende Konzepte – Das Einstichprobenproblem

1.1 Die klassische Analyse eines Datensatzes

1.1.1 Die Daten und einige Annahmen

Ziel der Statistik ist es, Strukturen und Gesetzmäßigkeiten in Datensätzen aufzudecken. Datensätze werden in der Regel mit einem bestimmten Ziel erhoben.

Schauen wir uns ein Beispiel an, das aus dem Buch Small Data Sets von Hand stammt.

Die Schmuckstücke an den Kleidungsstücken der Schoschonen sind rechteckig. Ein Forscher will nun untersuchen, ob diese Rechtecke nach dem goldenen Schnitt gefertigt wurden.

Ein Rechteck weist den goldenen Schnitt auf, wenn gilt

$$\frac{b}{l} = \frac{l}{b+l},$$

wobei b die Länge der kürzeren und l die Länge der längeren Seite ist.

Es muß gelten

$$\frac{b}{l} = \frac{\sqrt{5}-1}{2} \approx 0.618.$$

Dies sieht man folgendermaßen:

Aus

$$\frac{b}{l} = \frac{l}{b+l}$$

folgt

$$b^2 + bl = l^2$$

Dividieren wir beide Seiten durch l^2 , so erhalten wir

$$\left(\frac{b}{l}\right)^2 + \frac{b}{l} = 1$$

Diese quadratische Gleichung hat die positive Lösung

$$\frac{b}{l} = \frac{\sqrt{5} - 1}{2} \approx 0.618$$

Der Forscher bestimmt nun von 20 rechteckigen Schmuckstücken der Schoschonen das Verhältnis von b zu l .

Es ergaben sich folgende Zahlen:

0.693 0.662 0.690 0.606 0.570 0.749 0.672 0.628 0.609 0.844
0.654 0.615 0.668 0.601 0.576 0.670 0.606 0.611 0.553 0.933

Auf den ersten Blick fällt auf, daß die Zahlen sehr unterschiedlich sind.

Einen besseren Überblick erhält man, wenn man sich den geordneten Datensatz anschaut.

0.553 0.570 0.576 0.601 0.606 0.606 0.609 0.611 0.615 0.628
0.654 0.662 0.668 0.670 0.672 0.690 0.693 0.749 0.844 0.933

Der kleinste Wert ist 0.553 und der größte Wert ist 0.933. Der Wert 0.618 liegt im Bereich der Daten.

Man kann sich vorstellen, daß die Schoschonen eine Vorstellung von einem ästhetischen Verhältnis von Breite zu Länge bei den Rechtecken hatten und dieses Verhältnis auch erreichen wollten. Aufgrund der Unvollkommenheit der Fertigung wird dieses Verhältnis aber im Einzelfall nicht erreicht. Die einzelnen Rechtecke streuen um diesen charakteristischen Wert.

Wir wollen nun überprüfen, ob dieser Wert 0.618 ist.

Hierzu fassen wir das Verhältnis von Breite b zu Länge l als Zufallsvariable X auf.

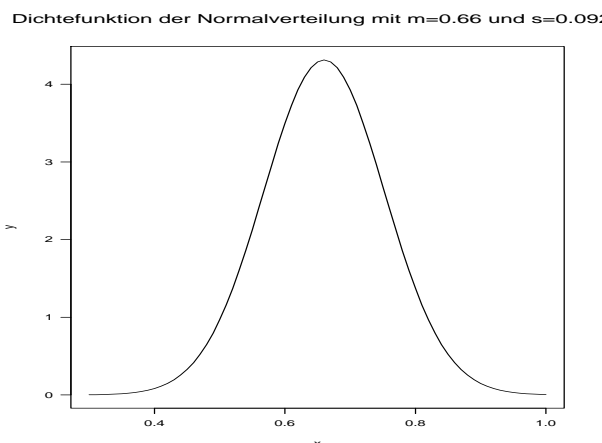
Die klassische Annahme ist, daß X mit den Parametern μ und σ^2 normalverteilt ist. Wir werden uns zunächst damit beschäftigen, wie wir vorzugehen haben, wenn diese Annahme erfüllt ist. Später beschäftigen wir uns mit Verfahren, die ohne die Annahme der Normalverteilung auskommen.

Die Dichtefunktion einer mit den Parametern μ und σ^2 normalverteilten Zufallsvariablen lautet:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{für } x \in \mathfrak{R}$$

Der Parameter μ ist der Erwartungswert und der Parameter σ^2 die Varianz der normalverteilten Zufallsvariablen X .

Die folgende Graphik zeigt die Dichtefunktion einer Normalverteilung mit den Parametern $\mu = 0.66$ und $\sigma = 0.0925$.



Warum die Werte der Parameter so gewählt wurden, wird später ersichtlich werden. Wir sehen, daß die Dichtefunktion der Normalverteilung symmetrisch hinsichtlich μ ist.

Wenn wir unterstellen, daß die Rechtecke nach dem goldenen Schnitt gefertigt wurden, so ist 0.618 der Wert, den wir für das Seitenverhältnis erwarten. Dieser Wert sollte also das Zentrum der Verteilung bilden.

Für das Zentrum der Verteilung können wir unterschiedliche Maßzahlen wählen.

In der Regel betrachtet man den **Erwartungswert** $E(X)$ oder den **Median** M . Wir wählen zunächst den Erwartungswert.

Wir nehmen also an, daß der Erwartungswert gleich 0.618 ist.

Wir wollen also überprüfen, ob der $E(X)$ gleich 0.618 ist.

Diese Annahme bezeichnet man als **statistische Hypothese**.

Um diese Hypothese zu überprüfen, entnehmen wir der Grundgesamtheit aller Rechtecke $n = 20$ Rechtecke.

Das Seitenverhältnis x_1 ist beim ersten Rechteck die Realisierung einer Zufallsvariablen X_1 , die genauso verteilt ist wie X , da wir dieses Rechteck der Grundgesamtheit entnommen haben.

Nach dem Zug ist der Wert bekannt, vor dem Zug ist aber jeder Wert der Grundgesamtheit möglich.

Dies gilt auch für die anderen Beobachtungen, wenn wir mit Zurücklegen ziehen, oder wenn die Grundgesamtheit so groß ist, daß das Fehlen einer oder mehrerer Beobachtungen keinen großen Einfluß auf die Verteilung der Grundgesamtheit hat.

Man spricht in diesem Fall von einer **Zufallsstichprobe**.

Wir fassen die Beobachtungen x_1, \dots, x_n also als Realisationen von unabhängigen, identisch mit Verteilungsfunktion $F_X(x)$ verteilten Zufallsvariablen X_1, \dots, X_n auf. Identisch verteilt bedeutet, daß alle Zufallsvariablen die gleiche Verteilung besitzen.

Auf der Basis der Stichprobe wollen wir die Annahme überprüfen. Dies geschieht mit Hilfe eines **statistischen Tests**. Mit statistischen Tests werden wir uns später detailliert beschäftigen.

Vorher schauen wir uns aber **Schätzer** an.

Hier wird aus der Stichprobe ein oder mehrere Werte für den unbekanntem Parameter gewonnen.

Im ersten Fall spricht man **Punktschätzung** und im zweiten von **Intervallschätzung**.

1.1.2 Schätzen

Es sollen also entweder ein Wert oder mehrere Werte für den unbekanntem Parameter angegeben werden.

Im ersten Fall spricht man von **Punktschätzung**, im zweiten Fall von **Intervallschätzung**.

Im folgenden gehen wir davon aus, daß die Verteilungsfunktion $F_X(x)$ und die Dichtefunktion bzw. Wahrscheinlichkeitsfunktion $f_X(x)$ von einem oder mehreren Parametern abhängt, den oder die wir mit θ bezeichnen.

Bei der Normalverteilung kann der interessierende Parameter der Erwartungswert μ oder die Varianz σ^2 sein.

Es kann aber auch der Vektor $\theta = (\mu, \sigma^2)$ sein.

Bei der Punktschätzung bestimmen wir aus den Beobachtungen x_1, \dots, x_n einen Wert $g(x_1, \dots, x_n)$, den wir **Schätzwert** nennen.

Die Beobachtungen x_1, \dots, x_n sind die Realisationen der Zufallsvariablen X_1, \dots, X_n .

Somit ist der Schätzwert $g(x_1, \dots, x_n)$ die Realisation der Zufallsvariablen $g(X_1, \dots, X_n)$.

Wir nennen $g(X_1, \dots, X_n)$ allgemein eine **Stichprobenfunktion**.

Im Fall der Punktschätzung heißt $g(X_1, \dots, X_n)$ **Schätzfunktion**.

Wir schreiben

$$\hat{\theta}_n = g(X_1, \dots, X_n).$$

Wie kann man nun eine geeignete Schätzfunktion finden?

Ein Verfahren zur Gewinnung einer geeigneten Schätzfunktion ist die **Maximum-Likelihood-Methode**, die wir mit M-L-Methode abkürzen.

Um die M-L-Methode zu motivieren, unterstellen wir, daß die Zufallsvariable X diskret ist.

Schauen wir uns ein ganz einfaches Beispiel an:

Eine Urne enthält 5 Kugeln, wobei es zwei mögliche Zusammensetzungen der Urne gibt:

Zusammensetzung I: 4 schwarze und 1 weiße Kugel

Zusammensetzung II: 2 schwarze und 3 weiße Kugeln

Die Zusammensetzung der Urne sei unbekannt.

Um sich ein Bild von der Zusammensetzung der Urne zu machen, entnimmt man ihr zufällig eine Kugel.

Die gezogene Kugel sei weiß.

Für welche Zusammensetzung der Urne spricht dieses Ergebnis?

Bei der ersten Zusammensetzung der Urne beträgt die Wahrscheinlichkeit 0.2, eine weiße Kugel zu ziehen.

Bei der zweiten Zusammensetzung beträgt diese Wahrscheinlichkeit 0.6.

Da es wahrscheinlicher ist, aus der zweiten Zusammensetzung eine weiße Kugel zu ziehen, ist es also viel plausibler, daß die Urne die zweite Zusammensetzung aufweist.

Dies ist die Entscheidungsregel der Maximum-Likelihood-Methode (M-L-Methode):

Wir entscheiden uns für den Zustand der Welt, bei dem die beobachtete Stichprobe am wahrscheinlichsten ist.

Versuchen wir nun, diese Vorgehensweise formaler darzustellen:

Sei p der Anteil der weißen Kugeln in der Urne.

Bei der ersten Zusammensetzung nimmt p den Wert 0.2, bei der zweiten Zusammensetzung p den Wert 0.6 an.

Unsere Entscheidung über die Zusammensetzung der Urne beruht auf der Farbe der gezogenen Kugel.

Wir betrachten die Zufallsvariable X : Anzahl der gezogenen weißen Kugeln. Die Zufallsvariable X kann die Werte 0 und 1 annehmen.

Ist die gezogene Kugel weiß, so nimmt sie den Wert 1 an, ansonsten den Wert 0.

Die Wahrscheinlichkeitsverteilung von X hängt natürlich vom Wert von p ab. Sie ist in der folgenden Tabelle zu finden:

p	0.2	0.6
x		
0	0.8	0.4
1	0.2	0.6

Jede Spalte der Tabelle stellt die Wahrscheinlichkeitsverteilung von X in Abhängigkeit von p dar.

Eine Zeile der Tabelle ist keine Wahrscheinlichkeitsverteilung. Sie sagt vielmehr aus, wie wahrscheinlich eine Beobachtung unter den verschiedenen Werten des Parameters ist.

Die Eintragungen in einer Zeile werden als **Likelihoods** des Parameters gegeben die Beobachtung bezeichnet.

Die gesamte Zeile heißt **Likelihoodfunktion**.

Das Maximum-Likelihood-Prinzip besagt nun, denjenigen Wert des Parameters zu wählen, für den die Likelihood am größten ist, für den die Likelihood also ihr Maximum annimmt.

Man kann das Maximum-Likelihood-Prinzip auch so beschreiben:

Wähle den Wert des Parameters, für den die Wahrscheinlichkeit des Auftretens der Stichprobe am größten ist.

Sehen wir uns an, wie die Entscheidungsregel aussieht, wenn zwei Kugeln mit Zurücklegen aus der Urne gezogen werden.

Sei X_i die Anzahl der beim i -ten Zug gezogenen weißen Kugeln, $i = 1, 2$.

Wir bestimmen die folgenden Wahrscheinlichkeiten:

$$P(X_1 = x_1, X_2 = x_2)$$

Liegt die erste Zusammensetzung der Urne vor, so gilt:

$$P(X_1 = 0, X_2 = 0) = P(X_1 = 0) \cdot P(X_2 = 0) = 0.8 \cdot 0.8 = 0.64$$

$$P(X_1 = 0, X_2 = 1) = P(X_1 = 0) \cdot P(X_2 = 1) = 0.8 \cdot 0.2 = 0.16$$

$$P(X_1 = 1, X_2 = 0) = P(X_1 = 1) \cdot P(X_2 = 0) = 0.2 \cdot 0.8 = 0.16$$

$$P(X_1 = 1, X_2 = 1) = P(X_1 = 1) \cdot P(X_2 = 1) = 0.2 \cdot 0.2 = 0.04$$

Liegt die zweite Zusammensetzung der Urne vor, so gilt:

$$P(X_1 = 0, X_2 = 0) = P(X_1 = 0) \cdot P(X_2 = 0) = 0.4 \cdot 0.4 = 0.16$$

$$P(X_1 = 0, X_2 = 1) = P(X_1 = 0) \cdot P(X_2 = 1) = 0.4 \cdot 0.6 = 0.24$$

$$P(X_1 = 1, X_2 = 0) = P(X_1 = 1) \cdot P(X_2 = 0) = 0.6 \cdot 0.4 = 0.24$$

$$P(X_1 = 1, X_2 = 1) = P(X_1 = 1) \cdot P(X_2 = 1) = 0.6 \cdot 0.6 = 0.36$$

Wir erhalten für die einzelnen Stichproben folgende Tabelle:

(x_1, x_2)	p	0.2	0.6
(0,0)		0.64	0.16
(0,1)		0.16	0.24
(1,0)		0.16	0.24
(1,1)		0.04	0.36

Sind beide Kugeln schwarz, beobachten wir also $(0, 0)$, so entscheiden wir uns aufgrund des Maximum-Likelihood-Prinzips für die erste Zusammensetzung der Urne. In allen anderen Fällen nehmen wir an, daß der zweite Zustand vorliegt.

Wir können nun das Maximum-Likelihood-Verfahren für den diskreten Fall allgemein formulieren:

X_1, \dots, X_n seien unabhängige, identisch verteilte diskrete Zufallsvariablen, deren Verteilung von einem unbekanntem Parameter θ abhängt.

Wir wollen θ auf der Basis der Realisationen x_1, \dots, x_n schätzen.

Dann ist

$$P(X_1 = x_1, \dots, X_n = x_n, \theta)$$

die Wahrscheinlichkeit für das Auftreten der Stichprobe x_1, \dots, x_n in Abhängigkeit von θ .

Diese Wahrscheinlichkeit fassen wir bei gegebenem x_1, \dots, x_n als Funktion von θ auf und nennen sie **Likelihood** $L(\theta)$.

Der Maximum-Likelihood-Schätzer $\hat{\theta}_n$ ist nun der Wert von θ , für den die Likelihood am größten ist:

$$L(\hat{\theta}_n) = \max_{\theta} L(\theta)$$

Aus technischen Gründen betrachtet man in der Regel den Logarithmus der Likelihoodfunktion.

Man erhält also die sogenannte **Loglikelihoodfunktion**:

$$l(\theta) = \ln L(\theta)$$

Da der Logarithmus eine monotone Transformation ist, nimmt die Loglikelihoodfunktion ihr Maximum an der gleichen Stelle an wie die Likelihoodfunktion.

Schauen wir uns das obige Beispiel für den Stichprobenumfang n an. Außerdem schränken wir die möglichen Werte von p nicht von vornherein ein. Es sind also alle Werte von p im Intervall $(0, 1)$ möglich.

X_1, \dots, X_n seien also unabhängige, identisch mit Parameter p bernoulliverteilte Zufallsvariablen.

Es gilt also

$$P(X_i = x_i) = p^{x_i} (1 - p)^{1-x_i} \quad \text{für } x_i = 0, 1$$

Die Likelihood lautet also

$$\begin{aligned} L(p) &= p^{x_1} (1-p)^{1-x_1} \dots p^{x_n} (1-p)^{1-x_n} \\ &= p^{\sum x_i} (1-p)^{n-\sum x_i} \\ &= p^{n\bar{x}} (1-p)^{n(1-\bar{x})} \end{aligned}$$

Die Loglikelihood ist:

$$l(p) = n\bar{x} \ln p + n(1-\bar{x}) \ln(1-p)$$

Zur Bestimmung des M-L-Schätzers bilden wir die erste Ableitung:

$$\begin{aligned} \frac{\delta}{\delta p} l(p) &= \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p} \\ &= \frac{n(\bar{x}-p)}{p(1-p)} \end{aligned}$$

Notwendige Bedingung für einen Extremwert ist, daß die erste Ableitung gleich 0 ist.

Es muß also gelten

$$\frac{n(\bar{x}-\hat{p})}{\hat{p}(1-\hat{p})} = 0$$

Diese Gleichung wird erfüllt von

$$\hat{p} = \bar{x}$$

Wir überprüfen noch die hinreichenden Bedingungen.

Es gilt

$$\frac{\delta^2}{\delta p^2} l(p) = -\frac{n\bar{x}}{p^2} - \frac{n(1-\bar{x})}{(1-p)^2}$$

Da gilt

$$0 \leq \bar{x} \leq 1$$

gilt

$$\frac{\delta^2}{\delta p^2} l(p) < 0$$

und es handelt sich um ein Maximum.

Somit ist der M-L-Schätzer von p :

$$\hat{p} = \bar{x}$$

Dies ist gerade die **relative Häufigkeit**.

Bei stetigen Zufallsvariablen ist die Likelihoodfunktion die gemeinsame Dichtefunktion der Zufallsvariablen X_1, \dots, X_n :

$$L(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

Wir unterstellen in der Regel, daß die Zufallsvariablen X_1, \dots, X_n unabhängig sind.

In diesem Fall ist die gemeinsame Dichtefunktion das Produkt der einzelnen Dichtefunktionen und die Likelihoodfunktion lautet:

$$L(\theta) = \prod_{i=1}^n f_{X_i}(x_i, \theta)$$

und für $l(\theta)$ gilt

$$l(\theta) = \sum_{i=1}^n \ln f_{X_i}(x_i, \theta)$$

Unterstellen wir Normalverteilung, so ist die Loglikelihoodfunktion von μ bei festem σ^2 gegeben durch

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

Notwendige Bedingung für einen Extremwert in $\hat{\mu}$ ist, daß die erste Ableitung an der Stelle $\hat{\mu}$ gleich 0 ist.

Die erste Ableitung ist gegeben durch

$$\begin{aligned} \frac{\delta}{\delta\mu} l(\mu) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ &= \frac{n}{\sigma^2} (\bar{x} - \mu) \end{aligned}$$

Für $\hat{\mu}$ muß also gelten

$$\frac{n}{\sigma^2} (\bar{x} - \hat{\mu}) = 0$$

Hieraus folgt

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Die zweite Ableitung lautet

$$\frac{\delta^2}{\delta\mu^2} l(\mu) = -\frac{n}{\sigma^2}$$

Da σ^2 größer als 0 ist, gilt

$$\frac{\delta^2}{\delta\mu^2} l(\mu) < 0$$

Der Mittelwert \bar{X} ist also der M-L-Schätzer von μ bei Normalverteilung.

Wir wollen nun den Mittelwert der Schoschonen-Daten mit R bestimmen.

Wir geben zunächst die Daten ein.

Dies geschieht folgendermaßen:

Wir definieren eine Variable `shosho`, indem wir ihr die Werte in Form eines Vektors zuweisen.

Die Zuweisung geschieht durch `<-` (das Kleinerzeichen und das Minuszeichen).

Die Funktion `c` erstellt aus einer Folge von Werten, die durch Komma getrennt sind, einen Vektor.

Die Werte sind die Argumente der Funktion `c`.

Argumente einer Funktion stehen in runden Klammern hinter dem Funktionsnamen und sind durch Kommata voneinander getrennt.

Wir geben also ein:

```
shosho <- c(0.693,0.662,0.690,0.606,0.570,0.749,0.672,
           0.628,0.609,0.844,0.654,0.615,0.668,0.601,
           0.576,0.670,0.606,0.611,0.553,0.933)
```

Die Daten kann man sich anschauen, indem man den Namen des Vektors eingibt:

```
shosho
[1] 0.693 0.662 0.690 0.606 0.570 0.749 0.672 0.628
[9] 0.609 0.844 0.654 0.615 0.668 0.601 0.576 0.670
[17] 0.606 0.611 0.553 0.933
```

Dabei geben hier `[1]`, `[9]` und `[17]` am jeweiligen Anfang der Ausgabe das 1., 9. und 17. Element des Vektors `shosho` an, das jeweils rechts des Indexes steht.

Auf die erste Komponente kann man zugreifen durch:

```
shosho[1]
[1] 0.693
```

Das 9. und 17. Element wird durch

```
shosho[c(9,17)]
[1] 0.609 0.606
```

ausgegeben. Entsprechend erhalten wir die anderen Komponenten.

Wir hatten am Anfang des Kapitels einen Überblick über den Datensatz `shosho` dadurch gewonnen, daß wir die Daten sortiert hatten.

Hierzu gibt es in R die Funktion `sort`.

Der Aufruf

```
sort(shosho)
[1] 0.553 0.570 0.576 0.601 0.606 0.606 0.609 0.611
[9] 0.615 0.628 0.654 0.662 0.668 0.670 0.672 0.690
[17] 0.693 0.749 0.844 0.933
```

liefert das gewünschte Ergebnis.

Das Maximum der Daten erhalten wir mit der Funktionen `max`

```
max(shosho)
[1] 0.933
```

und das Minimum mit der Funktion `min`

```
min(shosho)
[1] 0.553.
```

Um den Mittelwert zu bestimmen, benutzen wir die Funktion `mean`.

Das Argument von `mean` ist der Vektor der Daten, deren Mittelwert bestimmt werden soll.

```
mean(shosho)
[1] 0.6605
```

Das Ergebnis ist 0.6605.

Wir wollen nun noch die M-L-Schätzung des Parametervektors $\theta = (\mu, \sigma^2)$ bei Normalverteilung herleiten.

Die Loglikelihood lautet

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

Die notwendigen Bedingungen für einen Extremwert lauten:

$$\frac{\partial}{\partial \mu} l(\mu, \sigma^2) = \frac{n}{\sigma^2} (\bar{x} - \mu) = 0$$

und

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

Aus der ersten Gleichung folgt

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Setzt man diesen Ausdruck in die zweite Gleichung ein und löst diese nach σ^2 auf, so erhält man:

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

Man nennt dies auch die **mittlere quadratische Abweichung**.

Diese erhalten wir folgendermaßen in R :

```
mean((shosho-mean(shosho))^2)
[1] 0.00813035
```

Dabei wird zunächst der Befehl `shosho-mean(shosho)` ausgeführt. Von jeder Komponente des Vektors `shosho` wird der Mittelwert subtrahiert.

Man erhält also den zentrierten Vektor:

```
shosho-mean(shosho)
[1] 0.0325 0.0015 0.0295 -0.0545 -0.0905 0.0885
[7] 0.0115 -0.0325 -0.0515 0.1835 -0.0065 -0.0455
[13] 0.0075 -0.0595 -0.0845 0.0095 -0.0545 -0.0495
[19] -0.1075 0.2725
```

Dann wird jede Komponente dieses Vektors quadriert:

```
(shosho-mean(shosho))^2
[1] 0.00105625 0.00002225 0.00087025 0.00297025 0.00819025
[6] 0.00783225 0.00013225 0.00105625 0.00265225 0.03367225
[11] 0.00004225 0.00207025 0.00005625 0.00354025 0.00714025
[16] 0.00009025 0.00297025 0.00245025 0.01155625 0.07425625
```

Anschließend wird der Mittelwert dieses Vektors bestimmt:

```
mean((shosho-mean(shosho))^2)
[1] 0.00813035
```

Wie gut ist eigentlich eine Schätzfunktion?

Um diese Frage beantworten zu können, muß man sich erst einmal über die Verteilung der Schätzfunktion Gedanken machen.

Hierzu schauen wir uns ein einfaches Beispiel an.

Eine Urne enthält 10 Kugeln, von denen fünf 10 g und fünf 20 g wiegen.

Von Interesse ist das Durchschnittsgewicht der Kugeln in der Urne.

Da wir den Inhalt der Urne kennen, können wir den Wert des Durchschnittsgewichts θ problemlos bestimmen. Der Wert von θ ist 15.

Nun wollen wir aber davon ausgehen, daß der Wert von θ unbekannt ist, und wir nur eine Zufallsstichprobe vom Umfang $n=2$ ziehen können.

Wir ziehen also 2 Kugeln mit Zurücklegen aus der Urne.

Sei X_1 das Gewicht der ersten gezogenen Kugel und X_2 das Gewicht der zweiten gezogenen Kugel.

Dann gilt

$$P(X_1 = 10) = 0.5$$

$$P(X_1 = 20) = 0.5$$

$$P(X_2 = 10) = 0.5$$

$$P(X_2 = 20) = 0.5$$

Wir können folgende Stichproben beobachten:

(10, 10)

(10, 20)

(20, 10)

(20, 20)

Wählen wir \bar{X} als Schätzer für μ , so sind folgende Werte für \bar{X} möglich:

10 15 20.

Es gilt:

$$P(\bar{X} = 10) = P(X_1 = 10, X_2 = 10) = 0.25$$

$$P(\bar{X} = 15) = P(X_1 = 10, X_2 = 20) + P(X_1 = 20, X_2 = 10) = 0.5$$

$$P(\bar{X} = 20) = P(X_1 = 20, X_2 = 20) = 0.25$$

Wir können die Wahrscheinlichkeitsfunktion in R problemlos graphisch darstellen.

Wir erzeugen einen Vektor `x` mit den Positionen

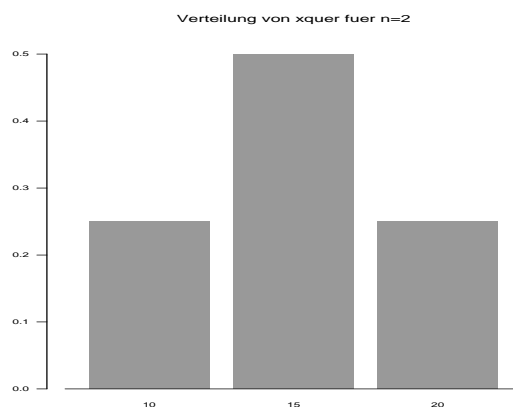
```
x <- c(10,15,20)
```

und einen Vektor `h` mit den Wahrscheinlichkeiten:

```
h <- c(0.25,0.5,0.25)
```

Die Graphik erzeugt man mit Hilfe der Funktion `barplot` durch folgenden Aufruf:

```
barplot(h, names=as.character(x),  
        main="Verteilung von xquer fuer n=2")
```



Das Argument `names` enthält die Namen der Positionen der Rechtecke. Es muß sich um einen Vektor handeln, dessen Elemente Zeichenketten sind. Einen Vektor, der eine Zeichenkette enthält, erzeugt man durch Zuweisung, z.B.:

```
st <- "Bielefeld"
```

Wir hätten also eingeben können:

```
xst <- c("10", "15", "20")
```

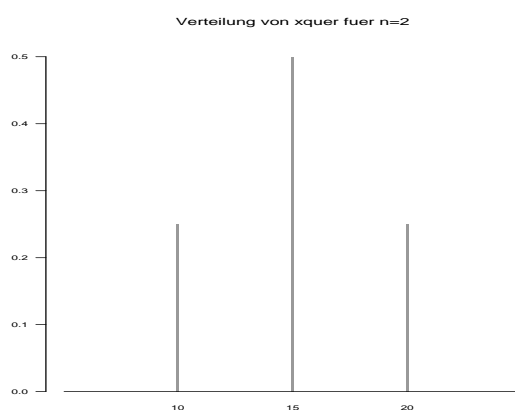
Zahlenvektoren kann man aber auch mit Hilfe der Funktion `as.character` in Zeichenkettenvektoren verwandeln, wie dies beim Aufruf von `barplot` geschieht.

Das Argument `main` versieht die Graphik mit einer Überschrift, die als Zeichenkette übergeben wird.

Die Graphik ist nicht sehr schön, da die Rechtecke sehr breit sind.

Der Parameter `space` gibt die Breite des Zwischenraums zwischen den Rechtecken in Abhängigkeit der Breite der Rechtecke an. Er ist standardmäßig auf 0.2 gesetzt. Wir erhöhen ihn auf 50 und erhalten folgendes Bild:

```
barplot(h, names=as.character(c(10, 15, 20)), space=50,
        main="Verteilung von xquer fuer n=2")
```



Die Verteilung von \bar{X} liegt um den wahren Wert 15 des Parameters θ .

Es gilt sogar:

$$E(\bar{X}) = 10 \cdot 0.25 + 15 \cdot 0.5 + 20 \cdot 0.25 = 15$$

Der Erwartungswert von \bar{X} ist im Beispiel θ .

Dies ist kein Zufall:

Sind nämlich X_1, X_2, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen mit $E(X_i) = \mu$ für $i = 1, \dots, n$.

Dann gilt

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n\mu$$

Für

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

gilt

$$E(\bar{X}) = \mu$$

Dies sieht man folgendermaßen:

$$\begin{aligned}
 E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n} n \mu \\
 &= \mu
 \end{aligned}$$

Man sagt auch, daß \bar{X} eine erwartungstreue Schätzfunktion für μ ist.

Definition 1.1.1 Eine Schätzfunktion $\hat{\theta}_n$ heißt erwartungstreu für den Parameter θ , wenn für alle θ gilt:

$$E(\hat{\theta}_n) = \theta$$

Die Schätzfunktion

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

ist nicht erwartungstreu für σ^2 .

Dies sieht man folgendermaßen:

Wegen

$$\text{Var}(X_i) = E(X_i^2) - E(X_i)^2$$

gilt

$$E(X_i^2) = \text{Var}(X_i) + E(X_i)^2 = \sigma^2 + \mu^2$$

Wegen

$$\text{Var}(\bar{X}) = E(\bar{X}^2) - E(\bar{X})^2$$

gilt

$$E(\bar{X}^2) = \text{Var}(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2$$

Also folgt

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)\end{aligned}$$

Somit folgt

$$\begin{aligned}E(\widehat{\sigma}^2) &= \frac{1}{n} \left(\sum_{i=1}^n E(X_i^2) - n E(\bar{X}^2) \right) \\ &= \frac{1}{n} (n\sigma^2 + n\mu^2 - n(\frac{\sigma^2}{n} + \mu^2)) \\ &= \frac{1}{n} (n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) \\ &= \frac{n-1}{n} \sigma^2\end{aligned}$$

Die mittlere quadratische Abweichung ist also nicht erwartungstreu für σ^2 .
Es gilt aber

$$E\left(\frac{n}{n-1} \widehat{\sigma}^2\right) = \sigma^2$$

Also ist

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

erwartungstreu für σ^2 .

Man nennt S^2 auch die **Stichprobenvarianz**.

In R können wir die Stichprobenvarianz mit Hilfe der Funktion `var` bestimmen

```
var(shosho)
[1] 0.008558263
```

Für die Schätzfunktion $\hat{\sigma}^2$ gilt:

$$\begin{aligned}\lim_{n \rightarrow \infty} E(\hat{\sigma}^2) &= \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 \\ &= \sigma^2 \lim_{n \rightarrow \infty} \frac{n-1}{n} \\ &= \sigma^2\end{aligned}$$

Definition 1.1.2 Eine Schätzfunktion $\hat{\theta}_n$ heißt asymptotisch erwartungstreu für den Parameter θ , wenn für alle Werte von θ gilt:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta.$$

Offensichtlich ist jede erwartungstreue Schätzfunktion auch asymptotisch erwartungstreu.

Die mittlere quadratische Abweichung zeigt, daß die Umkehrung nicht notwendigerweise gilt.

Die Erwartungstreue einer Schätzfunktion stellt sicher, daß das Zentrum der Verteilung der Schätzfunktion am wahren Wert des Parameters liegt. Bei einer asymptotisch erwartungstreuen Schätzfunktion gilt dies zumindest für große Stichprobenumfänge.

Da eine Schätzfunktion eine Zufallsvariable ist, kann für eine konkrete Stichprobe der Wert der Schätzfunktion weit weg vom wahren Wert des Parameters liegen. Dies sieht man sehr schön am obigen Beispiel.

Für $n = 2$ kann \bar{X} die Werte 10, 15 und 20 annehmen. In der Hälfte der Stichproben beobachtet man den wahren Wert 15, in den anderen Fällen entweder den Wert 10 oder den Wert 20.

Im Einzelfall kann man also mit der Schätzfunktion weit vom wahren Wert des Parameters entfernt sein.

Betrachten wir folgende Wahrscheinlichkeit:

$$P(13 < \bar{X} < 17)$$

Für $n = 2$ gilt:

$$P(13 < \bar{X} < 17) = P(\bar{X} = 15) = 0.5$$

Für $n=3$ kann man die Verteilung von \bar{X} noch einfach herleiten:

$$\begin{aligned}P(\bar{X} = 10) &= 0.125 \\ P(\bar{X} = 13.33) &= 0.375 \\ P(\bar{X} = 16.67) &= 0.375 \\ P(\bar{X} = 20) &= 0.125\end{aligned}$$

Also gilt:

$$P(13 < \bar{X} < 17) = 0.75$$

Im Beispiel wird diese Wahrscheinlichkeit mit wachsendem Stichprobenumfang immer größer.

Mit wachsendem Stichprobenumfang n wird es immer mühseliger, die Verteilung von \bar{X} herzuleiten.

Wir können aber mit Hilfe einer Simulation die Verteilung von \bar{X} schätzen. Hierzu erzeugen wir sehr viele Stichproben vom Umfang n aus der Verteilung der Grundgesamtheit. Für jede dieser Stichproben bestimmen wir den Wert der Schätzfunktion und schätzen die theoretische Verteilung der Schätzfunktion durch ihre empirische Verteilung.

Die Stichproben ziehen wir mit Hilfe eines Zufallszahlengenerators.

Von diesen sind sehr viele in R implementiert. Wir benötigen für unser Ziel aber nur einen, der auf $(0, 1)$ gleichverteilte Zufallszahlen erzeugt.

Der Zufallszahlengenerator für auf $(0, 1)$ gleichverteilte Zufallsvariablen heißt in R `runif` und hat als Argument die gewünschte Anzahl an Zufallszahlen.

Der Aufruf `runif(1)` erzeugt also eine auf $(0, 1)$ gleichverteilte Zufallszahl.

Die Dichtefunktion einer auf $(0, 1)$ gleichverteilten Zufallsvariablen U ist gegeben durch:

$$f_U(u) = \begin{cases} 1 & \text{für } 0 < u < 1 \\ 0 & \text{sonst} \end{cases}$$

Die Wahrscheinlichkeit für ein Teilintervall von $(0, 1)$ der Länge p beträgt somit p .

Mit Hilfe von gleichverteilten Zufallszahlen kann man Zufallszahlen aus jeder anderen Verteilung bestimmen.

Wir wollen das hier nicht vertiefen, aber für unser einfaches Beispiel zeigen.

Sei X_1 das Gewicht der ersten gezogenen Kugel.

Dann gilt

$$P(X_1 = 10) = 0.5$$

$$P(X_1 = 20) = 0.5$$

Wir können diesen einmaligen Zug dadurch simulieren, daß wir eine auf $(0, 1)$ gleichverteilte Zufallszahl u ziehen und uns folgendermaßen entscheiden:

Ist $u < 0.5$, so wiegt die gezogene Kugel 10 g, ansonsten wiegt sie 20 g.

Es gilt:

$$P(X_1 = 10) = P(U < 0.5) = 0.5$$

Dies können wir in R folgendermaßen realisieren:

```
u <- runif(1)
u
[1] 0.05707

if (u<0.5)
  x <- 10
else
  x <- 20

x
[1] 10
```

Einen Vektor v , der die Werte einer Stichprobe vom Umfang $n=2$ enthält, erhalten wir mit Hilfe einer Iteration.

Zunächst erzeugen wir einen Vektor v der Länge 2, der zwei Nullen enthält. Hierzu verwenden wir die Funktion `rep`.

Der Aufruf `rep(z,k)` erzeugt einen Vektor der Länge k aus dem Symbol z .

```
v <- rep(0,2)
v
[1] 0 0
```

Dann weisen wir in einer Schleife den Komponenten dieses Vektors Zufallszahlen aus der Verteilung zu:

```
for (i in 1:2){
  u <- runif(1)
  if (u<0.5) v[i] <- 10
  else v[i] <- 20
}

v
[1] 10 20
```

Nun haben wir eine Stichprobe erzeugt.

Um die Verteilung der Schätzfunktion zu schätzen, benötigen wir B Stichproben vom Umfang $n = 2$.

Für jede bestimmen wir den Mittelwert und weisen diesem der Komponente eines Vektors zu:

```
e <- rep(0,10000)

for (j in 1:10000){
  x <- rep(0,2)
  for (i in 1:2){
    u <- runif(1)
    if (u < 0.5) x[i] <- 10
    else x[i] <- 20
  }
  e[j] <- mean(x)
}
```

Den Vektor der absoluten Häufigkeiten erhalten wir mit Hilfe der Funktion `table`:

```
h <- table(e)

h
  10  15  20
2488 4942 2570
```

Das Ergebnis von `table` ist ein Vektor `h` der Länge drei, der die Zahlen 2488, 4942 und 2570 enthält.

Die erwarteten Häufigkeiten sind 2500, 5000 und 2500.

Die Zahlen 10, 15 und 20 sind Namen für die einzelnen Komponenten dieses Vektors `h`.

Auf diese Namen können wir mit der Funktion `dimnames` zugreifen:

```
dimnames(h)
[1] "10" "15" "20"
```

Mit der angegebenen Befehlsfolge können wir ohne Probleme größere Stichprobenumfänge simulieren.

Wir müssen die 2 durch den höheren Stichprobenumfang ersetzen.

Bevor wir dies aber tun, wollen wir die Befehlsfolge modifizieren.

Sie ist nicht falsch, aber nicht sehr geeignet für R .

Sie enthält nämlich sehr viele Iterationen. Da R eine interpretierende Sprache ist, führen Iterationen dazu, daß die Ausführung der Befehlsfolge sehr langwierig ist, da jeder Befehl immer wieder neu übersetzt werden muß, bevor er ausgeführt wird.

Man sollte also Iterationen vermeiden, wenn es geht.

Die innere Schleife können wir folgendermaßen umgehen (im folgenden ist $n=2$):

Wir erzeugen einen Vektor u der Länge n mit gleichverteilten Zufallszahlen:

```
u <- runif(2)

u
[1] 0.08879615 0.51526512
```

Dann bestimmen wir, welche der Komponenten größer als 0.5 sind:

```
u > 0.5
[1] FALSE TRUE
```

Das Ergebnis ist ein logischer Vektor, der ein `TRUE` enthält, wenn die Komponente kleiner als 0.5 ist, ansonsten ein `FALSE`.

Diese Wahrheitswerte können wir mit Hilfe der Funktion `as.numeric` in Zahlen umwandeln, wobei `T` gleich 1 und `F` gleich 0 wird.

```
as.numeric(u>0.5)
[1] 0 1
```

Nun addieren wir zu diesen Zahlen die 1 und indizieren mit dem Ergebnis den Vektor w , der aus den Zahlen 10 und 20 besteht:

```
1+as.numeric(u>0.5)
[1] 1 2

w <- c(10,20)

w[1+as.numeric(u>0.5)]
[1] 10 20
```

Wir können noch auf die Funktion `as.numeric` verzichten, da `T` und `F` bei einer numerischen Operation automatisch in 1 und 0 umgewandelt werden.

```
w[1+(u<0.5)]
[1] 10 20
```

Wir erhalten also folgende Befehlsfolge für die Simulation:

```
e <- rep(0,10000)

w <- c(10,20)

for (j in 1:10000){
  e[j] <- mean(w[1+(runif(2)>0.5)])
}
```

Wir führen die Simulation nun noch für $n=5$, $n=10$, $n=20$ und $n=30$ durch. Da wir nicht immer die gleiche Befehlsfolge eingeben wollen, schreiben wir uns eine kleine Funktion `simbsp1`.

Diese Funktion wird zwei Argumente haben:

- den Stichprobenumfang n (bisher $n=2$)
- die Anzahl der Simulationsläufe B (bisher $B=10000$)

Die Funktion wird dann definiert durch folgende Befehlsfolge:

```
simbsp1 <- function(n,b){
  e <- rep(0,b)
  w <- c(10,20)
  for (j in 1:b){
    e[j] <- mean(w[1+(runif(n)>0.5)])
  }
  e
}
```

Wir simulieren nun den Fall $n=5$:

```
e <- simbsp1(5,10000)
```

Wir bestimmen die absoluten Häufigkeiten mit Hilfe von `table`:

```
h <- table(e)
```

```
h
 10  12  14  16  18  20
320 1623 3107 3086 1543 321
```

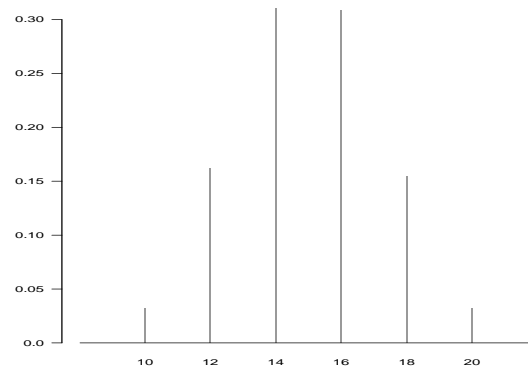
Wir bestimmen die relativen Häufigkeiten:

```
h <- h/sum(h)
```

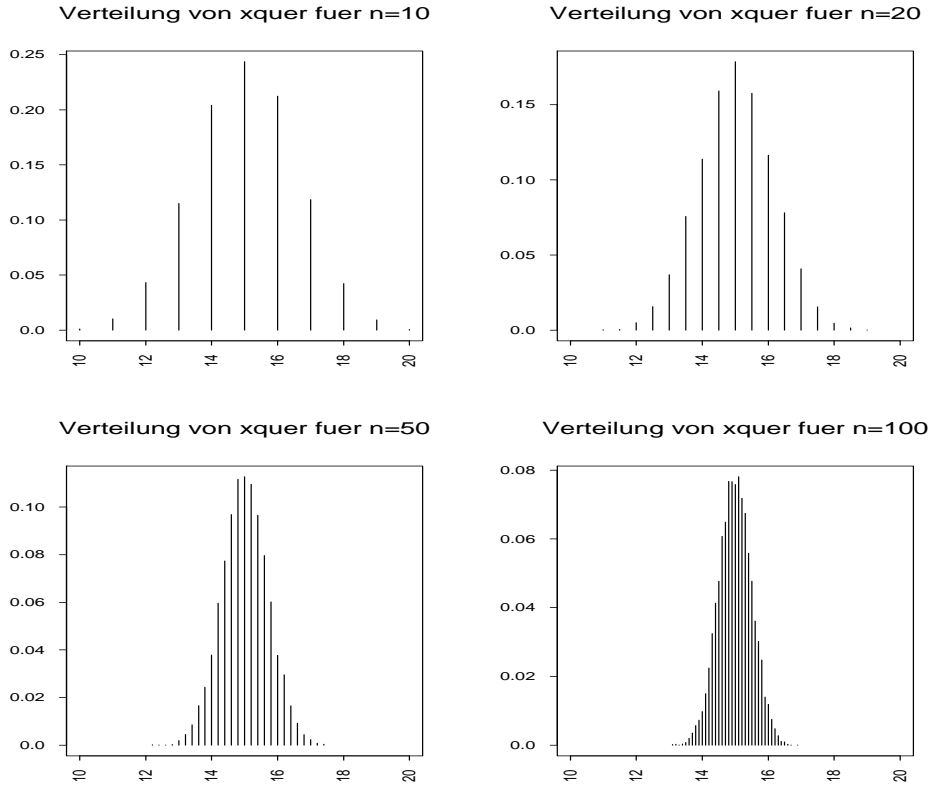
```
h
  10    12    14    16    18    20
0.032 0.1623 0.3107 0.3086 0.1543 0.0321
```

Wir zeichnen die geschätzte Wahrscheinlichkeitsfunktion:

```
barplot(h,space=50)
```



Nun simulieren wir noch die Fälle $n=10$, $n=20$, $n=50$ und $n=100$ und erstellen die Graphiken.



Wie wir sehen, konzentriert sich die Verteilung von \bar{X} mit wachsendem Stichprobenumfang immer stärker um den wahren Wert des Parameters. Dies hat zur Folge, daß Beobachtungen, die nicht in der Nähe des wahren Wertes des Parameters liegen, selten auftreten.

Ein beobachteter Wert von \bar{X} wird also ziemlich sicher in der Nähe des wahren Wertes des Parameters liegen, wenn der Stichprobenumfang nur groß genug ist.

Diese Eigenschaft einer Schätzfunktion kann man folgendermaßen formalisieren:

Definition 1.1.3 Eine Schätzfunktion $\hat{\theta}_n$ heißt **schwach konsistent** für den Parameter θ , wenn für jedes $\epsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P\left(|\hat{\theta}_n - \theta| < \epsilon\right) = 1$$

Der Nachweis der schwachen Konsistenz ist nicht sehr einfach.
 Eine andere Art der Konsistenz ist in der Regel aber leicht nachzuweisen.
 Hierzu benötigen wir den mittleren quadratischen Fehler.

Definition 1.1.4 *Der mittlere quadratische Fehler einer Schätzfunktion $\hat{\theta}_n$ bezüglich des Parameters θ ist definiert durch*

$$MSE(\hat{\theta}_n, \theta) = E \left[(\hat{\theta}_n - \theta)^2 \right]$$

Mit Hilfe des mittleren quadratischen Fehlers erhalten wir folgende Definition der Konsistenz:

Definition 1.1.5 *Eine Schätzfunktion $\hat{\theta}_n$ heißt eine im quadratischen Mittel konsistente Schätzfunktion für den Parameter θ , wenn gilt:*

$$\lim_{n \rightarrow \infty} MSE(\hat{\theta}_n, \theta) = 0$$

Die Konsistenz im quadratischen Mittel kann man sehr schön interpretieren, wenn man den MSE folgendermaßen umformt:

$$MSE(\hat{\theta}_n, \theta) = Var(\hat{\theta}_n) + \left[(E(\hat{\theta}_n) - \theta)^2 \right]$$

Hierzu benötigen wir folgende Aussage

Satz 1.1.1 *Sei W eine Zufallsvariable und a ein Skalar. Dann gilt*

$$E \left[(W - a)^2 \right] = Var(W) + (E(W) - a)^2$$

Beweis:

$$\begin{aligned} E \left[(W - a)^2 \right] &= E \left[(W - E(W) + E(W) - a)^2 \right] \\ &= E \left[(W - E(W))^2 + (E(W) - a)^2 + \right. \\ &\quad \left. + 2(W - E(W))(E(W) - a) \right] \\ &= E \left[(W - E(W))^2 \right] + E \left[(E(W) - a)^2 \right] + \\ &\quad + 2 E \left[(W - E(W))(E(W) - a) \right] \\ &= Var(W) + \left[(E(W) - a)^2 \right] + \\ &\quad + (E(W) - a)(E(W) - E(W)) \\ &= Var(W) + \left[(E(W) - a)^2 \right] \end{aligned}$$

Mit

$$W = \hat{\theta}_n$$

und

$$a = \theta$$

gilt

$$MSE(\hat{\theta}_n, \theta) = Var(\hat{\theta}_n) + \left[(E(\hat{\theta}_n) - \theta)^2 \right]$$

Der MSE ist also die Summe aus folgenden nichtnegativen Größen:

- der Varianz

$$Var(\hat{\theta}_n)$$

der Schätzfunktion

- dem quadrierten **Bias**

$$bias(\hat{\theta}_n, \theta) = E(\hat{\theta}_n) - \theta$$

der Schätzfunktion.

Bei einer im quadratischen Mittel konsistenten Schätzfunktion verschwindet also mit wachsendem n die Varianz und der Bias.

Die Verteilung von $\hat{\theta}_n$ konzentriert sich also mit wachsendem n immer mehr um θ .

Für \bar{X} gilt:

$$E(\bar{X}) = \mu$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Die letzte Eigenschaft kann man folgendermaßen zeigen:

Sind die Zufallsvariablen X_1, \dots, X_n unabhängig, so gilt:

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

Somit gilt

$$\begin{aligned}
 \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n^2} n \sigma^2 \\
 &= \frac{\sigma^2}{n}
 \end{aligned}$$

Also ist \bar{X} eine im quadratischen Mittel konsistente Schätzfunktion für $E(X)$. Man kann zeigen, daß eine im quadratischen Mittel konsistente Schätzfunktion auch schwach konsistent ist.

Hierzu benötigt man die Ungleichung von Markov.

Satz 1.1.2 *Sei Y eine nichtnegative Zufallsvariable und a eine positive reelle Zahl.*

Dann gilt

$$P(Y \geq a) \leq \frac{E(Y)}{a}$$

Beweis:

Wir zeigen den stetigen Fall

$$\begin{aligned}
 E(Y) &= \int_0^{\infty} y f_Y(y) dy \\
 &= \int_0^a y f_Y(y) dy + \int_a^{\infty} y f_Y(y) dy \\
 &\geq \int_a^{\infty} y f_Y(y) dy \\
 &\geq a \int_a^{\infty} f_Y(y) dy \\
 &\geq a P(Y \geq a)
 \end{aligned}$$

Sei nun $\hat{\theta}_n$ eine im quadratischen Mittel konsistente Schätzfunktion des Parameters θ .

Wir setzen

$$Y = (\hat{\theta}_n - \theta)^2$$

und

$$a = \epsilon^2$$

Dann gilt aufgrund von Satz 1.1.2:

$$P((\hat{\theta}_n - \theta)^2 \geq \epsilon^2) \leq \frac{E[(\hat{\theta}_n - \theta)^2]}{\epsilon^2}$$

und somit

$$P(|\hat{\theta}_n - \theta| \geq \epsilon) \leq \frac{MSE(\hat{\theta}_n, \theta)}{\epsilon^2}$$

Also gilt:

$$P(|\hat{\theta}_n - \theta| < \epsilon) \geq 1 - \frac{MSE(\hat{\theta}_n, \theta)}{\epsilon^2}$$

Da $\hat{\theta}_n$ eine im quadratischen Mittel konsistente Schätzfunktion des Parameters θ ist, gilt

$$\lim_{n \rightarrow \infty} MSE(\hat{\theta}_n, \theta) = 0$$

und somit folgt:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \epsilon) \geq 1 - \lim_{n \rightarrow \infty} \frac{MSE(\hat{\theta}_n, \theta)}{\epsilon^2}$$

Da ein Wahrscheinlichkeitsmaß kleiner gleich 1 ist, folgt

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1$$

Eine im quadratischen Mittel konsistente Schätzfunktion ist also auch konsistent.

Das arithmetische Mittel \bar{X} besitzt bei Normalverteilung aber noch folgende Eigenschaft als Schätzer von μ :

Unter allen erwartungstreuen Schätzfunktionen von μ ist \bar{X} die Schätzfunktion mit der kleinsten Varianz.

Man sagt auch, daß \bar{X} eine effiziente Schätzfunktion von μ bei Normalverteilung ist.

Bevor wir zeigen, warum dies der Fall ist, wollen wir mit Hilfe einer kleinen Simulationsstudie demonstrieren, daß bei Normalverteilung \bar{X} eine kleinere Varianz hat als der Median M .

Der Median M ist für eine Stichprobe X_1, \dots, X_n folgendermaßen definiert:

$$M = \begin{cases} X_{(0.5(n+1))} & \text{falls } n \text{ ungerade ist} \\ 0.5 (X_{(0.5n)} + X_{(0.5n+1)}) & \text{falls } n \text{ gerade ist} \end{cases}$$

Dabei ist $X_{(1)}, \dots, X_{(n)}$ die geordnete Stichprobe.

Man kann zeigen, daß der Median eine erwartungstreue Schätzfunktion für μ bei Normalverteilung ist.

Die Varianz des Medians ist nicht leicht zu bestimmen. Wir helfen uns mit einer Simulationsstudie.

Wir wählen den Stichprobenumfang $n=20$ und ziehen $b=10000$ Stichproben vom Umfang $n=20$ aus einer Standardnormalverteilung.

Hierzu verwenden wir die Funktion `rnorm`, deren Argument die Anzahl der standardnormalverteilten Zufallsvariablen ist.

Der Aufruf

```
rnorm(20)
```

liefert also 20 standardnormalverteilte Zufallszahlen.

Für jede dieser Stichproben bestimmen wir den Mittelwert und den Median.

```
mi <- rep(0,100000)

me <- rep(0,100000)

for (i in 1:10000){
    x <- rnorm(20)
    mi[i] <- mean(x)
    me[i] <- median(x)
}
```

Schließlich bestimmen wir die Varianz der Mittelwerte

```
var(mi)
[1] 0.004976833
```

und die Varianz der Mediane.

```
var(me)
[1] 0.007357627
```

Es gilt

```
var(me)/var(mi)
[1] 1.478375
```

Wir führen nun die Simulation für $n=50$ durch und erhalten folgendes Ergebnis:

```
var(me)/var(mi)
[1] 1.546751
```

Asymptotisch gilt bei Normalverteilung:

$$\frac{\text{Var}(M)}{\text{Var}(\bar{X})} = \frac{\pi}{2} = 1.57$$

Das arithmetische Mittel ist bei Normalverteilung aber nicht nur effizienter als der Median. Es ist unter allen erwartungstreuen Schätzfunktionen des Erwartungswerts am effizientesten.

Dies gilt aufgrund der Ungleichung von Rao-Cramer:

Satz 1.1.3 Seien X_1, \dots, X_n unabhängige, mit Dichtefunktion bzw. Wahrscheinlichkeitsfunktion $f_X(x)$ identisch verteilte Zufallsvariablen und $\hat{\theta}_n$ eine erwartungstreue Schätzfunktion für den Parameter θ .

Dann gilt

$$\text{Var}(\hat{\theta}_n) \geq \frac{1}{I(\theta)}$$

Dabei heißt $I(\theta)$ die Fisher-Information und ist folgendermaßen definiert:

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} l(\theta) \right)^2 \right]$$

Beweis:

Siehe Garthwaite, Jolliffe, Jones, S.12-13

Je größer die Fisher-Information ist, um so kleiner ist die Varianz der Schätzfunktion.

Nun können wir zeigen, daß \bar{X} eine effiziente Schätzfunktion von μ bei Normalverteilung ist.

Aus

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

folgt

$$\begin{aligned} \frac{\partial}{\partial \mu} l(\mu) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ &= \frac{n}{\sigma^2} (\bar{x} - \mu) \end{aligned}$$

Somit gilt

$$\begin{aligned} I(\mu) &= E \left[\left(\frac{\partial}{\partial \mu} l(\mu) \right)^2 \right] \\ &= E \left[\frac{n^2}{\sigma^4} (\bar{X} - \mu)^2 \right] \\ &= \frac{n^2}{\sigma^4} E \left[(\bar{X} - \mu)^2 \right] \\ &= \frac{n^2}{\sigma^4} \text{Var}(\bar{X}) \\ &= \frac{n^2}{\sigma^4} \frac{\sigma^2}{n} \\ &= \frac{n}{\sigma^2} \end{aligned}$$

Also gilt für jede erwartungstreue Schätzfunktion $\hat{\mu}$ von μ bei Normalverteilung:

$$\text{Var}(\hat{\mu}) \geq \frac{\sigma^2}{n}$$

Für \bar{X} gilt

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Also nimmt die Varianz von \bar{X} die untere Schranke an.

Da \bar{X} erwartungstreu für μ ist, ist \bar{X} also eine effiziente Schätzfunktion für μ bei Normalverteilung.

Dies hätte man auch anders zeigen können:

Satz 1.1.4 Die Varianz $\text{Var}(\hat{\theta}_n)$ einer erwartungstreuen Schätzfunktion $\hat{\theta}_n$ des Parameters θ nimmt genau dann die untere Schranke von Rao-Cramer an, wenn gilt

$$\frac{\partial}{\partial \theta} l(\theta) = I(\theta) (\hat{\theta}_n - \theta)$$

Beweis

Garthwaite, Jolliffe, Jones, S.14

Bei Normalverteilung gilt

$$\begin{aligned} \frac{\partial}{\partial \mu} l(\mu) &= \frac{n}{\sigma^2} (\bar{x} - \mu) \\ &= I(\mu) (\bar{x} - \mu) \end{aligned}$$

Wir betrachten nun noch ein weiteres Beispiel.

Beispiel 1.1.1 Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und identisch poissonverteilt mit Parameter λ . Es gilt also

$$f_{X_i}(x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

Also gilt

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

und es folgt

$$\begin{aligned} \frac{\delta}{\delta \lambda} l(\lambda) &= \frac{1}{\lambda} \sum_{i=1}^n x_i - n \\ &= \frac{n}{\lambda} (\bar{x} - \lambda) \end{aligned}$$

Also ist \bar{X} der M-L-Schätzer von λ bei Poissonverteilung.
Wegen

$$E(X) = \text{Var}(X) = \lambda$$

gilt

$$\begin{aligned} E(\bar{X}) &= \lambda \\ \text{Var}(\bar{X}) &= \frac{\lambda}{n} \end{aligned}$$

Also ist \bar{X} eine erwartungstreue und konsistente Schätzfunktion von λ .
Die Fisher-Information ist:

$$\begin{aligned} I(\lambda) &= E \left[\left(\frac{\partial}{\partial \lambda} l(\lambda) \right)^2 \right] \\ &= E \left[\left(\frac{n}{\lambda} (\bar{X} - \lambda) \right)^2 \right] \\ &= \frac{n^2}{\lambda^2} E \left[((\bar{X} - \lambda))^2 \right] \\ &= \frac{n^2}{\lambda^2} \text{Var}(\bar{X}) \\ &= \frac{n^2}{\lambda^2} \frac{\lambda}{n} \\ &= \frac{n}{\lambda} \end{aligned}$$

Also gilt für jede erwartungstreue Schätzfunktion $\hat{\lambda}$ von λ :

$$\text{Var}(\hat{\lambda}) \geq \frac{\lambda}{n}$$

Also ist \bar{X} auch effizient für λ .

Dies hätte man auch an der Faktorisierung der Likelihood erkennen können.

Wir betrachten nun noch ein Beispiel, das zeigt, warum es gerechtfertigt ist, eine Wahrscheinlichkeit mit Hilfe einer Simulation zu schätzen.

Beispiel 1.1.2 Sei p die Wahrscheinlichkeit eines interessierenden Ereignisses.

Ein Beispiel ist das zweimalige Ziehen mit Zurücklegen aus einer Urne, die 10 Kugeln enthält, von denen fünf 10 g und fünf 20 g wiegen.

Wir wissen

$$\begin{aligned} P(\bar{X} = 10) &= P(X_1 = 10, X_2 = 10) \\ &= P(X_1 = 10) P(X_2 = 10) \\ &= 0.5 \cdot 0.5 \\ &= 0.25 \end{aligned}$$

Nun wollen wir diese Wahrscheinlichkeit mit Hilfe einer Simulation schätzen. Hierzu erzeugen wir n -mal Zufallsstichproben vom Umfang 2 aus der Urne und betrachten folgende Zufallsvariablen:

$$U_i = \begin{cases} 1 & \text{wenn } \bar{X} = 10 \\ 0 & \text{sonst} \end{cases}$$

Es gilt

$$P(U_i = 1) = P(\bar{X} = 10) = p$$

Als Schätzfunktion für p wählen wir

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

Es gilt

$$\begin{aligned} E(\hat{p}) &= E \left[\frac{1}{n} \sum_{i=1}^n U_i \right] \\ &= \frac{1}{n} E \left[\sum_{i=1}^n U_i \right] \\ &= \frac{1}{n} \sum_{i=1}^n E(U_i) \\ &= \frac{1}{n} \sum_{i=1}^n p \\ &= p \end{aligned}$$

und

$$\begin{aligned} \text{Var}(\hat{p}) &= \text{Var} \left[\frac{1}{n} \sum_{i=1}^n U_i \right] \\ &= \frac{1}{n^2} \text{Var} \left[\sum_{i=1}^n U_i \right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(U_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n p(1-p) \\ &= \frac{p(1-p)}{n} \end{aligned}$$

Also ist \hat{p} eine erwartungstreue und konsistente Schätzfunktion für p .
Sie ist sogar varianzminimal, denn es gilt:

$$\frac{\partial}{\partial p} l(p) = \frac{n}{p(1-p)} (\bar{U} - p)$$

Kehren wir nun noch einmal zur M-L-Methode zurück.

Wir betrachten nun die Exponentialverteilung mit Dichtefunktion

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Es soll des M-L-Schätzer des Parameters λ bestimmt werden.

Die Log-Likelihood lautet:

$$\begin{aligned} l(\lambda) &= n \ln \lambda - \lambda \sum_{i=1}^n X_i \\ &= n \ln \lambda - \lambda n \bar{x} \end{aligned}$$

Die erste Ableitung ist:

$$\frac{\delta}{\delta \lambda} l(\lambda) = \frac{n}{\lambda} - n \bar{x}$$

Also ist der M-L-Schätzer von λ gegeben durch

$$\hat{\lambda}_{ML} = \frac{1}{\bar{X}}$$

Schauen wir uns die Verteilung von $\frac{1}{\bar{X}}$ bei Exponentialverteilung mit Hilfe einer Simulation einmal an.

Wir benötigen eine Funktion zur Erzeugung von Zufallszahlen, die mit dem Parameter λ exponentialverteilt sind.

Diese heißt in R `rexp` und hat als Argumente die gewünschte Anzahl n von Zufallszahlen und den Wert des Parameters λ .

Der Aufruf

```
rexp(5,1)
[1] 1.23626751 2.64235588 2.32868982 2.72578837 0.09081808
```

liefert 5 Zufallszahlen aus einer Exponentialverteilung mit Parameter $\lambda = 1$. Wir schauen uns die Verteilung von $\frac{1}{\bar{X}}$ bei Exponentialverteilung mit $\lambda = 1$ und $n = 5$ an.

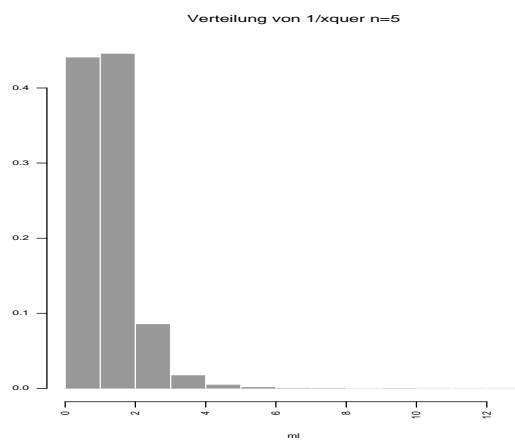
Die folgende Befehlsfolge leistet dies für 10000 Wiederholungen.

```
m1 <- rep(0,10000)
for (i in 1:10000){
  x <- rexp(5,1)
  m1[i] <- 1/mean(x)
}
```

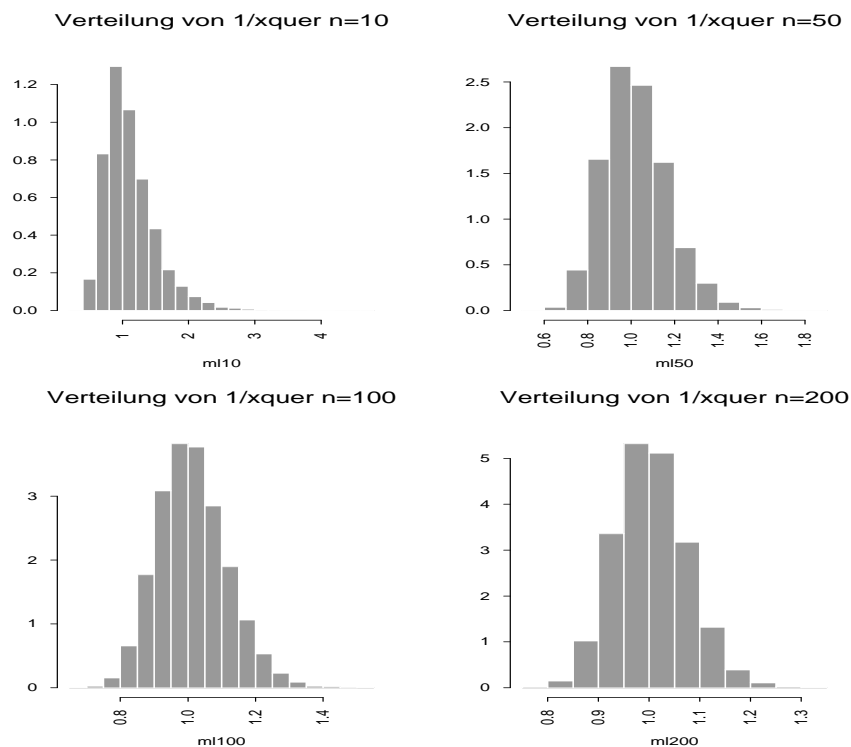
Der Mittelwert von $m1$ liegt beträchtlich über 1, so daß der Schätzer verzerrt zu sein scheint:

```
mean(m1)
[1] 1.257384
```

Das Histogramm deutet darauf hin, daß die Verteilung von $\frac{1}{\bar{X}}$ schief zu sein scheint.



Die nachfolgenden Bilder zeigen die Histogramme für $n = 10$, $n = 50$, $n = 100$ und $n = 200$:



Wie die Bilder zeigen, konzentriert sich die Verteilung von $\frac{1}{\bar{X}}$ mit wachsendem Stichprobenumfang n immer stärker um den Wert 1.

Dies zeigt auch die Folge der Mittelwerte von m_l in der nachstehenden Tabelle:

n	Mittelwert von m_l
10	1.110658
50	1.020347
100	1.012495
200	1.004606

Man kann zeigen, daß bei Exponentialverteilung mit Parameter λ gilt

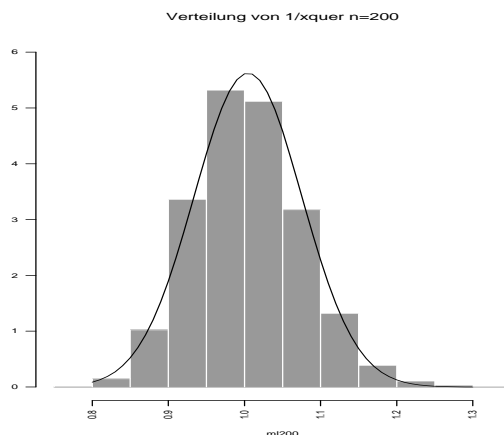
$$E\left[\frac{1}{\bar{X}}\right] = \frac{n}{n-1} \lambda$$

(Siehe dazu Mood, Graybill, Boes(1974), S.328)

Somit ist $\frac{1}{\bar{X}}$ eine asymptotisch erwartungstreue Schätzfunktion von λ .

Außerdem sieht die Verteilung von $\frac{1}{\bar{X}}$ mit wachsendem n immer symmetrischer und normalverteilter aus.

In das Histogramm für $n=200$ ist die Dichtefunktion der Normalverteilung mit den Parametern `mean(m1)` und `sqrt(var(m1))` eingezeichnet.



Man kann zeigen, daß unter bestimmten Regularitätsbedingungen der M-L-Schätzer $\hat{\theta}_{ML}$ des Parameters θ asymptotisch, d.h. für großes n , normalverteilt ist mit den Parametern $\mu = \theta$ und $\sigma^2 = I(\theta)^{-1}$, wobei $I(\theta)$ die Fisher-Information ist.

Die asymptotische Varianz ist also gerade die Inverse der Fisher-Information. Unter den Regularitätsbedingungen gilt

$$I(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} l(\theta) \right]$$

Man kann zeigen, daß bei der Exponentialverteilung die Regularitätsbedingungen erfüllt sind.

Nun gilt

$$\frac{\partial^2}{\partial \lambda^2} l(\lambda) = -\frac{n}{\lambda^2}$$

Also gilt

$$I(\lambda) = \frac{n}{\lambda^2}$$

Somit gilt für die Varianz des M-L-Schätzers approximativ:

$$\text{Var}(\hat{\lambda}_{ML}) = \frac{\lambda^2}{n}$$

Bei $n=100$ erwarten wir 0.01.

Wir simulieren und erhalten:

```
ml <- rep(0,10000)
for (i in 1:10000){
  x <- rexp(100,1)
  ml[i] <- 1/mean(x)
}
```

```
var(ml)
[1] 0.01077577
```

Nicht immer erhält man den Maximum-Likelihood-Schätzer durch differenzieren:

Wir betrachten die Gleichverteilung auf dem geschlossenen Intervall $[0, b]$ mit Dichtefunktion

$$f_X(x) = \begin{cases} \frac{1}{b} & \text{für } 0 \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

und Verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x}{b} & \text{für } 0 \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$$

Gesucht ist der M-L-Schätzer von b aus einer Zufallsstichprobe vom Umfang n .

Die Likelihoodfunktion lautet

$$L(b) = \begin{cases} \frac{1}{b^n} & \text{falls alle } x_i \leq b \\ 0 & \text{sonst} \end{cases}$$

Das Maximum wird hier am Rand angenommen und zwar gilt

$$\hat{b}_{ML} = \max\{X_1, \dots, X_n\}$$

Schauen wir uns diesen Schätzer noch genauer an.

Für die Verteilungsfunktion von \hat{b}_{ML} gilt:

$$\begin{aligned}
 F_{\hat{b}_{ML}}(b) &= P(\hat{b}_{ML} \leq b) \\
 &= P(\max\{X_1, \dots, X_n\} \leq b) \\
 &= P(\text{alle } X_i \leq b) \\
 &= P(X_1 \leq b, \dots, X_n \leq b) \\
 &= P(X_1 \leq b) \cdots P(X_n \leq b) \\
 &= F_X(b)^n
 \end{aligned}$$

Somit gilt

$$F_{\hat{b}_{ML}}(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x^n}{b^n} & \text{für } 0 \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$$

Die Dichtefunktion lautet

$$f_{\hat{b}_{ML}}(x) = \begin{cases} \frac{n x^{n-1}}{b^n} & \text{für } 0 \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Der M-L-Schätzer ist nicht erwartungstreu.

Es gilt:

$$\begin{aligned}
 E(\hat{b}_{ML}) &= \int_0^b x \frac{n x^{n-1}}{b^n} dx \\
 &= \frac{n}{b^n} \int_0^b x^n dx \\
 &= \frac{n}{b^n} \frac{b^{n+1}}{n+1} \\
 &= \frac{n}{n+1} b
 \end{aligned}$$

Wie wir sehen, ist der M-L-Schätzer nicht erwartungstreu, jedoch asymptotisch erwartungstreu.

Um die Konsistenz zu überprüfen, benötigen wir die Varianz.
Es gilt:

$$\begin{aligned} E(\hat{b}_{ML}^2) &= \int_0^b x^2 \frac{n x^{n-1}}{b^n} dx \\ &= \frac{n}{b^n} \int_0^b x^{n+1} dx \\ &= \frac{n}{b^n} \frac{b^{n+2}}{n+2} \\ &= \frac{n}{n+2} b^2 \end{aligned}$$

Somit folgt

$$\begin{aligned} Var(\hat{b}_{ML}^2) &= \frac{n}{n+2} b^2 - \left(\frac{n}{n+1} b\right)^2 \\ &= \frac{n}{(n+2)(n+1)^2} \end{aligned}$$

Offensichtlich ist der M-L-Schätzer konsistent.

1.1.3 Testen

Der t-Test

Kommen wir nun zur Überprüfung der Hypothese, daß die Rechtecke nach dem goldenen Schnitt gefertigt wurden. Dies bedeutet, daß das Verhältnis von Breite zu Länge 0.618 beträgt.

Wir können das Testproblem also folgendermaßen in der Sprache der mathematischen Statistik formulieren:

Das Seitenverhältnis der Rechtecke der Schoschonen ist eine Zufallsvariable X mit stetiger Verteilungsfunktion $F_X(x)$.

Anhand einer Zufallsstichprobe X_1, \dots, X_n aus $F_X(x)$ wollen wir überprüfen, ob der Erwartungswert von X gleich 0.618 ist.

Im Rahmen der parametrischen Statistik muß man nun $F_X(x)$ bis auf den Wert eines oder mehrerer Parameter spezifizieren, um das Problem lösen zu können.

Die klassische Annahme ist, daß $F_X(x)$ eine Normalverteilung ist.

Wir können also die obige Annahme noch spezieller formulieren:

Die Beobachtungen x_1, \dots, x_n sind Realisationen von unabhängigen Zufallsvariablen X_1, \dots, X_n , die identisch mit den Parametern μ und σ^2 normalverteilt sind.

Es soll getestet werden:

$$H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0$$

mit $\mu_0 = 0.618$.

Für einen Test benötigt man eine geeignete Teststatistik.

Da der Mittelwert \bar{X} eine erwartungstreue und konsistente Schätzung von μ ist, liegt es nahe die Entscheidung auf der Basis von \bar{X} zu fällen.

Ist μ_0 der wahre Wert von μ , so sollte \bar{X} Werte in der Nähe von μ_0 annehmen. Weicht \bar{X} stark von μ_0 ab, so spricht dies dagegen, daß μ_0 der wahre Wert von μ ist.

Wie soll man nun die Grenze wählen, ab der man H_0 ablehnt?

Bei der Wahl der Grenze unterstellen wir, daß die Nullhypothese H_0 zutrifft. Wir gehen also im Beispiel davon aus, daß die Rechtecke nach dem Goldenen Schnitt gefertigt wurden. In diesem Fall erwarten wir, daß wir Werte von \bar{X} beobachten, die in der Nähe von μ_0 liegen.

Es können sich natürlich auch weit von μ_0 entfernte Werte von \bar{X} realisieren, wenn μ_0 der wahre Wert von μ ist. In diesem Fall gehen wir aber davon aus, daß die starke Abweichung des arithmetischen Mittels von μ_0 nicht dadurch zustande gekommen ist, daß wir eine der extremen Stichproben beobachtet

haben, sondern daß ein anderer Wert als μ_0 der wahre Wert von μ ist. Wir würden uns daher gegen H_0 entscheiden.

Ist nun aber μ_0 der wahre Wert von μ , so würden wir bei unserer Entscheidung einen Fehler machen. Man nennt die fälschliche Entscheidung für H_1 einen Fehler 1. Art. Die Wahrscheinlichkeit α dieses Fehlers geben wir vor. In der Regel wählt man $\alpha = 0.05$ oder $\alpha = 0.01$.

Nun kann man natürlich fragen, warum man nicht gleich $\alpha = 0$ wählt. Dies würde bedeuten, daß man sich immer für H_0 entscheidet.

Dagegen spricht folgendes:

Bei dieser Festlegung bräuchte man überhaupt keine Daten erheben, da man sich ja immer für H_0 entscheidet. Man würde sich in diesem Fall fälschlicherweise sehr oft für H_0 entscheiden.

Diese Fehlentscheidung bezeichnet man als Fehler 2. Art.

Die folgende Tabelle stellt die Situation dar:

Realität	H_0 trifft zu	H_1 trifft zu
Entscheidung		
für H_0	richtige Entscheidung	Fehler 2.Art
für H_1	Fehler 1.Art	richtige Entscheidung

Fassen wir die obigen Überlegungen noch einmal zusammen:

Um zu überprüfen, ob μ_0 der wahre Wert von μ ist, wird einer normalverteilten Grundgesamtheit eine Zufallsstichprobe vom Umfang n entnommen.

Wir bestimmen dann die Teststatistik \bar{X} und vergleichen sie mit μ_0 . Weicht \bar{X} zu stark von μ_0 ab, so lehnen wir die Nullhypothese H_0 ab.

Die Werte von \bar{X} , für die man die Nullhypothese ablehnt, nennt man den Ablehnbereich.

Zur Bestimmung des Ablehnbereichs benötigt man die Verteilung von \bar{X} , wenn die Nullhypothese zutrifft. Man spricht auch von der Verteilung von \bar{X} unter H_0 .

Schauen wir uns dies für das Beispiel an:

Es soll getestet werden:

$$H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0$$

mit $\mu_0 = 0.618$.

Sind nun die Zufallsvariablen X_1, \dots, X_n unabhängig und identisch mit den Parametern μ_0 und σ^2 normalverteilt, so ist \bar{X} mit den Parametern μ_0 und $\frac{\sigma^2}{n}$ normalverteilt.

Also ist

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

standardnormalverteilt.

Ist die Varianz σ^2 bekannt, so erhalten wir als Ablehnbereich:

$$\bar{X} < \mu_0 - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

oder

$$\bar{X} > \mu_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

und entscheiden uns gegen H_0 . Dabei ist $z_{1-\alpha/2}$ das $1-\alpha/2$ -Quantil der Standardnormalverteilung.

In der Praxis ist σ^2 in der Regel unbekannt.

Es liegt nahe, σ^2 durch die erwartungstreue Schätzfunktion S^2 zu schätzen und dann in der obigen Teststatistik σ^2 durch $S = \sqrt{S^2}$ zu ersetzen.

Man erhält dann die Teststatistik:

$$t = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$$

Das Ersetzen von σ durch S führt dazu, daß die Teststatistik t nicht standardnormalverteilt ist, wenn die Nullhypothese zutrifft.

Dies zeigt folgende kleine Simulationsstudie:

Wir wollen testen:

$$H_0 : \mu = 0 \quad \text{gegen} \quad H_1 : \mu \neq 0$$

Wir erzeugen $B = 10000$ Stichproben vom Umfang $n = 6$ aus einer Normalverteilung mit den Parametern $\mu = 0$ und $\sigma = 1$ und bestimmen für jede den Wert der Teststatistik t .

```
B <- 10000
tv <- rep(0,B)
for (i in 1:B){
    zz <- rnorm(6)
    tv[i] <- (sqrt(length(zz))*mean(zz))/sqrt(var(zz))
}
```

Der Mittelwert der Teststatistiken ist in der Nähe von 0.

```
mean(tv)
[1] 0.003613457
```

Wir sehen, daß die Varianz aber sehr viel größer als 1 ist.

```
var(tv)
[1] 1.59729
```

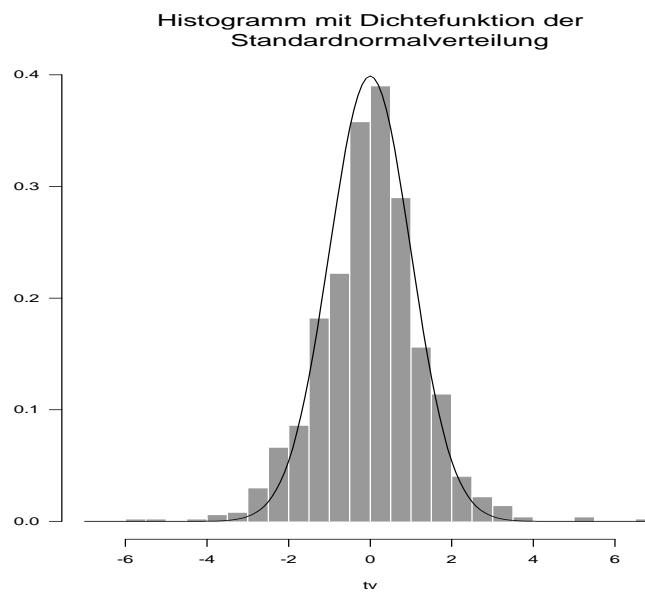
Die Teststatistik t streut also stärker als die Standardnormalverteilung. Die zusätzliche Streuung liegt daran, daß die Varianz σ^2 im Zähler der Teststatistik geschätzt werden muß.

Dies zeigt auch das Histogramm mit der überlagerten Dichtefunktion der Standardnormalverteilung.

Wir erstellen das Histogramm der Werte von t mit Hilfe der Funktion `hist`:

```
hist(tv,breaks=seq(-7,7,by=0.5),prob=T,ylim=c(0,0.4),main=
  "Histogramm mit Dichtefunktion der Standardnormalverteilung")

x <- seq(-7,7,by=0.1)
lines(x,dnorm(x))
```



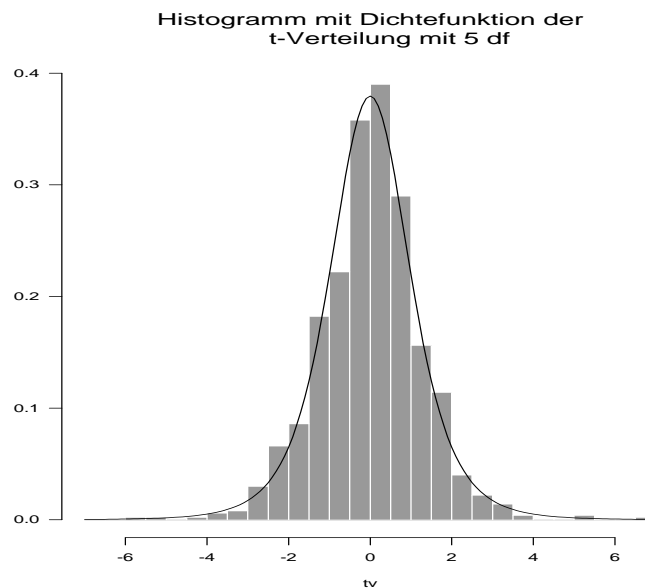
Die simulierte Verteilung hat mehr Wahrscheinlichkeitsmasse an den Rändern als die Standardnormalverteilung.

Man kann zeigen, daß t unter H_0 t -verteilt ist mit $n-1$ Freiheitsgraden. (siehe dazu Mood, Graybill, Boes, S. 250)

Die nachfolgende Graphik zeigt das Histogramm mit der überlagerten Dichtefunktion der t-Verteilung mit $n - 1 = 5$ Freiheitsgraden.

```
hist(tv,breaks=seq(-7,7,by=0.5),prob=T,ylim=c(0,0.4),main="
  Histogramm mit Dichtefunktion der t-Verteilung mit 5 df")

x <- seq(-7,7,by=0.1)
lines(x,dt(x,5))
```



Wir sehen, daß die Anpassung bedeutend besser ist.

Für das Testproblem

$$H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0$$

liegt es nun nahe, die Nullhypothese abzulehnen, wenn die Teststatistik t zu groß oder zu klein ist:

Lehne H_0 ab, wenn gilt

$$|t| > t_{n-1;1-\alpha/2}$$

wobei $t_{n-1;1-\alpha/2}$ das $1 - \alpha/2$ -Quantil der t-Verteilung mit $n-1$ Freiheitsgraden ist.

Für das Datenbeispiel gilt

- $\bar{x} = 0.6605$
- $s = 0.0925$
- $t = 2.0545$
- $t_{19;0.975} = 2.093$.

Wir lehnen also H_0 zum Signifikanzniveau $\alpha = 0.05$ nicht ab.

In R gibt es eine Funktion `t.test`, die man mit dem Datensatz und dem hypothetischen Wert des Parameters aufruft:

```
t.test(shosho,mu=0.618)
```

```
One-sample t-Test
```

```
data: shosho
t = 2.0545, df = 19, p-value = 0.0539
alternative hypothesis: true mean is not equal to 0.618
95 percent confidence interval:
 0.6172036 0.7037964
```

```
sample estimates:
mean of x
 0.6605
```

Diese Funktion liefert eine Reihe von Informationen:

- der Wert `t` der Teststatistik, der in unserem Fall 2.0545 beträgt,
- die Anzahl der Freiheitsgrade `df`, die hier 19 beträgt,
- die **Überschreitungswahrscheinlichkeit** (p-value), die 0.0539 beträgt.

Schauen wir uns die Überschreitungswahrscheinlichkeit genauer an.

Dies ist die Wahrscheinlichkeit, daß die Teststatistik unter H_0 Werte annimmt, die noch extremer als der beobachtete sind.

In unserem Fall bedeutet dies $2P(t > 2.0545)$.

Diese können wir mit R folgendermaßen berechnen:

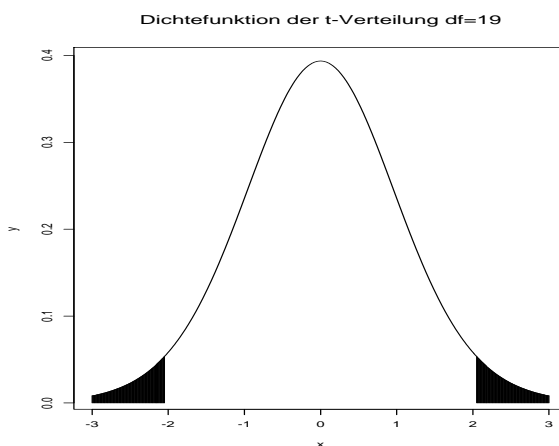
```
2*(1-pt(2.0545, 19))
[1] 0.05394381
```

Große Werte der Überschreitungswahrscheinlichkeit sprechen für H_0 , kleine gegen H_0 .

Ist die Überschreitungswahrscheinlichkeit kleiner als das vorgegebene Signifikanzniveau α , dann lehnen wir H_0 ab. Ansonsten lehnen wir H_0 nicht ab.

Die Funktion `pt(x, df)` berechnet die Verteilungsfunktion der t-Verteilung mit `df` Freiheitsgraden an der Stelle `x`.

Die folgende Graphik veranschaulicht die Überschreitungswahrscheinlichkeit. Der schraffierte Teil unter der Dichtefunktion entspricht der Überschreitungswahrscheinlichkeit.



Außerdem liefert die Funktion `t.test` noch ein Konfidenzintervall für den Erwartungswert μ bei Normalverteilung mit unbekannter Varianz zum Konfidenzniveau $1 - \alpha = 0.95$.

Für die Schoschonendaten ist dies:

[0.6172036, 0.7037964].

Mit Wahrscheinlichkeit 0.95 überdeckt dieses Intervall den wahren Wert von μ .

Schauen wir uns noch ein anderes Testproblem an.

Beispiel 1.1.3 *Das Beispiel stammt ursprünglich von Fisher und wurde von Neyman modifiziert.*

Eine englische Dame behauptet, daß sie erkennt, ob bei einer Tasse Tee mit Milch zuerst die Milch oder zuerst der Tee eingegossen wurde.

Wir wollen überprüfen, ob die Behauptung der Dame zutrifft.

Hierzu formulieren wir das Gegenteil der Behauptung der Dame als Hypothese und die Behauptung der Dame als Gegenhypothese.

Wir erhalten also:

*Hypothese: Die Dame rät
Gegenhypothese: Die Dame rät nicht*

Wir schreiben also

H_0 : Die Dame rät
 H_1 : Die Dame rät nicht

Versuchen wir nun auch die Hypothese und die Gegenhypothese in die Sprache der Statistik zu übersetzen. In der Statistik formuliert man Hypothesen in der Regel über Parameter.

Sei R das Ereignis, daß die Dame die richtige Wahl trifft, und \bar{R} das Ereignis, daß sie die falsche Wahl trifft.

Außerdem sei $p = P(R)$.

Rät die Dame, so ist $p = 0.5$, rät sie nicht, so gilt $p > 0.5$.

Die Hypothese und Gegenhypothese lauten also

H_0 : $p = 0.5$
 H_1 : $p > 0.5$

Um die Hypothesen zu überprüfen, beobachtet man den Sachverhalt. Es liegt nahe, der Dame eine Tasse zu reichen, in die ohne ihr Wissen zuerst die Milch gefüllt wurde. Sie soll dann entscheiden, welche Situation vorliegt. Es ist aber sicherlich fairer, ihr von vornherein die Möglichkeit zu geben, zu vergleichen. Wir reichen ihr also zwei Tassen, wobei in die eine zuerst die Milch und in die andere zuerst der Tee gefüllt wurde.

Auf der Basis der Beobachtung fällen wir nun die Entscheidung.

Man nennt die Beobachtung eine Teststatistik.

Folgende Entscheidungsregel liegt nahe:

Entscheidung für H_0 , wenn die Dame die Tassen falsch zugeordnet hat.

Entscheidung für H_1 , wenn die Dame die Tassen richtig zugeordnet hat.

Die Entscheidung ist fehlerbehaftet.

Wir können zwei Fehler begehen.

Entscheiden wir uns für H_1 , obwohl H_0 zutrifft, so begehen wir einen Fehler 1.Art.

In unserem Beispiel heißt dies, daß wir zu der Entscheidung kommen, daß die Dame differenzieren kann, obwohl sie in Wirklichkeit geraten hat.

Entscheiden wir uns für H_0 , obwohl H_1 zutrifft, so begehen wir einen Fehler 2.Art.

In unserem Beispiel heißt dies, daß wir zu der Entscheidung kommen, daß die Dame geraten hat, obwohl sie in Wirklichkeit differenzieren kann.

Die Wahrscheinlichkeit des Fehlers 1.Art ist

$$\alpha = P(\text{Entscheidung für } H_1 | H_0 \text{ trifft zu})$$

In unserem Beispiel gilt

$$\alpha = P_{H_0}(R) = 0.5$$

Die Wahrscheinlichkeit des Fehlers 2.Art ist

$$\beta = P(\text{Entscheidung für } H_0 | H_1 \text{ trifft zu})$$

In unserem Beispiel gilt

$$\beta = P_{H_1}(\bar{R})$$

Diese Wahrscheinlichkeit können wir nicht angeben, da sie davon abhängt, wie gut die Dame differenzieren kann.

Wir haben also nur den Fehler 1.Art unter Kontrolle. Dies ist die übliche Situation beim statistischen Test. Da wir die Wahrscheinlichkeit des Fehlers 1.Art unter Kontrolle haben, geben wir diese vor. In der Regel wählt man $\alpha = 0.05$. Hierdurch ist sichergestellt, daß man sich ziemlich sicher sein kann, wenn man sich für H_1 entscheidet, da die Wahrscheinlichkeit für einen Fehler klein ist. Nun ist auch klar, warum wir die Behauptung der Dame als Gegenhypothese formuliert haben.

In unserem Beispiel beträgt die Wahrscheinlichkeit für den Fehler 1.Art 0.5. Sie ist viel zu groß .

Wir können sie verkleinern, indem wir das Experiment mehrmals wiederholen.

Nehmen wir an, wir führen 6 Versuche durch und erhalten das Ergebnis

RRRRR \bar{R} .

Spricht dies für oder gegen die Fähigkeit der Dame zu differenzieren?

Wir brauchen eine geeignete Teststatistik. Es liegt nahe, die Anzahl S der Fälle zu wählen, bei denen die Dame sich richtig entschieden hat.

In unserem Fall beträgt sie 5.

Ist diese sehr groß, so würden wir sagen, daß die Dame differenzieren kann.

Was heißt groß?

Dies hängt von dem vorgegebenen Wert von α ab. Wir müssen den Wert, ab dem wir uns für die Gegenhypothese entscheiden so wählen, daß die Wahrscheinlichkeit für den Fehler 1. Art gleich α ist.

Hierzu benötigen wir die Verteilung der Teststatistik, wenn H_0 zutrifft.

Man spricht auch von der Verteilung der Teststatistik unter H_0 .

Hier kommt in unserem Beispiel die Binomialverteilung ins Spiel.

Wir haben sechsmal einen Bernoullivorgang beobachtet.

Wenn wir der Dame nicht nach jedem Versuch sagen, ob ihre Entscheidung richtig oder falsch ist, lernt sie bei den Versuchen nichts dazu.

Also sind die einzelnen Versuche unabhängig.

Außerdem bleibt $P(R) = 0.5$ konstant.

Wir beobachten also einen Bernoulliprozeß der Länge 6 mit Erfolgswahrscheinlichkeit $p = 0.5$.

Dann ist die Anzahl der Erfolge mit den Parametern $n = 6$ und $p = 0.5$ binomialverteilt.

Es gilt also

$$P(S = s) = \binom{6}{s} 0.5^6 \quad \text{für } s = 0, 1, \dots, 6.$$

Dies ist die Verteilung von S unter H_0 .

Wir haben gesagt, daß wir H_0 ablehnen, wenn S zu groß ist.

Wir bestimmen für das Beispiel die Wahrscheinlichkeit, den beobachteten Wert und noch extremere zu beobachten.

$$P_{H_0}(S \geq 5) = \left[\binom{6}{5} + \binom{6}{6} \right] 0.5^6 = 0.109375$$

Wir nennen diese Wahrscheinlichkeit Überschreitungswahrscheinlichkeit.

Wir können die Testentscheidung mit Hilfe der Überschreitungswahrscheinlichkeit durchführen.

Ist sie größer als $\alpha = 0.05$, so lehnen wir H_0 nicht ab.

Ist sie kleiner als $\alpha = 0.05$, so lehnen wir H_0 ab.

Wie gut ist die Entscheidungsregel?

Wir messen die Güte der Entscheidungsregel durch die Wahrscheinlichkeit, uns für H_1 zu entscheiden, wenn H_1 zutrifft.

Wir sprechen auch von der Gütefunktion $G(p)$:

$$G(p) = P(\text{Entscheidung für } H_1 | p)$$

Diese hängt vom Wert von p ab, der nicht bekannt ist.

Sinnvollerweise sollte die Güte mit wachsendem p zunehmen, da wir uns ja immer weiter von $p = 0.5$ entfernen.

In unserem Beispiel ist es einfach die Gütefunktion zu bestimmen.

Wir wollen diese für den Fall bestimmen, daß wir zum Niveau $\alpha = 0.015625$ testen.

Wir lehnen H_0 ab, wenn $S = 6$ gilt.

Offensichtlich gilt

$$P_{H_0}(S = 6) = 0.5^6 = 0.015625$$

Die Teststatistik S ist auch unter H_1 binomialverteilt. Nur der Wert von p ändert sich.

Es gilt also

$$P(S = s) = \binom{6}{s} p^s (1-p)^{6-s} \quad \text{für } s = 0, 1, \dots, n.$$

Die Gütefunktion lautet

$$\begin{aligned} G(p) &= P_{H_1}(S = 6) \\ &= p^6 \end{aligned}$$

Die folgende Tabelle zeigt die Gütefunktion für ausgewählte Werte von p .

p	$G(p)$
0.60	0.047
0.70	0.118
0.80	0.262
0.90	0.531
0.95	0.735

Wir sehen, daß mit wachsendem p die Güte immer größer wird.

Wir sind bei der Bestimmung der Gütefunktion davon ausgegangen, daß wir zum Signifikanzniveau $\alpha = 0.015625$ testen. Zum Niveau $\alpha = 0.05$ konnten wir keinen Test durchführen, da gilt $P(S \geq 5) = 0.109375$.

Wenn wir $S = 6$ ablehnen, beträgt $\alpha = 0.015625$, lehnen wir für $S \geq 5$ ab, so beträgt $\alpha = 0.109375$.

Da die Verteilung der Teststatistik diskret ist, kann das vorgegebene Signifikanzniveau $\alpha = 0.05$ in der Regel nicht voll ausgeschöpft werden. Wir müssen also zu einem kleineren α testen.

Es gibt aber einen Ausweg:

Man kann randomisieren.

Schauen wir uns dies für den Fall $n = 6$ an.

Die Wahrscheinlichkeitsfunktion von S erhalten wir in R mit der Funktion `dbinom`.

Diese ruft man mit der Stelle `x`, dem Stichprobenumfang `n` und der Erfolgswahrscheinlichkeit `p` auf.

Wir runden sie auf 5 Stellen nach dem Komma, um die Übersicht zu behalten:

```
> round(dbinom(0:6,size=6,prob=0.5),5)
[1] 0.01563 0.09375 0.23438 0.31250 0.23438 0.09375 0.01563
```

Das sieht immer noch sehr unübersichtlich aus. Deshalb schreiben wir den Vektor der Wahrscheinlichkeiten neben den Vektor der Werte und daneben noch den Vektor der kumulierten Wahrscheinlichkeiten, den wir mit der Funktion `pbinom` erhalten:

```
> cbind(0:6,round(cbind(dbinom(0:6,size=6,prob=0.5),
  pbinom(0:6,size=6,prob=0.5)),5))
  [,1] [,2] [,3]
[1,]  0 0.01563 0.01563
[2,]  1 0.09375 0.10938
[3,]  2 0.23438 0.34375
[4,]  3 0.31250 0.65625
[5,]  4 0.23438 0.89062
[6,]  5 0.09375 0.98438
[7,]  6 0.01563 1.00000
```

Wir lehnen H_0 ab, wenn die Teststatistik zu groß ist.

Um $\alpha = 0.05$ voll auszuschöpfen, dürfen wir H_0 nicht jedesmal ablehnen, wenn $S = 5$ gilt.

Vielmehr lehnen wir H_0 immer ab, wenn $S = 6$ gilt, und führen ein zusätzliches Zufallsexperiment durch, wenn $S = 5$ gilt.

So könnten wir eine auf $(0, 1)$ gleichverteilte Zufallszahl ziehen. Ist diese kleiner als 0.3666 , so entscheiden wir uns gegen H_0 , ansonsten für H_0 .

In diesem Fall schöpfen wir $\alpha = 0.05$ aus da gilt:

$$\begin{aligned} P(H_0 \text{ ablehnen} | H_0 \text{ trifft zu}) &= P_{H_0}(S = 6) + P_{H_0}(S = 5) \cdot 0.3666 = \\ &= 0.01563 + 0.09375 \cdot 0.3666 = 0.025 \end{aligned}$$

In der Praxis wird in der Regel nicht randomisiert. Man gibt die Überschreitungswahrscheinlichkeit an.

Dies geschieht auch in R .

Mit Hilfe der Funktion `binom.test` können wir den Binomialtest durchführen.

Wir rufen auf:

```
> binom.test(5,6,0.5,alternative="greater")
```

```
Exact binomial test
```

```
data: 5 out of 6
```

```
number of successes = 5, n = 6, p-value = 0.1094
```

```
alternative hypothesis: true p is greater than 0.5
```

Für große Werte von n ist es mühsam, die Wahrscheinlichkeiten der Binomialverteilung zu bestimmen. Hier kann man auf die Normalapproximation der Binomialverteilung zurückgreifen.

Es gilt approximativ:

$$S \sim N(0.5 n, 0.25 n)$$

Im Testproblem

$$H_0 : p = 0.5 \quad \text{gegen} \quad H_1 : p \neq 0.5$$

lehnt man H_0 also ab, wenn gilt

$$S \geq 0.5 n + z_{1-\alpha/2} 0.5 \sqrt{n}$$

oder

$$S \leq 0.5 n - z_{1-\alpha/2} 0.5 \sqrt{n}$$

Dabei ist z_p das p -Quantil der Standardnormalverteilung.

Die Entscheidungsregel im Beispiel beruht auf der Anzahl der Fehlversuche unter 6 Versuchen.

Der Likelihoodratiotest

Die Herleitung der Teststatistiken der bisher betrachteten Tests ist sehr heuristisch.

Auf der Basis der Loglikelihood gibt es eine Reihe von Möglichkeiten zur Konstruktion einer geeigneten Teststatistik.

Wir betrachten das Testproblem

$$H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta \neq \theta_0$$

Außerdem sei die Loglikelihood $l(\theta)$ gegeben.

Die Loglikelihood nimmt ihr Maximum am M-L-Schätzer $\hat{\theta}_{ML}$ an. Es soll überprüft werden, ob θ_0 der wahre Wert des Parameters θ ist. Da der M-L-Schätzer ein guter Schätzer ist, sollte $\hat{\theta}_{ML}$ in der Nähe des wahren Wertes von θ liegen. Wenn also θ_0 der wahre Wert des Parameters ist, so sollte $\hat{\theta}_{ML}$ in der Nähe von θ_0 liegen. Das sollte natürlich auch für die Loglikelihoods $l(\hat{\theta}_{ML})$ und $l(\theta_0)$ gelten.

Beim Likelihoodratiotest wird nun die Differenz aus $l(\hat{\theta}_{ML})$ und $l(\theta_0)$ als Testkriterium gewählt.

Aus technischen Gründen wird das doppelte dieser Differenz gewählt.

Die Teststatistik des Likelihoodratiotest lautet somit:

$$LR = 2 \left(l(\hat{\theta}_{ML}) - l(\theta_0) \right)$$

Dies ist unter H_0 approximativ chiquadratverteilt. Die Anzahl der Freiheitsgrade ist gleich der Dimension des Parameterraums ohne die Beschränkung der Nullhypothese minus der Dimension des Parameterraums unter der Nullhypothese.

Bei einem zweiseitigen Test auf einen eindimensionalen Parameter ist LR also chiquadratverteilt mit einem Freiheitsgrad.

Schauen wir uns den Test für ein Beispiel an:

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und identisch mit den Parametern μ und σ^2 normalverteilt, wobei σ^2 bekannt sei.

Es ist zu testen

$$H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0$$

Die Loglikelihood ist gegeben durch

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

Der M-L-Schätzer für μ ist \bar{X} . Somit gilt

$$l(\hat{\mu}) = l(\bar{x}) = -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2$$

Unter H_0 gilt:

$$l(\mu_0) = -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2$$

Somit gilt

$$\begin{aligned} LR &= 2 (l(\bar{x}) - l(\mu_0)) \\ &= 2 \left(-n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 + \right. \\ &\quad \left. + n \ln \sqrt{2\pi} + \frac{n}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2 \right) \\ &= \frac{1}{\sigma^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \bar{x})^2 \right) \\ &= \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i^2 + \sum_{i=1}^n \mu_0^2 - 2 \sum_{i=1}^n x_i \mu_0 - \sum_{i=1}^n x_i^2 - \sum_{i=1}^n \bar{x}^2 + 2 \sum_{i=1}^n x_i \bar{x} \right) \\ &= \frac{1}{\sigma^2} (n \mu_0^2 - 2 n \mu_0 \bar{x} - n \bar{x}^2 + 2 n \bar{x}^2) \\ &= \frac{n}{\sigma^2} (\mu_0^2 - 2 \mu_0 \bar{x} + \bar{x}^2) \\ &= \frac{n}{\sigma^2} (\bar{x} - \mu_0)^2 \end{aligned}$$

Dies ist das Quadrat der klassischen Statistik eines Tests auf μ bei Normalverteilung mit bekannter Varianz.

Da

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

standardnormalverteilt ist, wenn die Zufallsvariablen X_1, \dots, X_n unabhängig und identisch mit den Parametern μ und σ^2 normalverteilt sind, ist LR exakt

chiquadratverteilt mit einem Freiheitsgrad, da das Quadrat einer standardnormalverteilten Zufallsvariablen chiquadratverteilt mit einem Freiheitsgrad ist.

Wir betrachten nun noch das Beispiel des Tests auf p :

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und identisch mit dem Parameter p bernoulliverteilt.

Es gilt also

$$p(X_i = x_i) = p^{x_i} (1 - p)^{1-x_i}$$

Es ist zu testen

$$H_0 : p = p_0 \quad \text{gegen} \quad H_1 : p \neq p_0$$

Die Loglikelihood ist gegeben durch

$$\begin{aligned} l(p) &= \ln p \sum_{i=1}^n x_i + \ln 1 - p (n - \sum_{i=1}^n x_i) \\ &= n \bar{x} \ln p + n (1 - \bar{x}) \ln 1 - p \end{aligned}$$

Die erste Ableitung der Loglikelihood nach p ist:

$$\begin{aligned} \frac{\delta}{\delta p} l(p) &= \frac{n \bar{x}}{p} - \frac{n (1 - \bar{x})}{1 - p} \\ &= \frac{n (\bar{x} - p)}{p(1 - p)} \end{aligned}$$

Der M-L-Schätzer ist, wie man leicht nachrechnet, \bar{X} .

Somit gilt

$$l(\bar{x}) = n \bar{x} \ln \bar{x} + n (1 - \bar{x}) \ln 1 - \bar{x}$$

Unter H_0 gilt:

$$l(p_0) = n \bar{x} \ln p_0 + n (1 - \bar{x}) \ln 1 - p_0$$

Somit gilt

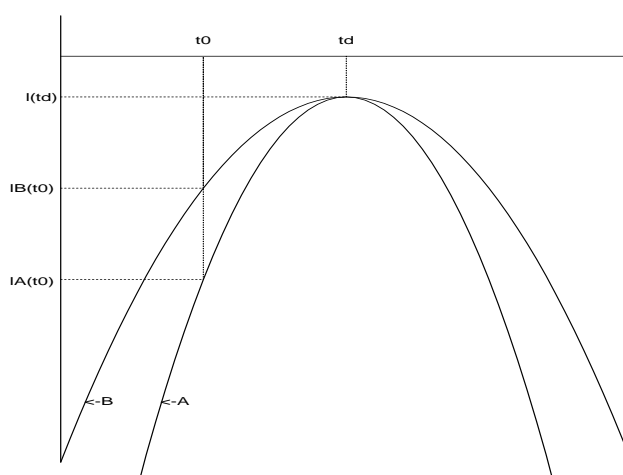
$$\begin{aligned} LR &= 2 (l(\bar{x}) - l(p_0)) \\ &= 2 (n \bar{x} \ln \bar{x} + n (1 - \bar{x}) \ln 1 - \bar{x} - \\ &\quad - n \bar{x} \ln p_0 - n (1 - \bar{x}) \ln 1 - p_0) \\ &= 2 \left(n \bar{x} \ln \frac{\bar{x}}{p_0} + n (1 - \bar{x}) \ln \frac{1 - \bar{x}}{1 - p_0} \right) \end{aligned}$$

Der Wald-Test

Anstatt die Differenz der Loglikelihoods zu betrachten, könnte man natürlich auch die Differenz aus $\hat{\theta}_{ML}$ und θ_0 als Teststatistik verwenden.

Hierbei muß man aber die negative Krümmung $C(\theta) = -\frac{\delta^2}{\delta\theta^2} l(\theta)$ der Loglikelihood berücksichtigen.

Die nachfolgende Graphik verdeutlicht, warum dies notwendig ist.



Die beiden Loglikelihoods weisen unterschiedliche Krümmung auf, haben an der gleichen Stelle ihr Maximum.

Im Fall A ist die stärker als im Fall B. Wie die Graphik zeigt, ist im Fall A aber auch die Differenz der Loglikelihood größer.

Je stärker also die Krümmung bei gleicher Differenz aus $\hat{\theta}_{ML}$ und θ_0 ist, um so mehr spricht diese Differenz für die Gegenhypothese.

Somit liegt es nahe, folgende Teststatistik zu betrachten:

$$W = (\hat{\theta}_{ML} - \theta_0)^2 C(\hat{\theta}_{ML})$$

Man nennt den zugehörigen Test auch Wald-Test. Anstatt der Krümmung wird oft der Erwartungswert der Krümmung verwendet.

Unter bestimmten Regularitätsbedingungen ist die aber die Fisher Information

$$I(\theta) = E(C(\theta)) = E\left(-\frac{\delta^2}{\delta\theta^2} l(\theta)\right)$$

Mit der Fisher-Information lautet die Teststatistik des Wald-Tests:

$$W = (\hat{\theta}_{ML} - \theta_0)^2 I(\hat{\theta}_{ML})$$

Die Teststatistik ist unter H_0 approximativ chiquadratverteilt. Die Anzahl der Freiheitsgrade ist gleich der Dimension des Parameterraums ohne die Beschränkung der Nullhypothese minus der Dimension des Parameterraums unter der Nullhypothese.

Bei einem zweiseitigen Test auf einen eindimensionalen Parameter ist W also chiquadratverteilt mit einem Freiheitsgrad.

Schauen wir uns den Wald-Test für das Beispiel der Normalverteilung an. Die Loglikelihood lautet:

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

Wir müssen nun noch die erste und zweite Ableitung bestimmen. Es gilt

$$S(\mu) = \frac{\partial}{\partial \mu} l(\mu) = \frac{n}{\sigma^2} (\bar{x} - \mu)$$

$$C(\mu) = -\frac{\partial^2}{\partial \mu^2} l(\mu) = \frac{n}{\sigma^2}$$

Wir erhalten somit

$$W = \frac{n}{2\sigma^2} l(\bar{x} - \mu_0)^2$$

Im Beispiel stimmen LR und W überein.

Für das Testproblem auf p ergibt sich folgende Loglikelihood:

$$l(p) = n \bar{x} \ln p + n(1 - \bar{x}) \ln 1 - p$$

Die erste Ableitung der Loglikelihood nach p ist:

$$S(p) = \frac{\delta}{\delta p} l(p) = \frac{n \bar{x}}{p} - \frac{n(1 - \bar{x})}{1 - p}$$

Die zweite Ableitung ist demnach:

$$C(p) = -\frac{\delta^2}{\delta p^2} l(p) = \frac{n \bar{x}}{p^2} + \frac{n(1 - \bar{x})}{(1 - p)^2}$$

Die Fisher-Information lautet:

$$\begin{aligned} I(p) &= E\left(-\frac{\delta^2}{\delta p^2} l(p)\right) \\ &= E\left[\frac{n\bar{x}n(1-\bar{x})}{p^2(1-p)^2}\right] \\ &= \frac{nE(\bar{x})}{p^2} + \frac{n(1-E(\bar{x}))}{(1-p)^2} \\ &= \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} \\ &= \frac{n}{p} + \frac{n}{1-p} \\ &= \frac{n}{p(1-p)} \end{aligned}$$

Also erhalten wir als Teststatistik des Wald-Tests

$$W = (\bar{x} - p_0)^2 \frac{n}{\bar{x}(1-\bar{X})}$$

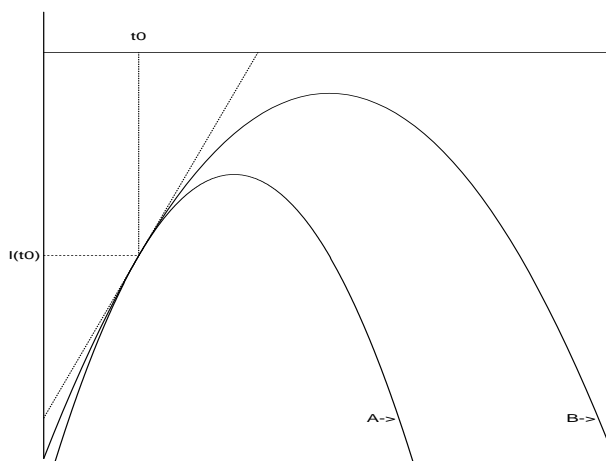
Der Score-Test

Beim Likelihoodratiotest und beim Wald-Test muß man den M-L-Schätzer bestimmen.

Das dritte Konstruktionsprinzip kommt ohne die Bestimmung des M-L-Schätzers aus.

Die Teststatistik beruht auf der Steigung der Loglikelihood in θ_0 . Diese sollte in der Nähe von 0 sein, wenn H_0 zutrifft. Starke Abweichungen der Steigung vom Wert 0 sprechen also gegen die Nullhypothese.

Wie das folgende Bild illustriert, muß man aber auch hier wieder die Krümmung der Loglikelihood berücksichtigen und zwar diesmal in θ_0 .



Auch hier hat die Loglikelihood im Fall A eine stärkere Krümmung als im Fall B. Hier spricht aber der Fall B mehr gegen die Nullhypothese, da die Differenz der Loglikelihood größer ist.

Man muß also die Steigung mit der inversen Krümmung gewichten und erhält

$$LM = S(\theta_0)^2 C(\theta_0)^{-1}$$

Man nennt den zugehörigen Test auch Lagrange-Multiplier-Test bzw. Score-Test.

Auch beim LM-Test wird oft die Fisher-Information verwendet:

$$LM = S(\theta_0)^2 I(\theta_0)^{-1}$$

Die Teststatistik ist unter H_0 approximativ chiquadratverteilt.

Die Anzahl der Freiheitsgrade ist gleich der Dimension des Parameterraums ohne die Beschränkung der Nullhypothese minus der Dimension des Parameterraums unter der Nullhypothese.

Bei einem zweiseitigen Test auf einen eindimensionalen Parameter ist LM also chiquadratverteilt mit einem Freiheitsgrad.

Für das Beispiel der Normalverteilung gilt:

$$S(\mu) = \frac{\partial}{\partial \mu} l(\mu) = \frac{n}{\sigma^2} (\bar{x} - \mu)$$

$$C(\mu) = -\frac{\partial^2}{\partial \mu^2} l(\mu) = \frac{n}{\sigma^2}$$

Also folgt

$$\begin{aligned} LM &= \frac{n^2 (\bar{x} - \mu)^2}{\sigma^4} \frac{\sigma^2}{n} \\ &= \frac{n (\bar{x} - \mu)^2}{\sigma^2} \end{aligned}$$

Für das Beispiel der Normalverteilung fallen die drei Teststatistiken zusammen.

Für den Test auf p erhalten wir

$$\begin{aligned} S(p) &= \frac{\delta}{\delta p} l(p) \\ &= \frac{n \bar{x}}{p} - \frac{n(1 - \bar{x})}{1 - p} \\ &= \frac{n(\bar{x} - p)}{p(1 - p)} \end{aligned}$$

Die Fisher-Information lautet:

$$I(p) = \frac{n}{p(1 - p)}$$

Also lautet die Teststatistik des Score-Tests:

$$\begin{aligned} LM &= \frac{n^2 (\bar{x} - p_0)^2}{p_0^2 (1 - p_0)^2} \frac{p_0 (1 - p_0)}{n} = \\ &= \frac{n (\bar{x} - p_0)^2}{p_0 (1 - p_0)} \end{aligned}$$

Schauen wir uns noch einmal die Teststatistik des LR-Tests und des Wald-Tests an:

$$LR = 2 \left(n \bar{x} \ln \frac{\bar{x}}{p_0} + (1 - \bar{x}) \ln \frac{1 - \bar{x}}{1 - p_0} \right)$$

$$W = (\bar{x} - p_0)^2 \frac{n}{\bar{x}(1 - \bar{x})}$$

Der Vergleich von LM-Test und Wald-Test zeigt, daß die Zähler identisch sind. Die Nenner sehen sehr ähnlich aus. In beiden Fällen steht die quadrierte Varianz des arithmetischen Mittels einer Zufallsstichprobe aus einer Bernoulliverteilung.

Aber beim Wald-Test wird diese Varianz an der Stelle \bar{x} bestimmt, während sie beim LM-Test an der Stelle p_0 bestimmt wird.

Welcher Zusammenhang besteht zur LR-Statistik?

Entwickelt man $f(x) = x \ln \frac{x}{x_0}$ in eine Taylorreihe um x_0 bis zum quadratischen Glied, so erhält man

$$f(x) = (x - x_0) + 0.5 (x - x_0)^2 \frac{1}{x_0}$$

Führen wir diese Approximation für die LR-Statistik mit $x = \bar{x}$ und $x_0 = p_0$ durch, so erhalten wir:

$$\begin{aligned} LR &\approx 2 \left(n \bar{x} \ln \frac{\bar{x}}{p_0} + n (1 - \bar{x}) \ln \frac{1 - \bar{x}}{1 - p_0} \right) = \\ &= 2 \left(n (\bar{x} - p_0) + 0.5 n \frac{(\bar{x} - p_0)^2}{p_0} + \right. \\ &\quad \left. + n (1 - \bar{x} - (1 - p_0)) + 0.5 n \frac{(1 - \bar{x} - (1 - p_0))^2}{1 - p_0} \right) = \\ &= 2 \left(0.5 n \frac{(\bar{x} - p_0)^2}{p_0} + 0.5 n \frac{(\bar{x} - p_0)^2}{1 - p_0} \right) = \\ &= n \frac{(\bar{x} - p_0)^2}{p_0 (1 - p_0)} \end{aligned}$$

Wir sehen, daß die Score-Statistik eine Approximation der LR-Statistik ist. Wir betrachten nun noch ein weiteres Beispiel.

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und identisch poissonverteilt mit Parameter λ .

Es gilt also

$$f_{X_i}(x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

Also gilt

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

und es folgt

$$\begin{aligned} S(\lambda) &= \frac{\delta}{\delta \lambda} l(\lambda) = \frac{1}{\lambda} \sum_{i=1}^n x_i - n = \\ &= \frac{n}{\lambda} (\bar{x} - \lambda) \end{aligned}$$

und

$$\begin{aligned} C(\lambda) &= -\frac{\delta^2}{\delta \lambda^2} l(\lambda) = \frac{n}{\lambda^2} \bar{x} \\ I(\lambda) &= E \left(-\frac{\delta^2}{\delta \lambda^2} l(\lambda) \right) = \frac{n}{\lambda^2} E(\bar{X}) = \\ &= \frac{n}{\lambda^2} \lambda = \frac{n}{\lambda} \end{aligned}$$

Somit erhalten wir folgende Teststatistiken:

$$\begin{aligned}
 LR &= 2(l(\bar{x}) - l(\lambda_0)) = \\
 &= 2\left(\ln \bar{x} \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i! - n\bar{x} - \right. \\
 &\quad \left. - \ln \lambda_0 \sum_{i=1}^n x_i + \sum_{i=1}^n \ln x_i! + n\lambda_0\right) = \\
 &= 2\left(n\bar{x} \ln \bar{x} - n\bar{x} - n\bar{x} \ln \lambda_0 + n\lambda_0\right) = \\
 &= 2n\bar{x} \ln \frac{\bar{x}}{\lambda_0} + 2n(\lambda_0 - \bar{x}) \\
 W &= (\bar{X} - \lambda_0)^2 \frac{n}{\bar{X}} \\
 LM &= \left(\frac{n}{\lambda_0}(\bar{x} - \lambda_0)\right)^2 \frac{\lambda_0}{n} = \\
 &= (\bar{x} - \lambda_0)^2 \frac{n}{\lambda_0}
 \end{aligned}$$

Beim Likelihood-Ratio-Test wird die Differenz der Loglikelihood an unterschiedlichen Werten des Parameters betrachtet.

Warum heißt der Test dann nicht Likelihood-Difference-Test?

Der Grund ist ganz einfach:

Ausgangspunkt ist das sogenannte Likelihoodverhältnis:

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta}_{ML})}$$

wobei $\hat{\theta}_{ML}$ der M-L-Schätzer von θ ist.

Man kann nun zeigen, daß $-2 \ln \Lambda$ approximativ chiquadratverteilt ist.

Nun gilt aber

$$-2 \ln \Lambda = 2(l(\hat{\theta}_{ML}) - l(\theta_0))$$

Bei der Normalverteilung haben wir den Likelihood-Ratio-Test für den Fall hergeleitet, daß die Varianz bekannt ist.

Wir wollen nun den realistischen Fall betrachten, daß die Varianz unbekannt ist.

In diesem Fall ist der Parameterraum $\Theta = (\mu, \sigma^2)$ zweidimensional. Unter $H_0 : \mu = \mu_0$ ist der Parameterraum gegeben durch $\Theta_0 = (\mu_0, \sigma^2)$.

Das Likelihoodverhältnis lautet

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$

Auch hier ist $-2 \ln \Lambda$ approximativ chiquadratverteilt.

Eine sehr schöne Beschreibung der unterschiedlichen Konstruktionsprinzipien von Tests, an der sich auch die obigen Ausführungen orientieren, ist bei Buse: The Likelihood Ratio, Wald, and Lagrange Multiplier Test: An expository note, The American Statistician, August 1982, S. 153-157 zu finden.

Ausblick

Der t-Test beruht auf der Annahme, daß die Zufallsvariablen X_1, \dots, X_n unabhängig und identisch mit den Parametern μ und σ^2 normalverteilt sind. Die Annahme der Normalverteilung ist in der Praxis in der Regel nicht erfüllt. Viele Datensätze stammen aus Verteilungen, die mehr Wahrscheinlichkeitsmasse an den Rändern haben als die Normalverteilung oder die sogar schief sind.

Ist die Verteilung der Grundgesamtheit keine Normalverteilung, so ist die Teststatistik des t-Tests nicht t-verteilt. Deshalb stimmen die vorgegebenen kritischen Werte nicht oder die Überschreitungswahrscheinlichkeit ist falsch. In diesen Fällen ist der Test entweder konservativ oder antikonservativ.

Ist der Test konservativ, so wird das vorgegebene Signifikanzniveau unterschritten, was die Güte des Tests vermindert. Man befindet sich aber in diesem Fall auf der sicheren Seite, hat also das Signifikanzniveau im Griff. Die Wahrscheinlichkeit für einen Fehler 1. Art ist kleiner als das vorgegebene α . Dies ist bei einem antikonservativen Test nicht der Fall.

Wie stark der t-Test auf Abweichungen von der Normalverteilung reagiert, ist bei Büning: Robuste und adaptive Tests zu finden.

Die Ergebnisse legen nahe, den t-Test nicht anzuwenden, wenn die Grundgesamtheit nicht normalverteilt ist. Zu dem gleichen Ergebnis kommen auch Staudte, Sheather: Robust Estimation and Testing, deren Kapitel 5.1 überschrieben ist mit: "Would W. S. Gosset Use the Student t-Test?"

Es stellen sich also folgende Fragen:

1. Wie kann man überprüfen, ob die Grundgesamtheit normalverteilt ist?
2. Welche Tests kann man anwenden, die entweder ohne die starke Annahme der Normalverteilung auskommen oder nicht zu stark auf Abweichungen von der Normalverteilung reagieren?

Mit der ersten Frage werden wir uns im nächsten Abschnitt beschäftigen.

In der zweiten Frage werden zwei Arten von Tests angesprochen.

Tests, die ohne eine spezielle Verteilungsannahme auskommen, heißen nicht-parametrische Tests. Mit diesen werden wir uns in den folgenden Kapiteln intensiv beschäftigen.

Tests, die nicht stark auf die Abweichung von der Normalverteilungsannahme reagieren, heißen robuste Tests. Sie sind nicht das Thema dieses Skripts.

Wir werden uns vielmehr mit robusten Schätzern beschäftigen.

Dies hat folgenden einfachen Grund:

In der Praxis haben sich robuste Schätzer und nichtparametrische Tests durchgesetzt. Robuste Schätzer sind einfach zu verstehen, während nichtparametrische Schätzer relativ kompliziert und schwer zu interpretieren sind. Bei den nichtparametrischen Tests braucht man keine Verteilungsannahme für die Grundgesamtheit, während die Verteilung der robusten Tests von der Verteilung der Grundgesamtheit abhängt.

1.2 Die Normalverteilungsannahme

Eine Möglichkeit die Normalverteilungsannahme zu überprüfen, besteht in der Erstellung einer Grafik.

Dabei sollen die Graphiken folgendes zeigen:

- Ist die Verteilung symmetrisch oder schief?
- Ist die Verteilung unimodal oder multimodal (eingipflig oder mehrgipflig)?
- Gibt es Ausreißer im Datensatz?

Bei den Graphiken muß man unterscheiden zwischen solchen, die sich speziell mit der Normalverteilungsannahme beschäftigen und solchen, die nur ein Bild der Daten geben.

Beginnen wir mit dem zweiten Fall.

Das populärste graphische Verfahren ist der **Boxplot**. Am Boxplot kann man leicht erkennen, ob die Daten aus einer symmetrischen Verteilung kommen, und ob der Datensatz Ausreißer enthält.

Zur Erstellung eines Boxplots benötigt man folgende 5 Statistiken eines Datensatzes:

- Minimum
- unteres Quartil $x_{0.25}$
- Median $x_{0.5}$
- oberes Quartil $x_{0.75}$
- Maximum

Das Maximum, Minimum und der Median sind eindeutig definiert, für die Schätzung der **Quartile** gibt es eine Reihe unterschiedlicher Vorschläge.

Die naheliegendste Idee stammt von Tukey:

Das untere Quartil $x_{0.25}$ teilt die untere Hälfte des geordneten Datensatzes in zwei gleiche Teile. Also liegt es nahe, als $x_{0.25}$ den Median der unteren Hälfte des geordneten Datensatzes zu wählen.

Wenn der Stichprobenumfang gerade ist, ist diese Definition eindeutig. Bei einem ungeraden Stichprobenumfang nimmt man den Median des Datensatzes zur unteren Hälfte mit dazu.

Für den geordneten Datensatz

```
0.553 0.570 0.576 0.601 0.606 0.606
```

ist das untere Quartil also 0.570, nämlich der Median von

```
0.553 0.570 0.576.
```

Für den geordneten Datensatz

```
0.553 0.570 0.576 0.601 0.606
```

ist das untere Quartil 0.570, nämlich der Median von

```
0.553 0.570 0.576.
```

Für das obere Quartil gilt die gleiche Regel.

Für den geordneten Datensatz

```
0.553 0.570 0.576 0.601 0.606 0.606
```

ist das obere Quartil also 0.606, nämlich der Median von

```
0.601 0.606 0.606.
```

Für den geordneten Datensatz

```
0.553 0.570 0.576 0.601 0.606 0.606
```

ist das obere Quartil also 0.606, nämlich der Median von

```
0.576 0.606 0.606.
```

Die folgenden R -Funktionen berechnen das untere und das obere Quartil nach der Tukey-Methode:

```
uquart <- function(x){  
# berechnet unteres Quartil nach Tukey  
# x ist Datensatz  
    x <- sort(x)  
    median(x[1:ceiling(length(x)/2)])  
}  
  
oquart <- function(x){  
# berechnet oberes Quartil nach Tukey  
# x ist Datensatz  
    x <- rev(sort(x))  
    median(x[1:ceiling(length(x)/2)])  
}
```

Für den Datensatz der Rechtecke der Schoschonen erhalten wir

- das Minimum durch

```
min(shosho)
[1] 0.553
```

- das untere Quartil durch

```
uquart(shosho)
[1] 0.606
```

- den Median durch

```
median(shosho)
[1] 0.641
```

- das obere Quartil durch

```
oquart(shosho)
[1] 0.681
```

das Maximum durch

```
max(shosho)
[1] 0.933
```

Mit diesen Zahlen könnte man nun ohne Probleme einen Boxplot erstellen. Die in R implementierte Boxplot-Funktion `boxplot` bestimmt das untere und obere Quartil jedoch nach einer anderen Methode. Um diese verstehen zu können, müssen wir weiter ausholen.

Das theoretische Quantil x_p ist definiert durch:

$$F_X(x_p) = p.$$

Es liegt also nahe, zur Bestimmung des empirischen Quantils die empirische Verteilungsfunktion $F_n(x)$ als Ausgangspunkt zu wählen.

Für eine Stichprobe x_1, \dots, x_n ist die empirische Verteilungsfunktion $F_n(x)$ definiert durch

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n H(x - x_i)$$

mit

$$H(u) = \begin{cases} 0 & \text{für } u < 0 \\ 1 & \text{für } u \geq 0 \end{cases}$$

Die empirische Verteilungsfunktion an der Stelle x ist also gleich der Anzahl der Beobachtungen, die x nicht übertreffen.

Mit Hilfe der folgenden Funktion kann man in R die empirische Verteilungsfunktion berechnen:

```
empdf <- function(x){
# berechnet die empirische Verteilungsfunktion
# fuer den Datensatz x
      m <- table(x)
      cumsum(m)/sum(m)
}
```

Für die Schoschonendaten erhalten wir:

```
e <- empdf(shosho)

e
[1] 0.553 0.570 0.576 0.601 0.606 0.609 0.611 0.615 0.628
[10] 0.654 0.050 0.100 0.150 0.200 0.300 0.350 0.400 0.450
[19] 0.500 0.550 0.662 0.668 0.670 0.672 0.690 0.693 0.749
[27] 0.844 0.933 0.600 0.650 0.700 0.750 0.800 0.850 0.900
[36] 0.950 1.000
```

Die nachstehende Funktion zeichnet die empirische Verteilungsfunktion

```
plot.empdf <- function(data, bereich){
# zeichnet die emprirische Verteilungsfunktion F(x)
# data ist der Datensatz
# bereich ist Vektor der Laenge 2
# 1. Komponente von bereich ist das Minimum
# 2. Komponente von bereich ist das Maximum des Bereichs,
# auf dem die Verteilungsfunktion gezeichnet werden soll

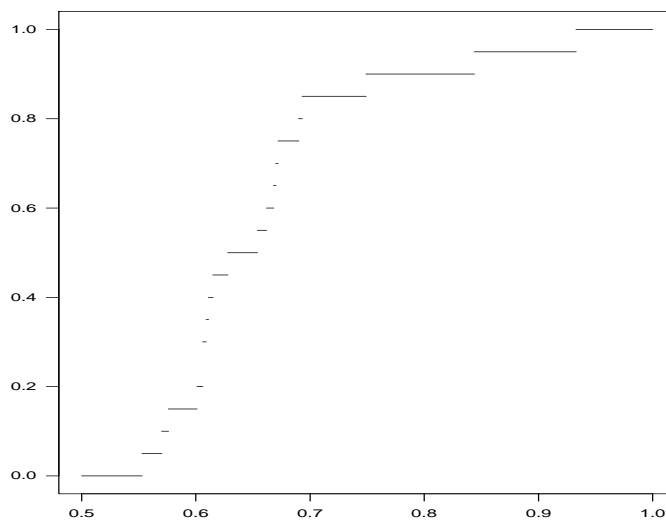
h <- empdf(data)
x <- as.numeric(names(h))
n <- length(x)
plot(c(bereich[1], x[1]), c(0, 0), xlab = "", ylab = "",
      xlim = bereich, ylim = c(0, 1), type = "l")
```



```
for(i in 1:(n - 1)){lines(x[i:(i + 1)], rep(h[i], 2))}
lines(c(x[n], bereich[2]), c(1, 1))
}
```

Wir zeichnen die empirische Verteilungsfunktion der Schoschonen-Daten.

```
plot.empdf(shosho, c(0.5, 1))
```



Wir sehen, daß die empirische Verteilungsfunktion eine Treppenfunktion ist. Deshalb ist ihre Inverse nicht eindeutig definiert, so daß man Quantile nicht eindeutig bestimmen kann.

Die übliche Vorgehensweise besteht darin, die empirische Verteilungsfunktion zu glätten, indem man sie durch eine stetige, stückweise lineare Funktion $\tilde{F}(x)$ ersetzt.

Hierbei muß man festlegen, welchen Wert die Funktion $\tilde{F}(x)$ an den Stellen $x_{(1)}, \dots, x_{(n)}$ annimmt, wobei die Stützstellen die geordneten Beobachtungen sind.

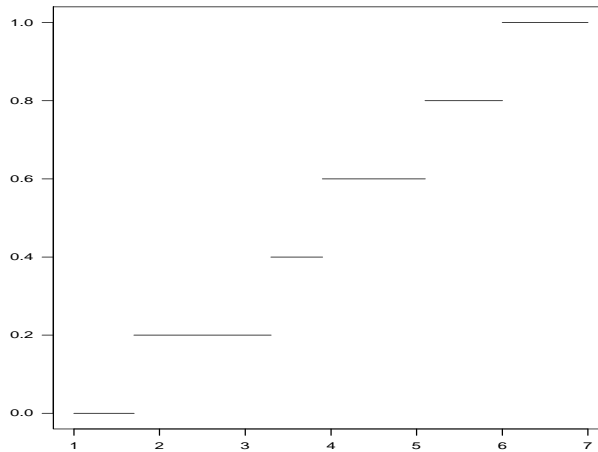
Betrachten wir hierzu folgenden Datensatz vom Umfang $n=5$:

3.3 3.9 1.7 6.0 5.1

Die geordneten Werte sind gegeben durch:

$$x_{(1)} = 1.7 \quad x_{(2)} = 3.3 \quad x_{(3)} = 3.9 \quad x_{(4)} = 5.1 \quad x_{(5)} = 6.0$$

Das folgende Bild zeigt die empirische Verteilungsfunktion:

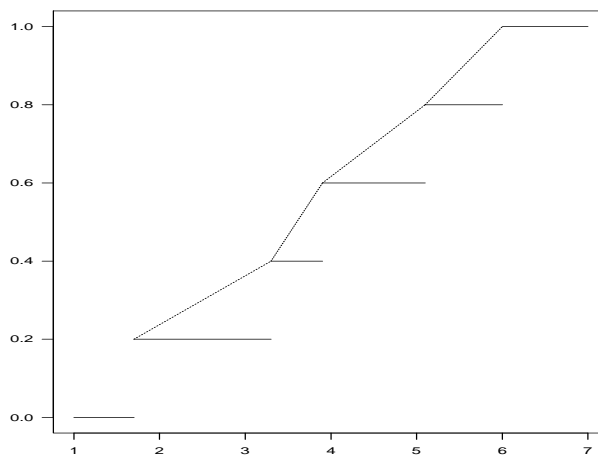


Es liegt nahe

$$\tilde{F}(x_{(i)}) = \frac{i}{n}$$

für $i = 1, \dots, n$ zu wählen und linear zu interpolieren.

Das folgende Bild zeigt die Approximation:



Als Quantilschätzer ergibt sich in diesem Fall:

$$\hat{Q}(p) = \begin{cases} x_{(1)} & \text{für } p < \frac{1}{n} \\ (1-g)x_{(j)} + gx_{(j+1)} & \text{für } \frac{1}{n} \leq p \leq 1 \end{cases}$$

mit $j = \lfloor np + 1 \rfloor$ und $g = np + 1 - j$.

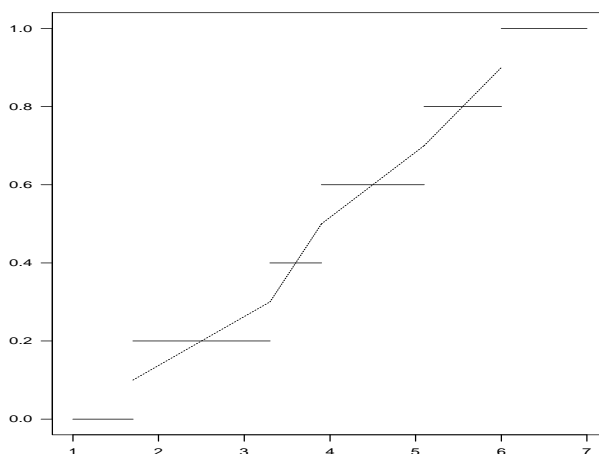
Der Nachteil dieses Schätzers ist, daß eine Vielzahl von Quantilen durch das Minimum geschätzt werden. Für $0 < p < \frac{1}{n}$ wird x_p durch das Minimum des Datensatzes geschätzt. Für $p > \frac{n-1}{n}$ wird hingegen jedem p ein anderes x_p zugeordnet. Die beiden Ränder der Verteilung werden also unterschiedlich behandelt.

Dieser Nachteil kann dadurch behoben werden, daß der Beitrag der i -ten Orderstatistik $x_{(i)}$ zu gleichen Teilen auf die Bereiche unterhalb und oberhalb von ihr aufgeteilt wird.

Somit gilt

$$\tilde{F}(x_{(i)}) = \frac{i - 0.5}{n}$$

Das folgende Bild zeigt die Approximation:



Als Quantilschätzer ergibt sich in diesem Fall:

$$\hat{Q}(p) = \begin{cases} x_{(1)} & \text{für } p < \frac{0.5}{n} \\ (1 - g)x_{(j)} + gx_{(j+1)} & \text{für } \frac{0.5}{n} \leq \frac{n-0.5}{n} \\ x_{(n)} & \text{für } p > \frac{n-0.5}{n} \end{cases}$$

mit $j = \lfloor np + 0.5 \rfloor$ und $g = np + 0.5 - j$.

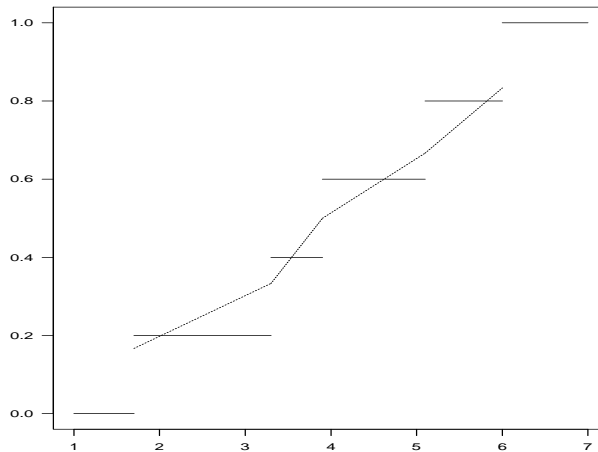
Eine andere Wahl geht von folgender Idee aus:

Die n Orderstatistiken teilen den Wertebereich der Verteilung in $n + 1$ Bereiche, von denen jeder im Mittel $100 \cdot \frac{1}{n + 1}$ Prozent der Beobachtungen enthält.

Dies legt nahe:

$$\tilde{F}(x_{(i)}) = \frac{i}{n + 1}$$

Das folgende Bild zeigt die Approximation:



Als Quantilschätzer ergibt sich in diesem Fall:

$$\hat{Q}(p) = \begin{cases} x_{(1)} & \text{für } p < \frac{1}{n+1} \\ (1 - g)x_{(j)} + gx_{(j+1)} & \text{für } \frac{1}{n+1} \leq \frac{n}{n+1} \\ x_{(n)} & \text{für } p > \frac{n}{n+1} \end{cases}$$

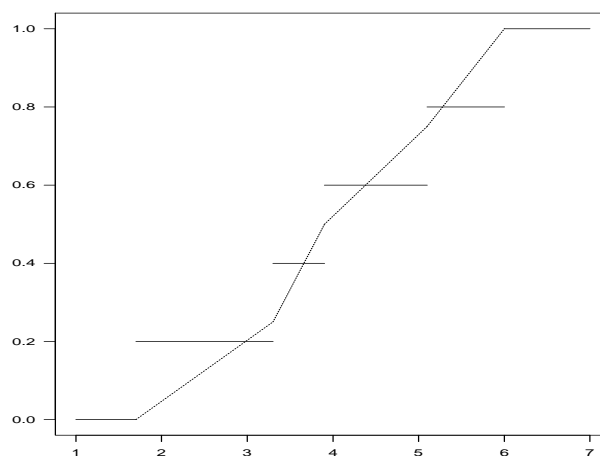
mit $j = \lfloor (n + 1)p \rfloor$ und $g = (n + 1)p - j$.

In R wird folgende Glättung der empirischen Verteilungsfunktion verwendet:

$$\tilde{F}(x_{(i)}) = \frac{i - 1}{n - 1}$$

Hier wird jedem p ein anderer Wert von x_p zugeordnet.

Das folgende Bild zeigt die Approximation:



Als Quantilschätzer ergibt sich in diesem Fall:

$$\hat{Q}(p) = (1 - g) x_{(j)} + g x_{(j+1)}$$

mit $j = \lfloor (n - 1)p + 1 \rfloor$ und $g = (n - 1)p + 1 - j$.

Dieser Schätzer hat den Vorteil, daß man Quantile, die zu kleinem oder großem p gehören, nicht ausschließlich durch das Minimum oder das Maximum schätzt.

Alle diese Schätzer sind Spezialfälle der folgenden Klasse von Quantilschätzern:

$$\hat{Q}(p) = \begin{cases} x_{(1)} & \text{für } p < \frac{1-\gamma}{n+1-\gamma-\delta} \\ (1-g)x_{(j)} + gx_{(j+1)} & \text{für } \frac{1-\gamma}{n+1-\gamma-\delta} \leq \frac{n-\gamma}{n+1-\gamma-\delta} \\ x_{(n)} & \text{für } p > \frac{n-\gamma}{n+1-\gamma-\delta} \end{cases}$$

mit $j = \lfloor (n+1-\gamma-\delta)p + \gamma \rfloor$ und $g = (n+1-\gamma-\delta)p + \gamma - j$. Es gilt:

- 1.Fall: $\gamma = 0$, $\delta = 1$
- 2.Fall: $\gamma = 0.5$, $\delta = 0.5$
- 3.Fall: $\gamma = 0$, $\delta = 0$
- 4.Fall: $\gamma = 1$, $\delta = 1$

Der 2. Fall liefert übrigens die Quartile nach Tukey.

Die folgende R Funktion liefert die Schätzung der Quantile nach dem oben angegebenen Schätzer:

```
quantil <- function(x, p, gamma, delta) {
# bestimmt fuer den Datenvektor x die Quantile xp

  xs <- sort(x)
  n <- length(x)
  j <- floor((n + 1 - gamma - delta) * p + gamma)
  g <- (n + 1 - gamma - delta) * p + gamma - j
  qua <- (1 - g) * xs[j] + g * xs[j + 1]
  qua[p < ((1 - gamma)/(n + 1 - gamma - delta))] <- xs[1]
  qua[p > ((n - gamma)/(n + 1 - gamma - delta))] <- xs[n]
  qua
}
```

Für den Datensatz `shosho` erhalten wir folgende Schätzung der Quantile

```
quantil(shosho, c(0.25, 0.5, 0.75), 0, 1)
[1] 0.606 0.628 0.672
```

```
quantil(shosho, c(0.25, 0.5, 0.75), 0.5, 0.5)
[1] 0.606 0.641 0.681
```

```
quantil(shosho,c(0.25,0.5,0.75),0,0)
[1] 0.6060 0.6410 0.6855
```

```
quantil(shosho,c(0.25,0.5,0.75),1,1)
[1] 0.6060 0.6410 0.6765
```

In R ist eine Funktion `quantile` implementiert, die die Quantile für den Fall $\gamma = 1$ und $\delta = 1$ bestimmt.

```
quantile(shosho,c(0.25,0.5,0.75))
 25%   50%   75%
0.606 0.641 0.6765
```

Ruft man diese Funktion nur mit dem Datenvektor auf, so werden die für den Boxplot benötigten Zahlen als Ergebnis geliefert:

```
quantile(shosho)
 0%   25%   50%   75%  100%
0.553 0.606 0.641 0.6765 0.933
```

Die Funktion `quantile` erlaubt in der aktuellen Version mit Hilfe des Arguments `type` die Umsetzung von neun verschiedene Quantilsalgorithmen. Details findet man unter `?quantile`.

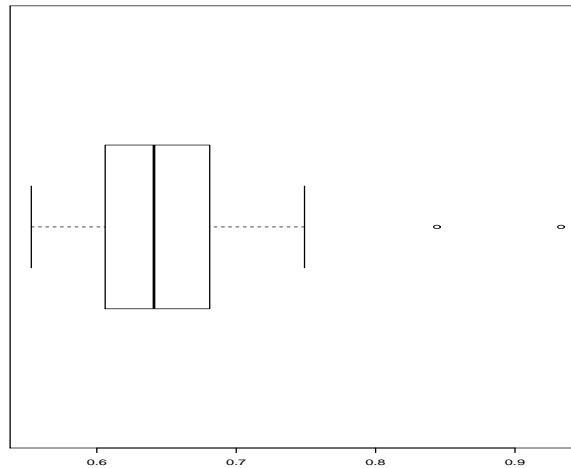
Beim Boxplot wird nun ein Kasten vom unteren Quartil bis zum oberen Quartil gezeichnet. Außerdem wird der Median als Linie in den Kasten eingezeichnet. Von den Rändern des Kastens bis zu den Extremen werden Linien gezeichnet.

Um Ausreißer zu markieren, wird der letzte Schritt in der Regel folgendermaßen modifiziert:

Sind Punkte mehr als das 1.5-fache der Kastenbreite von Quartilen entfernt, so wird die Linie nur bis zum 1.5-fachen der Kastenbreite gezeichnet. Alle Punkte, die außerhalb liegen, werden markiert.

Der folgende Aufruf liefert einen Boxplot in horizontaler Ausrichtung:

```
boxplot(shosho, horizontal=TRUE)
```



Am Boxplot ist zu erkennen, daß zwei Punkte Ausreißer sind. Ansonsten sieht er so aus, wie man ihn bei Normalverteilung erwarten würde. So liegt der Median genau in der Mitte des Kastens. Ohne die Ausreißer ist die Verteilung symmetrisch.

Die klassische Graphik zur Darstellung des Datensatzes eines stetigen Merkmals ist das **Histogramm**.

Zur Erstellung eines Histogramms geht man folgendermaßen vor:

Man wählt $k + 1$ Klassengrenzen $x_0^*, x_1^*, \dots, x_k^*$ und teilt den Wertebereich der Variablen in k Klassen auf:

$$[x_0^*, x_1^*], (x_1^*, x_2^*], \dots, (x_{k-1}^*, x_k^*]$$

Dann bestimmt man den Anteil h_i der Beobachtungen, die in die i -te Klasse fallen.

Das Histogramm $\hat{f}(x)$ an der Stelle x ist dann definiert durch:

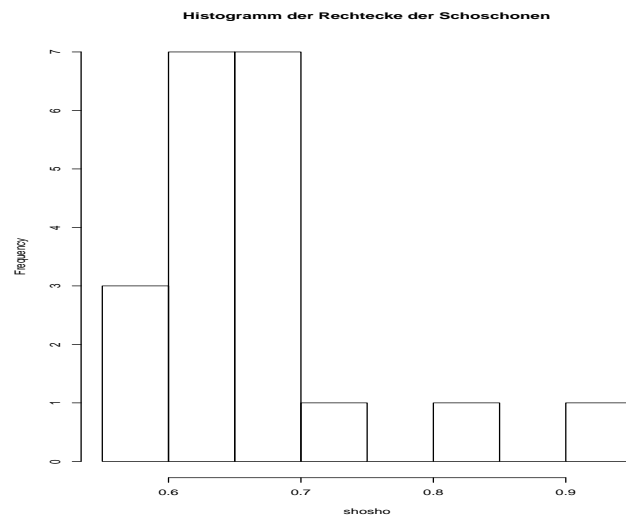
$$\hat{f}(x) = \begin{cases} \frac{h_i}{x_i^* - x_{i-1}^*} & \text{für } x_{i-1}^* < x \leq x_i^* \\ 0 & \text{sonst} \end{cases}$$

Das Histogramm besteht aus Rechtecken über den Klassen, deren Breite der relativen Häufigkeit der Klasse entspricht.

Es ist üblich, alle Intervalle gleich breit zu wählen. Dies ist auch die Vorgehensweise in der Funktion `hist` in `R`.

Der folgende Aufruf liefert ein Histogramm, bei dem die Klassengrenzen und Anzahl der Klassen automatisch gewählt werden.

```
hist(shosho,main="Histogramm der Rechtecke der Schoschonen")
```



Es ist möglich, `hist` so aufzurufen, daß kein Histogramm erstellt wird, sondern die Klassengrenzen und die Besetzungshäufigkeiten der Klassen ausgegeben werden.

```
hist(shosho,plot=F)
$breaks:
[1] 0.5 0.6 0.7 0.8 0.9 1.0
$count:
[1] 3 14 1 1 1
```

In R werden also runde Klassengrenzen gewählt.

Das Histogramm ist ein Schätzer der Dichtefunktion, der jedoch nicht glatt ist.

Rosenblatt hat als erster sogenannte **Kerndichteschätzer** vorgeschlagen.

Er geht davon aus, daß die Dichtefunktion $f_X(x)$ die Ableitung der Verteilungsfunktion $F_X(x)$ ist.

Es gilt also

$$f_X(x) = \frac{\delta}{\delta x} F_X(x) = \lim_{h \rightarrow 0} \frac{1}{2h} P(X - h < X < X + h) \quad (1.1)$$

Für jeden Wert von h können wir $P(X - h < X < X + h)$ durch den Anteil der Beobachtungen schätzen, die in das Intervall $(x - h, x + h)$ fallen:

$$\hat{P}(X - h < X < x + h) = \frac{1}{n} \cdot (\text{Anzahl von } X_1, \dots, X_n \text{ in } (x - h, x + h))$$

Somit erhalten wir als Schätzer der Dichtefunktion:

$$\hat{f}_X(x) = \frac{1}{2hn} \cdot (\text{Anzahl von } X_1, \dots, X_n \text{ in } (x - h, x + h))$$

Diesen Schätzer kann man auch folgendermaßen darstellen:

$$\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x - X_i}{h}\right)$$

mit

$$w(x) = \begin{cases} 0.5 & \text{für } |x| < 1 \\ 0 & \text{für } |x| \geq 1 \end{cases}$$

Also gilt

$$w\left(\frac{x - X_i}{h}\right) = 0.5,$$

wenn gilt

$$\left|\frac{x - X_i}{h}\right| < 1$$

Dies ist aber äquivalent zu

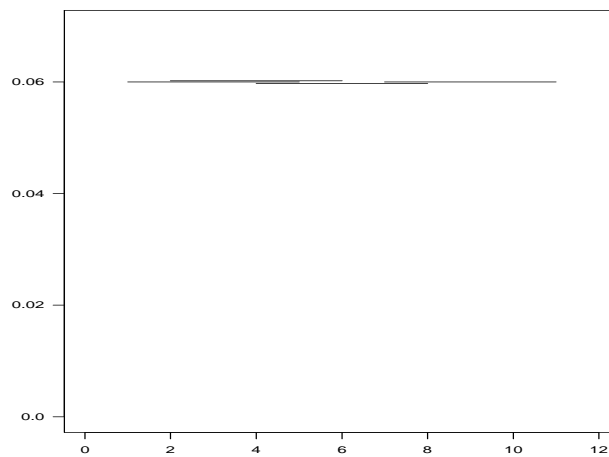
$$X_i - h < x < X_i + h$$

Jeder Punkt im Intervall $X_i - h < x < X_i + h$ liefert also zur Summe den Beitrag 0.5.

Der Schätzer wird also so konstruiert, daß man ein Rechteck der Breite $2h$ und Höhe $\frac{1}{2nh}$ um jede Beobachtung legt und die Höhen dieser Rechtecke aufsummiert.

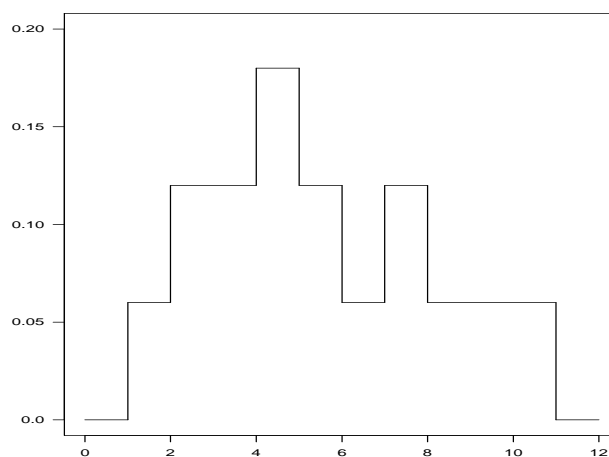
Das folgende Bild veranschaulicht den Dichteschätzer mit $h=2$ für den Datensatz:

3 4 6 9



Jede Gerade entspricht dem Beitrag einer Beobachtung zur Schätzung der Dichtefunktion in ihrer Umgebung.

Addieren wir die Geraden auf, so erhalten wir dann folgende Dichteschätzung:



Der Schätzer ist so konstruiert, daß alle Punkte in der Umgebung einer Beobachtung das gleiche Gewicht erhalten. Eine Beobachtung liefert also für alle Punkte in ihrer Umgebung den gleichen Beitrag zur Dichtefunktion.

Dies führt dazu, daß die geschätzte Dichtefunktion nicht glatt ist.

Um einen glatteren Verlauf der geschätzten Dichtefunktion zu erhalten, sollte man die Gewichtungsfunktion $w(u)$ so wählen, daß der Beitrag einer Beobachtung zur Dichteschätzung mit wachsendem Abstand von ihr abnimmt.

Wählt man dann als Gewichtungsfunktion eine Dichtefunktion, so besitzt auch die Dichteschätzung alle Eigenschaften einer Dichtefunktion.

Die Gewichtungsfunktion $w(u)$ heißt auch Kernfunktion.

In der Literatur gibt es eine Reihe von Vorschlägen für die Wahl der Kernfunktion.

Die klassische Wahl ist der Gauss-Kern:

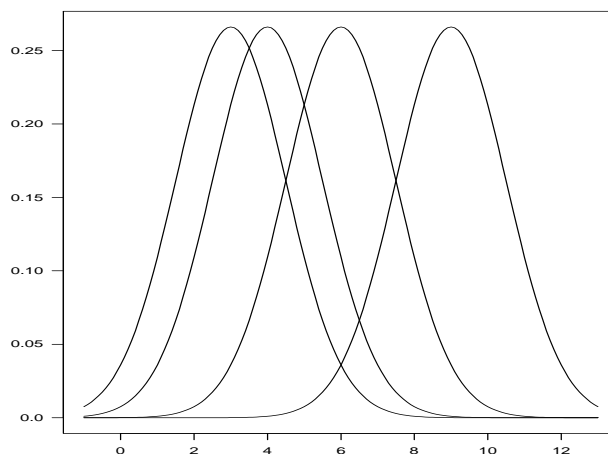
$$w(t) = \frac{1}{\sqrt{2\pi}} e^{-0.5t^2}$$

Dies ist gerade die Dichtefunktion der Standardnormalverteilung.

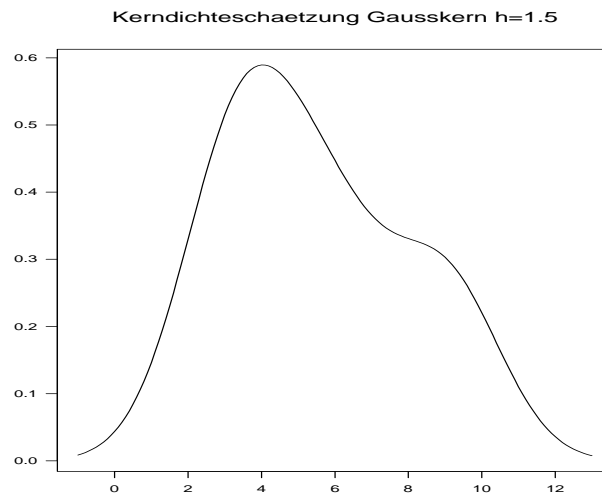
Weitere Kernfunktionen sind bei Silverman: Density Estimation zu finden.

Schauen wir uns an, was passiert, wenn wir für den obigen Datensatz die Dichtefunktion mit einem Gausskern mit $h = 1.5$ schätzen.

Wir legen um jede Beobachtung eine Dichtefunktion der Normalverteilung mit $\sigma = 1.5$ und erhalten folgendes Bild:

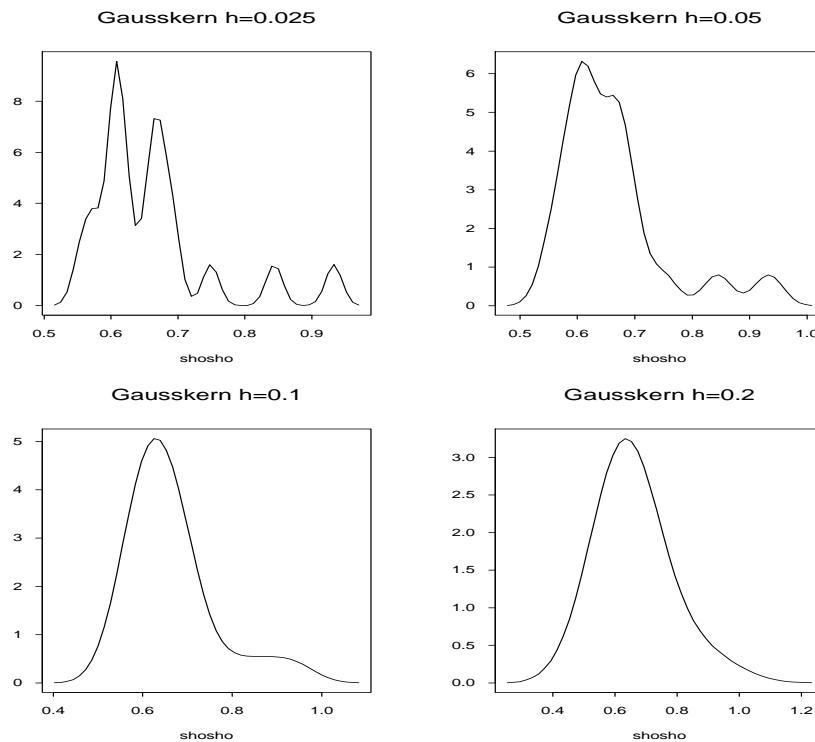


Dann addieren wir diese Kernfunktionen auf und erhalten folgendes Bild:



In R existiert eine Funktion `density` zur Schätzung einer Dichtefunktion. Bei dieser wird standardmäßig ein Gauss-Kern verwendet. Neben dem Datensatz muß man noch die Fensterbreite $2h$ übergeben.

Die folgenden Bilder zeigt den Dichteschätzer der Schoschonendaten mit Gausskern für unterschiedliche Werte von h :



Den Bildern kann man folgendes entnehmen:

Mit wachsendem h wird die Dichteschätzung immer glatter, dabei gehen aber lokale Informationen verloren. Für $h = 0.2$ erhalten wir fast eine Normalverteilung. Lokale Unterschiede werden für kleine Werte von h gut wiedergegeben, während die Kurve nicht sehr glatt wirkt.

In der Literatur gibt es eine Vielzahl von Vorschlägen für die Wahl von h . So findet man bei Silverman:

$$h = 0.9 \min\left(s, \frac{iqr}{1.34}\right) n^{-0.2}$$

Dabei sind:

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

und

$$iqr = x_{0.75} - x_{0.25}$$

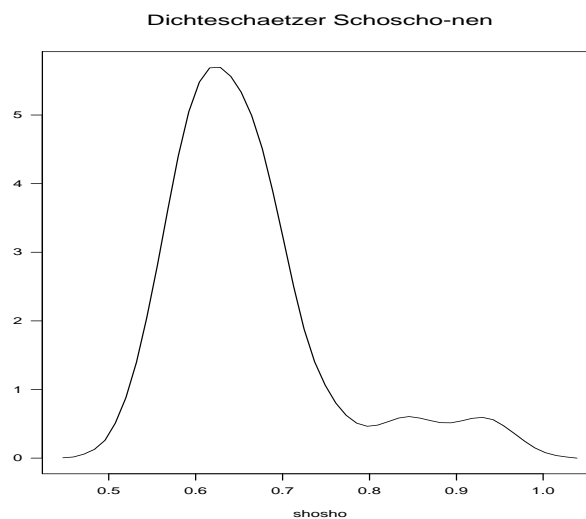
Die erste Größe ist also die Standardabweichung, und iqr ist der Interquartilsabstand zwischen dem oberen und dem unteren Quartil.

Den Interquartilsabstand erhält man durch:

```
iqr <- quantile(shosho)[4]-quantile(shosho)[2] iqr
75%
0.0705
```

Wir wählen hier $h = iqr$.

```
plot(density(shosho,width=2*iqr),xlab="shosho",ylab="",
      type="l",main="Dichteschätzer Schoschonen")
```



Auch der Dichteschätzer deutet auf Ausreißer und Symmetrie im Zentrum hin.

Die letzte Graphik, die wir noch betrachten wollen, dient dazu, Abweichungen von der Normalverteilung aufzudecken.

Sie wird **normal-probability-plot** genannt.

Es werden die geordneten Werte $x_{(1)}, \dots, x_{(n)}$ gegen spezielle Quantile z_p der Standardnormalverteilung gezeichnet.

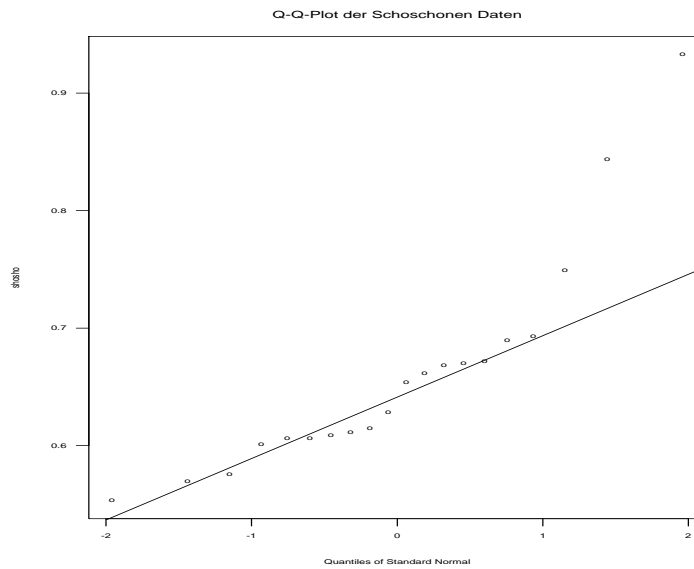
In Anlehnung an die Quantilschätzung wird $x_{(i)}$ gegen z_p mit $p = \frac{i-\gamma}{n+1-\gamma-\delta}$ gezeichnet.

Im Idealfall liegen alle Punkte auf einer Geraden. Große Abweichungen von dieser Geraden deuten darauf hin, dass die Normalverteilungsannahme verletzt ist.

```
qqnorm(shosho,main="Q-Q-Plot der Schoschonen Daten",cex=0.6)
```

Diese Idealgerade kann man in R der Grafik hinzufügen.


```
qqline(shosho)
```



Das Bild zeigt, daß die Normalverteilungsannahme für die meisten Beobachtungen gerechtfertigt ist. Es fallen aber wieder die zwei Ausreißer am rechten oberen Rand der Grafik auf.

1.3 Robuste Schätzer

Das arithmetische Mittel ist empfindlich gegenüber extremen Beobachtungen. Schätzer, bei denen dies nicht der Fall ist, heißen robust. Es ist naheliegend, extreme Beobachtungen aus der Stichprobe zu entfernen und den Mittelwert der übriggebliebenen Beobachtungen zu bestimmen. Dabei wird einer zuvor festgelegter Anteil α von beiden Rändern der geordneten Stichprobe $x_{(1)}, \dots, x_{(n)}$ entfernt.

Man spricht in diesem Fall von einem α -getrimmten Mittelwert.

Formal kann man einen α -getrimmten Mittelwert folgendermaßen beschreiben:

Man gibt einen Anteil α vor und berechnet

$$\bar{x}_\alpha = \frac{1}{n - 2 \lfloor n\alpha \rfloor} \sum_{i=1+\lfloor n\alpha \rfloor}^{n-\lfloor n\alpha \rfloor} x_{(i)}$$

Dabei ist $\lfloor u \rfloor$ der ganzzahlige Anteil von u mit $u \geq 0$, z.B. $\lfloor 2.3 \rfloor = 2$.

Der Datensatz `shosho` enthält 20 Beobachtungen. Wollen wir den 0.05-getrimmten Mittelwert bestimmen, so streichen wir das Maximum und das Minimum aus der Stichprobe und berechnen den Mittelwert der restlichen 18 Beobachtungen, da gilt

$$\lfloor 20 \cdot 0.05 \rfloor = 1.$$

In R kann man den α -getrimmten Mittelwert mit Hilfe der Funktion `mean` bestimmen.

Wir müssen das Argument `trim` auf den Anteil α setzen.

```
mean(shosho, trim=0.05)
[1] 0.6513333
```

Die folgende Tabelle zeigt \bar{x}_α in Abhängigkeit von α für den Datensatz shosho

α	\bar{x}_α
0	0.6605
0.05	0.6513
0.10	0.6444
0.15	0.6418
0.20	0.6405
0.25	0.6395
0.30	0.6396
0.35	0.6397
0.40	0.6398
0.45	0.6410
0.50	0.6410

Welcher der getrimmten Mittelwert gibt die Lage des Datensatzes am besten wieder?

Wie soll man also den Trimmanteil α wählen?

Wenn man statistisch argumentiert, so könnte man sagen, daß man den Schätzer wählen sollte, der am genauesten ist.

Ein Maß für die Genauigkeit ist die Varianz. Leider ist die Varianz der einzelnen Schätzer unbekannt. Man kann sie aber schätzen. Eine Möglichkeit dafür bietet der **Bootstrap**.

Wir haben schon weiter oben gesehen, daß man die Verteilung einer Statistik durch Simulation bestimmen kann. Hierzu erzeugt man Stichproben aus der Verteilung und bestimmt für jede Stichprobe den Wert der Statistik. Die empirische Verteilung der Statistik approximiert dann die theoretische Verteilung.

Nun ist in der Regel die Verteilung die Verteilungsfunktion $F_X(x)$ der Grundgesamtheit unbekannt.

Efron hat nun 1979 vorgeschlagen, die Stichproben nicht aus der unbekanntem Verteilungsfunktion $F_X(x)$ sondern aus der empirischen Verteilungsfunktion $F_n(x)$ zu ziehen.

Man muß also aus der Stichprobe x_1, \dots, x_n die Stichproben x_1^*, \dots, x_n^* mit Zurücklegen ziehen.

Ist man nun an der Verteilung einer Statistik $\hat{\theta} = g(X_1, \dots, X_n)$ interessiert,

wenn gilt $X_i \sim F_X(x)$, so approximiert der Bootstrap diese Verteilung durch die Verteilung von $\hat{\theta}^* = g(X_1^*, \dots, X_n^*)$, wobei gilt $X_i^* \sim F_n(x)$.

Die Bootstrap-Verteilung kann man nun mit Hilfe von Simulation bestimmen:

1. Erzeuge eine Bootstrap-Stichprobe x_1^*, \dots, x_n^* durch Ziehen mit Zurücklegen aus der Stichprobe x_1, \dots, x_n .
2. Bestimme $\hat{\theta}^* = g(x_1^*, \dots, x_n^*)$.
3. Wiederhole die Schritte 1. und 2. B-mal.
4. Approximiere die Verteilung von $\hat{\theta} = g(X_1, \dots, X_n)$ durch die Verteilung von $\hat{\theta}^* = g(X_1^*, \dots, X_n^*)$.

Wir wollen nun die Bootstrap-Verteilung des arithmetischen Mittels und des Medians für die Shoshonen-Daten vergleichen.

Mit Hilfe der Funktion `sample` kann man eine Stichprobe vom Umfang n aus den natürlichen Zahlen 1 bis N ziehen.

Der Aufruf

```
sample(49,6)
[1] 10 36  1 38 26 14
```

zieht ohne Zurücklegen 6 Zahlen aus den natürlichen Zahlen 1 bis 49. Das entspricht einer einfachen Lottoziehung.

Der Aufruf

```
sample(49,6,replace=T)
[1]  1 46 35  1 30 27
```

zieht mit Zurücklegen 6 Zahlen aus den natürlichen Zahlen 1 bis 49.

Wir können auch direkt aus einem Vektor w der Länge n eine Stichprobe vom Umfang n mit Zurücklegen ziehen durch

```
sample(w,replace=T)
```

Wir bestimmen nun zunächst die Bootstrap-Verteilung des Mittelwerts der Shoshonen-Daten. Dabei wählen wir $B = 1000$.

Zuerst erzeugen wir einen Vektor v der Länge 1000, der nur Nullen enthält. In diesen schreiben wir dann die einzelnen Realisationen.

In einer Schleife ziehen wir wiederholt mit Zurücklegen Stichproben vom Umfang 20 aus den Shoshonen-Daten und bestimmen für jede Stichprobe den Mittelwert.

In R sieht das dann so aus:

```
v <- rep(0,1000)
for(i in 1:1000){
    v[i] <- mean(sample(shosho,replace=T))
}
```

Wir können dann die Varianz von v bestimmen:

```
var(v)
[1] 0.0004221173
```

Nun bestimmen wir noch die Bootstrap-Verteilung des Medians.

```
v <- rep(0,1000)
for(i in 1:1000)
    v[i] <- median(sample(shosho,replace=T))
```

Die Varianz ist:

```
var(v)
[1] 0.0005086249
```

Wir sehen, daß die Varianz des Mittelwerts kleiner ist. Wir würden also den Mittelwert dem Median vorziehen.

Um nun einen geeigneten Schätzer für den Lageparameter der Schoschonen-Daten zu finden, führen wir die obigen Schritte für unterschiedliche Werte von α beim α -getrimmten Mittelwert durch.

Wir sollten in diesem Fall aber die Verteilung der Stichprobe symmetrisieren, da nur bei einer symmetrischen Verteilung der getrimmte Mittelwert eine erwartungstreue Schätzung des Erwartungswertes ist. In diesem Fall kann man dann auch die Schätzer hinsichtlich der Varianz vergleichen.

Wie symmetrisiert man die Stichprobe?

Im Buch von Davison und Hinkley findet man die Lösung.

Ist x_1, \dots, x_n die Stichprobe, so bestimmt man den Median $x_{0.5}$ und erweitert die Stichprobe um

$$x_{0.5} - (x_1 - x_{0.5}), \dots, x_{0.5} - (x_n - x_{0.5}).$$

Aus der Stichprobe

$$x_1, \dots, x_n, x_{0.5} - (x_1 - x_{0.5}), \dots, x_{0.5} - (x_n - x_{0.5})$$

zieht man dann die Bootstrapstichproben.

Wir bilden also zunächst die symmetrisierte Stichprobe:

```
shoshos <- c(shosho,2*median(shosho)-shosho)
```

```
shoshos
[1] 0.693 0.662 0.690 0.606 0.570 0.749
[7] 0.672 0.628 0.609 0.844 0.654 0.615
[13] 0.668 0.601 0.576 0.670 0.606 0.611
[19] 0.553 0.933 0.589 0.620 0.592 0.676
[25] 0.712 0.533 0.610 0.654 0.673 0.438
[31] 0.628 0.667 0.614 0.681 0.706 0.612
[37] 0.676 0.671 0.729 0.349
```

Dann ziehen wir wiederholt Stichproben aus der symmetrisierten Stichprobe und bestimmen für jede den 0.05-getrimmten Mittelwert.

```
v <- rep(0,1000)
for(i in 1:1000){
  v[i] <- mean(sample(shoshos,replace=T),trim=0.05)
}
```

Dann schätzen wir die Varianz.

```
var(v)
[1] 0.0001660668
```

Die folgende Tabelle gibt die geschätzten Varianzen wieder:

α	$\widehat{Var}(\bar{x}_\alpha)$
0	0.0003869
0.05	0.0001661
0.10	0.0001106
0.15	0.0000971
0.20	0.0001046
0.25	0.0001140
0.50	0.0002159

Die kleinste Varianz hat der 0.15-getrimmte Mittelwert, was angesichts der Ausreißer schon aus dem Boxplot ersichtlich war.

Die getrimmten Mittelwerte werden auch als L-Schätzer bezeichnet, da sie eine Linearkombination der Orderstatistiken sind.

Wir wollen noch eine weitere Klasse betrachten, die sogenannten M-Schätzer. Das arithmetische Mittel

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

erfüllt die Bedingung

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Wir hätten also das arithmetische Mittel auch so gewinnen können, daß wir von der Beziehung

$$\sum_{i=1}^n (x_i - a) = 0$$

ausgehen und den Wert von a bestimmen, für den diese Beziehung erfüllt ist. Wie können wir diesen Ansatz verallgemeinern?

Sei

$$\psi(t) = t.$$

Dann gilt

$$\sum_{i=1}^n \psi(x_i - \bar{x}) = 0$$

Gegeben sei der Datensatz

```
x <- c(1,4,5,8,17)
```

```
x
[1] 1 4 5 8 17
```

Wir definieren die (simple) Funktion

```
psi.mean <- function(x){
  x
}
```

Wir sehen, daß

$$\hat{\theta} = \bar{x}$$

Lösung von

$$\sum_{i=1}^n \psi(x_i - \hat{\theta}) = 0$$

ist.

```
sum(psi.mean(x-mean(x)))
[1] 0
```

Der Median $x_{0.5}$ erfüllt die Bedingung

$$\sum_{i=1}^n \psi(x_i - x_{0.5}) = 0$$

mit

$$\psi(t) = \begin{cases} -1 & \text{für } t < 0 \\ 0 & \text{für } t = 0 \\ 1 & \text{für } t > 0 \end{cases}$$

Wir bestimmen auch hier die Lösung für den obigen Datensatz mit R .

Wir definieren die Psi-Funktion des Medians:

```
psi.median <- function(x){
  as.numeric((x>0)-(x<0))
}
```

und berechnen

```
sum(psi.median(x-median(x)))
[1] 0
```

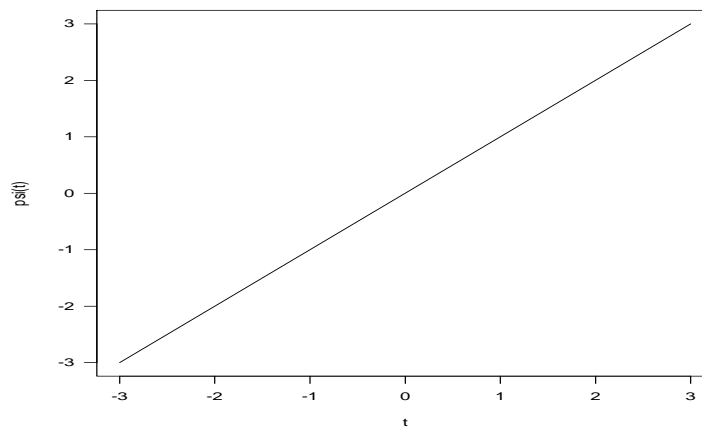
Der Mittelwert und der Median gehören zu einer speziellen Klasse von Schätzern, den sogenannten M-Schätzern.

Diese erfüllen die Bedingung

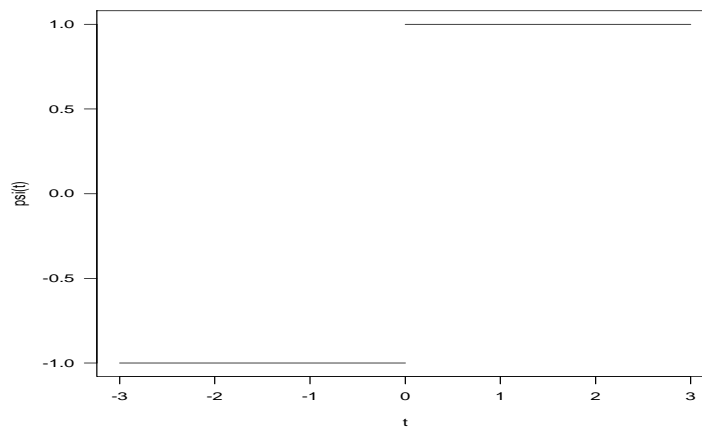
$$\sum_{i=1}^n \psi(x_i - \hat{\theta}) = 0,$$

wobei $\psi(t)$ eine Funktion ist, die als Psi-Funktion bezeichnet wird.

Die Psi-Funktion des Mittelwerts hat folgende Gestalt



und die Psi-Funktion des Medians die nachstehende Gestalt

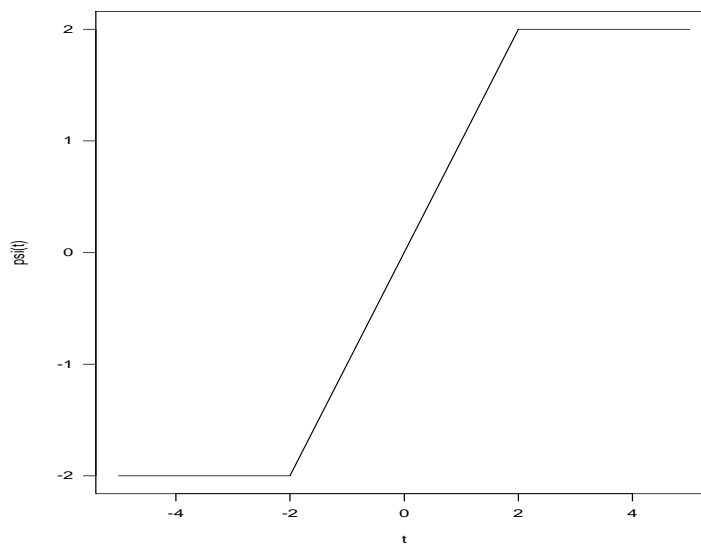


Von Huber wurde nun eine Psi-Funktion vorgeschlagen, die einen Übergang

zwischen Mittelwert und Median beschreibt.

$$\psi(t) = \begin{cases} -k & \text{für } t < -k \\ t & \text{für } -k \leq t \leq k \\ k & \text{für } t > k \end{cases}$$

Wir $k = 2$ ergibt sich folgender Funktionsverlauf.



Am Rand des geordneten Datensatzes gewichtet die Huber-Funktion wie der Median, während sie in der Mitte des Datensatzes wie der Mittelwert gewichtet.

Für $k=0$ erhalten wir den Median und für $k \rightarrow \infty$ ergibt sich der Mittelwert. Wir können uns ganz einfach die Psi-Funktion von Huber als R Funktion schreiben.

```
psi.huber <- function(x,k)
{ pmin(pmax(x, -k), k)
}
```

Wie können wir den Schätzer bestimmen?

Der M-Schätzer $\hat{\theta}$ erfüllt die Gleichung:

$$\sum_{i=1}^n \psi(x_i - \hat{\theta}) = 0,$$

Wir multiplizieren den Ausdruck in der Summe mit

$$\frac{x_i - \hat{\theta}}{x_i - \hat{\theta}}$$

und erhalten

$$\sum_{i=1}^n (x_i - \hat{\theta}) \frac{\psi(x_i - \hat{\theta})}{x_i - \hat{\theta}} = 0,$$

Mit

$$w_i = \frac{\psi(x_i - \hat{\theta})}{x_i - \hat{\theta}}$$

muß also gelten

$$\sum_{i=1}^n (x_i - \hat{\theta}) w_i = 0$$

Diesen Ausdruck können wir nun umformen:

$$\sum_{i=1}^n x_i w_i - \hat{\theta} \sum_{i=1}^n w_i = 0$$

Somit muß gelten

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

Nun hängt w_i aber von $\hat{\theta}$ ab.

Um zu einer Lösung zu kommen, müssen wir die Vorgehensweise iterieren.

Mit

$$w_{i,k-1} = \frac{\psi(x_i - \hat{\theta}_{k-1})}{x_i - \hat{\theta}_{k-1}}$$

erhalten wir als Schätzer auf der k-ten Stufe

$$\hat{\theta}_k = \frac{\sum_{i=1}^n x_i w_{i,k-1}}{\sum_{i=1}^n w_{i,k-1}}$$

wobei $\hat{\theta}_{k-1}$ der Schätzer auf der (k-1)-ten Stufe ist.

Auf der ersten Stufe wählen wir für $\hat{\theta}_0$ den Median der Beobachtungen.

Wir können dieses Verfahren implementieren, müssen aber berücksichtigen, daß $x_i - \hat{\theta}_{k-1}$ den Wert 0 annehmen kann. In diesem Fall würden wir durch 0 dividieren.

Wir setzen hier $w_{i,k-1}$ gleich 1.

Wir erhalten also folgende Funktion

```
w.huber <- function(w, k) {  
  i0 <- w == 0  
  ip <- w != 0  
  w[i0] <- 1  
  w[ip] <- psi.huber(w[ip], k)/w[ip]  
  w }  
}
```

Die folgenden Anweisungen zeigen eine Iterationsfolge.

Dabei ist `te.alt` der alte Schätzer, `te.neu` der neue Schätzer und `wn` der Vektor der Gewichte.

Wir starten mit dem Median.

```
te <- median(x)  
  
wn <- w.huber(x-te.alt,2)  
  
wn  
[1] 0.5000000 1.0000000 1.0000000 0.6666667 0.1666667  
  
te.neu <- sum(wn*x)/sum(wn)  
  
te.neu  
[1] 5.3  
  
te.alt <- te.neu  
  
wn <- w.huber(x-te.alt,2)  
  
wn  
[1] 0.4651163 1.0000000 1.0000000 0.7407407 0.1709402  
  
te.neu <- sum(wn*x)/sum(wn)  
  
te.neu  
[1] 5.418455  
  
te.alt <- te.neu  
  
wn <- w.huber(x-te.alt,2)
```

```
wn
[1] 0.4526469 1.0000000 1.0000000 0.7747300 0.1726885
```

```
ten <- sum(wn*x)/sum(wn)
```

```
te.neu
[1] 5.466422
```

Wir brechen ab und schreiben eine Funktion, die diese Schritte durchführt, bis sich der Schätzer nicht mehr ändert.

```
huber <- function(x, k, diff){
  te.alt <- median(x)
  repeat{
    wn <- w.huber(x - te.alt,k)
    te. neu <- sum(wn * x)/sum(wn)
    if(abs(te.alt - te.neu) < diff)
      break
    te.alt <- te.neu
  }
  te.alt
}
```

Wir rufen die Funktion mit dem Datensatz x für $k = 2$ und $\text{diff}=1e-007$ auf:

```
huber(x,2,1e-007)
[1] 5.5
```

Der Mittelwert und der Median sind skalenäquivalent.

Multiplizieren wir alle Beobachtungen mit einer Konstanten k , so nehmen auch Mittelwert und Median den k -fachen Wert an.

```
x2 <- x*2
```

```
mean(x)
[1] 7
mean(x2)
[1] 14
```

```
median(x)
[1] 5
median(x2)
[1] 10
```

Wie wir an unserem Beispiel feststellen können, besitzen M-Schätzer in der Regel nicht diese Eigenschaft.

huber(x2,2)
[1] 10

Wir müssen bei der Schätzung also einen Skalenparameter S berücksichtigen, d.h. die Lösung von

$$\sum_{i=1}^n \psi\left(\frac{x_i - \hat{\theta}}{S}\right) = 0,$$

zu bestimmen.

Es liegt nahe, als Skalenparameter die Stichprobenstandardabweichung

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

zu verwenden.

Diese ist aber keine robuste Schätzung des Skalenparameters.

Eine robuste Schätzung ist der MAD (median absolute deviation).

Dieser ist definiert durch

$$MAD = \text{Median}\{|x_1 - x_{0.5}|, \dots, |x_n - x_{0.5}|\}$$

wobei $x_{0.5}$ der Median ist.

Bei der Standardnormalverteilung nimmt der MAD den Wert 0.6745 an.

Dies ist gerade der 0.75-Prozentpunkt der Standardnormalverteilung.

Dies kann man folgendermaßen zeigen.

Sei Z standardnormalverteilt.

Gesucht ist die Verteilungsfunktion von $Y = |Z|$.

Es gilt

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(|Z| \leq y) \\ &= P(-y \leq Z \leq y) \\ &= \Phi(y) - \Phi(-y) \\ &= 2\Phi(y) - 1 \end{aligned}$$

Wir suchen den Wert $y_{0.5}$ von y , für den gilt

$$F_Y(y_{0.5}) = 0.5$$

Es muß also gelten

$$2\Phi(y_{0.5}) - 1 = 0.5$$

Hieraus folgt

$$\Phi(y_{0.5}) = 0.75$$

und somit

$$y_{0.5} = \Phi^{-1}(0.75) = 0.6745$$

Um einen konsistenten Schätzer für den MAD bei Normalverteilung zu erhalten, dividieren wir den *MAD* durch 0.6745.

In R ist der MAD in der Funktion `mad` implementiert.

Betrachten wir aber zunächst die einzelnen Schritte bei der Berechnung.

```
x-median(x)
[1] -4 -1  0  3 12
```

```
abs(x-median(x))
[1]  4  1  0  3 12
```

```
median(abs(x-median(x)))
[1] 3
```

```
median(abs(x-median(x)))/0.6745
[1] 4.447739
```

Der Aufruf von `mad` liefert das gleiche Ergebnis:

```
mad(x)
[1] 4.4478
```

Wenn wir noch den Skalenparameter berücksichtigen, müssen wir die Funktion Huber nur an einer Stelle modifizieren. Wir müssen bei der Bestimmung der Gewichte den Skalenparameter berücksichtigen.

$$w_{i,k-1} = \frac{\psi\left(\frac{x_i - \hat{\theta}_{k-1}}{S}\right)}{\frac{x_i - \hat{\theta}_{k-1}}{S}}$$

Dies geschieht durch

```
hubers <- function(x, k, diff){
  te.alt <- median(x)
  s <- mad(x)
  repeat{
    wn <- w.huber((x - te.alt)/s, k)
    te.neu <- sum(wn * x)/sum(wn)
    if(abs(te.alt - te.neu) < diff)
      break
    te.alt <- te.neu
  }
  te.alt
}
```

Für den Datensatz erhalten wir als Huber-Schätzer, wenn wir den MAD berücksichtigen:

```
hubers(x, 2)
[1] 6.7239
```

```
hubers(x2, 2)
[1] 13.4478
```

Für die Shoshonen erhalten wir für $k = 1.5$:

```
huber(shosho, k=1.5)
$mu:
[1] 0.6433751
$s:
[1] 0.0496671
```

und für $k = 2$:

```
huber(shosho, k=2)
$mu:
[1] 0.6468237
$s:
[1] 0.0496671
```


1.4 Nichtparametrische Tests

1.4.1 Der Vorzeichentest

In diesem Abschnitt werden wir uns mit Tests im Einstichprobenproblem beschäftigen, die ohne eine spezielle Verteilungsannahme für die Grundgesamtheit auskommen.

Solche Tests werden auch als **verteilungsfreie** Tests oder **nichtparametrische** Tests bezeichnet.

Der t-Test ist ein Test auf den Erwartungswert einer normalverteilten Grundgesamtheit. Der Erwartungswert muß aber nicht immer existieren. Dies ist zum Beispiel bei der Cauchyverteilung der Fall (siehe dazu Rice: Mathematical Statistics and Data Analysis, S.108).

Ein Lageparameter, der immer existiert ist der Median M .

Für den Median M gilt

$$P(X < M) \leq 0.5 \leq P(X \leq M) \quad (1.2)$$

Wir wollen im folgenden davon ausgehen, daß die Verteilungsfunktion $F_X(x)$ am Median M stetig ist.

Dies hat folgende Konsequenz:

$$\begin{aligned} P(X = M) &= \lim_{x \uparrow M} P(x < X \leq M) \\ &= F_X(M) - F_X(M) \\ &= 0 \end{aligned}$$

Ist $F_X(x)$ nicht stetig in M , so gilt $P(X = M) > 0$.

Schauen wir uns dazu ein Beispiel an:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.4 & \text{für } 0 \leq x < 1 \\ 0.8 & \text{für } 1 \leq x < 2 \\ 1 & \text{für } 2 \leq x \end{cases}$$

Es gilt also

$$P(X < 1) = F_X(0) = 0.4$$

und

$$P(X \leq 1) = F_X(1) = 0.8$$

Also ist der Median gleich 1.

Es gilt $P(X = 1) = 0.4$ und $P(X > 1) = 0.2$.

Ist aber $F_X(x)$ am Median stetig, so gilt immer

$$P(X < M) = P(X > M) = 0.5$$

und

$$P(X = M) = 0.$$

Diese Eigenschaft kann man nun benutzen, um einen einfachen Test auf den Median zu konstruieren.

Wir wollen testen

$$H_0 : M = M_0 \quad \text{gegen} \quad H_1 : M \neq M_0$$

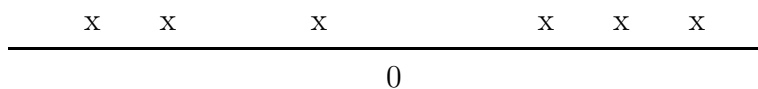
Wenn M_0 der wahre Wert des Medians M in der Grundgesamtheit ist, so erwarten wir, daß die Hälfte der Beobachtungen größer als M_0 ist.

Dies ist die Idee, die hinter dem Vorzeichentest steht.

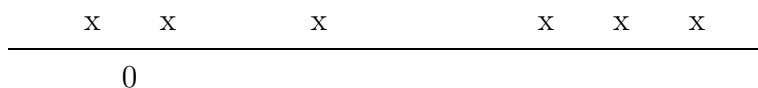
Wir zählen, wie viele der Beobachtungen größer als M_0 sind. Ist diese Anzahl zu groß oder zu klein, so spricht dies dagegen, daß M_0 der Wert des Medians in der Grundgesamtheit ist.

Die folgenden Bilder veranschaulichen die Situation.

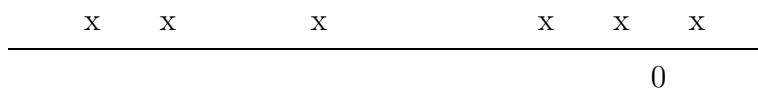
Die folgende Verteilung der Beobachtungen spricht dafür, daß der Median der Grundgesamtheit, aus der die Stichprobe gezogen wurde, gleich 0 ist.



Die folgende Verteilung der Beobachtungen spricht dagegen dafür, daß der Median der Grundgesamtheit, aus der die Stichprobe gezogen wurde, größer als 0 ist.



Die folgende Verteilung der Beobachtungen spricht dafür, daß der Median der Grundgesamtheit, aus der die Stichprobe gezogen wurde, kleiner als 0 ist.



Wir können ohne Beschränkung der Allgemeinheit M_0 gleich 0 setzen. Wollen wir nämlich überprüfen, ob M einen Wert M_0 annimmt, der ungleich 0 ist, so subtrahieren wir M_0 von allen Beobachtungen und zählen wieviele der Beobachtungen größer als 0 sind.

Ist nämlich M_0 der Median von X , so gilt:

$$P(X < M_0) \leq 0.5 \leq P(X \leq M_0)$$

Hieraus folgt

$$P(X - M_0 < 0) \leq 0.5 \leq P(X - M_0 \leq 0)$$

Also ist 0 der Median von $X - M_0$, wenn M_0 der Median von X ist.

Schreiben wir uns nun die Annahmen und die Vorgehensweise des Vorzeichentests formal auf:

Ausgangspunkt sind die Realisationen x_1, \dots, x_n der unabhängigen, identisch mit einer am Median M stetigen Verteilungsfunktion $F_X(x)$ verteilten Zufallsvariablen X_1, \dots, X_n .

Es soll getestet werden:

$$H_0 : M = 0 \quad \text{gegen} \quad H_1 : M \neq 0$$

Wir betrachten die Zufallsvariablen $s(X_1), \dots, s(X_n)$ mit

$$s(x) = \begin{cases} 1 & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Die Funktion $s(x)$ nimmt also den Wert 1 an, wenn x positiv ist, ansonsten ist sie 0.

Unter H_0 gilt

$$P(s(X_i) = 1) = P(X_i > 0) = 0.5$$

und

$$P(s(X_i) = 0) = P(X_i < 0) = 0.5.$$

Die Zufallsvariablen $s(X_1), \dots, s(X_n)$ sind also identisch bernoulliverteilt mit Parameter $p = 0.5$.

Außerdem sind sie unabhängig, da sie Funktionen der unabhängigen Zufallsvariablen X_1, \dots, X_n sind. (siehe dazu Mood, Graybill, Boes: Introduction to the theory of statistics, S.151).

Als Teststatistik wählen wir die Anzahl der Beobachtungen, die größer als 0 sind, d.h.

$$S = \sum_{i=1}^n s(X_i)$$

Diese ist unter H_0 binomialverteilt mit den Parametern n und $p = 0.5$, da $s(X_1), \dots, s(X_n)$ unabhängige, identisch mit Parameter $p = 0.5$ bernoulliverteilte Zufallsvariablen sind.

Es gilt also

$$\begin{aligned} P_{H_0}(S = s) &= \binom{n}{s} 0.5^s \cdot 0.5^{n-s} \\ &= \binom{n}{s} 0.5^n \end{aligned}$$

Zu große bzw. zu kleine Werte von S sprechen nun gegen H_0 . Sind nämlich fast alle Beobachtungen größer als 0, so spricht dies dafür, daß der wahre Wert des Medians größer als 0 ist. Sind hingegen fast alle Beobachtungen kleiner als 0, so spricht dies dafür, daß der wahre Wert des Medians kleiner als 0 ist.

Es gilt

$$\begin{aligned} P_{H_0}(S = n - s) &= \binom{n}{n-s} 0.5^s \cdot 0.5^{n-s} = \\ &= \binom{n}{s} 0.5^s \cdot 0.5^{n-s} \\ &= \binom{n}{s} 0.5^n \end{aligned}$$

Also ist die Verteilung von S unter H_0 symmetrisch bezüglich ihres Erwartungswertes. Somit wählen wir den Ablehnbereich beim zweiseitigen Test symmetrisch.

Wir lehnen also H_0 ab, wenn gilt:

$$S \leq s_{\alpha/2}$$

oder

$$S \geq s_{1-\alpha/2}$$

Dabei sind $s_{\alpha/2}$ und $s_{1-\alpha/2}$ das $\alpha/2$ -Quantil bzw. $1 - \alpha/2$ -Quantil einer Binomialverteilung mit den Parametern n und $p = 0.5$.

Natürlich können auch einseitige Tests durchgeführt werden. Im Testproblem

$$H_0 : M \leq 0 \quad \text{gegen} \quad H_1 : M > 0$$

wird H_0 abgelehnt, wenn S zu groß ist, d.h.

$$S \geq s_{1-\alpha}$$

Im Testproblem

$$H_0 : M \geq 0 \quad \text{gegen} \quad H_1 : M < 0$$

wird H_0 abgelehnt, wenn S zu klein ist, d.h.

$$S \leq s_{\alpha}$$

Mit Hilfe der Funktion `binom.test` können wir den Vorzeichentest durchführen:

```
anz <- sum(shosho>0.618)
n <- length(shosho)
```

```
binom.test(anz,n,0.5)
```

```
Exact binomial test
data:  anz out of n number of successes = 11, n = 20,
p-value =0.8238
alternative hypothesis: true p is not equal to 0.5
```

Die Argumente der Funktion sind:

- `anz`: Anzahl der Beobachtungen, die größer als 0.618 sind
- `n`: Anzahl der Beobachtungen
- `p`: 0.5

Wie wir sehen, sind 11 Beobachtungen größer als 0.618. Die Überschreitungswahrscheinlichkeit beträgt 0.8238.

Dies sieht man folgendermaßen:

$$\begin{aligned} P(S \leq 9) + P(S \geq 11) &= 1 - P(S = 10) \\ &= \binom{20}{10} 0.510 \\ &= 0.8238 \end{aligned}$$

Wir lehnen also H_0 nicht ab.

Für große Werte von n ist es mühsam, die Wahrscheinlichkeiten der Binomialverteilung zu bestimmen. Hier kann man auf die Normalapproximation der Binomialverteilung zurückgreifen.

Es gilt approximativ:

$$S \sim N\left(\frac{n}{2}, \frac{n}{4}\right)$$

Im Testproblem

$$H_0 : M = 0 \quad \text{gegen} \quad H_1 : M \neq 0$$

lehnt man also H_0 ab, wenn gilt

$$S \leq \frac{n}{2} - z_{1-\alpha/2} \frac{\sqrt{n}}{2}$$

oder

$$S \geq \frac{n}{2} + z_{1-\alpha/2} \frac{\sqrt{n}}{2}$$

Im Testproblem

$$H_0 : M \leq 0 \quad \text{gegen} \quad H_1 : M > 0$$

lehnt man also H_0 ab, wenn gilt

$$S \geq \frac{n}{2} + z_{1-\alpha} \frac{\sqrt{n}}{2}$$

Im Testproblem

$$H_0 : M \geq 0 \quad \text{gegen} \quad H_1 : M < 0$$

lehnt man also H_0 ab, wenn gilt

$$S \leq \frac{n}{2} - z_{1-\alpha} \frac{\sqrt{n}}{2}$$

Dabei ist z_p das p -Quantil der Standardnormalverteilung.

In der Praxis kann es vorkommen, daß eine oder mehrere Beobachtungen den Wert 0 annehmen, obwohl die Verteilung stetig ist. Dies dürfte eigentlich nicht passieren.

Ist man nur daran interessiert, daß gilt $P(X > 0) = P(X < 0)$, so besitzen die Beobachtungen, die den Wert 0 annehmen, keine Information für dieses Problem. Also sollte der Test nur für die Beobachtungen durchgeführt werden, die ungleich 0 sind. Man spricht in diesem Fall von einem konditionalen Vorzeichen-test.

Eine genaue Analyse des konditionalen Vorzeichen-tests ist bei Gibbons, Pratt: Concepts of Nonparametric Theory, S. 97-104 zu finden.

1.4.2 Fishers Permutationsprinzip

Den Vorzeichentest kann man auch unter folgendem Aspekt betrachten:

Wir gehen aus, daß $H_0 : M = 0$ zutrifft.

Dann gilt

$$P(X > 0) = P(X < 0) = 0.5.$$

Also ist das Vorzeichen jeder Beobachtung a priori mit Wahrscheinlichkeit 0.5 positiv und mit Wahrscheinlichkeit 0.5 negativ.

Da die Zufallsvariablen X_1, \dots, X_n unabhängig sind, sind auch die Vorzeichen der X_i unabhängig.

Sei $V = (V_1, \dots, V_n)$ der Vektor der Vorzeichen.

Dann gilt

$$P(V = v) = 0.5^n$$

Sei speziell $n = 4$.

Dann gibt es $2^4 = 16$ mögliche Vorzeichenkonfigurationen:

```

+ + + +
- + + +
+ - + +
+ + - +
+ + + -
- - + +
- + - +
- + + -
+ - + -
+ + - -
- - - +
- - + -
- + - -
+ - - -
- - - -

```

Jede dieser Konfigurationen hat unter $H_0 : M = 0$ die gleiche Wahrscheinlichkeit. Unter diesen Konfigurationen sprechen einige gegen die Nullhypothese und der Rest dafür. Über eine geeignete Teststatistik werden nun die Konfigurationen in solche aufgeteilt, die für die Nullhypothese sprechen und solche, die gegen die Nullhypothese sprechen.

Im Fall des Vorzeichentests ist die Teststatistik S die Anzahl der positiven Vorzeichen.

Dies kann man auch dadurch sehen, daß man das Plus durch eine Eins und das Minus durch eine Null ersetzt. Den Wert der Teststatistik erhält man als Summe der Nullen und Einsen. Dies entspricht der Anzahl der Einsen und somit der Anzahl der positiven Beobachtungen.

```

1 1 1 1 4
0 1 1 1 3
1 0 1 1 3
1 1 0 1 3
1 1 1 0 3
0 0 1 1 2
0 1 0 1 2
0 1 1 0 2
1 0 0 1 2
1 0 1 0 2
1 1 0 0 2
0 0 0 1 1
0 0 1 0 1
0 1 0 0 1
1 0 0 0 1
0 0 0 0 0

```

Die Verteilung der Teststatistik S unter H_0 ist somit:

s	P(S=s)
0	0.0625
1	0.2500
2	0.3750
3	0.2500
4	0.0625

Diese Verteilung hätten wir auch direkt über die Binomialverteilung bestimmen können:

$$P(S = s) = \binom{n}{s} 0.5^s \quad \text{für } s = 0, 1, 2, 3, 4$$

Der Vorzeichenstest ist ein spezieller Permutationstest. Diese wurden von Fisher vorgeschlagen.

Der Vorzeichenstest benutzt nur die Vorzeichen der Beobachtungen.

Die Verteilung der Teststatistik hängt nicht von der Verteilung der Grundgesamtheit ab, wenn die Nullhypothese zutrifft.

Es stellt sich die Frage, ob nicht auch die Werte der Beobachtungen verwendet werden können, um einen Test zu erhalten, bei dem die Verteilung der Teststatistik nicht von der Verteilung der Grundgesamtheit abhängt.

Ohne zusätzliche Annahmen über die Verteilung der Grundgesamtheit ist dies nicht möglich.

Nimmt man jedoch an, daß die Verteilung der Grundgesamtheit symmetrisch bezüglich 0 ist, so kann man verteilungsfreie Tests konstruieren.

Die Verteilungsfunktion einer stetigen Zufallsvariablen ist symmetrisch bezüglich 0, wenn für alle $x \in \mathfrak{R}$ gilt

$$F_X(x) = 1 - F_X(-x)$$

Für die Dichtefunktion gilt somit für alle $x \in \mathfrak{R}$

$$f_X(x) = f_X(-x)$$

Die Standardnormalverteilung erfüllt diese Bedingung.

Ist nun die Verteilung von X symmetrisch bezüglich 0, so ist das Vorzeichen einer Beobachtung unabhängig von ihrem Abstand vom Nullpunkt.

Dies gilt aufgrund des folgenden Satzes:

Satz 1.4.1 *Die Zufallsvariable X sei stetig und besitze die Verteilungsfunktion $F_X(x)$ mit*

$$F_X(x) = 1 - F_X(-x)$$

Außerdem sei

$$s(x) = \begin{cases} 1 & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Dann sind $|X|$ und $s(X)$ unabhängig.

Beweis:

Wir müssen zeigen

$$P(|X| \leq x, s(X) = 1) = P(|X| \leq x) \cdot P(s(X) = 1)$$

und

$$P(|X| \leq x, s(X) = 0) = P(|X| \leq x) \cdot P(s(X) = 0)$$

Wir zeigen nur die erste Beziehung. Die zweite ergibt sich analog.

$$\begin{aligned} P(|X| \leq x, s(X) = 1) &= P(-x \leq X \leq x, X > 0) \\ &= P(0 < X \leq x, X > 0) \\ &= F_X(x) - F_X(0) \\ &= F_X(x) - 0.5 \\ &= 0.5 \cdot (2F_X(x) - 1) \\ &= P(s(X) = 1) \cdot P(|X| \leq x) \end{aligned}$$

da aufgrund der Symmetrie gilt

$$\begin{aligned} P(|X| \leq x) &= P(-x \leq X \leq x) \\ &= F_X(x) - F_X(-x) \\ &= F_X(x) - (1 - F_X(x)) \\ &= 2F_X(x) - 1 \end{aligned}$$

Dieser Satz hat nun folgende wichtige Konsequenz:

Da die Vorzeichen und Absolutbeträge der Beobachtungen bei einer symmetrischen Verteilung unabhängig sind, ist jede Verteilung der Vorzeichen auf die Absolutbeträge der Beobachtungen gleichwahrscheinlich.

Wir können diese Eigenschaft benutzen, um mit Hilfe der Daten einen verteilungsfreien Test zu konstruieren.

Schauen wir uns dazu ein Beispiel an:

Es soll überprüft werden, ob der folgende Datensatz aus einer Verteilung mit Median 0 stammt, wobei wir unterstellen, daß diese Verteilung symmetrisch ist:

3 2 -5 6

Wir betrachten sämtliche Möglichkeiten, die Vorzeichen auf die Absolutbeträge der Beobachtungen zu verteilen und geben außerdem zu jeder Stichprobe den Wert von \bar{X} an.

-2	-3	-5	-6	-4
-2	-3	-5	6	-1
-2	-3	5	-6	-1.5
-2	-3	5	6	1.5
-2	3	-5	-6	-2.5
-2	3	-5	6	0.5
-2	3	5	-6	0
-2	3	5	6	3
2	-3	-5	-6	-3
2	-3	-5	6	0
2	-3	5	-6	-0.5
2	-3	5	6	2.5
2	3	-5	-6	-1.5
2	3	-5	6	1.5
2	3	5	-6	1
2	3	5	6	4

Nun können wir die Verteilung von \bar{X} unter $H_0 : M = 0$ durch Auszählen bestimmen.

Es gilt

\bar{x}	$P(\bar{X} = \bar{x})$
-4	0.0625
-3	0.0625
-2.5	0.0625
-1.5	0.1250
-1	0.0625
-0.5	0.0625
0	0.1250
0.5	0.0625
1	0.0625
1.5	0.1250
2.5	0.0625
3	0.0625
4	0.0625

Für die obige Stichprobe

$$3 \quad 2 \quad -5 \quad 6$$

nimmt \bar{X} den Wert 1.5 an.

Also gilt für die Überschreitungswahrscheinlichkeit:

$$\begin{aligned} P_{H_0}(|\bar{X}| \geq 1.5) &= 2(0.125 + 0.0625 + 0.0625 + 0.0625) \\ &= 0.625 \end{aligned}$$

Also lehnen wir zum Signifikanzniveau $\alpha = 0.05$ nicht ab, daß der Median gleich 0 ist.

Bei einer symmetrischen Verteilung der Grundgesamtheit bietet Fishers Permutationstest die Möglichkeit, unter Berücksichtigung der Werte der Beobachtungen einen verteilungsfreien Test auf den Median durchzuführen. Der Test hat jedoch den Nachteil, daß die Verteilung der Teststatistik für jeden Datensatz neu bestimmt werden muß. Es gibt 2^n unterschiedliche Möglichkeiten, die Vorzeichen auf die Absolutbeträge der Beobachtungen zu verteilen. Für großes n ist es sehr mühselig, die Permutationsverteilung zu bestimmen.

Der Übergang zu Rängen erlaubt es nun, einen Test anzugeben, der die Vorteile von Fishers Permutationstest aufweist, bei dem aber die Verteilung der Teststatistik nicht für jeden Datensatz neu bestimmt werden muß.

1.4.3 Ränge

Sei x_1, \dots, x_n eine Stichprobe vom Umfang n , und sei $x_{(1)}, \dots, x_{(n)}$ die geordnete Stichprobe mit $x_{(1)} \leq \dots \leq x_{(n)}$.

Wir gehen zunächst davon aus, daß keine identischen Beobachtungen in der Stichprobe vorliegen. Es gilt also $x_{(1)} < \dots < x_{(n)}$. In der Statistik spricht man davon, daß keine Bindungen auftreten.

Der Rang r_i von x_i gibt an, an welcher Position x_i in der geordneten Stichprobe steht, d.h. wieviele der Beobachtungen kleiner oder gleich x_i sind.

Beispiel 1.4.1 *Die Stichprobe lautet*

$$x_1 = 49.3 \quad x_2 = 48.7 \quad x_3 = 48.1 \quad x_4 = 48.6 \quad x_5 = 48.2$$

Die geordnete Stichprobe ist dann

$$x_{(1)} = 48.1 \quad x_{(2)} = 48.2 \quad x_{(3)} = 48.6 \quad x_{(4)} = 48.7 \quad x_{(5)} = 49.3$$

Da $x_1 = 49.3$ an der fünften Stelle in der geordneten Stichprobe steht, gilt also $r_1 = 5$. Wir können aber auch einfach nur zählen, wieviele der Beobachtungen kleiner oder gleich 49.3 sind und erhalten den gleichen Wert für den Rang von x_1 .

Analog erhalten wir $r_2 = 4, r_3 = 1, r_4 = 3$ und $r_5 = 2$.

Wir bilden aus den Rängen den Rangvektor $r = (r_1, \dots, r_n)$.

Wenn keine Bindungen vorliegen, ist die Rangvergabe eindeutig, und der Rangvektor ist eine Permutationen der natürlichen Zahlen $\{1, \dots, n\}$.

Der Rangvektor hängt von den Realisationen der Zufallsvariablen X_1, \dots, X_n ab und ist somit vor der Erhebung eine Zufallsvariable, die wir mit

$$R = (R_1, \dots, R_n)$$

bezeichnen.

Es gilt nun folgender

Satz 1.4.2 *Sind die Zufallsvariablen X_1, \dots, X_n unabhängig und identisch verteilt mit stetiger Verteilungsfunktion $F_X(x)$, so gilt*

$$P(R = r) = \frac{1}{n!}$$

Der Beweis ist z. B. bei Randles, Wolfe: Introduction to the theory of nonparametric statistics, S.37-38 zu finden.

Dieser Satz hat einige Konsequenzen für die angewandte Statistik.

Die wichtigste Konsequenz des Satzes ist:

Die Verteilung des Rangvektors und somit jeder Funktion des Rangvektors hängt nicht von der Verteilung der Grundgesamtheit ab, wenn diese stetig ist.

Man kann also mit Hilfe der Ränge verteilungsfreie Tests gewinnen. Wir werden in den folgenden Abschnitten immer wieder auf diese Eigenschaft des Satzes zurückgreifen.

Weitere Folgerungen gibt der nachstehende Satz an:

Satz 1.4.3 Sei $R = (R_1, \dots, R_n)$ der Rangvektor einer Zufallstichprobe aus einer stetigen Grundgesamtheit.

Dann gilt

1. $P(R_i = k) = \frac{1}{n}$ für $i = 1, \dots, n$ und $k = 1, \dots, n$
2. $P(R_i = k, R_j = l) = \frac{1}{n(n-1)}$ für $1 \leq i, j, k, l \leq n, i \neq j, k \neq l$
3. $E(R_i) = \frac{n+1}{2}$ für $i = 1, \dots, n$
4. $\text{Var}(R_i) = \frac{n^2-1}{12}$ für $i = 1, \dots, n$
5. $\text{Cov}(R_i, R_j) = -\frac{n+1}{12}$ für $i = 1, \dots, n, j = 1, \dots, n, i \neq j$
6. $\text{Corr}(R_i, R_j) = -\frac{1}{n-1}$ für $i = 1, \dots, n, j = 1, \dots, n, i \neq j$

Beweis:

1. Es gibt $(n-1)!$ Rangvektoren, bei denen gilt $R_i = k$. Da alle Rangvektoren gleichwahrscheinlich sind, gilt

$$P(R_i = k) = \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl möglicher Fälle}} = \frac{(n-1)!}{n!} = \frac{1}{n}$$

2. Es gibt $(n-2)!$ Rangvektoren, bei denen gilt $R_i = k$ und $R_j = l$.

Also gilt

$$P(R_i = k, R_j = l) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

3.

$$\begin{aligned}
 E(R_i) &= \sum_{k=1}^n k P(R_i = k) = \\
 &= \sum_{k=1}^n k \frac{1}{n} = \\
 &= \frac{1}{n} \frac{n(n+1)}{2} = \\
 &= \frac{n+1}{2}
 \end{aligned}$$

4.

$$\text{Var}(R_i) = E(R_i^2) - E(R_i)^2$$

Es gilt

$$\begin{aligned}
 E(R_i^2) &= \sum_{k=1}^n k^2 P(R_i = k) = \\
 &= \sum_{k=1}^n k^2 \frac{1}{n} = \\
 &= \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \\
 &= \frac{(n+1)(2n+1)}{6}
 \end{aligned}$$

Also gilt

$$\begin{aligned}
 \text{Var}(R_i) &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \\
 &= \frac{n+1}{2} \left(\frac{2n+1}{3} - \frac{n+1}{2}\right) = \\
 &= \frac{n+1}{2} \left(\frac{2n+1}{3} - \frac{n+1}{2}\right) = \\
 &= \frac{n^2 - 1}{12}
 \end{aligned}$$

5.

$$\text{Cov}(R_i, R_j) = E(R_i R_j) - E(R_i) E(R_j)$$

Es gilt

$$\begin{aligned} E(R_i R_j) &= \sum_{k=1}^n \sum_{l=1}^n k l P(R_i = k, R_j = l) \\ &= \sum_{k \neq l} k l \frac{1}{n(n-1)} = \\ &= \frac{1}{n(n-1)} \left(\sum_{k=1}^n \sum_{l=1}^n k l - \sum_{k=1}^n k^2 \right) \\ &= \frac{1}{n(n-1)} \left(\sum_{k=1}^n k \sum_{l=1}^n l - \frac{n(n+1)(2n+1)}{6} \right) \\ &= \frac{1}{n(n-1)} \left(\frac{(n+1)^2 n^2}{4} - \frac{n(n+1)(2n+1)}{6} \right) \\ &= \frac{n(n+1)}{n(n-1)} \left(\frac{(n+1)n}{4} - \frac{2n+1}{6} \right) \\ &= \frac{n+1}{n-1} \frac{3n^2 - n - 2}{12} \\ &= \frac{(n+1)(3n+2)}{12} \end{aligned}$$

Also gilt

$$\text{Cov}(R_i, R_j) = \frac{(n+1)(3n+2)}{12} - \frac{(n+1)^2}{4} = -\frac{n+1}{12}$$

6.

$$\text{Corr}(R_i, R_j) = \frac{\text{Cov}(R_i, R_j)}{\sqrt{\text{Var}(R_i)\text{Var}(R_j)}} = -\frac{1}{n-1}$$

In der Praxis findet man oft gebundene Beobachtungen in der Stichprobe.
In der Stichprobe

$$x_1 = 3.7 \quad x_2 = 1.5 \quad x_3 = 2.4 \quad x_4 = 3.7 \quad x_5 = 2.4$$

kommen der Wert 3.7 und der Wert 2.4 jeweils zweimal vor.

Bei Bindungen ist die Rangzuweisung nicht eindeutig.

Es gibt nun eine Reihe von Vorschlägen, wie man vorgehen soll, wenn Bindungen vorliegen.

- Es werden so lange Beobachtungen aus der Stichprobe entfernt, bis keine Bindungen mehr vorliegen.

Beim obigen Beispiel würde dies die Stichprobe

$$3.7 \quad 1.5 \quad 2.4$$

liefern.

Diese Vorgehensweise ist natürlich nur dann sinnvoll, wenn nicht zu viele Beobachtungen aus der Stichprobe entfernt werden müssen.

- Den gebundenen Beobachtungen werden zufällig Ränge zugeordnet.

Für die obige Stichprobe wäre eine zufällige Rangzuweisung

$$r_1 = 5 \quad r_2 = 1 \quad r_3 = 2 \quad r_4 = 4 \quad r_5 = 3$$

- Es wird das arithmetische Mittel der Rangzahlen bestimmt, die den gebundenen Werten insgesamt zugeordnet werden.

Im Beispiel sind die erste und vierte Beobachtung identisch und erhalten somit den Rang $\frac{4+5}{2} = 4.5$.

Die dritte und fünfte Beobachtung sind identisch und erhalten somit den Rang $\frac{2+3}{2} = 2.5$.

Für die obige Stichprobe ergeben sich also folgende Durchschnittsränge

$$r_1 = 4.5 \quad r_2 = 1 \quad r_3 = 2.5 \quad r_4 = 4.5 \quad r_5 = 2.5$$

In der Praxis werden in der Regel Durchschnittsränge vergeben.

In R gibt es eine Funktion `rank`, die für einen Datenvektor `x` den Vektor der Ränge liefert. Im Falle von Bindungen werden Durchschnittsränge bestimmt. Schauen wir uns diese Funktion für die beiden Datensätze an:

Der erste Datensatz ergibt folgendes Ergebnis

```
x1 <- c(49.3,48.7,48.1,48.6,48.2)
rank(x1)
[1] 5 4 1 3 2
```

und der zweite Datensatz folgendes Ergebnis

```
x2 <- c(3.7,1.5,2.4,3.7,2.4)
rank(x2)
[1] 4.5 1.0 2.5 4.5 2.5
```

1.4.4 Der Vorzeichen-Rangtest

Wir betrachten wiederum das Testproblem

$$H_0 : M = 0 \quad \text{gegen} \quad H_1 : M \neq 0.$$

Der Vorzeichentest benutzt von jeder Beobachtung nur ihr Vorzeichen. Der Permutationstest von Fisher benutzt außerdem noch den Betrag jeder Beobachtung. Die Teststatistik ist dabei der Mittelwert der mit den Vorzeichen versehenen Absolutbeträge der Beobachtungen:

$$T = \frac{1}{n} \left(\sum_{i=1}^n s(x_i) |x_i| - \sum_{i=1}^n s(-x_i) |x_i| \right)$$

da gilt

$$s(-x) = \begin{cases} 1 & \text{für } x < 0 \\ 0 & \text{für } x \geq 0 \end{cases}$$

Da gilt

$$\sum_{i=1}^n |x_i| = \sum_{i=1}^n s(x_i) |x_i| + \sum_{i=1}^n s(-x_i) |x_i|$$

folgt

$$\begin{aligned} T &= \frac{1}{n} \left(\sum_{i=1}^n s(x_i) |x_i| - \sum_{i=1}^n |x_i| + \sum_{i=1}^n s(x_i) |x_i| \right) \\ &= \frac{1}{n} \left(2 \sum_{i=1}^n s(x_i) |x_i| - \sum_{i=1}^n |x_i| \right) \\ &= \frac{2}{n} \sum_{i=1}^n s(x_i) |x_i| - \frac{1}{n} \sum_{i=1}^n |x_i| \end{aligned}$$

Da

$$\sum_{i=1}^n |x_i|$$

konstant ist, ist die Teststatistik

$$S = \sum_{i=1}^n s(x_i) |x_i|$$

äquivalent zu T.

Beide Teststatistiken hängen linear voneinander ab und lehnen also für die gleichen Stichproben ab.

Wilcoxon hat 1945 vorgeschlagen, in der Teststatistik S die Absolutbeträge der Beobachtungen durch die Ränge der Absolutbeträge zu ersetzen.

Er bildet also die Teststatistik

$$W^+ = \sum_{i=1}^n s(x_i) R(|x_i|)$$

Dabei ist $R(|x_i|)$ der Rang von $|x_i|$ unter $|x_1|, \dots, |x_n|$.

Beispiel 1.4.2 *Wir betrachten folgende Stichprobe:*

$$x_1 = 3 \quad x_2 = 2 \quad x_3 = -5 \quad x_4 = 6$$

Die Vorzeichen der Beobachtungen sind

$$s(x_1) = 1 \quad s(x_2) = 1 \quad s(x_3) = 0 \quad s(x_4) = 1$$

Die Absolutbeträge der Beobachtungen lauten

$$|x_1| = 3 \quad |x_2| = 2 \quad |x_3| = 5 \quad |x_4| = 6$$

Somit erhalten wir folgende Ränge der Absolutbeträge

$$R(|x_1|) = 2 \quad R(|x_2|) = 1 \quad R(|x_3|) = 3 \quad R(|x_4|) = 4$$

Also gilt

$$W^+ = 1 \cdot 2 + 1 \cdot 1 + 0 \cdot 3 + 1 \cdot 4 = 7$$

Die Verteilung von W^+ unter H_0 kann man durch Auszählen bestimmen.

Betrachten wir dazu den Fall $n = 4$.

Es gibt $2^4 = 16$ unterschiedlichen Teilmengen der Menge $\{1, 2, 3, 4\}$. Jede dieser Teilmengen beschreibt eine Konfiguration positiver Beobachtungen. So liegt die leere Menge \emptyset vor, wenn keine Beobachtung positiv ist, während $\{2, 3\}$ vorliegt, wenn die zweite und die dritte Beobachtung positiv ist.

Alle Möglichkeiten mit dem zugehörigen Wert von W^+ sind in der folgenden Tabelle zu finden:

Teilmenge	Wert von W^+
\emptyset	0
{1}	1
{2}	2
{3}	3
{4}	4
{1, 2}	3
{1, 3}	4
{1, 4}	5
{2, 3}	5
{2, 4}	6
{3, 4}	7
{1, 2, 3}	6
{1, 2, 4}	7
{1, 3, 4}	8
{2, 3, 4}	9
{1, 2, 3, 4}	10

Somit erhalten wir folgende Verteilung von :

w	$P(W^+ = w)$
0	0.0625
1	0.0625
2	0.0625
3	0.1250
4	0.1250
5	0.1250
6	0.1250
7	0.1250
8	0.0625
9	0.0625
10	0.0625

Da wir einen zweiseitigen Test der Hypothese $H_0 : M = 0$ durchführen, beträgt die Überschreitungswahrscheinlichkeit

$$P(W^+ \geq 7) + P(W^+ \leq 3) = 0.625.$$

In R wenden wir die Funktion `wilcox.test` an.

Der Aufruf entspricht dem Aufruf der Funktion `t.test`.

```
x <- c(3,2,-5,6)
wilcox.test(x)
  Exact Wilcoxon signed-rank test
data:  x
signed-rank statistic V = 7, n = 4, p-value = 0.625
alternative hypothesis: true mu is not equal to 0
```

Wir lehnen H_0 also nicht ab.

Schauen wir uns die Verteilung der Teststatistik W^+ noch genauer an.

Da die Zufallsvariablen X_1, \dots, X_n unabhängig sind, sind auch die Zufallsvariablen $s(X_1), \dots, s(X_n)$ unabhängig.

Unter H_0 gilt

$$P(s(X_i) = 0) = P(s(X_i) = 1) = 0.5$$

Somit sind $s(X_1), \dots, s(X_n)$ unabhängige, identisch mit Parameter $p = 0.5$ bernoulliverteilte Zufallsvariablen.

Außerdem sind unter H_0 die $s(X_i)$ und die $|X_i|$ und somit auch die $R(|X_i|)$ unabhängig.

Die Vorzeichen sind also unabhängig von den Rängen. Da es egal ist, in welcher Reihenfolge wir summieren, können wir auch die sortierten Ränge mit den Vorzeichen multiplizieren und aufsummieren.

Wir erhalten somit folgende Darstellung der Teststatistik, die die Bestimmung von Erwartungswert und Varianz erleichtert:

$$W^+ = \sum_{i=1}^n i s(X_i)$$

Die Teststatistik ist also eine Linearkombination von unabhängigen, identisch mit Parameter $p = 0.5$ bernoulliverteilten Zufallsvariablen. Wir erhalten also als Erwartungswert

$$\begin{aligned} E(W^+) &= E\left(\sum_{i=1}^n i s(X_i)\right) \\ &= \sum_{i=1}^n E(i s(X_i)) \\ &= \sum_{i=1}^n i E(s(X_i)) \\ &= \sum_{i=1}^n i 0.5 = \frac{n(n+1)}{4} \end{aligned}$$

und als Varianz

$$\begin{aligned}
 \text{Var}(W^+) &= \text{Var}\left(\sum_{i=1}^n i s(X_i)\right) \\
 &= \sum_{i=1}^n \text{Var}(i s(X_i)) \\
 &= \sum_{i=1}^n i^2 \text{Var}(s(X_i)) \\
 &= \sum_{i=1}^n i^2 \cdot 0.25 \\
 &= \frac{n(n+1)(2n+1)}{24}
 \end{aligned}$$

Man kann zeigen, daß W^+ unter H_0 approximativ normalverteilt ist.

Liegen Bindungen vor, so werden Durchschnittsränge bestimmt, bei der Normalapproximation wird die übliche Stetigkeitskorrektur verwendet, und die Varianz muß korrigiert werden.

Die Formel der korrigierten Varianz lautet

$$\text{Var}(W^+) = \frac{n(n+1)(2n+1)}{24} - \frac{1}{48} \sum_{j=1}^r (b_j^3 - b_j)$$

Dabei ist r die Anzahl der Gruppen mit Bindungen und b_j die Anzahl der Beobachtungen in der j -ten Bindungsgruppe.

Somit ist folgende Größe approximativ standardnormalverteilt

$$\frac{W^+ - 0.5 - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{1}{48} \sum_{j=1}^r (b_j^3 - b_j)}}$$

Ein Beispiel soll die Vorgehensweise verdeutlichen.

Es soll getestet werden, ob die folgende Stichprobe aus einer Grundgesamtheit mit Median 0 kommt:

```
3 5 -1 2 2 1 5 -4 2 -3 3 1 3 -1 5 -4 -4 2 -3 5
```

Wir bestimmen zunächst die Absolutbeträge der Beobachtungen:

```
3 5 1 2 2 1 5 4 2 3 3 1 3 1 5 4 4 2 3 5
```

Es gibt fünf Bindungsgruppen, nämlich

```
1 2 3 4 5
```

Also ist $r = 5$. Die 1 tritt unter den Absolutbeträgen 4-mal auf.

Also ist $b_1 = 4$.

Entsprechend erhalten wir

$$b_2 = 4 \quad b_3 = 5 \quad b_4 = 3 \quad b_5 = 4$$

Da gilt $n = 20$, erhalten wir folgende korrigierte Varianz

$$\frac{20 \cdot 21 \cdot 41}{24} - \frac{1}{48} [(4^3 - 4) + (4^3 - 4) + (5^3 - 5) + (3^3 - 3) + (4^3 - 4)]$$

Schauen wir uns die Vorgehensweise in R an.

Wir geben die Daten ein.

```
x <- c(3,5,-1,2,2,1,5,-4,2,-3,3,1,3,-1,5,-4,-4,2,-3,5)
```

Dann bestimmen wir die Ränge der Absolutbeträge

```
rank(abs(x))
[1] 11.0 18.5 2.5 6.5 6.5 2.5 18.5 15.0 6.5 11.0
    11.0 2.5 11.0 2.5 18.5 15.0 15.0 6.5 11.0 18.5
```

Der Wert von W^+ ist dann

```
sum((x>0)*rank(abs(x)))
[1] 138
```

Der Erwartungswert ist

```
n <- length(x)
n*(n+1)/4
[1] 105
```

Da die Ränge der Absolutbeträge bestimmt werden, müssen wir die Bindungsgruppen der Absolutbeträge bestimmen.

Mit Hilfe der Funktion `table` erhalten wir die absoluten Häufigkeiten:

```
h <- table(abs(x))
h
 1 2 3 4 5
4 4 5 3 4
```

Der Korrekturfaktor für die Varianz ist dann

```
sum(h^3-h)/48
[1] 6.75
```

Somit ist die Varianz gegeben durch

```
n*(n+1)*(2*n+1)/24-sum(h^3-h)/48
[1] 710.75
```

Die standardisierte Teststatistik ist somit

```
(138-105.5)/sqrt(710.75)
[1] 1.21906
```

und die Überschreitungswahrscheinlichkeit für den zweiseitigen Test ist

```
2*pnorm(abs((138-105.5)/sqrt(710.75)))
[1] 0.2228216
```

In R erhalten wir:

```
wilcox.test(x)
  Wilcoxon signed-rank test
data:  x
signed-rank normal statistic with correction Z = 1.2191,
p-value = 0.2228
alternative hypothesis: true mu is not equal to 0
Warning messages:
 cannot compute exact p-value with ties in:
 wil.sign.rank(dff, alternative, exact, correct)
```

Wir lehnen H_0 also nicht ab.

Wir wenden nun den Test auf die Shoshonen-Daten an.

```
wilcox.test(shosho,mu=0.618)
  Wilcoxon signed-rank test
data:  shosho
signed-rank normal statistic with correction Z =1.6988,
p-value = 0.0894
alternative hypothesis:
true mu is not equal to 0.618
Warning messages:
 cannot compute exact p-value with ties in:
 wil.sign.rank(dff, alternative, exact, correct)
```

Die Überschreitungswahrscheinlichkeit beträgt 0.0894, so daß zum Niveau $\alpha = 0.05$ die Nullhypothese nicht abgelehnt wird.

Die Funktion gibt auch eine Warnung, daß nicht die exakte Verteilung bestimmt wurde, da Bindungen vorliegen.

Bei den Rechtecken der Shoshonen ergeben sich folgende Überschreitungswahrscheinlichkeiten

t-Test	0.0539
Vorzeichentest	0.8238
Vorzeichen-Rangtest	0.0894

Alle drei Tests kommen zur identischen Entscheidung zum Niveau 0.05.

Dies muß wie wir im nächsten Kapitel sehen werden nicht bei allen Datensätzen der Fall sein.

Es stellt sich also die Frage, wann man welchen Test verwenden soll.

1.4.5 Vergleich der Test

Mit dem t-Test, dem Vorzeichen- und dem Vorzeichenrangtest haben wir drei Tests auf Lage im Einstichprobenproblem kennengelernt.

Es stellt sich die Frage, welchen Test man durchführen soll.

Der t-Test ist bei Normalverteilung am besten.

Wie sieht es aber mit den anderen Tests aus?

Ein guter Test sollte einen Lageunterschied mit einer großen Wahrscheinlichkeit aufdecken.

Im Testproblem

$$H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta = \theta_1$$

messen wir diese Wahrscheinlichkeit mit der Gütefunktion

$$P(\text{Entscheidung für } H_1 | \theta \text{ nimmt den Wert } \theta_1 \text{ an})$$

Im Testproblem

$$H_0 : M \geq 0 \quad \text{gegen} \quad H_1 : M < 0$$

betrachten wir einen kleinen Lageunterschied von $k \cdot \sigma$.

Wir gehen also davon aus, daß der wahre Wert vom M nicht 0, sondern $-k \cdot \sigma$ beträgt.

Mit Hilfe einer Simulation wollen wir die Gütefunktion der drei Tests an den Stellen $k = 0, 0.1, \dots, 1$ schätzen, wobei wir die folgenden drei Verteilungen zugrundelegen:

- Gleichverteilung auf $(-0.5, 0.5)$
- Normalverteilung
- Cauchyverteilung

Die Verteilungen decken ein großes Spektrum symmetrischer Verteilungen ab.

Die Dichtefunktion der Gleichverteilung auf $(-0.5, 0.5)$ ist gegeben durch:

$$f_X(x) = \begin{cases} 1 & \text{für } -0.5 < x < 0.5 \\ 0 & \text{sonst} \end{cases}$$

Die Dichtefunktion der Cauchyverteilung ist gegeben durch

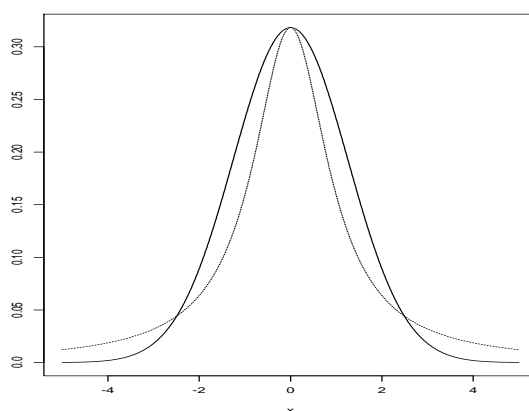
$$f_X(x) = \frac{1}{\pi \beta \left(1 + \left(\frac{x - \alpha}{\beta} \right)^2 \right)}$$

Der Parameter α ist ein Lageparameter und der Parameter β ist ein Skalensparameter.

Wir setzen in der Regel $\alpha = 0$ und $\beta = 1$.

Die nachstehende Graphik zeigt die Dichtefunktion der Standardnormalverteilung (durchgezogene Linie) und die Dichtefunktion der Cauchyverteilung mit $\alpha = 0$ und $\beta = \sqrt{2/\pi}$.

Der Parameter β wurde so gewählt, daß die Dichtefunktionen an der Stelle 0 die gleiche Höhe haben und so besser verglichen werden können.



Die Cauchyverteilung verläuft im Zentrum steiler als die Normalverteilung. Außerdem besitzt die Cauchyverteilung mehr Wahrscheinlichkeitsmasse an den Rändern. Deshalb treten extreme Beobachtungen bei der Cauchyverteilung häufiger auf als bei der Normalverteilung.

Wir wollen uns nun Gedanken über den Aufbau der Simulationsstudie machen.

Ziel ist eine Schätzung der Gütefunktion an den Stellen $-k \cdot \sigma$ mit $k = 0, 0.1, \dots, 1$ beim Stichprobenumfang $n = 10$.

Bei diesem kleinen Stichprobenumfang ist es beim Vorzeichentest und beim Vorzeichen-Rangtest notwendig zu randomisieren, da sonst die Tests nicht vergleichbar sind.

Hierzu benötigen wir die Verteilung der Teststatistik, wenn die Nullhypothese zutrifft.

Beim Vorzeichen-Rangtest finden wir diese mit der Funktion `psignrank`. Diese berechnet den Wert der Verteilungsfunktion der Teststatistik an der Stelle x .

Schauen wir uns dazu exemplarisch den Fall $n = 4$ an, den wir durch Auszählen bestimmt haben.

Beim Stichprobenumfang n ist der größte Wert der Teststatistik gleich

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

Im Beispiel ist dies 10.

Wir weisen dem Vektor x den Wertebereich zu und bestimmen die Verteilungsfunktion an den Stellen x .

```
x <- (0:10)

cbind(x,psignrank(x,4))

[1,] 0 0.0625
[2,] 1 0.1250
[3,] 2 0.1875
[4,] 3 0.3125
[5,] 4 0.4375
[6,] 5 0.5625
[7,] 6 0.6875
[8,] 7 0.8125
[9,] 8 0.8750
[10,] 9 0.9375
[11,] 10 1.0000
```

Wir müssen nun noch die Stelle y bestimmen, bei der die Verteilungsfunktion kleiner oder gleich 0.05 ist. Bis zu dieser Stelle y einschließlich lehnen wir ab. Bei der Stelle $y+1$ machen wir die Entscheidung von einem Zufallsexperiment abhängig, dessen Erfolgswahrscheinlichkeit wir noch bestimmen müssen.

Die folgende Befehlsfolge leistet dies für $n=10$:

```
x <- 0:(10*11/2)

p <- psignrank(x,10)
y <- sum(p<=0.05)-1

y
[1] 10

(0.05-p[y+1])/(p[y+2]-p[y+1])
[1] 0.7454545
```

Wir lehnen also für Werte von kleiner oder gleich 10 ab, während wir für gleich 11 eine auf (0,1) gleichverteilte Zufallszahl ziehen und ablehnen, wenn diese kleiner als 0.7454545 ist.

Für den Vorzeichentest ergibt sich:

```
x <- 0:10

p <- pbinom(x,10,0.5)

y <- sum(p<=0.05)-1

y
1

(0.05-p[y+1])/(p[y+2]-p[y+1])
[1] 0.8933333
```

Wir lehnen also für Werte von kleiner oder gleich 1 ab, während wir für S gleich 2 eine auf (0,1) gleichverteilte Zufallszahl ziehen und ablehnen, wenn diese kleiner als 0.893333 ist.

Die Durchführung der Simulationsstudie ist nun ganz einfach.

Wir ziehen 10000 Stichproben vom Umfang 10 aus einer der drei Verteilungen und bestimmen für jede dieser Stichproben die Entscheidung der drei Tests zum Signifikanzniveau 0.05.

Die Gütefunktion schätzen wir dann durch den Anteil der Stichproben, bei denen wir uns für die Gegenhypothese entscheiden. Der Fall $k=0$ schätzt dabei die Wahrscheinlichkeit des Fehlers 1. Art.

Die folgende Befehlsfolge simuliert bei Normalverteilung in 5000 Wiederholungen die Gütefunktion der Tests an der Stelle $-k \cdot \sigma$ bei $n = 10$.

```
# Initialisierung der Zaehlmatrix
  e <- matrix(0,16,3)
# Iteration ueber die Lagealternative
  for (k in 1:16)
# Iteration ueber die Stichproben
  for (i in 1:5000) {
# Erzeugung der Stichprobe
    x <- rnorm(10)-(k-1)*0.1
# t-Test
    e[k,1] <- e[k,1]+
      (t.test(x,alternative="less")$p.value<0.05)
```



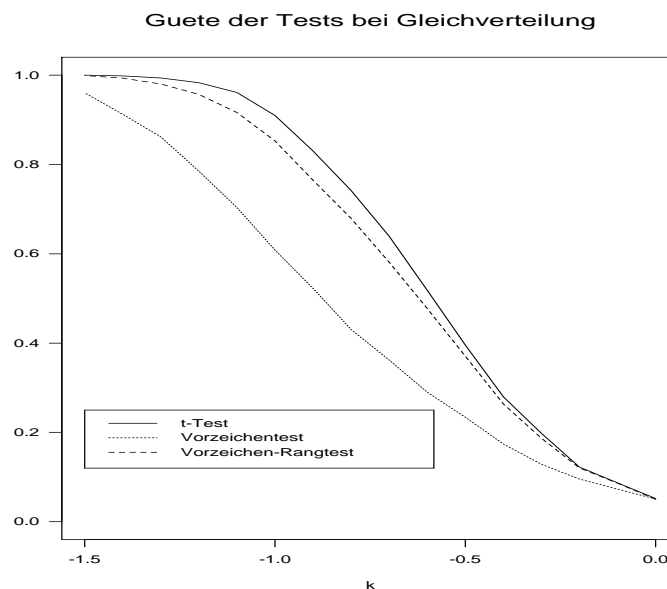
```

# Vorzeichentest
stest <- sum(x>0)
if (stest<=1)
    e[k,2] <- e[k,2]+1
else
    if (stest==2)
        e[k,2] <- e[k,2]+(runif(1)<0.8933)
# Vorzeichen-Rangtest
w <- sum((x>0)*rank(abs(x)))
if (w<=10)
    e[k,3] <- e[k,3]+1
else if (w==11)
    e[k,3] <- e[k,3]+(runif(1)<0.745454)
}

```

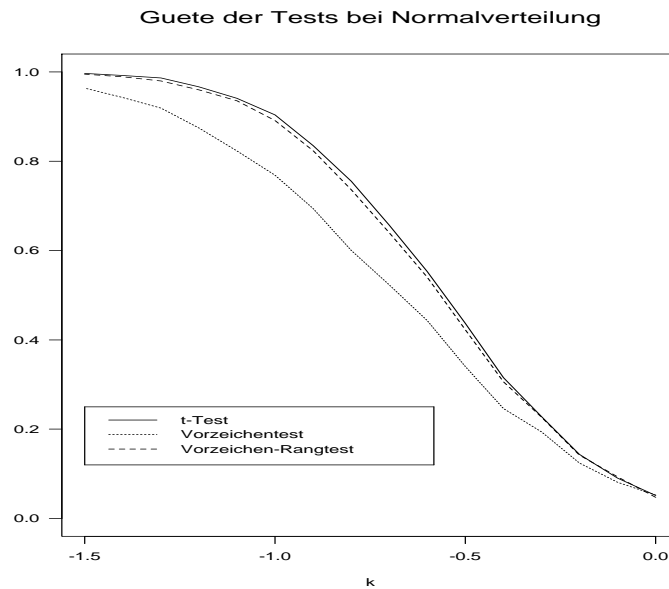
Die folgenden Graphiken zeigen die geschätzten Gütefunktionen der drei Tests bei den Verteilungen.

Bei Gleichverteilung erhalten wir folgendes Bild:



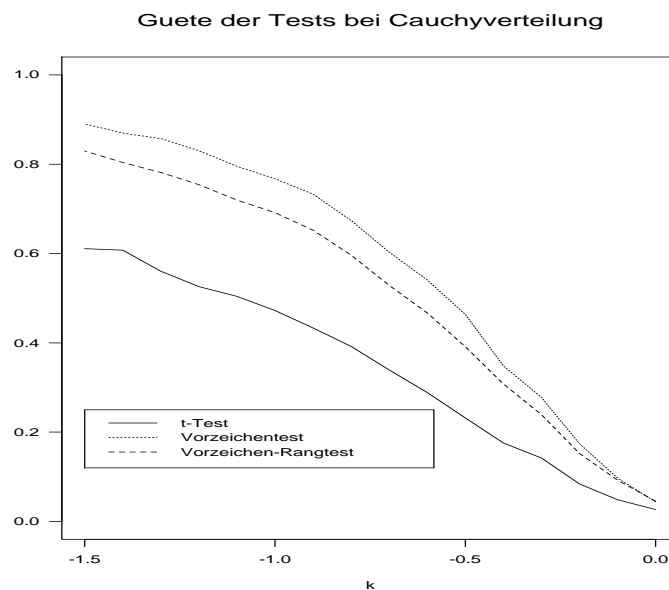
Bei der Gleichverteilung unterscheiden sich der t-Test und der Vorzeichen-Rangtest hinsichtlich der Güte kaum, während der Vorzeichentest sehr viel schlechter ist.

Bei Normalverteilung erhalten wir folgendes Bild:



Bei Normalverteilung ist der t-Test der beste Test, aber der Vorzeichen-Rangtest ist fast genauso gut, während der Vorzeichentest auch hier viel schlechter ist.

Bei Cauchyverteilung erhalten wir folgendes Bild:



Bei einer Verteilung mit hoher Wahrscheinlichkeitsmasse an den Rändern wie der Cauchyverteilung ist der Vorzeichentest am besten.

Aufgrund des Ergebnisses dieser kleinen Simulationsstudie und der Tatsache, daß t-Test und Vorzeichen-Rangtest die Symmetrie der Verteilung benötigen, bieten sich folgende Strategien für den Anwender an:

- Bei einer **schiefen** Verteilung sollte immer der Vorzeichentest angewendet werden.
- Bei einer **symmetrischen** Verteilung gibt es zwei Möglichkeiten:
 - immer den Vorzeichen-Rangtest anwenden, da dessen Gütefunktion immer in der Nähe des besten der Tests liegt
 - bei hoher Wahrscheinlichkeitsmasse an den Rändern den Vorzeichentest anwenden, ansonsten den t-Test.

Kapitel 2

Das Zweistichprobenproblem

Ein wichtiges Anwendungsgebiet der Statistik ist der Vergleich mehrerer Verfahren. Wir beschäftigen uns in diesem Kapitel den Vergleich von zwei Verfahren.

Wir wollen die unterschiedlichen Vorgehensweisen an einem Beispiel von Box, Hunter, Hunter illustrieren.

Es soll untersucht werden, welche von zwei Sorten von Schuhsohlen haltbarer ist. Wir bezeichnen die Sorten mit A und B.

Wie können wir herausfinden, welche der beiden haltbarer ist?

Eine Vorgehensweise besteht darin, $N = m+n$ Personen zufällig auszuwählen und zufällig auf zwei Gruppen aufzuteilen, so daß in der ersten Gruppe m und in der zweiten Gruppe n Personen sind.

Die m Personen der ersten Gruppe erhalten Schuhe mit Schuhsohle A und die n Personen der zweiten Gruppe erhalten Schuhe mit Schuhsohle B.

Nachdem die Personen die Schuhe eine vorgegebene Zeit die Schuhe getragen haben, wird die Abnutzung des Profils bestimmt.

Die Abnutzung in den beiden Gruppen ist:

Sohle A

13.2 8.2 10.9 14.3 10.7 6.6 9.5 10.8 8.8 13.3

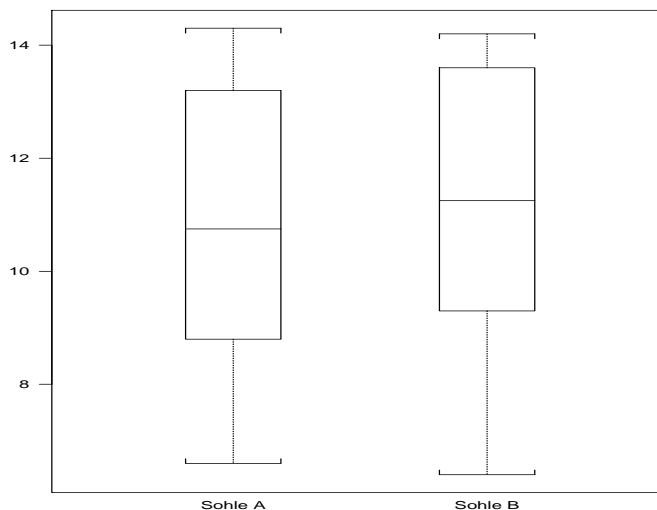
Sohle B

14.0 8.8 11.2 14.2 11.8 6.4 9.8 11.3 9.3 13.6

Deuten diese Zahlen auf einen Unterschied hin?

Wir wollen hier keinen Test durchführen, sondern die Daten nur graphisch darstellen.

Die Boxplots sehen folgendermaßen aus:



Wir sehen, daß die beiden Verteilungen sich nahezu vollständig überlappen. Dies deutet darauf hin, daß kein Unterschied zwischen den Verteilungen vorliegt.

Die Vorgehensweise nennt man ein **unverbundenes Zweistichprobenproblem**. Es wird so genannt, da in den beiden Gruppen unterschiedliche Personen sind.

Wir könnten aber auch anders vorgehen.

Da man nur an der Abnutzung der Schuhsohlen interessiert ist, wird man versuchen, alle anderen Einflußfaktoren möglichst auszuschließen oder konstant zu halten. Im unverbundenen Zweistichprobenproblem versucht man dies dadurch zu erreichen, daß man **randomisiert**.

Das bedeutet, daß man die Personen auf die beiden Gruppen zufällig aufteilt. Eine Verletzung des Prinzips der Randomisierung würde vorliegen, wenn die eine Gruppe nur aus Frauen und die andere nur aus Männern bestehen würde. Ein Unterschied zwischen den beiden Gruppen könnte dann nämlich an den unterschiedlichen Gruppenzusammensetzungen und/oder an der Art der Sohle liegen.

Man kann die anderen Einflußgrößen auch dadurch unter Kontrolle halten, daß man die beiden Behandlungen an einer Person vornimmt.

Man bildet also sogenannte **Blöcke** und wendet jede der beiden Behandlungen in jedem Block an. Hierdurch soll sichergestellt werden, daß alle anderen

Einflußgrößen der Abnutzung konstant gehalten werden.

Hier spricht man dann vom **verbundenen Zweistichprobenproblem**.

Im Beispiel ist dies einfach zu realisieren.

Jede Person zieht einen Schuh mit Schuhsohle A und einen Schuh mit Schuhsohle B an. Hierdurch wird sichergestellt, daß beide Sohlen den gleichen Belastungen ausgesetzt sind. Man könnte nun noch einwenden, daß der linke Fuß anders belastet wird als der rechte. Diesen Einfluß können wir aber dadurch kontrollieren, daß man bei jeder Person durch Münzwurf entscheidet, welche der Sohlen an den linken Fuß und welche an den rechten Fuß kommt. Diesen Effekt kontrollieren wir also wiederum durch **randomisieren**.

Wir haben zwei der **drei Prinzipien der Versuchsplanung** kennengelernt:

- **Randomisierung**
- **Blockbildung**

Beide dienen dazu, alle anderen Einflußgrößen unter Kontrolle, also konstant zu halten.

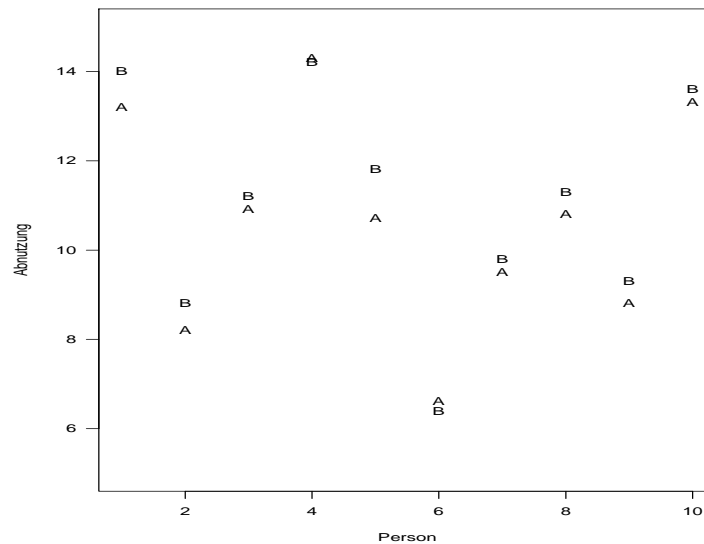
das dritte Prinzip der Versuchsplanung ist die **Wiederholung**. Dieses haben wir immer wieder benutzt, ohne explizit darauf hinzuweisen.

Kehren wir zum Beispiel zurück.

Nehmen wir an, daß die Werte bei den 10 Personen folgendermaßen aussehen:

Person	Sohle A	Sohle B
1	13.2	14.0
2	8.2	8.8
3	10.9	11.2
4	14.3	14.2
5	10.7	11.8
6	6.6	6.4
7	9.5	9.8
8	10.8	11.3
9	8.8	9.3
10	13.3	13.6

Nun wählen wir eine andere Art der graphischen Darstellung. Da die Werte bei einer Person zusammengehören, stellen wir sie auch zusammengehörend dar.



Wir sehen, daß in 8 von 10 Fällen die Abnutzung bei Sohle B größer ist als bei Sohle A. Ob dies signifikant ist, werden wir später sehen.

Verbunden vorzugehen ist aber nicht immer möglich. Will man zum Beispiel zwei Unterrichtsmethoden vergleichen, so kann man nicht eine Person zuerst nach der einen und dann nach der anderen Methode unterrichten. Beim Beginn des Unterrichts nach der zweiten Methode ist die Ausgangssituation nicht die gleiche. In diesem Fall muß man unverbunden vorgehen.

Wir werden im folgenden beide Fälle betrachten und folgende Notation verwenden:

X : Wirkung von Behandlung 1

Y : Wirkung von Behandlung 2

2.1 Verbundene Stichproben

2.1.1 Stetige Variablen

Im verbundenen Zweistichprobenproblem erhält jede Person beide Behandlungen.

Die Daten fallen also paarweise an.

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$$

Dabei ist

X_i : Wirkung von Behandlung 1 bei der i -ten Person

Y_i : Wirkung von Behandlung 2 bei der i -ten Person

Ein Beispiel

Betrachten wir hierzu ein Beispiel. Es soll untersucht werden, ob das Zusammensein mit einer anderen Ratte die Herzfrequenz HF (in Schlägen pro Minute) gegenüber dem Alleinsein verändert.

Dazu wurde von 10 Ratten die Herzfrequenz bestimmt, während sie allein waren und während sie mit einer anderen Ratte zusammen waren.

Es ergaben sich folgende Werte

Ratte	HF alleine	HF zusammen
1	463	523
2	462	494
3	462	461
4	456	535
5	450	476
6	426	454
7	418	448
8	415	408
9	409	470
10	402	437

Quelle: Latane, Cappell(1972): The effect of togetherness on heart rate in rats. Psychon. Science, 29, p.177-179

Es soll nun untersucht werden, ob sich die Herzfrequenz erhöht, wenn die Ratten nicht allein sind. ($\alpha = 0.05$).

Graphische Darstellungen

Wir erstellen zunächst einige Graphiken, um den Zusammenhang zwischen den beiden Variablen zu sehen.

Wir geben zunächst die Daten ein:

```
hfalone <- c(463,462,462,456,450,426,418,415,409,402)
hfnotalone <- c(523,494,461,535,476,454,448,408,470,437)
```

Die einfachste Graphik ist ein Streudiagramm der beiden Variablen.

Da man beide Meßwerte vergleichen will, wählt man einen quadratischen Bereich für die Zeichnung.

Hierzu dient der Befehl

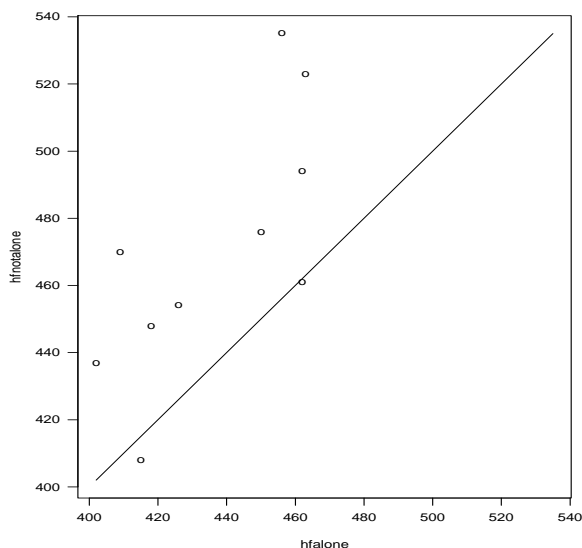
```
par(pty="s")
```

Dann bestimmen wir das Minimum und das Maximum aller Beobachtungen, um den Bereich festzulegen, auf dem wir zeichnen:

```
mi <- min(c(hfalone,hfnotalone))
ma <- max(c(hfalone,hfnotalone))
```

Das Streudiagramm erhalten wir dann durch

```
plot(hfalone,hfnotalone,xlim=c(mi,ma),ylim=c(mi,ma),pch="o")
lines(c(mi,ma),c(mi,ma))
```



Bei Punkten auf der Winkelhalbierenden ist die Herzfrequenz in beiden Situationen gleich. Bei Punkten oberhalb der Winkelhalbierenden ist die Herzfrequenz höher, wenn die Ratten nicht alleine sind.

Wir sehen, daß die meisten Punkte oberhalb der Winkelhalbierenden liegen. Außerdem fällt auf, daß Punkte oberhalb der Winkelhalbierenden weiter von der Winkelhalbierenden entfernt sind als Punkte unterhalb der Winkelhalbierenden.

Wir schreiben uns eine Funktion `scatter.paired`, die diesen Plot erstellt:

```
scatter.paired <- function(x, y, labx = "x", laby = "y") {
  par(pty = "s")
  mi <- min(x, y)
  ma <- max(x, y)
  plot(x, y, xlim = c(mi, ma), ylim = c(mi, ma),
       xlab = labx, ylab = laby, pch="o")
  lines(c(mi, ma), c(mi, ma))
  par(pty = "c")
}
```

Zur Analyse der Daten werden die Differenzen $d_i = y_i - x_i$, $i = 1, \dots, n$ betrachtet.

Dabei sind

- x_i Herzfrequenz der i -ten Ratte, wenn sie allein ist
- y_i Herzfrequenz der i -ten Ratte, wenn sie nicht allein ist

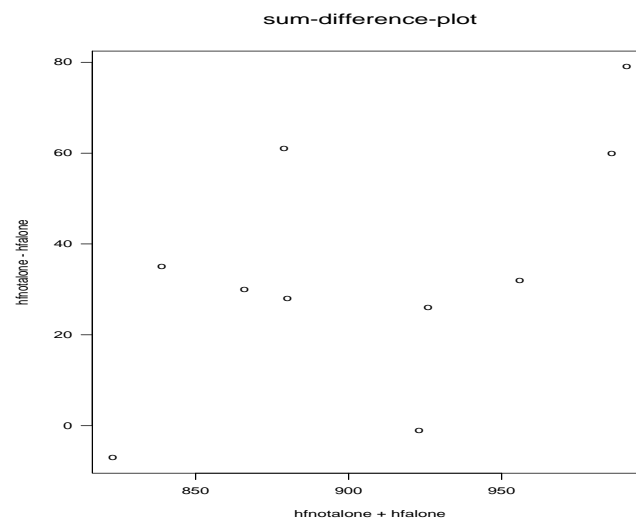
Tukey hat den **sum-difference-plot** vorgeschlagen, bei der die Differenz der Beobachtungen eines Paares gegen die Summe der Beobachtungen eines Paares gezeichnet wird. Die Summe sagt etwas über das Niveau der Daten aus.

Den sum-difference-plot erhält man aus dem Streudiagramm, indem man das Koordinatenkreuz im Urzeigersinn um 45 Grad dreht.

Auf der Winkelhalbierenden gilt im Streudiagramm $y - x = 0$. Auf der zur Winkelhalbierenden senkrecht stehenden Achse gilt $x + y = 0$.

Wir schauen uns diesen in R an:

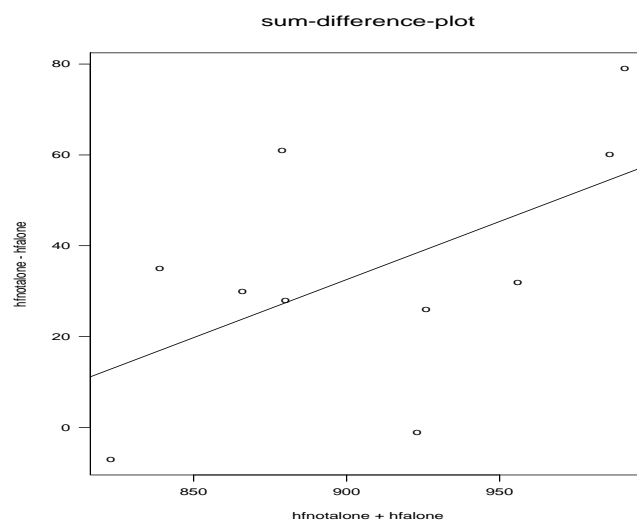
```
par(pty="c")
plot(hfnotalone+hfalone,hfnotalone-hfalone,
     main="sum-difference-plot",pch="o")
```



Es sieht so aus, als ob mit wachsendem Niveau die Differenzen immer größer werden.

Dies zeigt auch die Kleinst-Quadrate Gerade:

```
plot(hfnotalone+hfalone,hfnotalone-hfalone,
     main="sum-difference-plot",pch="o")
abline(lsfit(hfnotalone+hfalone,hfnotalone-hfalone))
```



Tests

Besteht kein Unterschied zwischen den beiden Behandlungen, so sollte gelten

$$E(X) = E(Y)$$

Dies ist aber äquivalent zu

$$\begin{aligned} E(D) &= E(Y - X) \\ &= E(Y) - E(X) \\ &= 0. \end{aligned}$$

Die Verteilung der Differenzen sollte also das Zentrum 0 besitzen.

Wir bilden also die Differenzen $Y_i - X_i$, wobei X_i die Herzfrequenz einer Ratte ist, die allein ist, und Y_i die Herzfrequenz einer Ratte ist, die nicht allein ist.

Die folgende Tabelle zeigt die Differenzen für das Beispiel:

Ratte i	d_i
1	60
2	32
3	-1
4	79
5	26
6	28
7	30
8	-7
9	61
10	35

Durch die Differenzenbildung haben wir es nur noch mit einer Stichprobe zu tun.

Wir können also die Tests des Einstichprobenproblems verwenden.

Die Analyse hängt nun von den Annahmen ab, die über die Differenzen D_i gemacht werden können.

Zunächst gehen wir davon aus, daß D_1, \dots, D_n unabhängige, normalverteilte Zufallsvariablen sind.

Das zweiseitige Testproblem lautet

$$H_0 : E(D) = 0 \quad \text{gegen} \quad H_1 : E(D) \neq 0$$

Wegen

$$E(D) = E(Y - X) = E(Y) - E(X)$$

ist H_0 äquivalent dazu, daß sich die Erwartungswerte der Wirkungen der beiden Behandlungen nicht unterscheiden.

Wir wissen, daß unter der Annahme der Normalverteilung der t-Test für dieses Testproblem am besten ist.

Die Teststatistik lautet

$$t_D = \frac{\sqrt{n} \bar{D}}{S_D}$$

mit

$$\begin{aligned} \bar{D} &= \frac{1}{n} \sum_{i=1}^n D_i \\ &= \frac{1}{n} \sum_{i=1}^n (Y_i - X_i) \\ &= \frac{1}{n} \sum_{i=1}^n Y_i - \frac{1}{n} \sum_{i=1}^n X_i \\ &= \bar{Y} - \bar{X} \end{aligned}$$

und

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$$

Wenn H_0 zutrifft, ist t_d t-verteilt mit $n-1$ Freiheitsgraden.

H_0 wird abgelehnt, wenn gilt

$$|t_D| > t_{n-1; 1-\alpha/2}$$

wobei $t_{n-1; 1-\alpha/2}$ das $1-\alpha/2$ -Quantil der t-Verteilung mit $n-1$ Freiheitsgraden ist.

Es können natürlich auch einseitige Tests durchgeführt werden.

$$H_0 : E(D) \leq 0 \quad \text{gegen} \quad H_1 : E(D) > 0$$

H_0 wird abgelehnt, wenn gilt

$$t_D > t_{n-1;1-\alpha}$$

wobei $t_{n-1;1-\alpha}$ das $1 - \alpha$ -Quantil der t-Verteilung mit $n-1$ Freiheitsgraden ist.

$$H_0 : E(D) \geq 0 \quad \text{gegen} \quad H_1 : E(D) < 0$$

H_0 wird abgelehnt, wenn gilt

$$t_D < -t_{n-1;1-\alpha}$$

wobei $t_{n-1;1-\alpha}$ das $1 - \alpha$ -Quantil der t-Verteilung mit $n-1$ Freiheitsgraden ist. Schauen wir uns das Beispiel unter der Annahme der Normalverteilung an. Wir wollen also überprüfen, ob das Zusammensein mit einer anderen Ratte die Herzfrequenz erhöht.

Es sollte unter der Alternativhypothese also gelten $E(Y) > E(X)$.

Dies führt zu folgendem Testproblem:

$$H_0 : E(D) \leq 0$$

gegen

$$H_1 : E(D) > 0$$

Dies ist ein einseitiger Test.

Es gilt

$$\bar{d} = 34.3$$

und

$$s_D = 26.78.$$

Also gilt

$$t_D = \frac{\sqrt{10} \cdot 34.3}{26.78}.$$

Wegen $t_{9;0.95} = 1.833$ wird H_0 zum Niveau $\alpha = 0.05$ abgelehnt.

In R sieht das so aus:

Wir rufen die Funktion `t.test` auf, wobei wir das Argument `paired` auf `T` setzen, um einen verbundenen t-Test durchzuführen.

Das Argument `alternative` wird auf `greater` gesetzt, um einen einseitigen Test mit der Alternativhypothese $E(Y) > E(X)$ durchzuführen.

```
t.test(hfnotalone,hfalone,alternative="greater",paired=T)
```

```
Paired t-Test
```

```
data: hfnotalone and hfalone
```

```
t = 4.0498,
```

```
df = 9,
```

```
p-value = 0.0014
```

```
alternative hypothesis:
```

```
true mean of differences is greater than 0
```

```
95 percent confidence interval:
```

```
18.77423      NA
```

```
sample estimates:
```

```
mean of x - y
```

```
34.3
```

Die Überschreitungswahrscheinlichkeit beträgt 0.0014, so daß die Nullhypothese zum Signifikanzniveau 0.05 abgelehnt wird.

Außerdem liefert die Funktion noch ein einseitiges Konfidenzintervall für die Differenz der Erwartungswerte als Ergebnis.

Unterstellt man keine Normalverteilung, so kann man den Vorzeichentest oder den Wilcoxon-Vorzeichen-Rangtest auf die Differenzen anwenden.

Zunächst der Vorzeichentest

```
anz <- sum(hfnotalone-hfalone>0)
```

```
n <- length(hfalone)
```

```
binom.test(anz,n,0.5,alternative="greater")
```

```
Exact binomial test
```

```
data:  anz out of n
```

```
number of successes = 8,
n = 10,
p-value = 0.0547
```

```
alternative hypothesis: true p is greater than 0.5
```

Die Überschreitungswahrscheinlichkeit beträgt 0.0547, so daß die Nullhypothese zum Signifikanzniveau 0.05 nicht abgelehnt wird.

Der Aufruf des Vorzeichen-Rangtest entspricht in den Argumenten dem des verbundenen t-Tests.

```
wilcox.test(hfnotalone,hfalone,paired=T,alternative="greater")
```

```
Exact Wilcoxon signed-rank test
```

```
data: hfnotalone and hfalone
signed-rank statistic V = 52,
n = 10,
p-value = 0.0049
```

```
alternative hypothesis: true mu is greater than 0
```

Die Überschreitungswahrscheinlichkeit beträgt 0.0049, so daß die Nullhypothese zum Signifikanzniveau 0.05 abgelehnt wird.

Wir stellen fest, daß zum Signifikanzniveau $\alpha = 0.05$ der Wilcoxontest und der t-Test ablehnen, während der Vorzeichentest nicht ablehnt.

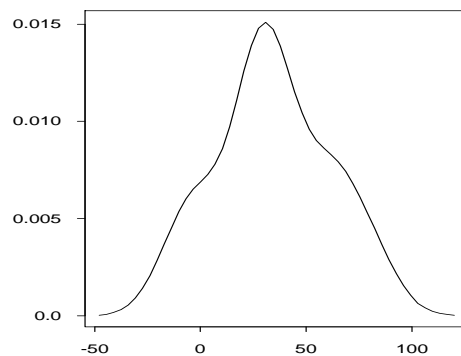
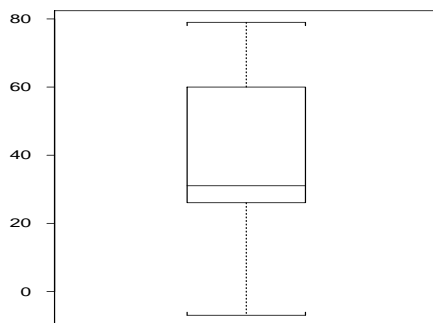
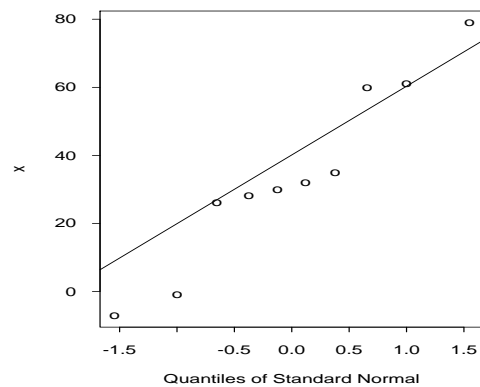
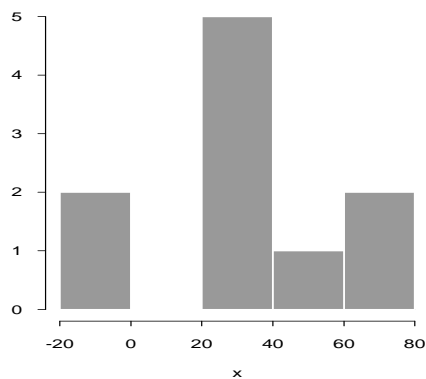
Um zu sehen, welcher Test für unsere Problemstellung am besten geeignet ist, erstellen wir das Histogramm, eine Dichteschätzung, den Boxplot und den normal probability plot der Differenzen.

Die nachstehende Funktion `eda.shape` erstellt alle 4 Grafiken:

```
eda.shape <- function(x){
  par(mfrow = c(2, 2))
  hist(x)
  qqnorm(x)
  qqline(x)
  boxplot(x)
  iqd <- summary(x)[5] -summary(x)[2]
  plot(density(x, width = 2 * iqd), xlab = "",
       ylab = "", type = "l")
}
```

Wir rufen die Funktion auf:

```
eda.shape(hfnotalone-hfalone)
```



Der Befehl `par(mfrow = c(2, 2))` setzt fest, daß in einem Fenster 4 Bilder erstellt werden.

Die Zeichnungen deuten auf Symmetrie hin, die Annahme der Normalverteilung scheint nicht gerechtfertigt zu sein, so daß man den Vorzeichen-Rang-Test anwenden sollte.

Wir kommen also zum Ergebnis, daß das Zusammensein mit einer anderen Ratte die Herzfrequenz erhöht.

Hinweis:

Die Graphiken sollte man natürlich vor der Auswahl eines Tests erstellen. In der Praxis wird man nur einen Test durchführen.

Die Schuhsohlen

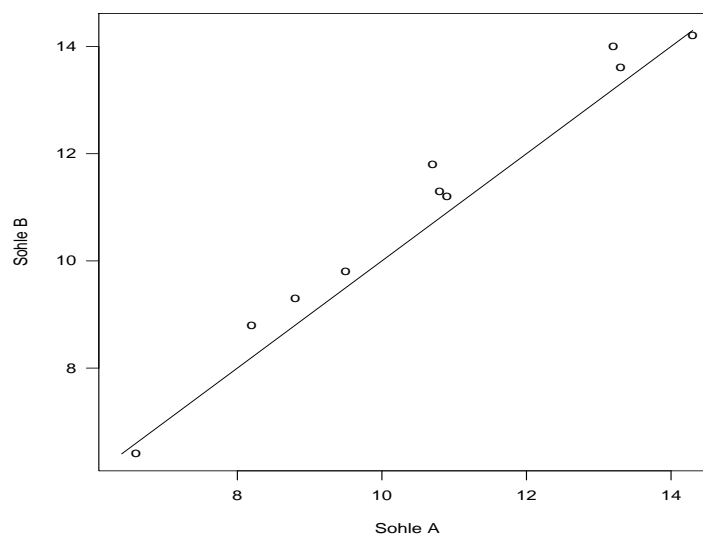
Wir wollen nun das Eingangsbeispiel mit den Schuhsohlen analysieren. Dazu geben wir die Daten ein:

```
sohle.A <- c(13.2,8.2,10.9,14.3,10.7,6.6,9.5,10.8,8.8,13.3)
```

```
sohle.B <- c(14.0,8.8,11.2,14.2,11.8,6.4,9.8,11.3,9.3,13.6)
```

Wir machen zunächst einige Bilder der Daten:

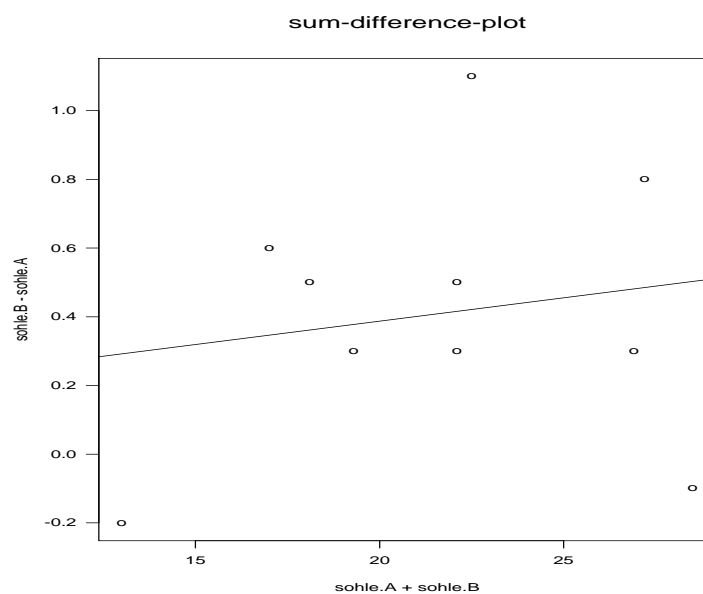
```
scatter.paired(sohle.A,sohle.B,"Sohle A","Sohle B")
```



Wir sehen, daß die meisten Punkte oberhalb der Winkelhalbierenden liegen.

Wir bilden die Differenzen D_i , wobei X_i die Sohle A und Y_i die Sohle B ist. Schauen wir uns den sum-difference-plot an.

```
plot(sohle.A+sohle.B,sohle.B-sohle.A,  
     main="sum-difference-plot",pch="o")  
  
abline(lsfit(sohle.A+sohle.B,sohle.B-sohle.A))
```



Der Plot deutet auf keinen systematischen Zusammenhang in den Differenzen hin.

Wir wollen testen, ob sich die beiden Sorten unterscheiden. Wir führen also einen zweiseitigen Test durch.

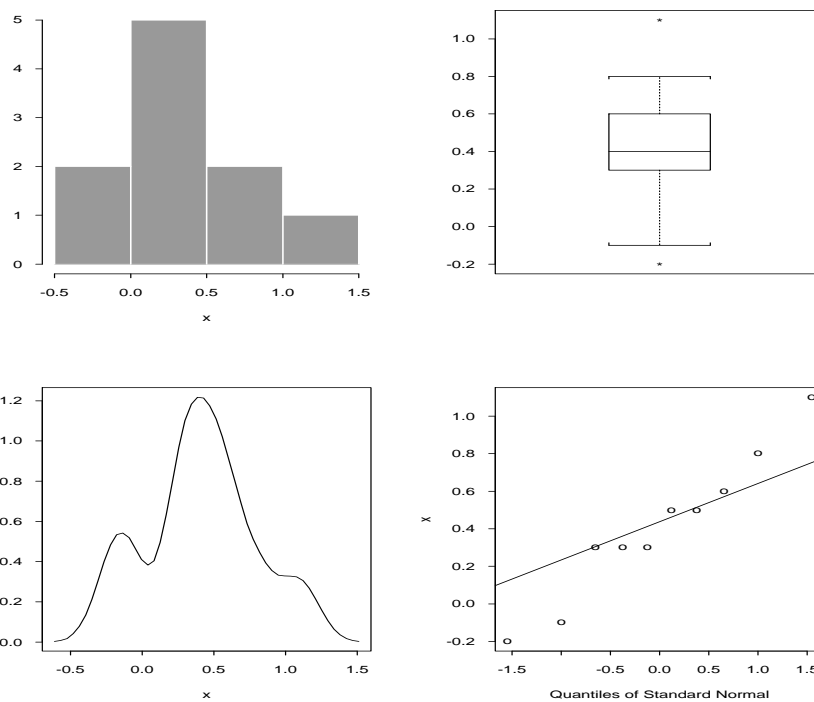
Das Testproblem in den Differenzen $D_i = Y_i - X_i$ lautet also

$$H_0 : E(D) = 0 \quad \text{gegen} \quad H_1 : E(D) \neq 0$$

Welcher Test ist geeignet?

Wir wenden die Funktion `eda.shape` an.

```
eda.shape(sohle.B-sohle.A)
```



Die Graphiken deuten eine symmetrischen Verteilung hin. Es sieht aber nicht nach der Normalverteilung aus.

Wir wenden also den Vorzeichenrangtest an.

```
> wilcox.test(sohle.A,sohle.B,paired=T)
```

Wilcoxon signed-rank test

data: sohle.A and sohle.B

signed-rank normal statistic with correction $Z = -2.4495$,

p-value = 0.0143

alternative hypothesis: true mu is not equal to 0

Warning messages: cannot compute exact p-value with ties in:
`wil.sign.rank(dff, alternative, exact, correct)`

Die beiden Sohlensorten unterscheiden sich also.

Noch ein Beispiel

Schauen wir uns noch ein Beispiel an.

Es soll untersucht werden, ob ein Medikament zur Senkung des diastolischen Blutdrucks führt.

Zunächst wird der Blutdruck von 10 Patienten bestimmt. Dann erhält jeder der Patienten das Medikament. Nach zwei Stunden wird bei jedem der Patienten der diastolische Blutdruck bestimmt.

Es ergaben sich folgende Werte:

Patient	Blutdruck vorher	Blutdruck nachher
1	130	125
2	122	121
3	124	121
4	104	106
5	112	101
6	102	98
7	98	90
8	119	98
9	106	110
10	107	103

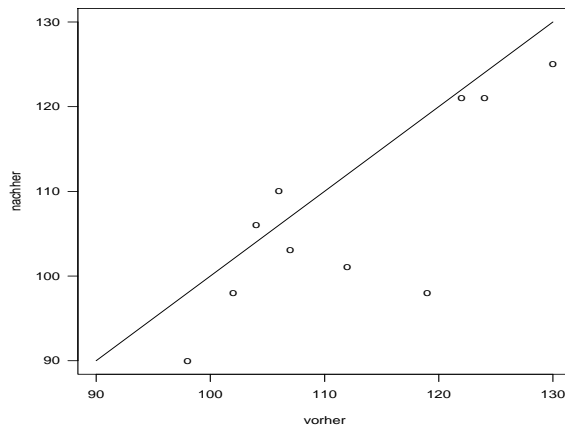
Wir geben die Daten ein:

```
vorher <- c(130,122,124,104,112,102,98,119,106,107)
```

```
nachher <- c(125,121,121,106,101,98,90,98,110,103)
```


Wir machen zunächst einige Bilder der Daten:

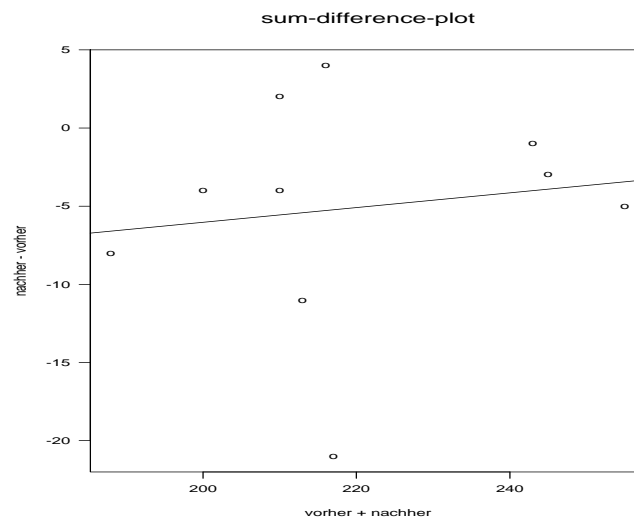
```
scatter.paired(vorher,nachher,"vorher","nachher")
```



Wir sehen, daß die meisten Punkte unterhalb der Winkelhalbierenden liegen. Wir bilden die Differenzen D_i , wobei X_i der Blutdruck vorher und Y_i der Blutdruck nachher ist.

Schauen wir uns den sum-difference-plot an.

```
plot(vorher+nachher,nachher-vorher,  
     main="sum-difference-plot",pch="o")  
abline(lsfit(vorher+nachher, nachher-vorher))
```



Der Plot deutet auf keinen systematischen Zusammenhang in den Differenzen hin.

Wir wollen testen, ob das Medikament zur Senkung des Blutdrucks führt.

Unter H_1 muß also $E(Y) < E(X)$ gelten.

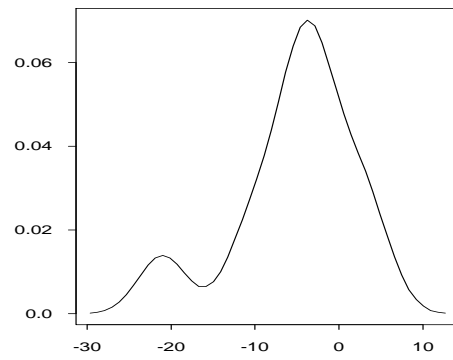
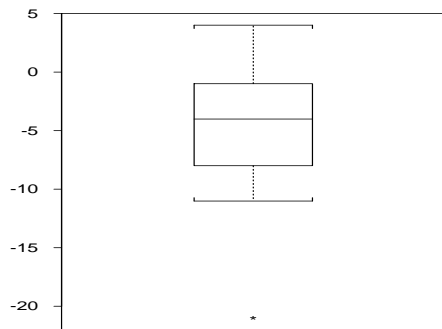
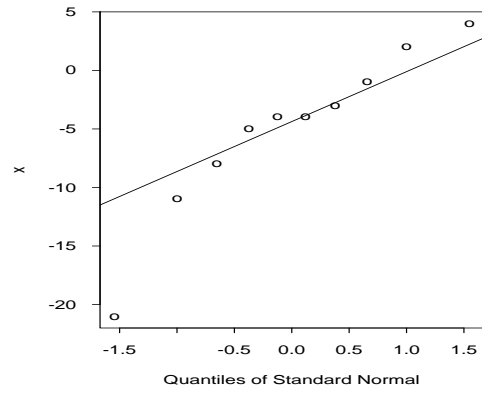
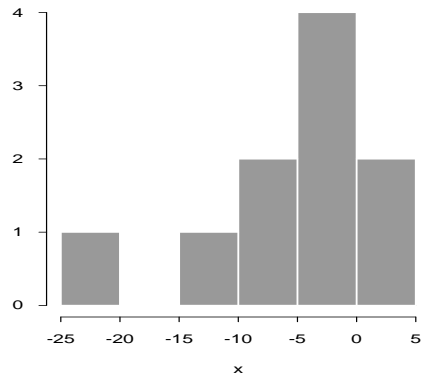
Das Testproblem in den Differenzen $D_i = Y_i - X_i$ lautet also

$$H_0 : E(D) \geq 0 \quad \text{gegen} \quad H_1 : E(D) < 0$$

Welcher Test ist geeignet?

Wir wenden die Funktion `eda.shape` an.

```
eda.shape(nachher-vorher)
```



Die Graphiken deuten auf einen Ausreißer bei einer symmetrischen Verteilung hin.

Wir wenden also den Vorzeichentest an.

```
anz <- sum(nachher-vorher>0)
n <- length(vorher)
binom.test(anz,n,0.5,alternative="less")
  Exact binomial test
data:  anz out of n
number of successes = 2, n = 10, p-value=0.0547
alternative hypothesis: true p is less than 0.5
```

Wir lehnen die Nullhypothese also nicht ab.

2.1.2 Binäre Variablen

Bisher haben wir nur stetige Variablen betrachtet. Bei vielen statistischen Untersuchungen werden aber binäre Variablen erhoben. Sehr oft werden zwei binäre Variablen an der gleichen Person erhoben.

Beispiel 2.1.1 *Um zu überprüfen, ob zwei Fragen vom gleichen Schwierigkeitsgrad sind, werden beide Fragen 20 Studenten gestellt.*

Seien

$$X_i = \begin{cases} 1 & \text{falls der } i.\text{te Student Frage 1 richtig beantwortet} \\ 0 & \text{sonst} \end{cases}$$

und

$$Y_i = \begin{cases} 1 & \text{falls der } i.\text{te Student Frage 2 richtig beantwortet} \\ 0 & \text{sonst} \end{cases}$$

Die Beobachtungen fallen als Tupel (x_i, y_i) , $i=1,2,\dots,n$ an.

Wir beobachten

$$\begin{array}{ccccc} (1, 1) & (1, 0) & (1, 0) & (0, 0) & (0, 0) \\ (0, 1) & (1, 1) & (1, 0) & (1, 1) & (0, 0) \\ (0, 1) & (1, 0) & (1, 1) & (0, 0) & (1, 0) \\ (0, 1) & (1, 0) & (1, 1) & (1, 0) & (1, 0) \end{array}$$

Wir können die Daten in einer Tabelle zusammenfassen:

(x_i, y_i)	Anzahl
(1,1)	5
(1,0)	8
(0,1)	3
(0,0)	4

Sei

$$p_{ij} = P(X_k = i, Y_k = j)$$

mit $i = 0, 1$, $j = 0, 1$ und $k = 1, \dots, n$.

Dann lautet das Testproblem

$$H_0 : p_{1.} = p_{.1} \quad \text{gegen} \quad H_1 : p_{1.} \neq p_{.1}$$

mit

$$p_{1.} = p_{11} + p_{10}$$

$$p_{.1} = p_{11} + p_{01}$$

Dabei ist $p_{1.}$ also die Wahrscheinlichkeit, die erste Frage richtig zu beantworten und $p_{.1}$ die Wahrscheinlichkeit, die zweite Frage richtig zu beantworten. Somit gilt unter H_0 :

$$p_{11} + p_{10} = p_{11} + p_{01}$$

Dies ist äquivalent zu

$$p_{10} = p_{01}$$

Es soll also überprüft werden, ob die Wahrscheinlichkeit, die erste Frage richtig und die zweite Frage falsch zu beantworten, gleich der Wahrscheinlichkeit ist, die erste Frage falsch und die zweite Frage richtig zu beantworten.

Es interessieren also nur die Personen, die genau eine Frage richtig beantwortet haben.

Es ist zu überprüfen, ob bei diesen gilt:

$$H_0 : p_{10} = p_{01} = 0.5$$

Sind die beiden Fragen gleich schwer, so erwarten wir, daß die Hälfte der Personen nur die erste Frage und der Rest nur die zweite Frage richtig beantwortet.

Sei N_{10} die Anzahl der Personen, die nur die erste Frage richtig beantwortet, und N_{01} die Anzahl der Personen, die nur die zweite Frage richtig beantwortet.

In unserem Beispiel gilt $n_{10} = 8$ und $n_{01} = 3$.

Bei beobachteter Anzahl $N_{10} + N_{01} = n$, ist N_{10} binomialverteilt mit den Parametern n und $p = 0.5$.

Wir lehnen H_0 ab, wenn N_{10} zu groß oder zu klein ist.

In unserem Beispiel gilt also $n_{10} = 8$ und $n=11$.

Somit erhalten wir als Überschreitungswahrscheinlichkeit für den zweiseitigen Test

$$\begin{aligned} 2 P_{H_0}(N_{10} \geq 8) &= 2 \left[\binom{11}{8} + \binom{11}{9} + \binom{11}{10} + \binom{11}{11} \right] 0.5^{11} = \\ &= 2(0.080566 + 0.026855 + 0.0053710 + 0.00048828) = \\ &= 0.2265625 \end{aligned}$$

Somit lehnen wir H_0 nicht zum Niveau $\alpha = 0.05$ ab.

Der Test heißt auch McNemar Test.

Der McNemar-Test ist nicht anderes als der konditionale Vorzeichentest.

Bilden wir nämlich die Differenzen $D_i = X_i - Y_i$, so erhalten wir für das Beispiel:

0 1 1 0 0 -1 0 1 0 0 -1 1 0 0 1 -1 1 0 1 1

Läßt man die Nullen weg, so bleiben 11 Zahlen übrig, von denen 8 positiv sind.

Da die Binomialverteilung für großes n durch die Normalverteilung approximiert werden kann, ist unter H_0

$$\begin{aligned} \frac{N_{10} - \frac{N_{10} + N_{01}}{2}}{\sqrt{\frac{N_{10} + N_{01}}{4}}} &= \frac{N_{10} - 0.5 N_{10} - 0.5 N_{01}}{0.5 \sqrt{N_{10} + N_{01}}} \\ &= \frac{N_{10} - N_{01}}{\sqrt{N_{10} + N_{01}}} \end{aligned}$$

approximativ standardnormalverteilt.

Da das Quadrat einer standardnormalverteilten Zufallsvariablen mit einem Freiheitsgrad chiquadratverteilt ist (siehe dazu z.B. Mood, Graybill, Boes, S.182), erhalten wir als äquivalente Teststatistik:

$$\frac{(N_{10} - N_{01})^2}{N_{10} + N_{01}}$$

In dieser Form ist die Teststatistik des McNemar Tests in vielen Büchern zu finden, wobei unter Umständen noch eine Stetigkeitskorrektur verwendet wird:

$$\frac{(|N_{10} - N_{01}| - 1)^2}{N_{10} + N_{01}}$$

Für unser Beispiel gilt

$$\frac{(N_{10} - N_{01})^2}{N_{10} + N_{01}} = .2727$$

und

$$\frac{(|N_{10} - N_{01}| - 1)^2}{N_{10} + N_{01}} = 1.4545$$

In R gibt es eine Funktion `mcnemar.test`, die die Daten in Form einer (2,2)-Matrix erwartet.

```
m <- matrix(c(5,3,8,4),2,2)
m
      [,1] [,2]
[1,]    5    8
[2,]    3    4
```

Der Aufruf der Funktion `mcnemar.test` mit der Matrix als Argument liefert das Ergebnis:

```
mcnemar.test(m)
      McNemar's chi-square test with
      continuity correction
data:  m
McNemar's chi-square = 1.4545, df = 1, p-value = 0.2278
```

Wir sehen, daß der approximative Test mit Stetigkeitskorrektur durchgeführt wird.

Die approximative Überschreitungswahrscheinlichkeit 0.2278 unterscheidet sich kaum von der weiter oben berechneten exakten 0.2265625.

2.2 Unverbundene Stichproben

2.2.1 Kategoriale Variablen

Binäre Variablen

Es sollen zwei Unterrichtsmethoden miteinander verglichen werden. In diesem Fall ist es nicht möglich, verbunden vorzugehen. Denn dann würde man ja jedes Kind nach jeder der beiden Methoden unterrichten, was dazu führt, daß die Ausgangssituation vor der zweiten Methode nicht dieselbe ist.

Deshalb muß man anders vorgehen.

Man teilt eine Gruppe von $m+n$ Kindern zufällig auf zwei Gruppen auf, wobei die m Kinder der ersten Gruppe nach der ersten Methode und die n Kinder der zweiten Gruppe nach der zweiten Methode unterrichtet werden. Am Ende müssen alle Kinder die gleiche Klausur schreiben.

Von Interesse ist, ob der Anteil der Kinder, die die Klausur bestehen, in beiden Gruppen gleich groß ist.

In jeder Gruppe seien 6 Kinder.

Von den Kindern der ersten Gruppe haben 5 die Klausur bestanden, während in der zweiten Gruppe nur 2 Kinder bestanden haben.

Wir können die Daten in einer Kontingenztabelle anordnen:

	Erfolg	bestanden	nicht bestanden
Methode			
1		5	1
2		2	4

Allgemein liegen die Daten in folgender Form vor:

	Erfolg	bestanden	nicht bestanden
Methode			
1		n_{11}	n_{12}
2		n_{21}	n_{22}

Bevor wir Tests durchführen, schauen wir uns die Daten erst näher einmal an.

Hierzu ergänzen wir die Tabelle um die Zeilensummen und Spaltensummen. Wir erhalten also folgende Tabelle in allgemeiner Form.

	Erfolg	bestanden	nicht bestanden	
Methode				
1		n_{11}	n_{12}	$n_{1.}$
2		n_{21}	n_{22}	$n_{2.}$
		$n_{.1}$	$n_{.2}$	

mit

$$n_{1.} = n_{11} + n_{12}$$

$$n_{2.} = n_{21} + n_{22}$$

$$n_{.1} = n_{11} + n_{21}$$

$$n_{.2} = n_{12} + n_{22}$$

Für das Beispiel erhalten wir

	Erfolg	bestanden	nicht bestanden	
Methode				
1		5	1	6
2		2	4	6
		7	5	

Wir schauen uns zunächst an, wie groß der Anteil der Leute, die die Klausur bestanden haben, in den beiden Gruppen ist.

Bei denen, die nach Methode 1 unterrichtet wurden, gilt

$$n_{1|1} = \frac{n_{11}}{n_{1.}} = \frac{5}{6} = 0.83$$

und bei denen, die nach Methode 2 unterrichtet wurden, gilt

$$n_{1|2} = \frac{n_{21}}{n_{2.}} = \frac{2}{6} = 0.33$$

Von den Kindern, die nach Methode 1 unterrichtet wurden, haben also 83 Prozent die Klausur bestanden, während von den Kindern, die nach Methode 2 unterrichtet wurden, nur 33 Prozent die Klausur bestanden haben.

Die Frage ist, ob dieser Unterschied signifikant ist.

Wir schauen uns nun einen für diese Fragestellung geeigneten Test an.

Sei p_1 die Wahrscheinlichkeit, daß ein Kind die Klausur besteht, wenn es nach Methode 1 unterrichtet wurde, und p_2 die Wahrscheinlichkeit, daß ein Kind die Klausur besteht, wenn es nach Methode 2 unterrichtet wurde.

Zu testen ist

$$H_0 : p_1 = p_2 \quad \text{gegen} \quad H_1 : p_1 \neq p_2$$

Mit Hilfe von Fishers Permutationsprinzip erhalten wir einen geeigneten Test. Dazu bezeichnen wir "bestanden" mit 1 und "nicht bestanden" mit 0.

Bei Fisher wird unterstellt, daß die Beobachtungen gegeben sind. In unserem Fall heißt dies, daß die Anzahl der Einsen und die Anzahl der Nullen fest sind.

Insgesamt wurden 12 Beobachtungen gemacht, wobei in jeder Gruppe 6 Beobachtungen anfielen. Von den 12 Beobachtungen nehmen 7 den Wert 1 und 5 den Wert 0 an.

Fishers Permutationsprinzip besagt nun, daß wir alle möglichen Verteilungen der Beobachtungen auf die beiden Stichproben vom Umfang 6 betrachten müssen.

Eine Verteilung der Beobachtungen auf die beiden gleich großen Stichproben liegt fest, wenn die Anzahl der Einsen in der ersten Gruppe bekannt ist.

Sind zum Beispiel in der ersten Gruppe 5 Einsen, so muß in dieser Gruppe eine Null sein. Außerdem müssen in der zweiten Gruppe 2 Einsen und 4 Nullen sein.

Dies ergibt die obige Kontingenztabelle.

Für das Beispiel sind somit folgende Kontingenztabellen möglich:

1	5
6	0

2	4
5	1

3	3
4	2

4	2
3	3

5	1
2	4

6	0
1	5

Die Anzahl N_{11} der Einsen in der ersten Gruppe kann also die Werte $1, \dots, 6$ annehmen.

Die Wahrscheinlichkeitsverteilung von N_{11} ergibt sich allgemein folgendermaßen:

Es gibt

$$\binom{n}{n_{1.}}$$

Möglichkeiten, aus den n Beobachtungen $n_{1.}$ für die erste Gruppe auszuwählen.

Unter diesen $n_{1.}$ Beobachtungen soll die Eins genau n_{11} -mal auftreten. Die restlichen $n_{1.} - n_{11}$ Beobachtungen dieser Gruppe sind dann Nullen.

Die n_{11} Einsen werden aus den $n_{1.}$ Einsen ausgewählt, und die $n_{1.} - n_{11}$ Nullen werden aus den $n - n_{1.}$ Nullen ausgewählt.

Es gibt

$$\binom{n_{1.}}{n_{11}}$$

Möglichkeiten, n_{11} Einsen aus den $n_{1.}$ Einsen auszuwählen.

Zu jeder dieser Möglichkeit gibt es

$$\binom{n - n_{1.}}{n_{1.} - n_{11}}$$

Möglichkeiten, $n_{1.} - n_{11}$ Nullen aus den $n - n_{1.}$ Nullen auszuwählen.

Somit gilt

$$P(N_{11} = n_{11}) = \frac{\binom{n_{1.}}{n_{11}} \binom{n - n_{1.}}{n_{1.} - n_{11}}}{\binom{n}{n_{1.}}}$$

N_{11} ist also hypergeometrisch verteilt.

Für die Gegenhypothese sprechen also zu große oder zu kleine Werte von N_{11} .

Für den Datensatz erhalten wir folgende Wahrscheinlichkeiten:

$$P(N_{11} = 1) = \frac{\binom{7}{1} \binom{5}{5}}{\binom{12}{6}} = 0.0076$$

$$P(N_{11} = 2) = \frac{\binom{7}{2} \binom{5}{4}}{\binom{12}{6}} = 0.1136$$

$$P(N_{11} = 3) = \frac{\binom{7}{3} \binom{5}{3}}{\binom{12}{6}} = 0.3788$$

$$P(N_{11} = 4) = \frac{\binom{7}{4} \binom{5}{2}}{\binom{12}{6}} = 0.3788$$

$$P(N_{11} = 5) = \frac{\binom{7}{5} \binom{5}{1}}{\binom{12}{6}} = 0.1136$$

$$P(N_{11} = 6) = \frac{\binom{7}{6} \binom{5}{0}}{\binom{12}{6}} = 0.0076$$

Für den Datensatz gilt $n_{11} = 5$. Somit beträgt die Überschreitungswahrscheinlichkeit beim zweiseitigen Test

$$P(N_{11} = 1) + P(N_{11} = 2) + P(N_{11} = 5) + P(N_{11} = 6) = 0.2424$$

In R gibt es eine Funktion `fisher.test`, die als Argument die Kontingenztafel erhält.

Wir bilden also erst die Tabelle als Matrix.

```
m <- matrix(c(5,2,1,4),2,2)
```

```
m
```

```
      [,1] [,2]
[1,]    5    1
[2,]    2    4
```

Dann rufen wir die Funktion `fisher.test` mit der Matrix auf.

```
fisher.test(m)
```

```
      Fisher's exact test
```

```
data:  m
p-value = 0.2424
alternative hypothesis: two.sided
```

Für sehr große Stichprobenumfänge kann man die Funktion `chisq.test` verwenden, auf die wir im nächsten Abschnitt eingehen.

Für unseren Datensatz liefert `chisq.test` eine ähnliche Überschreitungswahrscheinlichkeit.

```
chisq.test(m)
```

```
      Pearson's chi-square test
      with Yates' continuity correction
```

```
data:  m
X-squared = 1.3714,
df = 1,
p-value = 0.2416
```

```
Warning messages:
```

```
Expected counts < 5. Chi-squared approximation may not be
appropriate. in: chisq.test(m)
```

Kategoriale Variablen mit mehr als zwei Ausprägungen

In der ersten Statistik I Vorlesung im WS 96/97 wurden 255 Studenten nach ihrem Wahlverhalten und ihrem Geschlecht befragt.

Es ergab sich folgende Tabelle

	CDU/CSU	SPD	FDP	Grüne	gar keine	weiß nicht	
w	13	10	3	11	5	23	65
m	55	30	20	26	24	35	190
	68	40	23	37	29	58	255

Es soll überprüft werden, ob sich das Wahlverhalten der Männer und Frauen unterscheidet. Hierzu bestimmen wir zunächst die bedingten relativen Häufigkeiten in den beiden Gruppen.

Diese sind in der folgenden Tabelle zu finden.

	CDU/CSU	SPD	FDP	Grüne	gar keine	weiß nicht	
w	0.20	0.15	0.05	0.17	0.08	0.35	1.00
m	0.29	0.16	0.11	0.14	0.13	0.18	1.00
	0.27	0.16	0.09	0.15	0.11	0.23	1.00

Wir sehen, ist das Wahlverhalten bei den Frauen und Männern unterschiedlich. Große Unterschiede bestehen vor allem in den Kategorien 'FDP' und 'weiß nicht'.

Um zu überprüfen, ob diese Unterschiede signifikant sind, kann man den Chi-Quadrat-Homogenitätstest anwenden.

Seien n_{ij} die beobachteten absoluten Häufigkeiten in der i -ten Gruppe mit $i = 1$ für 'weiblich' und $i = 2$ für 'männlich' und der j -ten Kategorie j der zweiten Variablen mit $j = 1$ für 'CDU/CSU', $j = 2$ für 'SPD', $j = 3$ für 'FDP', $j = 4$ für 'Grüne', $j = 5$ für 'gar keine' und $j = 6$ für 'weiß nicht'.

Es soll überprüft werden, ob sich das Wahlverhalten der Männer und Frauen unterscheidet.

Sei p_{ij} die Wahrscheinlichkeit, daß eine Person die i -te Kategorie der ersten Variablen und die j -te Kategorie der zweiten Variablen aufweist. Wir betrachten die Vektoren der bedingten Wahrscheinlichkeiten in beiden Gruppen.

Die Wahrscheinlichkeitsverteilung in der ersten Gruppe ist gegeben durch

$$\left(p_{1|1}, \dots, p_{J|1}\right)$$

und die der zweiten Gruppe

$$\left(p_{1|2}, \dots, p_{J|2}\right)$$

mit

$$p_{j|i} = \frac{p_{ij}}{p_{i.}}$$

Es ist zu testen, daß die beiden Gruppen homogen sind:

$$H_0 : \frac{p_{ij}}{p_{i.}} = p_{.j}$$

für $i = 1, 2$ und $j = 1, \dots, J$.

Unter der Nullhypothese gilt

$$p_{ij} = p_{i.} p_{.j}$$

und somit

$$n p_{ij} = n p_{i.} p_{.j}$$

Die Wahrscheinlichkeiten sind unbekannt.

Wir schätzen sie durch die entsprechenden relativen Häufigkeiten:

$$\hat{p}_{i.} = \frac{n_{i.}}{n}$$

und

$$\hat{p}_{.j} = \frac{n_{.j}}{n}$$

Somit erhalten wir folgende geschätzten erwarteten Häufigkeiten

$$\hat{n}_{ij} = n \hat{p}_{i.} \hat{p}_{.j} = n \frac{n_{i.}}{n} \frac{n_{.j}}{n}$$

In R können wir die erwarteten Häufigkeiten folgendermaßen schätzen:
Wir geben die Kontingenztabelle ein:

```
h <- matrix(c(13,55,10,30,3,20,11,26,5,24,23,35),2,6)
h
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]   13   10    3   11    5   23
[2,]   55   30   20   26   24   35
```

Wir bestimmen die erwarteten Häufigkeiten mit Hilfe des äußeren Produkts.

```
m <- outer(apply(h,1,sum),apply(h,2,sum),FUN="*")/sum(h)

round(m,1)
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 17.3 10.2  5.9  9.4  7.4 14.8
[2,] 50.7 29.8 17.1 27.6 21.6 43.2
```

Unter der Nullhypothese sollten sich die beobachteten Häufigkeiten n_{ij} und die geschätzten erwarteten Häufigkeiten \hat{n}_{ij} nicht zu stark unterscheiden.

Als Test bietet sich der Chiquadrattest an.

Die Teststatistik des Chiquadrattests ist

$$X^2 = \sum_{i=1}^2 \sum_{j=1}^J \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$$

Diese erhalten wir in R durch

```
sum(((h-m)^2)/m)
[1] 10.85154
```

Unter der Nullhypothese ist diese Teststatistik approximativ chiquadratverteilt mit $J - 1$ Freiheitsgraden.

Wir lehnen die Nullhypothese ab, wenn gilt

$$X^2 \geq \chi_{J-1;1-\alpha}^2$$

wobei $\chi_{J-1;1-\alpha}^2$ das $1 - \alpha$ -Quantil einer Chiquadratverteilung mit $J - 1$ Freiheitsgraden ist.

Die Überschreitungswahrscheinlichkeit beträgt:

```
1-pchisq(sum(((m-h)^2)/m),ncol(h)-1)
[1] 0.05440427
```

In R können wir den Chi-Quadrat-Homogenitätstest mit der Funktion `chisq.test` durchführen.

```
chisq.test(h)
```

```
      Pearson's chi-square test without Yates'
      continuity correction
```

```
data:  h
X-squared = 10.8515,
df = 5,
p-value = 0.0544
```

Zum Niveau $\alpha = 0.05$ lehnen wir also die Nullhypothese nicht ab.

2.2.2 Stetige Variablen

Ein Beispiel

Es soll untersucht werden, ob ein aktives Einüben des Gehreflexes und des Plazierungsreflexes bei Neugeborenen dazu führt, daß die Kinder früher laufen. Dazu werden 10 Neugeborene zufällig auf zwei gleichgroße Gruppen aufgeteilt. Bei den Kindern der ersten Gruppe wird der Reflex nicht eingeübt, jedoch bei den Kindern der zweiten Gruppe.

Es ergab sich folgendes Alter in Monaten, in dem die Kinder zu laufen begannen:

ohne Einüben	mit Einüben
12.00	9.50
10.50	9.75
11.50	10.25
13.25	10.75
12.75	9.25

Das Modell

Wir gehen von folgender Situation aus:

Es werden $m+n$ Objekte zufällig ausgewählt und dann zufällig auf eine Gruppe mit m Beobachtungen und eine Gruppe mit n Beobachtungen aufgeteilt. Die Beobachtungen der ersten Gruppe erhalten dann die erste Behandlung und die Beobachtungen der zweiten Gruppe die zweite Behandlung.

Die den Beobachtungen der ersten Stichprobe zugrundeliegenden Zufallsvariablen sind

$$X_1, \dots, X_m$$

und die der zweiten Stichprobe sind

$$Y_1, \dots, Y_n$$

Wir gehen außerdem im folgenden davon aus, daß X_1, \dots, X_m identisch mit stetiger Verteilungsfunktion $F_X(x)$ verteilt sind und daß Y_1, \dots, Y_n identisch mit stetiger Verteilungsfunktion $F_Y(y)$ verteilt sind.

Außerdem unterstellen wir, daß

$$X_1, \dots, X_m, Y_1, \dots, Y_n$$

unabhängig sind.

Wir wollen überprüfen, ob die beiden Verteilungen sich hinsichtlich der Lage unterscheiden.

Das Testproblem lautet also

$$H_0 : F_X(z) = F_Y(z) \quad \text{für alle } z \in \mathfrak{R}.$$

gegen

$$H_1 : F_Y(z) = F_X(z - \Delta) \quad \text{mit } \Delta \neq 0.$$

Graphische Darstellung

Wir wollen uns zunächst ein Bild von den Daten machen.

Es liegt nahe, die Boxplots der beiden Stichproben nebeneinander zu zeichnen.

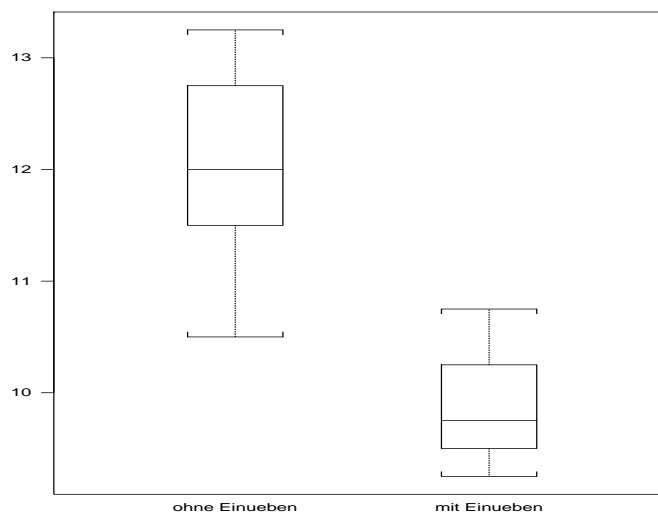
Wir weisen zunächst die beiden Stichproben den Variablen x und y zu:

```
x <- c(12.00, 10.50, 11.50, 13.25, 12.75)
```

```
y <- c(9.50, 9.75, 10.25, 10.75, 9.25)
```

und dann rufen wir die Funktion `boxplot` mit den beiden Variablen auf:

```
boxplot(x,y,boxcol=0,medline=T,medcol=1,outline=F,outpch="*",
medlwd=0.5, names=c("ohne Einueben", "mit Einueben"))
```



Die Boxplots deuten auf einen Lageunterschied hin. Es sieht aber auch so aus, daß sich die Variabilität in den beiden Gruppen unterscheidet.

Der t-Test

Die klassische Annahme ist, daß X normalverteilt ist mit den Parametern μ_X und σ_X^2 und Y normalverteilt mit den Parametern μ_Y und σ_Y^2 .

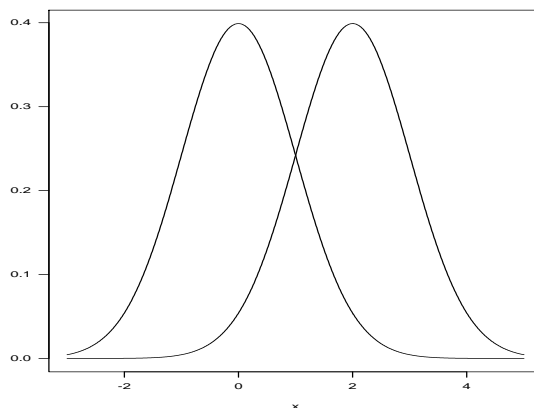
Zu testen ist

$$H_0 : \mu_x = \mu_Y \quad \text{gegen} \quad H_1 : \mu_x \neq \mu_Y$$

Da uns nur ein Lageunterschied interessiert, unterstellen wir, daß gilt

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2$$

Das folgende Bild zeigt die Dichtefunktionen von zwei normalverteilten Zufallsvariablen, die sich hinsichtlich der Lage unterscheiden.



Es liegt nahe, die Testentscheidung auf der Basis von $\bar{Y} - \bar{X}$ zu fällen. Unter den obigen Annahmen gilt

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma^2}{m}\right)$$

und

$$\bar{Y} \sim N\left(\mu_Y, \frac{\sigma^2}{n}\right)$$

Somit gilt

$$\bar{Y} - \bar{X} \sim N\left(\mu_Y - \mu_X, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)$$

Also gilt

$$\frac{\bar{Y} - \bar{X} - (\mu_Y - \mu_X)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0, 1)$$

Als Teststatistik bietet sich also an:

$$\frac{\bar{Y} - \bar{X} - (\mu_Y - \mu_X)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad (2.1)$$

Nun ist aber die Varianz σ^2 unbekannt.

Wir schätzen sie durch

$$\hat{\sigma}^2 = \frac{1}{m+n-2} \left(\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 \right)$$

Bei der Schätzung der Varianz berücksichtigen wir, daß die Daten aus unterschiedlichen Stichproben stammen.

Die Schätzfunktion $\hat{\sigma}^2$ ist eine Linearkombination der Stichprobenvarianzen

$$s_X^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

und

$$s_Y^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2,$$

Es gilt

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{m+n-2} \left(\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 \right) \\ &= \frac{1}{m+n-2} \left(\frac{m-1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2 + \frac{n-1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2 \right) \\ &= \frac{m-1}{m+n-2} s_X^2 + \frac{n-1}{m+n-2} s_Y^2 \end{aligned}$$

In der Teststatistik (2.1) ersetzen σ^2 durch $\hat{\sigma}^2$ und erhalten die geeignete Teststatistik

$$t = \frac{\bar{Y} - \bar{X} - (\mu_Y - \mu_X)}{\hat{\sigma} \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad (2.2)$$

Wenn

$$H_0 : \mu_X = \mu_Y$$

zutrifft, gilt

$$t = \frac{\bar{Y} - \bar{X}}{\hat{\sigma} \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

Wenn H_0 zutrifft, ist die Teststatistik t t-verteilt mit $m + n - 2$ Freiheitsgraden.

Die Entscheidungsregel beim zweiseitigen Test lautet somit Entscheidung für H_1 , wenn gilt

$$|t| > t_{m+n-2; 1-\alpha/2}.$$

Dabei ist $t_{m+n-2;p}$ das p-Quantil einer t-Verteilung mit $m + n - 2$ Freiheitsgraden.

Für das Datenbespiel gilt

$$\bar{x} = 12$$

$$\bar{y} = 9.9$$

$$s_X^2 = 1.15625$$

$$s_Y^2 = 0.3625$$

Also gilt

$$\begin{aligned} \hat{\sigma}^2 &= \frac{m-1}{m+n-2} s_X^2 + \frac{n-1}{m+n-2} s_Y^2 \\ &= \frac{4}{8} \cdot 1.15625 + \frac{4}{8} \cdot 0.3625 \\ &= 0.759375 \end{aligned}$$

Somit erhalten wir als Wert der Teststatistik

$$\begin{aligned} t &= \frac{9.9 - 12}{0.8714 \cdot 0.6324} \\ &= -3.81 \end{aligned}$$

Wegen $t_{8;0.975} = 2.306$ lehnen wir H_0 also ab.

Es können natürlich auch einseitige Tests durchgeführt werden.

$$H_0 : \mu_X \leq \mu_Y \quad \text{gegen} \quad H_1 : \mu_X > \mu_Y$$

Entscheidung für H_1 , wenn gilt

$$t < -t_{m+n-2;1-\alpha}.$$

oder

$$H_0 : \mu_X \geq \mu_Y \quad \text{gegen} \quad H_1 : \mu_X < \mu_Y$$

Entscheidung für H_1 , wenn gilt

$$t > t_{m+n-2;1-\alpha}.$$

In R kann man die Funktion `t.test` verwenden.

```
t.test(x,y,paired=FALSE)
```

```
Standard Two-Sample t-Test
```

```
data: x and y
```

```
t = 3.8103,
```

```
df = 8,
```

```
p-value = 0.0052
```

```
alternative hypothesis:
```

```
true difference in means is not equal to 0 95 percent
```

```
confidence interval:
```

```
0.82908 3.37092
```

```
sample estimates:
```

```
mean of x mean of y
```

```
12          9.9
```

Die Funktion liefert folgende Ergebnisse:

1. den Wert der Teststatistik $t=3.8103$
2. die Anzahl der Freiheitsgrade $df=8$
3. die Überschreitungswahrscheinlichkeit 0.0052
4. ein Konfidenzintervall für $\mu_X - \mu_Y$
5. die Mittelwerte \bar{X} und \bar{Y} .

Zum Signifikanzniveau $\alpha = 0.05$ wird die Nullhypothese abgelehnt, da die Überschreitungswahrscheinlichkeit 0.0052 kleiner als 0.05 ist.

Der Welch-Test

Der t-Test beruht auf der Annahme, daß die Varianzen der beiden Grundgesamtheiten gleich sind. Ist diese Annahme nicht erfüllt, so müssen wir die Varianzen der beiden Grundgesamtheiten getrennt schätzen.

Es liegt nahe, folgende Schätzer zu verwenden:

$$s_X^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

und

$$s_Y^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2$$

Es gilt:

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{m}\right)$$

und

$$\bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

Somit gilt

$$\bar{Y} - \bar{X} \sim N\left(\mu_Y - \mu_X, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}\right)$$

Also gilt

$$\frac{\bar{Y} - \bar{X} - (\mu_Y - \mu_X)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim N(0, 1)$$

Wir ersetzen σ_X^2 durch s_X^2 und σ_Y^2 durch s_Y^2 und erhalten folgende Teststatistik

$$t' = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}}$$

Diese ist unter H_0 nicht t-verteilt.

Von Welch wurde 1947 vorgeschlagen, die Freiheitsgrade der t-Verteilung so zu korrigieren, daß die Teststatistik approximativ t-verteilt ist.

Eine Herleitung dieses Vorschlags ist bei Miller: Beyond Anova, S.60-63 zu finden.

Die korrigierten Freiheitsgrade sind:

$$df = \frac{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{s_X^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{s_Y^2}{n}\right)^2}$$

Für das Datenbeispiel gilt

$$s_X^2 = 1.15625$$

$$s_Y^2 = 0.3625$$

Also erhalten wir als korrigierte Freiheitsgrade

$$\begin{aligned} df &= \frac{\left(\frac{1.15625}{5} + \frac{0.3625}{5}\right)^2}{\frac{1}{4} \left(\frac{1.15625}{5}\right)^2 + \frac{1}{4} \left(\frac{0.3625}{5}\right)^2} \\ &= 6.284 \end{aligned}$$

In R existiert die Möglichkeit, den Welch-Test durchzuführen.

Beim Aufruf der Funktion `t.test` muß man das Argument `var.equal` gleich `F` setzen.

```
t.test(x,y,var.equal=F)
```

```
Welch Modified Two-Sample t-Test
```

```
data: x and y
t = 3.8103,
df = 6.284,
p-value = 0.0081
```

```
alternative hypothesis:
true difference in means is not equal to 0
```

```
95 percent confidence interval:
0.7660406 3.4339594
```

```
sample estimates:
mean of x mean of y
12          9.9
```

Bei unserem Beispiel sind die Freiheitsgrade nicht mehr 8 wie unter der Annahme identischer Varianzen sondern 6.284. Die Überschreitungswahrscheinlichkeit ändert sich kaum.

Der F-Test

Der t-Test beruht auf der Annahme identischer Varianzen. Es liegt nahe, diese Annahme mit einem Test zu überprüfen.

Das Testproblem des zweiseitigen Tests lautet:

$$H_0 : \sigma_X^2 = \sigma_Y^2$$

$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

Es liegt nahe, die Stichprobenvarianzen

$$s_X^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

und

$$s_Y^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2$$

zu vergleichen.

Die Teststatistik des F-Tests ist gerade der Quotient dieser beiden Stichprobenvarianzen.

$$F = \frac{S_X^2}{S_Y^2}$$

Unter H_0 ist die Teststatistik F F-verteilt mit $m-1$ und $n-1$ Freiheitsgraden. Eine Herleitung ist bei Mood, Graybill, Boes (1974): Introduction to the theory of statistics, S.246 ff. zu finden.

In R gibt es für den F-Test die Funktion `var.test`.

```
var.test(x,y)
```

```
      F test for variance equality
```

```
data:  x and y
F = 3.1897,
num df = 4, denom df = 4,
p-value = 0.2874
```

alternative hypothesis:
true ratio of variances is not equal to 1

95 percent confidence interval:
0.332099 30.635138

sample estimates:
variance of x variance of y
1.15625 0.3625

Die Funktion liefert folgende Ergebnisse:

- den Wert der Teststatistik $F = 3.1897$
- die Anzahl der Freiheitsgrade des Zählers 4
- die Anzahl der Freiheitsgrade des Nenners 4
- die Überschreitungswahrscheinlichkeit 0.2874
- ein Konfidenzintervall für $\frac{\sigma_X^2}{\sigma_Y^2}$
- $s_X^2 = 1.15625$
- $s_Y^2 = 0.3625$

Obwohl die beiden Varianzen sich beträchtlich unterscheiden, lehnt der F-Test die Gleichheit nicht ab, da die Überschreitungswahrscheinlichkeit größer als α ist.

Der t-Test in der Praxis

Unter der Annahme der Normalverteilung liegt es nahe, bei einem Test auf Gleichheit der Erwartungswerte im unverbundenen Zweistichprobenproblem eine der drei folgenden Vorgehensweisen zu wählen:

1. immer den t-Test durchzuführen
2. immer den Welch-Test durchzuführen
3. erst den F-Test auf Gleichheit der Varianzen durchzuführen. Wird bei diesem die Nullhypothese abgelehnt, so wird der Welch-Test durchgeführt, ansonsten der t-Test.

Moser, Stevens[1992], American Statistician, S.19-21, haben die drei Vorgehensweisen miteinander verglichen.

Sie kommen zu folgendem Ergebnis:

Sind die beiden Stichprobenumfänge gleich, so unterscheiden sich die drei Vorgehensweisen nicht. Bei gleichem Stichprobenumfang sollte man also den t-Test durchführen, da dieser ein exakter Test ist.

Sind die beiden Stichprobenumfänge ungleich, so sollte man den Welch-Test durchführen, außer man weiß, daß das Verhältnis der Varianzen nahe am Wert 1 liegt.

Der Wilcoxon Rangsummentest

Der t-Test beruht auf der Annahme der Normalverteilung. Ist diese nicht gerechtfertigt, sollte man einen nichtparametrischen Test durchführen.

Der bekannteste ist der Wilcoxon-Rangsummentest.

Dieser beruht auf folgenden Annahmen:

Die Zufallsvariablen X_1, \dots, X_m seien unabhängig und identisch mit stetiger Verteilungsfunktion $F_X(x)$ verteilt sind und die Zufallsvariablen Y_1, \dots, Y_n seien unabhängig und identisch mit stetiger Verteilungsfunktion $F_Y(y)$. Das zweiseitige Testproblem lautet

$$H_0 : F_X(z) = F_Y(z) \quad \text{für alle } z \in \mathfrak{R}.$$

gegen

$$H_1 : F_Y(z) = F_X(z - \Delta) \quad \text{mit } \Delta \neq 0.$$

Unter der Nullhypothese kommen alle Beobachtungen aus einer Grundgesamtheit. Dies sollte sich in der Stichprobe dadurch zeigen, daß die Beobachtungen der beiden Stichproben gut gemischt sind. Es sollten also nicht alle Beobachtungen der einen Stichprobe an dem einen Ende und alle Beobachtungen der anderen Stichprobe nicht alle an dem anderen Ende der gemeinsamen geordneten Stichprobe liegen.

Schauen wir uns dazu den Fall $m=n=3$ an.

Die Konfiguration

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

deutet darauf hin, daß die Beobachtungen aus einer Grundgesamtheit kommen.

Die Konfiguration

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

und die Konfiguration

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

deuten darauf hin, daß sich die Grundgesamtheiten hinsichtlich der Lage unterscheiden.

Wie können wir diese Muster mit Hilfe einer geeigneten Teststatistik erkennen?

Der Wilcoxon Rangsummentest benutzt die Ränge $R(X_i)$ der X_i in der gemeinsamen Stichprobe $X_1, \dots, X_m, Y_1, \dots, Y_n$.

Der Rang $R(X_i)$ von X_i gibt an, wieviele von **allen** Beobachtungen kleiner oder gleich X_i sind.

Schauen wir uns dies für den Datensatz der Kinder an.

Es gilt

$$\begin{aligned}x_1 &= 12 \\x_2 &= 10.5 \\x_3 &= 11.5 \\x_4 &= 13.25 \\x_5 &= 12.75 \\y_1 &= 9.5 \\y_2 &= 9.75 \\y_3 &= 10.25 \\y_4 &= 10.75 \\y_5 &= 9.25\end{aligned}$$

Es gilt

$$R(x_1) = 8$$

denn 8 der Beobachtungen sind kleiner oder gleich 12.

Entsprechend erhalten wir

$$\begin{aligned}R(x_2) &= 5 \\R(x_3) &= 7 \\R(x_4) &= 10 \\R(x_5) &= 9\end{aligned}$$

Wie können wir die Ränge benutzen, um einen Lageunterschied aufzudecken?

Für

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

sind die Ränge der X_i gleich

$$1 \quad 4 \quad 5$$

Für

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

sind die Ränge der X_i gleich

$$1 \quad 2 \quad 3$$

Für

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

sind die Ränge der X_i gleich

$$4 \quad 5 \quad 6$$

Bildet man nun die Summe der Ränge der x -Werte, so ist diese im ersten Fall gleich 10, im zweiten Fall gleich 6 und im dritten Fall gleich 15.

Sehr kleine oder sehr große Werte der Summe der Ränge deuten also darauf hin, daß die Beobachtungen aus unterschiedlichen Verteilungen kommen.

Auf dieser Idee basiert der Wilcoxon Rangsummentest.

Seine Teststatistik lautet:

$$W = \sum_{i=1}^m R(X_i) \quad (2.3)$$

Im Beispiel gilt

$$W = 8 + 5 + 7 + 10 + 9 = 39.$$

Unter H_0 kann die exakte Verteilung von W für kleine Stichprobenumfänge einfach hergeleitet werden.

Es werden als Ränge die natürlichen Zahlen $1, 2, \dots, m + n$ vergeben.

Wenn H_0 zutrifft, stammen alle Beobachtungen aus der gleichen Grundgesamtheit, und jede Aufteilung der Ränge auf die beiden Stichproben ist gleichwahrscheinlich.

Für jede dieser Rangaufteilungen bestimmen wir den Wert von W .

Wir wollen dies für den Fall $m = n = 3$ durchführen.

Es gibt insgesamt

$$\binom{6}{3} = 20$$

Möglichkeiten, aus der Menge der Ränge $\{1, 2, 3, 4, 5, 6\}$ drei Ränge für die erste Stichprobe auszuwählen.

Alle diese Fälle und der zugehörige Wert von W sind in der folgenden Tabelle angegeben.

Rangkonfiguration	Wert von W
1,2,3	6
1,2,4	7
1,2,5	8
1,2,6	9
1,3,4	8
1,3,5	9
1,3,6	10
1,4,5	10
1,4,6	11
1,5,6	12
2,3,4	9
2,3,5	10
2,3,6	11
2,4,5	11
2,4,6	12
2,5,6	13
3,4,5	12
3,4,6	13
3,5,6	14
4,5,6	15

Durch einfaches Auszählen erhalten wir die Verteilung von W für $m = n = 3$:

w	P(W=w)
6	0.05
7	0.05
8	0.10
9	0.15
10	0.15
11	0.15
12	0.15
13	0.10
14	0.05
15	0.05

Für $m = n = 3$ gilt also $w_{0.05} = 6$ und $w_{0.10} = 7$.

In R gibt es eine Funktion `dwilcox`, die die exakte Verteilung des Wilcoxon-Rangsummentests liefert.

Wir müssen uns nur noch überlegen, was für die Stichprobenumfänge m und n der kleinste und der größte Wert von W sind.

Der kleinste Wert von W wird angenommen, wenn die kleinsten m Beobachtungen alle aus der 1. Stichprobe kommen.

In diesem Fall nimmt W den Wert

$$W = \sum_{i=1}^m i = \frac{m(m+1)}{2}$$

an.

Der größte Wert von W wird angenommen, wenn die größten m Beobachtungen alle aus der 1. Stichprobe kommen.

In diesem Fall nimmt W den Wert

$$\begin{aligned}
 W &= \sum_{i=n+1}^{m+n} i \\
 &= \sum_{i=1}^{m+n} i - \sum_{i=1}^n i \\
 &= \frac{(m+n)(m+n+1)}{2} - \frac{n(n+1)}{2} \\
 &= \frac{m(m+2n+1)}{2}
 \end{aligned}$$

an.

In R können wir diese Zahlen natürlich viel einfacher bestimmen:

```

m <- n <- 3

werte <- sum(1:m):sum((n+1):(m+n))

werte
[1] 6 7 8 9 10 11 12 13 14 15

cbind(werte,dwilcox(werte,m,n))

[1,] 6 0.05
[2,] 7 0.05
[3,] 8 0.10
[4,] 9 0.15
[5,] 10 0.15
[6,] 11 0.15
[7,] 12 0.15
[8,] 13 0.10
[9,] 14 0.05
[10,] 15 0.05

```

Die Entscheidungsregel beim zweiseitigen Test lautet:

Entscheidung für H_1 , wenn gilt

$$W \leq w_{\alpha/2} \quad \text{oder} \quad W \geq m(m+n+1) - w_{\alpha/2}$$

Im Datenbeispiel gilt

$$W = 39$$

Für $m = n = 5$ ist der maximale Wert von W gleich 40.

Nur die Rangkonfiguration $\{5, 7, 8, 9, 10\}$ liefert den Wert $W = 39$ und nur die Rangkonfiguration $\{6, 7, 8, 9, 10\}$ liefert den Wert $W = 40$.

Insgesamt gibt es

$$\binom{10}{5} = 252$$

Rangkonfigurationen.

Da wir einen zweiseitigen Test durchführen, beträgt die Überschreitungswahrscheinlichkeit

$$\frac{4}{252} = 0.0159$$

In R ist der Wilcoxon Rangsummentest in der Funktion `wilcox.test` implementiert:

```
wilcox.test(x,y)
  Exact Wilcoxon rank-sum test
data:  x and y
rank-sum statistic W = 39,
n = 5, m = 5,
p-value=0.0159

alternative hypothesis: true mu is not equal to 0
```

Sie liefert folgende Information

- den Wert der Teststatistik $W = 39$
- die Umfänge der beiden Stichproben
- die Überschreitungswahrscheinlichkeit 0.0159

Zum Signifikanzniveau $\alpha = 0.05$ wird H_0 also abgelehnt.

Für große Stichprobenumfänge ist es nicht einfach, die exakte Verteilung von W herzuleiten. In diesem Fall kann man auf die Asymptotik zurückgreifen.

Für große Werte von m und n ist die standardisierte Teststatistik

$$\frac{W - E(W)}{\sqrt{Var(W)}}$$

unter H_0 approximativ standardnormalverteilt.

Ein Beweis dieser Tatsache ist bei Randles, Wolfe: Introduction to the theory of Nonparametric Statistics zu finden.

Wir benötigen noch $E(W)$ und $Var(W)$.

Unter H_0 gilt

$$E(W) = \frac{m(N+1)}{2}$$

$$Var(W) = \frac{mn(N+1)}{12}$$

mit $N = m + n$.

Dies sieht man folgendermaßen:

In Satz 1.4.3 wurde gezeigt

$$E(R_i) = \frac{N+1}{2}$$

Also gilt

$$\begin{aligned} E(W) &= E\left(\sum_{i=1}^m R(X_i)\right) \\ &= \sum_{i=1}^m E(R(X_i)) \\ &= \sum_{i=1}^m \frac{N+1}{2} \\ &= \frac{m(N+1)}{2} \end{aligned}$$

Außerdem wurde in SATZ 1.4.3 gezeigt:

$$Var(R_i) = \frac{N^2 - 1}{12}$$

$$Cov(R_i, R_j) = -\frac{N+1}{12}$$

Also gilt

$$\begin{aligned}
 \text{Var}(W) &= \text{Var}\left(\sum_{i=1}^m R(X_i)\right) \\
 &= \sum_{i=1}^m \text{Var}(R(X_i)) + \sum_{i \neq j} \text{Cov}(R(X_i), R(X_j)) \\
 &= \sum_{i=1}^m \frac{N^2 - 1}{12} - \sum_{i \neq j} \frac{N + 1}{12} \\
 &= \frac{m(N^2 - 1)}{12} - \frac{m(m - 1)(N + 1)}{12} \\
 &= \frac{mn(N + 1)}{12}
 \end{aligned}$$

In vielen praktischen Anwendungen kommen Bindungen vor.

Bei dem folgenden Beispiel aus Schlittgen: Einführung in die Statistik ist dies der Fall.

Beispiel 2.2.1 *In einer Abteilung eines Krankenhauses wurden bei Stichproben aus zwei Patientenkollektiven, die an ähnlichen Krankheiten litten, die Verweildauer der Patienten im Krankenhaus in Tagen ermittelt.*

Es ergaben sich folgende Werte:

Kollektiv 1:

1	3	5	7	10	11	12	12	14	14	15
15	15	15	15	16	17	21	21	22	30	

Kollektiv 2:

1	1	1	2	2	4	4	4	5	5	6	7
9	9	10	10	10	10	11	11	14	15	19	23

Die 1 kommt viermal vor. Also erhält jede 1 den Rang

$$\frac{1 + 2 + 3 + 4}{4} = 2.5$$

Wir geben die Daten in R ein:

```
x <- c(1,3,5,7,10,11,12,12,14,14,15,15,15,15,16,17,21,21,22,30)
```

```
y <- c(1,1,1,2,2,4,4,4,5,5,6,7,9,9,10,10,10,10,11,11,14,15,19,23)
```

und bestimmen die Ränge der Beobachtungen des ersten Kollektivs:

```
rank(c(x,y))[1:length(x)]
```

```
[1]  2.5  7.0 12.0 15.5 21.0 25.0 27.5 27.5 30.0 30.0
     34.5 34.5 34.5 34.5 34.5 38.0 39.0 41.5 41.5 43.0 45.0
```

Der Wert von W ist

```
sum(rank(c(x,y))[1:length(x)])
```

```
[1] 618.5
```

Wir müssen wie schon im Einstichprobenproblem beim Wilcoxon-Vorzeichen-Rangtest die Varianz der Teststatistik modifizieren.

Es gilt

$$\text{Var}(W) = \frac{mn}{12} \left[m + n + 1 - \frac{1}{N(N-1)} \sum_{j=1}^r (b_j^3 - b_j) \right]$$

Dabei ist r die Anzahl der Gruppen mit Bindungen und b_j die Anzahl der Beobachtungen in der j -ten Bindungsgruppe.

Die erste Bindungsgruppe ist zum Beispiel die Zahl 1. Sie kommt viermal vor.

Deshalb ist

$$b_1 = 4.$$

Somit ist folgende Größe approximativ standardnormalverteilt, wobei wir die Stetigkeitskorrektur verwenden:

$$\frac{W - 0.5 - \frac{m(N+1)}{2}}{\sqrt{\frac{mn}{12} \left[m + n + 1 - \frac{1}{N(N-1)} \sum_{j=1}^r (b_j^3 - b_j) \right]}}$$

In R erhalten wir

```
wilcox.test(x,y)
```

```
Wilcoxon rank-sum test
```

```
data: x and y
```

```
rank-sum normal statistic with correction Z =3.0797,  
p-value =0.0021
```

```
alternative hypothesis: true mu is not equal to 0
```

```
Warning messages:
```

```
cannot compute exact p-value with ties in:
```

```
wil.rank.sum(x, y, alternative, exact, correct)
```

Sehr oft soll ein ordinalskaliertes Merkmal in zwei Gruppen verglichen werden. Die Daten liegen in der Regel in Form einer Kontingenztafel vor, was dazu führt, daß der Chi-Quadrat-Test durchgeführt wird. Die Anwendung des Wilcoxon-Rangsummentest wäre aber sinnvoller.

Schauen wir uns ein Beispiel an.

Beispiel 2.2.2 *Die Besucher des Films 'Titanic' wurden gefragt, wie ihnen der Film gefallen hat.*

Die folgende Tabelle zeigt das Ergebnis der Befragung für die Frauen und die Männer.

	<i>sehr gut</i>	<i>gut</i>	<i>mittelmäßig</i>
<i>weiblich</i>	13	10	2
<i>männlich</i>	6	7	7

Es soll untersucht werden, ob der Film von Frauen und Männern unterschiedlich eingeschätzt wird.

Wir geben die Daten als Matrix ein.

```
titanic <- matrix(c(13,6,10,7,2,7),2,3)
```

```
titanic
      [,1] [,2] [,3]
[1,]  13   10   2
[2,]   6    7   7
```

und führen den Chi-Quadrat-Test durch.

```
chisq.test(titanic)
```

```

Pearson's chi-square test
without Yates' continuity correction
```

```
data:  titanic
X-squared = 5.3972,
df = 2,
p-value = 0.0673
```

Warning messages:

```

Expected counts < 5. Chi-squared approximation may not be
appropriate. in: chisq.test(titanic)
```


Wir sehen, daß die Nullhypothese identischer Verteilungen bei Frauen und Männern zum Niveau $\alpha = 0.05$ nicht abgelehnt wird.

Wenn wir den Chiquadrattest auf den Datensatz anwenden, berücksichtigen wir nicht sämtliche in den Daten enthaltene Information.

Die Einschätzung ist nämlich ordinal.

Wir können das Testproblem auch als unverbundenes Zweistichprobenproblem auffassen, wobei die Daten ordinal mit sehr vielen Bindungen sind.

Kodierten wir 'sehr gut' mit 1, 'gut' mit 2 und 'mittelmäßig' mit 3, so liegen die Daten in folgender Form vor:

Bei den Frauen erhalten wir folgende Werte

```
1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3
```

und bei den Männern erhalten wir folgende Werte

```
1 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3
```

Wir können uns die Daten in R folgendermaßen erzeugen:

```
frauen <- rep(1:3,titanic[1,])
```

```
maenner <- rep(1:3,titanic[2,])
```

Wir führen den Wilcoxon-Rangsummentest durch:

```
wilcox.test(x,y)
```

```
Wilcoxon rank-sum test
```

```
data: x and y
```

```
rank-sum normal statistic with correction Z = -2.0213,
```

```
p-value = 0.0433
```

```
alternative hypothesis: true mu is not equal to 0
```

```
Warning messages:
```

```
cannot compute exact p-value with ties in:
```

```
wil.rank.sum(x, y,alternative, exact, correct)
```

Wir sehen, daß die Nullhypothese identischer Verteilungen bei Männern und Frauen beim Wilcoxon-Rangsummentest abgelehnt wird.

Der Median-Test

Der t-Test im unverbundenen Zweistichprobenproblem verwendet die Beobachtungen, während der Rangsummentest auf den Rängen basiert. Insofern entsprechen sie dem t-Test und dem Vorzeichenrang-Test im Einstichprobenproblem.

Bisher haben wir keinen Test, der dem Vorzeichentest entspricht. Bei diesem wird nur gezählt.

Dies leistet im unverbundenen Zweistichprobenproblem der Median-Test.

Er basiert auf folgender Idee:

Wenn alle Beobachtungen aus einer Grundgesamtheit kommen, erwarten wir, daß die Hälfte der Beobachtungen jeder Stichprobe kleiner als der gemeinsame Median M ist.

Wir bestimmen also die Werte folgender Zufallsvariablen:

N_{11} Anzahl der Beobachtungen in der ersten Stichprobe, die kleiner als der gemeinsame Median M sind

N_{21} Anzahl der Beobachtungen in der zweiten Stichprobe, die kleiner als der gemeinsame Median M sind

Dann können wir die Daten folgendermaßen in einer Kontingenztabelle zusammenstellen:

	$< M$	$> M$	
1. Stichprobe	N_{11}	$m - N_{11}$	m
2. Stichprobe	N_{21}	$n - N_{21}$	n
	$N_{11} + N_{21}$	$m + n - N_{11} - N_{21}$	$m + n$

Die Ausgangssituation und das Testproblem sind mit denen von Fishers-Test identisch.

Wir können also Fishers Test anwenden.

Im unverbundenen Zweistichprobenproblem mit stetigen Merkmalen heißt dieser Median-Test.

Schauen wir uns dies für das Datenbeispiel an.

Es gilt

$$\begin{aligned}x_1 &= 12 \\x_2 &= 10.5 \\x_3 &= 11.5 \\x_4 &= 13.25 \\x_5 &= 12.75 \\y_1 &= 9.5 \\y_2 &= 9.75 \\y_3 &= 10.25 \\y_4 &= 10.75 \\y_5 &= 9.25\end{aligned}$$

Der geordnete Datensatz lautet

9.25 9.50 9.75 10.25 10.50 10.75 11.50 12.00 12.75 13.25

Der Median ist 10.625.

In der ersten Stichprobe ist eine Beobachtung kleiner als 10.625, während in der zweiten Stichprobe 4 Beobachtungen kleiner als 10.625 sind.

Wir erhalten somit folgende Kontingenztabelle:

1	4
4	1

Auf diese können wir Fishers exakten Test anwenden.

In R geht das dann so: Wir bestimmen den Median beider Stichproben:

```
m <- median(c(x,y))
```

```
m
[1] 10.625
```

Dann zählen wir, wieviele der Beobachtungen in der ersten Stichprobe kleiner als m sind

```
sum(x<m)
[1] 1
```

und wieviele der Beobachtungen in der zweiten Stichprobe größer als m sind

```
sum(x>m)
[1] 4
```

Danach kommt die zweite Stichprobe dran

```
sum(y<m)
[1] 4
```

```
sum(y>m)
[1] 1
```

Danach bauen wir die Kontingenztabelle auf:

```
h <- matrix(c(sum(x<m),sum(x>m),sum(y<m),sum(y>m)),2,2)
```

```
h
      [,1] [,2]
[1,]    1    4
[2,]    4    1
```

Fishers Test für die Tabelle h ergibt dann das Ergebnis.

```
fisher.test(h)
```

```
Fisher's exact test
```

```
data: h
p-value = 0.2063
alternative hypothesis: two.sided
```

Kapitel 3

Das c -Stichprobenproblem

Im Zweistichprobenproblem soll überprüft werden, ob sich zwei Behandlungen hinsichtlich ihrer Wirkung unterscheiden. Sollen mehr als zwei Behandlungen verglichen werden so spricht man vom c -Stichprobenproblem.

Auch hier kann man auf zwei Arten vorgehen:

Man kann $N = n_1 + \dots + n_c$ Objekte auswählen, diese auf c Gruppen der Umfänge n_i , $i = 1, \dots, c$ aufteilen und alle Objekte einer Gruppe mit einer der Behandlungen versehen.

Man spricht in diesem Fall vom **unverbundenen c -Stichprobenproblem**.

Die andere Vorgehensweise besteht darin, n Blöcke zu bilden, die aus jeweils c ähnlichen Objekten bestehen. Jede Behandlung wird dann genau einem Objekt innerhalb eines Blockes zugeordnet, so daß innerhalb eines Blockes alle Behandlungen vorliegen. Dabei kann ein Block natürlich ein einzelnes Objekt sein, zum Beispiel eine Person, der die Behandlungen zu unterschiedlichen Zeitpunkten zugeordnet werden. In diesem Fall spricht man vom **verbundenen c -Stichprobenproblem**.

3.1 Unverbundene Stichproben

3.1.1 Einfaktorielle Varianzanalyse

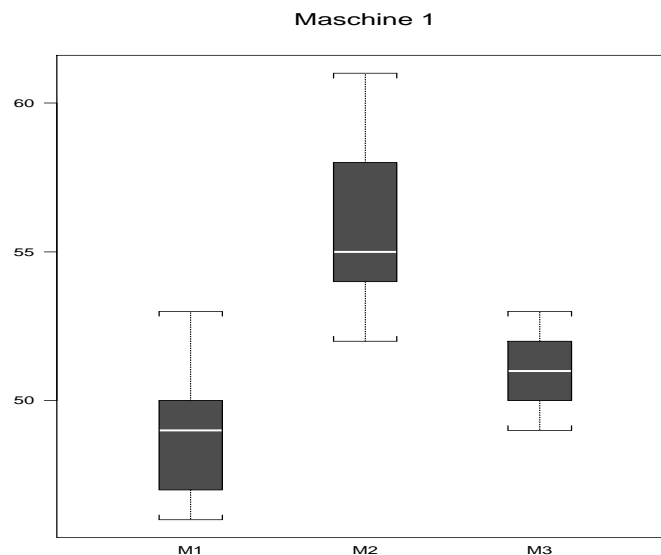
Drei Maschinen sollen hinsichtlich ihrer stündlichen Ausbringungsmenge verglichen werden. Da die stündliche Ausbringungsmenge Zufallsschwankungen unterliegt, wird überprüft, ob die durchschnittlichen Ausbringungsmengen der Maschinen gleich sind. Dazu werden bei jeder Maschine die Ausbringungsmengen von fünf unterschiedlichen Stunden bestimmt.

Sei x_{ij} die Ausbringungsmenge an Maschine i zum Zeitpunkt j , $i = 1, 2, 3$, $j = 1, 2, 3, 4, 5$.

Es ergaben sich folgende Werte:

Maschine	Ausbringungsmenge
1	47 53 49 50 46
2	55 54 58 61 52
3	52 50 51 53 49

Machen wir uns zunächst ein Bild von den Daten. Hierzu erstellen wir die Boxplots der drei Maschinen.



Wir sehen, daß sich die Ausbringungsmengen der drei Maschinen hinsichtlich der Lage unterscheiden.

Wir wollen nun überprüfen, ob dieser Unterschied signifikant ist.

Unser Ziel ist es also herauszufinden, ob sich die erwarteten Ausbringungsmengen μ_i der Maschinen unterscheiden.

Hierzu benötigen wir eine geeignete Teststatistik.

Ein Unterschied in den Erwartungswerten μ_i sollte sich natürlich auch in den Mittelwerten \bar{x}_i zeigen, für die gilt

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (3.1)$$

Wir schauen uns also die drei Mittelwerte an.

Es gilt

$$\bar{x}_1 = 49$$

$$\bar{x}_2 = 56$$

$$\bar{x}_3 = 51$$

Wir sehen, daß sich die drei Mittelwerte unterscheiden. Wie können wir diesen Unterschied durch eine geeignete Maßzahl beschreiben?

Bei zwei Stichproben ist dies einfach. Wir schauen uns die Differenz $\bar{x}_j - \bar{x}_i$ der beiden Mittelwerte an.

Bei mehr als drei Stichproben können wir alle paarweisen Vergleiche durchführen, also \bar{x}_1 mit \bar{x}_2 , \bar{x}_1 mit \bar{x}_3 und \bar{x}_2 mit \bar{x}_3 vergleichen.

Hierdurch erhalten wir aber kein globales Maß für den Vergleich aller drei Stichproben.

Um dieses zu erhalten, fassen wir die drei Mittelwerte als eine Stichprobe auf und schauen, wie stark diese um den Mittelwert aller drei Mittelwerte streuen.

Wir bilden also

$$\bar{x} = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{n_i} x_{ij} \quad (3.2)$$

Wegen

$$\sum_{j=1}^{n_i} x_{ij} = n_i \bar{x}_i \quad (3.3)$$

gilt

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{i=1}^c n_i \bar{x}_i \\ &= \sum_{i=1}^c \frac{n_i}{N} \bar{x}_i\end{aligned}$$

Im Beispiel gilt

$$\bar{x} = 52$$

Wir sehen, daß die ersten beiden Mittelwerte weit vom Gesamtmittel entfernt sind.

Nun schauen wir uns die Streuung der Stichprobenmittelwerte \bar{x}_i um das Gesamtmittel \bar{x} an, wobei wir die quadrierten Abweichungen der Stichprobenmittelwerte \bar{x}_i vom Gesamtmittelwert \bar{x} noch mit den Stichprobenumfängen n_i gewichten:

$$SS_A = \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})^2$$

Wir nennen SS_A auch die Streuung zwischen den Stichproben.

Im Beispiel gilt

$$\begin{aligned}SS_A &= 5 \cdot (49 - 52)^2 + 5 \cdot (56 - 52)^2 + 5 \cdot (51 - 52)^2 \\ &= 130\end{aligned}$$

SS_A allein ist aber noch keine geeignete Teststatistik zur Überprüfung der Gleichheit der Erwartungswerte.

In der folgenden Tabelle sind die Ausbringungsmengen von drei anderen Maschinen angegeben.

Maschine	Ausbringungsmenge
1	50 42 53 45 55
2	48 57 65 59 51
3	57 59 48 46 45

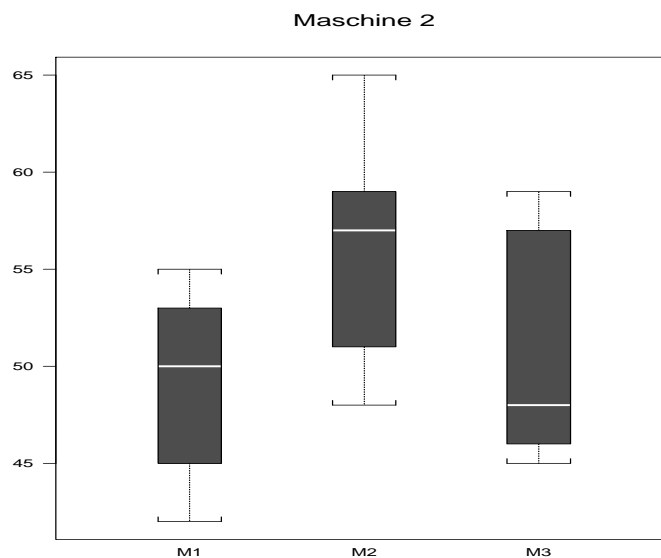
Für die Mittelwerte der drei Maschinen gilt ebenfalls

$$\bar{x}_1 = 49$$

$$\bar{x}_2 = 56$$

$$\bar{x}_3 = 51$$

Schauen wir uns die Boxplots der drei Maschinen an.



Die Streuungen der Ausbringungsmengen der drei Maschinen ist in diesem Fall sehr viel größer als im ersten Fall.

Die Streuungen der Werte der einzelnen Maschinen sind im zweiten Fall so groß, daß die drei Stichproben auch von einer Maschine stammen könnten. Die unterschiedlichen Mittelwerte können also allein aufgrund dieser großen Streuung zustande gekommen sein und nicht dadurch, daß die Maschine unterschiedliche Niveaus in der Produktion aufweisen.

Im ersten Fall hingegen ist die Streuung innerhalb der einzelnen Stichproben so gering, daß es plausibel erscheint, daß der Unterschied in den Mittelwerten an den unterschiedlichen Erwartungswerten liegt.

Eine Teststatistik eines Tests auf Gleichheit der Erwartungswerte sollte also nicht nur die Unterschiede zwischen den Mittelwerten sondern auch die Streuung innerhalb der Stichproben berücksichtigen.

Ein sinnvolles Maß für die Streuung innerhalb einer Stichprobe ist die Summe der quadrierten Abweichungen der Beobachtungen vom Stichprobenmittelwert:

$$\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Für das erste Beispiel gilt

$$\begin{aligned} \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2 &= (47 - 49)^2 + (53 - 49)^2 + (49 - 49)^2 + \\ &\quad + (50 - 49)^2 + (46 - 49)^2 \\ &= 30 \end{aligned}$$

$$\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 = 50$$

$$\sum_{j=1}^{n_3} (x_{3j} - \bar{x}_3)^2 = 10$$

Für das zweite Beispiel gilt:

$$\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2 = 118$$

$$\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 = 180$$

$$\sum_{j=1}^{n_3} (x_{3j} - \bar{x}_3)^2 = 170$$

Die Streuung innerhalb aller Stichproben berücksichtigen wir dadurch, daß wir die Streuungen innerhalb aller drei Stichproben addieren:

$$SS_R = \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Wir nennen SS_R auch die Streuung innerhalb der Stichproben.

Im ersten Beispiel gilt:

$$SS_R = 30 + 50 + 10 = 90$$

Im zweiten Beispiel gilt:

$$SS_R = 118 + 180 + 170 = 468$$

Betrachten wir nun die Situation allgemein:

Wir gehen aus von den unabhängigen Zufallsvariablen

$$X_{11}, \dots, X_{1n_1}, \dots, X_{c1}, \dots, X_{cn_c}$$

wobei

$$X_{i1}, \dots, X_{in_i}$$

identisch mit stetiger Verteilungsfunktion $F_i(z)$ verteilt sind.

Zu testen ist:

$$H_0 : F_1(z) = \dots = F_c(z) \quad \text{für alle } z \in \mathfrak{R}$$

gegen

$$H_1 : F_i(z) \neq F_j(z) \quad \text{für mindestens ein Paar } (i, j) \text{ mit } i \neq j$$

Die klassische Annahme ist, daß in den Grundgesamtheiten Normalverteilungen vorliegen, wobei gilt

$$X_{ij} \sim N(\mu_i, \sigma^2).$$

Es wird also unterstellt, daß die Varianz in allen Grundgesamtheiten gleich ist.

Das Testproblem lautet:

$$H_0 : \mu_1 = \dots = \mu_c$$

gegen

$$H_1 : \mu_i \neq \mu_j \quad \text{für mindestens ein Paar } (i, j) \text{ mit } i \neq j$$

Die Teststatistik für dieses Testproblem vergleicht nun die Streuung zwischen den Stichproben mit der Streuung innerhalb der Stichproben.

Wir haben bereits gesehen, daß dies eine sinnvolle Teststatistik ist, da es nicht ausreicht, die Streuung zwischen den Stichproben allein zu betrachten.

Wirft man die Beobachtungen aller Stichproben in einen Topf, so kann man die Streuung aller Beobachtungen um das Gesamtmittel bestimmen:

$$SS_T = \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

Im ersten Beispiel erhalten wir:

$$\begin{aligned} SS_T &= (47 - 52)^2 + (53 - 52)^2 + (49 - 52)^2 + (50 - 52)^2 + (46 - 52)^2 \\ &+ (55 - 52)^2 + (54 - 52)^2 + (58 - 52)^2 + (61 - 52)^2 + (52 - 52)^2 \\ &+ (52 - 52)^2 + (50 - 52)^2 + (51 - 52)^2 + (53 - 52)^2 + (49 - 52)^2 \\ &= 220 \end{aligned}$$

Hier gilt

$$SS_A = 130$$

und

$$SS_R = 30 + 50 + 10 = 90$$

Wir sehen, daß für das erste Beispiel gilt

$$SS_T = SS_A + SS_R$$

Im zweiten Beispiel gilt

$$\begin{aligned} SS_T &= (50 - 52)^2 + (42 - 52)^2 + (53 - 52)^2 + (45 - 52)^2 + (55 - 52)^2 \\ &+ (48 - 52)^2 + (57 - 52)^2 + (65 - 52)^2 + (59 - 52)^2 + (51 - 52)^2 \\ &+ (57 - 52)^2 + (59 - 52)^2 + (48 - 52)^2 + (46 - 52)^2 + (45 - 52)^2 \\ &= 598 \end{aligned}$$

Hier gilt ebenfalls

$$SS_A = 130$$

während die Streuung innerhalb der Stichproben viel größer ist

$$SS_R = 30 + 50 + 10 = 468$$

Wir sehen, daß auch für das zweite Beispiel gilt

$$SS_T = SS_A + SS_R$$

Diese Beziehung gilt allgemein, wie man folgendermaßen sieht:

$$\begin{aligned}
 SS_T &= \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 \\
 &= \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i + \bar{x}_i - \bar{x})^2 \\
 &= \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^c \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 \\
 &\quad + 2 \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) (\bar{x}_i - \bar{x}) \\
 &= \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})^2 \\
 &\quad + 2 \sum_{i=1}^c (\bar{x}_i - \bar{x}) \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) \\
 &= \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})^2 \\
 &= SS_A + SS_R
 \end{aligned}$$

da gilt

$$\begin{aligned}
 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) &= \sum_{j=1}^{n_i} x_{ij} - \sum_{j=1}^{n_i} \bar{x}_i = \\
 &= n_i \bar{x}_i - n_i \bar{x}_i = \\
 &= 0
 \end{aligned}$$

Beim Vergleich werden aber die mittleren Streuungen betrachtet, wobei der Mittelwert unter der Nebenbedingung bestimmt wird, wieviele der Summanden frei gewählt werden können.

Die Streuung zwischen den Stichproben setzt sich c Summanden zusammen, von denen aber nur $c - 1$ frei gewählt werden können, da sich der Mittelwert der c -ten Stichprobe aus

$$\bar{x}, \bar{x}_1, \dots, \bar{x}_{c-1}$$

ergibt.

Die Streuung innerhalb der Stichproben setzt sich aus N Summanden zusammen.

In der i -ten Stichprobe ergibt sich aber x_{in_i} aus der Kenntnis von

$$x_{i1}, \dots, x_{in_i-1}, \bar{x}_i.$$

Somit sind von den N Summanden nur $N - c$ frei wählbar.

Wir erhalten also

$$MSS_A = \frac{1}{c-1} \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})^2$$

und

$$MSS_R = \frac{1}{N-c} \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Die Teststatistik ist

$$\begin{aligned} F &= \frac{MSS_A}{MSS_R} \\ &= \frac{\frac{1}{c-1} \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})^2}{\frac{1}{N-c} \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \end{aligned}$$

Ist die mittlere Streuung zwischen den Stichproben groß im Verhältnis zur mittleren Streuung innerhalb der Stichproben, so wird die Nullhypothese identischer Erwartungswerte abgelehnt.

Unter der Nullhypothese ist die Teststatistik F-verteilt mit $c - 1$ und $N - c$ Freiheitsgraden.

In der Regel werden die Ergebnisse einer Varianzanalyse in einer ANOVA-Tabelle zusammengestellt:

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Stichproben	SS_A	$c-1$	MSS_A	$\frac{MSS_A}{MSS_R}$
Rest	SS_R	$N-c$	MSS_R	
Gesamt	SS_T	$N-1$		

Für das erste Beispiel ergibt sich folgende ANOVA-Tabelle:

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Stichproben	130	2	65	8.7
Rest	90	12	7.5	
Gesamt	220	14		

Wegen $F_{2,12;0.95} = 5.89$ wird die Nullhypothese zum Niveau $\alpha = 0.05$ abgelehnt.

Für das zweite Beispiel ergibt sich folgende ANOVA-Tabelle:

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Stichproben	130	2	65	1.7
Rest	468	12	39	
Gesamt	598	14		

Wegen $F_{2,12;0.95} = 5.89$ wird die Nullhypothese zum Niveau $\alpha = 0.05$ nicht abgelehnt.

In R kann man mit Hilfe der Funktion `aov` eine Varianzanalyse durchführen. Hierzu muß man folgendes leisten:

Man erzeugt eine Variable `anzahl` mit den Werten aller drei Maschinen:

```
anzahl <- c(47,53,49,50,46,55,54,58,61,52,53,50,51,52,49)
```

Die ersten 5 Beobachtungen in `anzahl` stammen von Maschine 1, die nächsten 5 von Maschine 2 und die letzten 5 von Maschine 3.

Dies teilt man R mit durch

```
maschine <- factor(rep(1:3,rep(5,3)))
maschine
[1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
```

Mit den beiden Variablen erzeugen wir einen `dataframe`

```
maschine1.df <- data.frame(maschine,anzahl)
```

den wir uns durch Aufruf anschauen können:

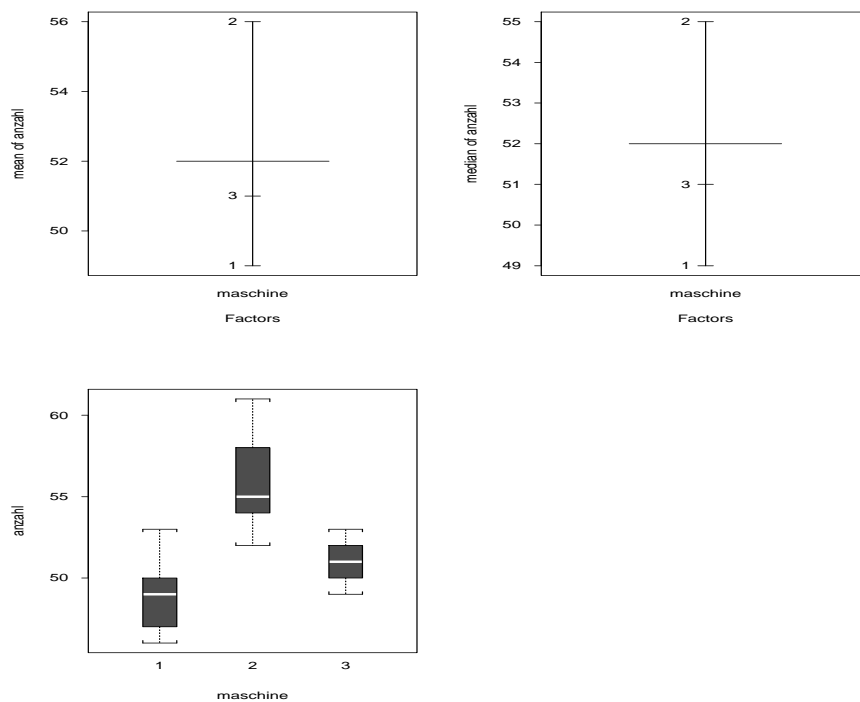
```
maschine1.df
  machine anzahl
1       1     47
2       1     53
3       1     49
4       1     50
5       1     46
6       2     55
7       2     54
8       2     58
9       2     61
10      2     52
11      3     53
12      3     50
13      3     51
14      3     52
15      3     49
```


Vor der Durchführung einer Varianzanalyse sollte man sich die Daten ansehen. Hierzu gibt es eine Reihe von Möglichkeiten in R .

Die Befehlsfolge

```
par(mfrow=c(2,2))
plot.design(maschine1.df)
plot.design(maschine1.df,fun=median)
plot.factor(maschine1.df)
```

erzeugt die folgenden Bilder:



In der ersten Graphik werden die Mittelwerte und in der zweiten die Mediane der Stichproben gezeichnet.

Die letzte Graphik zeigt die Boxplots der drei Gruppen, in denen sich ein Lageunterschied der Gruppen zeigt.

Die Streuungen in den einzelnen Gruppen unterscheiden sich nicht zu stark.

Jetzt kann man die Funktion `aov` aufrufen durch

```
maschine1.aov <- aov(anzahl~maschine,maschine1.df)
```

Die ANOVA-Tabelle erhält man durch

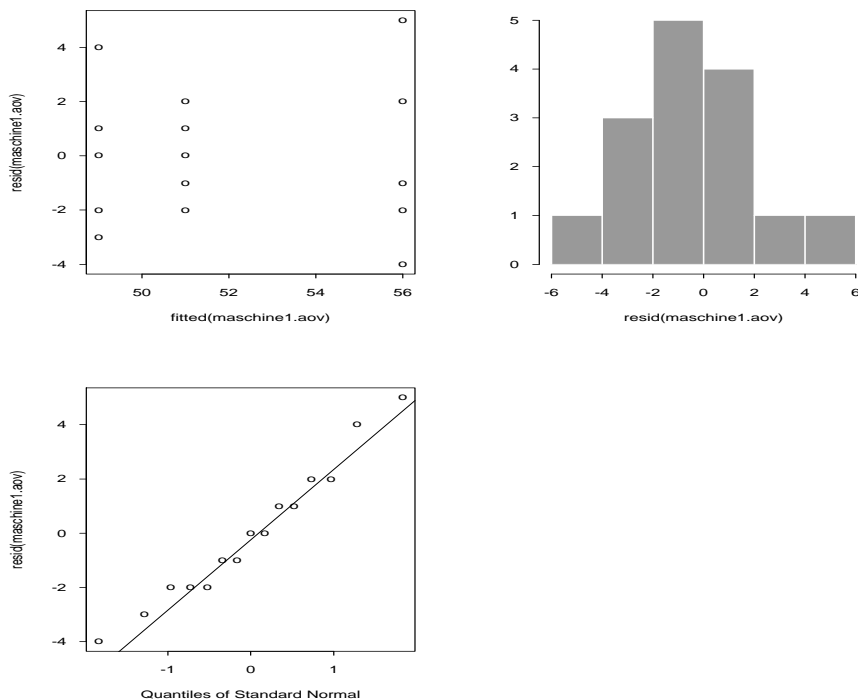
```
summary(maschine1.aov)
```

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
maschine	2	130	65.0	8.666667	0.004687259
Residuals	12	90	7.5		

Die Güte der Anpassung kann man nun wieder mit einigen Bildern überprüfen.

```
par(mfrow=c(2,2))
plot(fitted(maschine1.aov),resid(maschine1.aov))
hist(resid(maschine1.aov))
qqnorm(resid(maschine1.aov))
qqline(resid(maschine1.aov))
```

zeichnet die Residuen gegen die Gruppenmittelwerte, das Histogramm und den normal probability plot.



3.1.2 Kruskal-Wallis-Test

Ist die Annahme der Normalverteilung nicht gerechtfertigt, so sollte man einen nichtparametrischen Test durchführen.

Am bekanntesten ist der Kruskal-Wallis-Test, der eine Verallgemeinerung des Rangsummentests für das c-Stichprobenproblem darstellt.

Schauen wir uns diesen für das Beispiel der drei Maschinen an.

Hier sind noch einmal die Daten.

Maschine	Ausbringungsmenge
1	47 53 49 50 46
2	55 54 58 61 52
3	52 50 51 53 49

Da Bindungen vorliegen, nehmen wir von jeder Maschine zunächst die ersten drei Beobachtungen.

Wir erhalten also folgende Stichproben:

$$\begin{aligned} x_{11} &= 47 & x_{12} &= 53 & x_{13} &= 49 \\ x_{21} &= 55 & x_{22} &= 54 & x_{23} &= 58 \\ x_{31} &= 52 & x_{32} &= 50 & x_{33} &= 51 \end{aligned}$$

Beim Wilcoxon-Rangsummentest werden die Ränge aller Beobachtungen in der gemeinsamen Stichprobe bestimmt. Danach werden die Rangsummen der Ränge der einzelnen Stichproben bestimmt.

Genauso wird beim Kruskal-Wallis-Test vorgegangen.

Wir bestimmen für jede Beobachtung X_{ij} der Rang R_{ij} in der gemeinsamen Stichprobe

$$X_{11}, \dots, X_{1n_1}, \dots, X_{c1}, \dots, X_{cn_c}.$$

Die erste Beobachtung x_{11} der ersten Stichprobe ist die kleinste aller Beobachtungen. Also erhält sie den Rang 1.

Es gilt also

$$r_{11} = 1$$

Analog erhalten wir die Ränge der anderen Beobachtungen.

$$\begin{aligned} r_{11} &= 1 & r_{12} &= 6 & r_{13} &= 2 \\ r_{21} &= 8 & r_{22} &= 7 & r_{23} &= 9 \\ r_{31} &= 5 & r_{32} &= 3 & r_{33} &= 4 \end{aligned}$$

Beim Rangsummentest ist die Teststatistik die Rangsumme der ersten Stichprobe. Bei zwei Stichproben ist dies sinnvoll, da die Summe aller Ränge konstant ist. Bei mehr als zwei Stichproben müssen wir anders vorgehen.

Wir bilden die Summen der Ränge in den Stichproben:

$$R_i = \sum_{j=1}^{n_i} R_{ij}$$

Im Beispiel erhalten wir

$$\begin{aligned} R_1 &= 1 + 6 + 2 \\ &= 9 \\ R_2 &= 8 + 7 + 9 \\ &= 24 \\ R_3 &= 5 + 3 + 4 \\ &= 12 \end{aligned}$$

Diese Rangsummen vergleichen wir nun mit ihren Erwartungswerten $E(R_i)$ unter H_0 .

Bei N Beobachtungen werden die Ränge $1, 2, \dots, N$ vergeben, wenn keine Bindungen vorliegen.

Der erwartete Rang $E(R_{ij})$ einer Beobachtung ist also

$$E(R_{ij}) = \frac{N+1}{2}.$$

Die erwartete Rangsumme der i -ten Gruppe ist somit

$$\begin{aligned} E(R_i) &= E\left(\sum_{j=1}^{n_i} R_{ij}\right) \\ &= \sum_{j=1}^{n_i} E(R_{ij}) \\ &= \sum_{j=1}^{n_i} \frac{N+1}{2} \\ &= \frac{n_i(N+1)}{2} \end{aligned}$$

Unter H_0 sollten die Rangsummen R_i nicht zu stark von ihren Erwartungswerten abweichen.

Wir bilden die Summe der quadrierten Abweichungen der Rangsummen von ihren Erwartungswerten:

$$T = \sum_{i=1}^c \left(R_i - \frac{n_i(N+1)}{2} \right)^2$$

Kruskal und Wallis modifizieren diese Teststatistik nun so, daß sie asymptotisch chiquadratverteilt ist und erhalten:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^c \frac{1}{n_i} \left(R_i - \frac{n_i(N+1)}{2} \right)^2$$

Unter H_0 ist der Erwartungswert von H gleich $c - 1$. Dies ist gerade der Erwartungswert einer mit $c - 1$ Freiheitsgraden chiquadratverteilten Zufallsvariablen.

Dies sieht man folgendermaßen:

$$\begin{aligned} E(H) &= \frac{12}{N(N+1)} \sum_{i=1}^c \frac{1}{n_i} E \left[\left(R_i - \frac{n_i(N+1)}{2} \right)^2 \right] \\ &= \frac{12}{N(N+1)} \sum_{i=1}^c \frac{1}{n_i} \text{Var}(R_i) \\ &= \frac{12}{N(N+1)} \sum_{i=1}^c \frac{1}{n_i} \frac{n_i(N+1)(N-n_i)}{12} \\ &= \frac{1}{N} \sum_{i=1}^c (N - n_i) \\ &= \frac{cN - N}{N} \\ &= c - 1 \end{aligned}$$

Somit besitzt H denselben Erwartungswert wie eine mit $c - 1$ Freiheitsgraden chiquadratverteilte Zufallsvariable.

Man kann auch zeigen, daß H asymptotisch chiquadratverteilt mit $c - 1$ Freiheitsgraden ist.

Wir lehnen die Nullhypothese ab, wenn H zu groß ist.

Man kann H nun so modifizieren, daß die Berechnung leichter ist.

Es gilt

$$H = \frac{12}{N(N+1)} \sum_{i=1}^c \frac{R_i^2}{n_i} - 3(N+1)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} H &= \frac{12}{N(N+1)} \sum_{i=1}^c \frac{1}{n_i} \left(R_i - \frac{n_i(N+1)}{2} \right)^2 \\ &= \frac{12}{N(N+1)} \sum_{i=1}^c \frac{1}{n_i} \left(R_i^2 - 2 R_i \frac{n_i(N+1)}{2} + \frac{n_i^2(N+1)^2}{4} \right) \\ &= \frac{12}{N(N+1)} \left(\sum_{i=1}^c \frac{R_i^2}{n_i} - 2 \frac{N+1}{2} \sum_{i=1}^c R_i + \frac{(N+1)^2}{4} \sum_{i=1}^c n_i \right) \\ &= \frac{12}{N(N+1)} \left(\sum_{i=1}^c \frac{R_i^2}{n_i} - 2 \frac{N+1}{2} \frac{N(N+1)}{2} + \frac{N(N+1)^2}{4} \right) \\ &= \frac{12}{N(N+1)} \left(\sum_{i=1}^c \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right) \\ &= \frac{12}{N(N+1)} \sum_{i=1}^c \frac{R_i^2}{n_i} - 3(N+1) \end{aligned}$$

Für das Datenbeispiel gilt

$$R_1 = 9$$

$$R_2 = 24$$

$$R_3 = 12$$

Also gilt

$$\begin{aligned} H &= \frac{12}{9 \cdot 10} \left(\frac{81}{3} + \frac{576}{3} + 1443 \right) - 3 \cdot 10 \\ &= 5.6 \end{aligned}$$

Aus der Tabelle der exakten Verteilung im Buch von Büning und Trenkler entnehmen wir

$$P_{H_0}(H \geq 5.6) = 0.05$$

Also lehnen wir die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ ab.

In R gehen wir folgendermaßen vor:

```
x <- c(47,53,49,55,54,58,52,50,51)
```

```
g <- rep(1:3,rep(3,3))
```

```
kruskal.test(x,g)
```

```
    Kruskal-Wallis rank sum test
```

```
data:  x and g
```

```
Kruskal-Wallis chi-square = 5.6, df = 2,
```

```
p-value = 0.0608
```

```
alternative hypothesis: two.sided
```

R rechnet beim Kuskal-Wallis-Test nicht die exakte Verteilung unter H_0 aus. Die Überschreitungswahrscheinlichkeit wird mit Hilfe der Chiquadratverteilung bestimmt. Dies sieht man, indem Eins minus Verteilungsfunktion der Chiquadratverteilung mit 2 Freiheitsgraden an der Stelle 5.6 bestimmt:

```
1-pchisq(5.6,2)
```

```
[1] 0.06081006
```

Für kleine Stichprobenumfänge kann man die exakte Verteilung auch selber herleiten.

Wir betrachten den Fall $n_1 = 2$, $n_2 = 2$ und $n_3 = 1$.

1. Stichprobe	2. Stichprobe	3. Stichprobe	H
1 2	3	4	2.7
1 2	4	3	2.7
1 3	2	4	1.8
1 3	4	2	1.8
1 4	2	3	0.3
1 4	3	2	0.3
2 3	1	4	2.7
2 3	4	1	2.7
2 4	1	3	1.8
2 4	3	1	1.8
3 4	1	2	2.7
3 4	2	1	2.7

Somit erhalten wir durch einfaches Auszählen die exakte Verteilung:

h	$P(H = h)$
0.3	1/6
1.8	1/3
2.7	1/2

Der Datensatz für Maschine 1 wies Bindungen auf. Wie beim Wilcoxon-Test muß die Teststatistik modifiziert werden, um Bindungen zu berücksichtigen:

$$H^* = \frac{H}{1 - \frac{1}{N^3 - N} \sum_{j=1}^r (b_j^3 - b_j)}$$

Dabei ist r die Anzahl der Gruppen mit Bindungen und b_j die Anzahl der Beobachtungen in der j -ten Bindungsgruppe.

Schauen wir uns die für alle Beobachtungen der drei Maschinen an:

$$\begin{aligned} x_{11} &= 47 & x_{12} &= 53 & x_{13} &= 49 & x_{14} &= 50 & x_{15} &= 46 \\ x_{21} &= 55 & x_{22} &= 54 & x_{23} &= 58 & x_{24} &= 61 & x_{25} &= 52 \\ x_{31} &= 52 & x_{32} &= 50 & x_{33} &= 51 & x_{34} &= 53 & x_{35} &= 49 \end{aligned}$$

Die Ränge sind:

$$\begin{aligned} r_{11} &= 2 & r_{12} &= 10.5 & r_{13} &= 3.5 & r_{14} &= 5.5 & r_{15} &= 1 \\ r_{21} &= 13 & r_{22} &= 12 & r_{23} &= 14 & r_{24} &= 15 & r_{25} &= 8.5 \\ r_{31} &= 8.5 & r_{32} &= 5.5 & r_{33} &= 7 & r_{34} &= 10.5 & r_{35} &= 3.5 \end{aligned}$$

Für die Rangsummen gilt

$$\begin{aligned} R_1 &= 2 + 10.5 + 3.5 + 5.5 + 1 \\ &= 22.5 \\ R_2 &= 13 + 12 + 14 + 15 + 8.5 \\ &= 62.5 \\ R_3 &= 8.5 + 5.5 + 7 + 10.5 + 3.5 \\ &= 35 \end{aligned}$$

Also erhalten wir für H :

$$\begin{aligned} H &= \frac{12}{15 \cdot 16} \left(\frac{506.25}{5} + \frac{3906.25}{5} + 12255 \right) - 3 \cdot 16 \\ &= 8.375 \end{aligned}$$

In der folgenden Tabelle sind die Bindungsgruppen mit den jeweiligen Besetzungshäufigkeiten:

x_j	b_j
46	1
47	1
49	2
50	2
51	1
52	2
53	2
54	1
55	1
58	1
61	1

Wir erhalten also als Nenner in H^* :

$$1 - \frac{1}{N^3 - N} \sum_{j=1}^r (b_j^3 - b_j) = 1 - \frac{1}{15^3 - 15} \cdot 24$$

$$= 0.9928571$$

Also gilt

$$H^* = \frac{8.375}{0.9928571}$$

$$= 8.435252$$

In R erhalten wir das gleiche Ergebnis:

```
attach(maschine1.df)

kruskal.test(anzahl,maschine)

Kruskal-Wallis rank sum test

data:  anzahl andmaschine
Kruskal-Wallis chi-square = 8.4353,
df=2,
p-value = 0.0147

alternative hypothesis: two.sided
```

Ein wichtiger Spezialfall für einen Kruskal-Wallis-Test mit Bindungen liegt bei einer Kontingenztafel vor, bei der eine Kategorie geordnet ist.

Schauen wir uns dazu ein Beispiel an.

100 Personen wurden gebeten, die wirtschaftliche Lage einzuschätzen, wobei als Antworten 'schlecht', 'normal' und 'gut' vorgegeben waren.

Außerdem wurden sie gefragt, welche Partei sie wählen.

Es ergab sich folgende Tabelle:

	schlecht	normal	gut
SPD	20	17	5
CDU	10	15	15
FDP	1	3	4
Grüne	5	3	2

Wir wollen überprüfen, ob die Einschätzung der wirtschaftlichen Lage bei den Wählern der unterschiedlichen Parteien gleich ist.

Es liegt nahe, den Chi-Quadrat-Unabhängigkeitstest anzuwenden.

Wir geben die Daten als Matrix `h` ein

```
h <- matrix(c(20,10,1,5,17,15,3,3,5,15,4,2),4,3)
h
      [,1] [,2] [,3]
[1,]   20   17   5
[2,]   10   15  15
[3,]    1    3   4
[4,]    5    3   2
```

und rufen die Funktion `chisq.test` mit dieser Matrix auf:

```
chisq.test(h)

Pearson's chi-square test without Yates' continuity correction

data:  h
X-squared = 12.0852,
df = 6,
p-value = 0.0601
```

```
Warning messages: Expected counts < 5. Chi-squared approximation may not be appropriate in: chisq.test(h)
```

Der Test lehnt zum Niveau $\alpha = 0.05$ die Nullhypothese nicht ab.

Wir können aber auch den Kruskal-Wallis-Test durchführen.

Die Einschätzung der wirtschaftlichen Lage ist eine geordnete Variable, die wir mit 1=schlecht, 2=normal und 3=gut kodieren können.

Dann haben 20 der SPD Wähler den Wert 1, bei den CDU Wählern sind es 10, u.s.w.

Wir erzeugen also zunächst einen Datenvektor mit den Werten:

```
as.vector(t(h))
[1] 20 17  5 10 15 15  1  3  4  5  3  2

d <- rep(rep(1:3,4),as.vector(t(h)))

d
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
    2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3
    3 3 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2
[60] 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3
    3 3 3 1 2 2 2 3 3 3 3 1 1 1 1 1 2 2 2 3 3
```

Dann rufen wir die Funktion `kruskal.test` auf, wobei wir berücksichtigen, daß man auch anstatt der Gruppenzugehörigkeit angeben kann, wieviele in jeder Gruppe sind.

```
kruskal.test(d,rep(1:4,apply(h,1,sum)))
```

```
Kruskal-Wallis rank sum test
```

```
data: d and rep(1:4, apply(h, 1, sum))
Kruskal-Wallis chi-square= 11.2088,
df = 3,
p-value = 0.0106
```

```
alternative hypothesis:
two.sided
```

Der Kruskal-Wallis-Test lehnt im Gegensatz zum Chiquadrattest die Hypothese ab.

3.1.3 Der Jonckheere-Test

Der Kruskal-Wallis-Test überprüft, ob die Verteilungen der Grundgesamtheiten sich bezüglich der Lage unterscheiden. Sehr oft hat man aber vor der Analyse eine Vorstellung über die Beziehung der Lageparameter unter der Alternativhypothese.

Das folgende Beispiel macht dies deutlich.

Beispiel 3.1.1 *Es soll untersucht werden, ob Koffein einen Einfluß auf die Konzentrationsfähigkeit hat. Dazu werden 30 Studenten zufällig auf drei gleichgroße Gruppen aufgeteilt. Die 10 Studenten der ersten Gruppe erhalten ein Getränk ohne Koffein, die der zweiten Gruppe eins mit 100 mg Koffein und die der dritten Gruppe eins mit 200 mg Koffein. Danach müssen die Studenten 1 Minute mit dem Zeigefinger auf den Tisch klopfen. Es wird die Anzahl der Schläge bestimmt.*

Es ergaben sich folgende Werte:

0 mg:	242	245	244	248	247	248	242	244	246	242
100 mg:	248	246	245	247	248	250	247	246	243	244
200 mg:	246	248	250	252	248	250	246	248	245	250

In diesem Datensatz liegen eine Reihe von Bindungen vor. Da wir zunächst den Fall ohne Bindungen betrachten wollen, wählen wir aus jeder Gruppe drei Beobachtungen aus:

0 mg:	242	245	244
100 mg:	248	247	243
200 mg:	246	250	252

Es soll hier nicht nur überprüft werden, ob die Konzentrationsfähigkeit von der Dosis Koffein abhängt, sondern sogar, ob sie mit wachsender Dosis zunimmt.

Wir wollen also testen

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

gegen

$$H_1 : \mu_1 < \mu_2 < \mu_3$$

Wir betrachten nun einen geeigneten Test.

Wir wollen überprüfen, ob mit wachsendem Koffeingehalt die Anzahl der Schläge pro Minute zunimmt.

Hierzu vergleichen wir zuerst die erste Stichprobe mit der zweiten, dann die erste Stichprobe mit der dritten und dann die zweite Stichprobe mit der dritten.

Beginnen wir mit dem Vergleich der ersten mit der zweiten Stichprobe. Kommt die zweite Stichprobe aus einer Grundgesamtheit, deren Verteilung bezüglich der Lage größer als die Verteilung, aus der die erste Stichprobe gezogen wurde, so wird man erwarten, daß die meisten Beobachtungen der ersten Stichprobe kleiner sind als die Beobachtungen der zweiten Stichprobe. Wir vergleichen also jeden Wert der zweiten Stichprobe mit jedem Wert der ersten Stichprobe und zählen, wie oft ein Wert der zweiten Stichprobe größer ist als ein Wert der ersten Stichprobe.

Dies ist die Teststatistik des Vergleichs der ersten mit der zweiten Stichprobe. Schauen wir uns für das Beispiel an.

Die Daten sind

0 mg: 242 245 244
100 mg: 248 247 243

Die erste Beobachtung 248 der zweiten Stichprobe ist größer als alle drei Beobachtungen der ersten Stichprobe. Also trägt diese Beobachtung den Wert 3 zur Teststatistik bei.

Die zweite Beobachtung 247 der zweiten Stichprobe ist ebenfalls größer als alle drei Beobachtungen der ersten Stichprobe. Also trägt diese Beobachtung auch den Wert 3 zur Teststatistik bei.

Die dritte Beobachtung 244 der zweiten Stichprobe ist größer als eine Beobachtung der ersten Stichprobe. Also trägt diese Beobachtung den Wert 1 zur Teststatistik bei.

Die Teststatistik nimmt also den Wert

$$3 + 3 + 1 = 7$$

an.

Der Vergleich der ersten mit der dritten Stichprobe liefert den Wert 9 und der Vergleich der zweiten mit der dritten Stichprobe liefert den Wert 7.

Um alle drei zu vergleichen, addiert man die Werte der drei Vergleiche und erhält den Wert 23.

Spricht dieser Wert für die Gegenhypothese?

Schauen wir uns die Vorgehensweise allgemein an.

Um die i -te mit der j -ten Stichprobe zu vergleichen, bilden wir die Teststatistik

$$U_{ij} = \sum_{t=1}^{n_j} \sum_{s=1}^{n_i} D_{st}$$

mit

$$D_{st} = \begin{cases} 1 & \text{für } X_{is} < X_{jt} \\ 0 & \text{sonst} \end{cases}$$

Es wird also für jede Beobachtung der j -ten Stichprobe bestimmt, wieviele der Beobachtungen in der i -ten Stichprobe kleiner sind.

Vergleicht man c Stichproben unter der Alternative, daß die Lageparameter mit wachsender Stichprobennummer immer größer werden, so bildet man die Teststatistik

$$V = \sum_{i < j} U_{ij}$$

Der zugehörige Test wird Jonckheere Test genannt.

Unter der Nullhypothese identischer Verteilungen gilt:

$$E(V) = \frac{1}{4} \left(N^2 - \sum_{i=1}^c n_i^2 \right)$$

und

$$Var(V) = \frac{1}{72} \left(N^2 (2N + 3) - \sum_{i=1}^c (n_i^2 (2n_i + 3)) \right)$$

Dabei ist n_i der Stichprobenumfang der i -ten Stichprobe.

Außerdem ist

$$N = n_1 + \dots + n_c.$$

Für große Stichprobenumfänge ist

$$J = \frac{V - E(V)}{\sqrt{Var(V)}}$$

approximativ standardnormalverteilt.

Wir lehnen H_0 ab, wenn gilt

$$J \geq z_{1-\alpha}$$

wobei $z_{1-\alpha}$ das $1 - \alpha$ -Quantil der Standardnormalverteilung ist.

Für das Datenbeispiel gilt

$$n_1 = n_2 = n_3 = 3.$$

Also gilt $N = 9$.

Wir erhalten

$$\begin{aligned} E(V) &= \frac{1}{4} (9^2 - (3^2 + 3^2 + 3^2)) \\ &= 13.5 \end{aligned}$$

und

$$\begin{aligned} Var(V) &= \frac{1}{72} \left(N^2 (2N + 3) - \sum_{i=1}^c (n_i^2 (2n_i + 3)) \right) \\ &= \frac{1}{72} (9^2 (2 \cdot 9 + 3) - 3 \cdot 3^2 (2 \cdot 3 + 3)) \\ &= 20.25 \end{aligned}$$

Es gilt also

$$\begin{aligned} J &= \frac{23 - 13.5}{\sqrt{20.25}} \\ &= 2.11 \end{aligned}$$

Wegen $z_{0.95} = 1.645$ lehnen wir H_0 ab.

Wie können wir dies in R implementieren?

Wir erstellen zunächst einen Vektor mit den Werten der neun Personen:

```
ka <- c(242,245,244,248,247,243,246,250,252)
ka
[1] 242 245 244 248 247 243 246 250 252
```

und einen Vektor mit der Gruppenzugehörigkeit:

```
gr <- rep(1:3,c(3,3,3))
gr
[1] 1 1 1 2 2 2 3 3 3
```

Mit Hilfe der Funktion `outer` können wir das äußere Produkt von zwei Vektoren bezüglich eines Operators bilden.

Wir wählen zunächst die beiden Stichproben aus

```
x1 <- ka[gr==1]
x1
[1] 242 245 244

x2 <- ka[gr==2]
x2
[1] 248 247 243
```

und bilden dann das äußere Produkt von `x1` und `x2` bezüglich des Operators `<`

```
outer(x1,x2,FUN="<")
      [,1] [,2] [,3]
[1,]    T    T    T
[2,]    T    T    F
[3,]    T    T    F
```

Das Ergebnis ist eine Matrix `m`.

Das Element `m[i,j]` ist T, wenn `x1[i]` kleiner als `x2[j]` ist, ansonsten ist es F.

Dem T entspricht die 1 und dem F die 0.

Summieren wir nun alle Werte der Matrix auf, so erhalten wir den Wert von U_{12} .

```
sum(outer(x1,x2,FUN("<")))
[1] 7
```

U_{12} ist die Teststatistik des Mann-Whitney-Tests.

Wir schreiben eine Funktion in R :

```
mw.stat <- function(x1, x2) {
    sum(outer(x1,x2, FUN = "<"))
}
```

```
mw.stat(x1,x2)
[1] 7
```

Nehmen wir nun an, daß 3 Stichproben vorliegen. Es soll überprüft werden, ob die erste Stichprobe aus der Grundgesamtheit mit dem kleinsten Lageparameter, die zweite Stichprobe aus der Grundgesamtheit mit dem zweitkleinsten Lageparameter und die dritte Stichprobe aus der Grundgesamtheit mit dem größten Lageparameter kommt.

Es liegt nahe, U_{12} , U_{13} und U_{23} zu bestimmen und als Teststatistik

$$V = U_{12} + U_{13} + U_{23}$$

zu verwenden.

```
x3 <- ka[gr==3]
x3
[1] 246 250 252
```

```
v <- mw.stat(x1,x2)+mw.stat(x1,x3) + mw.stat(x2,x3)
```

```
v
[1] 23
```

Die folgende Funktion führt einen Jonckheere Test für den Fall durch, daß keine Bindungen vorliegen.

```

joheere.test <- function(y, g) {
# Jonckheere Test ohne Bindungen
# Quelle: Sprent :Applied Nonparametric Statistical Methods,
      2.Auflage, S.141-143
# Bestimmung der Teststatistik
  s <- 0
  anz <- max(g)
  for(i in 1:(anz - 1))
    for(j in (i + 1):anz)
      s <- s + mw.stat(y[g == i], y[g == j])
  n <- length(g)
  ni <- apply(outer(1:anz, g, FUN = "=="), 1, sum)
  # Erwartungswert
  m <- (n^2 - sum(ni^2))/4
  # Varianz
  v <- ((n^2) * (2 * n + 3) -
        sum((ni^2) * (2 * ni + 3)))/72
  # \("Überschreitungswahrscheinlichkeit
  1 - pnorm((s - m)/sqrt(v))
}

```

Sie erhält als Argumente den Datenvektor und den Vektor mit der Gruppenzugehörigkeit.

```

joheere.test(ka,gr)
[1] 0.01738138

```

Der ursprüngliche Datensatz soll in dem dataframe `kaffee.df` stehen:

```
kaffee.df
```

```
  gruppe kaffee
1      1    242
2      1    245
3      1    244
4      1    248
5      1    247
6      1    248
7      1    242
8      1    244
9      1    246
10     1    242
11     2    248
12     2    246
13     2    245
14     2    247
15     2    248
16     2    250
17     2    247
18     2    246
19     2    243
20     2    244
21     3    246
22     3    248
23     3    250
24     3    252
25     3    248
26     3    250
27     3    246
28     3    248
29     3    245
30     3    250
```

Er erhält eine Vielzahl von Bindungen.

Diese können wir in der Teststatistik dadurch berücksichtigen, daß die Funktion D_{st} den Wert 0.5 annimmt, wenn beim Vergleich zwei Werte identisch sind:

$$D_{st} = \begin{cases} 1 & \text{für } X_{is} < X_{jt} \\ 0.5 & \text{für } X_{is} = X_{jt} \\ 0 & \text{sonst} \end{cases}$$

Wir erhalten also folgende Funktion

```
mw.ties.stat <- function(x1, x2) {
  sum(outer(x1,x2, FUN = "<")+0.5*outer(x1,x2, FUN = "=="))
}
```

Die Teststatistik für den Vergleich der i -ten mit der j -ten Gruppe ist dann wieder

$$U_{ij} = \sum_{t=1}^{n_j} \sum_{s=1}^{n_i} D_{st}$$

und als Teststatistik des Jonckheere Tests erhalten wir:

$$V = \sum_{i < j} U_{ij}$$

Der Erwartungswert ändert sich nicht, hingegen die Varianz:

$$\text{Var}(V) = \frac{U_1}{72} + \frac{U_2}{36 N (N - 1) (N - 2)} + \frac{U_3}{8 N (N - 1)}$$

mit

$$U_1 = N(N-1)(N-2) - \sum_{i=1}^c n_i(n_i-1)(2n_i+5) -$$

$$- \sum_{j=1}^r b_j(b_j-1)(2b_j+5)$$

$$U_2 = \left(\sum_{i=1}^c n_i(n_i-1)(n_i-2) \right) \left(\sum_{j=1}^r b_j(b_j-1)(b_j-2) \right)$$

$$U_3 = \left(\sum_{i=1}^c n_i(n_i-1) \right) \left(\sum_{j=1}^r b_j(b_j-1) \right)$$

Dabei gibt b_j die Anzahl der Beobachtungen in der j -ten Bindungsgruppe an, $j = 1, \dots, r$.

Die nachfolgende Funktion führt einen Jonckheere Test mit Bindungen durch:

```

joheere.ties.test <- function(y, g) {
# Jonckheere Test mit Bindungen
# Quelle: Sprent :Applied Nonparametric Statistical Methods,
      2.Auflage, S.141-143
# Bestimmung der Teststatistik
  su <- 0
  anz <- max(g)
  for(i in 1:(anz - 1))
    for(j in (i + 1):anz)
      su <- su + mw.ties.stat(y[g== i], y[g == j])
  n <- length(g)
  # Bestimmung der Stichprobenumf\ange
  ni <- apply(outer(1:anz, g, FUN = "=="), 1, sum)
  # Bestimmung der Besetzungszahlen der Bindungsgruppen
  b <- haeuf(y)[,2]
  # Erwartungswert
  mi <- (n^2 - sum(ni^2))/4
  # Varianz
  u1 <- n * (n - 1) * (2 * n + 5) -
    sum(ni * (ni - 1) * (2 * ni + 5)) -
    sum(b * (b - 1) * (2 * b + 5))
  u2 <- sum(ni * (ni - 1) * (ni - 2)) *
    sum(b * (b - 1) * (b - 2))
  u3 <- sum(ni * (ni - 1)) * sum(b * (b - 1))
  v <- u1/72 + u2/(36 * n * (n - 1) * (n - 2)) +
    u3/(8 * n * (n - 1))
  return(1 - pnorm((su - mi)/sqrt(v)))
}

```

Wir wenden sie auf den Datensatz `kaffee.df` an.

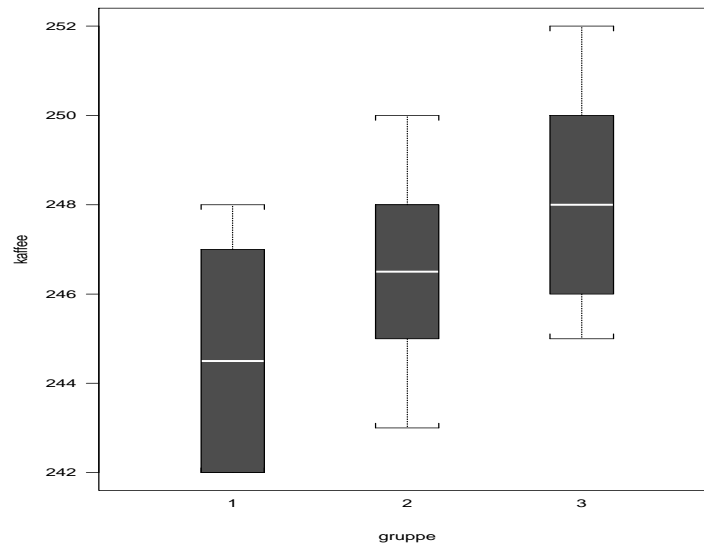
```

attach(kaffee.df)
joheere.test.ties(kaffee, as.numeric(gruppe))
[1]0.001521493

```

Der Test kommt zu dem Ergebnis, daß die Konzentrationsfähigkeit mit wachsendem Koffeingehalt zunimmt.

Das zeigt sich in den nachfolgenden Boxplots:



Bei einigen Anwendungen treten sehr viele Bindungen auf.

901 Personen wurden nach ihrem Einkommen befragt. Außerdem sollten sie angeben, wie zufrieden sie sind.

Es ergab sich folgende Tabelle:

	sehr unzufrieden	unzufrieden	zufrieden	sehr zufrieden
bis 6000 DM	20	24	80	82
6000 bis 25000 DM	35	66	185	238
über 25000 DM	7	18	54	92

Wir können auf diesen Datensatz den Chiquadrattest anwenden.

```
m <- matrix(c(20,35,7,24,66,18,80,185,54,82,238,92),3,4)
```

```
m
      [,1] [,2] [,3] [,4]
[1,]  20  24  80  82
[2,]  35  66 185 238
[3,]   7  18  54  92
```

```
chisq.test(m)
  Pearson's chi-square test
  without Yates' continuity correction
```

```
data:  m
X-squared = 10.2624,
df = 6,
p-value = 0.114
```

Die Hypothese, daß die Zufriedenheit vom Einkommen unabhängig ist, wird auf dem Niveau $\alpha = 0.05$ nicht abgelehnt.

Bei diesem Datensatz sind beide Variablen geordnet. Es stellt sich die Frage, ob die Zufriedenheit mit wachsendem Einkommen zunimmt.

Wir kodieren **sehr unzufrieden** mit 1, **unzufrieden** mit 2, **zufrieden** mit 3 und **sehr zufrieden** mit 4.

Die Daten für den Jonckheere Test erhalten wir dann folgendermaßen aus der Matrix m:

```
dat <- rep(rep(1:4,3),as.vector(t(m)))
gru <- rep(1:3,apply(m,1,sum))
```


Wir führen nun den Jonckheere Test durch:

```
jonckheere.test.ties(dat,gru)
[1] 0.002681505
```

Die Funktion `joheere.ties.test` ist bei vielen Bindungen nicht effizient. Die nachstehende Funktion `joheere.ties` operiert direkt auf der Kontingenztafel und ist viel schneller.

```
joheere.ties <- function(m) {
  co <- ncol(m)
  ro <- nrow(m)
  summe <- 0
  for(j in 1:(co - 1))
    for(i in 1:(ro - 1))

      summe <- summe +
        m[i, j] * (0.5 * sum(m[(i + 1):ro, j]) +
          sum(m[(i + 1):ro, (j + 1):co]))
  for(k in 1:(ro - 1))
    summe <- summe +
      m[k, co] * 0.5 * sum(m[(k + 1):ro, co])
  n <- sum(m)
  nip <- apply(m, 1, sum)
  npj <- apply(m, 2, sum)
  mi <- (n^2 - sum(nip^2))/4
  u1 <- n * (n - 1) * (2 * n + 5) -
    sum(nip * (nip - 1) * (2 * nip + 5)) -
    sum(npj * (npj - 1) * (2 * npj + 5))
  u2 <- sum(nip * (nip - 1) * (nip - 2)) *
    sum(npj * (npj - 1) * (npj - 2))
  u3 <- sum(nip * (nip - 1)) * sum(npj * (npj - 1))
  v <- u1/72 + u2/(36 * n * (n - 1) * (n - 2)) +
    u3/(8 * n * (n - 1))
  return(1 - pnorm((summe - mi)/sqrt(v)))
}
```

```
jonckheere.ties(m)
[1] 0.002681505
```

3.2 Verbundene Stichproben

3.2.1 Zweifaktorielle Varianzanalyse

Vier Reifensorten sollen hinsichtlich des Reifenabriebs nach 20000 km verglichen werden.

Wie soll man vorgehen?

Sicherlich ist es nicht sinnvoll, von jeder Reifensorte nur einen Reifen in den Vergleich einzubeziehen. Die Reifen einer Sorte werden sicherlich nicht den gleichen Abrieb nach 20000 km aufweisen. Also sollte man von jeder Reifensorte mehrere Reifen untersuchen. Wir entscheiden uns für vier Reifen von jeder Sorte. Der Reifenabrieb soll nun in einem Feldversuch bestimmt werden. Das bedeutet, daß die Reifen auf Autos montiert werden, und jedes Auto 20000 km zurücklegt.

Da wir von jeder der vier Reifensorten vier Reifen verwenden, benötigen wir vier Autos.

Bei der Verteilung der Reifen auf die Autos kann man nun unterschiedlich vorgehen.

1. Möglichkeit

Man ordnet alle Reifen einer Sorte einem Auto zu.

Dies ist aber nicht sinnvoll, da man nach Durchführung des Versuchs nicht unterscheiden kann, ob der unterschiedliche Abrieb an den Autos oder den Reifensorten liegt.

2. Möglichkeit

Man ordnet die 16 Reifen den 16 Positionen an den Autos zufällig zu. Hierbei kann es natürlich passieren, daß alle vier Reifen einer Sorte einem Auto zugeordnet werden. Dies ist aber unwahrscheinlich.

Durch die zufällige Aufteilung der Reifen zu den 16 Rädern versucht man den Einfluß anderer Faktoren wie Auto oder Fahrer auszuschalten. Eine solche Vorgehensweise nennt man einen randomisierten Versuchsplan.

Wir haben es dann mit einem unverbundenen c-Stichprobenproblem zu tun.

Nach Beendigung des Versuchs seien folgende Daten angefallen:

Reifensorte	1	2	3	4	\bar{x}_i
1	13	14	12	13	13
2	14	13	10	11	12
3	17	10	13	12	13
4	12	11	9	8	10

Wir speichern die Daten in dem `data.frame` `reifen1.df`:

```
gruppe <- rep(1:4,rep(4,4))
abrieb <- c(13,14,12,13,14,13,10,11,17,10,13,12,12,11,9,8)
reifen1.df <- data.frame(gruppe,abrieb)
```

und führen eine Varianzanalyse durch

```
summary(aov(abrieb~gruppe,reifen1.df))
```

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
gruppe	3	24	8	2	0.1678316
Residuals	12	48	4		

Wir erhalten also folgende ANOVA-Tabelle:

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Reifensorten	24	3	8	2
Rest	48	12	4	
Gesamt	72	15		

Die Überschreitungswahrscheinlichkeit beträgt 0.1678.

Somit ist der Unterschied des Abriebs der Reifen zum Niveau $\alpha = 0.05$ nicht signifikant.

3. Möglichkeit

Bei der Zuweisung der Reifen zu den einzelnen Untersuchungseinheiten kann man aber auch berücksichtigen, daß jedes Auto eine Einheit bildet und die Abnutzung der Reifen bei den einzelnen Autos unterschiedlich sein wird.

Wie kann man diesen Autoeffekt in den Griff bekommen?

Man kann von jeder Reifensorte genau einen Reifen zufällig auswählen und diese 4 Reifen dann zufällig den vier Rädern eines Autos zuweisen.

Hierdurch ist jede der Reifensorten den gleichen Bedingungen ausgesetzt, wenn man davon ausgeht, daß die Position des Reifens am Auto keinen Einfluß auf die Abnutzung hat.

Ein Auto bildet einen sogenannten Block. Die Untersuchungseinheiten in einem Block sind die vier Positionen. Diesen Untersuchungseinheiten werden nun zufällig die vier Behandlungen, d.h. Reifensorten, zugeordnet.

Man spricht deshalb von einem zufälligen Blockplan (**randomized block design**).

Wie kann man einen zufälligen Blockplan auswerten?

Schauen wir uns dazu zunächst noch einmal das Modell des randomisierten Versuchsplans an.

Es gilt

$$X_{ij} = \mu_i + \epsilon_{ij}$$

für $i = 1, \dots, c$ und $j = 1, \dots, n_i$.

Die ϵ_{ij} sind unabhängige, identisch mit Erwartungswert 0 und Varianz σ^2 normalverteilte Zufallsvariablen.

Wir spalten μ_i auf in $\mu_i = \mu + \alpha_i$.

Dabei nennen wir μ das Gesamtmittel und α_i den Effekt der i -ten Stufe des Faktor A.

In unserem Beispiel ist der Faktor A die Reifensorte und die einzelnen Stufen sind die Ausprägungen der Reifensorte, also die unterschiedlichen Reifensorten.

Das Modell lautet also

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

für $i = 1, \dots, c$ und $j = 1, \dots, n_i$.

Der Vorteil dieser Art der Parametrisierung ist, daß man das Modell leicht durch Hinzunahme weiterer Faktoren erweitern kann.

Bei einem zufälligen Blockplan hat man neben dem interessierenden Faktor A noch die Blöcke, also den Faktor B mit J Faktorstufen.

In unserem Beispiel sind dies die Autos.

Wir erhalten also folgendes Modell

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

für $i = 1, \dots, c$ und $j = 1, \dots, J$.

X_{ij} ist der Abrieb der i -ten Reifensorte innerhalb des j -ten Blocks.

Um die Übersicht zu behalten, numerieren wir folgendermaßen: $i = 1, \dots, I$ und $j = 1, \dots, J$.

Auf der Ebene der Erwartungswerte lautet das Modell

$$E(X_{ij}) = \mu + \alpha_i + \beta_j$$

Der Parameter α_i ist der Effekt der i -ten Stufe des Faktors A und der Parameter β_j ist der Effekt der j -ten Stufe des Faktors B.

Ist zum Beispiel $\alpha_1 > 0$, so ist der Abrieb der ersten Reifensorte im Mittel höher als der Durchschnitt aller Reifensorten.

Die Daten fallen also in folgender Form an:

Faktor A	Faktor B	1	...	J
1		x_{11}	...	x_{1J}
\vdots		\vdots	...	\vdots
I		x_{I1}	...	x_{IJ}

Nun ist es nicht nur wichtig, in welcher Zeile ein Element steht, sondern auch in welcher Spalte.

Für das Datenbeispiel unterstellen wir nun, daß ein zufälliger Blockplan vorliegt. Wir erhalten also

Reifensorte	Auto	1	2	3	4	$\bar{x}_{i.}$
1		13	14	12	13	13
2		14	13	10	11	12
3		17	10	13	12	13
4		12	11	9	8	10
$\bar{x}_{.j}$		14	12	11	11	12

Dabei ist

$$\bar{x}_{i.} = \frac{1}{J} \sum_{j=1}^J x_{ij}$$

und

$$\bar{x}_{.j} = \frac{1}{I} \sum_{i=1}^I x_{ij}$$

Wir können wie auch schon bei der einfaktoriellen Varianzanalyse die Residuenquadratsumme bestimmen, die sich aufgrund der unterschiedlichen Reifensorten ergibt.

Wir bezeichnen sie mit SS_A :

$$SS_A = J \sum_{i=1}^I (\bar{x}_{i.} - \bar{x})^2$$

Im Beispiel gilt

$$SS_A = 24$$

Da in den Spalten der Tabelle die Werte des Abriebs der Reifen bei einem Auto zu finden ist, können wir auch die Residuenquadratsumme bestimmen, die sich aufgrund der unterschiedlichen Autos ergibt.

Diese bezeichnen wir, da sie sich auf den Faktor B bezieht, mit SS_B :

$$SS_B = I \sum_{j=1}^J (\bar{x}_{.j} - \bar{x})^2$$

Im Beispiel gilt

$$\begin{aligned} SS_B &= 4 [(14 - 12)^2 + (12 - 12)^2 + (11 - 12)^2 + (11 - 12)^2] \\ &= 24 \end{aligned}$$

Wir können die Gesamtstreuung nun folgendermaßen zerlegen:

$$\begin{aligned}
SS_T &= \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x})^2 \\
&= \sum_{i=1}^I \sum_{j=1}^J (x_{i.} - \bar{x} + \bar{x}_{.j} - \bar{x} + x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 \\
&= \sum_{i=1}^I \sum_{j=1}^J (x_{i.} - \bar{x})^2 + \sum_{i=1}^I \sum_{j=1}^J (\bar{x}_{.j} - \bar{x})^2 \\
&\quad + \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 \\
&\quad + \sum_{i=1}^I \sum_{j=1}^J 2(x_{i.} - \bar{x})(\bar{x}_{.j} - \bar{x}) \\
&\quad + \sum_{i=1}^I \sum_{j=1}^J 2(x_{i.} - \bar{x})(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) \\
&\quad + \sum_{i=1}^I \sum_{j=1}^J 2(\bar{x}_{.j} - \bar{x})(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) \\
&= \sum_{i=1}^I J(x_{i.} - \bar{x})^2 + \sum_{j=1}^J I(\bar{x}_{.j} - \bar{x})^2 + \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2
\end{aligned}$$

da gilt

$$\sum_{i=1}^I \sum_{j=1}^J 2(x_{i.} - \bar{x})(\bar{x}_{.j} - \bar{x}) = 0$$

$$\sum_{i=1}^I \sum_{j=1}^J 2(x_{i.} - \bar{x})(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) = 0$$

$$\sum_{i=1}^I \sum_{j=1}^J 2(\bar{x}_{.j} - \bar{x})(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) = 0$$

Es gilt somit

$$SS_T = SS_A + SS_B + SS_R$$

mit

$$SS_A = \sum_{i=1}^I J(x_{i.} - \bar{x})^2$$

$$SS_B = \sum_{j=1}^J I(\bar{x}_{.j} - \bar{x})^2$$

$$SS_R = \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2$$

Wir können nun wieder wie im Fall der einfaktoriellen Varianzanalyse die mittlere Residuenquadratsummen bestimmen.

Für SS_A haben wir $I - 1$ Freiheitsgrade, für SS_B haben wir $J - 1$ Freiheitsgrade und für SS_T haben wir $IJ - 1$ Freiheitsgrade.

Also hat SS_R

$$\begin{aligned} IJ - 1 - (I - 1) - (J - 1) &= IJ - I - J + 1 \\ &= (I - 1)(J - 1) \end{aligned}$$

Freiheitsgrade.

Wir erhalten also folgende ANOVA-Tabelle:

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Stufen von A	SS_A	I-1	MSS_A	$\frac{MSS_A}{MSS_R}$
zwischen den Stufen von B	SS_B	J-1	MSS_B	$\frac{MSS_B}{MSS_R}$
Rest	SS_R	$(I - 1)(J - 1)$	MSS_R	
Gesamt	SS_T	N-1		

Durch die Betrachtung eines weiteren Faktors haben wir die Reststreuung verkleinert.

Ist die Verkleinerung nun so groß, daß der Abrieb der einzelnen Reifensorten unterschiedlich ist?

Wir testen also

$$H_0 : \alpha_1 = \dots = \alpha_I = 0$$

gegen

$$H_1 : \alpha_i \neq \alpha_j \quad \text{für mindestens ein Paar } (i, j) \text{ mit } i \neq j$$

Wenn wir überprüfen wollen, ob sich die Behandlungen unterscheiden, ist die Teststatistik

$$F_A = \frac{MSS_A}{MSS_R}$$

relevant. Diese ist unter der Nullhypothese F-verteilt mit $I - 1$ und $(I - 1)(J - 1)$ Freiheitsgraden.

Ist der Faktor B kein Block, sondern ein eigenständiger Faktor, so verwenden wir zur Überprüfung der Hypothesen

$$H_0 : \beta_1 = \dots = \beta_J = 0$$

gegen

$$H_1 : \beta_i \neq \beta_j \text{ für mindestens ein Paar } (i, j) \text{ mit } i \neq j$$

die Teststatistik

$$F_B = \frac{MSS_B}{MSS_R}$$

Diese ist unter der Nullhypothese F-verteilt mit $J - 1$ und $(I - 1)(J - 1)$ Freiheitsgraden.

Im Beispiel gilt

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Stufen von A	24	3	8	3
zwischen den Stufen von B	24	3	8	3
Rest	24	9	2.67	
Gesamt	72	15		

Wegen $F_{3,9;0.95} = 3.863$ lehnen wir die Nullhypothese, daß der Abrieb der Reifensorten sich nicht unterscheidet, nicht ab.

In R gehen wir folgendermaßen vor:

Wir vergeben zunächst Namen für die Stufen der beiden Faktoren.

```
fnames <- list(block=paste("Block ",1:4),
               sorte=paste("Sorte ",1:4))
fnames
$block: [1] "Block 1" "Block 2" "Block 3" "Block 4"
$sorte: [1] "Sorte 1" "Sorte 2" "Sorte 3" "Sorte 4"
```

Danach bauen wir das Design auf:

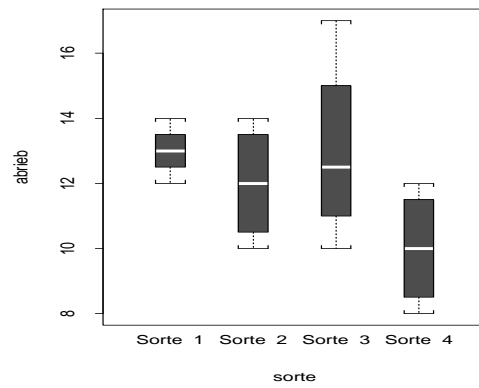
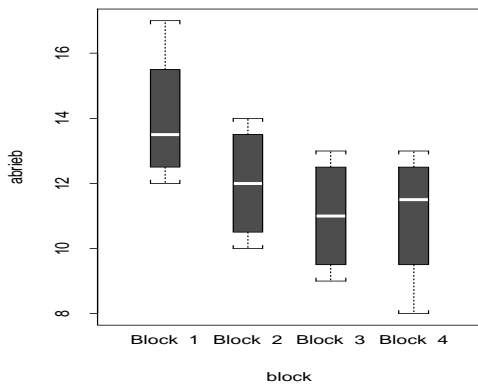
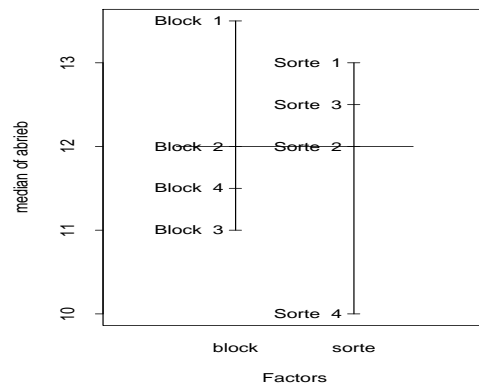
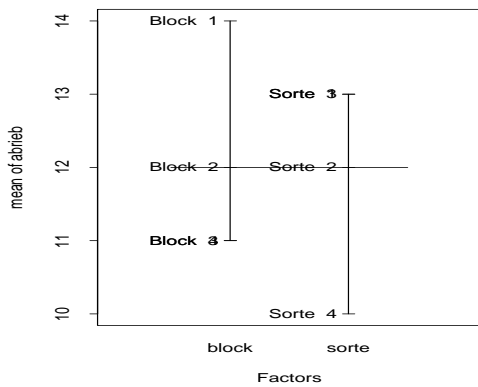
```
reifen.design <- fac.design(c(4,4),fnames)
reifen.design
  block  sorte
1 Block  1 Sorte  1
2 Block  2 Sorte  1
3 Block  3 Sorte  1
4 Block  4 Sorte  1
5 Block  1 Sorte  2
6 Block  2 Sorte  2
7 Block  3 Sorte  2
8 Block  4 Sorte  2
9 Block  1 Sorte  3
10 Block 2 Sorte  3
11 Block 3 Sorte  3
12 Block 4 Sorte  3
13 Block 1 Sorte  4
14 Block 2 Sorte  4
15 Block 3 Sorte  4
16 Block 4 Sorte  4
```

Das Design und die Daten schreiben wir in den Dataframe `reifen2.df`:

```
reifen2.df <- data.frame(reifen.design,abrieb=reifen1.df[[2]])
reifen2.df
  block  sorte abrieb
1 Block  1 Sorte  1    13
2 Block  2 Sorte  1    14
3 Block  3 Sorte  1    12
4 Block  4 Sorte  1    13
5 Block  1 Sorte  2    14
6 Block  2 Sorte  2    13
7 Block  3 Sorte  2    10
8 Block  4 Sorte  2    11
9 Block  1 Sorte  3    17
10 Block 2 Sorte  3    10
11 Block 3 Sorte  3    13
12 Block 4 Sorte  3    12
13 Block 1 Sorte  4    12
14 Block 2 Sorte  4    11
15 Block 3 Sorte  4     9
16 Block 4 Sorte  4     8
```

Wir können uns wie beim randomisierten Versuchsplan zuerst die Daten anschauen:

```
par(mfrow=c(2,2))
plot.design(reifen2.df)
plot.design(reifen2.df,fun=median)
plot.factor(reifen2.df)
```



Die Boxplots der einzelnen Blöcke deuten auf einen Lageunterschied der Blöcke hin, wobei die Streuung innerhalb der Blöcke nicht zu groß ist.
 Die Boxplots der Reifensorten deuten auf einen Lageunterschied hin, wobei die Streuungen sich relativ stark unterscheiden.

Nun können wir die Varianzanalyse durchführen.

Wir rufen die Funktion `aov` folgendermaßen auf:

```
aov.reifen2 <- aov(abrieb~sorte+block,reifen2.df)
```

Die Anova-Tabelle erhalten wir durch:

```
summary(aov.reifen2)
```

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
sorte	3	24	8.000000	3	0.08771291
block	3	24	8.000000	3	0.08771291
Residuals	9	24	2.666667		

Im Beispiel hat die Blockbildung zu einer großen Verkleinerung der Residuenquadratsumme geführt. Diese ist aber nicht so groß, daß die Gleichheit des Abriebs der einzelnen Reifensorten abgelehnt wird.

Das Modell

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

ist additiv in den Parametern α_i und β_j .

Es existieren also keine Wechselwirkungen.

Eine Wechselwirkung zwischen den Faktoren A und B liegt dann vor, wenn durch das gemeinsame Auftreten der i -ten Stufe des Faktors A und der j -ten Stufe des Faktors B ein zusätzlicher Effekt entsteht.

Schauen wir uns ein Beispiel an.

Ein Unternehmen betrachtet unterschiedliche Preisstrategien und Kommunikationsstrategien.

Der Faktor A sei die Preisstrategie mit den Stufen $i = 1$ gleich Niedrigpreispolitik und $i = 2$ gleich "Hochpreispolitik".

Der Faktor B sei die Kommunikationsstrategie mit den Stufen $j = 1$ gleich Postwurfsendungen und $j = 2$ gleich "Anzeigenwerbung".

Sei X_{ij} die abgesetzte Menge bei Preisstrategie i und Kommunikationsstrategie j .

Wir unterstellen das Modell

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

mit $i = 1, 2$ und $j = 1, 2$.

Es gilt

$$E(X_{ij}) = \mu + \alpha_i + \beta_j$$

In der folgenden Tabelle sind die hypothetische Erwartungswerte der X_{ij} zu finden.

	Postwurfsendungen	Anzeigenwerbung
Niedrigpreis	60	50
Hochpreis	40	30

Zwischen den beiden Faktoren besteht keine Interaktion.

Geht man nämlich bei Postwurfsendungen von einer Niedrigpreispolitik zu einer Hochpreispolitik über, so sinkt die erwartete abgesetzte Menge um 20.

Geht man bei Anzeigenwerbung von einer Niedrigpreispolitik zu einer Hochpreispolitik über, so sinkt die erwartete abgesetzte Menge ebenfalls um 20.

Die Wirkung der Preispolitik ist unabhängig davon, welche Kommunikationspolitik angewendet wird.

Mit

$$\mu_{ij} = E(X_{ij})$$

gilt also:

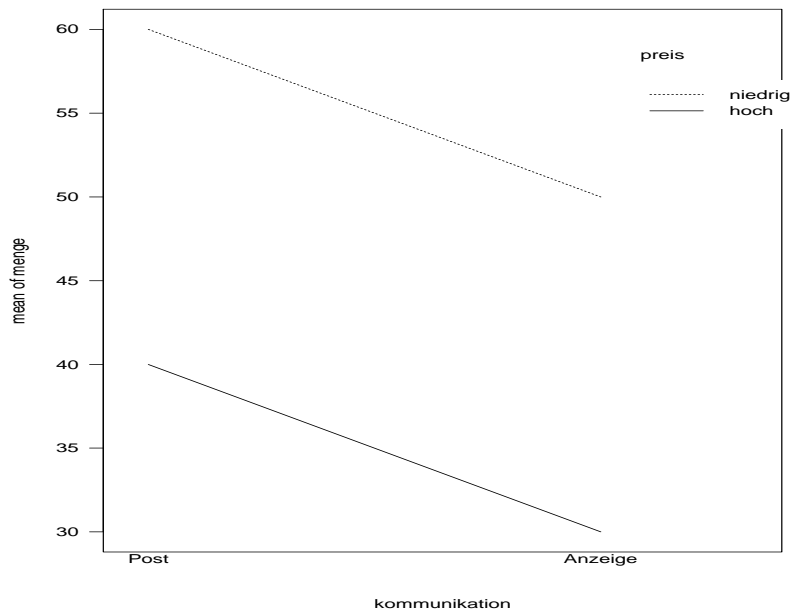
$$\begin{aligned} \mu_{21} - \mu_{11} &= \mu_{22} - \mu_{12} \\ &= \alpha_2 - \alpha_1 \end{aligned}$$

Geht man bei Postwurfsendungen von einer Niedrigpreisstrategie zu einer Hochpreisstrategie über, so vermindert sich die abgesetzte Menge um 20.

Geht man bei Anzeigenwerbung von einer Niedrigpreisstrategie zu einer Hochpreisstrategie über, so vermindert sich die abgesetzte Menge ebenfalls um 20.

Die Verringerung der abgesetzten Menge beim Übergang von einer Niedrigpreisstrategie zu einer Hochpreisstrategie ist unabhängig von der Kommunikationsstrategie.

In der folgenden Graphik sind noch einmal die Erwartungswerte graphisch dargestellt, wobei die Erwartungswerte einer Stufe des Faktors A durch eine Gerade verbunden sind.



Wenn keine Interaktion vorliegt, laufen die Geraden parallel.

Die folgende Tabelle gibt eine andere Situation wieder:

	Postwurfsendungen	Anzeigenwerbung
Niedrigpreis	60	50
Hochpreis	40	60

Hier liegt Interaktion vor.

Geht man nämlich bei Postwurfsendungen von einer Niedrigpreispolitik zu einer Hochpreispolitik über, so sinkt die erwartete abgesetzte Menge um 20.

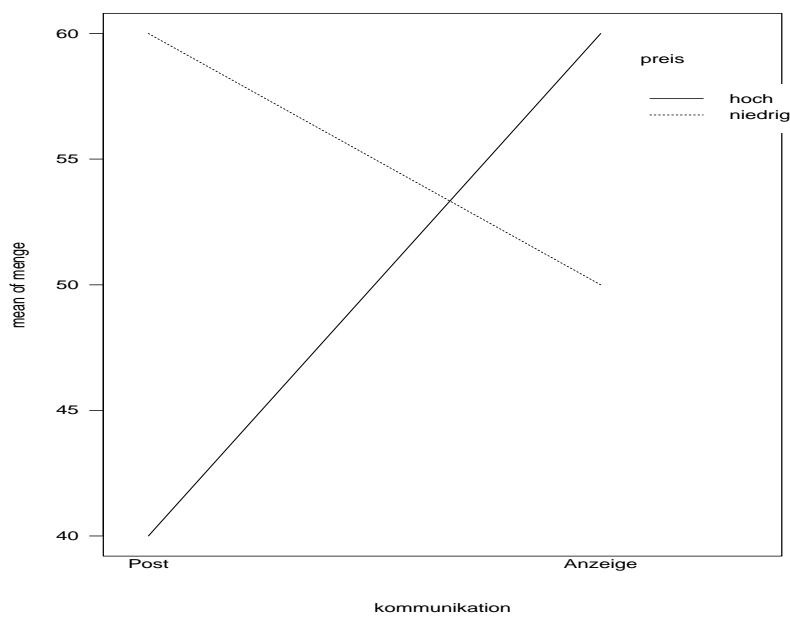
Geht man bei Anzeigenwerbung von einer Niedrigpreispolitik zu einer Hochpreispolitik über, so steigt die erwartete abgesetzte Menge hingegen um 10.

Die Wirkung der Preispolitik ist abhängig davon, welche Kommunikationspolitik angewendet wird. Es gibt noch einen gemeinsamen Effekt von Kommunikationspolitik und Preispolitik.

Im Modell berücksichtigen wir diesen als zusätzlichen Faktor

$$X_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

In der folgenden Graphik sind noch einmal die Erwartungswerte graphisch dargestellt, wobei die Erwartungswerte einer Stufe des Faktors A durch eine Gerade verbunden sind.



Wie man sieht, verlaufen die Geraden nicht parallel.

In R besteht die Möglichkeit, für die Daten einen Interaktionsplot zu erstellen.

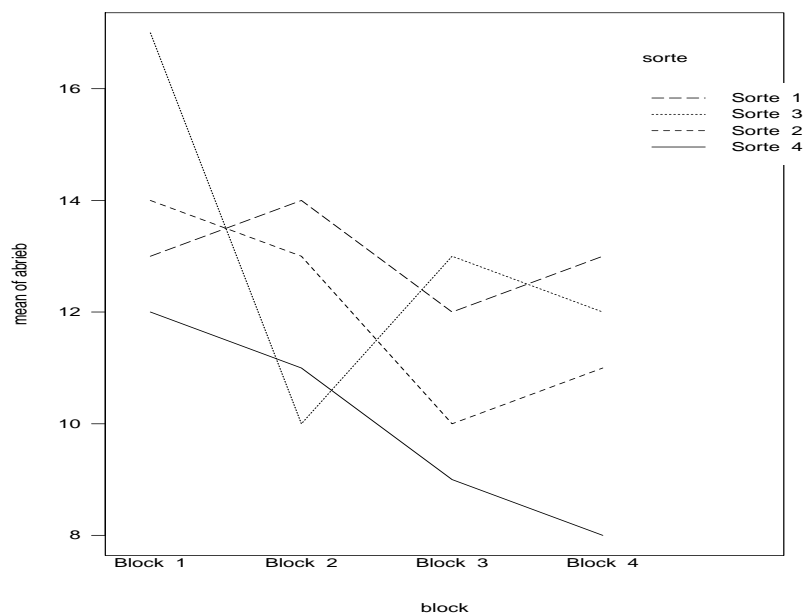
Dieser bezieht sich natürlich nicht auf die Erwartungswert, sondern auf die Beobachtungen.

Hierzu muß der `dataframe` aufgerufen werden, um auf die einzelnen Komponenten zugreifen zu können.

```
attach(reifen2.df)
```

Danach ruft man die Funktion `interaction.plot` auf:

```
interaction.plot(block,sorte,abrieb)
```



Bei den Reifensorten 1, 2 und 4 verlaufen die Geraden weitgehend parallel. Aus dem Rahmen fällt nur die Reifensorte 3.

Wenn aber bei den einzelnen Merkmalskombinationen keine Wiederholungen vorliegen, so kann man ohne zusätzliche Annahmen keinen Test auf Interaktion durchführen.

Von Tukey wurde 1949 folgendes Modell vorgeschlagen, um ohne Meßwiederholungen auf Interaktion zu testen:

$$X_{ij} = \mu + \alpha_i + \beta_j + \theta \alpha_i \beta_j + \epsilon_{ij}$$

Es wird also unterstellt, daß die Interaktion multiplikativ ist.

Die Quadratsumme bezüglich des Interaktionseffekts ist gegeben durch:

$$SS_{AB} = \frac{\left(\sum_{i=1}^I \sum_{j=1}^J \hat{\alpha}_i \hat{\beta}_j x_{ij} \right)^2}{\left(\sum_{i=1}^I \hat{\alpha}_i^2 \right) \left(\sum_{j=1}^J \hat{\beta}_j^2 \right)}$$

mit

$$\hat{\alpha}_i = \bar{x}_{i.} - \bar{x}$$

und

$$\hat{\beta}_j = \bar{x}_{.j} - \bar{x}$$

Zu testen ist

$$H_0 : (\alpha \beta)_{ij} = 0$$

gegen

$$H_1 : (\alpha \beta)_{ij} \neq 0$$

Die Teststatistik lautet

$$F_{AB} = \frac{((I-1)(J-1)-1)SS_{AB}}{SS_R - SS_{AB}}$$

Sie ist unter der Nullhypothese F-verteilt mit 1 und $(I-1)(J-1)-1$ Freiheitsgraden.

Im Beispiel gilt $F_{AB} = 0.148$.

Wegen $F_{1,9} = 5.318$ wird die Nullhypothese nicht abgelehnt.

Die Daten deuten also auf keine Interaktion hin.

Das folgende Beispiel stammt aus Schuchart-Fischer: Multivariate Analysemethoden.

Beispiel 3.2.1 *Es soll untersucht werden, welchen Einfluß unterschiedliche Preisstrategien und Kommunikationsstrategien auf die abgesetzten Mengeneinheiten einer Magarinesorte haben. Dazu werden an 2 Tagen die abgesetzten Mengen bestimmt.*

Es ergaben sich folgende Werte:

	<i>Postwurfsendungen</i>	<i>Anzeigenwerbung</i>
<i>Niedrigpreis</i>	<i>62</i>	<i>52</i>
	<i>66</i>	<i>56</i>
<i>Normalpreis</i>	<i>49</i>	<i>45</i>
	<i>53</i>	<i>49</i>
<i>Hochpreis</i>	<i>36</i>	<i>48</i>
	<i>40</i>	<i>44</i>

Der Faktor A sei die Preisstrategie und der Faktor B die Kommunikationsstrategie.

Sei X_{ijk} die abgesetzte Menge bei der i -ten Preisstrategie und der j -ten Kommunikationsstrategie am k -ten Tag.

Wir unterstellen folgendes Modell

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

Dabei ist

- α_i der Effekt der i -ten Stufe des Faktors A
- β_j der Effekt der j -ten Stufe des Faktors B
- $(\alpha\beta)_{ij}$ der Effekt des gemeinsamen Auftretens der i -ten Stufe des Faktors A und der j -ten Stufe des Faktors B

Als Schätzer für die Parameter erhalten wir

$$\begin{aligned}\hat{\alpha}_i &= \bar{x}_{i..} - \bar{x} \\ \hat{\beta}_j &= \bar{x}_{.j.} - \bar{x} \\ \widehat{(\alpha\beta)}_{ij} &= \bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}\end{aligned}$$

mit

$$\bar{x}_{i..} = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K x_{ijk}$$

$$\bar{x}_{.j.} = \frac{1}{IK} \sum_{i=1}^I \sum_{k=1}^K x_{ijk}$$

$$\bar{x}_{ij.} = \frac{1}{K} \sum_{i=1}^I \sum_{k=1}^K x_{ijk}$$

$$\bar{x} = \frac{1}{IJK} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}$$

Im Beispiel gilt

$$\begin{aligned} \bar{x} &= 50 & \bar{x}_{.1.} &= 51 & \bar{x}_{.2.} &= 49 \\ \bar{x}_{1..} &= 49 & \bar{x}_{2..} &= 49 & \bar{x}_{3..} &= 42 \\ \bar{x}_{11.} &= 64 & \bar{x}_{21.} &= 51 & \bar{x}_{31.} &= 38 \\ \bar{x}_{12.} &= 54 & \bar{x}_{22.} &= 47 & \bar{x}_{32.} &= 46 \end{aligned}$$

Die Quadratsummen lauten

$$SS_A = JK \sum_{i=1}^I (x_{i..} - \bar{x})^2$$

$$SS_B = IK \sum_{j=1}^J I (\bar{x}_{.j.} - \bar{x})^2$$

$$SS_{AB} = K \sum_{i=1}^I \sum_{j=1}^J I (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x})^2$$

$$SS_R = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \bar{x}_{ij.})^2$$

$$SS_T = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \bar{x})^2$$

Im Beispiel gilt

$$\begin{aligned} SS_A &= 2 \cdot 2 \cdot ((59 - 50)^2 + (49 - 50)^2 + (42 - 50)^2) \\ &= 584 \end{aligned}$$

$$\begin{aligned} SS_B &= 3 \cdot 2 \cdot ((51 - 50)^2 + (49 - 50)^2) \\ &= 12 \end{aligned}$$

$$\begin{aligned} SS_{AB} &= 2 \cdot ((64 - 59 - 51 + 50)^2 + (54 - 59 - 49 + 50)^2 \\ &+ (51 - 49 - 51 + 50)^2 + (47 - 49 - 49 + 50)^2 \\ &+ (38 - 42 - 51 + 50)^2 + (46 - 42 - 49 + 50)^2) \\ &= 168 \end{aligned}$$

$$\begin{aligned} SS_R &= (62 - 64)^2 + (66 - 64)^2 + (52 - 64)^2 + (56 - 64)^2 \\ &+ (49 - 51)^2 + (53 - 51)^2 + (45 - 47)^2 + (49 - 47)^2 \\ &+ (36 - 38)^2 + (40 - 38)^2 + (48 - 46)^2 + (44 - 46)^2 \\ &= 48 \end{aligned}$$

Wir können die ANOVA-Tabelle aufstellen:

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Stufen von A	SS_A	$I - 1$	MSS_A	$\frac{MSS_A}{MSS_R}$
zwischen den Stufen von B	SS_B	$J - 1$	MSS_B	$\frac{MSS_B}{MSS_R}$
Interaktion	SS_{AB}	$(I - 1)(J - 1)$	MSS_{AB}	$\frac{MSS_{AB}}{MSS_R}$
Rest	SS_R	$IJ(K - 1)$	MSS_R	
Gesamt	SS_T	$N - 1$		

Für das Beispiel erhalten wir folgende ANOVA-Tabelle:

Quelle der Variation	Quadratsummen	Freiheitsgrade	Mittlere Quadratsummen	F
zwischen den Stufen von A	584	2	292	36.5
zwischen den Stufen von B	12	1	12	1.5
Interaktion	168	2	84	10.5
Rest	48	6	8	
Gesamt	812	11		

Im Rahmen der zweifaktoriellen Varianzanalyse mit Meßwiederholungen wird zunächst auf Interaktion getestet:

$$H_0 : (\alpha\beta)_{ij} = 0 \quad \text{für alle } i, j$$

gegen

$$H_1 : (\alpha\beta)_{ij} \neq 0 \quad \text{für mindestens ein Paar } (i, j)$$

Zur Überprüfung dieser Hypothesen verwenden wir die Teststatistik

$$F_{AB} = \frac{MSS_{AB}}{MSS_R}$$

Diese ist unter der Nullhypothese F-verteilt mit $(I-1)(J-1)$ und $IJ(K-1)$ Freiheitsgraden.

Wir lehnen die Nullhypothese ab, wenn gilt

$$F_{AB} \geq F_{(I-1)(J-1), IJ(K-1); 1-\alpha}$$

Dabei ist $F_{(I-1)(J-1), IJ(K-1); 1-\alpha}$ das $(1-\alpha)$ -Quantil der F-Verteilung mit $(I-1)(J-1)$ und $IJ(K-1)$ Freiheitsgraden.

Im Beispiel gilt

$$F_{AB} = 10.5.$$

Wegen $F_{2,6;0.95} = 5.14$ lehnen wir die Nullhypothese ab.

Die Daten sprechen also für Interaktion.

In diesem Fall ist es nicht sinnvoll zu testen, ob Effekte von A oder B signifikant sind, da diese nicht vom Interaktionseffekt getrennt betrachtet werden können.

Wir werden die Tests aus Gründen der Vollständigkeit aber beschreiben und auch durchführen.

Wir betrachten nun noch die anderen Tests:

Überprüfung, ob ein Einfluß des Faktors A vorliegt:

$$\begin{aligned} H_0 &: \alpha_i = 0 \quad \text{für alle } i \\ H_1 &: \alpha_i \neq 0 \quad \text{für mindestens ein } i \end{aligned}$$

Zur Überprüfung dieser Hypothesen verwenden wir die Teststatistik

$$F_A = \frac{MSS_A}{MSS_R}$$

Diese ist unter der Nullhypothese F-verteilt mit $I - 1$ und $IJ(K - 1)$ Freiheitsgraden.

Wir lehnen die Nullhypothese ab, wenn gilt

$$F_A \geq F_{I-1, IJ(K-1); 1-\alpha}$$

Dabei ist $F_{I-1, IJ(K-1); 1-\alpha}$ das $(1 - \alpha)$ -Quantil der F-Verteilung mit $I - 1$ und $IJ(K - 1)$ Freiheitsgraden.

Überprüfung, ob ein Einfluß des Faktors B vorliegt:

$$\begin{aligned} H_0 &: \beta_j = 0 \quad \text{für alle } j \\ H_1 &: \beta_j \neq 0 \quad \text{für mindestens ein } j \end{aligned}$$

Zur Überprüfung dieser Hypothesen verwenden wir die Teststatistik

$$F_B = \frac{MSS_B}{MSS_R}$$

Diese ist unter der Nullhypothese F-verteilt mit $J - 1$ und $IJ(K - 1)$ Freiheitsgraden.

Wir lehnen die Nullhypothese ab, wenn gilt

$$F_B \geq F_{J-1, IJ(K-1); 1-\alpha}$$

Dabei ist $F_{J-1, IJ(K-1); 1-\alpha}$ das $(1 - \alpha)$ -Quantil der F-Verteilung mit $J - 1$ und $IJ(K - 1)$ Freiheitsgraden.

In R gehen wir folgendermaßen vor:

Wir erzeugen das Design

```
fnames <- list(kommunikation=c("Post","Anzeige"),
              preis=c("niedrig","normal","hoch"))
magarine.design <- fac.design(c(2,3),fnames,rep=2)
magarine.design
kommunikation  preis
1             Post niedrig
2             Anzeige niedrig
3             Post  normal
4             Anzeige normal
5             Post   hoch
6             Anzeige hoch
7             Post niedrig
8             Anzeige niedrig
9             Post  normal
10            Anzeige normal
11            Post   hoch
12            Anzeige hoch
```

Die Daten müssen nun passend zum Design eingegeben werden:

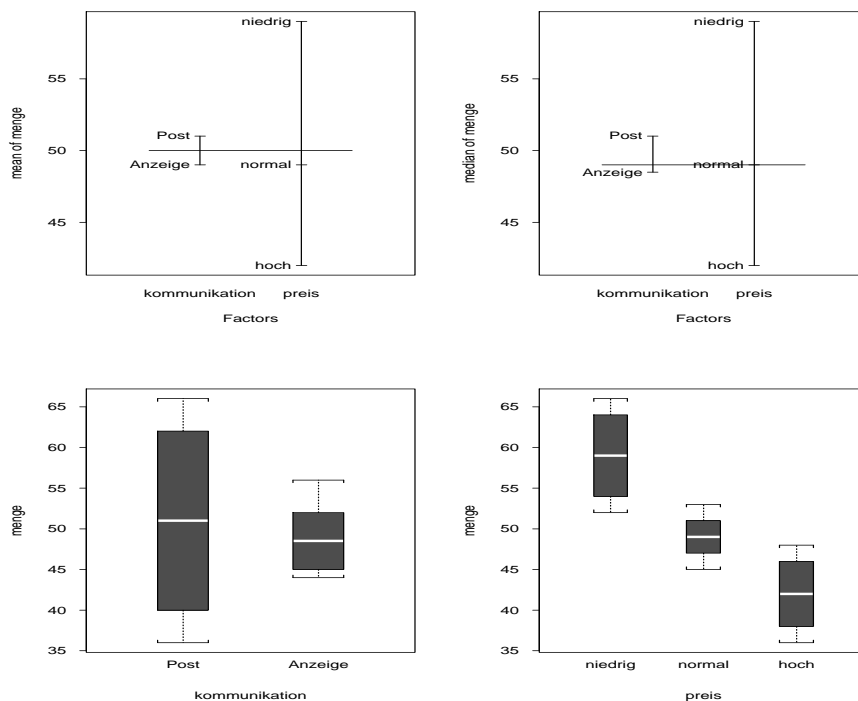
```
menge <- c(62,52,49,45,36,48,66,56,53,49,40,44)
```

Dann bringen wir Design und Daten zusammen:

```
magarine.df <- data.frame(magarine.design,menge)
magarine.df
kommunikation  preis  menge
1             Post niedrig   62
2             Anzeige niedrig  52
3             Post  normal   49
4             Anzeige normal  45
5             Post   hoch    36
6             Anzeige hoch   48
7             Post niedrig   66
8             Anzeige niedrig  56
9             Post  normal   53
10            Anzeige normal  49
11            Post   hoch    40
12            Anzeige hoch   44
```


Wir können uns wieder die Daten anschauen.
Zunächst die Mittelwerte und die Boxplots:

```
par(mfrow=c(2,2))
plot.design(magarine.df)
plot.design(magarine.df,fun=median)
plot.factor(magarine.df)
```

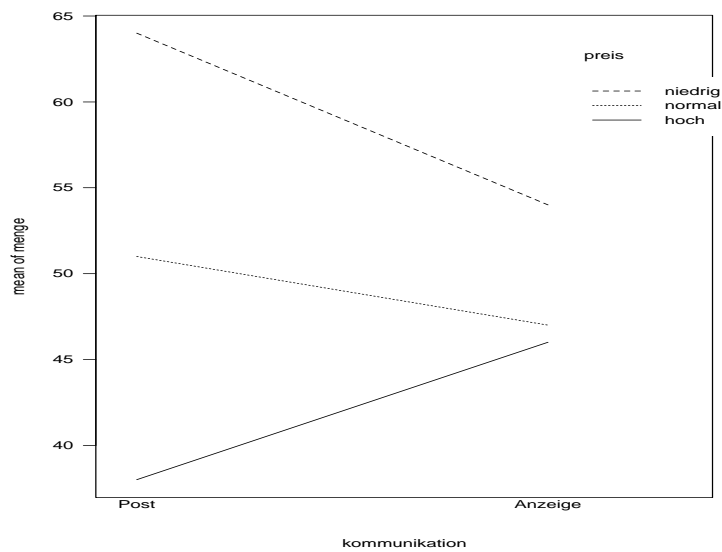


Wir sehen, daß die beiden Kommunikationsstrategien sich kaum hinsichtlich der abgesetzten Menge unterscheiden, während die Preisstrategien sich beträchtlich unterscheiden.

Die Streuungen der Preisstrategien sind ähnlich, während die Streuungen der beiden Kommunikationsstrategien sehr unterschiedlich sind.

Schauen wir uns noch den Interaktionsplot an.

```
attach(magarine.df)
interaction.plot(kommunikation,preis,menge)
```



Die Graphik deutet auf Interaktion hin. Bei Postwurfsendungen ist die abgesetzte Menge bei den drei Preisstrategien sehr unterschiedlich, wobei bei niedrigen Preisen am meisten abgesetzt wird. Bei Anzeigenwerbung ist der Unterschied zwischen den Preisstrategien fast vernachlässigbar.

Schließlich führen wir die Varianzanalyse durch:

```
aov.magarine <- aov(menge~preis*kommunikation,magarine.df)
summary(aov.magarine)
```

	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
preis	2	584	292	36.5	0.0004381
kommunikation	1	12	12	1.5	0.2665697
preis:kommunikation	2	168	84	10.5	0.0109739
Residuals	6	48	8		

3.2.2 Der Friedman-Test

Kann keine Normalverteilung unterstellt werden, so sollte man einen nichtparametrischen Test anwenden. Der bekannteste ist der Friedman-Test. Dieser wird angewendet, wenn keine Meßwiederholungen vorliegen.

Schauen wir uns diesen für den Vergleich der Reifensorten an.

Wir schauen uns noch einmal kurz die Tabelle an:

Reifensorte	Auto	1	2	3	4
1		13	14	12	13
2		14	13	10	11
3		17	10	13	12
4		12	11	9	8

Es soll überprüft werden, ob sich die Reifensorten hinsichtlich der Abnutzung unterscheiden. Ist dies der Fall, und wäre zum Beispiel die erste Reifensorte am besten, so würde man für diese bei jedem Auto den kleinsten Wert unter den 4 Reifensorten erwarten. Dies würde bedeuten, daß unter den 4 Beobachtungen bei einem Auto die erste Reifensorte den Rang 1 erhielte.

Dies ist die Idee des Friedman-Tests.

Um zu überprüfen, ob die Behandlungen sich unterscheiden, werden die Ränge innerhalb der Blöcke vergeben. Dabei ist R_{ij} der Rang der i -ten Behandlung innerhalb des j -ten Blocks.

Beim ersten Auto erhält die erste Reifensorte den Rang 2, die zweite Reifensorte den Rang 3, die dritte Reifensorte den Rang 4 und die vierte Reifensorte den Rang 1.

Wir erhalten somit folgende Tabelle der Ränge:

Reifensorte	Auto	1	2	3	4	R_i
1		2	4	3	4	13
2		3	3	2	2	10
3		4	1	4	3	12
4		1	2	1	1	5

In der letzten Spalte haben wir noch die Rangsumme R_i der i -ten Reifensorte bestimmt:

$$R_i = \sum_{j=1}^J R_{ij}$$

Unterscheiden sich die Reifensorten nicht, so sollten diese Rangsummen ähnliche Werte annehmen.

Um zu überprüfen, ob sich diese Rangsummen nicht unterscheiden, vergleichen wir jede einzelne mit ihrem Erwartungswert unter H_0 .

Wenn keine Bindungen vorliegen, nehmen innerhalb eines Blockes die Ränge die natürlichen Zahlen $1, 2, \dots, I$ an.

Unter H_0 gilt:

$$P(R_{ij} = k) = \frac{1}{I}$$

Also gilt

$$E(R_{ij}) = \frac{I+1}{2}$$

und wir erhalten

$$\begin{aligned} E(R_i) &= E\left(\sum_{j=1}^J R_{ij}\right) \\ &= \sum_{j=1}^J E(R_{ij}) \\ &= \sum_{j=1}^J \frac{I+1}{2} \\ &= \frac{J(I+1)}{2} \end{aligned}$$

Naheliegender ist es, folgende Teststatistik zu verwenden:

$$T = \sum_{i=1}^I \left(R_i - \frac{J(I+1)}{2} \right)^2$$

Friedman hat 1936 folgende Modifikation von T vorgeschlagen:

$$F_I = \frac{12}{JI(I+1)} \sum_{i=1}^c \left(R_i - \frac{J(I+1)}{2} \right)^2$$

Man kann zeigen, daß unter gilt:

$$E(F_I) = I - 1$$

Somit hat F_I denselben Erwartungswert wie eine mit $I - 1$ Freiheitsgraden chiquadratverteilte Zufallsvariable.

Man kann auch zeigen, daß F_I approximativ chiquadratverteilt ist mit $I - 1$ Freiheitsgraden.

Wir lehnen die Nullhypothese identischer Behandlungen ab, wenn gilt

$$F_I \geq \chi_{I-1;1-\alpha}^2$$

Wie auch die Teststatistik des Kruskal-Wallis-Tests kann man F_I so umformen, daß die Berechnung einfacher ist.

Es gilt

$$F_I = \frac{12}{JI(I+1)} \sum_{i=1}^I R_i^2 - 3J(I+1)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} F_I &= \frac{12}{JI(I+1)} \sum_{i=1}^I \left(R_i - \frac{J(I+1)}{2} \right)^2 \\ &= \frac{12}{JI(I+1)} \sum_{i=1}^I \left(R_i^2 - 2R_i \frac{J(I+1)}{2} + \frac{J^2(I+1)^2}{4} \right) \\ &= \frac{12}{JI(I+1)} \left(\sum_{i=1}^I R_i^2 - 2 \frac{J(I+1)}{2} \sum_{i=1}^I R_i + \frac{IJ^2(I+1)^2}{4} \right) \\ &= \frac{12}{JI(I+1)} \left(\sum_{i=1}^I R_i^2 - J(I+1) \frac{JI(I+1)}{2} + \frac{IJ^2(I+1)^2}{4} \right) \\ &= \frac{12}{JI(I+1)} \left(\sum_{i=1}^I R_i^2 - \frac{IJ^2(I+1)^2}{4} \right) \\ &= \frac{12}{JI(I+1)} \sum_{i=1}^I R_i^2 - 3J(I+1) \end{aligned}$$

Für das Beispiel gilt:

$$\begin{aligned} F_I &= \frac{12}{4 \cdot 4 \cdot 5} (13^2 + 10^2 + 12^2 + 5^2) - 3 \cdot 4 \cdot 5 \\ &= 5.7 \end{aligned}$$

Wegen $\chi_{3;0.95}^2 = 7.82$ lehnen wir die Nullhypothese nicht ab.

Wir können dies auch direkt in R bestimmen.

```
m <- matrix(reifen2.df[[3]],4,4,b=T)
```

```
m
      [,1] [,2] [,3] [,4]
[1,]  13  14  12  13
[2,]  14  13  10  11
[3,]  17  10  13  12
[4,]  12  11   9   8
```

In den Spalten stehen die Blöcke.

Die Ränge innerhalb jedes Blockes erhalten wir, indem wir die Funktion `rank` auf die zweite Dimension, d.h. die Spalten, von `m` anwenden.

```
r <- apply(m,2,FUN="rank")
```

```
r
      [,1] [,2] [,3] [,4]
[1,]    2    4    3    4
[2,]    3    3    2    2
[3,]    4    1    4    3
[4,]    1    2    1    1
```

Die Rangsummen für die einzelnen Gruppen erhalten wir dann durch:

```
rs <- apply(r,1,sum)
```

```
rs
[1] 13 10 12  5
```

Für die Teststatistik erhalten wir dann

```
(12*sum(rs^2))/(4*4*5)-3*4*5
[1] 5.7
```

In R können wir dies auch durch die Funktion `friedman.test` erreichen, nachdem wir den Dataframe `reifen2.df` attached haben.

Hierdurch können wir auf die Variablen mit ihrem Namen zugreifen.

```
attach(reifen2.df)

friedman.test(abrieb,sorte,block)
```

```
Friedman rank sum test
```

```
data: abrieb and sorte and block
Friedman chi-square = 5.7,
df =3,
p-value = 0.1272
```

```
alternative hypothesis: two.sided
```

Oft fallen die Daten schon bei der Erhebung in Form von Rängen an:

Fünf Studenten sollten die 10 Paarvergleiche zwischen 5 Politikern nach Ähnlichkeit der Größe nach ordnen.

Es ergaben sich folgende Ergebnisse.

Fischer	-	Kinkel	8	7	10	9	1
Fischer	-	Kohl	7	8	1	7	7
Fischer	-	Lafontaine	2	3	2	4	2
Fischer	-	Waigel	3	10	8	10	3
Kinkel	-	Kohl	5	2	4	2	8
Kinkel	-	Lafontaine	9	5	9	5	4
Kinkel	-	Waigel	4	4	5	3	5
Kohl	-	Lafontaine	10	6	6	6	10
Kohl	-	Waigel	1	1	3	1	9
Lafontaine	-	Waigel	6	9	7	8	6

Wir geben die Daten ein

```
fnames <- list(student=paste("student",1:5),
               paar=paste("paar",1:10))

fnames $student:
[1] "student 1" "student 2" "student 3"
    "student4" "student 5"

$paar:
[1] "paar 1" "paar 2" "paar 3" "paar 4" "paar 5"
    "paar 6" "paar 7" "paar 8" "paar 9" "paar 10"

pol1.design <- fac.design(c(5,10),fnames)

rang <- c(8,7,10,9,1,7,8,1,7,7,2,3,2,4,2,3,10,8,
          10,3,5,2,4,2,8,9,5,9,5,4,4,4,5,3,5,10,
          6,6,6,10,1,1,3,1,9,6,9,7,8,6)

pol1.df <- data.frame(pol1.design,rang)

friedman.test(rang,paar,student)

      Friedman rank sum test

data:  rang and paar and student

Friedman chi-square = 16.5491,
df= 9,
p-value = 0.0563

alternative hypothesis: two.side
```


Oft weisen die Daten Bindungen auf:

Wir betrachten noch einmal die Fragestellung des vorherigen Beispiels. Nur sollen jetzt die 5 Studenten die 10 Paare auf einer Skala von 1 bis 7 vergleichen. Dabei ist 1 "sehr ähnlich" und 7 "total verschieden".

Es ergaben sich folgende Ergebnisse:

Fischer	- Kinkel	6	4	6	6	6
Fischer	- Kohl	5	6	2	5	5
Fischer	- Lafontaine	2	3	2	3	1
Fischer	- Waigel	3	7	7	7	7
Kinkel	- Kohl	4	2	4	1	2
Kinkel	- Lafontaine	6	4	6	5	6
Kinkel	- Waigel	4	4	1	3	3
Kohl	- Lafontaine	7	5	4	4	5
Kohl	- Waigel	2	1	3	2	1
Lafontaine	- Waigel	4	6	7	4	7

In diesem Fall muß die Teststatistik von Friedman modifiziert werden zu

$$F_I = \frac{12 \sum_{i=1}^I \left(R_i - \frac{J(I+1)}{2} \right)^2}{JI(I+1) - \frac{1}{I-1} \sum_{j=1}^J \left[\left(\sum_{i=1}^{r_j} b_{ij}^3 \right) - I \right]}$$

Dabei ist r_j die Anzahl der Bindungsgruppen im j -ten Block und b_{ij} die Anzahl der gebundenen Werte in der i -ten Bindungsgruppe des j -ten Blocks. In unserem Beispiel ist der erste Block der erste Student. In diesem Block gibt es $r_1 = 6$ Bindungsgruppen, nämlich 2, 3, 4, 5, 6, 7 und es gilt

$$\begin{aligned} b_{11} &= 2 & b_{21} &= 1 & b_{31} &= 3 \\ b_{41} &= 1 & b_{51} &= 2 & b_{61} &= 1 \end{aligned}$$

Wir geben die Daten in R ein:

```
wert <- c(6,4,6,6,6,5,6,2,5,5,2,3,2,3,1,3,7,7,
          7,7,4,2,4,1,2,6,4,6,5,6,4,4,1,3,3,7,
          5,4,4,5,2,1,3,2,1,4,6,7,4,7)
```

Wir schreiben die Daten mit dem Design, das sich nicht geändert hat, in die Datei pol2.df:

```
pol2.df <- data.frame(pol1.design,wert)
```

Die Funktion `friedman.test` liefert das Ergebnis:

```
friedman.test(wert, paar, student)
```

```
Friedman rank sum test
```

```
data: wert and paar and student
```

```
Friedman chi-square = 28.9054,
```

```
df= 9,
```

```
p-value = 0.0007
```

```
alternative hypothesis: two.sided
```