Statistik – Methoden zum Vergleich von zwei Gruppen

Unterstützende Signifikanztests

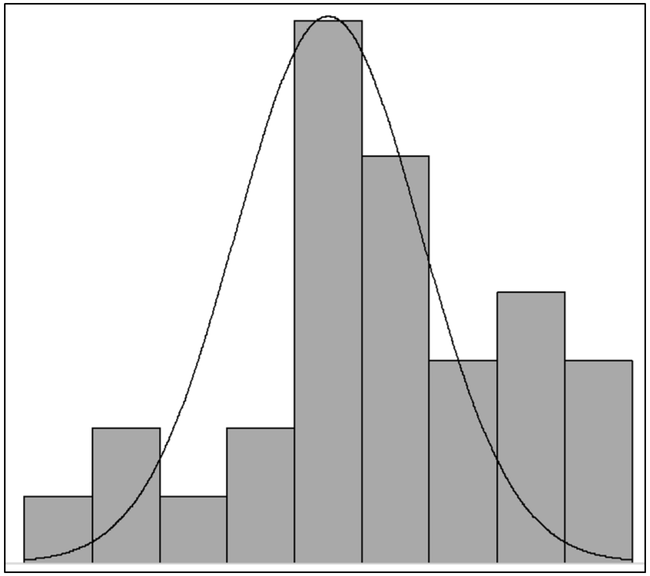
Verteilungstests

- Für verschiedene statistische Testverfahren (z.B. z-Test, t-Test) ist es notwendig, dass die zu untersuchenden Daten einer Normalverteilung folgen
- Gilt besonders für kleine Stichproben ($n < 30$)
- Wir benötigen Werkzeuge, die die Normalverteilungsannahme überprüfen können
- Des weiteren können einige der aufgeführten Testverfahren auch andere Verteilungsformen identifizieren, zum Teil können auch Stichproben gegeneinander geprüft werden (nicht Bestandteil des RCommander)

Verteilungstests

- Ein grafischer Vergleich der Daten mit einer Normalverteilung liefert Hinweise, wie nah man der geforderte Normalverteilung ist, macht aber keine statistisch signifikante Aussage

- Trotzdem sollte immer eine grafische Überprüfung erfolgen!!!



Verteilungstests

- Einige im RCommander genutzte Verteilungstests (nur Normalverteilung)

Test	
Shapiro-Wilk (SW)	Standardtest
χ^2-Test (Chi)	Datenklassifizierung erforderlich
Anderson-Darling (AD)	Six Sigma / englisch sprachiger Raum
Kolmogorov-Smirnov (KS)	Klassiker
Cramer von Mises (CM)	Starke Abhängigkeit von der Stichprobengröße
Shapiro-Francia (SF)	SW-Abwandlung für platokurtische Verteilungen
(Epps-Pulley)	Tolerant gegen Abweichungen
(Jarque-Bera)	Tolerant gegen Abweichungen

Verteilungstests

- Es gibt nicht den einen Test auf Normalverteilung, alle aufgeführten Tests haben Schwächen und Stärken, sie führen auch zum Teil zu divergierenden Aussagen

- Je nach Stichprobengröße sind einzelne Testverfahren besser oder schlechter geeignet

- Kleine Stichproben
  + Shapiro-Wilk, Anderson-Darling, Kolmogorow-Smirnov

- Große Stichproben
  + Shapiro-Wilk, χ^2-Test

Verteilungstests

Für alle aufgeführten Testverfahren gelten die gleichen Hypothesen

- Wichtig für die Auswertung des p-Wertes !!!

H_0 Die Daten folgen der zugrunde gelegten Verteilung (z.B. der Normalverteilung)

H_1 Die Daten folgen **nicht** der zugrunde gelegten Verteilung (z.B. der Normalverteilung)

- Das Signifikanzniveau wird im allgemeinen zu α = 5% gesetzt

Beispiel: Anderson-Darling-Test

- Der Anderson-Darling-Test kann für verschiedene Verteilungsformen eingesetzt werden
- Verwendung für die Überprüfung der Normalverteilungsannahme
- Hohe Trennschärfe

Beispiel: Anderson-Darling-Test

Das Testverfahren

- Zu untersuchenden Daten werden der Größe nach sortiert
- Anschließend erfolgt eine Transformation der sortierten Daten in eine Gleichverteilung
- Anschließend wird die Übereinstimmung mit einer Gleichverteilungsfunktion überprüft
- Das Ergebnis der Prüfung kann in einen p-Wert überführt werden

Beispiel: Anderson-Darling-Test

Die Testgröße

$AD^2 = -n - S$ (AD steht für Anderson Darling)

$$S = \sum_{k=1}^{n} \frac{2k-1}{n} \left[\ln F(Y_k) + ln(1 - F(Y_{n+1-k}))\right]$$

n: Stichprobengröße

F: Die Funktion der F-Verteilung

Beispiel: Anderson-Darling-Test

Der p-Wert

- $z = AD^2\left(1 + \frac{0{,}75}{n} + \frac{2{,}25}{n^2}\right)$

z	p-Wert
$z \leq 0{,}2$	$1 - exp(-13{,}436 + 101{,}14 * z - 223{,}73 * z^2)$
$0{,}2 < z \leq 0{,}34$	$1 - exp(-8{,}318 + 42{,}796 * z - 59{,}938 * z^2)$
$0{,}34 < z \leq 0{,}6$	$exp(0{,}9177 - 4{,}279 * z - 1{,}38 * z^2)$
$0{,}6 < z$	$exp(1{,}2937 - 5{,}709 * z + 0{,}0186 * z^2)$

Beispiel: Anderson-Darling-Test

Voraussetzungen für den Test

- Metrisch skalierte Daten
- Daten lassen sich in eine Gleichverteilung transformieren
- Stichprobengröße $n \geq 8$

Verteilungstests

Beispiel

Datei *Übung_Deskriptiv.xlsx*, Daten NV und NNV

Führen Sie die verschiedenen Testverfahren durch und vergleichen Sie die Werte.

Verteilungstests

Beispiel

Datei *Übung_Deskriptiv.xlsx*, Daten NV und NNV

Test	p-Wert (NV)	p-Wert (NNV)
SW	0,3076	3,333e-10
Chi	0,8508	8,707e-11
AD	0,1914	1,015e-14
KS	0,1367	1,556e-11
CM	0,1525	9,585e-10
SF	0,3076	0,000000005747

Ryan-Joiner-Test

- Korrelationsverfahren zur Überprüfung der Normalverteilung

- Der Ryan-Joiner-Test bestimmt einen Koeffizienten für die Korrelation der Daten und den normalverteilten Werten der Daten

- Liegt dieser Koeffizient nahe 1, so liegen die Daten am Wahrscheinlichkeitsnetz der Normalverteilung

Ryan-Joiner-Test

Teststatistik

$$\boldsymbol{R_p} = \frac{\sum(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}})\boldsymbol{b_i}}{\sqrt{\boldsymbol{s}^2(\boldsymbol{n} - \mathbf{1})\sum \boldsymbol{b_i^2}}}$$

mit

Y_i geordnete Beobachtungen
b_i normalverteilte Werte der geordneten Beobachtungen
s^2 Stichprobenvarianz

QQ - Diagramm

- Quantil-Quantil-Diagramm

- Grafische Überprüfung von Daten auf Normalverteilung (auch für andere Verteilungsformen)

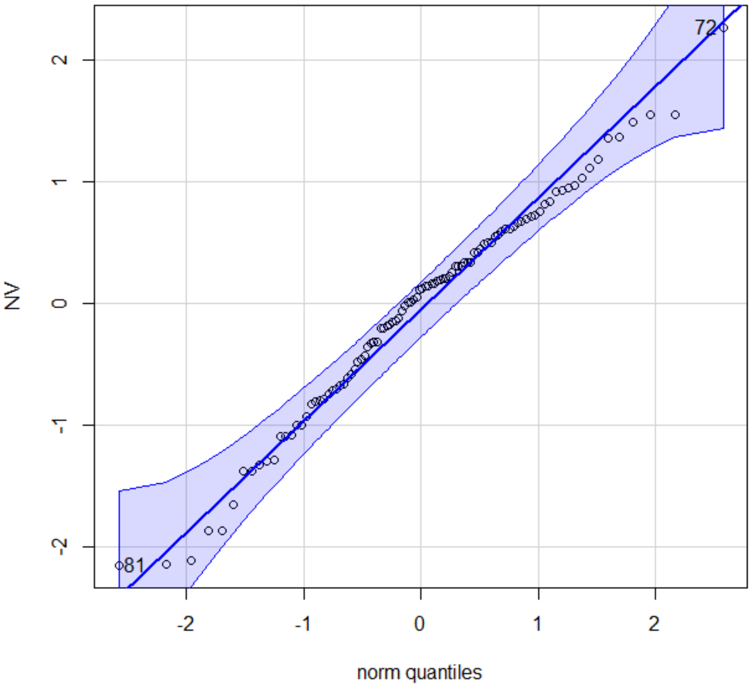
- Vergleichbar mit dem Wahrscheinlichkeitsnetz

- Es ersetz nicht den Verteilungstest, aber grafische Werkzeuge sollten immer parallel eingesetzt werden!

QQ - Diagramm

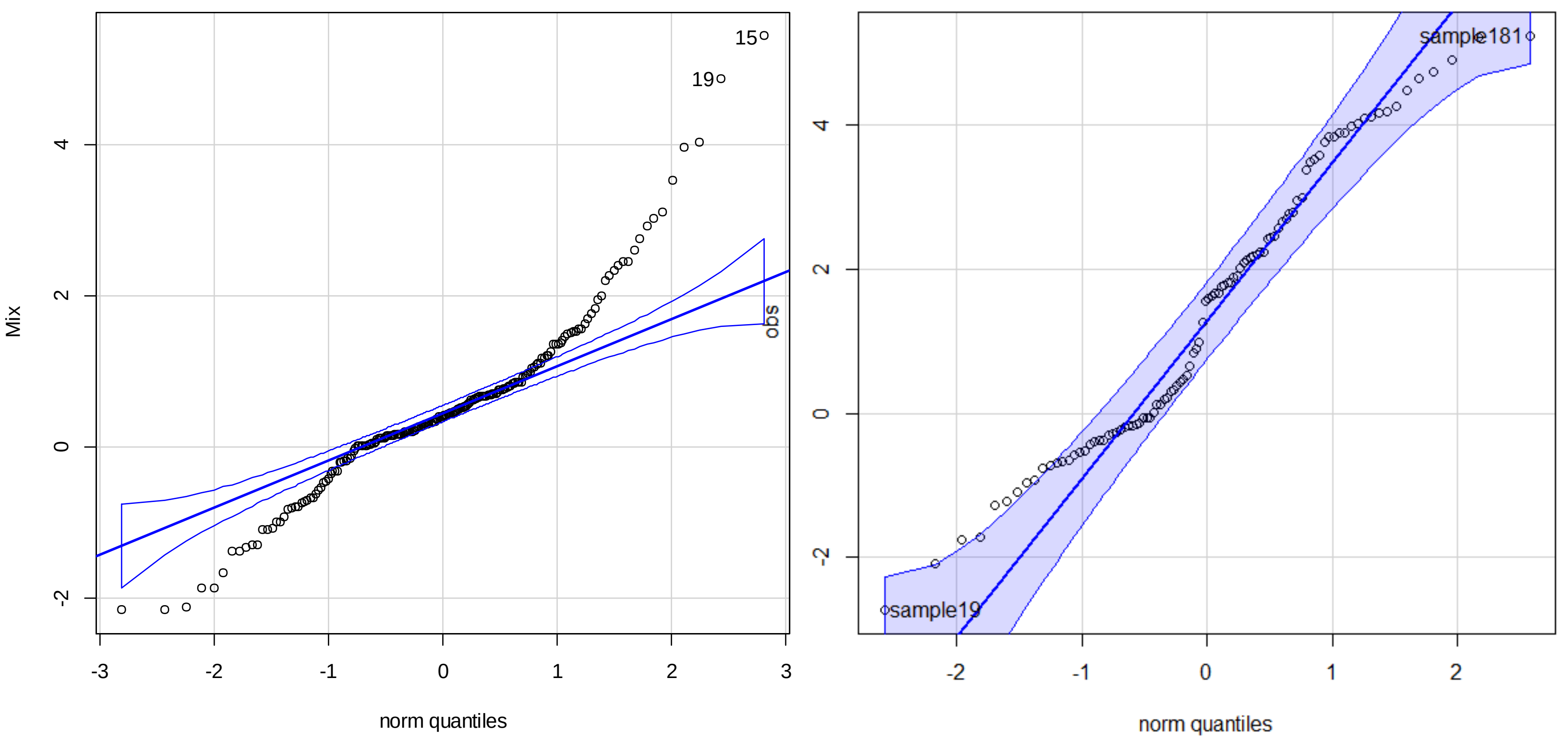
- Die Gerade stellt hier den Verlauf der Normalverteilung dar

- Sind die Daten normalverteilt, folgen die Datenpunkte möglichst genau der Geraden



QQ - Diagramm

- Überlagerung von Verteilungen



Test auf Varianzhomogenität

- Signifikanztests zur Prüfung der Gleichheit von Varianzen
- Die vorgestellten Tests können auf zwei (F-Test) oder mehr Gruppen (Bartlett/Levene) gleichzeitig angewandt werden
- Für alle Tests gelten die Voraussetzungen:
  - Unabhängige Beobachtungen
  - Metrisches Skalenniveau
  - Normalverteilung (nur Bartlett und F-Test), der Levene-Test ist auch für Nichtnormalverteilungen geeignet

Test auf Varianzhomogenität

- Verschiedene Tests und ihr Einsatzgebiet

Test	Normalverteilung	Stichproben	Ein/zweiseitig?
Levene-Test	Nein	≥ 2	zweiseitig
Bartlett-Test	Ja	≥ 2	zweiseitig
F-Test	Ja	$= 2$	ein-/zweiseitig

Test auf Varianzhomogenität

Hypothesen für Levene und Bartlett

Nullhypothese

$H_o: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$

Alternativhypothese

$H_1: \sigma_i^2 \neq \sigma_j^2$ für mindestens ein Gruppenpaar i,j mit i ≠ j

Der F-Test kann ein- und zweiseitig genutzt werden, entsprechend müssen hier die Alternativhypothesen eingestellt werden

Test auf Varianzhomogenität

Anmerkungen

- Bartlett-Test: Trennschärfer als Levene, setzt aber Normalverteilung voraus und sollte annähernd gleiche Stichprobengrößen aufweisen

- Levene-Test: Robust bei nicht-normalverteilten Daten und bei unterschiedlich großen Stichproben

- F-Test: Höchste Trennschärfe, aber sehr empfindlich bei Nicht-Normalität

Test auf Varianzhomogenität

Beispiel

Beispiel_Var.xlsx

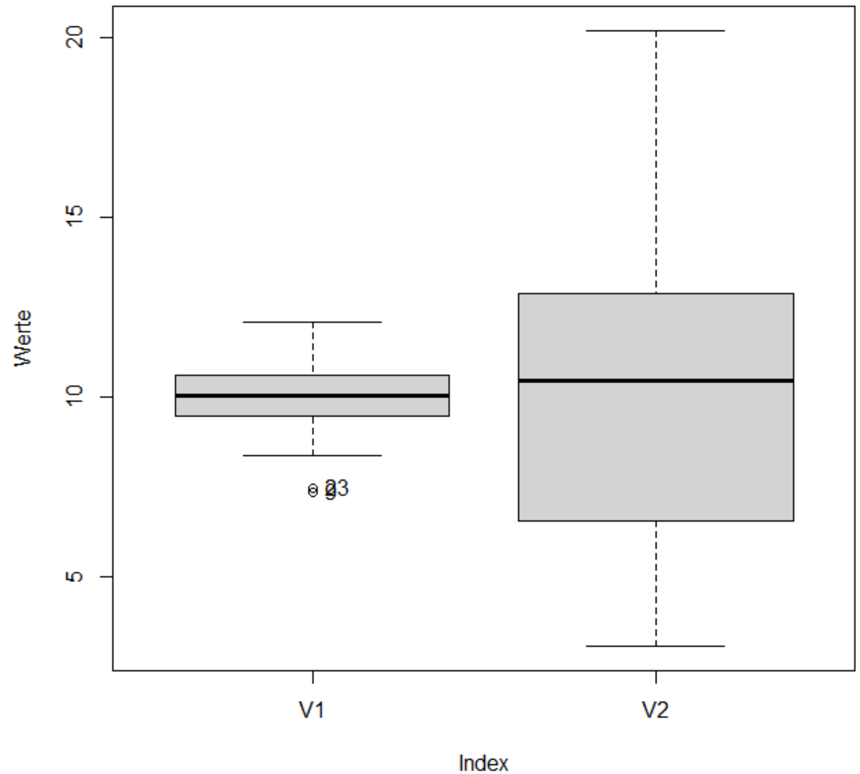
Untersuchung auf ungleiche Streuung / Varianz

$H_0 \qquad \sigma_1^2 = \sigma_2^2$
$H_1 \qquad \sigma_1^2 \neq \sigma_2^2$

Test auf Varianzhomogenität

Beispiel

- **Unterschiedliche Streuungen**



Test auf Varianzhomogenität

Beispiel

```
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value    Pr(>F)
group  1  67.007 1.017e-12 ***
      98
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ` 1
```

Der p-Wert wird in der R-Ausgabe als Pr bezeichnet

Test auf Varianzhomogenität

Beispiel

```
Bartlett test of homogeneity of variances

data:  Werte by Index
Bartlett's K-squared = 79.646, df = 1, p-value < 2.2e-16
```

Test auf Varianzhomogenität

Beispiel

```
        F test to compare two variances

data:  Werte by Index
F = 0.053739, num df = 49, denom df = 49, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.03049562 0.09469829
sample estimates:
ratio of variances
        0.05373903
```

Signifikanztest für Korrelationen

- Korrelation auf Signifikanz prüfen

- Gibt es in zwischen Gruppen eine Korrelation, kann geprüft werden, ob diese auch statistisch signifikant ist

- Ab welchem Betrag ist ein Korrelationskoeffizient statistisch signifikant?

Signifikanztest für Korrelationen

- Testverfahren: t-Tests, eine Stichprobe
- Weicht der Korrelationskoeffizient signifikant von 0 ab
- Prüfung auf lineare Unabhängigkeit

Nullhypothese

H_0 $r = 0$, es besteht keinen Zusammenhang zwischen den Gruppen

Alternativhypothesen

H_1 $r \neq 0$, es besteht ein Zusammenhang zwischen den Gruppen

Signifikanztest für Korrelationen

Teststatistik

$$t = \frac{r * \sqrt{n-2}}{\sqrt{1-r^2}}$$

- Diesen Wert kann man mit einem kritischen Wert t_{krit} vergleichen, der z.B. einer t-Tabelle entnommen werden kann

- Anzahl der Freiheitsgrade $df = n - 2$

Ausreißer

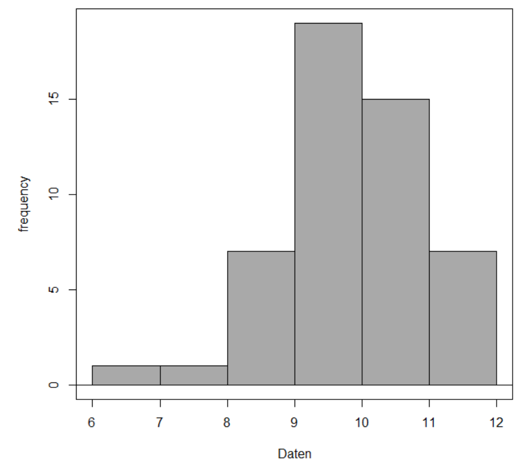
- Bei der ersten Analyse unserer Daten fallen immer wieder Werte auf, die scheinbar nicht zum Datensatz passen

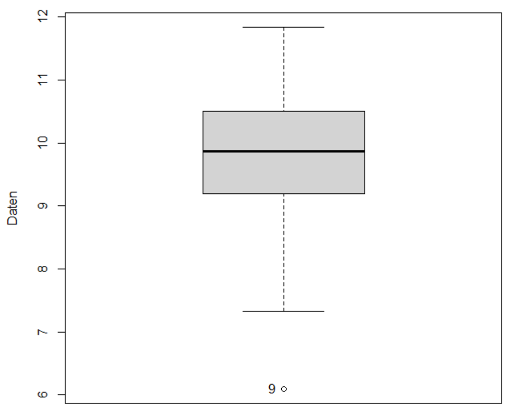
- Wir neigen dazu, diese Werte per se als Ausreißer zu identifizieren

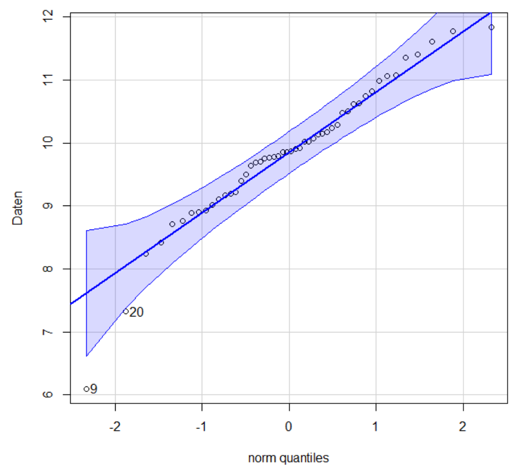
- Sie verhindern teilweise, dass wir parametrische Verfahren zur Datenanalyse nutzen, wir müssen auf nicht-parametrische Verfahren ausweichen

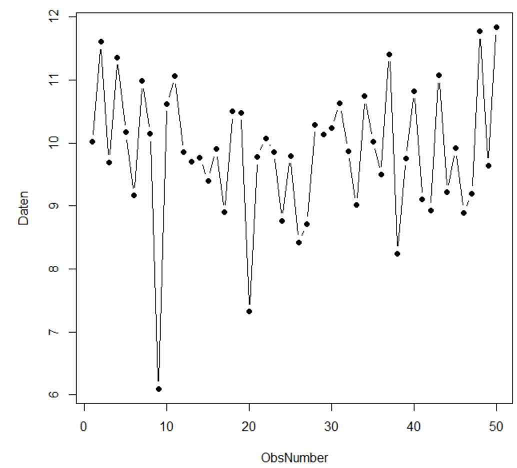
Ausreißer

- Datenpunkt 9 ist in mehreren Grafiken auffällig
- (Beispiel_Ausreisser.xlsx / Daten)



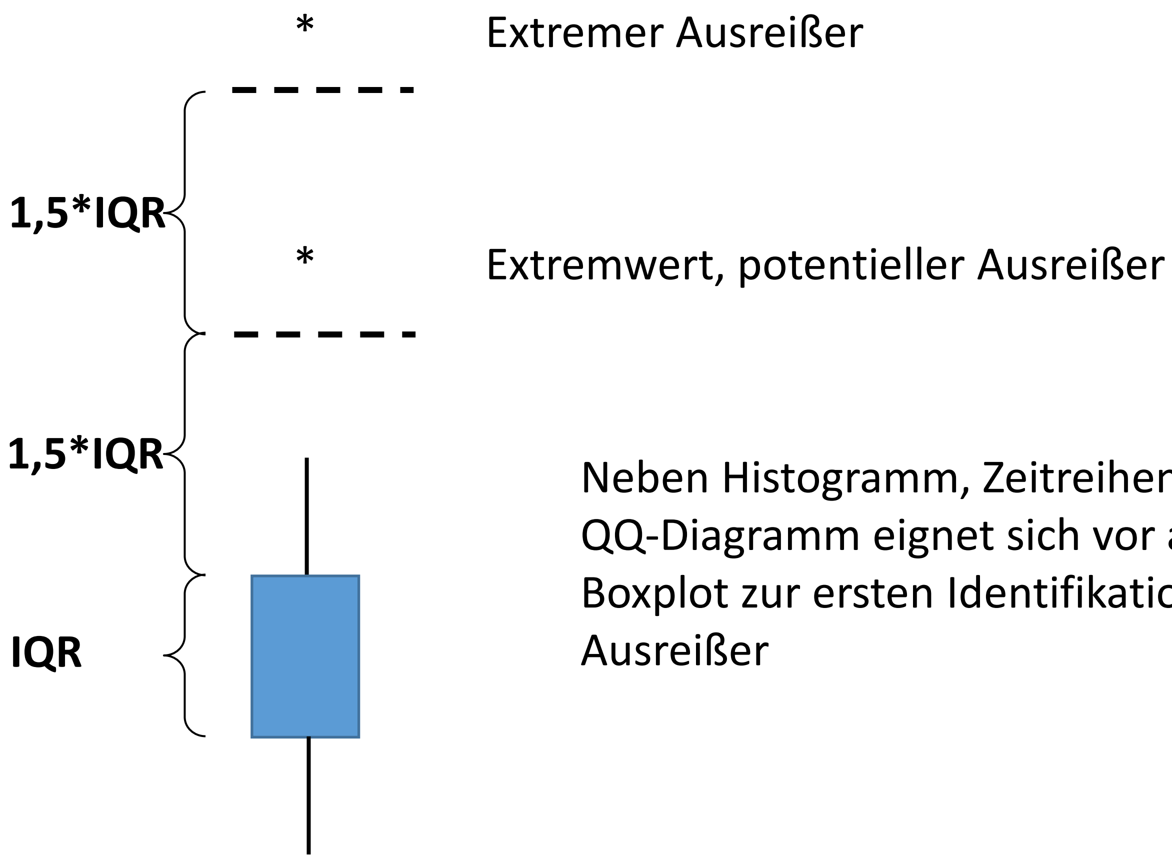






Ausreißer

- Definition von Ausreißern im Boxplot



Neben Histogramm, Zeitreihendiagramm und QQ-Diagramm eignet sich vor allem ein Boxplot zur ersten Identifikation möglicher Ausreißer

Ausreißer

- Führt man einen Test auf Normalverteilung durch, sieht man möglicherweise eine Nichtnormalverteilung

```
        Shapiro-Wilk normality test

data:  Daten
W = 0.95365, p-value = 0.04826
```

- Ohne auffällige Werte können wir unter Umständen von einer Normalverteilung ausgehen

```
        Shapiro-Wilk normality test

data:  DatenoA
W = 0.98526, p-value = 0.7924
```

Ausreißer

- Andererseits wissen wir, dass Normalverteilungen durchaus eine größere Bandbreite haben können

- Wir benötigen also Verfahren, die Werte hinsichtlich ihrer ungewöhnlichen Lage überprüfen

- Ausreißer-Test
  - Grubb‘s Test
  - David-Hartley-Pearson-Test
  - Dixon‘s r-Statistiken

- Voraussetzung für all diese Verfahren:
  Normalverteilte Daten (ohne Ausreißer)

Ausreißer

- Belegt ein Ausreißer-Test das Vorliegen eines Ausreißers, stellt sich die Frage: Wie mit diesem Wert umgehen?

- Mögliches Vorgehen:
  - Identifizierte Ausreißer auf Mess-, Rechen-, Schreib- oder Datenerfassungsfehler (systematischer Fehler) überprüfen und ggf. korrigieren/entfernen

  - Entscheiden, ob der Wert überhaupt zur Stichprobe gehört

  - Gegebenenfalls wird der Wert aus der Stichprobe entfernt

Ausreißer: Grubb‘s Test

- Testverfahren zur Überprüfung des Minimal- und/oder Maximalwertes des Datensatzes

Voraussetzungen:

- Die Stichprobe (ohne Ausreißer) ist normalverteilt

- Es befinden sich maximal zwei Ausreißer in der Stichprobe (ansonsten muss der Test mehrfach wiederholt werden, wobei identifizierte Ausreißer entfernt werden)

- Grubbs‘ Ausreißertest kann ein- oder zweiseitig angesetzt werden

Ausreißer: Grubb‘s Test

- Hypothesen

H_0 Der Datensatz enthält keine Ausreißer
H_1 Es gibt mindestens einen Ausreißer im Datensatz

- Teststatistik

$$G = \frac{max|x_i - \bar{x}|}{s}$$

mit $\bar{x}$: Stichprobenmittelwert
s: Standardabweichung der Stichprobe

Ausreißer: Grubb‘s Test

- Kritischer Wert G_{krit}
  - Beispiel: Zweiseitiger Test

$$G_{krit} = \frac{(n-1)}{\sqrt{n}} \sqrt{\frac{\left(t_{\frac{\alpha}{2n},n-2}\right)^2}{n-2+\left(t_{\frac{\alpha}{2n},n-2}\right)^2}}$$

$t_{\frac{\alpha}{2n},n-2}$ erhält man aus einer t-Verteilung mit $\mathrm{df} = n-2$ Freiheitsgraden und einem Signifikanzniveau von $\alpha/2n$

Für $G > G_{krit}$ wird die Nullhypothese verworfen

Ausreißer: Grubb‘s Test

- Kritischer Wert G_{krit}
  - Für den einseitigen Test wird das Signifikanzniveau zu α/n gewählt

Ausreißer: Grubb‘s Test

Berechnung in R

- Im RCommander sind keine Ausreißer-Tests hinterlegt

- Installation des Package „Outliers“

- install.packages(„outliers“)
- library(outliers)

Ausreißer: Grubb‘s Test

Berechnung in R

```
grubbs.test(x, type = 10, opposite = FALSE, two.sided = FALSE)
```

mit

x	Datenvektor
type	Testvariante (10 – ein Ausreißer; 11 – zwei Ausreißer auf gegenüberliegenden Seiten des Datensatzes; 20 – zwei Ausreißer auf einer Seite des Datensatzes)
opposite	FALSE – Überprüfung des Datenpunktes mit dem größten Mittelwertabstand TRUE – Überprüfung des gegenüberliegenden Punktes
two.sided	FALSE – einseitiger Test TRUE – zweiseitiger Test

Ausreißer: Grubb‘s Test

Berechnung in R

Vorsicht, nur folgende Kombinationen machen Sinn

`type = 10` **und** `two.sided = FALSE`

`type = 11` **und** `two.sided = TRUE`

andernfalls wird der α-Wert falsch interpretiert

Ausreißer: Grubb‘s Test

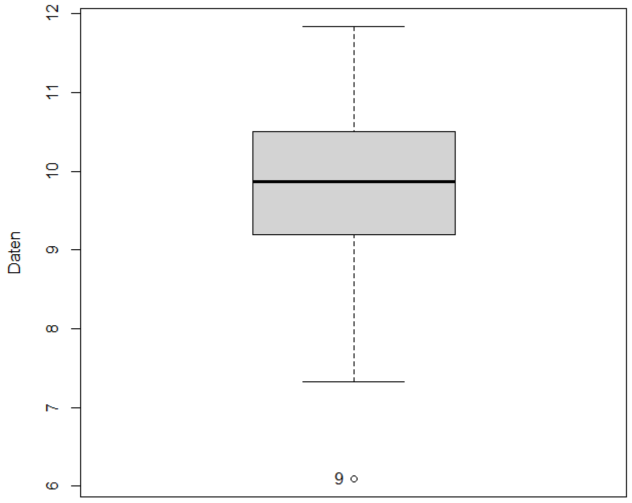
- Beispiel_Ausreisser.xlsx (Daten)

- Datenpunkt 9 ist ein potentieller Ausreißer

```
Shapiro-Wilk normality test

data:  Daten
W = 0.95365, p-value = 0.04826
```

- Der Originaldatensatz ist nicht normalverteilt



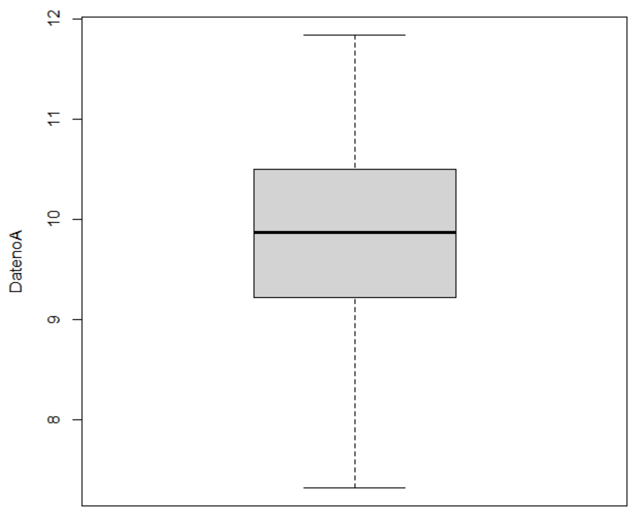
Ausreißer: Grubb‘s Test

- Beispiel_Ausreisser.xlsx (DatenoA)
- Überprüfung ohne Punkte 9

```
Shapiro-Wilk normality test

data:  DatenoA
W = 0.98526, p-value = 0.7924
```

- Der veränderte Datensatz ist normalverteilt



Ausreißer: Grubb‘s Test

- Beispiel_Ausreisser.xlsx (Daten)
- Hier: Test auf einen Ausreißer

```
> with(Dataset,grubbs.test(Daten,type=10,opposite=FALSE,two.sided=FALSE))


        Grubbs test for one outlier

data:  Daten
G = 3.46375, U = 0.75015, p-value = 0.005479
alternative hypothesis: lowest value 6.1004376 is an outlier
```

$p < 0{,}05$: Alternativhypothese – Datenpunkt 9 ist ein Ausreißer

Ausreißer: Grubb‘s Test

- Beispiel_Ausreisser.xlsx (Daten)
- Hier: Test auf zwei Ausreißer (gegenüberliegende Seiten)

```
> with(Dataset,grubbs.test(Daten,type=11,opposite=FALSE,two.sided=TRUE))


        Grubbs test for two opposite outliers

data:  Daten
G = 5.29914, U = 0.68528, p-value = 0.1517
alternative hypothesis: 6.1004376 and 11.833236 are outliers
```

p > 0,05: Nullhypothese – Datenpunkte 9 und 50 sind keine Ausreißer

Ausreißer: Grubb‘s Test

- Beispiel_Ausreisser.xlsx (Daten2A)
- Hier: Test auf zwei Ausreißer (gegenüberliegende Seiten)

```
> with(Dataset,grubbs.test(Daten2A,type=11,opposite=FALSE,two.sided=TRUE))


        Grubbs test for two opposite outliers

data:  Daten2A
G = 5.97107, U = 0.63081, p-value = 0.00871
alternative hypothesis: 6.1004376 and 12.833236 are outliers
```

p < 0,05: Alternativhypothese – Datenpunkte 9 <u>und</u> 50 sind Ausreißer