UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

STATISTISCHE TESTS
TEIL 2

KONFIDENZINTERVALLE

Bahrenberg I: Kap. 5;
Ernste Anh. B;
Ewing I: Kap. 10

T-Test für den Mittelwert

(Vergleich von Stichprobemittel und mit fixem Wert,
z.B. bekanntem Mittel der GG oder Grenzwert)

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Hypothesen:

H_0: Erwartete Sept.-Nov. Mitteltemperatur beträgt 9ºC ($x = \mu_0$)

H_A: Die Temperatur weicht signifikant von 9ºC ab ($x \neq \mu_0$)

Teststatistik: $T = \frac{\bar{x} - \mu_0}{s_x}$ x=8.75ºC, s_x=0.10ºC, n=111, T=-2.40

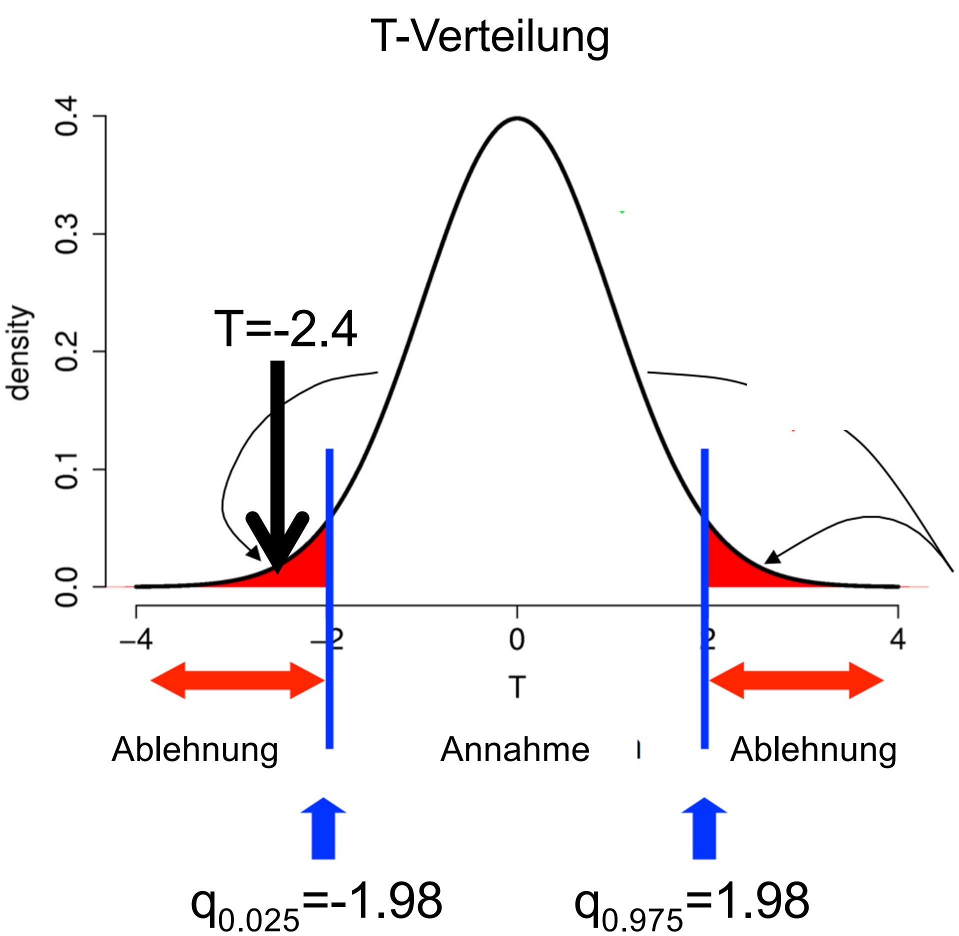
ähnlich der Standardisierung / Z-Transformation / Standardnormalverteilung

mit dem Standardfehler $s_x = \sqrt{\frac{s^2}{n}}$

T-Test für den Mittelwert

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH



Test

> Wert von T (= –2.40) ist im Ablehnungsbereich

> Stichprobenmittelwert unterscheidet sich von Erwartungswert (9ºC) auf dem 5% Signifikanzlevel.

> Test p-Wert von 0.018

p-Wert

> Das kleinste α für das H_0 abgelehnt würde

Testtheorie
Hypothesen aufstellen

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Wie im Gericht: Beratungen und Überlegungen

- Angenommen H_0 ist wahr, wie wahrscheinlich ist es zufällig diesen oder einen grösseren Testwert zu erhalten?
- Der p-Wert ist die numerische Antwort
- Je kleiner p, desto stärker die Beweise gegen H_0 (oft 0.05 als Grenze)
- p-Wert sagt **NICHT**: Wie wahrscheinlich ist es, dass die H_0 wahr ist
- p-Wert sagt: Wie wahrscheinlich die Daten sind, wenn H_0 wahr ist

Testtheorie
Hypothesen aufstellen

b
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Wie im Gericht: Das Urteil

wichtig!

Option 1:

> grosser p-Wert ($p > \alpha$, z.B. mit $\alpha=0.05$ für 95%ige Sicherheit)
> wir schliessen daraus -> Daten sind konsistent mit H_0

Option 2:

> kleiner p-Wert ($p < \alpha$)
> wir schliessen daraus -> es gibt ausreichend Beweise, um H_0 abzulehnen und H_A anzunehmen (statistisch signifikant).

Konfidenzintervalle

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Eine Masszahl für die Unsicherheit der Parameterschätzung
- Wertebereich eines Parameters, für den H_0 nicht abgelehnt wird
- Eng verknüpft mit dem Signifikanzlevel α !

Beispiel:

$$T = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} \quad \Rightarrow \quad \mu \in \left\{ \bar{x} \pm q_{T,0.975} \cdot s_{\bar{x}} \right\}$$

in diesem Beispiel das 0.975 Quantil der T-Verteilung mit n-1 Freiheitsgraden.
n: Stichprobengrösse

Konfidenzintervalle

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

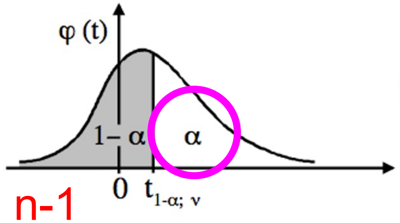
- Wie repräsentativ sind unsere Stichprobenergebnisse für die Grundgesamtheit?
- Wir wollen den Unsicherheitsbereich um den Stichprobenmittelwert bestimmen, sogenanntes Konfidenzintervall:

$$\mu_{Obergrenze} = \bar{x} + q_{1-\alpha} \cdot \sqrt{\frac{s^2}{n}}$$

$$\mu_{Untergrenze} = \bar{x} - q_{1-\alpha} \cdot \sqrt{\frac{s^2}{n}}$$

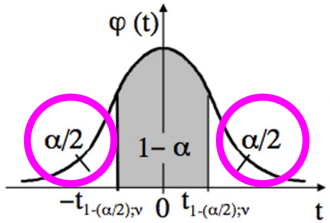
(identisch mit vorheriger Folie, nur inklusive Berechnung der Standardfehlers!)

- q abzulesen aus t-Tabelle normalerweise für den zweiseitigen Test (wenn Ober- und Untergrenze des Konfidenzintervals bestimmt werden), n-1 Freiheitsgrade und α oder aus
- **R:** `t.test(x, alternative = ("two.sided" ODER "less‟ ODER "greater"), conf.level = 0.95, ...)`
- **Excel** Funktion T.INV bzw. T.INV.2T, Excel Funktion T.TEST (zur Bestimmung des p-Wertes).



$t_{1-\alpha} = -t_{\alpha}$

n-1



$t_{1-(\alpha/2)} = -t_{\alpha/2}$

t-Verteilungs Tabelle

FG

ν	Statistische Sicherheit 1-α					
	0,90	0,95	0,975	0,99	0,995	0,999
1	3,078	6,314	12,71	31,82	63,66	318,3
2	1,886	2,920	4,303	6,965	9,925	22,33
3	1,638	2,353	3,182	4,541	5,841	10,21
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
11	1,363	1,796	2,201	2,718	3,106	4,025
12	1,356	1,782	2,179	2,681	3,055	3,930
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,337	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	1,734	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,485
24	1,318	1,711	2,064	2,592	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,396
30	1,310	1,697	2,042	2,457	2,750	3,385
40	1,303	1,684	2,021	2,443	2,704	3,307
50	1,299	1,676	2,009	2,403	2,678	3,261
60	1,296	1,671	2,000	2,390	2,660	3,232
70	1,294	1,667	1,994	2,381	2,648	3,211
80	1,292	1,664	1,990	2,374	2,639	3,195
90	1,291	1,662	1,987	2,368	2,632	3,183
100	1,290	1,660	$\boxed{1,984}$	2,364	2,626	3,174
200	1,286	1,652	1,972	2,345	2,601	3,131
500	1,283	1,648	1,965	2,334	2,586	3,107
∞	1,282	1,645	1,960	2,326	2,576	3,090

oder entsprechend -1.98

Statistische Sicherheit 1-α							ν
0,80	0,90	0,95	0,98	0,99	0,998	0,999	
3,078	6,314	12,71	31,82	63,66	318,3	636,6	1
1,886	2,920	4,303	6,965	9,925	22,33	31,60	2
1,638	2,353	3,182	4,541	5,841	10,21	12,92	3
1,533	2,132	2,776	3,747	4,604	7,173	8,610	4
1,476	2,015	2,571	3,365	4,032	5,893	6,869	5
1,440	1,943	2,447	3,143	3,707	5,208	5,959	6
1,415	1,895	2,365	2,998	3,499	4,785	5,408	7
1,397	1,860	2,306	2,896	3,355	4,501	5,041	8
1,383	1,833	2,262	2,821	3,250	4,297	4,781	9
1,372	1,812	2,228	2,764	3,169	4,144	4,587	10
1,363	1,796	2,201	2,718	3,106	4,025	4,437	11
1,356	1,782	2,179	2,681	3,055	3,930	4,318	12
1,350	1,771	2,160	2,650	3,012	3,852	4,221	13
1,345	1,761	2,145	2,624	2,977	3,787	4,140	14
1,341	1,753	2,131	2,602	2,947	3,733	4,073	15
1,337	1,746	2,120	2,583	2,921	3,686	4,015	16
1,333	1,740	2,110	2,567	2,898	3,646	3,965	17
1,330	1,734	2,101	2,552	2,878	3,610	3,992	18
1,328	1,729	2,093	2,539	2,861	3,579	3,883	19
1,325	1,725	2,086	2,528	2,845	3,552	3,850	20
1,323	1,721	2,080	2,518	2,831	3,327	3,819	21
1,321	1,717	2,074	2,508	2,819	3,505	3,792	22
1,319	1,714	2,069	2,500	2,807	3,485	3,768	23
1,318	1,711	2,064	2,492	2,797	3,467	3,745	24
1,316	1,708	2,060	2,485	2,787	3,450	3,725	25
1,315	1,706	2,056	2,479	2,779	3,435	3,707	26
1,314	1,703	2,052	2,473	2,771	3,421	3,690	27
1,313	1,701	2,048	2,467	2,763	3,408	3,674	28
1,311	1,699	2,045	2,462	2,756	3,396	3,659	29
1,310	1,697	2,042	2,457	2,750	3,385	3,646	30
1,303	1,684	2,021	2,423	2,704	3,307	3,551	40
1,299	1676	2,009	2,403	2,678	3,261	3,496	50
1,296	1,671	2,000	2,390	2,660	3,232	3,460	60
1,292	1,664	1,990	2,374	2,639	3,195	3,416	80
1,290	1,660	$\boxed{1,984}$	2,364	2,626	3,174	3,390	100
1,286	1,652	1,972	2,345	2,601	3,131	3,340	200
1,283	1,648	1,965	2,334	2,586	3,107	3,310	500
1,282	1,645	1,960	2,326	2,576	3,090	3,291	∞

nur der positive Wert angegeben,
beim 2-seitgen Test ist der negative
kritische Grenzwert entsprechend -1.98

Konfidenzintervalle

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

95% Konfidenzintervall der erwarteten Temperatur aus dem vorherigen Beispiel:

- alle Werte, für die T kleiner ist als der kritische Wert mit:
- x = 8.75ºC, s_x = 0.10ºC, n =111, $q_{T,0.95(\text{zweiseitig})}$ =1.98

Konfidenzintervall:

- Untergrenze: 8.75-1.98*0.1=8.55ºC
- Obergrenze: 8.75+1.98*0.1=8.95ºC

Anmerkung:

- 9ºC liegt nicht im Konfidenzintervall, was im Einvernehmen mit der vorherigen Ablehnung von H_0 (μ = 9) steht

Daumenregel:

- Konfidenzintervall = Stichprobenergebnis ± 2 Standardfehler

Was sind die Voraussetzungen, damit der t-Test angewendet werden darf?

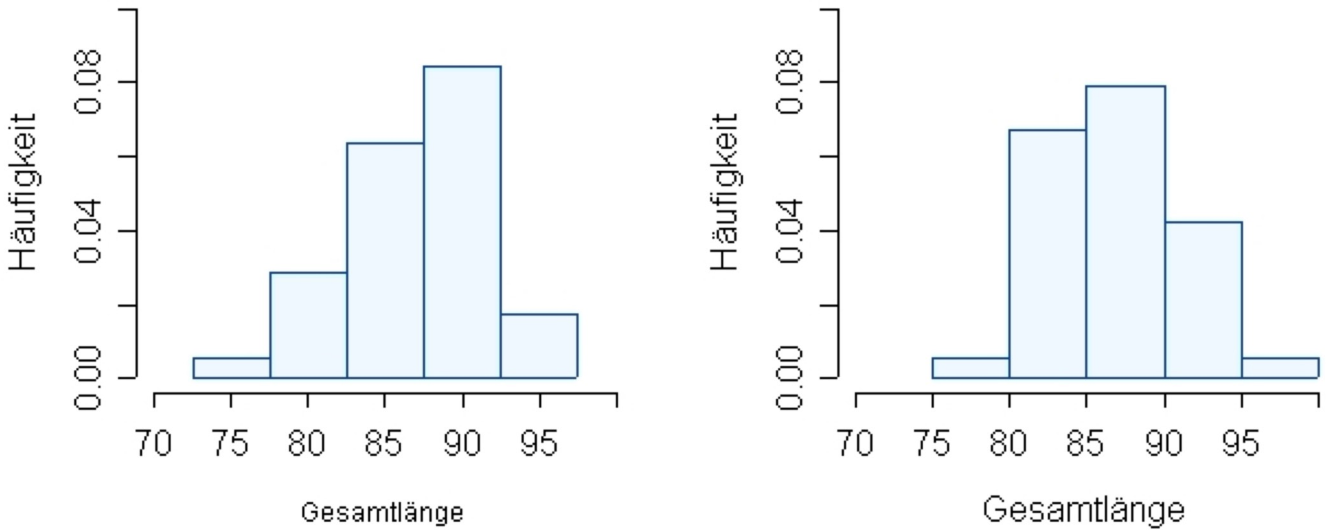
Was sind die Voraussetzungen, damit der t-Test angewendet werden darf?

Für den T-Test alles, was Voraussetzung für das arithmetische Mittel ist, d.h. metrische Daten, keine Ausreisser in den Daten, und symetrische Verteilung

Überprüfung auf Normalverteilung

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

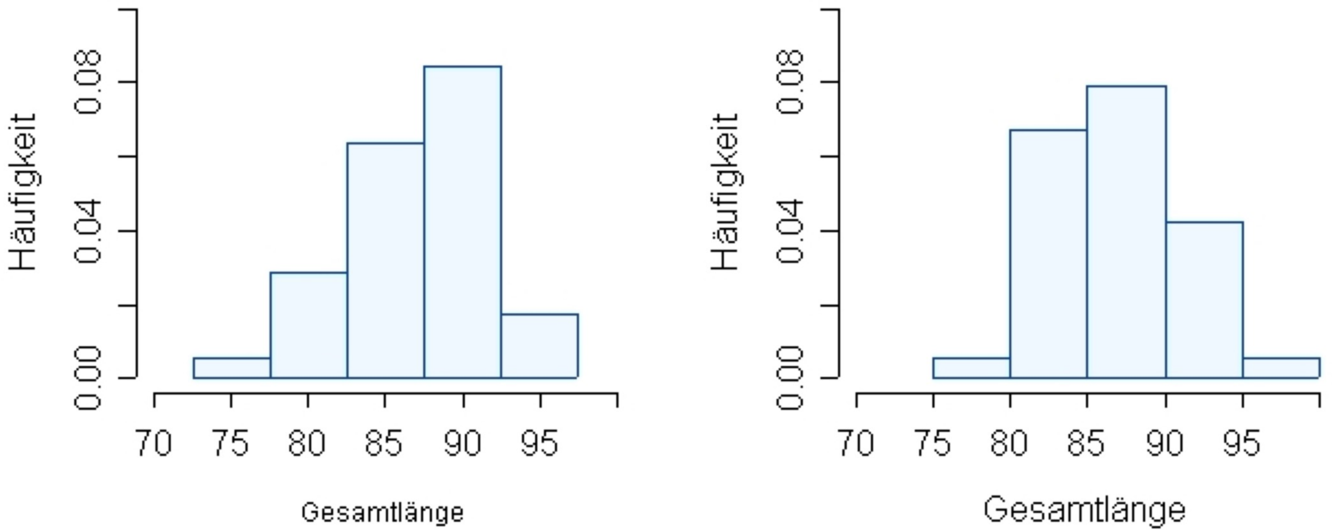


Sind diese Daten normalverteilt?

Überprüfung auf Normalverteilung

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH



Gleiche Daten, aber

visuell unterschiedliche Histogramme

Hypothesen aufstellen

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

> Wie lautet H_0 und H_A beim Test, ob die Daten normalverteilt sind?

Hypothesen aufstellen

b
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

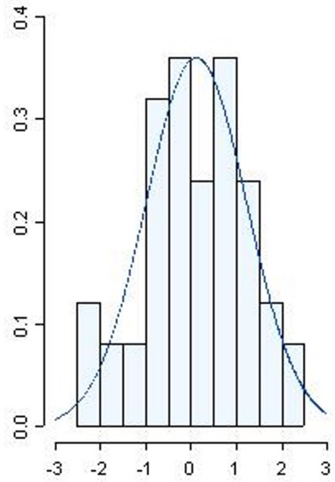
Warum ist Normalverteilung Nullhypothese, wenn man doch gerade beweisen will, dass Daten normalverteilt sind?

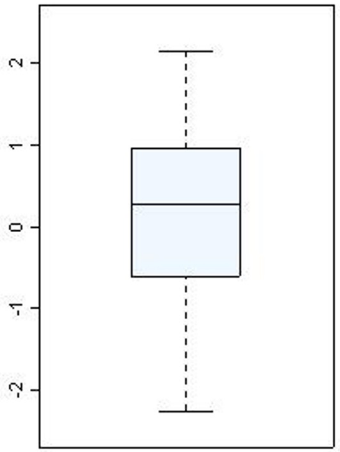
Ein Test sagt ja nur gleich oder ungleich bzw. es gibt (nicht) ausreichend Beweise gegen H_0. Wenn wir nun testen, ob H_0 nicht z.B. gleichverteilt,... ist. Wenn H_0 verworfen werden kann, heisst das aber noch nicht, dass die Daten normalverteilt sind, könnten auch irgendwie anders verteilt sein.

Der einzige Weg, um herauszubekommen, ob die Daten normalverteilt sind, ist also zu prüfen ob man die Hypothese der Normalverteilung widerlegen kann!

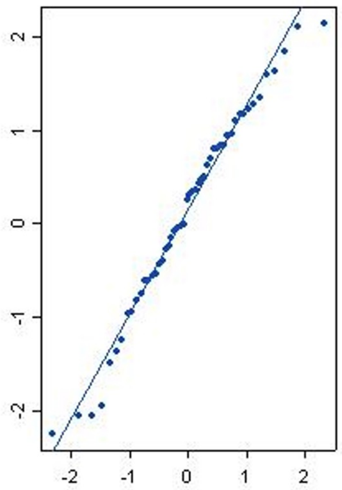
Daten sind gleich der Normalverteilung
Das = steht also wieder in H_0

Überprüfung auf Normalverteilung





*



Kenngröße	Messreihe A	Normalverteilung
(Mittelwert - Median) / s	-0,14	0
IQR / s	1.36	1,34
# 1s Intervall / n	72 %	68 %
# 2s Intervall / n	98 %	95 %
# 3s Intervall / n	100 %	99,73 %

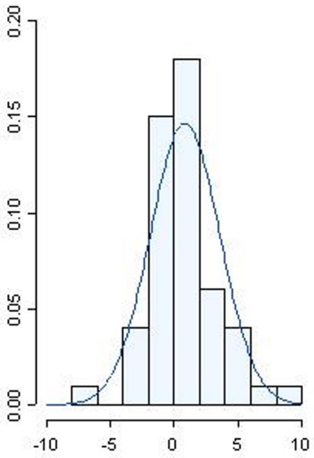
Test	p-Wert
Shapiro-Wilks	0,4544
Anderson-Darling	0,6796
Cramér-von Mises	0,7521

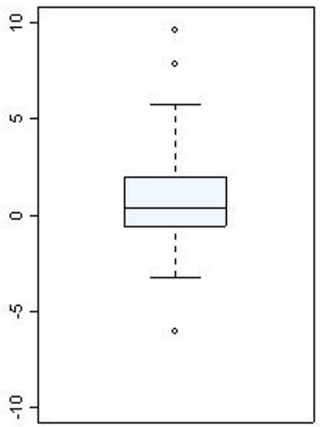
Die p-Werte der Tests auf Normalverteilung sind alle deutlich größer als 0.05, somit wird die Nullhypothese der Normalverteilung beibehalten.

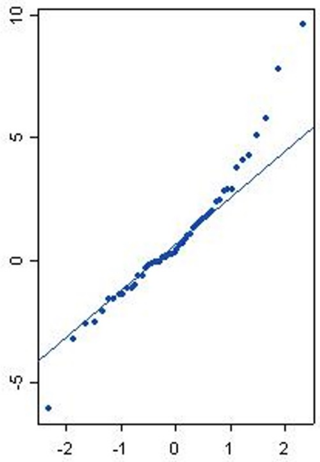
Beim Test auf Normalverteilung wird immer in der Nullhypothese angenommen, dass die Messreihe aus einer Normalverteilung stammt.

*QQ-Plot: Quantile zweier statistischer Variablen gegeneinander abgetragen werden, um ihre Verteilungen zu vergleichen

Überprüfung auf Normalverteilung







Kenngröße	Messreihe B	Normalverteilung
(Mittelwert - Median) / s	0,18	0
IQR / s	0,94	1,34
# 1s Intervall / n	76 %	68 %
# 2s Intervall / n	94 %	95 %
# 3s Intervall / n	98 %	99,73 %

Test	p-Wert
Shapiro-Wilks	0,0303
Anderson-Darling	0,0285
Cramér-von Mises	0,0331

Die p-Werte der Tests auf Normalverteilung lehnen in jedem Fall die Nullhypothese zum Niveau α=0,05=5% ab

p-Werte sind deutlich kleiner als 0,05, d. h. die Nullhypothese "Messreihe ist normalverteilt" wird verworfen

Überprüfung auf Normalverteilung in R

Test	Vorteile	Nachteile
Chi-Quadrat-Test	- geeignet für beliebig skalierte Variablen	- Gruppierung der Beobachtungen notwendig - ungeeignet für kleine Stichproben - quadratische Testgrösse, d.h. sensibel auf Ausreisser
Kolmogorov-Smirnov Test	- geeignet für kleine Stichproben - wie Chi-Quadrat auch zum Vergleich anderer Verteilungen geeignet - nicht-parametrischer Test, d.h. NICHT sensibel auf Ausreisser	- geringe Teststärke im Vergleich zu den folgenden Tests
Cramér-von-Mises-Test	- höhere Güte als KS-Test	- quadratische Testgrösse
Lilliefors-Test	- bessere Trennschärfe als KS-Test - nicht-parametrischer Test, d.h. NICHT sensibel auf Ausreisser	- nur zum Test auf Normalverteilung
Anderson-Darling-Test	- sehr hohe Güte bei Test auf Normalverteilung	- keine kategorialen Daten - quadratische Testgrösse, d.h. sensibel auf Ausreisser
Shapiro-Wilk-Test	- Test mit höchster Güte	- ausschließlich Test auf Normalverteilung - manuell schlecht durchführbar - sensibel auf Ausreisser und viele identische Werte

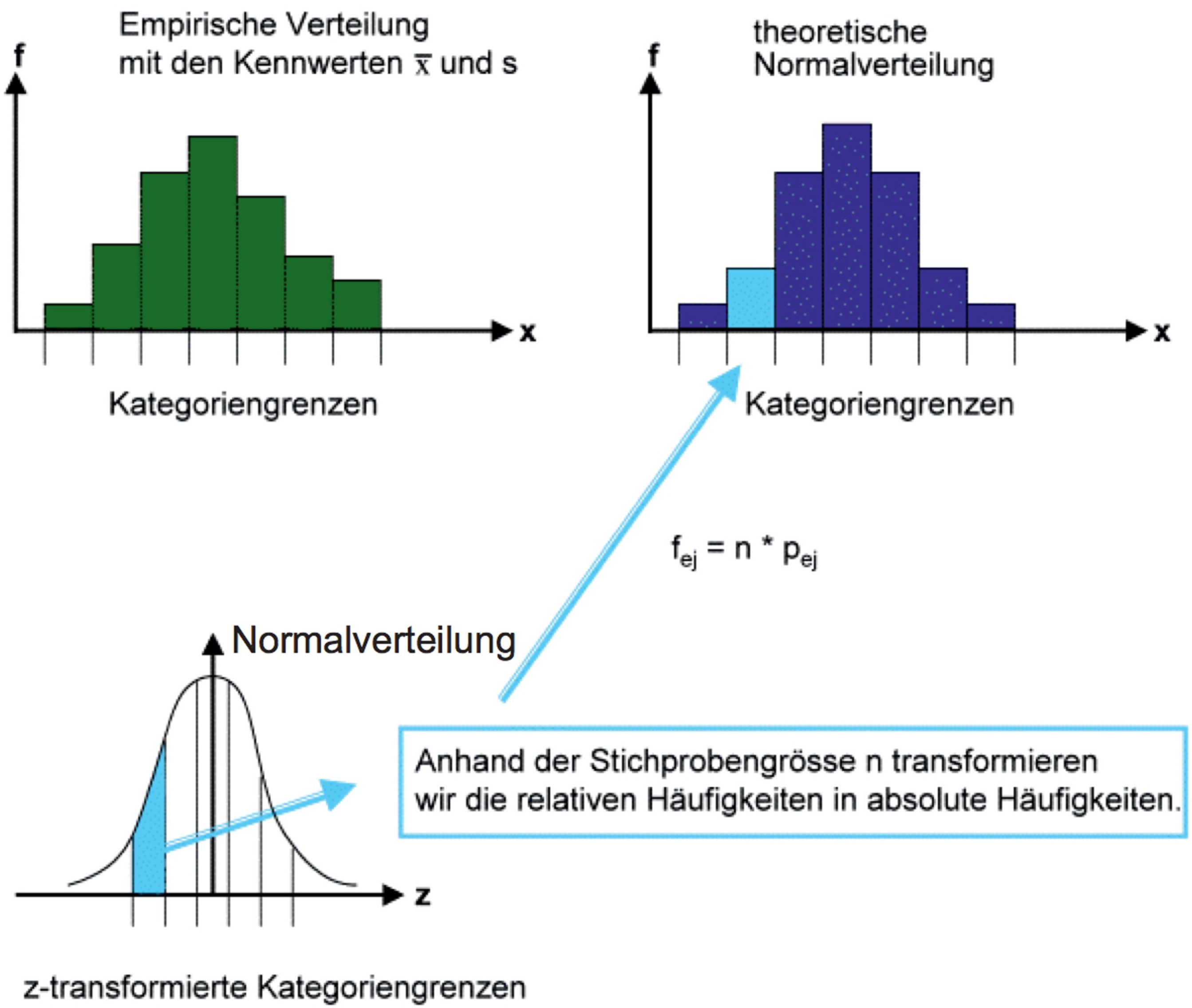
Überprüfung auf Normalverteilung in R

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

```
> data <- rnorm(500) # 500 normalverteile Zufallszahlen
> data2 <- data^2    # quadrierte Zufallszahlen
> hist(data)         # Histogramm
> plot(density(data))# Dichtefunktion
> boxplot(data)      # Boxplot
> qqnorm(data)       # Quantil-Quantil Plot
> qqline(data, col = 2) # Erwartung bei Normalverteilung
> shapiro.test(data) # Shapiro Normalverteilungstest
```

χ^2 Test auf Normalverteilung



Anhand der Stichprobengrösse n transformieren wir die relativen Häufigkeiten in absolute Häufigkeiten.

Allgemeiner Test für Verteilungen
χ^2-Verteilungstest

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Summe der normierten quadrierten Abweichungen:

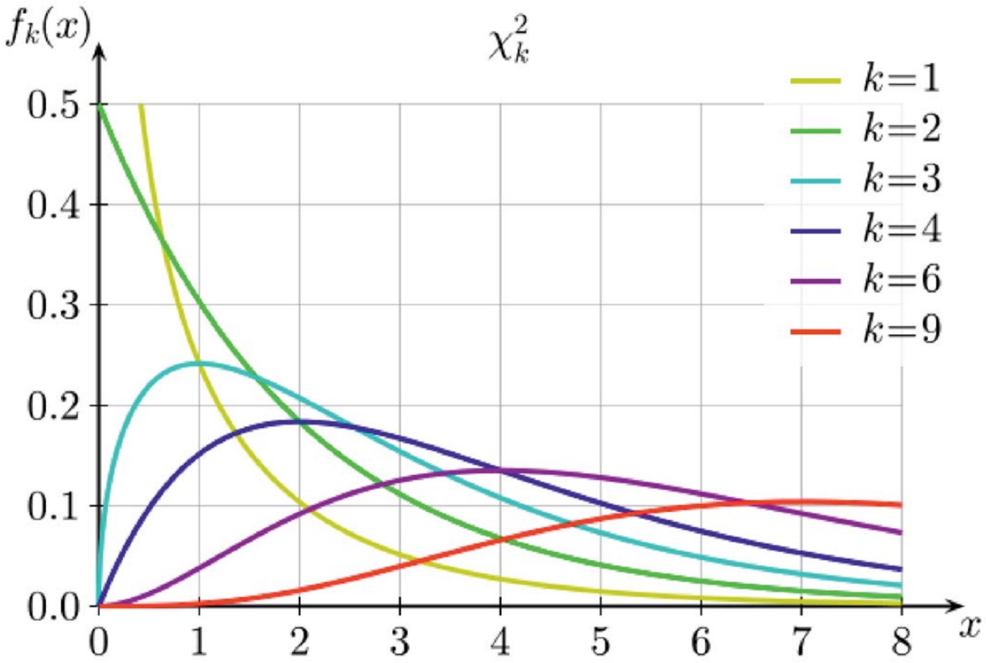
$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - n_i)^2}{n_i}$$

- N beobachtete Häufigkeit
- n erwartete Häufigkeit
- i Klassen
- k Anzahl an Klassen

χ^2-Verteilung

UNIVERSITÄT
BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Stetige Wahrscheinlichkeitsverteilung mit der Anzahl Freiheitsgrade k als einzigem Parameter.
- Verteilung der Summe der Quadrate von k unabhängigen
- und standardnormalverteilten Zufallsvariablen.

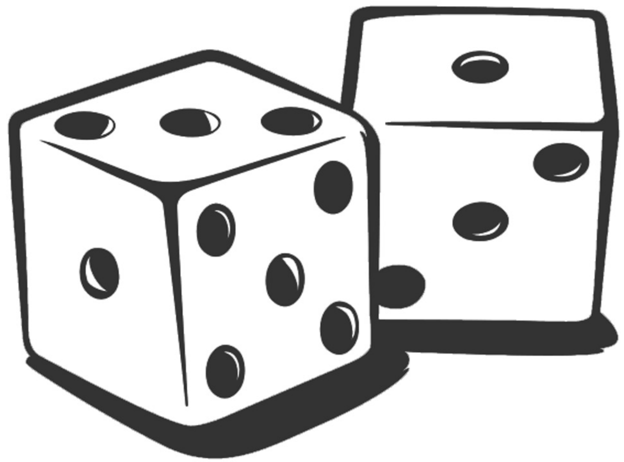


Würfelergebnisse

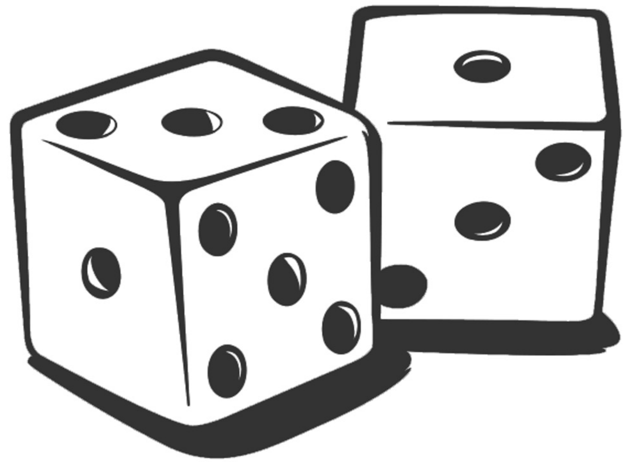
UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

> Was sind die Null-, und Alternativhypothese unseres Würfelexperiments?



Würfelergebnisse



- Was sind die Null-, und Alternativhypothese unseres Würfelexperiments?

- H_0: Die Verteilung der gewürfelten Zahlen entspricht den theoretischen Wahrscheinlichkeiten

- H_A: Die Verteilung der gewürfelten Zahlen entspricht NICHT den theoretischen Wahrscheinlichkeiten

Richtigen Test und statistische Verfahren finden

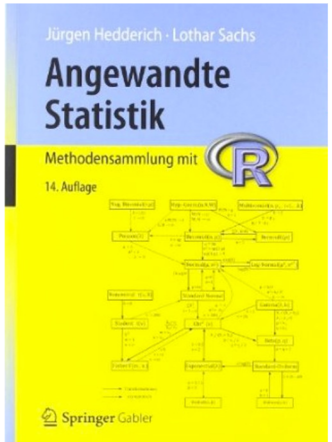
b
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Chatbot?

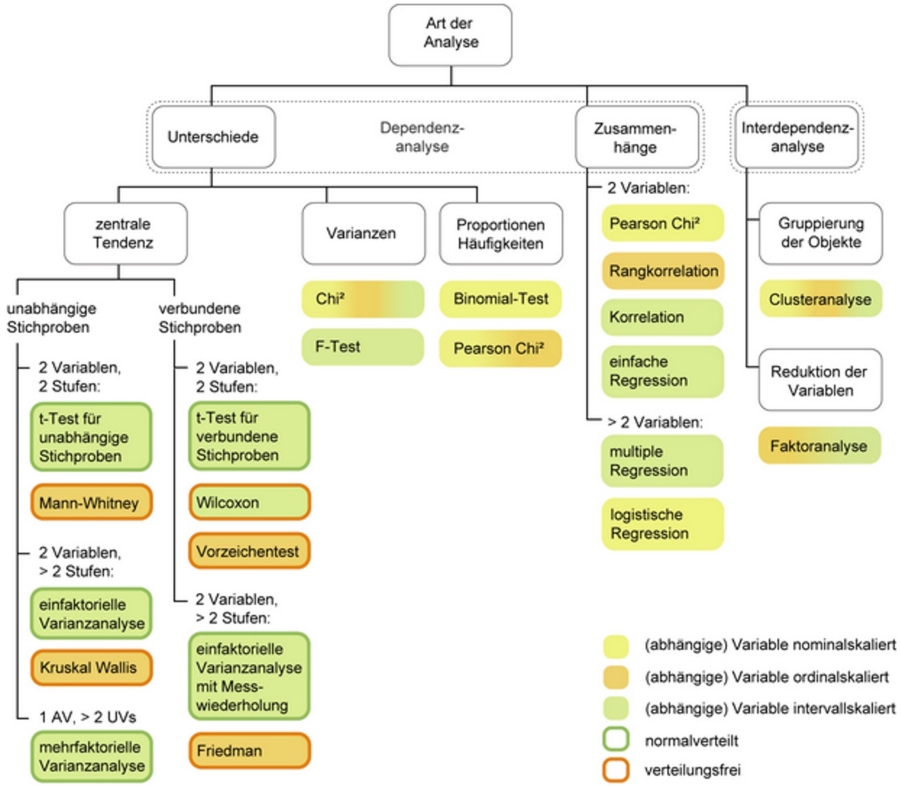
Bücher:

- Ernste S. 21 ff.
- Jürgen Hedderich, Lothar Sachs, 2012: Angewandte Statistik: Methodensammlung mit R. Springer



Entscheidungsbäume

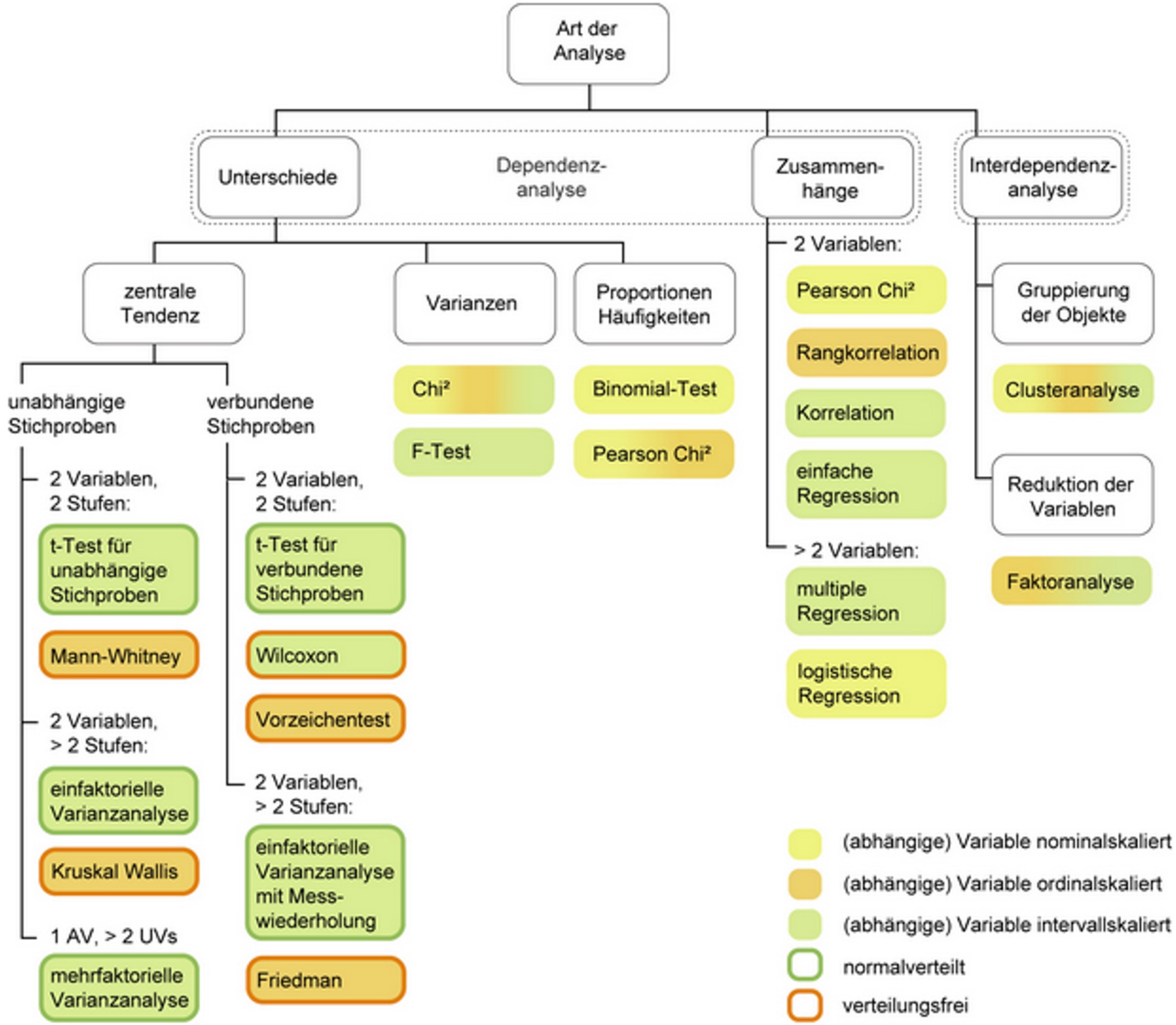
- http://www.methodenberatung.uzh.ch/index.html
- http://etools.fernuni.ch/entscheidungsbaum/



Angenommen ihr wollt Winter- und Sommermittelwerte von nicht-normalverteilten Niederschlagssummen vergleichen. Welchen Test müsst ihr nehmen?

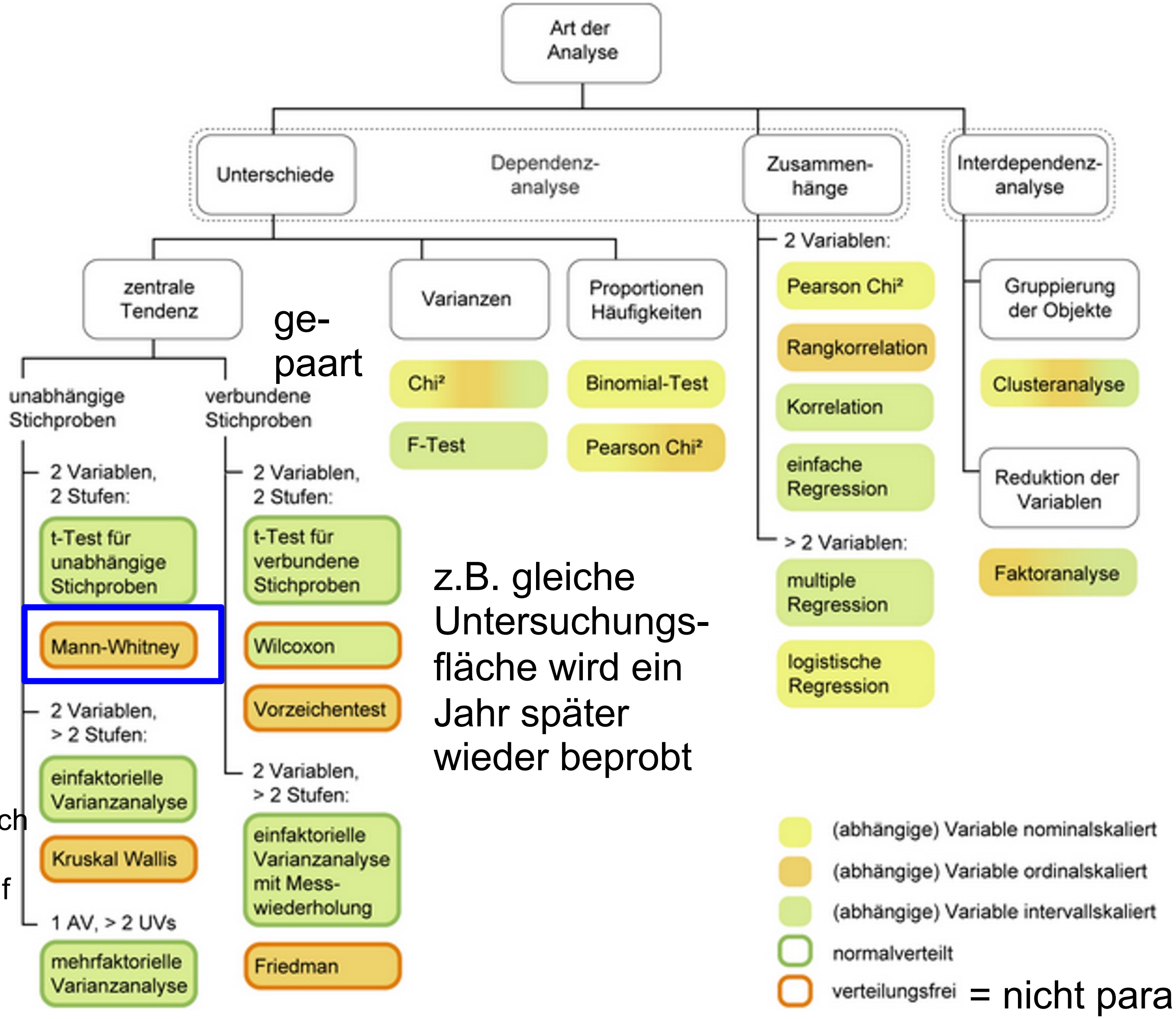
UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH



UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH



unge-paart

ge-paart

verteil-ungsfrei

ANOVA=
z.B. Vergleich vieler Medi-kamente auf die Genes-ungsdauer

z.B. gleiche Untersuchungs-fläche wird ein Jahr später wieder beprobt

Verfahren kann oft auch für metrische Daten verwendt werden, die nicht normalverteilt sind

verteilungsfrei = nicht parametrische Tests

AV=abhängige Variable
UV=unabhängige Variable

Take-home Message

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Standardvorgehen bei statistischen Tests:

1. Hypothesen aufstellen

2. Schwellenwert festlegen (z.B. $\alpha = 0.05$, d.h. bei jedem 20. Test eine falsche Entscheidung)

3. Ziehung der Stichprobe

4.1. Traditionelle und manuelle Zusammenfassung der Daten in einem Wert (einer Teststatistik) und Vergleich ob im Annahme- oder Ablehnungsbereich mithilfe einer Tabelle **oder**

4.2. p(robability)-Wert durch Software ermitteln. Wenn der p-Wert < als z.B. 0.05 ist bedeutet dies, dass es eine sehr kleine (<5%) Wahrscheinlichkeit gibt, die Stichprobe zu erhalten, wenn die Nullhypothese wahr ist **oder**

4.3. Konfidenzintervalle bestimmen. Diese beschreiben (z.B. für den Mittelwert) einer Stichprobe, den Bereich in dem z.B. mit 95% Sicherheit der entsprechende Wert (hier Mittelwert) Grundgesamtheit liegt.

Schliessende Statistik beruht meist auf Wahrscheinlichkeiten und liefert uns KEINE 100% sicheren Ergebnisse!

Enger mathematischer Zusammenhang zwischen Tests und Konfidenzintervallen

Alle Tests funktionieren nach ähnlichem Muster: Wenn ihr einen Verstanden habt, könnt ihr jeden beliebigen Test in eurer eigenen Arbeit anwenden. Webseiten und Bücher helfen den richtigen Test zu finden, enthalten Informationen über Freiheitsgrade, etc.

Statistisch signifikant

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

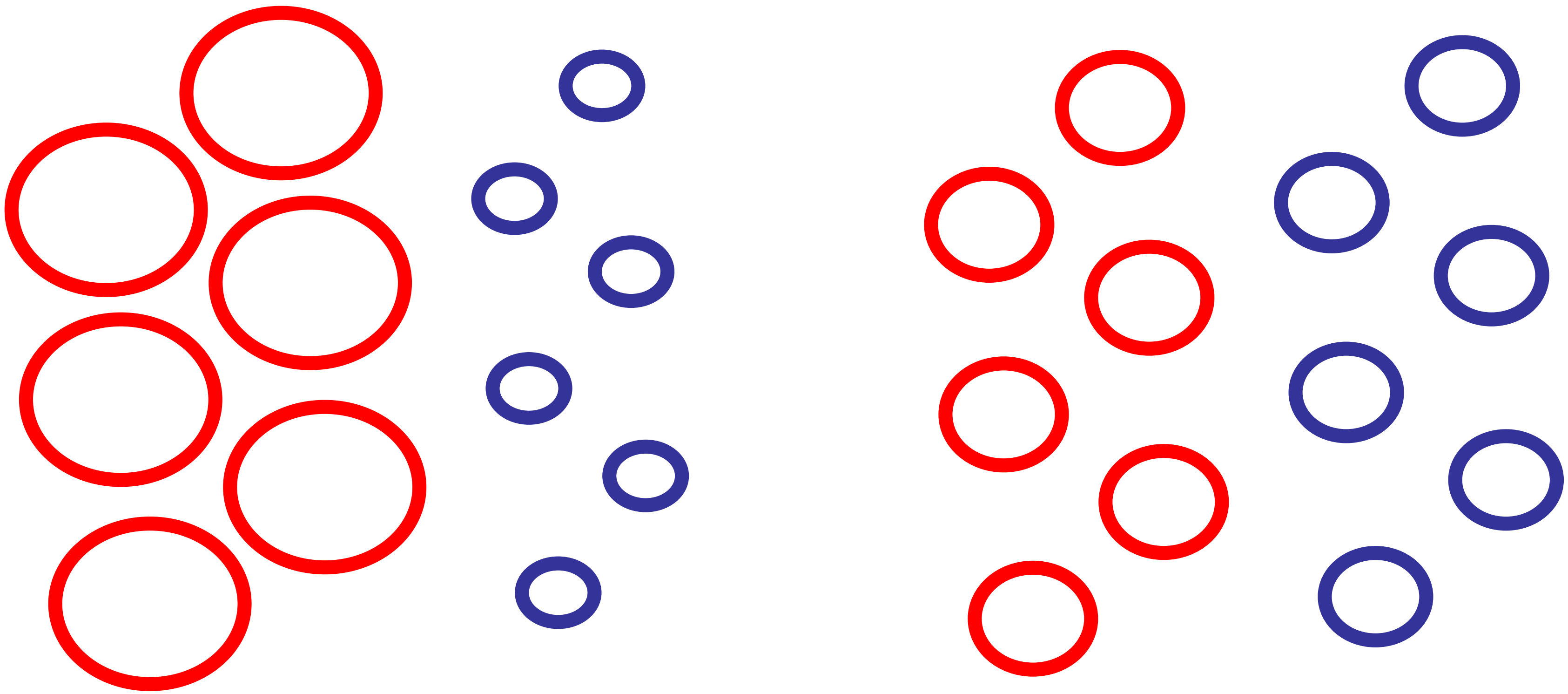
- Eine Studie mit 10000 Personen zeigt, dass ein neues Medikament die Dauer einer Grippe statistisch signifikant ($\alpha<0.05$) um eine Stunde verkürzt.
- Was haltet ihr von dem Medikament?

ACHTUNG:
Bei einer grossen Stichprobe kann auch ein kleiner Unterschied statistisch signifikant sein

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Ob die Grösse des Unterschieds eine Bedeutung hat, müsst ihr interpretieren:
- **Grösse des Effekts**



UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Beispiel Prüfungsfrage

- Was versteht man in der Statistik unter dem “Standardfehler”?
  - Abweichungen der Stichprobenwerte vom Stichprobenmittelwert
  - Abweichungen der Stichprobenmittelwerte vom Mittelwert der Grundgesamtheit
  - Die Varianz der Mittelwerte von mehreren Stichproben der gleichen Grundgesamtheit
  - Die Varianz der Stichprobenwerte von Stichproben aus verschiedenen Grundgesamtheiten
- Interpretiert die R Ausgabe mit dem Ergebnis eines t-Tests (siehe Übung aus Teil 1)
- Welcher statistische Test eignet, um die Mittelwerte nicht normalverteilter, gepaarter Stichproben zu vergleichen (mit Hilfe eines Entscheidungsdiagramms)?