HTW Chur

Hochschule für Technik und Wirtschaft
University of Applied Sciences

Parametrische und nicht-parametrische Testverfahren

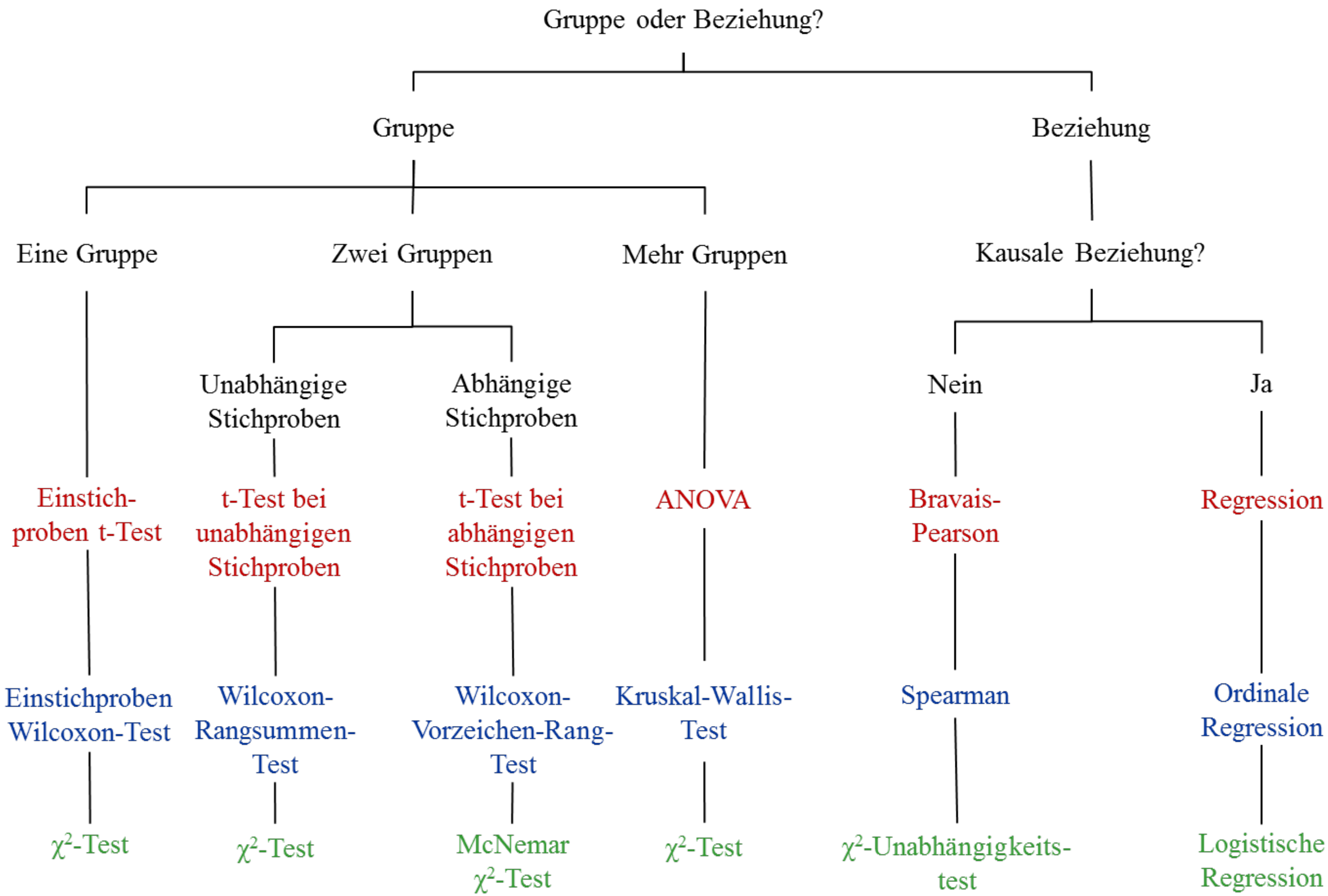
Quantitative Forschungsmethoden

Prof. Dr. Franz Kronthaler

Lernziele

- verstehen des Unterschieds zwischen parametrischen und nicht-parametrischen Testverfahren
- wissen wann parametrische oder nicht-parametrische Testverfahren sachgemäss sind
- wissen wie der t-Test für unabhängige Stichproben berechnet und interpretiert wird
- wissen wie der Wilcoxon-Rangsummen-Test berechnet und interpretiert wird
- wissen wie der t-Test für abhängige Stichproben berechnet und interpretiert wird
- wissen wie der Wilcoxon-Vorzeichen-Rang-Test berechnet und interpretiert wird
- verstehen wie das angemessene Testverfahren basierend auf der Fragestellung, der Daten und der Stichprobengrösse auszuwählen ist

Einleitung



Variable(n) von Interesse sind:
- metrisch und normal verteilt
- ordinal oder metrisch und nicht-normal verteilt
- nominal

Parametrische und nicht-parametrische Testverfahren

- parametrische Testverfahren erfordern Kenntnis über die Verteilung der Testdaten, typischerweise wird der Ausdruck für Tests verwendet, die metrisch und normalverteilte Daten voraussetzen
- Beispiele für parametrische Testverfahren sind
  - der t-Test bei unabhängigen Stichproben
  - der t-Test bei abhängigen Stichproben
  - die ANOVA
- nicht-parametrische Testverfahren sind Testverfahren, die keine spezifische Verteilung der Daten voraussetzen
- nicht-parametrische Testverfahren werden auch als verteilungsfreie Tests bezeichnet
- Beispiele für nicht-parametrische Testverfahren sind
  - die χ^2-Tests
  - der Wilcoxon-Rangsummen-Test
  - der Wilcoxon-Vorzeichen-Rang-Test
  - der Kruskal-Wallis-Test

Parametrische und nicht-parametrische Testverfahren

- Gründe für die Anwendung eines parametrischen Testverfahrens
  - die Testvariable ist metrisch und normalverteilt
  - die Testvariable ist metrisch und nicht-normalverteilt, aber die Stichprobengrösse ist gross (parametrische Testverfahren sind bei grosser Stichprobe relativ robust)
- Gründe für die Anwendung eines nicht-parametrischen Testverfahrens
  - die Testvariable ist metrisch, aber nicht-normalverteilt
  - die Testvariable ist ordinal
  - die Testvariable ist nominal
  - der Sachverhalt von Interesse wird am Besten durch den Median repräsentiert (nicht durch den arithmetischen Mittelwert)
  - die Stichprobengrösse ist klein (bei kleiner Stichprobengrösse ist es schwierig zu entscheiden, ob die Testvariable normalverteilt ist)
  - Ausreisser (Ausreisser können das Ergebnis eines parametrischen Tests stark beeinflussen)

t-Test bei unabhängigen Stichproben

t-Test bei unabhängigen Stichproben

- der t-Test bei unabhängigen Stichproben ist ein parametrischer Test
- der t-Test bei unabhängigen Stichproben vergleicht die Mittelwerte zwei Stichproben und testet, ob es einen signifikanten Unterschied zwischen den Gruppen in der Grundgesamtheit gibt
- die Testvariable sollte metrisch und normalverteilt sein
- Wie ist der Test durchzuführen?
  - Schritt 1: Formuliere die Forschungsfrage, H_0, H_A und α
  - Schritt 2: Berechne die Teststatistik
  - Schritt 3: Prüfe auf Signifikanz

t-Test bei unabhängigen Stichproben

- Beispiel:
  - Schritt 1:
    - Gibt es einen Unterschied im Lernerfolg zwischen Klasse A und Klasse B?
    - H_0: $\mu_A - \mu_B = 0$
    - H_A: $\mu_A - \mu_B \neq 0$
    - $\alpha = 0.05$
  - Schritt 2:
    - um die Hypothese zu prüfen werden Stichprobendaten für die beiden Klassen erhoben:

	Punkte in der Prüfung (max. 60)									$\bar{x}$	s
Klasse A	53	49	60	34	45	49	56	44	57	49.67	8.00
Klasse B	45	54	57	25	50	45	42	40	39	44.11	9.44

    - Es gibt einen Unterschied in der Stichprobe, aber ist dieser Unterschied signifikant?

t-Test bei unabhängigen Stichproben

- die Teststatistik ist die Differenz der Mittelwerte geteilt durch die Standardabweichung der Stichprobenverteilung

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sigma_{\bar{x}_A, \bar{x}_B}} = 1.35$$

mit

$$\sigma_{\bar{x}_\mathrm{A}, \bar{x}_\mathrm{B}} = \sqrt{\left[\frac{(n_\mathrm{A} - 1)s_\mathrm{A}^2 + (n_\mathrm{B} - 1)s_\mathrm{B}^2}{n_\mathrm{A} + n_\mathrm{B} - 2}\right]\left[\frac{n_A + n_B}{n_A \times n_B}\right]} = 4.12$$

– Schritt 3

- vergleiche den Testwert mit dem kritischen Wert der t-Verteilung, $Fg = n_A + n_B - 2 = 16$
- der kritische Wert ist ± 2.120
- H_0 wird nicht verworfen, es gibt keinen Unterschied im Lernerfolg zwischen den beiden Gruppen

t-Test bei unabhängigen Stichproben

- Anmerkungen
  - soweit haben wir gleiche Varianz für beide Gruppen angenommen
  - die Teststatistik ist leicht anders, wenn ungleiche Varianzen vorliegen
  - d. h. wir müssen auf Gleichheit der Varianzen prüfen
    - wir können uns die Varianzen der Stichproben ansehen $var = s^2$
      - ❖ $var_A = s_A^2 = 8.00^2 = 64.00$
      - ❖ $var_B = s_B^2 = 9.44^2 = 89.11$
    - wir können mit Hilfe des Tests von Levene auf gleiche Varianzen testen
      - ❖ H_0: Die Varianzen in beiden Gruppen sind gleich.
      - ❖ H_A: Die Varianzen in beiden Gruppen sind nicht gleich.
      - ❖ $\alpha = 0.05$

Beispiel t-Test bei unabhängigen Stichproben

- im Folgenden führen wir den t-Test bei unabhängigen Stichproben mit Hilfe des R Commanders durch
  - der Datensatz ist growth.Rdata
  - der Datensatz ist simuliert und stammt aus dem Lehrbuch Kronthaler 2016
  - der Datensatz enthält Informationen zu neu gegründeten Unternehmen und den Gründern

  # lade den Datensatz “growth.Rdata”

Beispiel t-Test bei unabhängigen Stichproben

- Schritt 1:
  - Gibt es einen Unterschied im Gründungsalter zwischen Gründerinnen und Gründern?
  - H_0: $\mu_M - \mu_F = 0$
  - H_A: $\mu_M - \mu_F \neq 0$
  - $\alpha = 0.05$
- Schritt 2:
  - bevor wir den Test durchführen sollten wir die Daten vorbereiten und die deskriptive Statistiken ansehen
    - ❖ teile R mit, dass die Variable sex eine Faktorvariable ist

```
# nutze die Benutzeroberflache des R Commanders oder den folgenden Befehl
growth$f_sex<-as.factor(growth$sex)
```

Beispiel t-Test bei unabhängigen Stichproben

- prüfe die Testvariable auf Normalverteilung

```
# nutze die Benutzeroberflache des R Commanders oder die folgenden Befehle
attach(growth)
hist(age, scale="frequency", breaks="Sturges", col="deeppink")
qqPlot(age, dist="norm", id.method="y", id.n=2, labels=rownames(growth))
shapiro.test(age)
detach(growth)
```

Beispiel t-Test bei unabhängigen Stichproben

- prüfe ob die Gruppen gleiche Varianzen haben

  # nutze die Benutzeroberflache des R Commanders oder die folgenden Befehle

  ```
  numSummary(growth[,"age"], groups=growth$f_sex, statistics=c("mean", "sd"),   quantiles=c(0,.25,.5,.75,1))
  ```

  ```
  leveneTest(age ~ f_sex, data=growth, center="mean")
  ```

- betrachte die deskriptiven Statistiken, insbesondere den arithmetischen Mittelwert und die Standardabweichung

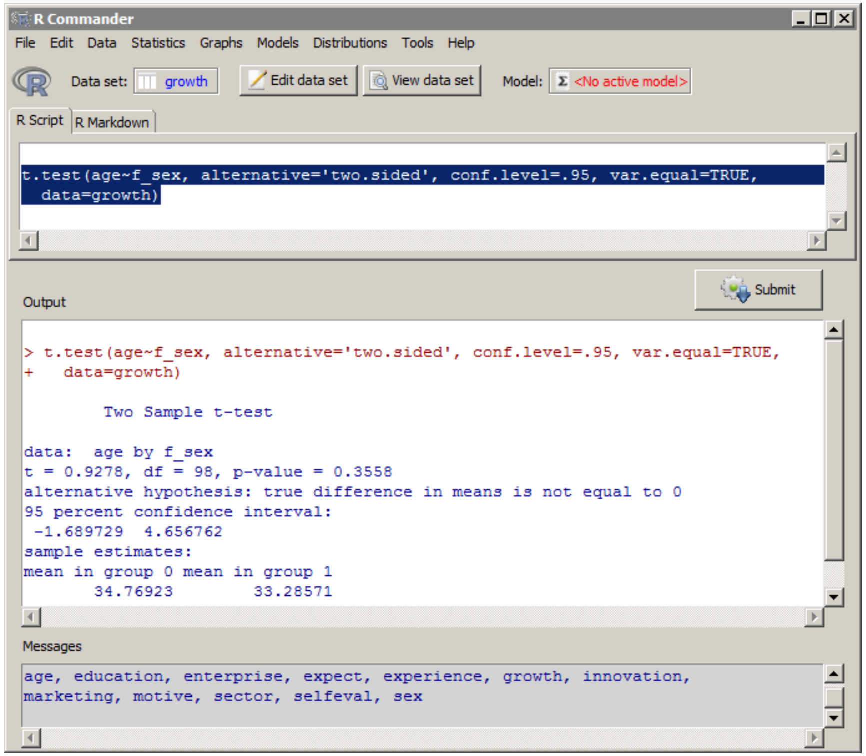
- führe den Test durch

  # nutze die Benutzeroberflache des R Commanders oder den folgenden Befehl

  ```
  t.test(age~f_sex, alternative='two.sided', conf.level=.95, var.equal=TRUE,
    data=growth)
  ```

Beispiel t-Test bei unabhängigen Stichproben

– Schritt 3:



Anwendungen

- Die Frage ist, ob neu gegründete Unternehmen des Dienstleistungssektors oder neu gegründete Unternehmen des Industriesektors höhere Umsatzwachstumsraten erzielen. Die folgenden Stichprobenwerte sind gegeben:

  $n_{ind} = 34,\ \bar{x}_{ind} = 6.35,\ sd_{ind} = 5.03$

  $n_{ser} = 66,\ \bar{x}_{ser} = 7.49,\ sd_{ser} = 5.77$

  Teste zum 5% Signifikanzniveau. Berechne den Test von Hand.

- Nutze den Datensatz growth.Rdata. Teste mit Hilfe des R Commanders, ob Unternehmen des Dienstleistungssektors oder Unternehmen des Industriesektors höhere Wachstumsraten erzielen. Teste zum Signifikanzniveau von 10%.

Wilcoxon-Rangsummen-Test

Wilcoxon-Rangsummen-Test

- der Wilcoxon-Rangsummen-Test (Wilcoxon–Mann–Whitney-Test) ist die nicht-parametrische Alternative zum t-Test bei unabhängigen Stichproben
- die Testvariable ist ordinal oder metrisch und nicht normalverteilt
- der Wilcoxon-Rangsummen-Test testet, ob die zwei Gruppen die selbe zentrale Tendenz haben
- Wie wird der Test durchgeführt?
  - Schritt 1: Formuliere die Forschungsfrage, H_0, H_A und α
  - Schritt 2: Berechne die Teststatistik
  - Schritt 3: Prüfe auf Signifikanz

Wilcoxon-Rangsummen-Test

- Beispiel:
  - Schritt 1:
    - Hat Klasse A bessere Noten als Klasse B?
    - H_0: $me_A - me_B \leq 0$
    - H_A: $me_A - me_B > 0$
    - $\alpha = 0.05$
  - Schritt 2:
    - um die Hypothese zu prüfen wird eine Stichprobe für beide Klassen gezogen und es werden die Noten erhoben:

	Noten										Median
Klasse A	5.5	4.3	5.1	3.6	5.8	2.5	3.9	5.9	6.0	4.4	4.75
Klasse B	3.7	5.7	4.8	4.6	3.5	4.0	5.6	4.1	4.5	5.3	4.50

    - Es gibt einen Unterschied in der Stichprobe, aber ist der Unterschied signifikant?

Wilcoxon-Rangsummen-Test

- um die Teststatistik zu berechnen sind zuerst die Rangwerte für beide Stichproben zu vergeben

Noten	2.5	3.5	3.6	3.7	3.9	4.0	4.1	4.3	4.4	4.5	4.6	4.8	5.1	5.3	5.5	5.6	5.7	5.8	5.9	6.0
Klassen	A	B	A	B	A	B	B	A	A	B	B	B	A	B	A	B	B	A	A	A
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

- wenn die zwei Klassen identische Verteilungen aufweisen, würden wir dieselbe Rangsumme erwarten
  - ❖ $W_A = 1 + 3 + 5 + 8 + 9 + 13 + 15 + 18 + 19 + 20 = 111$
  - ❖ $W_B = 2 + 4 + 6 + 7 + 10 + 11 + 12 + 14 + 16 + 17 = 99$
- der Wilcoxon-Rangsummen-Test vergleicht nicht die zwei Rangsummen, sondern eine der Rangsummen mit der durchschnittlichen Rangsumme
- in diesem Fall ist Rangsumme W der ersten Stichprobe die Teststatistik

Wilcoxon-Rangsummen-Test

- W hat den Mittelwert

  $$\mu_W = \frac{n_1(N+1)}{2}$$

  und die Standardabweichung

  $$\sigma_W = \sqrt{\frac{n_1 n_2 (N+1)}{12}}$$

- der Wilcoxon-Rangsummen-Test verwirft die Hypothese, dass die zwei Gruppen identische Verteilungen haben, wenn die Rangsumme deutlich vom Mittelwert abweicht

Wilcoxon-Rangsummen-Test

- in unserem Beispiel haben wir die folgenden Zahlen
  - $n_A = 10,\ n_B = 10$
  - $N = n_A + n_B = 20$
  - $\mu_W = \frac{n_A(N+1)}{2} = \frac{10(20+1)}{2} = 105$
  - $\sigma_W = \sqrt{\frac{n_1 n_2(N+1)}{12}} = \sqrt{\frac{10*10(20+1)}{12}} = \sqrt{175} = 13.23$
  - $W_A = 111$
- W_A weicht von μ_W ab, ist der Unterschied aber gross genug?
  - die Differenz ist 6
  - es ist weniger als eine halbe Standardabweichung ${}^{6}/_{13.23} = 0.45$
  - wir erwarten, dass der Unterschied nicht gross genug ist um H_0 zu verwerfen

Wilcoxon-Rangsummen-Test

– Schritt 3:
  - um zu entscheiden brauchen wir zusätzliche Informationen über die Testverteilung
  - die Testverteilung hängt von der Stichprobengrösse n_1 und n_2 ab
  - wenn $n_1 \geq 10$ und $n_2 \geq 10$ ist die Teststatistik W annähernd normalverteilt und wir können die Standardnormalverteilung nutzen
  - wenn nicht, nutzen wir die Wilcoxon Rangsummen-Verteilungstabelle (suche im Internet nach der Tabelle)
  - in unserem Fall ist die standardisierte Rangsummenstatistik

$$z = \frac{W - \mu_W}{\sigma_W} = \frac{111 - 105}{13.23} = 0.45$$

    - ❖ $z = 0.45$ hat eine Wahrscheinlichkeit von 32.64% (vergleiche die Standardnormalverteilungstabelle)
    - ❖ der Wert ist deutlich über dem spezifizierten Signifikanzniveau von 5%
    - ❖ wir verwerfen H_0 nicht

Beispiel Wilcoxon-Rangsummen-Test

- um den Wilcoxon-Rangsummen-Test durchzuführen nutzen wir den folgenden Beispieldatensatz
  - der Datensatz ist growth.Rdata
  - der Datensatz ist simuliert und stammt aus dem Lehrbuch Kronthaler 2016
  - der Datensatz enthält Informationen zum
    - Geschlecht der Unternehmensgründer/innen
    - zur Selbstbewertung der Gründer hinsichtlich ihrer Berufserfahrung

```
# lade den Datensatz "growth.Rdata"
```

Beispiel Wilcoxon-Rangsummen-Test

– Schritt 1:
  - Gibt es einen Unterschied in der Selbstbewertung zwischen Gründern und Gründerinnen?
  - H_0: $me_M - me_F = 0$
  - H_A: $me_M - me_F \neq 0$
  - $\alpha = 0.05$

– Schritt 2:
  - bevor wir den Test durchführen müssen wir die Daten vorbereiten
  - teile R mit, dass die Variable sex eine Faktorvariable ist

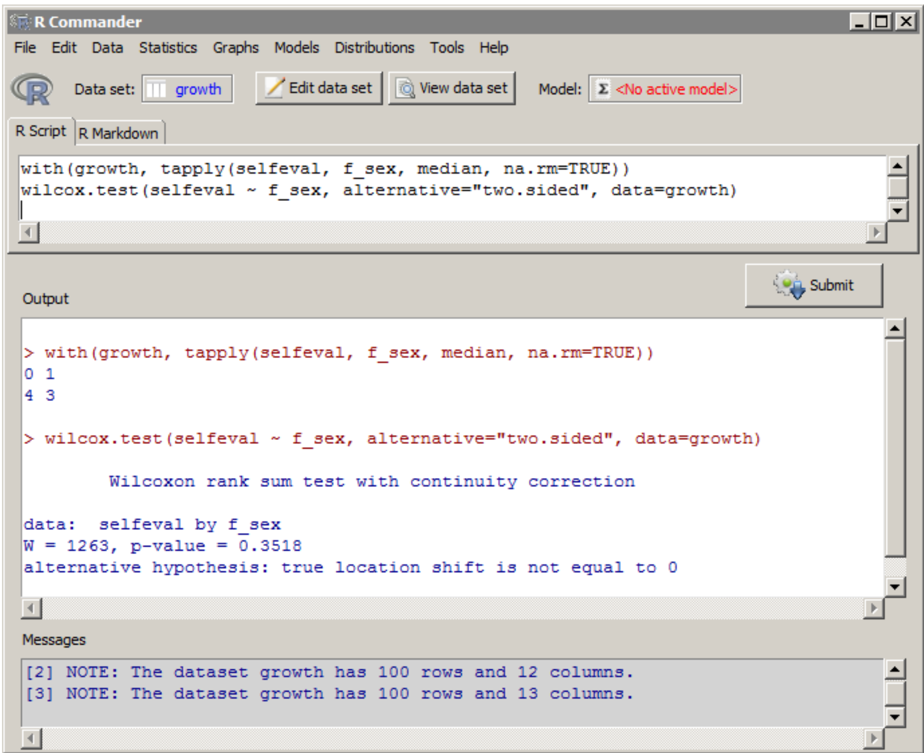
    ```
    # nutze die Benutzeroberflache des R Commanders oder den folgenden Befehl
    growth$f_sex<-as.factor(growth$sex)
    ```

  - führe den Test durch

    ```
    # nutze die Benutzeroberflache des R Commanders oder die folgenden Befehle
    with(growth, tapply(selfeval, f_sex, median, na.rm=TRUE))
    wilcox.test(selfeval ~ f_sex, alternative="two.sided", data=growth)
    ```

Beispiel Wilcoxon-Rangsummen-Test

– Schritt 3:



Anwendungen

- Nutze den Datensatz growth.Rdata. Teste mit der Hilfe des R Commanders ob es einen Unterschied in der Ausbildung zwischen Gründern und Gründerinnen gibt.
- Nutze den Datensatz gssft.RData (Lehrbuch von Norusis 2008) um zu prüfen, ob es einen Unterschied in der Arbeitszeit (Variable hrs1) zwischen Männern und Frauen gibt (Variable sex).
  - Führe den t-Test bei unabhängigen Stichproben durch.
  - Führe den Wilcoxon-Rangsummen-Test durch.
  - Vergleiche die Ergebnisse.
  - Welchen Test ziehen Sie vor? Warum?

t-Test bei abhängigen Stichproben

t-Test bei abhängigen Stichproben

- der t-Test bei abhängigen Stichproben ist ein parametrisches Testverfahren
- der t-Test bei abhängigen Stichproben testet die Wirksamkeit einer Massnahme
- er vergleicht den Stichprobenmittelwert vor und nach einer Massnahme, z. B.
  - Beeinflusst das Trinken von Bier die Fahrfähigkeit von Personen?
  - Hat ein Medikament einen Einfluss auf den Gesundheitszustand von Personen?
  - Hat eine Werbemassnahme einen Einfluss auf das Image eines Produktes bei Personen?
- die Testvariable ist metrisch und normalverteilt
- Wie wird der Test durchgeführt?
  - Schritt 1: Formuliere die Forschungsfrage, H_0, H_A, und α
  - Schritt 2: Kalkuliere die Teststatistik
  - Schritt 3: Prüfe auf Signifikanz

t-Test bei abhängigen Stichproben

- Beispiel:
  - Schritt 1:
    - Hilft die Trainingsmassnahme die Produktivität zu erhöhen?
    - H_0: $\mu_{pre} - \mu_{post} \geq 0$
    - H_A: $\mu_{pre} - \mu_{post} < 0$
    - $\alpha = 0.1$
  - Schritt 2:
    - um die Hypothese zu prüfen wurde eine Stichprobe vor und nach der Trainingsmassnahme beobachtet

	produzierte Einheiten pro Stunde									$\bar{x}$	s
Arbeiter	1	2	3	4	5	6	7	8	9		
pre	160	180	150	130	150	110	150	140	130	140.0	20.6
ex post	180	170	170	140	160	140	150	160	120	150.0	16.6

    - Es gibt einen Unterschied, aber ist der Unterschied signifikant?

t-Test bei abhängigen Stichproben

- die Teststatistik ist $t = \dfrac{\sum d_i}{\sqrt{\dfrac{n \sum d_i^2 - \left(\sum d_i\right)^2}{n-1}}}$

d_i ist die Differenz zwischen vor und nach der Massnahme

n ist die Zahl der Paarbeobachtungen

Arbeiter	pre	ex_post	di	di²
1	120	140	-20	400
2	180	170	10	100
3	150	170	-20	400
4	130	140	-10	100
5	150	160	-10	100
6	110	140	-30	900
7	150	150	0	0
8	140	160	-20	400
9	130	120	10	100
Summe			-90	2500

- $t = \dfrac{\sum d_i}{\sqrt{\dfrac{n \sum d_i^2 - \left(\sum d_i\right)^2}{n-1}}} = \dfrac{-90}{\sqrt{\dfrac{9*2500-(-90)^2}{8}}} = \dfrac{-90}{42.43} = -2.12$

t-Test bei abhängigen Stichproben

- Schritt 3:
  - vergleiche den Testwert mit dem kritischen Wert der t-Tabelle mit $Fg = n - 1 = 9 - 1 = 8$
  - der kritische Wert ist -1.397
  - H_0 wird verworfen, die Massnahme erhöht die Produktivität

Beispiel t-Test bei abhängigen Stichproben

- um den t-Test bei abhängigen Stichproben mit dem R Commander durchzuführen nutzen wir den folgenden Beispieldatensatz
  - training.RData
  - der Datensatz ist simuliert
  - er enthält das eben diskutierte Beispiel

  # lade den Datensatz “training.Rdata”

Beispiel t-Test bei abhängigen Stichproben

- Schritt 1:
  - Hilft die Trainingsmassnahme die Produktivität zu erhöhen?
  - H_0: $\mu_{pre} - \mu_{post} \geq 0$
  - H_A: $\mu_{pre} - \mu_{post} < 0$
  - $\alpha = 0.1$
- Schritt 2:
  - betrachte die deskriptiven Statistiken

```
# nutze die Benutzeroberfläche des R Commanders oder die folgenden Befehle
attach(training)
mean(pre)
mean(ex_post)
sd(pre)
sd(ex_post)
detach(training)
```

Beispiel t-Test bei abhängigen Stichproben

- prüfe/ teste, ob die Testvariable normalverteilt ist

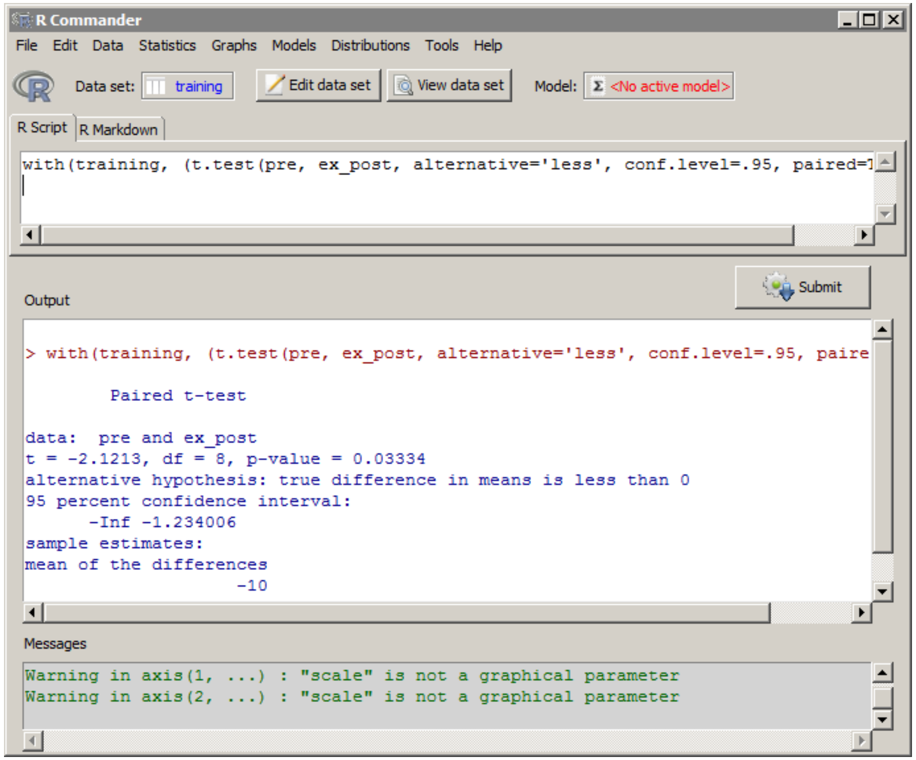
```
# nutze die Benutzeroberfläche des R Commanders oder die folgenden Befehle
attach(training)
hist(di, scale="frequency", breaks="Sturges", col="red")
qqPlot(di, dist="norm", id.method="y", id.n=2, labels=rownames(training))
shapiro.test(di)
detach(training)
```

- führe den Test durch

```
# nutze die Benutzeroberfläche des R Commanders oder den folgenden Befehl
with(training, (t.test(pre, ex_post, alternative='less', conf.level=.95,
paired=TRUE)))
```

Beispiel t-Test bei abhängigen Stichproben

– Schritt 3:



Anwendungen

- Wir sind daran interessiert, ob ein Schulungsprogramm für Buchhaltung die Effizienz bei der Durchführung der Buchhaltung erhöht. Wir beobachten 15 Teilnehmer und erhalten die folgenden Daten:

Teilnehmer	pre	ex post	di	d_i^2
1	17	19	-2	4
2	9	7	2	4
3	21	15	6	36
4	30	25	5	25
5	29	31	-2	4
6	33	33	0	0
7	5	4	1	1
8	18	10	8	64
9	28	30	-2	4
10	11	8	3	9
11	14	14	0	0
12	6	5	1	1
13	5	7	-2	4
14	12	11	1	1
15	13	15	-2	4

Teste zum 10%-Signifikanzniveau. Kalkuliere die Werte von Hand.

Anwendungen

- Wir sind daran interessiert, ob Energy-Drinks die Leistung von Athleten steigert. Nutze den simulierten Datensatz energy.Rdata. Teste mit Hilfe des R Commanders zum 1%-Signifikanzniveau.

Wilcoxon-Vorzeichen-Rang-Test

Wilcoxon-Vorzeichen-Rang-Test

- der Wilcoxon-Vorzeichen-Rang-Test ist die nicht-parametrische Alternative zum t-Test bei abhängigen Stichproben
- die Testvariable ist ordinal oder metrisch, aber nicht normalverteilt
- der Wilcoxon-Vorzeichen-Rang-Test testet, ob die zentrale Tendenz zweier gepaarter Stichproben gleich ist
- Wie wird der Test durchgeführt?
  - Schritt 1: Formuliere die Forschungsfrage, H_0, H_A und α
  - Schritt 2: Kalkuliere die Teststatistik
  - Schritt 3: Prüfe auf Signifikanz

Wilcoxon-Vorzeichen-Rang-Test

- Beispiel:
  - Irgendwo im Norden von Europa gibt es eine Schule, welche Schülern und Schülerinnen die Möglichkeit gibt, während der Schulpausen mit Fitnessgeräten zu trainieren. Die Schule macht dies, um das Übergewicht bei Kindern zu reduzieren.
  - Schritt 1:
    - Reduzieren aktive Schulpausen mit Fitnessgeräten das Gewicht von Kindern?
    - H_0: $me_{pre} - me_{post} \leq 0$
    - H_A: $me_{pre} - me_{post} > 0$
    - $\alpha = 0.05$
  - Schritt 2:
    - um die Hypothese zu testen führen wir ein Experiment durch und beobachten die Kinder vor und einige Zeit nach der Einführung aktiver Schulpausen

	Gewicht									$\bar{x}$	s
Kinder	1	2	3	4	5	6	7	8	9		
pre	30.0	35.0	33.0	28.0	60.0	37.0	32.0	29.0	31.0	35.0	9.8
ex post	27.0	30.0	34.0	28.0	62.0	33.0	28.5	28.5	26.5	33.1	11.2

Wilcoxon-Vorzeichen-Rang-Test

- die Daten sind metrisch, warum also ein nicht-parametrischer Test, z. B. weil
  - die Stichprobe klein ist
  - die Daten nicht normalverteilt sind
  - es Ausreisser gibt
- für die Teststatistik sind zuerst die paarweisen Differenzen zu berechnen

Gewicht

Kinder	1	2	3	4	5	6	7	8	9
pre	30.0	35.0	33.0	28.0	60.0	37.0	32.0	29.0	31.0
ex post	27.0	30.0	34.0	28.0	62.0	33.0	28.5	28.5	26.5
Differenz	3.0	5.0	-1.0	0.0	-2.0	4.0	3.5	0.5	4.5

- wir betrachten nur die Differenzen ungleich null und ordnen diese in aufsteigender Reihenfolge (das Vorzeichen wird nicht berücksichtigt)

Differenz	0.5	-1.0	-2.0	3.0	3.5	4.0	4.5	5.0
Rang	1	2	3	4	5	6	7	8

Wilcoxon-Vorzeichen-Rang-Test

- um die Teststatistik zu berechnen summieren wir die Ränge aller positiven und aller negativen Differenzen

  $T_+ = 1 + 4 + 5 + 6 + 7 + 8 = 31$

  $T_- = 2 + 3 = 5$

- zudem gibt es folgende Beziehung

  $$T_+ + T_- = \frac{n(n+1)}{2}$$

  mit n ist die Zahl der paarweisen Differenzen ungleich null

- die Beziehung führt zu folgender Gleichung

  $$\mu_T = \frac{n(n+1)}{4}$$

- die Teststatistik ist der grössere Wert

  $w = \mathrm{m}ax(T_+; T_-)$

- je weiter weg der Wert der Teststatistik vom durchschnittlichen Wert ist, desto eher verwerfen wir H_0

Wilcoxon-Vorzeichen-Rang-Test

- in unserem Fall haben wir die folgenden Werte

  $T_{+} = 31$

  $$\mu_T = \frac{n(n+1)}{4} = \frac{8(8+1)}{4} = 18$$

– Schritt 3:

- um die Signifikanz zu berechnen betrachten wir den Mittelwert μ_T und den Standardfehler SE_{μ_T}

  $$\mu_T = \frac{n(n+1)}{4} = \frac{8(8+1)}{4} = 18$$

  $$SE_{\mu_T} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{8(8+1)(16+1)}{24}} = \sqrt{51} = 7.14$$

- mit Hilfe der Teststatistik, dem Mittelwert und dem Standardfehler können wir den $z - Wert$ berechnen und auf Signifikanz prüfen

  $$z = \frac{T - \mu_T}{SE_{\mu_T}} = \frac{31 - 18}{7.14} = 1.82$$

Wilcoxon-Vorzeichen-Rang-Test

  - der kritische $z - Wert$ ist $z_{crit} = 1.645$ (vergleiche Standardnormalverteilung)
  - wir verwerfen H_0
- Bemerkung:
  - tatsächlich ist die Testverteilung die Standardnormalverteilung, wenn $n \geq 25$
  - wenn $n < 25$ ist die Testverteilung die w-Verteilung

Beispiel Wilcoxon-Vorzeichen-Rang-Test

- machen wir den Test mit dem R Commander
- der Datensatz ist
  - schoolbreak.RData
  - der Datensatz ist simuliert
  - der Datensatz enthält das eben diskutierte Beispiel

```
# lade den Datensatz “schoolbreak.Rdata”
```

Beispiel Wilcoxon-Vorzeichen-Rang-Test

- Schritt 1:
  - Reduzieren aktive Schulpausen mit Fitnessgeräten das Gewicht von Kindern?
  - H_0: $me_{pre} - me_{post} \leq 0$
  - H_A: $me_{pre} - me_{post} > 0$
  - $\alpha = 0.05$
- Schritt 2:
  - betrachte die deskriptiven Statistiken

```
# nutze die Benutzeroberfläche des R Commanders oder die folgenden Befehle
attach(schoolbreak)
mean(pre)
mean(ex_post)
sd(pre)
sd(ex_post)
detach(schoolbreak)
```

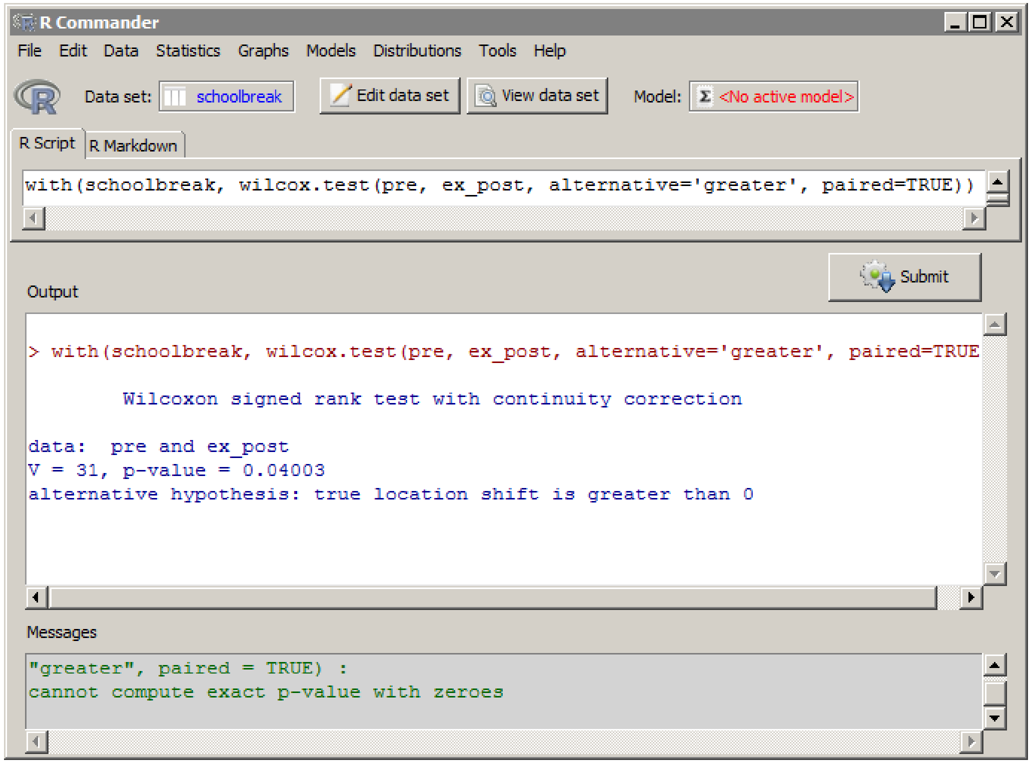
Beispiel Wilcoxon-Vorzeichen-Rang-Test

- prüfe auf Normalverteilung (wir lassen diesen Schritt aus, da wir aufgrund der kleinen Stichprobe und des Ausreissers entschieden haben den Wilcoxon-Vorzeichen-Rang-Test zu benutzen)
- führe den Test durch

```
# nutze die Benutzeroberfläche des R Commanders oder den folgenden Befehl
with(schoolbreak, wilcox.test(pre, ex_post, alternative='greater', paired=TRUE))
```

Beispiel Wilcoxon-Vorzeichen-Rang-Test

– Schritt 3:



Anwendungen

- Nutze den Datensatz siqss.RData (der Datensatz stammt aus dem Lehrbuch Norusis 2008). Wir sind daran interessiert, ob die Zahl der erhaltenen E-Mails (Variable emrec) grösser ist als die Zahl der gesendeten E-Mails (Variable emsent).
  - Führe den t-Test bei abhängigen Stichproben durch.
  - Führe den Wilcoxon-Vorzeichen-Rang-Test durch.
  - Vergleiche die Testergebnisse.
  - Welcher Test ist vorzuziehen? Warum?