

Aufgabe 1: Diskutieren Sie die folgenden Begriffe in Hinblick auf Tests: Verteilung, Lage- und Streu-Parameter, Hypothese, p-Wert

Aufgabe 2: Statistische Analysen zum *sleep* Datensatz

In einem klinischen Versuch sollte die Wirksamkeit eines Schlafmittels getestet werden. Dazu wurden von den Patienten, die in einer Klinik mit Schlafstörungen behandelt werden, zufällig 10 Patienten herausgesucht, die ein neuartiges Schlafmittel erhielten (Behandlungsgruppe) und 10 Patienten die eine wie Schlafmittel aussehende und schmeckende wirkungslose Pille erhielten (Kontrollgruppe). Um die Wirksamkeit zu bestimmen wurden jeweils zunächst die Schlafdauer in der Nacht vor der ersten Einnahme bestimmt. Vor der zweiten Nacht wurde den Patienten das Schlafmittel oder eine nach Schlafmittel aussehende wirkungslose Pille gegeben. Die Schlafdauer der zweiten Nacht wurde ebenfalls ermittelt.

Als Daten haben wir die Information, welches Mittel der Patient erhalten hat (2 = Schlafmittel, 1 = Wirkungsloses Placebo), und die Information, wieviel länger die Patienten in der zweiten Nacht geschlafen haben (in Stunden).

Laden Sie den Datensatz *sleep*!

```
data(sleep)
sleep
# Um ein angenehmeres Arbeiten zu ermöglichen, spalten wir den Datensatz einmal
# auf:
placebo <- sleep$extra[ sleep$group == 1 ]
mittel <- sleep$extra[ sleep$group == 2 ]
```

- (1) Mal angenommen, die Extra-Schlafdauer nach Einnahme des Placebo wäre normalverteilt. Welchen Test könnten Sie dann zur Überprüfung der Hypothese verwenden, dass die Verabreichung des Placebo keinen Einfluss auf die Schlafdauer hat (d.h. dass der Erwartungswert der Extra-Schlafdauer gleich Null ist)? Begründen Sie!

Tip: Welche Art von Verteilungsparameter wird bemüht? Welche Voraussetzungen müssen erfüllt sein? Handelt es sich um einen Ein-, Zwei- oder Mehrstichprobentest?

- (2) Nachfolgend finden Sie den Shapiro-Wilk-Test, der hier prüft, ob die Extra-Schlafzeit der Placebo-Gruppe normalverteilt ist. Interpretieren Sie die Ergebnisse und wenden Sie den Test auch auf die Schlafmittel-Gruppe an! Spricht etwas gegen die Anwendung des Tests?
-
-

```

# Shapiro-Wilk-Test
shapiro.test( placebo )
% # Ein-Stichproben-Kolmogorow-Smirnow-Test
% ks.test(sleep[,1][1:10], "pnorm", mean=0, sd=1)

```

Tip: Welche Voraussetzungen müssen erfüllt sein? Gibt es Besonderheiten im Datensatz wie Ausreißer oder Bindungen?

- (3) Einer der beteiligten Ärzte versucht Ihnen gerade weiszumachen, dass die Extra-Schlafzeit in der Placebo-Gruppe im Mittel nicht signifikant von Null abweicht. Überprüft haben will er dies mittels eines Ein-Stichproben-t-Tests. Ein Kollege von ihm kommt hinzu und postuliert, dass ein t-Test hier ja gar nicht in Frage kommt. Stattdessen schlägt er einen Vorzeichen-test vor, der überprüft, ob der Median mit der Null übereinstimmt. Was sagen Sie unter Berücksichtigung ihrer bisherigen Ergebnisse dazu? Wenden Sie den von Ihnen favorisierten Test sowohl auf die Placebo- als auch auf die Schlafmittel-Gruppe an und interpretieren Sie das Ergebnis!

```

# Ein-Stichproben-t-Test
t.test( placebo , mu=0)
# Vorzeichen-Test
binom.test( table( placebo <= 0 ) )

```

Tip: Was sind die Voraussetzungen der Tests?

- (4) Haben Sie in obigen Tests bereits nachgewiesen, dass das Schlafmittel besser ist, als das Placebo? Begründen Sie ihre Antwort bzw. geben Sie an, aus welchem Test oder welchen Tests dieses hervorgeht!

Tip: Warum vergibt man überhaupt ein Placebo?

- (5) Jemand möchte gerne wissen, ob die Verteilung der Extra-Schlafdauer beider Behandlungsgruppen identisch ist. Was können Sie unter Anwendung des folgenden Tests diesbezüglich aussagen ?

```
# Zwei-Stichproben-Kolmogorow-Smirnow-Test
ks.test( placebo, mittel )
```

- (6) Dürfen Sie den F-Test auf die Daten anwenden, um zu überprüfen, ob die Streuung in den Gruppen identisch ist? Begründen Sie ihre Antwort unter Berücksichtigung ihrer bisherigen Erkenntnisse! Führen sie den Test davon unabhängig durch und interpretieren Sie das Ergebnis!

```
# F-Test
var.test( placebo , mittel )
var.test( extra~group, data=sleep )      # geht genauso
```

- (7) Bei dem Gedanken an die gekerbten Boxplots, die Sie irgendwann einmal zu diesem Datensatz erstellt haben, überkommt Sie das Verlangen zu überprüfen, ob die Lage der Extra-Schlafdauer beider Gruppen übereinstimmt. Treffen Sie basierend auf den bereits erfolgten Tests eine geeignete Auswahl (unter den hier aufgeführten), begründen Sie diese und formulieren Sie das Ergebnis!

```
# Zwei-Stichproben-t-Test
t.test( placebo, mittel, var.equal = TRUE )
# Welchs t-Test
t.test( placebo, mittel )
# Wilcoxon-Rangsummen-Test
wilcox.test( placebo, mittel )
```

- (8) Was wird man mit diesem Test nachweisen wollen? Ist das Resultat dasselbe wie schon bei den gekerbten Boxplots?

Aufgabe 3: Statistische Analysen zum *OldFaithful* Datensatz

Laden Sie die Datei '*OldFaithful*' von der Website, wo sich auch die Übungsblätter befinden. Lesen Sie anschließend den Datensatz mittels des Befehls `read.table()` in *R* ein!

Der Geysir *Old Faithful* im Yellowstone-Nationalpark ist bekannt für die vergleichsweise geringe Variationsbreite der Wartezeit, die zwischen zwei Wasser-Eruptionen vergeht. Der Ihnen vorliegende Datensatz enthält zwei Messreihen an Wartezeiten (angegeben in Minuten), die in zwei verschiedenen Perioden ermittelt wurden.

(1) Welche Skala haben die Variablen?

(2) Handelt es sich um Stichproben und/oder Zufallsexperimente?

(3) Wofür könnten diese Daten repräsentativ sein ?

(4) Was ist an diesem Datensatz besonders problematisch in Hinblick auf die Repräsentativität ?

Tip: Lesen Sie sich die Beschreibung durch!

Für alle nachfolgenden Aufgaben wird angenommen, die Daten seien repräsentativ!!!

(5) Beschreiben Sie die Verteilung der Wartezeiten in der ersten Zeitperiode!

(6) Erschließen Sie mit geeigneten Graphiken, welche Eigenschaft(en) des Datensatzes für Tests problematisch sein könnten! Geben Sie diese zusammen mit der Graphik an, aus der diese ersichtlich ist.

- (7) Welche statistische Graphik und welcher Test würde sich eignen, um zu ersehen, ob die Wartezeiten in der ersten Zeitperiode normalverteilt sein könnten?

-
- (8) Überprüfen Sie, ob die Wartezeiten in den beiden Zeitperioden derselben Wahrscheinlichkeitsverteilung folgen könnten! Und?

-
- (9) Würden Sie, um zu testen, ob die Streuung in beiden Datensätzen identisch ist, den F-Test, oder den Fligner-Test wählen? Begründen Sie unter Verweis auf ein konkretes Test-Ergebnis (Führen Sie einen entsprechenden Test durch, falls Sie dieses müssen)!

```
# F-Test
var.test( wtime~period ,data=OldFaithful )
# Fligner-Test
fligner.test( wtime~period, data=OldFaithful )
```

- (10) Wählen Sie auf Grundlage ihrer bisherigen Ergebnisse einen Test, mit dem Sie überprüfen, ob die mittleren Wartezeiten in beiden Perioden übereinstimmen! Was ist das Ergebnis ?

-
- (11) Formulieren Sie das Ergebnis bitte für ihre Oma!

-
- (12) Formulieren Sie das Ergebnis bitte für ihren Statistikprofessor!
-
-