

Übung 04 - Datenanalyse und Statistik WS 2010/2011

Bei Großhandels-Verpackungsdienstleister XY gehen jeden Tag tausende ungekennzeichnete Kisten ein, die entweder mit Äpfeln, Birnen, Bananen, oder anderen Obstsorten gefüllt sind. Um mit Äpfeln gefüllte Kisten zu separieren greift sich in einem ersten Schritt ein Automat je 10 Objekte aus einer Kiste heraus und misst deren Länge und Durchmesser. Von Schlaubi Schlumpf weiß der Automat zudem, dass der typische oder auch durchschnittliche Apfel ein Längen zu Durchmesser Verhältnis von 4:3 hat und führt einen statistischen Test durch, anhand dessen er mit einer Wahrscheinlichkeit von 95 Prozent Kisten mit apfelförmigen Objekten richtig identifizieren kann.

Aufgabe 1: Was genau ist die Testgröße?

Aufgabe 2: Welcher Art von Parameter / Aspekt gehört diese Testgröße an?

Aufgabe 3: Formulieren Sie die Nullhypothese und die Alternativhypothese zu diesem Test!

Aufgabe 4: Wie lautet das Signifikanzniveau α dieses Tests?

Aufgabe 5: Was ist der p-Wert eines Tests?

Aufgabe 6: Wann wird die Alternativhypothese angenommen?

Aufgabe 7: Wie wird das Ergebnis des Tests üblicherweise für die folgenden Objekte lauten (Geben Sie an, wie sich der p-Wert zum Signifikanzniveau verhält, und formulieren Sie das Ergebnis in Prosa für den Professor)

Bananen

Äpfel

Apfelsinen

Statistische Analysen zum *sleep* Datensatz

In einem klinischen Versuch sollte die Wirksamkeit eines neuartigen Schlafmittels getestet werden. Dazu wurden unter den Patienten, die sich aufgrund von Schlafstörungen in einer Klinik behandeln ließen, zufällig 10 Personen ausgewählt, die das Schlafmittel erhielten (Behandlungsgruppe) sowie 10 weitere Patienten, die ein wie das Schlafmittel aussehendes und schmeckendes Placebo (Kontrollgruppe) erhielten.

Um die Wirksamkeit zu bestimmen, bildete man jeweils die Differenz aus der Schlafdauer in der ersten Nacht nach Einnahme des Mittels bzw. des Placebos und der Schlafdauer in der Nacht vor der Einnahme desselben (Extra-Schlafdauer).

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

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 ]
```

Aufgabe 8: 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 dieses Tests?

```
# Shapiro-Wilk-Test
shapiro.test( placebo )
```

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

Aufgabe 9: Einer der beteiligten Ärzte versucht Ihnen gerade weiszumachen, dass das Placebo nicht zu einer längeren Schlafdauer führt. 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, sondern ein Vorzeichentest verwendet werden muss. Was sagen Sie unter Berücksichtigung ihrer bisherigen Ergebnisse dazu (Formulieren und überprüfen Sie die jeweiligen Voraussetzungen der Tests!)?

Tip: Welche Art/Aspekt von Verteilungsparameter wird bemüht? Handelt es sich um einen Ein-, Zwei- oder Mehrstichprobentest? Bindungen, Ausreißer vorhanden? ???

Aufgabe 10: Wenden Sie den von Ihnen favorisierten Test sowohl auf die Placebo- als auch auf die Schlafmittel-Gruppe an und geben Sie die Ergebnisse für die Ärzte wieder! Haben Sie hiermit bereits nachgewiesen, dass das Schlafmittel besser wirkt als das Placebo?

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

Aufgabe 11: Jemand möchte gerne wissen, ob die Verteilung der Extra-Schlafdauer beider Behandlungsgruppen klar unterschiedlich ist. Was können Sie unter Anwendung des folgenden Tests diesbezüglich aussagen?

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

Aufgabe 12: 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 )      # Alternative Eingabe
```

Aufgabe 13: Bei dem Gedanken an die gekerbten Boxplots, die Sie irgendwann einmal zu diesem Datensatz erstellt haben, überkommt Sie das Verlangen noch einmal nachzuprüfen, ob die Verabreichung des Schlafmittels nicht doch im Mittel zu einer längeren Schlafdauer führt als bei der Verabreichung des Placebos. 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 , alternative="less" )
# Welchs t-Test
t.test( placebo, mittel , alternative="less" )
# Wilcoxon-Rangsummen-Test
wilcox.test( placebo, mittel , alternativ="less" )
```

Aufgabe 14: Was wird man mit diesem Test nachweisen wollen? Ist das Resultat dasselbe wie schon bei den gekerbten Boxplots?
