Übungsblätter zur VU

Statistik für Naturwissenschaftler

LV-Nr. 431.010

Ass.-Prof. Dr. Wolfgang Trutschnig
www.trutschnig.net
Fachbereich Mathematik
Paris Lodron Universität Salzburg
Hellbrunner Strasse 34
A-5020 Salzburg

1. Übung am 16. März 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe www.trutschnig.net/courses]

Downloaden und Installieren Sie R und RStudio, siehe http://cran.r-project.org/ und http://www.rstudio.com/products/rstudio/download/ und lösen Sie dann die nachfolgenden Aufgaben.

Übungsaufgabe 1 (R) Absolvieren Sie die ersten zwei Kapitel im (frei verfügbaren) R-online-training auf http://tryr.codeschool.com/.

Übungsaufgabe 2 Bestimmen Sie die Zahl x:

1. Die Partei A verliert x Prozent Ihrer Stimmen und fällt von 33.4 auf 25.6 Prozentpunkte.
2. Der Preis P_1 eines Autos erhöht sich zuerst um x Prozent auf P_2. Danach wird P_2 um 20% reduziert und man erhält den Ausgangspreis P_1.

Übungsaufgabe 3 (R) Lesen Sie den Artikel 'Frühpensionisten sterben früher' abrufbar unter dem Link http://sciencev1.orf.at/news/141680.html und versuchen Sie eine logische (!) Erklärung für die dort beschriebene Beobachtung zu finden.
Hinweis: Gehen Sie Zeile für Zeile durch die folgenden Programmzeilen und finden Sie heraus, was die Befehle bewirken (und worin der Zusammenhang mit der Aufgabe besteht):

```
x<-sample(50:100,size=20,replace=TRUE)
x
mean(x)
mean(x[x>60])
mean(x[x>70])
```

Übungsaufgabe 4 Insgesamt befinden sich im Mittelstufen Schultyp A und Schultyp B 20000 Kinder, wobei drei mal so viele Kinder eine Schule des Typs A besuchen. Im Schultyp A wird bei 2400 Kindern eine Leseschwäche festgestellt, im Schultyp B nur bei 900 Schülern. Welchen Typus würden Sie wählen ?

Übungsaufgabe 5 Drei Volksschüler aus gutem Hause haben jeweils 100 Euro Taschengeld erhalten und kaufen sich zu dritt ein ferngesteuertes Flugzeug um 300 Euro. Den Verkäufer packt während dem Ausdrucken der Rechnung/Quittung das schlechte Gewissen und er beschliesst, den Schülern 50 Euro zurückzugeben, woraufhin die drei beschließen, jeweils 10 Euro zu behalten und sich (für den Rest) gleich im Geschäft zusammen Süssigkeiten im Wert von 20 Euro zu kaufen. Jeder der drei hat also in Summe 90 Euro ausgegeben und die Süssigkeiten kosten 20 Euro, i.e. $3 \cdot 90 + 20 = 290 \neq 300$. Wohin sind die verbleibenden 10 Euro verschwunden ?

2. Übung am 06. April 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Ankreuzliste siehe www.trutschnig.net/courses,
mit Stern versehene Aufgaben sind freiwillig]

Übungsaufgabe 6 (R) Absolvieren Sie das dritte und das vierte Kapitel im (frei verfügbaren) R-online-training auf http://tryr.codeschool.com/.

Übungsaufgabe 7 (R) Kopieren Sie den R-Code 'R-codes_GEO_01.R'† ins RStudio, gehen Sie den Code Zeile für Zeile durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt. Wie kann Zeile 22 abgeändert werden, sodass das erstellte Histogramm eine feinere Unterteilung (mehr Intervalle) hat ? Verwenden Sie weiters den Befehl subset um ein (separates) Histogramm der abgehobenen Geldmengen im Jahr 2007, im Jahr 2008 und im Jahr 2009 zu erstellen. Verwenden Sie den Code in den Zeilen 35-37 um die ecdf der an Montagen abgehobenen Geldmengen (im gesamten Zeitraum) und der an Dienstagen abgehobenen Geldmengen zu plotten.

Übungsaufgabe 8

(1) Lesen Sie den Artikel 'Wärmster Winter aller Zeiten', abrufbar unter dem Link - Wärmster Winter.
   (a) Welche Aussage fällt ob ihrer Sinnlosigkeit auf ?
   (b) Auf Wetterseiten finden sich Aussagen wie 'Regenwahrscheinlichkeit in Salzburg morgen: 70%' - was bedeutet diese Aussage ?

(2*) Lesen Sie den Artikel 'Vegetarier vorurteilsfreier', abrufbar unter dem Link - Vegetarier vorurteilsfreier
Der Text klingt plausibel, trotzdem ist Vorsicht geboten - warum ?

Übungsaufgabe 9 (R) Gegeben ist die Stichprobe $x_1 = 0.5, x_2 = 1.5, x_3 = 2$. Skizzieren Sie die empirische Verteilungsfunktion $\hat{F}_3$ mit der Hand und überprüfen Sie dann Ihre Skizze mit Hilfe von R. An welcher Stelle springt die Funktion zum ersten Mal über 0.5 ?

Übungsaufgabe 10 @Astronautenartikel: Berechnen Sie (unter Verwendung von Schulwissen) die Wahrscheinlichkeit dafür, dass keiner der 100 Astronauten an Krebs erkrankt.
Hinweis: Falls die Schulkenntnisse nicht mehr voll aktiv sind, tasten Sie sich mit Hilfe des wiederholten Münzwurfs heran - Bsp: Die Wahrscheinlichkeit dafür, beim dreimaligen Werfen einer Münze kein Mal Kopf zu bekommen ist $(\frac{1}{2})^3$.

Frohe Ostern und erholsame Feiertage !

† wie alle Datensätze der Vorlesung abrufbar auf www.trutschnig.net/courses

3. Übung am 13. April 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe www.trutschnig.net/courses]

Übungsaufgabe 11 (R) Absolvieren Sie das fünfte und das sechste Kapitel im (frei verfügbaren) R-online-training auf http://tryr.codeschool.com/.

Übungsaufgabe 12 Lesen Sie den Artikel 'Den 62 Reichsten gehrt die halbe Welt', abrufbar unter dem Link - Artikel. Wo liegt der Denkfehler ?

Übungsaufgabe 13 Gegeben ist die Stichprobe $x_1 = 0.5, x_2 = 0.5, x_3 = 1, x_4 = 2$. Skizzieren Sie die empirische Verteilungsfunktion $\hat{F}_4$ und berechnen Sie das Quantil z_p für jedes $p \in (0, 1]$.

Übungsaufgabe 14 (R) Kopieren Sie den R-Code 'R-codes_GEO_02.R' ins RStudio, gehen Sie den Code Zeile für Zeile durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt.

Übungsaufgabe 15 (R) Verwenden Sie die Zeilen 36-40 im R-Code 'R-codes_GEO_02.R' um den RTR-Datensatz zu laden und erstellen Sie einen boxplot der Variable 'rtr_speed_dl' (Download-Geschwindigkeit) (i) für jedes Monat und (ii) für jedes Monat und jeden operator. Welchen operator würden Sie wählen, wenn alle den selben Tarif anbieten ?

Übungsaufgabe 16 (R) Der RTR-Datensatz enthält neben den gps-Koordinaten auch eine Spalte namens 'iso_adm2', die den (politischen) Bezirk, in dem die Messung stattgefunden hat, angibt. Folgende iso_adm2-Werte entsprechen den Bezirken Salzburgs:

iso_adm2	Bezirk	Bundesland
at0501	Salzburg (Stadt)	Salzburg
at0502	Hallein	Salzburg
at0503	Salzburg Umgebung	Salzburg
at0504	Sankt Johann im Pongau	Salzburg
at0505	Tamsweg	Salzburg
at0506	Zell am See	Salzburg

Beantworten Sie mit Hilfe von Boxplots die Frage, welcher operator im Februar 2014 den besten download-speed (Variable 'rtr_speed_dl') in Hallein hatte. Vergleichen Sie die erstellten boxplots mit den boxplots für Hallein im Februar 2015 - was ist zu beobachten ?
Hinweis: Verwenden Sie 'subset' um auf Hallein und die entsprechenden Monate zu filtern

4. Übung am 20. April 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe www.trutschnig.net/courses]

Übungsaufgabe 17 Gegeben ist die Stichprobe $x_1 = 0, x_2 = 0.5, x_3 = 0, x_4 = 3, x_5 = 1$. Skizzieren Sie die empirische Verteilungsfunktion $\hat{F}_5$ (händisch oder mit R) und berechnen Sie unteres und oberes Quartil, sowie den Median der Stichprobe.

Übungsaufgabe 18 (R) Die Datensätze 'A1.RData' und 'A2.RData' enthalten Lebensdauern von (lt. Werbung gleichwertigen) Bohrköpfen zweier verschiedener Hersteller $A1$ und $A2$. Für welchen der beiden Anbieter entscheiden Sie sich, falls beide Hersteller zum selben Preis anbieten ?
Hinweis: Laden Sie die Datensätze wie gewohnt mittels
dir <− url("http://www.trutschnig.net/A1.RData"); load(dir)
und arbeiten Sie mit boxplots und/oder ecdfs.

Übungsaufgabe 19 (R) Kopieren Sie den R-Code 'R-codes_GEO_04.R' ins RStudio, gehen Sie den Code Zeile für Zeile durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt.

Übungsaufgabe 20 Ein Würfel wird zwei Mal geworfen. (a) Berechnen Sie die Wahrscheinlichkeit dafür, dass die Augensumme aus erstem und zweitem Wurf fünf ist. Geben Sie dazu Ω und E explizit an und berechnen Sie $\mathbb{P}(E) = \frac{\#E}{\#\Omega}$.
(b) Wiederholen Sie (a) für Augensumme neun sowie für Augensumme zehn.
(c) Berechnen Sie die Wahrscheinlichkeit dafür, dass die Augensumme aus erstem und zweitem Wurf höchstens acht ist.

Übungsaufgabe 21 (R) Überprüfen Sie die Ergebnisse von Aufgabe 20 mit Hilfe von Simulationen in R, i.e. wiederholen Sie das Experiment 100.000 mal in R und berechnen Sie, in wie vielen Fällen die Augensumme fünf (neun, zehn, höchstens acht) ist.
Hinweis: Orientieren Sie sich an 'R-codes_GEO_04.R' und verwenden Sie 'ifelse'

Übungsaufgabe 22 (*R) In einer Packung mit sechs Riesen-Schoko-Ostereiern mit verschiedenen Füllungen (die von außen aber nicht zu erkennen sind) gibt es genau zwei Eier mit Ihrer Lieblingsfüllung. Sie wählen zufällig zwei der Ostereier - wie groß ist die Wahrscheinlichkeit dafür, dass Ihre Lieblingsfüllung nicht dabei ist ? Beantworten Sie die Frage entweder rechnerisch oder mit Hilfe von Simulationen mit R.

5. Übung am 27. April 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe www.trutschnig.net/courses]

Übungsaufgabe 23 (R) Kopieren Sie den R-Code 'R-codes_GEO_05.R' ins RStudio, gehen Sie die Zeilen 1-28 Schritt für Schritt durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt. Ändern Sie die Zeilen 11-28 so ab, dass 40% den Kandidaten wählen.

Übungsaufgabe 24 Ein hochmotivierter Amateurfahrer nimmt an einem Radmarathon teil, bei dem zwei Runden zu jeweils 80km (und 550 Höhenmeter) zu bewältigen sind. Er möchte den Marathon mit einer Durchschnittsgeschwindigkeit von 28 km/h absolvieren, schont sich die erste Runde, und absolviert selbige nur mit einer Durchschnittsgeschwindigkeit von 14 km/h. Mit welcher Durchschnittsgeschwindigkeit muss er die zweite Runde fahren, um insgesamt noch auf die angestrebten 28 km/h zu kommen?

Übungsaufgabe 25 Lesen Sie den Artikel 'Generationen-Problem! 98 Prozent weniger ÖSV-Siege', abrufbar unter dem Link - Artikel. Wo liegt dieses Mal (wieder) der Denkfehler?

Übungsaufgabe 26 Jeder Flugzeugmotor fällt unabhängig von den anderen während eines Fluges mit Wahrscheinlichkeit $p = 0.1$ aus. Voraussetzung für einen erfolgreichen Flug ist, dass mindestens die Hälfte der Motoren nicht versagt. Ist eine zweimotorige oder eine viermotorige Maschine vorzuziehen?

Übungsaufgabe 27 Eine wildcat Ölbohrung hat (unter bestimmten Rahmenbedingungen) eine Erfolgswahrscheinlichkeit von 12%. Wie oft muss gebohrt werden, um mit mindestens 50%-tiger Wahrscheinlichkeit mindesten ein mal erfolgreich zu sein ? Wie oft muss gebohrt werden, um mit mindestens 90%-tiger Wahrscheinlichkeit mindesten ein mal erfolgreich zu sein ? Wie oft muss gebohrt werden, um mit mindestens 99.9%-tiger Wahrscheinlichkeit mindesten ein mal erfolgreich zu sein ?
Hinweis: $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0)$, letzteres lässt sich für binomialverteiltes X leicht berechnen.

Übungsaufgabe 28 @Astronautenartikel: Berechnen Sie die Wahrscheinlichkeit dafür, dass mindestens einer der 100 Astronauten an Krebs erkrankt. Wie hoch ist die Wahrscheinlichkeit dafür, dass mindestens zwei erkranken ?

Übungsaufgabe 29 (*R) Kopieren Sie den R-Code 'R-codes_GEO_05.R' ins RStudio und finden Sie heraus, was die Zeilen 30-43 bewirken.

6. Übung am 04. Mai 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe www.trutschnig.net/courses]

Übungsaufgabe 30 (R) Kopieren Sie den R-Code 'R-codes_GEO_06.R' ins RStudio, gehen Sie die Zeilen 1-47 Schritt für Schritt durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt.

Übungsaufgabe 31 (R) Kopieren Sie den R-Code 'R-codes_GEO_06.R' ins RStudio, gehen Sie die Zeilen 49-93 Schritt für Schritt durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt.

Übungsaufgabe 32 Angenommen $X \sim Pois(\lambda)$. Bestimmen Sie jenen Wert von λ, sodass gilt

(a) $\mathbb{P}(X = 0) = \mathbb{P}(X = 1)$

(b) $\mathbb{P}(X \leq 1) = \mathbb{P}(X = 2)$

(c) $\mathbb{P}(X = 0) = \mathbb{P}(X \geq 1)$

Übungsaufgabe 33 (R) Sei $X \sim Pois(\lambda)$, $n = 10000$, und $x_1, \ldots, x_n$ eine Stichprobe von X. Überprüfen Sie durch Modifikation der Zeilen 36-46 im R-Code 'R-codes_GEO_06.R', ob auch die Stichprobenvarianz von $x_1, \ldots, x_n$ ein guter Schätzer von λ ist und experimentieren Sie dabei mit verschiedenen Werten von λ. Welcher Schätzer ist der bessere - die Stichprobenvarianz oder das Stichprobenmittel?

Übungsaufgabe 34 (R) Laden Sie 'Erdbeben.RData' mit Hilfe von 'R-codes_GEO_06.R' und finden Sie heraus, (i) wo die zehn stärksten Erdbeben stattgefunden haben, (ii) in welchem Jahr am meisten Erdbeben der Stärke ≥ 8 verzeichnet sind, (iii) in welchem Monat (Jänner, Februar, etc.) insgesamt am meisten und am wenigsten Erdbeben stattgefunden haben, und (iv) ob im Datensatz Tage enthalten sind, an denen es mehr als ein Erdbeben gab.

Übungsaufgabe 35 Aus den letzten Jahren weiss man, dass auf der Südosttangente in Wien Montags zwischen 07:00 und 10:00 im Mittel 1,25 Unfälle passieren, bei denen die Rettung ausrücken muss. Berechnen Sie die Wahrscheinlichkeit dafür, dass am kommenden Montag zwischen 07:00 und 10:00 zwei Mal ein Rettungswagen ausrücken muss unter der Annahme, dass die Anzahl der Unfälle X, bei denen ein Rettungswagen erforderlich ist, Poisson-verteilt mit Parameter λ ist.

7. Übung am 11. Mai 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe www.trutschnig.net/courses]

Übungsaufgabe 36 Lesen Sie den Artikel 'Dreimal so viele Österreicher verurteilt wie Ausländer', abrufbar unter dem Link - Artikel. Wo liegt dieses Mal (wieder) der Denkfehler?

Übungsaufgabe 37 Angenommen, die Fehler X und Y im Hangrutschbeispiel aus der Vorlesung erfüllen nun $\mathbb{P}(X = -40) = \mathbb{P}(Y = -40) = \mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = \mathbb{P}(X = 40) = \mathbb{P}(Y = 40) = \frac{1}{3}$. Gilt auch in diesem Fall $\mathbb{E}(|100+Y-X|) > 100$? Beantworten Sie die Frage indem Sie $\mathbb{E}(|100 + Y - X|)$ wie in der Vorlesung berechnen.

Übungsaufgabe 38 (R) Angenommen, die Fehler X und Y im Hangrutschbeispiel aus der Vorlesung erfüllen nun $\mathbb{P}(X = -100) = \mathbb{P}(Y = -100) = \mathbb{P}(X = -50) = \mathbb{P}(Y = -50) = \mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = \mathbb{P}(X = 50) = \mathbb{P}(Y = 50) = \mathbb{P}(X = 100) = \mathbb{P}(Y = 100) = \frac{1}{5}$. Gilt auch in diesem Fall $\mathbb{E}(|100 + Y - X|) > 100$? Beantworten Sie die Fragen in dem Sie $\mathbb{E}(|100 + Y - X|)$ berechnen.
Überprüfen Sie das rechnerisch erhaltene Resultat mittels Simulation in R (es reicht, 'R-codes_GEO_07.R', entsprechend anzupassen.)

Übungsaufgabe 39 Ein Spiel mit zufälligem Ausgang X heißt *fair*, wenn $\mathbb{E}(X) = 0$ gilt. Wir betrachten folgendes Spiel: Sie würfeln; wenn eine Sechs kommt, erhalten Sie $X = 5$ Euro, wenn keine Sechs kommt verlieren Sie einen Euro, d.h. $X = -1$. Wir erhalten also $\mathbb{P}(X = 5) = \frac{1}{6}$ und $\mathbb{P}(X = -1) = \frac{5}{6}$ und damit $\mathbb{E}(X) = 5 \cdot \frac{1}{6} + (-1) \cdot \frac{5}{6} = 0$. Das Spiel ist also fair.

(a) Sie würfeln und erhalten 2 Euro falls das Ergebnis eine gerade Zahl ist; Sie verlieren 2 Euro falls das Ergebnis eine ungerade Zahl ist. Ist das Spiel fair ?

(b) In einer Urne liegen 3 weiße und 1 schwarze Kugel. Sie ziehen zufällig eine Kugel und gewinnen 6 Euro falls die gezogene Kugel schwarz ist, und verlieren x Euro falls sie weiß ist. Für welches x ist das Spiel fair ?

(c) Vier Münzen werden geworfen. Sie gewinnen 5 Euro wenn höchstens 1 Mal Zahl kommt; sie verlieren 2 Euro, wenn mehr als ein Mal Zahl kommt.

Übungsaufgabe 40 (R) Wie groß ist die Wahrscheinlichkeit, dass in einer Lehrveranstaltung mit 25 Teilnehmern mindestens zwei am gleichen Tag Geburtstag haben (unter der Annahme, dass alle Geburtstage gleich wahrscheinlich sind)? Beantworten Sie die Frage entweder rechnerisch oder (einfacher) mit Hilfe von Simulationen in R.
Hinweis: Für letzteres können Sie zum Beispiel x<−sample(1:365,25,replace=TRUE) und max(table(x)) verwenden.

8. Übung am 18. Mai 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe www.trutschnig.net/courses]

Übungsaufgabe 41 Werfen Sie einen Blick auf die Grafik am Ende des Artikels 'Kreuzweh, Allergien, Verdauung, Stress: Das plagt die Österreicher', abrufbar unter dem Link - Artikel. und lesen Sie dann erst den Artikel. Welche Information deutet die Grafik (auf den ersten Blick) an und welche essentielle Info fehlt komplett?

Übungsaufgabe 42 (R) Kopieren Sie den R-Code 'R-codes_GEO_08.R' ins RStudio, gehen Sie den Code Zeile für Zeile durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt. Verwenden Sie weiters die Funktion f im Code um $\mathbb{E}(X)$ und $\mathbb{V}(X)$ für die folgende Zufallsvariable X zu verwenden:

Wert von X:	1	2	3	4	5	6	7	8	9	10	11	12
Wahrscheinlichkeit:	0.02	0.03	0.13	0.15	0.13	0.18	0.14	0.10	0.06	0.02	0.03	0.01

Übungsaufgabe 43 (R) @Hangrutschbeispiel: Verwenden Sie die Funktion f am Ende von 'R-codes_GEO_08.R' um für $a = 0.25$ die Verzerrung $\mathbb{E}(|100 + Y - X|) - 100$ zu berechnen, unter der Annahme (siehe slides page 25), dass die Fehler X und Y (unabhängig voneinander) nur drei Werte annehmen können:

$$\begin{aligned} \mathbb{P}(X = 0) &= \mathbb{P}(Y = 0) = a \in [0, 1] \\ \mathbb{P}(X = 70) &= \mathbb{P}(Y = 70) = \mathbb{P}(X = -70) = \mathbb{P}(Y = -70) = \tfrac{1-a}{2} \end{aligned} \tag{1}$$

Wiederholen Sie die Berechnung für $a = 0, a = 0.1, a = 0.2,$ und $a = 0.3$ - für welches a ist die Verzerrung maximal?
Hinweis: Es reicht, sich zu überlegen, welche Werte $|100 + Y - X|$ mit welchen Wahrscheinlichkeiten annimmt, und dann f zu verwenden.

Übungsaufgabe 44 Berechnen Sie für die Zufallsvariable X aus dem vorigen Beispiel $\mathbb{V}(X)$ für beliebiges $a \in [0, 1]$. Für welches a ist $\mathbb{V}(X)$ maximal?

Übungsaufgabe 45 (R) Gehen Sie auf Link-homepage, speichern Sie den Datensatz 'EISENERZ.TXT' auf Ihrem Rechner und lesen Sie den Datensatz in R ein. Der Datensatz enthält 'time series of the composition and tonnage of the output from the benefidation plant for a 21-week period in 1996.' Plotten Sie 'hour' versus jede einzelne der anderen Variablen (insgesamt also 7 plots). Versuchen Sie weiters, die durchnummerierten Stunden in Tagesstunden (Werte 1 bis 24) umzurechnen und dann mit Hilfe von boxplots herauszufinden, ob die Fördermengen zu jeder Tageszeit ähnlich waren.

9. Übung am 25. Mai 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe http://www.trutschnig.net/courses]

Übungsaufgabe 46 (R) Kopieren Sie den R-Code 'R-codes_GEO_09.R' ins RStudio, gehen Sie im Code die Zeilen 1-30 durch und finden Sie heraus, die Funktion 'hangrutsch' genau macht.

Übungsaufgabe 47 Die Funktion f sei gegeben durch

$$f(x) = \begin{cases} a\,x^2 & \text{falls } x \in [-1, 1], \\ 0 & \text{sonst.} \end{cases}$$

Bestimmen Sie a so, dass f eine Wahrscheinlichkeitsdichte ist und skizzieren Sie f (zu Fuß oder mit R). Berechnen Sie weiters $\mathbb{E}(X)$ und $\mathbb{V}(X)$ für eine Zufallsvariable X mit Dichte f.

Übungsaufgabe 48 Die Funktion f sei gegeben durch

$$f(x) = \begin{cases} a + x & \text{falls } x \in [-1, 0], \\ a - x & \text{falls } x \in [0, 1], \\ 0 & \text{sonst.} \end{cases}$$

Bestimmen Sie a so, dass f eine Wahrscheinlichkeitsdichte ist und skizzieren Sie f (zu Fuß oder mit R). Berechnen Sie weiters $\mathbb{E}(X)$ und $\mathbb{V}(X)$ für eine Zufallsvariable X mit Dichte f.

Übungsaufgabe 49 (R) Finden Sie heraus (wikipedia, etc.), was eine exponential-verteilte Zufallsvariable X ist, welche Werte sie annehmen kann, wie $\mathbb{E}(X)$ von λ abhängt (muss nicht nachgerechnet werden), wie man für den Fall, dass X exponential-verteilt mit Parameter $\lambda = 2$ ist, $\mathbb{P}(X \leq 1)$ berechnen kann, und wie Stichproben von X in R generiert werden können.

Übungsaufgabe 50 (R) Kopieren Sie den R-Code 'R-codes_GEO_09.R' ins RStudio, gehen Sie im Code die Zeilen 31-42 durch und finden Sie heraus, was hier gemacht wird. Wiederholen Sie den Berechungen für anderen Funktionen f. Was ist zu beobachten?

10. Übung am 01. Juni 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe http://www.trutschnig.net/courses]

Übungsaufgabe 51 (R) Kopieren Sie den R-Code 'R-codes_GEO_10.R' ins RStudio, gehen Sie die Zeilen 1-38 durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt. Was passiert wenn für die Zeilen 10-21 n sehr klein (Bsp: $n = 10$) gewählt wird?

Übungsaufgabe 52 (R) Sie X normalverteilt mit Parametern $\mu = 0$ und $\sigma = 1$. Finden Sie heraus, wie Sie mit Hilfe der Funktion pnorm † $\mathbb{P}(X \leq 0.5)$ sowie $\mathbb{P}(X \leq 2)$ und $\mathbb{P}(X \leq -2)$ berechnen können. Berechnen Sie weiters mit Hilfe von pnorm $\mathbb{P}(X \in [2,3])$ sowie $\mathbb{P}(|X| \leq 1)$.
Hinweis: $\mathbb{P}(X \in [2,3]) = \mathbb{P}(X \leq 3) - \mathbb{P}(X \leq 2)$

Übungsaufgabe 53 (R) (a) Laden Sie mit Hilfe der letzten Zeilen in 'R-codes_GEO_10.R' den Datensatz 'data_norm' einer normalverteilten Zufallsvariable $X \sim \mathcal{N}(\mu, \sigma^2)$. Weder μ noch σ^2 sind bekannt - wie würden Sie μ und σ^2 aus den Daten schätzen ?
(b) Simulieren Sie für von Ihnen gewählte Parameter *mu* und *sigma* Daten mit Hilfe von rnorm(10000,mu,sigma) und überprüfen Sie, ob die in (a) vorgeschlagenen Schätzer nahe bei den echten Werten liegen.
Hinweis: Arbeiten Sie mit $\overline{x}_n$ und s_n^2.

Übungsaufgabe 54 (R) (a) Laden Sie mit Hilfe der letzten Zeilen in 'R-codes_GEO_10.R' den Datensatz 'data_unif' einer stetig gleichverteilten Zufallsvariable $X \sim \mathcal{U}(a, b)$. Weder a noch b sind bekannt - wie würden Sie a und b aus den Daten schätzen ?
(b) Simulieren Sie für von Ihnen gewählte Parameter a und b Daten mit Hilfe des Befehls runif(10000,a,b) und überprüfen Sie, ob die in (a) vorgeschlagenen Schätzer nahe bei den echten Werten liegen.

Übungsaufgabe 55 (R) Vier Seiten eines Würfels werden mit der Zahl 1 beschriftet, die verbleibenden zwei Seiten mit der Zahl 0. Der Würfel wird $n = 1000$ mal geworfen. Berechnen Sie mit Hilfe des ZGWS die Wahrscheinlichkeit dafür, dass mindestens 650 Mal 1 gewürfelt wird.
Hinweis: Gehen Sie analog zum in den Folien gerechneten Schulbeispiel vor, i.e. berechnen Sie $\mathbb{E}(X)$ und $\mathbb{V}(X)$ für einen Wurf und hanteln Sie sich dann bis zur Standardnormalverteilung durch.

---

† ignorieren Sie lower.tail = TRUE und log.p = FALSE

11. Übung am 08. Juni 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe http://www.trutschnig.net/courses]

Übungsaufgabe 56 (R) Kopieren Sie den R-Code 'R-codes_GEO_11.R' ins RStudio, gehen Sie den Code Zeile für Zeile genau durch und finden Sie heraus, was jeder der verwendeten Befehle bewirkt.

Übungsaufgabe 57 (R) Finden Sie mit Hilfe von 'R-codes_GEO_11.R' heraus, ob die Schätzung der zugrundeliegenden Parameter (a und b im linearen Fall, b, a_1, a_2 im quadratischen Fall) nur 'funktioniert', wenn die simulierten Fehler normalverteilt sind, oder ob die einzige einschränkende Eigenschaft für die Fehler Erwartungswert gleich null ist.
Experimentieren Sie dafür zum Beispiel mit

1. mit Fehlern, die stetig gleichverteilt auf $[-1, 1]$ sind; zum Beispiel error<−runif(n,−1,1)
2. mit diskreten Fehlern; zum Beispiel error<−sample(c(−1,0,1),n,replace=TRUE)

Übungsaufgabe 58 (R) Laden Sie den Datensatz reg_data.RData mit Hilfe von dir <− url("http://www.trutschnig.net/reg_data.RData"); load(dir); A<−reg_data, plotten Sie die Daten, passen Sie (i) ein lineares Modell, (ii) ein quadratisches Modell, (iii) ein kubisches Modell und (iv) ein Modell vierten Grades an, und berechnen Sie für jedes der vier Modelle das Bestimmheitsmaß R^2. Für welches der vier Modelle würden Sie sich entscheiden? Prognostizieren Sie weiters (separat für jedes Modell) den Wert an der Stelle $x = 1.1$.
Hinweis: Verwenden Sie (als bequeme Alternative zu Ausdrücken der Form I(x^2) etc.) model<−lm(y ~ poly(x,2,raw=TRUE)) mit entsprechendem Grad des Polynoms.

Übungsaufgabe 59 (R) Der Datensatz brainhead.txt † auf http://www.trutschnig.net enthält 'Brain weight (grams) and head size (cm3) for 237 adults'. Passen Sie eine lineare Regression mit 'cm3' als erklärender und 'weight' als abhängiger Variable an und berechnen Sie das entsprechende R^2. Berechnen Sie weiters die grössten 10 Residuen und finden Sie heraus, wie viele Frauen und wie viele Männer in diesen 'top-ten' sind.

Übungsaufgabe 60 (R*) Der Datensatz beer.txt auf http://www.trutschnig.net enthält die Schaumhöhe (h) verschiedener Biere zu verschiedenen Zeiten (t). Es ist bekannt, dass die Schaumhöhe dem folgenden mathematischen Modell mit Parametern a und b folgt

$$h = ae^{-bt}$$

Passen Sie das Modell an die Daten an und schätzen Sie damit a und b.
Hinweis: Durch Logarithmieren beider Seiten des Modells erhalten Sie ein lineares Modell, das wie gewohnt gehandhabt werden kann.

---

† Verwenden sie 'read.table' zum Laden von txt-files und 'load' zum Laden von RData-files; in diesem Fall lässt sich der Datensatz einlesen durch A<−read.table("http://www.trutschnig.net/brainhead.txt",head=TRUE)

12. Übung am 15. Juni 2016

[LVA 431.010 VU Statistik für Naturwissenschaftler,
Link zur Ankreuzliste siehe http://www.trutschnig.net/courses]

Übungsaufgabe 61 (R) Der Datensatz Patients2.RData auf http://www.trutschnig.net enthält BMI (body mass index), Grösse, Gewicht und Cholesterol-Wert von 174 Patienten. Passen Sie ein lineares Modell mit 'BMI' als erklärter Variable und 'Grösse' und 'Gewicht' als erklärender Variable an und berechnen Sie das entsprechende R^2. Warum ist das Resultat überraschend?†

Übungsaufgabe 62 (R) Der Datensatz 'stackloss' ist standardmäßig in R enthalten, die ersten Zeilen können mit A<−stackloss; head(A) betrachtet werden.
Verwenden Sie help(stackloss) um herauszufinden, was der Datensatz enthält. Passen Sie dann ein lineares Modell mit 'stack.loss' als erklärter und 'Air.Flow', 'Water.Temp', und 'Acid.Conc.' als erklärende Variable an (i.e. dreidimensionale lineare Regression). Verwenden Sie dann das leaps-package um jenes Modell mit kleinstem BIC zu berechnen. Wie groß ist R^2 für dieses Modell ? Prognostizieren Sie weiters mit Hilfe dieses Modells den Wert von 'stack.loss' für den Fall, dass Air.Flow=72, Water.Temp=20, Acid.Conc.=85.

Übungsaufgabe 63 (R) Der Datensatz moisture.txt auf http://www.trutschnig.net enthält Wassergehaltsmessung von Bodenproben in Abhängigkeit der Tiefe (Einheit: 'grams water per 100 grams dry weight'). Plotten Sie die Daten (x-Achse: Tiefe), berechnen Sie mittels sm.regression den Kernregression-Schätzer (kurz: Kernregression) mit 'Tiefe' als erklärender Variable und ergänzen Sie selbigen im plot. Prognostizieren Sie den Wassergehalt in 5, 10 und 15 Meter Tiefe. Wie würden Sie den Wassergehalt in 25 Meter Tiefe schätzen?
Hinweis: Orientieren Sie sich an 'R-codes_GEO_12.R'

Übungsaufgabe 64 (R - Fortsetzung von Aufgabe 63) Angenommen, aus physikalischen Modellen ist bekannt, dass (unter speziellen Bedingungen) der Zusammenhang zwischen Tiefe T und Wassergehalt W gegeben ist durch folgende Formel

$$W = 125 - 110(1 - e^{-ax}),$$

wobei $a > 0$ eine von Bodenbedingungen abhängige (und aus den Stichproben zu schätzende) Variable ist. Wie würden Sie a für die vorliegenden Daten wählen?

Übungsaufgabe 65 (R) Der Datensatz SBP.RData enthält Blutdruck (SBP), BMI (und Alter) von 8215 Patienten einer Klinik. Plotten Sie die Daten (x-Achse: BMI, y-Achse: SBP), berechnen Sie mittels sm.regression den Kernregression-Schätzer (kurz: Kernregression) mit 'BMI' als erklärender Variable und ergänzen Sie selbigen im plot. Prognostizieren Sie den (durschnittlichen) Blutdruck für Patienten mit BMI = 20,21,22,...,34,35.

†Lesen Sie nach (wikipedia), wie der BMI berechnet wird