Aufgabe 7 (ca. 10,5 Punkte)

In 41 US-amerikanischen Städten wurde die Schwefeldioxid-Konzentration in der Luft in Abhängigkeit von klimatischen und geographischen Variablen untersucht. U. a. wurde auch ein multiples lineares Regressionsmodell mit den folgenden drei erklärenden Variablen gerechnet: `temp` (Jahresdurchschnittstemperatur in Grad Fahrenheit), `anzahl_unt` (Anzahl der produzierenden Unternehmen mit mehr als 20 Arbeitern), `wind` (jährliche durchschnittliche Windgeschwindigkeit in Meilen pro Stunde). Die abhängige Variable war die logarithmierte jährliche durchschnittliche Schwefeldioxidkonzentration in Mikrogramm pro Quadratmeter `log(sd_qm)`. Mit R ergaben sich folgende Werte

```
summary(model)
##
## Call:
## lm(formula = log(sd_qm) ~ temp + anzahl_unt + wind)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7890 -0.3329 -0.1623  0.3150  1.4729
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.1230847  1.0019137      NA       NA
## temp        -0.0526558  0.0121583      NA       NA
## anzahl_unt   0.0005650  0.0001497      NA       NA
## wind        -0.1412816  0.0610668      NA       NA
##
## Residual standard error: 0.5144 on 37 degrees of freedom
## Multiple R-squared:  0.5037,Adjusted R-squared:  0.4635
## F-statistic: 12.52 on 3 and 37 DF,  p-value: 8.456e-06
```

a) Welche Voraussetzungen müssen allgemein erfüllt sein, um auf Signifikanz der Regressionsparameter β_i testen zu können?

b) Formulieren Sie Hypothese und Alternative um zu überprüfen ob die Jahresdurchschnittstemperatur einen signifikanten Einfluss auf die abhängige Variable hat. Führen Sie anschließend einen geeigneten Test zum Signifikanzniveau $\alpha = 0.05$ durch.

c) Als Alternative zum obigen Modell betrachtet man ein Modell, welches nur `temp` als Kovariable hat.

```
model_temp <- lm(log(sd_qm) ~ temp)
```

(i) Welche Modellwahlkriterien sind für beide Modelle in der Angabe enthalten?

(ii) Welches dieser Kriterien können Sie zum Vergleich der beiden Modelle verwenden? Begründen Sie Ihre Antwort.

(iii) Welches Modell würden Sie bzgl. dieses Kriteriums bevorzugen? Begründen Sie Ihre Antwort.

Hinweis:

```
c(summary(model_temp)$r.squared, summary(model_temp)$adj.r.squared)
## [1] 0.2776821 0.2591611
c(qt(0.95, df = 40), qt(0.975, df = 40), qt(0.95, df = 37), qt(0.975, df = 37))
## [1] 1.683851 2.021075 1.687094 2.026192
```

Technische Universität München — MA9712 Statistik für BWL
Fakultät für Mathematik — Modulprüfung

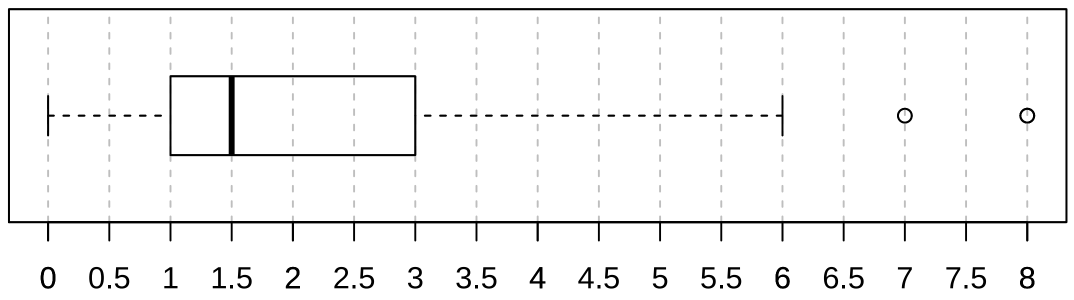
Claudia Czado, Stephan Haug — 22. Juli, 11:00 - 12:30 Uhr

Allgemeine Hinweise: Bis auf einen nicht-programmierbaren Taschenrechner sind keine Hilfsmittel zugelassen. Lösungen sind ausschließlich in das beigefügte **Lösungsblatt** einzutragen. Insgesamt sind 40 Punkte zu erreichen. Runden Sie numerische **Endergebnisse** auf **drei Nachkommastellen**. Runden Sie **keine Zwischenergebnisse**.

Gruppe A

Aufgabe 1 (ca. 2,5 Punkte)

Die nachfolgende Grafik zeigt den Boxplot einer Urliste $x_1, \ldots, x_n$.



a) Geben Sie den Median, den Interquartilsabstand sowie die Spannweite der Urliste an.

b) Beschreiben Sie die Form der Verteilung der Urliste $x_1, \ldots, x_n$.

Hinweis: Zur Konstruktion der Box wurden das untere und obere Quartil verwendet.

Aufgabe 2 (ca. 5 Punkte)

Eine Maschine befüllt Milch in 350 ml Packungen. Es kann angenommen werden, dass die Abfüllmenge normalverteilt ist mit Erwartungswert ϑ. Es wird vermutet, dass die Maschine nicht korrekt arbeitet und dadurch die erwartete Abfüllmenge vom Wert $\vartheta_0 = 350$ abweicht. Es wurde eine Stichprobe der Länge $n = 16$, bestehend aus Abfüllmengen von Milchpackungen dieser Maschine, erstellt. Die Stichprobenrealisierung $x_1, \ldots, x_{16}$ ergab die Werte $\bar{x}_{16} = 346.667$ und $s_{16}^2 = 17.541$.

a) Geben Sie Nullhypothese und Alternative des Testproblems, welches zur Überprüfung der Vermutung aus der Aufgabenstellung verwendet werden kann, an.

b) Wählen Sie einen geeigneten Test zur Überprüfung des Testproblems aus Teil a). Geben Sie den Wert der zugehörigen Teststatistik an.

c) Kann die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ verworfen werden? Begründen Sie Ihre Antwort.

```
round(c(qt(1 - alpha/2, df = n - 1), qt(1-alpha/2, df=n),
        qt(1 - alpha, df = n - 1), qt(1 - alpha, df = n)), 3)
## [1] 2.131 2.120 1.753 1.746
```

Aufgabe 3 (ca. 3 Punkte)

Entscheiden Sie, ob die folgenden Aussagen **Wahr** oder **Falsch** sind. Ihre Antworten müssen **nicht** begründet werden.

a) Median, empirischer Mittelwert und Spannweite sind alles Lageparameter zur Beschreibung der Verteilung eines metrischen Merkmals.

b) Für ein diskretes Merkmal ist der Modus stets eindeutig.

c) Der oder die Ausprägungen eines diskreten Merkmals mit der größten Häufigkeit bezeichnet man als Modus.

d) Die empirische Varianz kann als mittlere quadratische Abweichung vom Mittelwert aufgefasst werden.

e) Beim Boxplot liegen ungefähr $\frac{2}{3}$ der Beobachtungen innerhalb der Box.

f) Die Summe der Abweichungen $(x_i - \overline{x}_n)$ vom empirischen Mittelwert ist stets Null.

Antworten Sie mit **W** für Wahr oder **F** für Falsch.

Aufgabe 4 (ca. 9 Punkte)

Seien $X_1, \dots, X_n \overset{uiv}{\sim} U_{[-\vartheta,0]}$ mit $\vartheta \in (0, \infty)$ und $x_1, \dots, x_n$ ein Stichprobenergebnis.

a) Zeigen Sie für $x \in [-\vartheta, 0]$, dass die Verteilungsfunktion von X_1 gegeben ist durch $F^{X_1}(x) = \left(\frac{x}{\vartheta} + 1\right)$.

b) Zeigen Sie, dass der Maximum-Likelihood-Schätzer für ϑ gegeben ist durch
$$\widehat{\vartheta}(X_1, \dots, X_n) = -X_{(1)}\,,$$
wobei $X_{(1)} = \min\{X_1, \dots, X_n\}$.
Hinweis: Überlegen Sie sich, ob die Likelihoodfunktion differenzierbar ist.

c) Zeigen Sie für $x \in [0, \vartheta]$, dass die Verteilungsfunktion von $-X_{(1)}$ gegeben ist durch $F^{-X_{(1)}}(x) = \left(\frac{x}{\vartheta}\right)^n$.
Hinweis: Verwenden Sie in Ihrer Rechnung in einem Schritt die Wahrscheinlichkeit $P(X_{(1)} \geq -x)$.

d) Zeigen Sie, dass die Riemann-Dichte von $-X_{(1)}$ gegeben ist durch $f^{-X_{(1)}}(x) = \frac{n \cdot x^{n-1}}{\vartheta^n} \cdot \mathbb{1}_{(0,\vartheta)}(x)$.

e) Zeigen Sie, dass der Maximum-Likelihood-Schätzer $\widehat{\vartheta}(X_1, \dots, X_n)$ nicht erwartungstreu ist.

Hinweis: Jede Teilaufgabe kann eigenständig gelöst werden. Dabei dürfen die Ergebnisse der vorherigen Teilaufgaben ohne Nachweis verwendet werden.

Aufgabe 5 (ca. 6 Punkte)

Eine Brauerei produziert ein neues alkoholfreies Bier. In einem Geschmackstest erhalten $n = 10$ Personen je ein Glas alkoholfreies bzw. gewöhnliches Bier, und sie sollen versuchen, das alkoholfreie Bier zu identifizieren. 6 Personen können die Getränke richtig unterscheiden. Testen Sie anhand dieser Daten zum Signifikanzniveau $\alpha = 0.05$ die Nullhypothese, alkoholfreies und gewöhnliches Bier sind geschmacklich nicht zu unterscheiden, gegen die Alternative, dass man mit einer Wahrscheinlichkeit von mehr als 0.5 das alkoholfreie Bier erkennt, d.h. testen Sie

$$H_0: \ \vartheta = 0.5 \qquad H_1: \ \vartheta > 0.5$$

mit $\vartheta = \{\text{Wahrscheinlichkeit sich richtig für alkoholfreies Bier zu entscheiden}\} \in (0, 1)$.

a) Geben Sie den Wert der Teststatistik T des zugehörigen exakten Tests an und entscheiden Sie ob die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ verworfen werden kann. Begründen Sie Ihre Antwort.

b) Beschreiben Sie allgemein in Worten den Fehler 1. Art.

c) Unabhängig vom Ergebnis in a) nehmen wir nun an, dass die Wahrscheinlichkeit für das Ereignis $K = \{\text{alkoholfreies Bier wird korrekt erkannt}\}$ gleich 0.5 ist. Es ist bekannt, dass der Verzehr von Weißbrot vor der jeweiligen Bierprobe hilfreich dabei ist, sich korrekt zu entscheiden. Dies wird allerdings nicht von allen Testern durchgeführt. Man weiß, dass 74% der Tester, die sich korrekt entscheiden, Weißbrot essen. Andererseits essen nur 20% der Tester, die sich falsch entscheiden, Weißbrot. Geben Sie die Wahrscheinlichkeit an, dass ein Tester korrekt das alkoholfreie Bier erkennt, wenn er zuvor Weißbrot gegessen hat.

```
round(pbinom(0:n, size = n, prob = 0.5), 3)
##  [1] 0.001 0.011 0.055 0.172 0.377 0.623 0.828 0.945 0.989 0.999 1.000
```

Aufgabe 6 (ca. 4 Punkte)

In einem Experiment zur Wirkung von Alkohol auf die Reaktionszeit wurden insgesamt 480 Versuchspersonen zufällig in zwei Gruppen aufgeteilt. Eine der beiden Gruppen erhielt dabei eine standardisierte Menge Alkohol. Abschließend ergab sich folgende Kontingenztafel

```
table(alkohol, reaktion)
##               reaktion
## alkohol        gut mittel stark verzoegert
##   mit Alkohol   84     69               87
##   ohne Alkohol  90     49              101
```

Anhand dieser Daten soll der Zusammenhang zwischen Alkoholkonsum und Reaktionszeit untersucht werden.

a) Berechnen Sie den korrigierten Kontingenzkoeffizienten. Beurteilen Sie die Stärke der Assoziation zwischen den Merkmalen Alkoholkonsum und Reaktionszeit.

b) Geben Sie die Nullhypothese des χ^2-Unabhängigkeitstest an.

c) Entscheiden Sie, ob für diese Daten die Nullhypothese des χ^2-Unabhängigkeitstest zum Niveau 0.05 verworfen werden kann. Begründen Sie Ihre Antwort.

Hinweis:

```
chisq.test(table(alkohol, reaktion), correct = FALSE)$statistic
## X-squared
##     4.639
round(qchisq(c(0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99), df = 2), 3)
##  [1] 4.605 4.816 5.051 5.319 5.627 5.991 6.438 7.013 7.824 9.210
```