Aufgabe 1

Der Student Paul fährt jeden Tag mit dem Fahrrad zur Uni. Die Strecke kreuzt drei Fußgängerüberwege mit Bedarfsampeln, die durch Fußgänger aktiviert werden. Wird Paul an einer Bedarfsampel gestoppt, verzögert dies seine Weiterfahrt um 90 Sekunden. Wird Paul nicht aufgehalten, benötigt er 17 Minuten, bis er den Hörsaal erreicht.

Nehmen Sie an, die Zufallsvariable X: "Anzahl der Bedarfsampeln, die auf Pauls Weg zur Uni ein rotes Signal zeigen" ist binomialverteilt mit den Parametern $n = 3$ und $p = 0.4$, d.h. $X \sim \text{Bin}(3, 0.4)$.

1. Berechnen Sie die Wahrscheinlichkeit, dass
   (a) genau zwei Bedarfsampeln auf Pauls Weg zur Uni das Haltesignal zeigen.
   (b) höchstens zwei Bedarfsampeln auf Pauls Weg zur Uni das Haltesignal zeigen.
   (c) mindestens zwei Bedarfsampeln auf Pauls Weg zur Uni **nicht** das Haltesignal zeigen.
2. Geben Sie das 90%-Quantil der Binomialverteilung mit Parametern $n = 3$ und $p = 0.4$ an. Interpretieren Sie dieses Quantil inhaltlich.
   (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden, dass das gesuchte Quantil den Wert 3 hat)
3. Geben Sie Modus, Erwartungswert und Varianz der Zufallsvariablen X an.
   (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden, den Wert 2 für den Modus, den Wert $\frac{5}{4}$ für den Erwartungswert und den Wert 3.1 für die Varianz)

Betrachten Sie zusätzlich die Zufallsvariable Y: "Fahrtzeit in Minuten, die Paul für seinen Weg zur Uni benötigt". Paul erkennt, dass er die Zufallsvariablen X und Y linear ineinander überführen kann:

$$Y = 17 + 1.5 \cdot X$$

4. Geben Sie Modus, Erwartungswert und Varianz der Zufallsvariablen Y an.
5. Pauls Lieblingsübung beginnt pünktlich um 8:00 Uhr. Wann muss Paul spätestens losfahren, wenn er mit Wahrscheinlichkeit von 0.9 rechtzeitig ankommen möchte?

6. Geben Sie an, ob die folgenden Aussagen "wahr" oder "falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und keine Antwort wird mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

   (a) Die Poissonverteilung ist eine stetige Verteilung.

   (b) Der Parameter λ der Poissonverteilung bestimmt den zugehörigen Erwartungswert.

   (c) Die Normalverteilung liefert eine geeignete Approximation der Binomialverteilung, wenn für deren Parameter gilt $n \cdot p \cdot (1-n) > 9$.

   (d) Die Wahrscheinlichkeitsfunktion der Poissonverteilung nimmt negative Werte an, wenn der Parameter $\lambda < 1$ ist.

Aufgabe 2

In der Tiefgarage eines Einkaufszentrums müssen die Kunden einen Aufzug zu den Geschäften benutzen. Die Zufallsvariable X: "Wartezeit vor dem Aufzug" sei exponentialverteilt mit der Verteilungsfunktion $F_{Exp}(x;\lambda) = 1 - e^{-\lambda x}$, für $x \geq 0$. Durchschnittlich warten die Kunden 1.25 Minuten vor dem Aufzug.

1. Bestimmen Sie den Erwartungswert $E(X)$ und dadurch den Parameter λ.
   (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden λ =0.9)

2. Berechnen Sie die Wahrscheinlichkeit, dass ein Kunde mehr als 3 Minuten auf den Aufzug warten muss.

3. Berechnen Sie die Wahrscheinlichkeit, dass ein Kunde gar nicht warten muss.

4. Mit einer Wahrscheinlichkeit von 60% liegt die Wartezeit im Intervall $[a, 4[$. Berechnen Sie den Wert von a.

5. Die Zufallsvariable Y: "Stromverbrauch (in kW) des Aufzugs pro Stunde" sei normalverteilt. Gegeben ist die Varianz $\sigma_Y^2 = 2$. Eine Stichprobe von 32 solchen Aufzügen wurde geprüft. Das Stichprobenmittel ist $\overline{y}_{32} = 17.5$. Berechnen Sie ein realisiertes 90%-Konfidenzintervall für den mittleren stündlichen Stromverbrauch μ_Y.

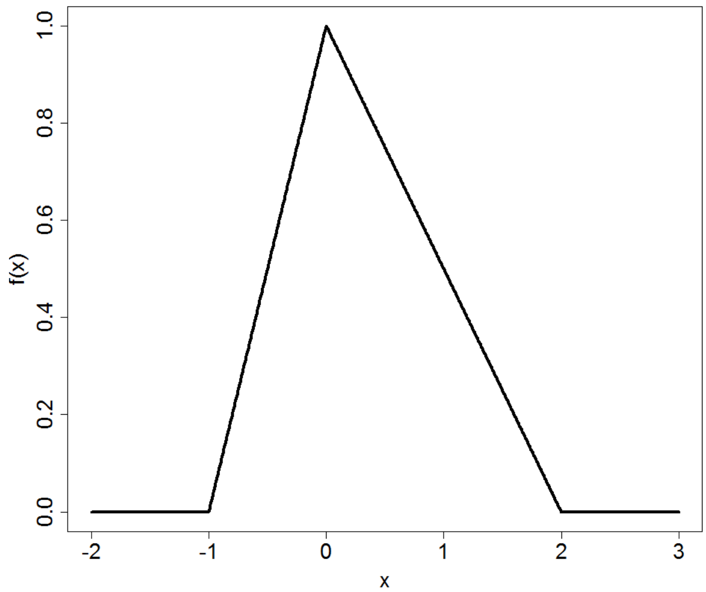
6. Das Einkaufszentrum hat 3 baugleiche Aufzüge, die voneinander unabhängig sind. Der stündliche Stromverbrauch eines Aufzugs sei $Y_i \sim N(17, 2), i = 1, 2, 3$. Der Strompreis beträgt c Euro/kWh. Gegeben ist Zufallsvariable K: "Gesamte stündliche Stromkosten der 3 Aufzüge". Wie ist $K = c \cdot (Y_1 + Y_2 + Y_3)$ verteilt?

7. Geben Sie an, ob die folgenden Aussagen "wahr" oder "falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und keine Antwort wird mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

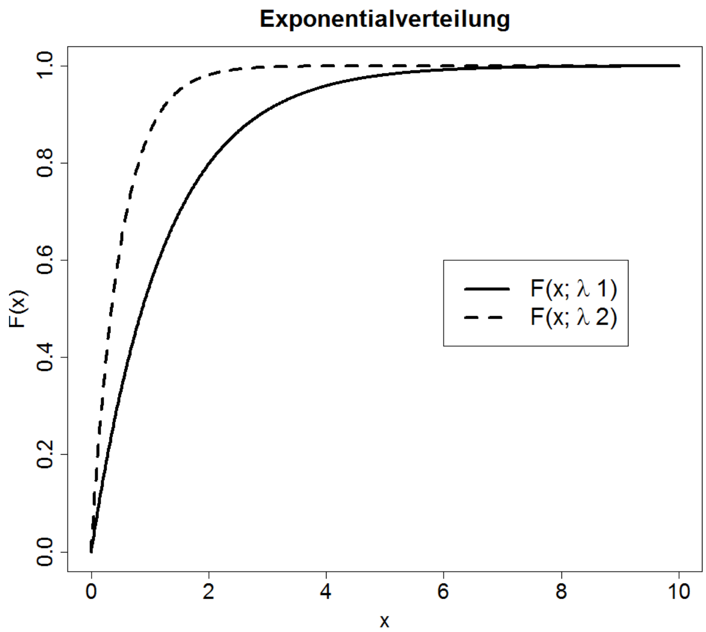
   (a) Die Dichtefunktion $f_X(x)$ einer stetigen Zufallsvariablen kann Werte größer als 1 annehmen.

   (b) Für die Verteilungsfunktion einer stetigen Zufallsvariablen gilt $F_X(x) = \int_{-\infty}^{x} f(s)ds$.

(c) Die im Folgenden skizzierte Funktion $f(x)$ kann als Dichtefunktion interpretiert werden.



(d) Aus folgendem Plot kann man folgern, dass $\lambda 1 > \lambda 2$ ist.



Aufgabe 3

Rennwagen **R1** und **R2** trainieren abwechselnd auf einer Strecke, d.h. die Autos fahren unabhängig voneinander. Ein Beobachter notiert die auf der Strecke erreichten Zeiten (in Minuten) x_i von R1 und y_i von R2 für $i = 1, 2, \ldots, 10$ Runden und berechnet folgende Größen: $\overline{x}_{10} = 25.23$, $\sum_{i=1}^{10} x_i^2 = 6399.79$, $\overline{y}_{10} = 25.49$ und $\sum_{i=1}^{10}(y_i - \overline{y})^2 = 9.269$.

1. Bestimmen Sie die Stichprobenvarianzen der erreichten Zeit von R1 und R2.
   (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden $s^2_{X;10} = 4$ und $s^2_{Y;10} = 1$.)

Die Zufallsvariablen X: "Erreichte Zeit (in Min.) von R1" und Y: "Erreichte Zeit (in Min.) von R2" seien in der Grundgesamtheit normalverteilt: $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$.

2. Testen Sie mit einer Irrtumswahrscheinlichkeit von 5% die Vermutung, dass R1 im Durchschnitt besser abschneidet als R2, d.h. $\mu_X < \mu_Y$, wenn es gilt $\sigma_X^2 = \sigma_Y^2$.

3. Zusätzlich wird die Hypothese $H_0 : \mu_X = 25$ gegen $H_1 : \mu_X \neq 25$ mit $\alpha = 0.05$ getestet. Der p-Wert des Tests beträgt 0.71.

   (a) Treffen Sie eine Testentscheidung und begründen Sie diese kurz (keine Berechnung erforderlich).

   (b) Geben Sie den theoretischen kritischen Bereich des Tests an.

R1 trainiert anschließend allein auf einer Strecke der Länge 120 km. Auf dem ersten Streckenabschnitt fährt R1 mit einer Geschwindigkeit von 280 km/h. Die folgenden zwei gleich langen Abschnitte bewältigt R1 mit jeweils 300 km/h und 295 km/h.

4. Bestimmen Sie die durchschnittliche Geschwindigkeit des R1 auf der Strecke. Bekannt ist, dass der erste Abschnitt $\frac{2}{3}$ und die letzten beiden jeweils $\frac{1}{6}$ der Strecke ausmachen.

5. Geben Sie an, ob die folgenden Aussagen "wahr" oder "falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und keine Antwort wird mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

   (a) Eine einfache Zufallsstichprobe liegt vor, wenn alle X_i für $i = 1, \ldots, n$ stochastisch unabhängig, identisch verteilt sind.

(b) Sind die Zufallsvariablen $X_1, \ldots, X_n$ stochastisch unabhängig, normalverteilt mit μ_X und σ_X^2, dann ist das Stichprobenmittel $\overline{X}_n$ χ^2-verteilt mit $n-1$ Freiheitsgraden.

(c) Für die erreichte Zeit von R1 (x_i) wird ein Mittelwerttest bei bekannter Varianz ($H_0 : \mu_X \leq 25$ gegen $H_1 : \mu_X > 25$) mit $\alpha = 0.05$ durchgeführt. Hierbei ergibt sich eine Teststatistik von 0.32.

i. Die kritische Schranke des Tests lautet $\lambda_{0.95}$.

ii. Der p-Wert des Tests ≈ 0.37.

Aufgabe 4

Der R-Data Frame `Unternehmen` enthält die folgenden Merkmale von 40 deutschen Unternehmen des Dienstleistungssektors (aus dem Jahr 2010).

Merkmal A:	Umsatz in Mio. Euro	Spalte `Umsatz`
Merkmal B:	Branche	Spalte `Branche`
Merkmal C:	Gewinn in Mio. Euro	Spalte `Gewinn`
Merkmal D:	Umsatzrentabilität in % (=Gewinn/Umsatz*100)	Spalte `Um_Rent`

Gehen Sie im Folgenden davon aus, dass alle Einträge im Data Frame `Unternehmen` den passenden Datentyp besitzen.

1. Welchen Merkmalstyp besitzt das Merkmal D?
2. Welchen Datentyp hat die Spalte `Um_Rent` in R?

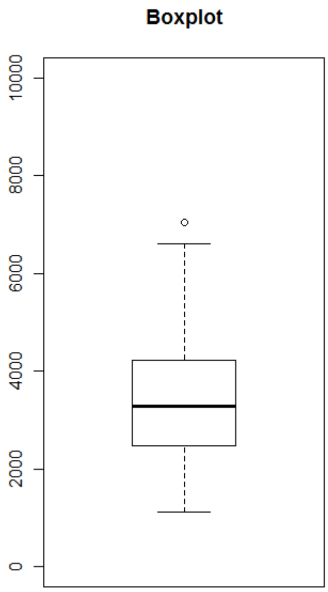
Betrachten Sie nun den zum Data Frame `Unternehmen` gehörenden R-Output und die Abbildung 1.

```
> table(Unternehmen$Branche)/nrow(Unternehmen)

 Finanzen    Handel  Logistik Tourismus
    0.275     0.225     0.425     0.075
>
> Log=Unternehmen[Unternehmen$Branche=="Logistik","Umsatz"]
> Fin=Unternehmen[Unternehmen$Branche=="Finanzen","Umsatz"]
>
> length(Log)
[1] 17
>
> quantile(Log,0.5,type=1)
 50%
3272
> quantile(Log,1,type=1)
100%
7033
```

```
> quantile(Fin,0.5,type=1)
 50%
4793
> max(Fin)
[1] 33333
```

Abbildung 1



3. Welche Ausprägungen des Merkmals Branche kommen bei den 40 Unternehmen vor?

4. Geben Sie den Modus des Merkmals Branche an.

5. Geben Sie den `R`-Befehl an, um

   (a) den durchschnittlichen Umsatz der 40 Unternehmen zu berechnen.

   (b) die Häufigkeitsverteilung des Merkmals Gewinn geeignet graphisch darzustellen.

6. Stellt der Befehl

   ```
   quantile(unternehmen$Branche,0.6,type=1)
   ```

   eine sinnvolle Operation dar? (mit kurzer Begründung)

7. Vervollständigen Sie die folgende `R`-Funktion so, dass damit die Varianz eines Datenvektors `x` berechnet wird.

   ```
   Varianz=function() { sum(x)/length(x) }
   ```

   Hinweis: Schreiben Sie die vollständige Funktion in Ihren Lösungsbogen!

8. Geben Sie an, ob folgende Aussagen „wahr" oder „falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und **keine** Antwort mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

   (a) Der Median von `Fin` ist größer als der von `Log`.

   (b) Alle Unternehmen in der Branche Finanzen hatten einen höheren Umsatz als die Unternehmen der Branche Logistik.

   (c) Die Abbildung 1 stellt den Boxplot des Merkmals Umsatz in der Branche Finanzen dar.

   (d) 50% der Unternehmen in der Branche Logistik erzielten einen Umsatz von maximal 3272 Mio. Euro.

Es soll nun mit Hilfe der Funktion `t.test` bei einer Irrtumswahrscheinlichkeit von 5% die Vermutung getestet werden, dass die durchschnittliche Umsatzrentabilität deutscher Dienstleistungsunternehmen größer als 3% ist.
Hierfür muss vorausgesetzt werden, dass es sich bei den 40 vorliegenden Unternehmen um eine einfache Zufallsstichprobe deutscher Dienstleistungsunternehmen handelt.

9. Welche weitere Voraussetzung bzgl. der Verteilung der Umsatzrentabilität deutscher Dienstleistungsunternehmen muss erfüllt sein?
   (**Hinweis:** Gehen Sie im Folgenden davon aus, alle Voraussetzungen seien erfüllt.)

10. Vervollständigen Sie den folgenden Befehl so, dass der oben beschriebene Test durchgeführt wird.

    ```
    t.test(Unternehmen$Um_Rent,y=NULL,mu=3,conf.level=0.95)
    ```

11. Der korrigierte Befehl ergab folgenden Output:

```
        One Sample t-test

data:  Unternehmen$Um_Rent t = 2.0081, df = 39, p-value = 0.0258
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 3.241423      Inf
sample estimates: mean of x
      4.5
```

Wie lautet die Testentscheidung? (mit <u>kurzer</u> Begründung)

Lösung Aufgabe 1

1. Punktwahrscheinlichkeiten aus Tabelle

   (a) $P(X = 2) = f_{Bin}(2; 3, 0.4) = 0.288$

   (b) $P(X \leq 2) = 1 - P(X > 2) = 1 - P(X = 3) = 1 - 0.064 = 0.936$

   (c) $P(3 - X \geq 2) = P(X \leq 1) = 0.648$

2. $Q_{90\%}^{\text{Bin}(3,0.4)} = 2$ Mit Wahrscheinlichkeit von 90% wird eine Fahrt von Paul zur Uni von höchstens zwei (Ersatzlösung: 3) Rotphasen verzögert.

3. $X_{mod} = 1$,
   $\mathrm{E}[X] = np = 1.2$,
   $\mathrm{Var}[X] = npq = 1.2 \cdot 0.6 = 0.72$.

4. $Y_{mod} = 17 + 1.5 \cdot X_{mod} = 18.5$ (Ersatzlösung: 20),
   $\mathrm{E}[Y] = 17 + 1.5 \cdot \mathrm{E}[X] = 18.8$ (Ersatzlösung: 18.875),
   $\mathrm{Var}[Y] = (1.5)^2 \cdot \mathrm{Var}[X] = 1.62$ (Ersatzlösung: 6.975).

5. 90% -Quantil aus Aufgabe 2: 2 (Ersatzlösung: 3)

$$8:00 \text{ Uhr} \quad - \quad \left(17 + 1.5 \cdot Q_{90\%}^{\text{Bin}(3,0.4)}\right) \text{Minuten “ = “ } 7:40 \text{ Uhr}$$

(Ersatzlösung: 7 : 38 Uhr und 30 Sekunden)

6. (a) falsch

   (b) wahr: $P \sim \text{Poiss}(\lambda) \Rightarrow \mathrm{E}[P] = \lambda$.

   (c) falsch: $n \in \mathbb{N}, p \in (0,1) \Rightarrow np(1-n) \leq 0$

   (d) falsch

Lösung Aufgabe 2

1.
$$E(X) = 1.25$$
$$\lambda = \frac{1}{E(X)} = \frac{1}{1.25} = 0.8$$

2. $P(X > 3) = 1 - P(X \leq 3) = 1 - F(3) = e^{-0.8 \cdot 3} = 0.0907$

3. 0, da Punktwahrscheinlichkeit.

4. $P(a \leq X < 4) = F(4) - F(a) = 1 - e^{-0.8 \cdot 4} - (1 - e^{-0.8 \cdot a}) = 0.6$
   $a = 0.5563$

5.
$$\begin{aligned} KI &= \left[\bar{Y}_n - \lambda_{1-\frac{\alpha}{2}} \frac{\sigma_Y}{\sqrt{n}}, \bar{Y}_n + \lambda_{1-\frac{\alpha}{2}} \frac{\sigma_Y}{\sqrt{n}}\right] \\ &= \left[17.5 - 1.6448 \cdot \sqrt{\frac{2}{32}}, 17.5 + 1.6448 \cdot \sqrt{\frac{2}{32}}\right] \\ &= [17.0888, 17.9112] \end{aligned}$$

6. $K \sim \mathcal{N}(51c, 6c^2)$

7. wahr, wahr, falsch, falsch

Lösung Aufgabe 3

NEU

1. $s^2_{X;10} = \frac{n}{(n-1)}\overline{x^2} - \overline{x}^2 = \frac{(\sum_{i=1}^{10} x_i^2) - n*\overline{x}^2}{(n-1)} = (6399.79 - 10*(25.23)^2)/9 = 3.8068$
   $s^2_{Y;10} = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{9.269}{9} = 1.0299$

2. (a) Ein Mittelwertdifferenztest für unverbundene Stichproben, da x_i und y_i paarweise stochastisch unabhängig sind (verschiedene Autos).

   (b) $H_0: \ \mu_x - \mu_y \geq 0$ gegen $H_1: \mu_x - \mu_y < 0$. $\Rightarrow$
   Kritischer Bereich: $\frac{\overline{X}_n - \overline{Y}_m}{\overline{S}}\sqrt{\frac{mn}{n+m}} < -t_{1-\alpha;n+m-2}$.

   (c) kritische Schranke: $-t_{1-\alpha;n+m-2} = -t_{0.95;18} = -1.734$

   (d)

$$\overline{S}^2 = \frac{(n-1)s^2_X + (m-1)s^2_Y}{n+m-2} = \frac{9*3.8068 + 9*1.0299}{18} = 2.41 \text{ bzw. } \left[\frac{9*4 + 9*1.0299}{18} = 2.515\right]$$

   (e) Teststatistik: $T = \frac{25.23-25.49}{1.5551}\sqrt{\frac{100}{20}} = -0.3739$ bzw. $\left[\frac{25.23-25.49}{1.5859}\sqrt{\frac{100}{20}} = -0.3666\right]$

   (f) Testentscheidung: $T = -0.3739[-0.3666] \not< -1.734 = -t_{0.95;18}$, d.h die Teststatistik liegt nicht im kritischen Bereich, somit wird H_0 nicht abgelehnt.

3. (a) Die H_0 wird nicht abgelehnt, da $p > \alpha$ ist, somit liegt die Teststatistik nicht im kritischen Bereich.

   (b) Der kritische Bereich des Tests lautet: $T_n = \left|\frac{\overline{X}_n - \mu_0}{S_{X;n}/\sqrt{n}}\right| > t_{1-\alpha/2;n-1}$.

4. <u>Berechnung der durchschnittlichen Geschwindigkeit:</u>
   $x_1 = \frac{2}{3} * 120 = 80$ und $x_2 = x_3 = \frac{1}{6} * 120 = 20$.
   Harmonisches Mittel, da eine Verhältniszahl und benötigte Zeit (Nenner) unbekannt. (Alternativ arithmetisches Mittel möglich)

$$\overline{v}_H = \left(\frac{1}{n}\sum_{i=1}^{k}\frac{n_i}{x_i}\right)^{-1} = \left(\frac{1}{120}\left(\frac{80}{280} + \frac{20}{300} + \frac{20}{295}\right)\right)^{-1} = 285.5935 \text{ km/h}$$

5. wahr, falsch, wahr, wahr.

Lösung Aufgabe 4

1. quantitativ
2. numeric
3. Finanzen, Handel, Logistik, Tourismus
4. Logistik
5. Befehle:
   (a) `mean(Unternehmen$Umsatz)`
   (b) `hist(Unternehmen$Gewinn,prob=T)`
6. nein, denn
   Vektor `unternehmen$Branche` ist in `R` nicht definiert, da zwischen Groß- und Kleinschreibung unterschieden wird, d.h. Zugriff auf Spalte `Branche` des Data Frame nur mit `Unternehmen$Branche`
   ODER
   Quantile nur sinnvoll bei mindestens ordinalskalierten Merkmalen, Merkmal Branche ist nur nominalskaliert
7. z.B.

   ```
   Varianz=function(x) { mean(x^2)-(sum(x)/length(x))^2}
   ```
8. Geben Sie an, ob folgende Aussagen „wahr“ oder „falsch“ sind.
   (a) w
   (b) f
   (c) f
   (d) w
9. Umsatzrentabilität muss in der GG normalverteilt sein
10. `alternative="greater"`
11. H_0 wird abgelehnt, da p-Wert kleiner als vorgegebene Irrtumswahrscheinlichkeit, Vermutung somit bestätigt