Aufgabe 1

Für die Securance-Versicherung liegen Ihnen die gemeinsamen absoluten Häufigkeiten der Merkmale X: "Schadenshöhe" und Y: "Versicherungsart" für die letzten 500 gemeldeten Schäden vor.

$y_i \downarrow x_j \rightarrow$	gering	mittel	hoch	$n_{i\cdot}$
Hausrat	200	25	0	225
KFZ	0	10	75	85
Unfall	20	35	90	145
Reiserücktritt	40	5	0	45
$n_{\cdot j}$	260	75	165	500

1. Interpretieren Sie n_{13}.

2. Geben Sie jeweils den Merkmalstyp und das Skalenniveau für das Merkmal X und für das Merkmal Y an.

3. Berechnen Sie für das Merkmal Y ein geeignetes Lagemaß.

4. Berechnen Sie für das Merkmal X zwei geeignete Lagemaße.

Die Securance-Versicherung hat im Durchschnitt 1 Schadensfall pro Tag. Betrachtet wird die Zufallsvariable S_k: "Anzahl der Schäden in k Tagen".

5. Nennen Sie die Annahmen des Poisson-Prozesses.

6. Gehen Sie im Folgenden davon aus, S_k sei Poisson-verteilt mit Parameter λ_k.

   (a) Bestimmen Sie den Parameter λ_5 für die Zufallsvariable S_5 (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden $S_5 \sim \text{Poi}(10)$).

   (b) Mit welcher Wahrscheinlichkeit geschehen innerhalb von 5 Tagen mehr als 5 Schadensfälle?

   (c) Mit welcher Wahrscheinlichkeit geschehen innerhalb von 5 Tagen mehr als 3 aber weniger als 8 Schadensfälle?

(d) Bestimmen Sie Mittelwert, Median und Modi von S_5.

Schmierpapier:

Aufgabe 2

Die Securance-Versicherung modelliert die Zeitdauer zwischen zwei Unfällen mithilfe der Zufallsvariable X: "Wartezeit bis zum nächsten Unfall in Stunden". X sei exponentialverteilt mit Dichtefunktion:

$$f_X(x) = \lambda \exp(-\lambda x), \quad x > 0, \lambda > 0.$$

1. Wie groß ist die Wahrscheinlichkeit, dass die Wartezeit zwischen zwei Unfällen mehr als drei Stunden beträgt, wenn der Parameter $\lambda = 1/2$ ist?

2. Es wird der Maximum-Likelihood Schätzer für den unbekannten Parameter λ gesucht.

   (a) Zeigen Sie, dass die logarithmierte Likelihoodfunktion für den unbekannten Parameter λ gegeben ist durch:

$$\ln L(\lambda; x_i) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

(b) Zeigen Sie, dass der Maximum-Likelihood Schätzer für den unbekannten Parameter λ lautet:

$$\hat{\lambda}_{ML} = \frac{1}{\bar{x}},$$

wobei $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. (Keine Berechnung der zweiten Ableitung nötig.)

3. Sie notieren die Wartezeit zwischen 20 Unfällen und fassen die Daten als Realisationen einer Stichprobe $X_1, \ldots, X_{20}$ aus einer exponentialverteilten Grundgesamtheit auf. Anschließend berechnen Sie $\sum_{i=1}^{20} x_i = 40$.

   (a) Tragen Sie die Werte der logarithmierten Likelihoodfunktion für $\lambda = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$ in untenstehende Tabelle ein und zeichnen Sie die entsprechenden Werte in ein geeignetes Koordinatensystem.

λ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1

(b) Berechnen Sie $\hat{\lambda}_{ML}$ aufgrund der Ergebnisse der Stichprobe.

(c) Der Vorstand des Unternehmens vermutet, dass die Wartezeit zwischen zwei Unfällen mehr als zwei Stunden beträgt. Sie formulieren folgendes Hypothesenpaar:

$$H_0 : \lambda \leq \lambda_0 = 1/2 \quad H_1 : \lambda > \lambda_0 = 1/2.$$

Zur Überprüfung obiger Hypothesen ziehen Sie folgende Prüfgröße heran $T_n = 2\lambda_0 \sum_{i=1}^{n} X_i$. Testen Sie auf Basis des obigen Stichprobenbefundes bei einer Irrtumswahrscheinlichkeit von 5% die Vermutung des Vorstandes. Sie wissen, dass die Prüfgröße T_n unter Gültigkeit der Nullhypothese χ^2-verteilt mit $2n$ Freiheitsgraden ist. Vervollständigen Sie das folgende Testverfahren.

Kritischer Bereich:

Kritische Schranke:

Berechnung der Teststatistik:

Testentscheidung:

4. Vervollständigen Sie folgende Aussagen:

   (a) $P(T_n > 40 | \lambda_0 = 1/2) = 0.4702$ ist der p-Wert des Hypothesentests aus Teilaufgabe 3 (c). Auf Basis dieses Wertes ist die Nullhypothese bei einem Signifikanzniveau von 5% ______________________________.

   (b) $P(T_n \neq 40 | \lambda_0 = 1/2) =$ ____________.

5. Berechnen Sie die Wahrscheinlichkeit, dass die Prüfgröße T_n unter Gültigkeit der Nullhypothese aus Teilaufgabe 3 (c) zwischen 25.8 und 51.8 liegt.

Aufgabe 3

Bei einem Elfmeterschießen, in dem ein Spieler auf einen Torwart schießt, gibt es für den Schützen nur die Möglichkeiten eines Schusses nach links, rechts oder in die Mitte. Der Torwart kann sich ebenfalls links, rechts oder mittig positionieren. Ein Elfmeter gilt als gehalten, wenn sich der Torwart in Schussrichtung positioniert.

1. Für den Verlauf des Elfmeterschießens gelte folgende Wahrscheinlichkeitstabelle:

			Schussrichtung S des Schützen			
			Links	Mitte	Rechts	
			$S=1$	$S=2$	$S=3$	
Position T des Torwarts	Links	$T=1$				1/3
	Mitte	$T=2$				1/3
	Rechts	$T=3$	1/18	1/18	2/9	1/3
			1/3	1/3		1

   (a) Vervollständigen Sie die Wahrscheinlichkeitstabelle. Zusätzlich wissen Sie, dass
   - der Schütze mit einer Wahrscheinlichkeit von 3/10 in die Mitte schießt, wenn sich der Torwart links positioniert,
   - $P(T=2,\ S=1) = P(T=2,\ S=3)$ gilt.

(b) Berechnen Sie $P(T = 3|S = 1)$. Überprüfen Sie, ob die Ereignisse "Der Schütze schießt nach links" und "Der Torwart positioniert sich rechts" stochastisch unabhängig sind.

Hinweis: Wenn Sie Teilaufgabe 1 (a) nicht lösen konnten, verwenden Sie im Folgenden

$$P(T = 1,\ S = 1) = P(T = 1,\ S = 3) = P(T = 2,\ S = 2) = P(T = 2,\ S = 3) = \frac{1}{18},$$
$$P(T = 1,\ S = 2) = P(T = 2,\ S = 1) = \frac{2}{9}.$$

(c) Berechnen Sie die Wahrscheinlichkeit dafür, dass ein Tor fällt (also dass Schussrichtung des Schützen und Position des Torwarts nicht übereinstimmen).

2. Ein Schütze, der mit Wahrscheinlichkeit 40% trifft, hat nun 5 Versuche.

(a) Welche Annahme muss getroffen werden, damit die Zufallsvariable X_5: "Anzahl der Treffer bei 5 Schüssen" binomialverteilt ist?

Gehen Sie von nun an davon aus, dass die benötigte Annahme erfüllt ist.

(b) Geben Sie die Verteilungsparameter der Zufallsvariable X_5 an.

(**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden $X_5 \sim \text{Bin}(n = 9,\ p = 0.25)$).

(c) Berechnen Sie die Wahrscheinlichkeit, dass der Schütze bei 5 Schüssen mindestens 3 Mal trifft.

(d) Wie viele Treffer kann der Schütze erwarten?

3. Nun sind zwei weitere Spieler (A und B) am Zug. Für die Zufallsvariablen
A_2: "Anzahl der Treffer bei 2 Schüssen von Spieler A" und
B_2: "Anzahl der Treffer bei 2 Schüssen von Spieler B"
gelte

$$A_2 \sim \text{Bin}(n = 2, p = 0.4) \qquad \text{und} \qquad B_2 \sim \text{Bin}(n = 2, p = 0.5).$$

(a) Wie hoch ist die Wahrscheinlichkeit, dass Spieler B bei 2 Schüssen genau 2 Treffer erzielt?

(b) Berechnen Sie $P(A_2 = 0)$.

(c) Spieler A und B veranstalten einen Wettkampf:

> Es gewinnt, wer nach 5 Schüssen mehr Treffer hat. Bei gleicher Trefferanzahl einigen sie sich auf Unentschieden.
>
> Nachdem die beiden Spieler A und B jeweils 3 von 5 Elfmeterschüssen durchgeführt haben, führt Spieler A mit 3 : 1.

Wie hoch ist die Wahrscheinlichkeit für ein Unentschieden, wenn Sie davon ausgehen, dass A_2 und B_2 stochastisch unabhängig sind und nicht vom Spielstand beeinflusst werden?

Schmierpapier:

Schmierpapier:

Aufgabe 4

Die Securance-Versicherung bietet u.a. Kfz-Versicherungen an. Von den versicherten Fahrzeugen liegen im `R`-Data Frame `Vers` die folgenden Merkmale vor:

Merkmal *Wert*:	Aktueller Wert des Autos (in Euro)	Spalte `Wert`
Merkmal *Kilometer*:	Aktueller Kilometerstand	Spalte `Kilometer`
Merkmal *Marke*:	Herstellermarke des Fahrzeugs	Spalte `Marke`

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

Ihnen liegt folgender `R`-Output vor:

```
> VAR=function(z){
+   sum((z-mean(z))^2)/length(z)
+   }
>
> table(Vers$Marke)/nrow(Vers)
  Audi    BMW Nissan   Opel Toyota     VW
  0.19   0.14   0.22   0.16   0.16   0.13
> nrow(Vers)
[1] 100
> Wert=Vers$Wert
> mean(Wert)
[1] 36396
> quantile(Wert,0.5,type=1)
  50%
36150
> auto1=Vers[Vers$Marke=="BMW","Wert"]
> length(auto1)
[1] 14
> quantile(auto1,0.5,type=1)
  50%
38750
> auto2=Vers[Vers$Kilometer>40000,"Wert"]
> mean(auto2)
[1] 44067.72
> auto3=Vers[Vers$Wert>40000,"Kilometer"]
> mean(auto3)
[1] 167388.2
```

1. Wie viele Fahrzeuge beinhaltet der Datensatz der Securance-Versicherung?

2. Wie hoch ist der Anteil der Marke BMW unter allen versicherten Fahrzeugen?

3. Stellt der Befehl

```
> hist(Vers$Wert, prob=TRUE)
```

eine sinnvolle Operation dar? (mit <u>kurzer</u> Begründung)

4. Vervollständigen Sie die folgenden Aussagen:

   (a) Der Median des Merkmals *Wert* beträgt ________________ Euro.

   (b) _____ % der versicherten BMW sind höchstens 38750 Euro wert.

   (c) Ein Fahrzeug mit einem Kilometerstand von mehr als 40000 km hat durchschnittlich einen Wert von ________________ Euro.

5. Geben Sie den (die) `R`-Befehl(e) an, um in `R` die empirische Verteilungsfunktion des Merkmals *Kilometer* zu zeichnen.

6. Geben Sie den (die) R-Befehl(e) an, um den Variationskoeffizienten ($v_X = s_X/\bar{x}$) des Merkmals *Wert* zu berechnen.

Die Versicherung interessiert sich für die Zufallsvariable X: "Schadenshöhe (in Euro) bei einem 'Parkplatz'-Unfall" und nimmt an $X \sim N(\mu_X, \sigma_X^2 = 100)$.
Der R-Vektor `Schaden` enthält die Schadenshöhe von 50 zufällig ausgewählten 'Parkplatz'-Unfällen.

7. Mit folgendem Befehl wird ein Hypothesentest durchgeführt:

```
> t.test(Schaden, mu=250, alternative="two.sided", conf.level=0.95)
        One Sample t-test
data:  Schaden t = 1.1579, df = 49, p-value = 0.2525
alternative hypothesis: true mean is not equal to 250
95 percent confidence interval:
 243.3960 274.5606
sample estimates:
mean of x
 258.9783
```

   (a) Geben Sie Null- und Alternativhypothese des durchgeführten Tests an.

   (b) Treffen und begründen Sie die Testentscheidung bei $\alpha = 0.05$.

8. Vervollständigen Sie die folgende `R`-Funktion so, dass damit aus einer Stichprobe `stich` ein zweiseitiges Konfidenzintervall für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz `var` zum Niveau `alpha` berechnet werden kann.

```
Mittelwert_KI=function(        ,var,alpha){
Xn = mean(stich)
n =
lambda =
u = Xn-lambda*sqrt(var/n)
o = Xn+lambda*sqrt(var/n)
KI=c(u,o)
return(KI)}
```

9. Mit der vollständigen (und korrekten) Funktion `Mittelwert_KI` berechnen Sie

```
> Mittelwert_KI(Schaden,100,0.05)
[1] 256.2065 261.7501
```

Vergleichen Sie das Ergebnis mit dem `R`-Output aus Teilaufgabe 7. Warum unterscheiden sich die berechneten Konfidenzintervalle?

Aufgabe 1

1. $n_{13} = 0$: keiner der Schadensfälle wurde als hoch eingestuft und betrifft eine Hausratsversicherung.

2. - Merkmal X, Merkmalstyp: komperativ, Skalenniveau: Ordinalskala
   - Merkmal Y, Merkmalstyp: qualitativ, Skalenniveau: Nominalskala

3. Der Modus liegt bei der Merkmalsausprägung 'Hausrat'.

4. Modus: häufigste Ausprägung 'gering'
   Median: $260/500 = 0.52 \geq 0.5 \Rightarrow x_{(0.5)} =$ 'gering'.

5. Annahmen des Poisson-Prozesses:

   - Die Wahrscheinlichkeit für einen Ereigniseintritt in einem Intervall hängt von dessen Länge ab, nicht von dessen Lage.
   - Die Anzahl der Ereignisse in zwei sich nicht überlappenden Intervallen sind voneinander unabhängig.
   - Die Wahrscheinlichkeit für mehr als ein Ereignis in einem sehr kurzen Intervall ist vernachlässigbar gering.

6. (a) $S_1 \sim Pois(\lambda_1 = 1) \Rightarrow S_5 \sim Pois(\lambda_5 = 5)$

   (b) $P(S_5 > 5) = 1 - P(S_5 \leq 5) = 1 - F_{Pois}(5;\ 5) = 1 - 0.6160 = 0.3840$

   (c) $P(3 < S_5 < 8) = F_{Pois}(7;\ 5) - F_{Pois}(3;\ 5) = 0.8666 - 0.2650 = 0.6016$

   (d) Erwartungswert: $E[S_5] = \lambda_5 = 5$,
   Median: $F_{Pois}(4;\ 5) = 0.4405 < 0.5$ und $F_{Pois}(5;\ 5) = 0.6160 \geq 0.5$, so dass der Median bei $s_{(0.5)} = 5$ liegt
   Modi: $f_{Pois}(4;\ 5) = f_{Pois}(5;\ 5) = 0.1755 > f_{Pois}(n;\ 5) \forall n \in \mathbb{N}_0 \backslash \{4, 5\}$, d.h. 4 und 5 sind die beiden Modi.

Aufgabe 2

1. $P(X > 3) = 1 - F_{Exp(1/2)}(3) = 0.22.$

2. (a)

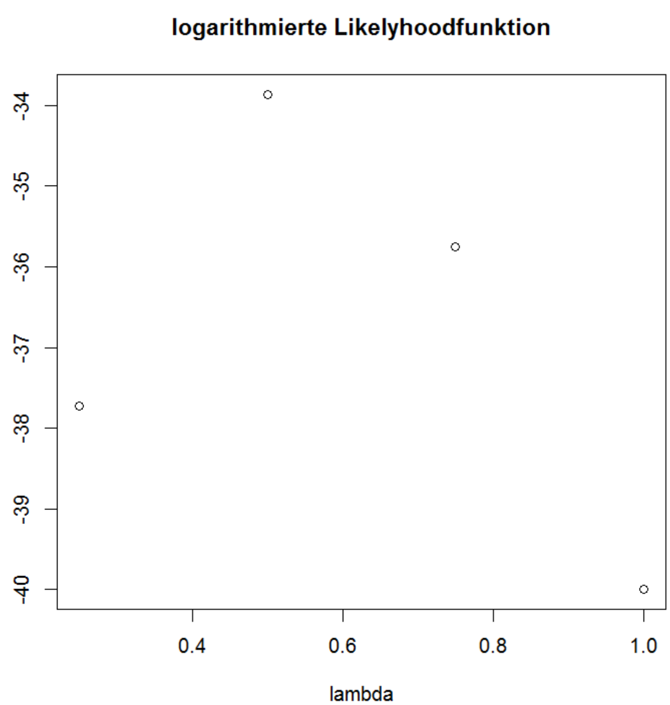
$$\begin{aligned} L(\lambda, x_i) &= \prod_{i=1}^{n} f(\lambda, x_i) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) \\ &= \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i). \\ \ln L(\lambda, x_i) &= n \ln \lambda - \lambda \sum_{i=1}^{n} x_i. \end{aligned}$$

   (b)

$$\begin{aligned} \frac{\partial \ln L(\lambda, x_i)}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \stackrel{!}{=} 0. \\ \Leftrightarrow & \quad \frac{n}{\lambda} = \sum_{i=1}^{n} x_i \\ \Leftrightarrow & \quad \hat{\lambda}_{ML} = 1/\bar{x}. \end{aligned}$$

3. (a)

λ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$\ln L(\lambda; xi)$	-37.72589	-33.86294	-35.75364	-40.00000



(b) $\hat{\lambda}_{ML} = 1/(1/20 \cdot 40) = 1/2$.

(c)
- Kritischer Bereich:
  Falls $T_n > KS$ ist H_0 abzulehnen.
- Kritische Schranke:
  95%-Quantil der $\chi^2(2 \cdot 20)$-Verteilung aus Tabelle ablesen: 55.76=KS.
- Berechnung der Teststatistik:
  $T_n = 2 \cdot 1/2 \cdot 40 = 40$.
- Testentscheidung: $T_n \not> KS$, d.h. H_0 nicht ablehnen.

4. Vervollständigen Sie folgende Aussagen:

   (a) $P(T > 40|\lambda_0 = 1/2) = 0.4702$ ist der p-Wert des obigen Hypothesentests. Auf Basis dieses Wertes ist die Nullhypothese <u>nicht abzulehnen</u>.

   (b) $P(T \neq 40|\lambda_0 = 1/2) = \underline{1}$.

5.
$$P(25.8 < T_n < 51.8) = F_{\chi^2(40)}(51.8) - F_{\chi^2(40)}(25.8) = 0.9 - 0.04 = 0.86.$$

Aufgabe 3

			Schussrichtung S			
			Links	Mitte	Rechts	
			$S=1$	$S=2$	$S=3$	
Position T	Links	$T=1$	$1/5$	$1/10$	$1/30$	$1/3$
	Mitte	$T=2$	$7/90$	$8/45$	$7/90$	$1/3$
	Rechts	$T=3$	$1/18$	$1/18$	$2/9$	$1/3$
			$1/3$	$1/3$	$1/3$	1

oder

			Schussrichtung S			
			Links	Mitte	Rechts	
			$S=1$	$S=2$	$S=3$	
Position T	Links	$T=1$	0.2	0.1	$0.0\overline{3}$	$1/3$
	Mitte	$T=2$	$0.0\overline{7}$	$0.1\overline{7}$	$0.0\overline{7}$	$1/3$
	Rechts	$T=3$	$1/18$	$1/18$	$2/9$	$1/3$
			$1/3$	$1/3$	$0.\overline{3}$	1

1. (a) Laut Angabe gilt

$$P(S=2|T=1) = \frac{3}{10} = \frac{P(S=2 \cap T=1)}{P(T=1)},$$

durch Umformen dann $1/10$ berechnen und $8/45$ erschließen. Es gilt:

$$P(T=2 \cap S=1) + P(T=2 \cap S=2) + P(T=2 \cap S=3) = \frac{1}{3}$$

und daher mit der Angabe

$$2P(T=2 \cap S=1) = \frac{1}{3} - (T=2 \cap S=2) = \frac{1}{3} - \frac{8}{45} = \frac{7}{45},$$

also

$$P(T=2 \cap S=1) = P(T=2 \cap S=3) = \frac{7}{90}$$

(b) Aus Tabelle

$$P(T=3|S=1) = \frac{P(T=3 \cap S=1)}{P(S=1)} = \frac{1/18}{1/3} = \frac{1}{6}$$

Die Ereignisse $T=3$ und $S=1$ sind stochastisch unabhängig, wenn gilt:

$$\frac{1}{6} = P(T=3|S=1) \overset{!}{=} P(T=3) = \frac{1}{3}$$

Die Ereignisse sind nicht stochastisch unabhängig.

(c) Für p_X gilt:

$$p_X = 1 - \left(\underbrace{P(S = 1 \cap T = 1) + P(S = 2 \cap T = 2) + P(S = 3 \cap T = 3)}_{\text{"Torwart hält den Schuss"}} \right).$$

Damit ist für Werte aus Vorgabe

$$p_X = 1 - (\frac{1}{18} + \frac{1}{18} + \frac{2}{9}) = \frac{2}{3},$$

und für korrekt berechnete Werte aus TA 1.

$$p_X = 1 - (\frac{1}{5} + \frac{8}{45} + \frac{2}{9}) = \frac{2}{5}.$$

2. (a) Die Versuche müssen voneinander unabhängige Bernoulli-Experimente mit Erfolgswahrscheinlichkeit p_X sein. Alternativ: Stichprobe mit Zurücklegen.

   (b) $X_5 \sim \text{Bin}(n = 5, p = 0.4)$ – die Parameter sind also $n = 5$ und $p = 0.4$

   (c)

$$\begin{aligned} P(X_5 \geq 3) = 1 - P(X_5 \leq 2) = 1 - F_{Bin(n=5,p=0.4)}(2) = \\ = 1 - 0.68256 = 0.31744 \end{aligned}$$

   Richtiges Ergebnis bei anderen Parametern 0.3993

   (d) Erwartungswert einer binomialverteilten Zufallsvariablen:

$$\mathbb{E}[X_5] = np = 5 \cdot 0.4 = 2$$

   Richtiges Ergebnis bei anderen Parametern 1.25, 3.6 oder 2.25

3. (a)

$$P(B_2 = 2) = \binom{2}{2} 0.5^2 (1 - 0.5)^0 = 1 \cdot 0.25 \cdot 1 = 0.25$$

   (b)

$$P(A_2 = 0) = \binom{2}{0} 0.4^0 (1 - 0.4)^2 = 1 \cdot 1 \cdot 0.36 = 0.36$$

   (c) Für ein Unenschieden bei 2 verbleibenden Versuchen muss Spieler B zweimal treffen, und Spieler 1 darf nie verwandeln. Gesucht ist $P(B_2 = 2 \cap A_2 = 0)$, und wegen stochastischer Unabhängigkeit von A_2 und B_2 gilt

$$P(B_2 = 2 \cap A_2 = 0) = P(B_2 = 2)P(A_2 = 0) = 0.25 \cdot 0.36 = 0.09$$

Aufgabe 4

1. 100
2. 0.14
3. ja, *Wert* ist quasi-stetiges Merkmal, damit Histogramm zur Darstellung relativer Häufigkeiten geeignet
4. (a) 36150

   (b) 50 %

   (c) 44067.72
5. `plot(ecdf(Vers$Kilometer))`
6. `sqrt(VAR(Wert))/mean(Wert)`
7. (a) $H_0: \ \mu_X = \mu_0 = 250$ gegen $H_1: \ \mu_X \neq \mu_0$

   (b) Nullhypothese wird nicht abgelehnt, da p-Wert größer als vorgegebenes Signifikanzniveau von $\alpha = 0.05$
8.
```
Mittelwert_KI=function(stich,var,alpha){
Xn = mean(stich)
n = length(stich)
lambda = qnorm(1-alpha/2)
u = Xn-lambda*sqrt(var/n)
o = Xn+lambda*sqrt(var/n)
KI=c(u,o)
return(KI)}
```
9. Funktion `t.test` berechnet KI für Mittelwert bei unbekannter Varianz, d.h. mit Stichprobenvarianz und Quantilen der t-Verteilung. Funktion `Mittelwert_KI` verwendet "wahre" Varianz (die von der Stichprobenvarianz i.d.R. abweicht) und somit Quantile der Normalverteilung