Name Matr.-Nr. Platznr.

## Aufgabe 1 *(15 Punkte )*

*Die Aufgabe besteht aus vier Teilaufgaben.*

Beantworten Sie die folgenden Fragen indem Sie JA oder NEIN zuordnen. Begründen Sie die Antworten! (Für unbegründete korrekte Antworten gibt es keine Punkte)

1. Hat der Parameter $\pi$ der Binomialverteilung bei konstantem Parameter $n$ Einfluss auf die Genauigkeit der Binomialverteilungsapproximation mit der Normalverteilung?

2. Sind bei 1000 unabhängigen Tests zum Niveau $\alpha = 0.05$ mit einer Wahrscheinlichkeit von 95 Prozent mindestens 38 Ablehnungen der Nullhypothese zu erwarten, wenn die Nullhypothese immer "wahr" ist?

   *Lösung:* Ein 95%–Prognoseintervall für die Anzahl der Ablehnungen der Nullhypothese bei 1000 unabhängigen Tests zum Niveau $\alpha = 0.05$ ist gegeben durch das Intervall

   $$\left[50 - 1.6445 \cdot \sqrt{1000 \cdot 0.05 \cdot 0.95}\,;\, \infty\right] \approx [38.67; \infty] \ ,$$

   wenn die Nullhypothese "in Wahrheit" immer angenommen werden müsste.

3. Wird das Prognose–Intervall für die Anzahl der Erfolge bei $n$ unabhängigen Wiederholungen eines Bernoulli–Versuchs umso breiter, je größer $n$ wird?

   *Lösung:* JA
   Es gilt:
   $$n\,\pi - \tau\sqrt{n\,\pi\,(1-\pi)} \leq y \leq n\,\pi - \tau\sqrt{n\,\pi\,(1-\pi)}$$
   Je größer $n$ ist, desto größer ist $n\,\pi$ und auch $\tau\sqrt{n\pi\,(1-\pi)}$ MAX 4 Punkte: kein Punkt ohne Begründung. 2 Punkte für das Aufstellen des Prognoseintervalls, 1 Punkte für das Betrachten der Anzahl und den richtigen Schluß. 1 Punkt für die Erwähnung der Approximation.

4. Ist die Länge des Konfidenzintervall für den Erwartungswert von normalverteilten Zufallsvariablen bei unbekannter Varianz immer breiter als bei bekannter Varianz?

   *Lösung:* NEIN. Zwar ist der Wert des Fraktils der t–Verteilung im betragsmäßig größer als das der Normalverteilung und somit das Intervall bei unbekannter Varianz größer, wenn die geschätzte Varianz nahe der wahren Varianz liegt. Aber es kann durchaus vorkommen (zwar selten), daß die Varianzschätzung sehr weit neben der wahren Varianz liegt (in unserer Betrachtung viel kleiner) und somit das Konfidenzintervall sehr klein wird. Wie gesagt, ein seltenes Ereignis, aber möglich.

MAX 4 Punkte: Es gibt jeweils ein Punkt für die t–Verteilung (ungekannte Varianz) und die Normalverteilung (bekannte Varianz), 2 Punkte für die Antwort NEIN, aber selten kleiner.

## Aufgabe 2 *(15 Punkte)*

*Die Aufgabe besteht aus drei Teilaufgaben.*

In einem Biotechnologie–Labor wurde durch Einsatz ionisierender Strahlung das bisher unbekannte Bakterium *Randomia variabilis* (**RV**) erzeugt.

Aus den zu dem neuen Organismus durchgeführten Studien ist Folgendes bekannt:

- Durchschnittlich zeigt jede zweite Kultur des RV–Bakteriums UV–Fluoreszenz (**U**).
- Im Schnitt stellen 80% aller RV–Bakterienkulturen das Wachstum ein, wenn das Nährmedium mit Zink versetzt wird, d.h. diese RV–Bakterienkulturen sind nicht Zink–resistent (**Z**).

Um eine Kultur verkaufen zu können, sollte sie UV-Fluoreszenz zeigen und keine Wachstumshemmung durch Zink aufweisen (Zink-Resistenz). Vom Labor erfahren Sie, dass die Wahrscheinlichkeit, dass eine Kultur des neuen Bakteriums *mindestens eins* der gewünschten Merkmale (Zink-Resistenz oder UV-Fluoreszenz) besitzt, 60% beträgt.

MAX 1 Punkte. 1 Punkt für saubere Notationen.

1. Berechnen Sie die Wahrscheinlichkeit, dass eine zufällig ausgewählte Kultur beide gewünschten Merkmale (Zink–Resistenz und UV–Fluoreszenz) aufweist!

   *Lösung:* $F$ : Die Kultur zeigt Fluoreszenz, $Z$ : Die Kultur ist zinkresistent.

$$\begin{aligned}
\mathrm{P}(F) &= \frac{1}{2}, \\
\mathrm{P}(\overline{Z}) &= \frac{8}{10}, \\
\mathrm{P}(F \cup Z) &= \frac{6}{10}, \\
\\
\mathrm{P}(F \cap Z) &= \mathrm{P}(F) + \mathrm{P}(Z) - \mathrm{P}(F \cup Z) \\
&= \frac{1}{2} + \left(1 - \frac{8}{10}\right) - \frac{6}{10} \\
&= \frac{5+2-6}{10} \\
&= \frac{1}{10} = 0,1.
\end{aligned}$$

   Die Wahrscheinlichkeit, dass eine Kultur beide Merkmale aufweist, ist 0,1.

   MAX 5 Punkte. 3 Punkte für die Werte aus dem Text, 1 Punkt für die Formel, 1 Punkt für die Rechnung

Leider stellt sich heraus, dass durchschnittlich drei von zehn RV–Bakterienkulturen giftige Stoffe (**G**) bilden, die tödlich für Säugetiere aller Art sind. Weiter ist bekannt, dass im Schnitt 70% der RV–Bakterienkulturen UV–fluoreszent sind, wenn die RV–Bakterienkulturen nicht tödlich sind.

2. Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte RV–Kultur unter UV-Licht fluoresziert *und* nicht tödlich auf Säugetiere wirkt?

   *Lösung:* Nach dem Satz von Bayes gilt:

   $T$: Bakterienkultur wirkt tötlich.

$$\begin{aligned}
P\left(F \mid \overline{G}\right) &= \frac{P\left(F \cap \overline{G}\right)}{P\left(\overline{G}\right)} \\
P\left(F \mid \overline{G}\right) \cdot P\left(\overline{G}\right) &= P\left(F \cap \overline{G}\right) \\
\frac{7}{10} \cdot \frac{7}{10} &= P\left(F \cap \overline{G}\right) \\
\frac{49}{100} &= P\left(F \cap \overline{G}\right) \\
0,49 &= P\left(F \cap \overline{G}\right)
\end{aligned}$$

   Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Kultur des Bakteriums nicht tötlich ist und fluoresziert, beträgt $0,49$. MAX 4 Punkte. 2 Punkte für Satz von Bayes, 2 Punkte für die Berechnung.

3. Wie groß ist die Wahrscheinlichkeit, dass eine tödliche RV–Bakterienkultur durch UV–Licht zu Fluoreszenz angeregt werden kann?

   *Lösung:* Nach dem Satz der totalen Wahrscheinlichkeit gilt:

$$\begin{aligned}
P(F) &= P\left(F \mid \overline{T}\right) \cdot P\left(\overline{T}\right) + P(F \mid T) \cdot P(T) \\
\frac{P(F) - P\left(F \mid \overline{T}\right) \cdot P\left(\overline{T}\right)}{P(T)} &= P(F \mid T) \\
\frac{\frac{1}{2} - \frac{7}{10} \cdot \frac{7}{10}}{\frac{3}{10}} &= P(F \mid T) \\
\frac{\frac{1}{2} - \frac{49}{100}}{\frac{3}{10}} &= P(F \mid T) \\
\frac{1}{30} &= P(F \mid T) \\
0,033 &\approx P(F \mid T)
\end{aligned}$$

   Mit einer Wahrscheinlichkeit von etwa $3\%$ kann eine tötliche Bakterienkultur zum Leuchten angeregt werden. MAX 5 Punkte. 2 Punkte für die Formel (Satz der Totalen Wahrscheinlichkeit), 2 Punkte für die Werte aus dem Text, 2 Punkte für die Berechnung.

**Hinweis**: Bitte verwenden Sie die im Aufgabentext angegebenen Bezeichnungen (**U**, **Z**, **G**) für die Ereignisse.

## Aufgabe 3 (20 Punkte)

*Die Aufgabe besteht aus fünf Teilaufgaben.*

Da Paul seit Jahren nur Fruchtgummis einer Marke ißt, weis Paul, daß es drei verschiedenen Tütenabfüllungen gibt:

| Abfüllung | Anzahl der Fruchtgummis insgesamt | Anzahl der grünen Fruchtgummis |
|---|---|---|
| $A$ | 100 | 10 |
| $B$ | 100 | 15 |
| $C$ | 100 | 25 |

Die Anzahl der grünen Fruchtgummis interessieren Paul, weil er diese am liebsten ißt. Da Paul nett sein will, hat er sich vorgenommen, immer dann von seinen Fruchtgummis abzugeben, wenn er annimmt, daß eine Tüte keine 25 grünen Fruchtgummis enthält (Typ $A$ oder $B$).

Da der Tüteninhalt nicht von aussen sichtbar ist und die einzelnen Fruchtgummis stark aneinander kleben, ist es nicht möglich die Anzahl der grünen Fruchtgummis in einer Packung zu zählen (es sein den, man nimmt sie einzeln heraus).

Paul überlegt sich daher, die ersten 10 Fruchtgummis alleine zu essen und dabei die Anzahl der grünen Fruchtgummis zu zählen, um die Anzahl der grünen Fruchtgummis in der Tüte zu schätzen.

**Hinweis:** Beachten Sie die tabellierten Werte am Ende der Klausur!

1. Wie ist die Anzahl der grünen Fruchtgummis in den ersten 10 Fruchtgummis verteilt?

   *Lösung:* Sei $X$ die Anzahl der grünen Grauelche in den ersten 10 Fruchtgummis, $N = 100$ die Gesamtanzahl von Fruchtgummis in einer Tüte, $n = 10$ die Anzahl der gezogenen Fruchtgummis, $G$ die Gesamtanzahl der grünen Grauelche in einer Tüte. Dann ist $X$ hypergeometrisch verteilt

   $$P(X = x) = \frac{\binom{G}{x}\binom{100-G}{n-x}}{\binom{100}{10}}.$$

   MAX 4 Punkte: 2 Punkte für die Hypergeometrische Verteilung und 2 Punkte für die richtigen Parameterwerte.

2. Welches Schätzprinzip kann Paul zur Schätzung der Anzahl grüner Fruchtgummis in der Tüte benutzen? Worauf ist dabei zu achten?

   *Lösung:* Das Likelihood–Prinzip. Es wird der Parameter ausgewählt, der am plausibelsten ist. Dabei ist zu beachten, daß wir nur drei Parameterwerte für $G$ untersuchen müssen. MAX 4 Punkte: 2 Punkte für Likelihood und 2 Punkte für den eingeschränkten Parameterraum.

3. Geben Sie ausgehend von den ersten 10 Fruchtgummis eine Schätzfunktion für die Anzahl der grünen Fruchtgummis in der Tüte an.

*Lösung:* Aufstellen der Likelihood–Funktion

$$L\left(G\,||\,x_1,\ldots,x_{10}\right)=c\cdot\binom{G}{x}\binom{100-G}{10-x},$$

wobei $x=\sum x_i$ und $x_i=\begin{cases}1 & \text{ite Fruchtgummi ist gruen}\\ 0 & \text{sonst}\end{cases}$.

Der Schätzer für $G$ läßt sich dann schreiben als

$$\widehat{G}:=\operatorname{argmax}_G L\left(G\,||\,\sum x_i\right).$$

MAX 4 Punkte: 2 Punkte für den Übergang von der Wahrscheinlichkeitsfunktion zur Likelihoodfunktion, 1 Punkte für die Anstellung von $x_i$ und 1 Punkt für die Konstruktion des Schätzer (kann auch verbal erfolgen).

4. Paul hat nun 3 grüne Fruchtgummis in den ersten 10 Fruchtgummis gezählt. Gegeben Sie eine Schätzung für die Anzahl der grünen Fruchtgummis in der Tüte an. Falls Sie Aufgabenteil 3 nicht gelöst haben, begründen Sie Ihre Schätzung.

   *Lösung:* Berechnung der Likelihood für die drei möglichen Parameterwerte:

$$\begin{aligned}L\left(10\,||3\right)&=0.0518\\L\left(15\,||3\right)&=0.1178\\L\left(25\,||3\right)&=0.2305\end{aligned}$$

   Damit ist $\widehat{G}=25$. MAX 4 Punkte: 3 Punkte für die Berechnung der Likelihood/Wahrscheinlichkeit und ein Punkt für den Schätzwert.

5. Angenommen Paul entscheidet sich, dass Typ $C$ vorliegt (25 grüne Fruchtgummis in der Tüte), wenn in den ersten 10 Fruchtgummis mindestens 2 grüne sind. Wie groß ist die Wahrscheinlichkeit, daß Paul seinen Freunden Fruchtgummis anbietet, wenn in der Tüte wirklich 25 grüne Fruchtgummis sind.

   *Lösung:* Die falsche Entscheidung ist nur dann möglich, wenn $G=25$. Damit haben wird die Wahrscheinlichkeitsfunktion vollständig gegeben und müssen nur noch die Wahrscheinlichkeit für $x=\{0,1\}$ berechnen und aufsummieren.

$$\begin{aligned}P\left(X=0\right)&=\frac{\binom{25}{0}\binom{100-25}{10}}{\binom{100}{10}}\\&=0.2369\\P\left(X=1\right)&=\frac{\binom{25}{1}\binom{100-25}{9}}{\binom{100}{10}}\\&=0.0767\end{aligned}$$

   Damit ist $P\left(X<2\right)=0.3136$.

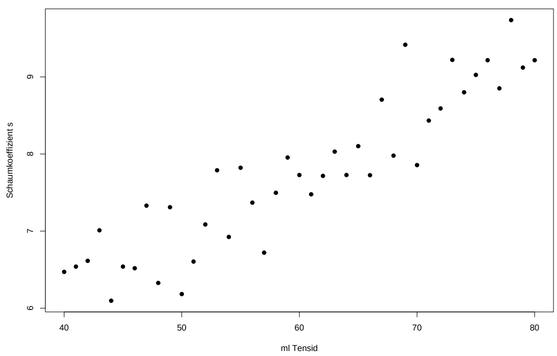
MAX 4 Punkte: 2 Punkte für die Idee und 2 Punkte für die Rechnung.

## Aufgabe 4 *(20 Punkte)*

*Diese Aufgabe besteht aus acht Teilaufgaben.*

Die Firma *BadeWonne* möchte ihren Verkaufsschlager *Cream* durch Zusatz des neuen Tensids T2a weiter verbessern. Aus Konsumentenbefragungen ist bekannt, daß die Kunden beim Baden besonderen Wert auf die Schaumbildung legen. Die Forschungsabteilung untersucht daher die Schaumbildung während eines Vollbades bei unterschiedlichen Mengen zugesetzten Tensids T2a. Die Schaumbildung wird dabei durch den sogenannten Schaumkoeffizienten $s$ (angegeben ohne Einheit) quantifiziert.

Die Untersuchungen umfassen 41 Experimente, wobei die zugesetzte Menge Tensid T2a [ml] von 40 bis 80 Milliliter reicht. Die Ergebnisse sind in der folgenden Graphik dargestellt.




Die leitende Chemikerin geht von einem linearen Zusammenhang zwischen Tensidmenge $t$ und Schaumkoeffizient $s$ aus. Sie stellt das lineare Modell

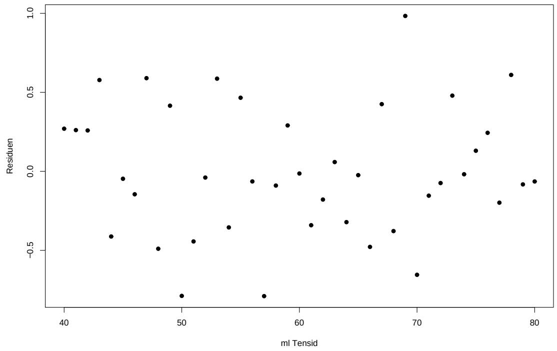
$$s_i\ (t) = \beta_0 + \beta_1 \cdot t_i + \varepsilon_i$$

auf, wobei $s_i$ der Schaumkoeffizient ist und $t_i$ die zugesetzte Menge Tensid T2a im i-ten Versuch.

1. Geben Sie die Modell– und Verteilungsannahmen an.

   *Lösung:* Unabhängig identisch normalverteilte Fehler.

MAX 1 Punkte. Als Ergebnisse der Regressionsanalyse ergab sich die Schätzung $\widehat{\beta_0} = 3.1222$ mit einer geschätzten Standardabweichung von 0.3341, $\widehat{\beta_1} = 0.0770$ mit einer geschätzten Standardabweichung von 0.0055 sowie ein Bestimmtheitsmaß von $R^2 = 0.8354$. Ausserdem lieferte das verwendete Statistikprogramm folgenden Residuenplot.




2. Ist die Annahme eines linearen Zusammenhangs aufgrund der Daten und der Regressionsergebnisse plausibel? Verwenden Sie dazu auch den Residuenplot.

   *Lösung:* JA. Im Residuenplot sind keine Unregelmäßigkeiten zu erkennen, die gegen einen linearen Zusammenhang sprechen.

MAX 1 Punkte.

3. Interpretieren Sie die berechneten Werte (Parameter, Bestimmtheitsmaß).

   *Lösung:* Der Intercept kann als Werte für $s$ interpretiert werden, wenn man das Tensid T2a nicht hinzufügt (zu beachten ist allerdings, daß die Versuche mit relativ hohen Tensidmengen durchgeführt wurde). Der Parameter $\beta_1$ gibt an, um wieviel der Schaumkoeffizient $s_i$ steigt, wenn die Tensidmenge um 1ml erhöht wird. Hier also 0.077 pro ml Tensid. Das Bestimmheitsmaß gibt an, wie gut das Modell zu den Daten paßt, wenn der Zusammenhang ein linearer Zusammenhang ist. Mit einem Wert von 0.83 und dem plausiblen linearen Zusammenhang (siehe Residuenplot) ist das Modell geeignet.

MAX 3 Punkte. Je 1 Punkt für $\beta_0$, $\beta_1$ und $R^2$. Die Geschäftsleitung will nun wissen, ob davon auszugehen ist, daß die Zugabe des sehr teuren Tensid T2a einen Anstieg des Schaumkoeffizienten zur Folge hat. Die Zielsetzung der Geschäftsleitung ist das Tensid T2a nur dann beizumischen, wenn pro ml Tensid T2a der Schaumkoeffizient $s$ um mehr als 0.05 steigt.

4. Formalisieren Sie die Fragestellung der Geschäftsleitung in einer Hypothese und einer Alternative. (Begründen)

   *Lösung:* Die zu überprüfende Hypothese lautet

$$\begin{aligned} H_0: \quad \beta_1 \leq & \quad 0.05, \\ H_1: \quad \beta_1 > & \quad 0.05. \end{aligned}$$

   Die Firma will sicher sein, daß das Tensid die gewünschte Wirkung hat. Damit ist die Hypothese abzulehnen, daß das das Tensid die gewünschte Wirkung nicht hat, denn nur die Wahrscheinlichkeit für den Fehler erster Art ist zu kontrollieren.

MAX 3 Punkte. 1 Punkt für die einseitigen Hypothesen, 1 Punkt für den richtigen Hypothesenwert und 1 Punkt für die richtige Richtung und deren Begründung.

5. Sollte unter der Zielsetzung der Geschäftsleitung die Hypothese zum Niveau $\alpha = 0.01$ oder zum Niveau $a = 0.1$ überprüft werden. (Begründen)

   *Lösung:* Das unternehmen will sicher sein, nicht unnötigerweise das Tensid beizumischen (Kostenfrage). Daher ist $\alpha = 0.01$ zu wählen.

MAX 1 Punkte.

6. Wie kann die Hypothese überprüft werden? Geben Sie die Methode, die dazu betrachtete Größe und deren Verteilung an.

   *Lösung:* Gaußtest auf Parameterwerte bei unbekannter Varianz. Die Prüfgröße ist

$$\frac{\widehat{\beta}_1 - 0.05}{\sqrt{\widehat{\text{var}}\left(\widehat{\beta}_1\right)}}$$

   t–verteilt mit $(n-2)$ Freiheitsgraden.

MAX 3 Punkte. 1 Punkte für die Prüfgröße, 1 Punkte für die Verteilung, 1 Punkte für den Gaußtest auf Parameterwerte (bei unbekannter Varianz).

7. Führen Sie die Überprüfung anhand der verfügbaren Werte zu einem Niveau von $\alpha = 0.05$ durch.

   *Lösung:* Die Prüfgröße realisiert sich zu

$$\frac{0.077 - 0.05}{0.0055} = 4.91.$$

   Die kritische Region ist $[-\infty, t_{0.95}(n-2)]$ mit $t_{0.95}(39) \approx \tau_{0.95} = 1.645$. Die Nullhypothese ist also abzulehnen.

MAX 3 Punkte. 1 Punkte für die Berechnung der Prüfgröße. 1 Punkte für die Angabe der kritischen Region, 1 Punkte für die Berechnung des Quantils. Die Approximation mit der Normalverteilung muss in der Lösung noch überprüft werden.

8. Interpretieren Sie das Ergebnis.

   *Lösung:* Die Firma kann davon ausgehen, daß das Tensid die gewünschte Wirkung hat, da die Nullhypothese abgelehnt werden konnte. MAX 1 Punkte.

Name Matr.-Nr. Platznr.

Zu Aufgabe:

Name Matr.-Nr. Platznr.

Zu Aufgabe:

Name Matr.-Nr. Platznr.

Zu Aufgabe: