

Matrikel-Nr.					
--------------	--	--	--	--	--

Modulprüfung

Prüfungsfach: Statistk II für Betriebswirte
 Prüfer: PD Dr. Frank Heyde
 Tag: 06. März 2014

Aufgabe	1	2	3	4	5	6
erreichbare Punkte	5	4	5	5	5	6
erreichte Punkte						

**Geben Sie zu allen Aufgaben einen nachvollziehbaren Lösungsweg an!
 Beantworten Sie die Fragen möglichst kurz und mit eigenen Worten!**

1. Aufgabe: Eine sächsische Molkerei füllt Milch in Tetrapacks ab. Die Füllmenge ist normalverteilt mit Standardabweichung $\sigma = 1 \text{ ml}$. In einer Stichprobe von 12 Tetrapacks wurden folgende Füllmengen gemessen:

499,7	501,3	499,1	500,6	500,2	499,5
500,0	499,1	500,5	499,6	498,7	501,7

- a) Bestimmen Sie zum Konfidenzniveau von 99 % ein zentrales Konfidenzintervall für die erwartete Füllmenge.
- b) Wie groß muss n mindestens sein, damit die Länge des 99 % Konfidenzintervalles höchstens $0,5 \text{ ml}$ ist?
- c) Mit Statgraphics wurde aus den Daten noch folgendes Testergebnis erzielt:

Tests for Normality for Füllmenge

<i>Test</i>	<i>Statistic</i>	<i>P-Value</i>
Shapiro-Wilk W	0,963671	0,779243

Welche Hypothesen werden getestet und wie lautet die Testentscheidung bei $\alpha = 0,1$?

2. Aufgabe: Im Rahmen der statistischen Qualitätskontrolle wird ein Posten mit höchstens 1,7% Ausschussanteil als gut angesehen. Ein Posten mit mehr als 5% Ausschussanteil ist hingegen ein schlechter Posten. Der Produzent fordert, dass ein guter Posten höchstens mit Wahrscheinlichkeit von 0,01 abgelehnt wird. Der Konsument fordert, dass ein schlechter Posten höchstens mit Wahrscheinlichkeit 0,05 angenommen wird.

a) Prüfen Sie, ob für die Annahmezahl $c = 14$ ein n bestimmbar ist, so dass die Forderungen des Produzenten und auch die des Konsumenten erfüllt sind. Falls ein n bestimmbar ist, dann geben Sie alle möglichen (n, c) -Pläne an.

Hinweis: Nutzen Sie die Poissonapproximation.

b) Für die Annahmezahl $c = 16$ gibt es folgende Pläne, welche beide Forderungen (die vom Konsument und die vom Produzent) erfüllen:

$$\begin{aligned} (n, c) &= (487, 16) \\ (n, c) &= (488, 16) \\ &\vdots \\ (n, c) &= (523, 16) \end{aligned}$$

Welchen dieser Pläne würde der Konsument bevorzugen, um sein Risiko kleinstmöglich zu halten? Begründen Sie kurz!

3. Aufgabe: In einer Studie soll der Erfolg von 3 verschiedenen Düngemitteln verglichen werden. Diese wurden auf 22 Erdbereifeldern verwendet. Dabei wurden folgende Erträge in Kilogramm erzielt. (Für die einzelnen Düngemittel sind die Erträge schon der Größe nach geordnet.)

Düngemittel 1	44	45	55	61	63	67	78		
Düngemittel 2	62	68	75	79	80	81	86	90	92
Düngemittel 3	72	85	89	100	101	121			

Die Erträge sind **nicht** normalverteilt. Testen Sie zum Niveau $\alpha = 0,01$, ob die erwarteten Erträge der 3 Düngemittel gleich sind oder sich signifikant voneinander unterscheiden.

4. Aufgabe: Für 29 PKWs wurden die Merkmale

X_1 - Alter,

X_2 - Leistung und

X_3 - Verbrauch erfasst.

Aus der Stichprobe erhält man folgende Schätzung der Korrelationen:

$$r_{X_1, X_2} = -0,18 \quad ; \quad r_{X_1, X_3} = 0,39 \quad \text{und} \quad r_{X_2, X_3} = 0,51.$$

a) Schätzen Sie die partielle Korrelation ($\rho_{X_2, X_3 | X_1}$) zwischen Leistung und Verbrauch unter Partialisierung des Alters.

b) Testen Sie, ob die partielle Korrelation zwischen Leistung und Verbrauch unter Partialisierung des Alter signifikant ($\alpha = 0,01$) größer als 0 ist.

5. **Aufgabe:** Für 29 PKWs wurden die Merkmale **Preis** Y (in €), gefahrene **Kilometer** X_1 , **Leistung** X_2 (in Ps) und **Alter** X_3 (in Jahren) erfasst. Aus den Daten erhält man das folgende Ergebnis.

Multiple Regression - Preis

Dependent variable: Preis

Independent variables:

Kilometer

Leistung

Alter

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	16646,2	771,963	21,5635	0,0000
Kilometer	-0,0157827	0,00253198	-6,23334	0,0000
Leistung	5,92975	5,19015	1,1425	0,2641
Alter	-907,708	44,4003	-20,4437	0,0000

- Wie lautet die geschätzte Regressionsfunktion?
- Welche der drei Einflussgrößen würden Sie am ehesten aus dem Modell entfernen? Begründen Sie Ihre Entscheidung kurz.
- Im Folgenden wurden für zwei weitere Regressionsmodelle die Ergebnisse mit Statgraphics erstellt.

Modell 1:

Simple Regression - Preis vs. Kilometer

Dependent variable: Preis

Independent variable: Kilometer

Linear model: $Y = a + b \cdot X$

Coefficients

	<i>Least Squares</i>	<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
Intercept	15966,7	1626,16	9,81863	0,0000
Slope	-0,0460363	0,00863454	-5,33165	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	1,93448E8	1	1,93448E8	28,43	0,0000
Residual	1,8374E8	27	6,8052E6		
Total (Corr.)	3,77188E8	28			

R-squared = 51,2868 percent

Standard Error of Est. = 2608,68

Modell 2:

Simple Regression - Preis vs. Alter

Dependent variable: Preis

Independent variable: Alter

Linear model: $Y = a + b \cdot X$

Coefficients

	<i>Least Squares</i>	<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
Intercept	15775,4	450,764	34,9971	0,0000
Slope	-1071,45	54,8556	-19,5323	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	3,52258E8	1	3,52258E8	381,51	0,0000
Residual	2,49299E7	27	923328,		
Total (Corr.)	3,77188E8	28			

- i) Bestimmen Sie für das Modell 2 das Bestimmtheitsmaß.
- ii) Bestimmen Sie für das Modell 2 die Schätzung für die Standardabweichung des Fehlers.
- iii) Welches der beiden Modelle würden Sie dem anderen Modell vorziehen? Begründen Sie Ihre Wahl kurz.

6. Aufgabe:

- a) Bei einer Mittelwertkarte mit Sollwert $a = 10$ ist die Gütefunktion an der Stelle $\mu = 11$ gleich 0,25, d.h. $g(11) = 0,25$.
 - i) Wie groß ist die erwartete Lauflänge, bis diese Abweichung vom Sollwert erkannt wird?
 - ii) Wie groß ist die Wahrscheinlichkeit für den Fehler 2. Art, d.h. für unterlassenen Alarm?
 - iii) Wie verändert sich $g(11)$, falls der Stichprobenumfang für jede Prüfung erhöht wird?
- b) Beschreiben Sie kurz die Funktionsweise eines 2-stufigen Stichprobenplanes mit: $n_1 = 50, n_2 = 50, c_1 = 0, c_2 = 4$ und $c_3 = 5$.
- c) Sei $F_{(X,Y)}(x, y)$ die gemeinsame Verteilungsfunktion des Zufallsvektors (X, Y) an der Stelle (x, y) mit $x, y \in \mathbb{R}$. Geben Sie den richtigenen Ausdruck für
 - i. $\lim_{x \rightarrow \infty} F_{(X,Y)}(x, y)$
 - ii. $\lim_{x \rightarrow -\infty} F_{(X,Y)}(x, y)$

an.

Zur Auswahl stehen dabei dabei folgende Ausdrücke:

$-1, 0, 1, -\infty, \infty, F_Y(y)$ und $F_X(x)$.