

Name _____

**Klausur zur Veranstaltung
Mathematisch-Naturwissenschaftliche Grundlagen**

**WS 15/16, erster Prüfungstermin
02.02.2016**

Dozentin: Prof. Dr. Ulrike Grömping

Erlaubte Hilfsmittel

- Uhr
- Zwei selbstbeschriebene Blätter
- Taschenrechner

ET: +1

Verboten und als Täuschungsversuch gewertet

- jegliche kommunikationsfähigen Geräte

Bitte schreiben Sie Ihre Ergebnisse auf leere Blätter,

wobei Sie den Lösungsweg wenigstens skizzieren sollten.

Kennzeichnen Sie jeweils deutlich die bearbeitete Aufgabe / Teilaufgabe,

und markieren Sie gegebenenfalls deutlich das Endergebnis.

Wichtig: Unbegründete Ergebnisse bringen keine Punkte, selbst wenn sie richtig sind!

Viel Erfolg!

Gesamtpunktzahl	72 Punkte
Sehr Gut-Garantie	ab 64 Punkte
Bestehensgarantie	ab 32 Punkte

Aufgabe	1	2	3	4	5	6	7
Max. Punktzahl	10	14	14	7	7	12	8
	9,5	14	14	7	7	12	8

Ich bin damit einverstanden,

dass meine Note mit Matrikel-Nr. in Moodle veröffentlicht wird.

71,5 Punkte
+1: 77,5

Sehr gut (1,0)

Unterschrift

Aufgabe 1 (Quantile und Boxplot)

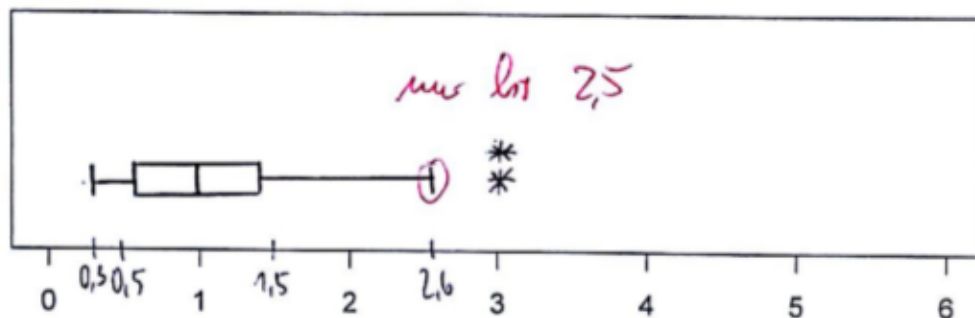
10

Diese Aufgabe basiert auf den Druckkosten pro Seite, die unter www.druckkosten.de für eine große Anzahl von Schwarzweiß-Druckern angegeben sind.

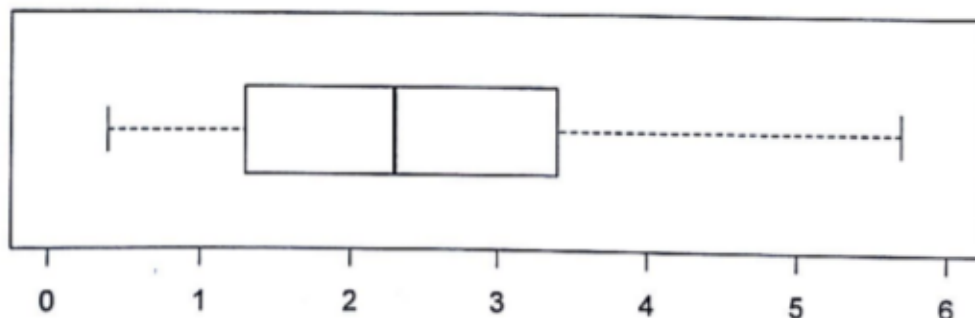
Die Tabelle zeigt Druckkosten pro Seite in ct für die 166 S/W-Drucker mit Anschaffungspreis ab 400€ bei Druck von 500 Seiten pro Monat über drei Jahre (sortierte Werte):

1	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3
21	0,4	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6
41	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6
61	0,7	0,7	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8
81	1	1	1	1	1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1
101	1,2	1,2	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3
121	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4
141	1,7	1,7	1,7	1,7	1,7	2	2	2	2	2	2	2	2	2	2	2	2	2	2
161	2,5	2,5	2,5	2,5	3	3													

- (i) Berechnen Sie den Median, die beiden Quartile (25% und 75%) und das 90%-Quantil – wichtig ist nicht nur das richtige Ergebnis, sondern auch die richtige Formell! 4
- (ii) Erstellen Sie einen Boxplot. 4
- Wenn Sie diesen in das Blanko-Feld unten eintragen, ist er für Teil (iii) direkt mit dem Plot weiter unten vergleichbar.
Machen Sie auf jeden Fall auch Ihre Überlegungen zu Ausreißern sichtbar.



- (iii) Der Boxplot unten zeigt die Druckkosten pro Seite für die 254 Drucker mit Anschaffungspreis unter 1000 € (darin ist der günstigere Teil der obigen Drucker enthalten). 2



Vergleichen Sie die beiden Verteilungen.

Aufgabe 2 (Beschreibung klassierter Daten)

14

Betrachtet wird das Merkmal „Anschaffungspreis in €“ für 300 Schwarz-Weiß-Drucker.

Die folgende Tabelle liegt vor (modifiziert zwecks einfacherer Rechnung):

j	Euro von über	Bis	Anzahl (n_j)	Klassenbreite b_j	rel. Hfgd $h_j = \frac{n_j}{n}$	Hfgkdichte $d_j = \frac{h_j}{b_j}$	emp. Vkgp. Fkt. $\hat{F}(x_j^0)$
1	0	400	144	400	0,48	0,0012	0,48
2	400	1000	93	600	0,31	0,000516	0,79
3	1000	2000	39	1000	0,13	0,00013	0,92
4	2000	4000	21	2000	0,07	0,000035	0,99
5	4000	7000	3	3000	0,01	~~0,0000033~~	1
		Gesamt	300			0,0000033	

- (i) Bestimmen Sie relative Häufigkeiten, Häufigkeitsdichten und die empirische Verteilungsfunktion. 6
- (ii) Zeichnen Sie das Histogramm der Verteilung des Merkmals Anschaffungspreis. (nicht zu ordentlich!) 2
- (iii) Berechnen Sie den Modus, den Median und den arithmetischen Mittelwert. 4
- (iv) Beschreiben Sie die Form der Verteilung unter Verwendung eines Fachbegriffs. 2

Aufgabe 3 (Regression I)

14

Der Zusammenhang zwischen zwei Merkmalen X und Y soll untersucht werden. Es sind fünf Paare von Merkmalswerten beobachtet worden.

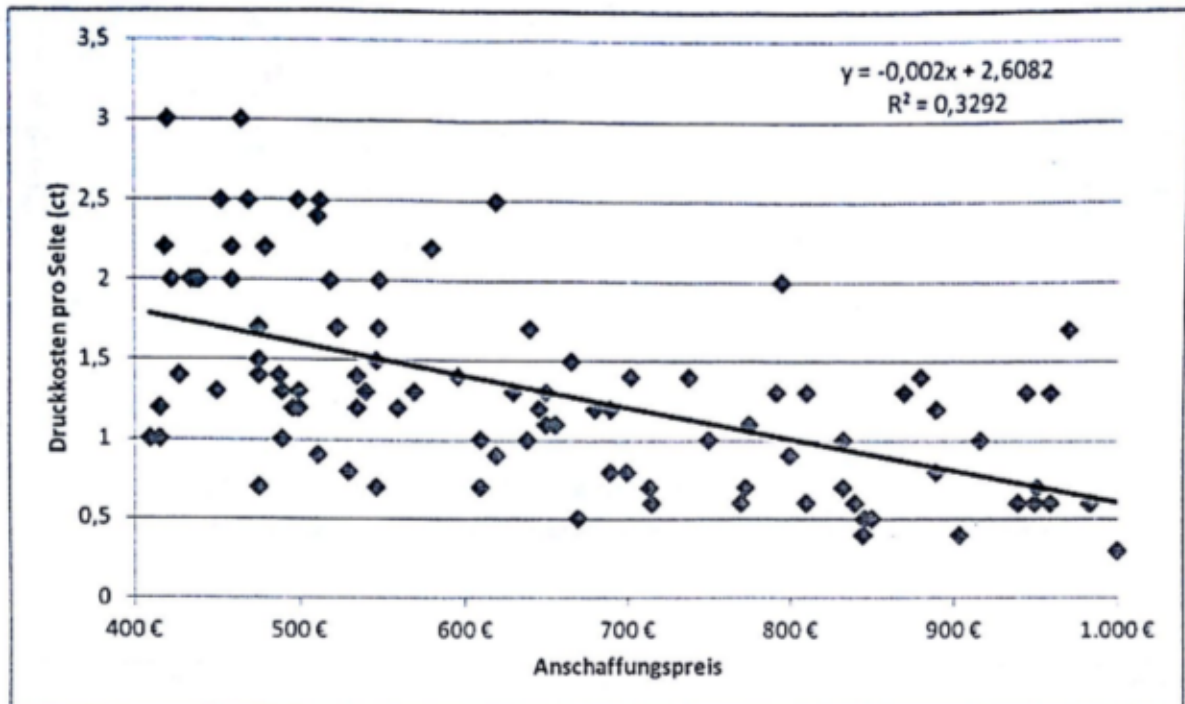
x_i	3	1	6	9	1
y_i	-3	-8	6	8	-3

- (i) Erstellen Sie ein Streudiagramm der Wertepaare, in das Sie später auch die Regressionsgerade einzeichnen können. 2
- (ii) Berechnen Sie die arithmetischen Mittelwerte und Varianzen sowie die Kovarianz der beiden Merkmale. 6
- (iii) Berechnen Sie die Regressionsgerade. 2
- (iv) Zeichnen Sie die Gerade in das Diagramm aus (i) ein. 2
- (v) Berechnen Sie die Korrelation zwischen den x - und y -Werten, und geben Sie das Bestimmtheitsmaß der Regression an. 2

Aufgabe 4 (Regression II)

7

Die Grafik zeigt eine Regressionsanalyse der Druckkosten pro Seite (y) auf den Anschaffungspreis.



- (i) Machen Sie aufgrund der angegebenen geschätzten Geradengleichung eine Vorhersage für die Druckkosten pro Seite bei einem Drucker, der nur 100€ in der Anschaffung kostet. 3
- (ii) Kommentieren Sie die Brauchbarkeit dieser Vorhersage. 2
- (iii) Geben Sie die Korrelation zwischen den beiden Merkmalen an. 2

Aufgabe 5 (Normalverteilung, Tabelle s. Beiblatt)

7

Das Gewicht [g] einer Sorte A4-Papier der Spezifikation 80g/m² schwankt gemäß einer Normalverteilung von $N(4.99, 0.0001)$.

- (i) Mit welcher Wahrscheinlichkeit wiegt ein Blatt mehr als 5g? 2
- (ii) Geben Sie ein Gewicht an, das von höchstens 10% der Blätter überschritten wird. 2
- (iii) Wie ist die Verteilung des Gewichts von 2500 Blatt dieses Papiers (so viel wird in einer Box verkauft), wenn wir Unabhängigkeit der Blattgewichte annehmen? 3

Wahrscheinlichkeitstabelle für Aufgabe 5Tabelliert ist $P(Z \leq z)$.Die Vorkommastellen von z finden Sie in den Zeilen,
die Nachkommastellen in den Spalten.

z	0.00	0.33	0.67
-3	0.0013		
-2	0.0228	0.0099	0.0038
-1	0.1587	0.0912	0.0478
-0	0.5000	0.3694	0.2525
+0	0.5000	0.6306	0.7475
+1	0.8413	0.9088	0.9522
+2	0.9772	0.9901	0.9962
+3	0.9987		

Quantiltabelle für Aufgaben 5 und 6Quantile der Normalverteilung und der t-Verteilung (Freiheitsgrade im Index der Überschrift)

p	z_p	$t_{3,p}$	$t_{4,p}$	$t_{5,p}$
0.005	-2.5758	-5.8409	-4.6041	-4.0321
0.01	-2.3263	-4.5407	-3.7469	-3.3649
0.025	-1.9600	-3.1824	-2.7764	-2.5706
0.05	-1.6449	-2.3534	-2.1318	-2.0150
0.1	-1.2816	-1.6377	-1.5332	-1.4759
0.9	1.2816	1.6377	1.5332	1.4759
0.95	1.6449	2.3534	2.1318	2.0150
0.975	1.9600	3.1824	2.7764	2.5706
0.99	2.3263	4.5407	3.7469	3.3649
0.995	2.5758	5.8409	4.6041	4.0321

Aufgabe 6 (Konfidenzintervalle, Tests)

12

Die Gleichmäßigkeit der Papierstärke von A4-Blättern wird in einem ad-hoc Verfahren untersucht, indem diese in vier gleich große Stücke zerschnitten werden und die Spannweite der Gewichte ermittelt wird (fiktiv!). Der Merkmalswert x_i für das i -te Blatt Papier ist also der Gewichtsunterschied zwischen dem leichtesten und schwersten Viertel.

Für eine Papiersorte hat sich in einer Stichprobe von 1000 Blatt ein mittlerer Gewichtsunterschied von $\bar{x} = 0.025\text{g}$ ergeben, bei einer empirischen Standardabweichung von $s = 0.003$.

- (i) Geben Sie je eine Punktschätzung für den erwarteten Gewichtsunterschied μ sowie für die (theoretische) Varianz σ^2 des Gewichtsunterschiedes an. 2
- (ii) Von einer Normalverteilung kann nicht ausgegangen werden. Geben Sie trotzdem ein zweiseitiges 95%-Konfidenzintervall für den erwarteten Gewichtsunterschied μ an, oder begründen Sie, warum dies nicht möglich ist. 4
- (iii) Sie möchten mit einem statistischen Test nachweisen, dass der erwartete Gewichtsunterschied mindestens 0.0245g beträgt. Stellen Sie ein geeignetes Testproblem (H_0 und H_1) auf. 2
- (iv) Führen Sie für das Testproblem aus (iii) einen geeigneten Test zum Signifikanzniveau 1% durch, und interpretieren Sie das Ergebnis. 4

Aufgabe 7 (χ^2 -Test)

8

Im ersten Semester der Studiengänge „Druck- und Medientechnik“ (DMT) und BWL wurde jeweils in einer Umfrage die „Angst vor Mathe“ erfragt (Daten fiktiv, Kategorien ++, +, o, -, -; vergrößert zu + o -).

		Angst vor Mathe			
		Viel (+)	Normal (o)	Wenig (-)	
Studiengang	BWL	20	25	10	55
	DMT	10	25	10	45
		30	50	20	100

Es soll die Nullhypothese untersucht werden, dass die Merkmale „Angst vor Mathe“ und „Studiengang“ (DMT oder BWL) unabhängig sind.

- (i) Berechnen Sie erwartete Häufigkeiten für die Kombinationen bei gegebenen Randhäufigkeiten. 4

		Angst vor Mathe			
		Viel (+)	Normal (o)	Wenig (-)	
Studiengang	BWL	16,5	22,5	11	55
	DMT	13,5	22,5	9	45
		30	50	20	100

- (ii) Berechnen Sie den Beitrag der linken oberen Zelle zur Teststatistik des χ^2 -Tests. 2
- (iii) Wie viele Freiheitsgrade hat die χ^2 -Verteilung, mit der Sie die Teststatistik vergleichen müssen? 1
- (iv) Aus einem Statistik-Programm erhalten Sie den p-Wert des Tests. Dieser lautet 0.3078. Kann die Nullhypothese zum Signifikanzniveau 5% abgelehnt werden? 1

Aufgabe 1

i) $n \cdot p = 0,25 \cdot 166 = 41,5 \leadsto k = 42$

$x_{0,25} = x_{(42)} = \underline{\underline{0,6}}$ $\uparrow$
nicht ganzz.,
aufrunden

$n \cdot p = 0,75 \cdot 166 = 124,5 \leadsto k = 125$

$x_{0,75} = x_{(125)} = \underline{\underline{1,4}}$ $\uparrow$
nicht ganzz.,
aufrunden

$n \cdot p = 0,5 \cdot 166 = 83$ (ganzz.)

$x_{0,5} = \frac{x_{(k)} + x_{(k+1)}}{2} = \frac{x_{(83)} + x_{(84)}}{2} = \frac{1+1}{2} = \underline{\underline{1}}$

$n \cdot p = 0,9 \cdot 166 = 149,4 \leadsto k = 150$

$x_{0,9} = x_{(150)} = \underline{\underline{2}}$ $\uparrow$
nicht ganzz.,
aufrunden

ii) Boxplot

Stämme: Boxbreite = $x_{0,75} - x_{0,25} = 1,4 - 0,6 = \underline{\underline{0,8}}$

rechter Stamm: $x_{0,75} - \text{Boxbreite} \cdot 1,5 = 0,6 - 1,5 \cdot 0,8 = -0,6$ (Stamm geht bis $x_{\min} = 0,3$)

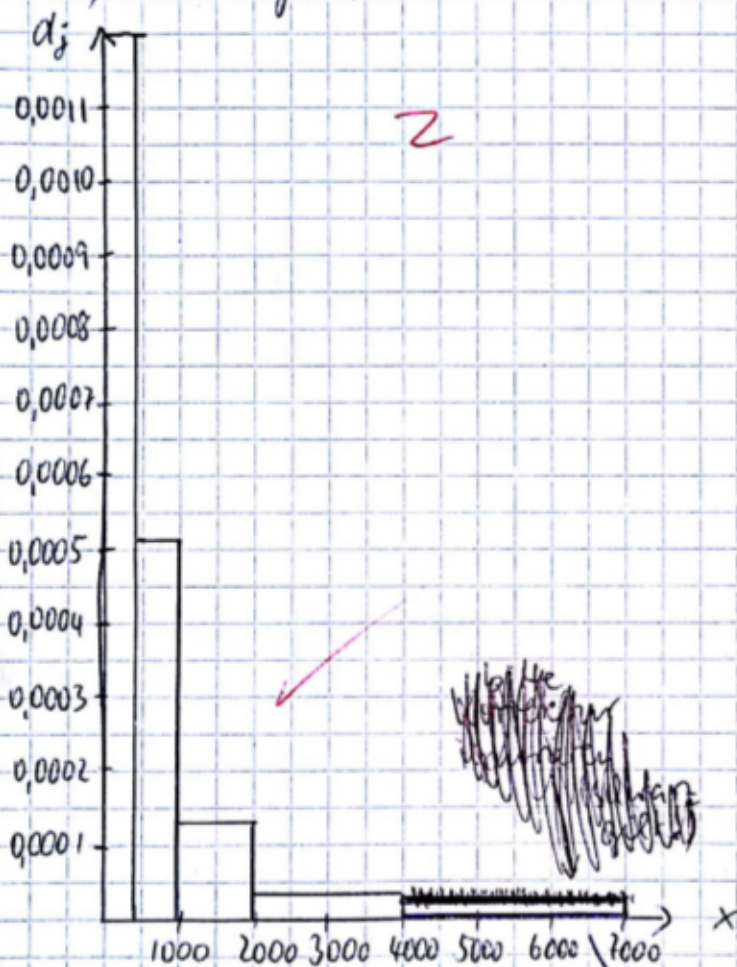
linker Stamm: $x_{0,25} + \text{Boxbreite} \cdot 1,5 = 0,6 + 0,8 \cdot 1,5 = 1,8$

iii) Der Großteil der Druckkosten von (iii) ist größer, als die Druckkosten von (ii). Die 50/50-Grenze von (iii) liegt bei ca. 2,3, wohingegen sie bei (ii) bei 1 liegt. Zudem ist die Verteilung bei (iii) erheblich breiter gestreut mit einem Maximum von ca. 5,8, während bei (ii) die "tuerste" Seite 3ct kostet. Die mittleren 50%, also zwischen 25% und 75% liegen bei (ii) zwischen 0,6 und 1,4; bei (iii) zwischen

ca. 1,3 und 3,4, was auch zeigt, dass die Druckkosten in (iii) höher sind als z in (ii).
9,5

Aufgabe 2

ii) Histogramm



~~Druckkosten = 10000~~

blauer Balken gilt!
 (sorry!)

(i) 6

iii)

$$x_{\text{mod}} = \frac{400 - 0}{2} = \underline{200} \text{ (Klasse 1)}$$

$$x_{0,5} = 400 + \frac{0,5 - 0,48}{0,31} \cdot 600 = \underline{438,7097} \text{ (Klasse 2)}$$

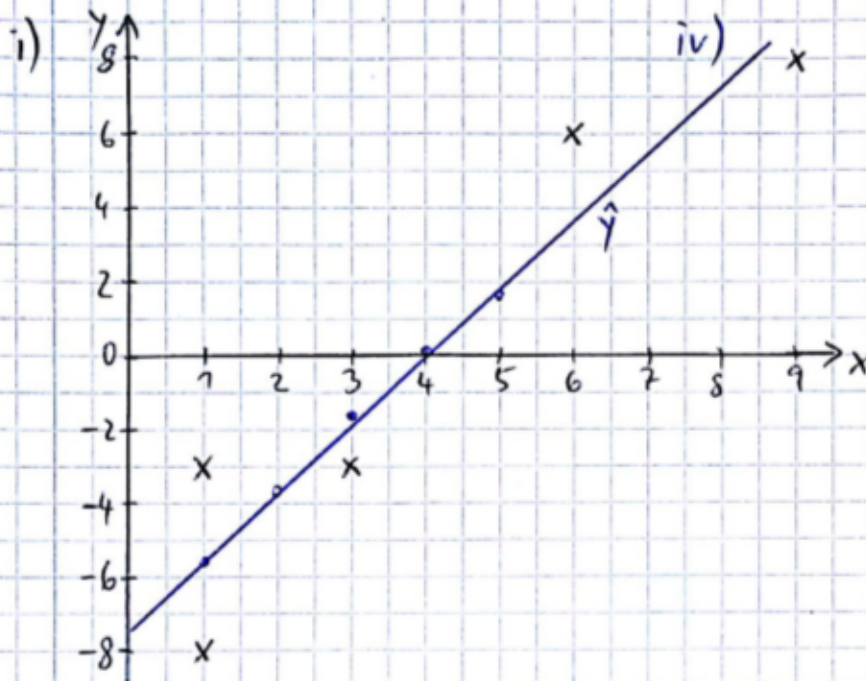
$$\begin{aligned} \bar{x} &= \sum_{i=1}^k m_j \cdot h_j \\ &= \left(\frac{400}{2} \cdot 0,48 \right) + \left(\frac{1400}{2} \cdot 0,31 \right) \\ &\quad + \left(\frac{3000}{2} \cdot 0,13 \right) + \left(\frac{6000}{2} \cdot 0,07 \right) \\ &\quad + \left(\frac{11000}{2} \cdot 0,01 \right) \\ &= 96 + 217 + 195 + 210 + 55 \\ &= \underline{773} \text{ (2. Klasse)} \end{aligned}$$

iv) Die Form der Verteilung ist linkssteil / rechtsschief.

14

Aufgabe 3

x_i	3	1	6	9	1
y_i	-3	-8	6	8	-3
$x_i - \bar{x}$	-1	-3	2	5	-3
$y_i - \bar{y}$	-3	-8	6	8	-3



(i) ✓

(ii) ✓

$$\text{ii) } \bar{x} = \frac{1}{5} \cdot (3 + 1 + 6 + 9 + 1) = \underline{4}$$

$$\bar{y} = \frac{1}{5} \cdot ((-3) + (-8) + 6 + 8 + (-3)) = \underline{0}$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{4} \cdot (1 + 9 + 4 + 25 + 9) = \underline{12}$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{4} \cdot (9 + 64 + 36 + 64 + 9) = \underline{45,5}$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) = \frac{1}{4} \cdot (3 + 24 + 12 + 40 + 9) = \underline{22}$$

$$\text{iii) } \hat{y} = \hat{a} + \hat{b}x \quad \hat{b} = \frac{s_{xy}}{s_x^2} = \frac{22}{12} = \underline{1,8\bar{3}} \quad \hat{a} = \bar{y} - \hat{b} \cdot \bar{x}$$

$$\underline{\underline{\hat{y} = (-7,3\bar{3}) + 1,8\bar{3}x}} \quad \underline{\underline{= 0 - 1,8\bar{3} \cdot 4}} \\ \underline{\underline{= -7,3\bar{3}}}$$

iv)

x	1	2	3	4	5
y	-5,5	-3,7	-1,8	0,002	1,84

$$v) \quad r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{22}{\sqrt{12} \cdot \sqrt{45,5}} = \underline{0,9415}$$

$$R^2 = r_{xy}^2 = 0,9415^2 = \underline{0,88645} \quad \checkmark$$

(14)

Aufgabe 4

$$i) \quad \hat{y} = -0,002 \cdot (100) + 2,6082 = \underline{2,4082} \quad \checkmark$$

ii) Die Schätzung von 2,4082 ct/Seite für einen Druckes mit 100€ Anschaffungspreis ist wenig brauchbar, da das Merkmal weit außerhalb des beobachteten Bereichs liegt und das Bestimmtheitsmaß von $R^2 = 32,92\%$ nur wenig Vertrauen in die Beziehung zwischen Druckkosten und Anschaffungspreis gibt. $\checkmark$

$$iii) \quad r_{xy} = \sqrt{R^2} = \sqrt{0,3292} = \underline{\pm 0,5738}$$

Die Steigung ist negativ, d. heißt die Korrelation muss ein negatives Vorzeichen haben, somit $\underline{r_{xy} = -0,5738} \quad \checkmark$

Aufgabe 5

$$i) \quad X \sim N(4,99; 0,0001)$$

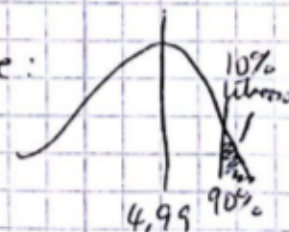
$$P(Y > 5) = 1 - \Phi\left(\frac{5 - 4,99}{\sqrt{0,0001}}\right) = 1 - \Phi(1) = 1 - 0,8413 = \underline{0,1587} \quad \checkmark$$

ii) ges: 90%-Quantil

$$z_{0,9} = 1,2816$$

Skizze:

$$X = 4,99 + \sqrt{0,0001} \cdot 1,2816 = \underline{5,002816} \quad \checkmark$$



$$iii) \quad Y = \sum_{i=1}^{2500} x_i \sim N(12475g; 0,25g^2)$$

$$\left(\sum_{i=1}^n x_i\right) \sim N(\mu \cdot n; \sigma^2 \cdot n) \quad \checkmark$$

(7)

Aufgabe 6

$$n = 1000 \quad \bar{x} = 0,025 \quad \hat{s} = 0,003$$

$$i) \quad \hat{\mu} \hat{=} \bar{x} = \underline{0,025} \quad \hat{\sigma}^2 \hat{=} s^2 = 0,003^2 = \underline{0,000009} \quad \checkmark$$

ii) KI 2-seitig, σ unbekannt, n groß, also $\hat{\sigma} \hat{=} s$

$$\bar{x} \pm \sqrt{\frac{\hat{\sigma}^2}{n}} \cdot z_{0,975} = 0,025 \pm \sqrt{\frac{0,003}{1000}} \cdot 1,96$$

$$\rightarrow \underline{[0,02481 ; 0,02519]} \quad \checkmark$$

iii) $H_0: \mu < 0,0245g$ $H_1: \mu \geq 0,0245g$ (✓) Stimmt zu meiner Hypothese

iv) $T = \sqrt{n} \frac{\bar{x} - \mu_0}{\hat{s}}$ verworfe H_0 g.d.w. $T > t_{n-1; 1-\alpha}$ man kann das "aber" wie in H_1 prüfen.

$$= \sqrt{1000} \frac{0,025 - 0,0245}{0,003} \quad T > t_{n-1; 1-\alpha}$$
$$= 5,2705 \quad t_T > z_{1-\alpha} \quad (da n \text{ groß})$$
$$z_{0,99} = 2,3263$$

$$5,2705 > 2,3263 \quad \text{w. A.}$$

Damit kann H_0 zum Signifikanzniveau von 1% verworfen werden. (12)

Aufgabe 7

$$i) \quad E_{11} = \frac{30 \cdot 55}{100} = 16,5 \quad E_{12} = \frac{50 \cdot 55}{100} = 27,5$$

Rest auffüllen:

$$E_{21} = 30 - 16,5 = 13,5 \quad E_{22} = 50 - 27,5 = 22,5$$

$$E_{13} = 55 - 27,5 - 16,5 = 11 \quad E_{23} = 20 - 11 = 9 \quad \checkmark$$

$$ii) \quad \text{Summand}_{11} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(20 - 16,5)^2}{16,5} = \underline{0,7424} \quad \checkmark$$

$$iii) \quad df = (r-1)(c-1) = (2-1)(3-1) = \underline{2} \quad \checkmark$$

$$iv) \quad p = 0,3078 \quad \alpha = 0,05 \quad \alpha \neq p$$

Die Nullhypothese, dass "Angst vor Mathe" und "Studiengang" unabhängig ist, kann nicht abgelehnt werden, da p nicht kleiner als α ist. (8)