Name:

Klausur zur Veranstaltung
Mathematisch-Naturwissenschaftliche Grundlagen

SoSe 2015, Zweiter Prüfungstermin
24.9.2015

Dozentin: Prof. Dr. Ulrike Grömping

Erlaubte Hilfsmittel

- Zwei selbstbeschriebene Blätter
- Taschenrechner
- Ein Tafelwerk Ihrer Wahl

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.

Viel Erfolg!

Gesamtpunktzahl	76 Punkte	Klausur: 2,3
Sehr Gut-Garantie	ab 64 Punkte	gesamt: 2,0
Bestehensgarantie	ab 32 Punkte	

Aufgabe	1	2	3	4	5	6	7
Max. Punktzahl	12	12	18	6	10	10	8
	10	11	17	2	5	2,5	5

52,5 Punkte

Ich bin damit einverstanden,
dass meine Note mit Matrikel-Nr. in Moodle veröffentlicht wird.

Unterschrift

Aufgabe 1 (Quantile, Boxplot) 12

Für 500 Nutzer/innen sind die Verweildauern in Sekunden auf einer Webseite protokolliert worden. Die folgende Häufigkeitstabelle gibt das Ergebnis wieder:

Sekunden	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Häufigkeit	**44**	**56**	**38**	**33**	**21**	**19**	**10**	**19**	**14**	**13**	**15**	**5**	**12**	**7**	**11**
Kumuliert	*44*	*100*	*138*	*171*	*192*	*211*	*221*	*240*	*254*	*267*	*282*	*287*	*299*	*306*	*317*

Sekunden	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**≥30**
Häufigkeit	**7**	**7**	**8**	**11**	**2**	**7**	**1**	**1**	**4**	**2**	**2**	**6**	**3**	**4**	**118**
Kumuliert	*324*	*331*	*339*	*350*	*352*	*359*	*360*	*361*	*365*	*367*	*369*	*375*	*378*	*382*	*500*

Die Zeile „Kumuliert" kann Ihnen helfen, die richtigen geordneten Werte mit wenig Arbeit zu finden!

(i) Was ist das Merkmal, und wie groß ist n? 2
Hinweis: Wenn Sie für n 30, 60 oder 90 herausbekommen, denken Sie noch einmal gründlich nach!

(ii) Berechnen Sie den Median, die beiden Quartile (25% und 75%), und das 0.1%-Quantil. 4

(iii) Erstellen Sie einen Boxplot 4
(volle Punktzahl nur,
wenn Ihre Überlegungen zu etwaigen Ausreißern erkennbar sind).

(iv) Können Sie den Modus x_{mod} angeben? 2
(Versuchen Sie es, und diskutieren Sie ggf. etwaige Probleme.)

Aufgabe 2 (Beschreibung klassierter Daten) 12

Betrachtet werden ΔE-Werte (räumliche Abweichungen im Farbraum Lab) für 100 Proben in Abweichung von ihrer Vorlage.
(Für die Bearbeitung der Aufgabe ist ein Verständnis des Lab-Farbraums nicht erforderlich.)

Die folgende Tabelle liegt vor:

j	ΔE Von über	ΔE bis	Anzahl (n_j)	relative Häufigkeit $h_j = n_j / n$	Klassenbreite $b_j = x_j^o - x_j^u$	Häufigkeitsdichte $d_j = h_j / b_j$	Verteilungsfunktion $F(x_j^o)$	
1	0	1	21	0,21	1,0	0,21	0,21	
2	1	1.5	18	0,18	0,5	0,36	0,39	← Modus
3	1.5	2	16	0,16	0,5	0,32	0,55	← Median
4	2	3.5	35	0,35	1,5	$0,2\overline{3}$	0,90	
5	3.5	5.5	10	0,10	2,0	0,05	1,00	
		Gesamt	100	1				

(i) Bestimmen Sie relative Häufigkeiten, Häufigkeitsdichten und die empirische Verteilungsfunktion. 6

(ii) Zeichnen Sie das Histogramm der Verteilung des Merkmals (nicht zu ordentlich!). 2

(iii) Berechnen Sie den Modus und den Median. 3

(iv) Kommentieren Sie die Form der Verteilung. 1

Aufgabe 3 (Regression) 18

Die folgenden Wertepaare sind beobachtet worden:

x_i	2	4	-2	0	1
y_i	-2	2	-9	-6	0

(i) Erstellen Sie ein Streudiagramm der Wertepaare, in das Sie später auch die Regressionsgerade einzeichnen können. 2

(ii) Berechnen Sie die Mittelwerte, Varianzen und die Kovarianz s_{xy}. 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

(vi) Es wird eine Vorhersage für y benötigt, wenn x den Wert 20 annimmt. Bestimmen Sie diese aufgrund der Geraden. 2

(vii) Diskutieren Sie die Qualität der Vorhersage aus (vi). 2

Aufgabe 4 (Binomialwahrscheinlichkeiten) 6

An einem Kopierer tritt das Problem auf, dass ein ausgegebenes Blatt mit Wahrscheinlichkeit 5% leer ist, statt den gewünschten Aufdruck zu zeigen. Es werden 8 brauchbare Kopien benötigt. Mit welcher Wahrscheinlichkeit sind unter 9 Kopien weniger als 8 brauchbare dabei ?

Achtung, bei geschicktem Vorgehen müssen Sie nur wenige Einzelwahrscheinlichkeiten berechnen!

Hinweise:

Für $y \in \{0,1,\ldots,n\}$: $P(Y=y)=\binom{n}{y}p^y(1-p)^{n-y}$,

und $\binom{n}{0}=\binom{n}{n}=1$, $\binom{n}{1}=\binom{n}{n-1}=n$.

Aufgabe 5 (Normalverteilung; Wahrscheinlichkeiten und Quantile s. Beiblatt) 10

Das Auftragsvolumen an einem Tag sei normalverteilt mit Erwartungswert 2000€ und Varianz σ^2=10000€^{2}.

(i) Berechnen Sie die Wahrscheinlichkeit, dass an einem Tag Aufträge von weniger als 1750€ Volumen entstehen. 2

(ii) Berechnen Sie die Wahrscheinlichkeit für ein Auftragsvolumen zwischen 1850€ und 2150€. 2

(iii) Sie möchten das Auftragsvolumen ermitteln, das mit nur 5% Wahrscheinlichkeit überschritten wird. An Tagen, an denen das Auftragsvolumen mindestens so hoch ist, gibt es für alle Angestellten eine kleine Überraschung. Ab welchem Auftragsvolumen ist dies der Fall? 2

(iv) Unterstellen Sie, dass die Auftragsvolumina der Tage eines Monats unabhängig sind (evtl. etwas gewagt). Wie ist unter dieser Voraussetzung das monatliche Auftragsvolumen verteilt (rechnen Sie mit 25 Arbeitstagen)? 2

(v) Bestimmen Sie die Wahrscheinlichkeit, dass das monatliche Auftragsvolumen 49750€ überschreitet. 2

Aufgabe 6 (Konfidenzintervalle; Quantile s. Beiblatt) 10

Es soll untersucht werden, wie groß der Markt für ein neues Produkt ist.
Als Zielgruppe wurden 25 bis 45jährige Männer identifiziert. Es werden 10000 25 bis 45jährige Männer befragt, ob sie an diesem Produkt Interesse haben (Antwortmöglichkeiten nur JA oder NEIN). 900 Männer antworten mit JA.

(i) Schätzen Sie den Anteil der 25 bis 45jährigen Männer, die sich für das Produkt interessieren (Punktschätzung). 2

(ii) Geben Sie ein zweiseitiges 95%-Konfidenzintervall für diesen Anteil an. 4

(iii) Die Firmenchefin ist sehr konservativ und möchte nur ein neues Produkt auf den Markt bringen, wenn sich *mindestens* 4% der Männer dafür interessieren. Welches Konfidenzintervall könnte der Marketingchef ihr vorlegen, wenn er sie überzeugen möchte? Bestimmen Sie auch dieses (ebenfalls 95%-Niveau), und erklären Sie den Unterschied zu (ii). 4

Aufgabe 7 (Testen und p-Werte) 8

Das Auftragsvolumen an einem Tag sei normalverteilt mit Erwartungswert μ (in €) und unbekannter Varianz σ^2. Es wurden Auftragsvolumina über 25 Tage aufgezeichnet; diese werden als unabhängig angenommen. Das durchschnittliche Auftragsvolumen der 25 Tage betrug 2100€, die empirische Varianz s^2 betrug 8100€^{2}.

(i) Stellen Sie ein Testproblem auf, mit dem nachgewiesen werden kann, dass das erwartete Auftragsvolumen mehr als 2000€ beträgt. 2

(ii) Führen Sie einen geeigneten Test für das Testproblem aus (i) durch, wobei Sie das Signifikanzniveau 5% ansetzen sollen. 4

(iii) Was können Sie aufgrund von (ii) über den p-Wert des Tests sagen? 2

①

i) Merkmal = Verweildauer in Sekunden

$n = 500$

2

ii.) $x_p = x_{0,5} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(\frac{500}{2})} + x_{(\frac{500}{2}+1)}}{2}$

$= \frac{x_{(250)} + x_{(251)}}{2} = \frac{9 + 9}{2} = \underline{\underline{9}}$ ✓

$x_p = x_{0,25} = \frac{x_{(k)} + x_{(k+1)}}{2}$ NR: $k = n \cdot p = 500 \cdot 0,25 = 125$

$= \frac{x_{(125)} + x_{(126)}}{2} = \frac{3 + 3}{2} = \underline{3}$ ✓

$x_p = x_{0,75} = \frac{x_{(k)} + x_{(k+1)}}{2}$ NR: $k = n \cdot p$, $k = 500 \cdot 0,75 = 375$

$= \frac{x_{(375)} + x_{(376)}}{2} = \frac{27 + 28}{2} = \underline{\underline{27,5}}$ ✓

$x_p = x_{0,1} = \frac{x_{(k)} + x_{(k+1)}}{2}$ NR: $k = n \cdot p$, $500 \cdot 0,1 = 50$

$x_{0,001}$ gesucht

$= \frac{x_{(50)} + x_{(51)}}{2} = \frac{2 + 2}{2} = \underline{\underline{2}}$

3

iii.) Boxbreite $= x_{0,75} - x_{0,25} = 27,5 - 3 = \underline{\underline{24,5}}$

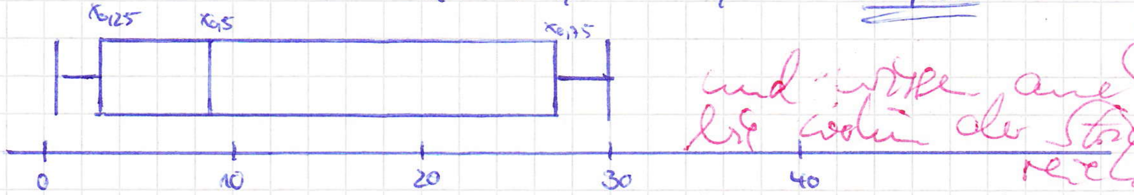
Linker Zaun $= x_{0,25} - 1,5 \cdot$ Boxbreite

$= 3 - 1,5 \cdot 24,5 = \underline{\underline{-33,75}}$

rechter Zaun $= x_{0,75} + 1,5 \cdot$ Boxbreite

$= 27,5 + 1,5 \cdot 24,5 = \underline{\underline{64,25}}$

4



und Whisker an, bis wohin die Striche reichen nicht,

- keine Ausreißer vorhanden

bzw. wir wissen es nicht

iv) $x_{mod} \rightarrow$ 30 Sekunden

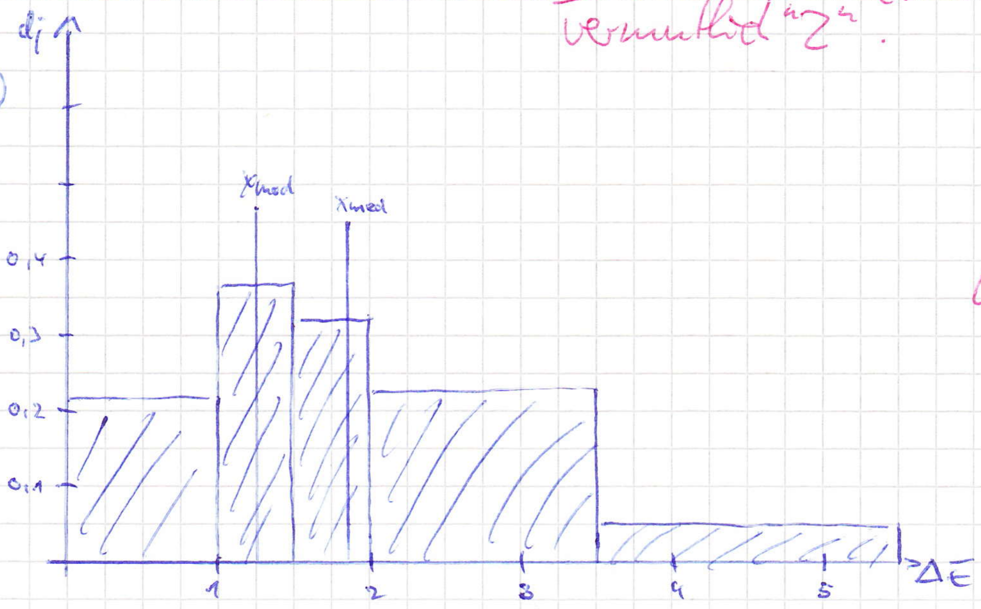
↳ der Modus ist der Wert unserer Datenreihe mit der größten Häufigkeit

Ja, aber >30 sind vermutl(ic) viele verschiedene Werte. Der realistische Modus ist vermutlich "2". 1

10

5.

ii)



iii) $x_{mod} = \frac{x_i^u + x_i^o}{2} = \frac{1 + 1{,}5}{2} = \underline{1{,}25}$

der Modus befindet sich in Klasse 2

$x_{0,5} = x_i^u + \frac{x_p - F}{h_i} \cdot (x_i^o - x_i^u) = 1{,}5 + \frac{0{,}5 - 0{,}39}{0{,}16} \cdot (2 - 1{,}5)$

$= 1{,}8438 \approx \underline{1{,}84}$ der Median befindet sich in Klasse 3

iv) - die Form der Verteilung lässt sich nicht genau beschreiben, da in der vorliegenden Tabelle nicht ersichtlich ist, wie die einzelnen Daten in den Klassen verteilt sind
- für Modus und Median ist nur eine Annäherung möglich

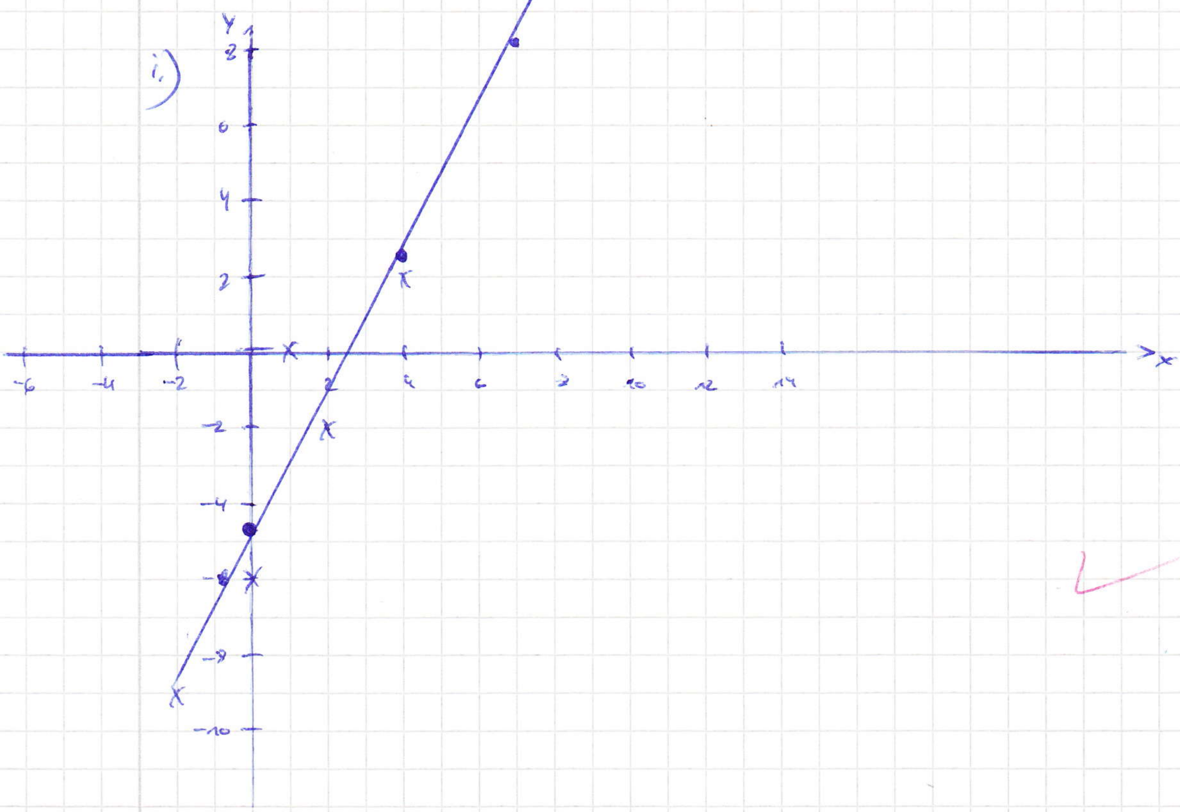
Man sieht aber z.B. an der Zeichnung, dass die Vtg rechtsschief ist. 1

11

$x_i \quad 2 \quad 4 \quad -2 \quad 0 \quad 1 \qquad n = 5$

$y_i \quad -2 \quad 2 \quad -9 \quad -6 \quad 0$

i)



ii) $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{4} x_i = \frac{1}{5} \cdot (2 + 4 + (-2) + 0 + 1) = \underline{1}$

$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{4} y_i = \frac{1}{5} \cdot (-2 + 2 + (-9) + (-6) + 0) = \underline{\underline{-3}}$

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

$+ (0-1)^2 + (1-1)^2 \Big)$

$= \underline{\underline{5}}$

$\overset{1}{(-2+3)^2} + \overset{25}{(2+3)^2} + \overset{36}{(-9+3)^2} + \overset{9}{(-6+3)^2} + \overset{9}{(0+3)^2}$

$s_y^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{4} (y_i - \bar{y})^2 = \frac{1}{4} \Big((-2-(-3))^2 + (2-(-3))^2$

$+ (-9-(-3))^2 + (-6-(-3))^2$

$+ (0-(-3))^2 \Big)$

$= \underline{\underline{20}}$

$\bar{x} = 1$
$\bar{y} = -3$

zu 3 i) $s_{xy} = \frac{1}{n-1} \cdot \sum_{i=1} (x_i - \bar{x}) \cdot (y_i - \bar{y})$

$= \frac{1}{4} \cdot \big((2-1) \cdot (-2-(-3)) + (4-1) \cdot (2-(-3))$

$+ (-2-1) \cdot (-9-(-3)) + (0-1) \cdot (-6-(-3))$

$+ (1-1) \cdot (0-(-3)) \big)$

$= \underline{\underline{9{,}25}}$ ✓

ii.) $\hat{y} = \hat{a} + \hat{b} \cdot x = \underline{\underline{-4{,}85 + 1{,}85 \cdot x}}$

$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{9{,}25}{5} = \underline{\underline{1{,}85}}$ ✓

$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x} = (-3) - 1{,}85 \cdot 1 = \underline{\underline{-4{,}85}}$ ✓

NR: $-4{,}85 + 1{,}85 \cdot 0 = -4{,}85 = \hat{y}$

$-4{,}85 + 1{,}85 \cdot 4 = 2{,}55 = \hat{y}$

$7 = 8{,}1 = \hat{y}$

v.) $r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{9{,}25}{\sqrt{5 \cdot 20}} = \underline{\underline{0{,}925}}$ ✓

$R^2 = r_{xy}^2 = 0{,}8556 \approx \underline{\underline{85{,}56\,\%}}$ ✓

vi.) $\hat{y} = -4{,}85 + 1{,}85 \cdot 20 = \underline{\underline{32{,}15}}$ ✓

Vor allem war nach der Vorhersagegüte (vi) gefragt. Da ist noch unklar, dass x=20 weit abseits liegt!

iii.) Anhand des Diagramms lässt sich erkennen, dass die Punkte annähernd miteinander korrelieren, denn die Abstände zur Geraden sind relativ klein. Das Bestimmtheitsmaß gibt an, dass die x- und y-Werte zu ca. 85,56 % miteinander korrelieren. Das Bestimmtheitsmaß liegt näher an 1 als an 0. Da der Korrelationskoeffizient ~~fast bei~~ liegt fast bei eins. Ein positives Vorzeichen besagt, dass die Gerade positiv steigend ist.

17

4

$p = 5\% = 0{,}05 \;\hat{=}\; 0{,}95$

$n = 9$

$k = 8$

$$P(X=k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \binom{9}{8} \cdot 0{,}05 \cdot (1-0{,}95)^{9-8}$$

$$= 0{,}2985 \approx 29{,}85\%$$

Mit einer Wahrscheinlichkeit von 29,85% sind ~~weniger als~~ 8 brauchbare Kopien dabei.

genau 8   P von "weniger als 8": $1 - P(Y=8) - P(Y=9)$

2

---

5. i) $P(X \leq 1750)$

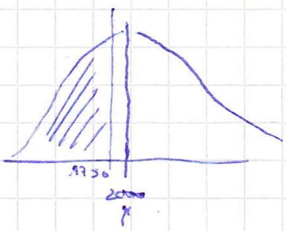
$X \sim N(2000, 10000)$, μ, σ^2

$\sqrt{10.000} = 100 = \sigma$

$$= P\left(Z \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{1750-2000}{100}\right)$$

$$= P(Z \leq -2{,}5) = F(c) = F(2{,}5)$$

$$= 0{,}0062 = \underline{0{,}62\%}$$



2

ii.) $P(1850 \leq X \leq 2150)$

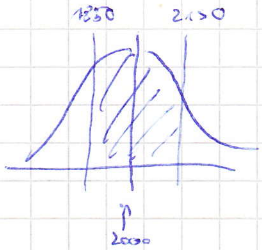
$$= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

$$= P\left(\frac{1850-2000}{100} \leq Z \leq \frac{2150-2000}{100}\right)$$

$$= P(\underset{a}{-1{,}5} \leq Z \leq \underset{b}{1{,}5})$$

$$= F(b) - F(a) = 0{,}9332 - 0{,}0668 = 0{,}8664$$

$$= \underline{86{,}64\%}$$



2

iii) $z_p = 5\% = 0{,}05$

$$x_p = \mu + \sigma \cdot z_p = 2000 + 100 \cdot \underset{(-1{,}6449)}{z_{0{,}05}}$$

$$= 2000 + 100 \cdot (-1{,}6449)$$

$$= \underline{1835{,}51}$$

1

Dieses Volumen wird mit ~~95% W~~ 95% W'keit überschritten. Prämie ist zu teuer!

(iv) ?, (v) ?

5

6. i) $\hat{p} = \frac{k}{n} = \frac{\text{Anzahl Erfolge}}{\text{Anzahl Versuche}} = \frac{900}{10.000} = 0{,}09 = \underline{\underline{9\%}}$ 2

ii) $KI = 95\% = \alpha = 0{,}05$

$n = 10000$ 0,5

2,5

7. $n = 25$ $\alpha = 5\% = 0{,}05$

$\bar{x} = 2100$

$s^2 = 8100 = \sqrt{s^2} = 90$

i) $H_0: \mu \geq \mu_0$ vs. $H_1: \mu < \mu_0$ $\rightarrow \mu_0 = 2000$ 1

$\mu \leq \mu_0$ $\mu > \mu_0$

$T = \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{s} = \underset{(5)}{\sqrt{25}} \cdot \frac{2100 - 2000}{90} = \underline{\underline{5{,}56}}$

~~$T < t_{n-1,\alpha} = 5{,}56 < t_{25-1,0{,}05} = -t_{n-1,1-\alpha}$~~

~~$= 5{,}56 < -1{,}7081$~~

$T < t_{n-1,\alpha} = t_{n-1,1-\alpha}$

$5{,}56 < t_{24,0{,}05}$ ~~$= t_{24,0{,}95}$~~

$5{,}56 < -1{,}7109$ ~~$= 1{,}7109$~~

Lehne H_0 nicht ab

(ii) ?

4

5