FernUniversität in Hagen

LOTSE

Vorläufige Musterlösungen (Stand: 21. März 2013) zur Klausur

zu Modul M1 im BA-Studiengang
„Politikwissenschaft, Verwaltungswissenschaft, Soziologie“

und zu Modul 3 im BA-Studiengang
„Soziologie“

Termin: **5. März 2013, 14.00 - 18.00 Uhr**

Erstprüfer: **apl. Prof. Dr. H.-J. Mittag**

Zweitprüfer: **Dr. H.-G. Sonnenberg**

Rückfragen zur Klausur und zur Musterlösung
sind stets **an den Erstprüfer** zu richten.

Multiple-Choice-Aufgaben

Lösung zu Aufgabe 1 (Aussagenlogik):

A, C, D – vgl. zu dieser Aufgabe auch den Abschnitt 5.1.3 in Kurs 33210, insbesondere die letzte der dort wiedergegebenen beiden Wahrheitstabellen.

Kommentar:

Wenn man die Tabelle um die fehlenden 12 Werte ergänzt und letztere durch Fettdruck betont, erhält man:

a	b	P1: $a \wedge b$	P2: $\neg b$	K: $(a \wedge b) \vee (\neg b)$
w	w	**w**	**f**	**w**
w	f	**f**	**w**	**w**
f	w	**f**	**f**	**f**
f	f	**f**	**w**	**w**

Anhand dieser ergänzten Tabelle kann man die fünf Aussagen sofort bezüglich ihres Wahrheitsgehalts beurteilen.

Lösung zu Aufgabe 2 (Merkmalsklassifikationen) (5 Punkte)

A, B, C – vgl. zu dieser Aufgabe auch Kromrey, Abschnitt 5.4.3, oder Kurs 33209, Abschnitt 2.2.

Zu D: Metrisch skalierte Merkmale sind immer quantitativ; vgl. Kurs 33209, Abschnitt 2.2.

Zu E: Die Ausprägungen ordinalskalierter Merkmale lassen sich in eine Rangfolge bringen, bei nominalskalierten Merkmalen gilt dies nicht.

Lösung zu Aufgabe 3 (Nominal- und Realdefinitionen) (5 Punkte)

A, B, E.

Zu A: Kromrey, Abschnitt 3.5.4.

Zu B: Kromrey, Abschnitt 3.5.1.

Zu C: Realdefinitionen sind niemals vollständig, sie heben vielmehr nur auf das Wesentliche des Definiendums ab (vgl. Kromrey, Abschnitt 3.5.4.

Zu D: Für Nominaldefinitionen trifft die Aussage nicht zu; vgl. Kromrey, Abschnitt 3.5.3.

Zu E: Kromrey, Abschnitt 3.5.1.

Lösung zu Aufgabe 4 (Informationsgewinnung, Datenerhebung) (5 Punkte)

A, B, C.

Zu A - B: vgl. Kromrey, Abschnitt 7.3.3

Zu C: vgl. Kromrey, Abschnitt 2.2, oder Kurs 33209, Abschnitt 2.3.

Zu D: Es ist gerade der Vorzug des Random-Route-Verfahrens, keine Namens- oder Adressdatei zu benötigen – vgl. Kromrey, Abschnitt 6.5.3.

Zu E: vgl. Kromrey, Abschnitt 6.4.3, und Kurs 33209, Abschnitt 3.2.

Lösung zu Aufgabe 5 (Messen / Stichprobenverfahren) (5 Punkte)

E.

Zu A: Die Validität (Gültigkeit) ist das Gütekriterium für Messungen, mit dem beschrieben wird, inwieweit wirklich das gemessen wird, was gemessen werden soll. Eine Messung, die nicht-valide ist, kann trotzdem reliabel sein - vgl. Kromrey, Abschnitt 5.7 (ganz am Ende).

Zu B: Auswahleinheiten sind hier die Schulen - vgl. Kromrey, Abschnitt 6.1.

Zu C: Die Klumpenauswahl ist eine zweistufige Auswahlprozedur, bei der nur in der ersten Stufe zufällig ausgewählt werden (Zufallsauswahl von Klumpen). In der zweiten Verfahrensstufe werden dann alle Elemente der ausgewählten Klumpen herangezogen, nicht nur eine Zufallsauswahl der Elemente – vgl. hierzu Kromrey, Abschnitt 6.5.2, oder Kurs 33209, Abschnitt 3.2.

Zu D: Der Prozentsatz (Auswahlsatz; Quote) kann von Schicht zu Schicht variieren. Nur bei einer proportional geschichteten Stichprobenauswahl – nicht aber allgemein – ist die Aussage richtig. Vgl. Kromrey, Abschnitt 6.5.2 oder Kurs 33209, Abschnitt 3.2.

zu E: vgl. Kurs 33209, Abschnitt 3.2.

Anmerkung:

Der letzte Satz bei Aussage D sollte beinhalten, dass der Prozentsatz von Schicht zu Schicht gleich bleibt. Die Aussage D ist daher in der obigen Musterlösung als unzutreffend ausgewiesen. Da man aber den Satz u. U. auch anders interpretieren konnte, wurde generell 1 Punkt vergeben.

Lösung zu Aufgabe 6 (absolute und relative Häufigkeiten) (5 Punkte)

B, D, E - vgl. auch Beispiel 8.2 in Kurs 33209. Die Lösungen verifiziert man am einfachsten anhand einer Vierfeldertafel oder anhand eines Baumdiagramms (beides nachstehend wiedergegeben).

Zu A: Die Aussage bezieht sich auf die absolute Häufigkeit $h_{21} = 8,0901$, trifft also nicht zu (s. die erste der nachstehenden Vierfeldertafeln oder das Baumdiagramm).

Zu B: Die Aussage bezieht sich bei dem folgenden Baumdiagramm auf die absolute Häufigkeit $h_{12} = 7,4844$, trifft also zu.

Zu C: Die Aussage bezieht sich auf die relative Häufigkeit $f_{11} = 0,2871$ (s. die zweite der nachstehenden Vierfeldertafeln oder das Baumdiagramm). Dies entspricht 28,71 %. Die Aussage ist also nicht zutreffend.

Zu D: Die Aussage bezieht sich auf die (in Millionen ausgewiesene) Randhäufigkeit $h_{\cdot 1} = h_{11} + h_{21}$, also auf $h_{\cdot 1} = 10,3356 + 8,091 = 18,4257$, die oberhalb von $18,3$ liegt.

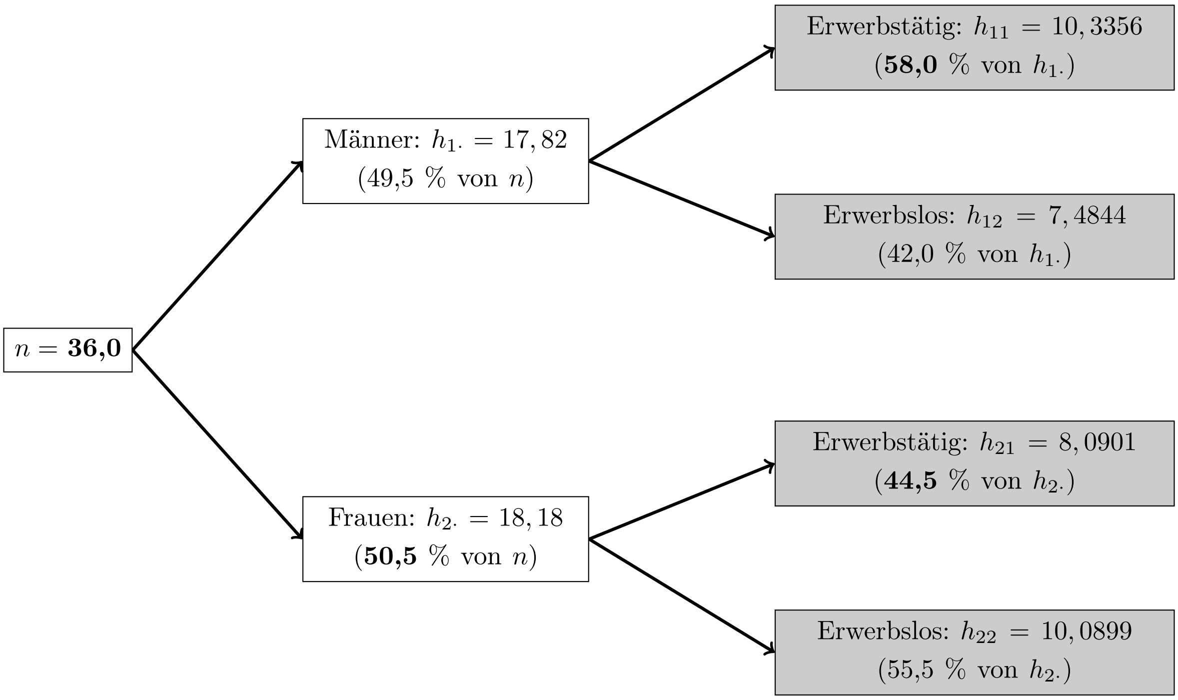
Zu E: Vgl. hierzu Tabelle 8.2 oder Tabelle 8.3 in Kurs 33209.

Die vom Statistischen Amt kommunizierten Angaben lassen sich anhand einer Vierfeldertafel für absolute bzw. – nach Division aller Werte durch n – für relative Häufigkeiten darstellen:

	Erwerbstätige	Erwerbslose	Zeilensummen
Männer	$10,3356$	$7,4844$	$17,82$
Frauen	$8,0901$	$10,0899$	$18,18$
Spaltensummen	$18,4257$	$17,5743$	$36,0$

	Erwerbstätige	Erwerbslose	Zeilensummen
Männer	$0,2871$	$0,2079$	$0,495$
Frauen	$0,2247$	$0,2803$	$0,505$
Spaltensummen	$0,5118$	$0,4882$	1

Alternativ kann man die Angaben auch in Form eines Baumdiagramms darstellen:



Die in der Aufgabe vorgegebenen Zahlen sind beim Baumdiagramm durch Fettdruck hervorgehoben (vgl. das analoge Beispiel 8.2 in Kurs 33209).

Lösung zu Aufgabe 7 (Kenngrößen von Datensätzen) (5 Punkte)

C, E.

Zu A: Der Datensatz hat den – eindeutig bestimmten – Modalwert $4,3$. Dieser Wert tritt doppelt auf, alle anderen Werte nur einmal.

Zu B: Der Median des Datensatzes des Umfangs $n = 12$ ist $\widetilde{x} = 5,4$ (Mittelwert aus dem sechsten Wert $5,3$ und siebten Wert $5,5$ der nach aufsteigender Größe geordneten Urliste), während für den Mittelwert $\overline{x} = 5,2$ gilt.

Zu C: Die Begrenzungen der Box sind durch die in den Interquartilsabstand eingehenden beiden Quartile definiert. Innerhalb der Box oder auf den Begrenzungslinien der Box liegen hier die „mittleren 50 %“ des nach Größe geordneten Datensatzes. Bei dem nach aufsteigender Größe geordneten Datensatz sind dies die Werte $4,3$, $4,3$, $5,3$, ..., $6,2$.

Zu D: Wenn man den ersten Wert $(5,3)$ der Urliste um $0,2$ senkt und den letzten Wert $(6,2)$ um $0,2$ erhöht, bleibt die Merkmalssumme konstant und damit auch der Mittelwert. Der Median, der sich wieder ($n = 12$) als Mittelwert aus dem sechsten Wert – nun auf $5,1$ reduziert – und dem siebten Wert $(5,5)$ bestimmt, hat jetzt den Wert $5,3$.

Zu E: Wenn man den Wert $6,2$ streicht, sinkt der Mittelwert. Man erkennt dies auch ohne Rechnung, denn es wurde ja ein oberhalb des ursprünglichen Mittelwerts liegender Wert aus dem Datensatz herausgenommen. Man errechnet für den neuen Mittelwert $\overline{x} \approx 5,11$. Der neue Median ergibt sich nun (Datensatz des Umfangs $n = 11$) als sechster Wert der neuen, wieder nach aufsteigender Größe geordneten Urliste. Man erhält $\widetilde{x} = 5,3$.

Lösung zu Aufgabe 8 (Konzentrationsmessung) (5 Punkte)

B, D, E.

Zu A - B: Die Merkmalssumme beträgt $p_5 = 255$ und die Summe $p_{(2)}$ der beiden kleinsten Merkmalswerte ist $p_{(2)} = 60$. Nach Formel (6.3) in Kurs 33209 erhält man dann $v_2 = \frac{60}{255} \approx 0,235$ (Anteil der beiden kleinsten Merkmalsträger an der Summe $p_5$aller Merkmalswerte).

Zu C: Die gewichtete Merkmalsssumme q_5 aus Formel (6.4) in Kurs 33209 beträgt 910. Nach (6.5) errechnet sich dann für unnormierten Gini-Koeffizienten bei Rundung auf drei Dezimalstellen der Wert $G = \frac{2 \cdot 910}{5 \cdot 255} - \frac{6}{5} \approx 0,227$.

Zu D: Für den normierten Gini-Koeffizienten G^* gilt $G^* = \frac{G}{G_{max}} \approx \frac{0{,}227}{0{,}8} \approx 0,284$.

Zu E: Eine Halbierung der Merkmalswerte führt zu einer Halbierung sowohl der Merkmalssumme p_5 als auch der nach (6.4) errechneten gewichteten Merkmalssumme q_5. Man erkennt aus der Formel (6.5) in Kurs 33209, dass die gleichzeitige Halbierung von q_5 und p_5 keinen Effekt auf G und damit auch keinen Effekt auf G^* hat.

Lösung zu Aufgabe 9 (Kontingenztafeln; Randverteilungen) (5 Punkte)

A, E - vgl. auch die Beispiele 8.1 und 8.2 in Kurs 33209.

Zu A: Der Wert 520 ist das erste Element der Randverteilung des Merkmals Y.

Zu B: Es ist $h_{22} = 117$ und $f_{22} = \frac{117}{520} \approx 0,225$. Dies entspricht 22,5 %.

Zu C: Durch die gemeinsamen Häufigkeiten h_{ij} sind die Randhäufigkeiten eindeutig bestimmt. Die Umkehrung gilt aber nicht, d. h. man kann *nicht* eindeutig bei gegebenen Randverteilungen auf die gemeinsamen Häufigkeiten h_{ij} zurückschließen (vgl. auch in Kurs 33209 den Text unmittelbar vor Beispiel 8.1).

Zu D: Gesucht ist die bedingte relative Häufigkeit $f_Y(b_2|a_2)$. Man erhält $f_Y(b_2|a_2) = \frac{117}{217} \approx 0,539$. Dies entspricht 53,9 %.

Zu E: Gesucht ist hier die bedingte relative Häufigkeit $f_X(a_1|b_1)$. Man erhält $f_X(a_1|b_1) = \frac{179}{501} \approx 0,357$. Dies entspricht 35,7 %.

Lösung zu Aufgabe 10 (Venn-Diagramme, Kombinatorik) (5 Punkte)

B, D, E.

Zu A: Die gesuchte Wahrscheinlichkeit ist durch den Wert der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 8$ und $p = 0,5$ an der Stelle $x = 3$ gegeben, nach Tabelle 19.1 also durch $F(3) = 0,3633$.

Zu B: Die Wahrscheinlichkeit dafür, genau dreimal „Zahl“ zu erhalten, ergibt sich als Differenz der Werte $F(3)$ und $F(2)$ der Verteilungsfunktion der Binomialverteilung mit $n = 8$ und $p = 0,5$, also nach Tabelle 19.1 als $F(3) - F(2) = 0,3633 - 0,1445 = 0,2188$.

Zu C: Es ist der Erwartungswert $\mu = E(X)$ einer Binomialverteilung mit $n = 10$ und $p = \frac{1}{3}$ zu bestimmen. Der Parameter p bezeichnet die Wahrscheinlichkeit dafür, dass bei einem Wurf des Würfels eine der beiden Augenzahlen 1 und 6 erscheint. Für μ errechnet man nach Formel (11.20) aus Kurs 33209 den Wert $\mu = 10 \cdot \frac{1}{3} \approx 3,33$.

Zu D - E: vgl. Kurs 33209, dort Abschnitt 10.1 und Aufgabe 10.1 oder auch Kurs 33210, Kapitel 5.

Lösung zu Aufgabe 11 (Wahrscheinlichkeits- und Verteilungsfunktion) (5 Punkte)

A, B, E.

Zu A - C: Die gesuchte Wahrscheinlichkeit ist gegeben durch die Treppenfunktion

$$F(x) = \begin{cases} 0 & \text{für } x < 1 \\ 0,1 & \text{für } 1 \leq x < 2 \\ 0,1 + 0,2 = 0,3 & \text{für } 2 \leq x < 3 \\ 0,3 + 0,1 = 0,4 & \text{für } 3 \leq x < 4 \\ 0,4 + 0,2 = 0,6 & \text{für } 4 \leq x < 5 \\ 0,6 + 0,2 = 0,8 & \text{für } 5 \leq x < 6 \\ 0,8 + 0,1 = 0,9 & \text{für } 6 \leq x < 7 \\ 1 & \text{für } x \geq 7 \end{cases}$$

- vgl. Formel (11.3) in Kurs 33209 sowie auch Abbildung 11.1, die sich allerdings auf eine diskrete Zufallsvariable mit sechs Ausprägungen und gleichen Eintrittswahrscheinlichkeiten bezieht. Es gilt also $F(3) = 0,4$ und $F(3,5) = 0,4$. Die Verteilungsfunktion $F(x)$ ist auch für $x > 7$ definiert. Sie erreicht an der Stelle $x = 7$ den Wert 1, den sie dann für $x > 7$ beibehält.

Zu D: Der Erwartungswert errechnet sich nach (11.6) als Summe der mit den Eintrittswahrscheinlichkeiten gewichteten Ausprägungen:

$$\mu = E(X) = 1 \cdot 0,1 + 2 \cdot 0,2 + 3 \cdot 0,1 + 4 \cdot 0,2 + 5 \cdot 0,2 + 6 \cdot 0,1 + 7 \cdot 0,1 = 3,9.$$

Zu E: Die Richtigkeit der Aussage erschließt sich aus (11.12) aus Kurs 33209 – man setze dort speziell $a = -1$.

Lösung zu Aufgabe 12 (Zusammenhangsmessung) (5 Punkte)

E

Zu A: Nicht die Kovarianz, sondern der aus ihr durch Normierung hervorgehende Korrelationskoeffizient ρ liegt zwischen -1 und $+1$ – vgl. Kurs 33209, Abschnitt 13.2.

Zu B: Der Korrelationskoeffizient r nach Bravais-Pearson misst die Stärke eines linearen, nicht aber den eines nicht-linearen Zusammenhangs zwischen zwei Merkmalen X und Y – vgl. auch Abbildung 9.2 in Kurs 33209.

Zu C: Wenn $r = 0$ ist, liegt kein linearer Zusammenhang vor (möglicherweise aber ein nicht-linearer) – vgl. auch Abbildung 9.2 in Kurs 33209.

Zu D: vgl. (9.15) in Kurs 33209.

Zu E: Setzt man in (9.2) in Kurs 33209 den Wert $M = min(3; 2) = 2$ und $n = 28$ ein, folgt, dass der χ^2-Koeffizient nicht größer als 28 werden kann.

Anmerkung:

Aussage B sollte beinhalten, dass der Korrelationskoeffizient von Bravais-Pearson sowohl einen linearen als auch einen nicht-linearen Zusammenhang zwischen X und Y widerspiegelt. Letzteres trifft nicht zu und deshalb ist die Aussage in der obigen Musterlösung als nicht zutreffend ausgewiesen. Da man aber den Satz u. U. auch anders interpretieren konnte („oder“ im Sinne der Aussagenlogik verlangt nur die Gültigkeit von mindestens einer der beiden genannten Bedingungen), wurde generell 1 Punkt vergeben.

Aufgabe 13 (Punkt- und Intervallschätzungen) (5 Punkte)

A, E.

Zu A : vgl. Formel (14.6) in Kurs 33209.

Zu B: Es ist genau umgekehrt – vgl. (13.6) in Kurs 33209.

Zu C: Eine unverzerrte (= erwartungstreue) Schätzung erhält man, wenn man die genannte Summe nicht durch n, sondern durch $n - 1$ dividiert – vgl. (14.8) und (14.9) in Kurs 33209.

Zu D: Ein Konfidenzintervall für μ enthält den unbekannten Parameter μ nicht immer– vgl. Abbildung 14.3 in Kurs 33209, bei der die dunkel markierten Konfidenzintervalle Gegenbeispiele darstellen.

Zu E: Dass die Länge des Konfidenzintervalls mit zunehmendem n abnimmt, geht aus (14.15) bzw. (14.17) hervor (vgl. auch Abbildung 14.3 in Kurs 33209).

Lösung zu Aufgabe 14 (Stetige Verteilungen) (5 Punkte)

A, B, E.

Zu A: Es gilt $P(X \leq x_{0,05}) = 0,05$ und somit $P(X > x_{0,05}) = 0,95$.

Zu B: Sowohl die Dichtekurve einer standardnormalverteilten Zufallsvariablen wie auch die Dichtekurve jeder t-verteilten Zufallsvariablen ist symmetrisch bezüglich des Erwartungswertes. Letzterer hat bei beiden Verteilungen den Wert 0. Hieraus folgt in Verbindung mit der genannten Symmetrieeigenschaft, dass die Aussage zutrifft.

Zu C: Eine t-verteilte Zufallsvariable T ist – wie die Standardnormalverteilung – symmetrisch bezüglich des Nullpunkts und wird der Standardnormalverteilung mit zunehmender Anzahl von Freiheitsgraden immer ähnlicher (vgl. Abbildung 12.6 in Kurs 33209). Ihre Varianz $V(T) = \frac{n}{n-2}$ nimmt mit zunehmender Anzahl der Freiheitsgrade ab (Konvergenz gegen den Grenzwert 1).

Zu D: Die Varianz von $Y = X_1 + X_2$ ist nach (13.14) durch

$$V(Y) = V(X_1) + V(X_2) + 2 \cdot Cov(X_1; X_2)$$

gegeben. Nur bei Unabhängigkeit von X_1 und X_2 (Spezialfall) entfällt der Kovarianzterm. Die Aussage D gilt also nicht allgemein.

Zu E: Der Wert, den eine mit $m = 5$ und $n = 15$ Freiheitsgraden F-verteilte Zufallsvariable X nur mit Wahrscheinlichkeit $0,05$ überschreitet, ist das $0,95$-Quantil $F_{5;15;0,95}$ der genannten Verteilung. Aus Tabelle 19.6 in Kurs 33209 kann man ablesen, dass $F_{5;15;0,95} = 2,9$ ist.

Lösung zu Aufgabe 15 (Testen, Fehler beim Testen) (5 Punkte)

C, D, E.

Zu A - B: Die Ablehnung der Nullhypothese erfolgt bei dem hier zugrunde gelegten zweiseitigen Test, wenn die Realisation der Prüfgröße außerhalb des Intervalls $[z_{0,005}; z_{0,995}] = [-z_{0,995}; z_{0,995}]$ liegt, also außerhalb des Intervalls $[-2,578; 2,578]$ – vgl. auch Abbildung 15.2 in Kurs 33209. Im Falle $z = -2,3$ kann H_0 also nicht verworfen werden. Die

Überschreitung des $0,99$-Quantils $z_{0,99} = 2,3263$ begründet ebenfalls noch nicht eine Verwerfung von H_0.

Zu C - E: Ein Fehler 1. Art kann hier nur für $\mu = \mu_0$ auftreten (dort nimmt die Gütefunktion ihr Minimum an), ein Fehler 2 Art nur für $\mu \neq \mu_0$ – vgl. auch Abbildung 15.6 in Kurs 33209.

Lösung zu Aufgabe 16 (Regressions- und Varianzanalyse) (5 Punkte)

B, E.

Zu A: Es gilt nach (16.6) und (16.7) bei Beachtung von $\overline{x} = 3$ und $\overline{y} = 7$

$$\widehat{\beta} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{8}(x_i - \overline{x})^2} = \frac{2,3}{4,1} \approx 0,561.$$

Zu B: Nach (16.18) gilt für das Bestimmtheitsmaß

$$R^2 = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = \frac{2,3^2}{4,1 \cdot 12} = \frac{5,29}{49,2} \approx 0,108.$$

Zu C: Mit $\overline{x} = 3$ und $\overline{y} = 7$ erhält man zunächst $\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} \approx 7 - 0,561 \cdot 3 = 5,317$. Die nach der KQ-Methode geschätzte Regressionsgerade

$$\widehat{y} = \widehat{\alpha} + \widehat{\beta} \cdot \overline{x} = 5,317 + 0,561 \cdot x$$

nimmt daher für $x = 1$ den Wert $\widehat{y} = 5,317 + 0,561 = 5,878$ an.

Zu D: Bei der KQ-Methode wird die Regressionsgerade dadurch bestimmt, dass die Summe der quadrierten Residuen minimiert wird. Sie ist aber i. a. nicht Null — sie ist nur dann Null, wenn alle Punkte des Datensatzes exakt auf einer Geraden liegen.

Zu E: Bei der Varianzanalyse wird die unabhängige Variable X (eine Einflussgröße im einfaktoriellen Fall) als diskret modelliert, nicht aber die abhängige Variable Y.

Numerische Aufgaben

Lösung zu Aufgabe 41 (Rangkorrelationskoeffizient) (3 Punkte)

0,771.

Herleitung:

Wenn man den Beurteilungen der beiden Tester jeweils Ränge $rg(x_i)$ bzw. $rg(y_i)$ zuordnet und auch noch die Rangdifferenzen $d_i = rg(x_i) - rg(y_i)$ ausweist, erhält man die folgende erweiterte Tabelle:

	Tester 1		Tester 2		
Wein i	Bewertung x_i	$rg(x_i)$	Bewertung y_i	$rg(y_i)$	d_i
1	7	**4**	7	**4**	0
2	8	**3**	9	**2**	1
3	6	**5**	3	**6**	-1
4	4	**6**	5	**5**	1
5	10	**1**	8	**3**	-2
6	9	**2**	10	**1**	1

Bei der Zusammenhangsmessung anhand des Rangkorrelationskoeffizienten r_{SP} von Spearman kann anstelle von (9.14) aus Kurs 33209 die vereinfachte Formel (9.16) angewendet werden, weil die Bewertungen nicht mit der Mehrfachbelegung eines Rangplatzes verbunden sind. Mit (9.16) resultiert

$$r_{SP} = 1 - \frac{6 \cdot [0^2 + 1^2 + (-1)^2 + 1^2 + (-2)^2 + 1^2]}{6 \cdot (36 - 1)} = 1 - \frac{8}{35} \approx 1 - 0,229 = 0,771.$$

Anmerkungen:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,763; 0,779]$ als richtig anerkannt.

Wenn anstelle von $r_{SP} \approx 1 - 0,229 = 0,771$ fälschlich der Wert $0,229$ als Lösung angegeben wurde, wurden für diesen Wert 2 P. gewährt. Letzteres gilt auch für jeden Wert aus dem Intervall $[0,226; 0,232]$.

Lösung zu Aufgabe 42 (Auswahlwahrscheinlichkeit) (3 Punkte)

0,2090.

Herleitung:

Die im Text der Aufgabe enthaltenen Informationen sind in der nachstehenden Kontingenztafel mit Randverteilungen durch Kursivschrift betont, die für die Lösung benötigten Werte sind fett ausgewiesen:

	weiblich (B)	männlich $\overline{B}$	Zeilensummen
MATH (A)	**28**	*48*	*76*
PHYS ($\overline{A}$)	26	*32*	*58*
Spaltensummen	*54*	80	**134**

Mit (10.5) aus Kurs 33209 folgt, dass die Wahrscheinlichkeit $P(A \cap B)$ dafür, dass per Zufallsauswahl aus der Population aller 134 Klausurteilnehmer eine Frau mit Studienfach MATH gewählt wird, den Wert $P(A \cap B) = \frac{28}{134} \approx 0,2090$ hat (entspricht 20,9 %).

Anmerkung:

Anstelle von $0,2090$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,207; 0,211]$ als richtig anerkannt.

Lösung zu Aufgabe 43 (Wahrscheinlichkeiten beim Roulettespiel) (3 Punkte)

0,2367

Herleitung:

Die Wahrscheinlichkeit für den Ausgang „Impair“ beträgt bei jedem Spiel $\frac{18}{37}$. Die Wahrscheinlichkeit dafür, das „Impair“ nicht auftritt (Komplementärwahrscheinlichkeit) ist folglich durch $\frac{19}{37}$ gegeben. Die Wahrscheinlichkeit dafür, bei zwei Spielen jeweils den Ausgang „Impair“ zu verfehlen, ist dann aufgrund der Unabhängigkeit der beiden Einzelereignisse nach (10.16) gegeben durch

$$P = \frac{19}{37} \cdot \frac{19}{37} = \frac{361}{1369} \approx 0,2637.$$

Anmerkungen:

Anstelle von $0,2637$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,261; 0,266]$ als richtig anerkannt.

Wenn man bei der Berechnung der gesuchten Wahrscheinlichkeit den Ausgang „0“ fälschlich nicht berücksichtigte, errechnete sich die Wahrscheinlichkeit dafür, bei beiden Spielen den Ausgang „Impair“ zu verfehlen, nach

$$P = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0,25.$$

Für dieses Ergebnis wurde 1 Punkt gewährt. Ebenfalls 1 Punkt wurde gewährt, wenn nicht die Wahrscheinlichkeit für "2 x Pair", sondern fälschlich die Wahrscheinlichkeit

$$P = \frac{18}{37} \cdot \frac{18}{37} = \frac{324}{1369} \approx 0,2367$$

für "2 x Impair"berechnet wurde. Neben $0,2367$ wurde bei der maschinellen Auswertung auch jeder Wert aus dem Intervall $[0,234; 0,239]$ als richtig anerkannt.

Lösung zu Aufgabe 44 (Varianzberechnung) (3 Punkte)

0,1914

Herleitung:

Die Zufallsvariable X ist hypergeometrisch verteilt, weil ohne Zurücklegen gezogen wird. Ihr Erwartungswert $\mu = E(X)$ bestimmt sich nach Formel (11.24) des Kurses 33209, die Varianz $\sigma^2 = V(X)$ nach (11.25), wobei $n = 7$ sowie $M = 3$ und $N = 100$ einzusetzen ist. Für die zu berechnende Varianz folgt

$$\sigma^2 = 7 \cdot \frac{3}{100} \cdot (1 - \frac{3}{100}) \cdot \frac{93}{99} \approx 0,1914.$$

Anmerkungen:

Anstelle von $0,1914$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,189; 0,193]$ als richtig anerkannt.

Wenn man fälschlich X als binomialverteilt angenommen hätte (bei Ziehen mit Zurücklegen wäre diese Verteilung zutreffend), würde man den Erwartungswert nach (11.20) bestimmen mit $n = 7$ und $p = 0,03$ und die Varianz nach (11.21). Das Ergebnis für den Erwartungswert wäre dasselbe wie bei der hypergeometrischen Verteilung, während sich für die Varianz ein im Vergleich zur Varianz der hypergeometrischen Verteilung etwas kleinerer Wert ergäbe, nämlich $0,2037$. Der Unterschied ist durch den in (11.25) auftretenden Faktor $\frac{N-n}{N-1} = \frac{93}{99} \approx 0,939$ definiert. Für den Wert $0,2037$ bzw. für jeden Wert aus dem Intervall $[0,202; 0,206]$ gab es 1 Punkt.

Lösung zu Aufgabe 45 (Korrelationsmessung) (3 Punkte)

0,9375

Herleitung:

Wenn man die KQ-Schätzungen manuell und nicht mit dem Taschenrechner berechnen will, kann man eine Arbeitstabelle anlegen:

i	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	y_i	$y_i - \overline{y}$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
1	25	-25	625	15	-5	25	125
2	35	-15	225	10	-10	100	150
3	55	5	25	25	5	25	25
4	100	50	2500	35	15	225	750
5	35	-15	225	15	-5	25	75
6	50	0	0	20	0	0	0
Σ	300	0	3600	120	0	400	1125

Für den Korrelationskoeffizienten r erhält man dann nach (9.11) aus Kurs 33209:

$$r = \frac{1125}{\sqrt{3600} \cdot \sqrt{400}} = \frac{1125}{1200} = 0,9375.$$

Anmerkung:

Anstelle von $0,9375$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,928; 0,947]$ als richtig anerkannt.

Lösung zu Aufgabe 46 (Normalverteilung) (3 Punkte)

0,6826

Herleitung:

Die Berechnung ist analog zur Lösung von Aufgabe 12.2a in Kurs 33209. Man erhält mit (12.23) und (12.20) für die $N(5; 1^2)$-verteilte Zufallsvariable X zunächst

$$P = \Phi\left(\frac{6-5}{1}\right) - \Phi\left(\frac{4-5}{1}\right) = \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(1)]$$

und daraus mit Tabelle 19.2

$$P = 0,8413 - [1 - 0,8413] = 0,8413 - 0,1587 = 0,6826.$$

Anmerkung:

Anstelle von $0,6826$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,675; 0,690]$ als richtig anerkannt. Die Differenz $\Phi(1) - \Phi(-1) \approx 0,6826$ lässt sich – zumindest in sehr guter Näherung – anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und die Schieber auf $-1,00$ und $+1,00$ einstellen - dies ist schon die Voreinstellung.)

Lösung zu Aufgabe 47 (Standardnormalverteilung) (3 Punkte)

0,8416

Herleitung:

Der gesuchte Wert $z = a$ hat die Eigenschaft, dass die Verteilungsfunktion $\Phi(z)$ für $z = a$ den Wert $0,8$ besitzt (vgl. auch Abbildung 12.3 in Kurs 33209). Zu bestimmen ist also das $0,8$-Quantil der Standardnormalverteilung. Dieses hat nach Tabelle 19.3 in Kurs 33209 den Wert $0,8416$.

Anmerkungen:

Der gesuchte Wert kann direkt aus Tabelle 19.3 in Kurs 33209 abgelesen werden. Etwas ungenauer lässt sich das gesuchte Quantil auch aus Tabelle 19.2 ermitteln, nämlich als ein Wert, der zwischen $0,84$ und $0,85$ liegen muss. Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,833; 0,850]$ als richtig anerkannt.

Der Wert $0,8416$ lässt sich auch anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und im Menüfeld den Wert $0,8$ einstellen).

Aufgabe 48 (Gauß-Test) (3 Punkte)

Die Wahrscheinlichkeit beträgt ca. $0,1446$.

Herleitung:

Die Wahrscheinlichkeit für die Verwerfung der Nullhypothese H_0 ist für jeden Wert des unbekannten Parameters μ, auf den sich der Test bezieht, durch den Wert $G(\mu)$ der Gütefunktion (15.11) gegeben (Gütefunktion des Gauß-Tests im rechtsseitigen Fall). Man erhält nach Einsetzen von $\mu = 0,3$ sowie $\mu_0 = 0$, $\sigma = 1$, $n = 9$ und $z_{0,975} \approx 1,96$ in (15.11):

$$G(0,3) \approx 1 - \Phi(1,96 - 0,3 \cdot 3) = 1 - \Phi(1,06) \approx 1 - 0,8554 = 0,1446.$$

Anmerkungen:

Die Aussage $\Phi(1,06) \approx 0,8554$ lässt sich auch – zumindest in sehr guter Näherung – anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und Schieber auf $z = 1,06$ einstellen.

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,143; 0,146]$ als richtig anerkannt.

Wenn $0,8554$ anstelle von $1-0,8554 = 0,1446$ als Lösung eingetragen wurde, wurden 2 P. vergeben. Die Vergabe von 2 P. wurde auf jeden Wert aus dem Intervall $[0,847; 0,864]$ angewendet.

Wurde anstelle der Gütefunktion (15.11) (Gütefunktion für den *rechtsseitigen* Fall) fälschlich die Gütefunktion (15.12) angewendet (Gütefunktion für den *linksseitigen* Fall), resultierte

$$G(0,3) \approx \Phi(-1,96 - 0,3 \cdot 3) = 1 - \Phi(2,86) \approx 1 - 0,9979 = 0,0023.$$

Wenn dieser Wert bzw. ein Wert aus dem Intervall $[0,0022; 0,0024]$ als Lösung angegeben wurde, gab es 1 Punkt.