

Vorläufige Musterlösungen – Stand: 16. 9. 2013 – zur Klausur
zu Modul M1 im BA-Studiengang
„Politikwissenschaft, Verwaltungswissenschaft, Soziologie“
und zu Modul 3 im BA-Studiengang
„Soziologie“

Termin: 3. September 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):

B, 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: $a \vee \neg b$	P2: $\neg a$	K: $\neg a \wedge \neg b$
w	w	w	f	f
w	f	w	f	f
f	w	f	w	f
f	f	w	w	w

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

Lösung zu Aufgabe 2 (Merkmalsklassifikationen)

(5 Punkte)

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

zu A: Die Ausprägungen ordinalskalierteter Merkmale lassen sich in eine Rangfolge bringen, bei nominalskalierten Merkmalen gilt dies nicht.

Zu E: Metrisch skalierte Merkmale sind immer quantitativ.

Lösung zu Aufgabe 3 (Definitionen, Operationalisierung) (5 Punkte)

A, D, E.

A: vgl. Kromrey, Abschnitt 3.5.1.

B: Eine Nominaldefinition beinhaltet lediglich, dass zwei Gegenstände oder Sachverhalte (das Definiendum und das Definiens) bedeutungsgleich sind. Sie hat keinen empirischen Informationsgehalt (vgl. Kromrey, Abschnitt 3.5.1).

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

D: Realdefinitionen sind Behauptungen über die Beschaffenheit oder das Wesen eines Gegenstands oder Sachverhalts und können richtig oder falsch sein; vgl. Kromrey, Abschnitt 3.5.1.

E: vgl. Kromrey, Abschnitt 4.3.

Lösung zu Aufgabe 4 (Messen, Datenerhebung, Stichprobenverfahren) (5 Punkte)

B, D, E

A: Die *Reliabilität* sagt aus, inwieweit ein Messinstrument bei wiederholter Messung die gleichen Messwerte liefert (Zuverlässigkeit im technischen Sinne). Das Gütekriterium für Messungen, mit dem beschrieben wird, inwieweit wirklich das gemessen wird, was gemessen werden soll, ist die *Validität* (Gültigkeit) – vgl. Kromrey, Abschnitt 5.7 oder Kurs 33209, Abschnitt 2.3.

B - C: Die Klumpenauswahl ist eine zweistufige Auswahlprozedur, bei der die Klumpen diejenigen Einheiten darstellen, auf die sich der Stichprobenauswahlplan bezieht. Die Klumpen stellen also bei diesem Verfahren die *Auswahleinheiten* dar. Es sind die Elemente der Klumpen (z. B. alle Schüler der an einer Schule zufällig ausgewählten Klassen), die die *Erhebungseinheiten* repräsentieren.

D: vgl. Kurs 33209, Abschnitt 3.1.

E: vgl. Kromrey, Abschnitt 6.4.3, oder Kurs 33209, Abschnitt 3.2.

Lösung zu Aufgabe 5 (Kenngrößen von Datensätzen)

(5 Punkte)

B, E.

Zu A: Der Datensatz hat zwei Modalwerte, nämlich 4, 2 und 4, 8.

Zu B: Der Median des Datensatzes ist $\tilde{x} = 4,8$ (fünfter Wert der nach aufsteigender Größe geordneten Urliste des Umfangs $n = 9$), während für den Mittelwert $\bar{x} \approx 5,22$ gilt.

Zu C: Wenn man den kleinsten Wert (3, 9) der Urliste um 1,8 erhöht, vergrößert sich die Summe der Merkmalswerte und damit auch der Mittelwert. Er beträgt dann $\bar{x} \approx 5,42$. Der Median bleibt hingegen unverändert.

Zu D: Nach Streichung des Wertes 3, 9 hat man einen Datensatz des Umfangs $n = 8$. Der Median ist dann der Mittelwert aus dem vierten und fünften Wert der geordneten Liste, also $\tilde{x} = \frac{1}{2}(4,8 + 4,8) = 4,8$. Der Mittelwert wird hingegen größer; für ihn gilt nach Streichung von 3, 9 nun $\bar{x} \approx 5,46$.

Zu E: Da sich mit Streichung des Wertes 6, 5 keiner der beiden Extremwerte des Datensatzes (3, 9 und 7, 6), aus denen sich die Spannweite errechnet, ändert, besitzt die Spannweite nach wie vor den Wert 3, 7.

Lösung zu Aufgabe 6 (Charakterisierung eines Datensatzes)

(5 Punkte)

A, B, D.

Zu C: Der Median ist hier nicht erklärt, weil das Merkmal „Konfessionszugehörigkeit“ nominalskaliert ist und daher keine Rangfolge definiert ist.

Zu E: Für die Konstruktion von Boxplots benötigt man u. a. das untere und obere Quartil. Beide sind hier ebenfalls und aus gleichem Grund nicht erklärt.

Lösung zu Aufgabe 7 (Konzentrationsmessung)

(5 Punkte)

A, C, E.

Zu A: Die Konzentrationsmessung bezieht sich auf die Originaldaten $x_1 = 6$, $x_2 = 24$, $x_3 = x_4 = 15$ (Wert der vier Aktienpakete in Millionen Euro). Die geordnete Liste ist dann $x_{(1)} = 6$, $x_{(2)} = x_{(3)} = 15$, $x_{(4)} = 24$. Mit (6.2) – (6.3) aus Kurs 33209 folgt mit $p_4 = 60$ dann $v_3 = \frac{21}{60} = 0,35$.

Zu B: Der Wert, den man für den Ordinatenwert v_3 errechnet, gibt an, welcher Anteil des gesamten Aktienpakets auf diejenigen 3 Aktionäre entfällt, die die *kleinsten* Anteile am gesamten Aktienpaket halten.

Zu C: Nach (6.6) ist für den unnormierten Gini-Koeffizienten G im Falle $n = 4$ mit $G_{max} = 0,75$ eine obere Schranke gegeben.

Zu D: Die gewichtete Merkmalssumme q_4 aus Formel (6.4) in Kurs 33209 beträgt 177. Nach (6.5) errechnet sich dann für den unnormierten Gini-Koeffizienten der Wert

$$G = \frac{2 \cdot 177}{4 \cdot 60} - \frac{5}{4} = 0,225.$$

Für den normierten Gini-Koeffizienten G^* gilt somit

$$G^* = \frac{G}{G_{max}} = \frac{0,225}{0,75} = 0,3.$$

E: Der neue geordnete Datensatz lautet $x_{(1)} = x_{(2)} = 15$, $x_{(3)} = 30$. Die gewichtete Merkmalssumme q_4 beträgt dann 135 und nach (6.5) errechnet sich nun für den unnormierten Gini-Koeffizienten der Wert

$$G = \frac{2 \cdot 135}{3 \cdot 60} - \frac{4}{3} = \frac{1}{6}.$$

Ferner ist jetzt $G_{max} = \frac{2}{3}$ und folglich

$$G^* = \frac{G}{G_{max}} = \frac{1}{6} \cdot \frac{3}{2} = 0,25.$$

Der Wert des Gini-Koeffizienten wurde also trotz eines Schrumpfens der Aktionärsgruppe etwas kleiner. Dieses Beispiel zeigt, dass der Gini-Koeffizient bei sehr kleinen Datensätzen Schwächen aufweist.

Lösung zu Aufgabe 8 (Randverteilungen, bedingte Wahrscheinlichkeiten) (5 Punkte)

C, D.

Es ist zweckmäßig die im Text enthaltenen Informationen in einer Kontingenztafel mit Randverteilungen zusammenfassen (vgl. in Kurs 33209 die Tabelle 10.2 oder die Lösung zu Teil e von Aufgabe 10.5). Man erhält folgende Vierfeldertafel, bei der die Vorgaben dieser Aufgabe kursiv gesetzt sind (Codierung: Präferenz von Designvariante $D_1 = A$, Präferenz von Variante $D_2 = \bar{A}$, weiblich = B , männlich = \bar{B}):

	weiblich (B)	männlich (\bar{B})	Zeilensummen
Präferenz von D_1 (= A)	48	36	84
Präferenz von D_2 (= \bar{A})	18	45	63
Spaltensummen	66	81	147

Aus der Vierfeldertafel folgt dann:

Zu A: Von den an der Befragung beteiligten 66 Frauen bevorzugten 18 die Designvariante D_2 , also ca. 27,3 %.

Zu B: Die Wahrscheinlichkeit $P(\bar{A})$ dafür, dass von den 147 befragten Personen per Zufallsauswahl eine der 63 Personen gewählt wird, die die Designvariante D_2 bevorzugten, beträgt nach der Formel (10.5) aus Kurs 33209 $P(\bar{A}) = \frac{63}{147} = 0,429$, also ca. 42,9 %.

Zu C: Die Wahrscheinlichkeit $P(\bar{A} \cap B)$ dafür, dass per Zufallsauswahl eine der 18 Frauen gewählt wird, die D_2 bevorzugten, beträgt $P(\bar{A} \cap B) = \frac{18}{147} = 0,122$.

Zu D: von den befragten 63 Personen mit Präferenz für D_2 waren 18 Personen weiblichen Geschlechts, also ca. 28,6 %.

Zu E: Von den 66 weiblichen Befragten bevorzugten 48 Personen D_1 , also ca. 72,7 %, während bei den 81 männlichen Befragten 36 die Variante D_1 bevorzugten, also ca. 44,4 %. Der erstgenannte Prozentsatz ist im Vergleich zum letztgenannten Prozentsatz zwar deutlich größer (ca. 1,6-mal größer), nicht aber doppelt so groß.

Lösung zu Aufgabe 9 (diskrete Zufallsvariablen / Venn-Diagramme) (5 Punkte)

B, C, D, E

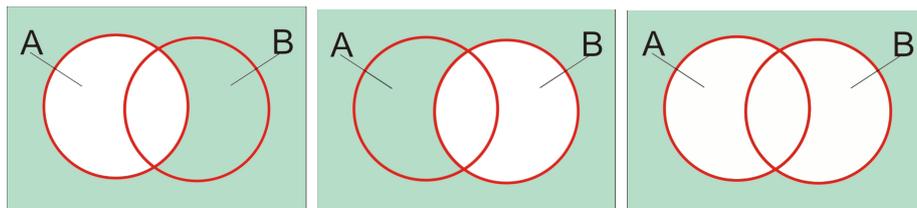
Zu A: Die Wahrscheinlichkeit $P(X \geq 3)$ dafür, mindestens drei Mal „Zahl“ beim 6-maligen Münzwurf zu erhalten, ist die Komplementärwahrscheinlichkeit von $P(X \leq 2)$. Letztere ist durch den Wert der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 6$ und $p = 0,5$ an der Stelle $x = 2$ gegeben, nach Tabelle 19.1 also durch $F(2) = 0,3438$. Es gilt also $P(X \geq 3) = 1 - 0,3438 = 0,6562$.

Zu B: Die Wahrscheinlichkeit dafür, genau drei Mal „Zahl“ zu erhalten, ergibt sich als Differenz der Werte $F(3)$ und $F(2)$ der Verteilungsfunktion der Binomialverteilung mit $n = 6$ und $p = 0,5$, nach Tabelle 19.1 somit als $F(3) - F(2) = 0,6563 - 0,3438 = 0,3125$.

Zu C: Vgl. hierzu die letzten Zeilen von Beispiel 10.1 in Kurs 33209. Von den 36 Elementarereignissen, die den Ereignisraum Ω beim Würfeln mit zwei Würfeln definieren, führen 3 Elementarereignisse, nämlich die Ausgänge $(5; 6)$, $(6; 5)$ und $(6; 6)$ zu einer Augensumme, die nicht kleiner als 11 ist. Die Wahrscheinlichkeit P für die Erzielung einer Augensumme, die mindestens 11 ist, hat somit den Wert $P = \frac{3}{36} = \frac{1}{12} \approx 0,083$.

Zu D: Die Augensumme ist durch $Y = X_1 + X_2$ gegeben, wenn X_1 und X_2 die Augenzahlen der beiden einzelnen Würfel bezeichnen (unabhängige Zufallsvariablen). Der Erwartungswert von X_1 und X_2 ist jeweils 3,5 (vgl. Beispiel 11.2). Für den Erwartungswert $E(Y)$ resultiert dann nach (11.13) der Wert 7.

Zu E: Dass Aussage E zutrifft, lässt sich leichter nachvollziehen, wenn man zunächst je ein Venn-Diagramm für \bar{A} und \bar{B} und erst danach die Schnittmenge $\bar{A} \cap \bar{B}$ zeichnet. Man erhält dann die folgenden drei Diagramme, bei denen die dunkler gefärbten Flächen die drei genannten Mengen repräsentieren:



Das letzte Venn-Diagramm stimmt offenbar mit dem Venn-Diagramm aus Aufgabenteil E überein.

Lösung zu Aufgabe 10 (Wahrscheinlichkeits- und Verteilungsfunktion) (5 Punkte)**A, C.***Kommentar:*

Zu A: Die Zufallsvariable X folgt einer diskreten Gleichverteilung mit $p = 0,1$. Die Wahrscheinlichkeitsfunktion $f(x)$ ist durch (11.1) und die Verteilungsfunktion $F(x)$ durch (11.3) definiert mit $p = 0,1$ und $k = 10$. Für die Wahrscheinlichkeitsfunktion $f(x)$ gilt hier

$$f(0) = f(1) = f(2) = \dots = f(9) = 0,1.$$

Zu B und C: Wenn man die beiden Funktionen $f(x)$ und $F(x)$ grafisch darstellt, resultiert eine Abbildung analog zu Abbildung 11.1. Die Balken im linken Teil von Abbildung 11.1 (Darstellung der Wahrscheinlichkeitsfunktion) setzen aber nun in den Punkten $x_1 = 0, x_2 = 1, \dots, x_{10} = 9$ an und haben je die Länge $0,1$. Entsprechend hat man bei der Darstellung der Verteilungsfunktion $F(x)$ – s. rechter Teil der Abbildung 11.1 – Sprünge in $x_1 = 0, x_2 = 1, \dots, x_{10} = 9$, jeweils mit einer Sprunghöhe von $0,1$. Es gilt somit insbesondere $F(0) = 0,1, F(1) = 0,2$ und $F(2) = 0,3$. Die Treppenfunktion erreicht in $x = 9$ ihr Endniveau 1 .

Zu D: Der Erwartungswert bestimmt sich nach (11.6) zu $\mu = \frac{1}{10} \cdot 45 = 4,5$.

Zu E: Die Varianz von Y hat nach (11.12) einen Wert $V(Y)$, der einem *Viertel* der Varianz $V(X)$ der Zufallsvariablen X entspricht.

Lösung zu Aufgabe 11 (Zusammenhangsmessung)

(5 Punkte)

A, D, E

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

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 . Im Falle $r = 0$ liegt zwar kein linearer Zusammenhang vor, es kann aber durchaus ein nicht-linearer Zusammenhang gegeben sein – vgl. auch Abbildung 9.2 in Kurs 33209.

Zu C: vgl. (9.2) in Kurs 33209. Da die Spaltenzahl m und die Zeilenzahl k der Kontingenztafel hier übereinstimmen ($k = m = 2$), ist in (9.2) für M der Wert 2 einzusetzen.

Zu D: Wenn man den Beurteilungen der Sachbearbeiter A und Bank B jeweils Ränge $rg(x_i)$ bzw. $rg(y_i)$ zuordnet und auch die Rangdifferenzen $d_i = rg(x_i) - rg(y_i)$ ausweist, erhält man die folgende erweiterte Tabelle:

Kreditantrag i	Sachbearbeiter A		Sachbearbeiter B		d_i
	Bewertung x_i	$rg(x_i)$	Bewertung y_i	$rg(y_i)$	
1	4	4	5	4	0
2	7	3	9	1	2
3	9	1	8	2	-1
4	8	2	6	3	-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 hier bei beiden Sachbearbeitern nicht mit der Mehrfachbelegung eines Rangplatzes verbunden ist. Mit (9.16) resultiert

$$r_{SP} = 1 - \frac{6 \cdot [0^2 + 2^2 + (-1)^2 + (-1)^2]}{4 \cdot (16 - 1)} = 1 - 0,6 = 0,4.$$

Zu diesem Ergebnis kommt man natürlich auch – allerdings deutlich umständlicher – bei Anwendung der Formel (9.14).

Zu E: Die Richtigkeit der Aussage folgt aus (13.14) in Verbindung mit (13.13).

Lösung zu Aufgabe 12 (Stetige Verteilungen)

(5 Punkte)

A, C, D, E.

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

Zu B: Nicht die Dichtefunktion, sondern die Verteilungsfunktion der Standardnormalverteilung nimmt an der Stelle 0 den Wert 0,5 an. Diese Aussage lässt sich anhand eines auch auf mobilen Endgeräten lauffähigen statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren. Der Wert $z = 0$ ist hier bereits voreingestellt. Man erkennt, dass die Dichtekurve an der Stelle $z = 0$ einen Wert annimmt, der deutlich unterhalb von 0,5 liegt.

Zu C: Die Wahrscheinlichkeit dafür, dass eine standardnormalverteilte Zufallsvariable den Wert 0 oder – allgemeiner – einen bestimmten Wert x_0 annimmt, besitzt nach (12.8) den Wert 0.

Zu D: Eine t -verteilte Zufallsvariable ist – wie die Standardnormalverteilung – symmetrisch bezüglich des Nullpunkts. Ihre Dichte wird der Standardnormalverteilung mit zunehmender Anzahl von Freiheitsgraden immer ähnlicher. Die Dichtekurve der t -Verteilung wird steiler mit zunehmender Anzahl von Freiheitsgraden steiler, mit abnehmender Anzahl flacher (vgl. Abbildung 12.6 in Kurs 33209).

Zu E: Es gilt $P(X \leq x_{0,05}) = 0,05$ sowie $P(X > x_{0,05}) = 0,95$ und folglich

$$P(x_{0,05} \leq X \leq x_{0,95}) = 0,90.$$

Die letztgenannte Wahrscheinlichkeit wird durch die vom Punkt $x_{0,05}$ bis zum Punkt $x_{0,95}$ gerechnete Fläche zwischen und der x -Achse und der Dichtekurve repräsentiert.

Lösung zu Aufgabe 13 (Vermischte Aussagen zur schließenden Statistik) (5 Punkte)

B, C, D, E.

Zu A: Der Erwartungswert $\mu = E(X)$ von X ergibt sich nach (11.24) mit Einsetzen von $n = 6$, $M = 6$ und $N = 45$ – vgl. auch Beispiel 11.4 in Kurs 33209. Es resultiert $\mu = 0,80$.

Zu B: Die Richtigkeit der Aussage folgt aus der Symmetrie der Dichte der t -Verteilung bezüglich des Nullpunkts.

Zu C: Ein Fehler 1. Art beinhaltet die Ablehnung der Nullhypothese H_0 , obwohl sie zutrifft. Die Nullhypothese trifft hier nur für $\mu = \mu_0$ zu, d. h. nur hier kann ein Fehler 1. Art eintreten. Ein Fehler 2. Art kann nur eintreten, wenn H_0 nicht zutrifft, bei diesem Test also für alle Werte μ mit $\mu \neq \mu_0$. Für ein beliebig gewähltes μ kann nur entweder $\mu = \mu_0$ gelten oder aber $\mu \neq \mu_0$.

Zu D: vgl. hierzu Abbildung 12.6 aus Kurs 33209.

Zu E: Der Erwartungswert der t -Verteilung ist wie der Erwartungswert der Standardnormalverteilung 0 – gleich, wie man die Anzahl der Freiheitsgrade der t -Verteilung wählt.

Anmerkung:

In Kurs 33209 geht aus den Ausführungen in Abschnitt 12.4 hervor, dass der Erwartungswert einer t -verteilten Zufallsvariablen 0 ist. Nicht explizit erwähnt wurde hier, dass dies nur unter der Prämisse gilt, dass für die Anzahl n der Freiheitsgrade der t -Verteilung die Bedingung $n > 1$ gilt. Die Übereinstimmung des Erwartungswerts der t -Verteilung mit dem Erwartungswert der Standardnormalverteilung trifft daher streng genommen nur unter dieser Prämisse zu. Aussage E wurde aus diesem Grunde stets mit 1 Punkt bewertet – unabhängig davon, ob sie als zutreffend oder unzutreffend gekennzeichnet wurde.

Lösung zu Aufgabe 14 (Punkt- und Intervallschätzungen) (5 Punkte)

A, E.

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

Zu B: 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 C: Ein Konfidenzintervall für μ enthält den unbekannt Parameter μ nicht immer

– vgl. etwa Abbildung 14.3 in Kurs 33209, bei der die dunkel markierten Konfidenzintervalle Gegenbeispiele darstellen.

Zu D: Dass die Aussage nicht korrekt ist, kann man direkt aus (14.15) oder auch (14.17) erschließen. Man kommt aber auch mit folgender Überlegung zum selben Ergebnis: Wenn man die Nicht-Überdeckungswahrscheinlichkeit α für den unbekanntem, zu schätzenden Parameter verkleinert und damit weniger Unsicherheit bezüglich der Nicht-Überdeckung zulässt, müssen die Konfidenzintervalle länger sein, damit man „sicherer“ den Parameter μ trifft.

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 15 (Testen, Fehler beim Testen)

(5 Punkte)

B, C, D.

Zu A: Die Ablehnung der Nullhypothese erfolgt bei dem hier zugrunde gelegten linksseitigen Test, wenn für die Realisation z der Prüfgröße Z die Bedingung $z < z_\alpha$ gilt, wegen $\alpha = 0,05$ also bei Unterschreitung von $z_{0,05}$. Wegen $z_{0,05} = -z_{0,95}$ und $z_{0,95} \approx 1,6499$ ist die Nullhypothese im Falle $z < -1,6499$ abzulehnen. Wenn für die Prüfgröße Z der Wert $-1,58$ errechnet wird, erfolgt demnach *keine* Ablehnung von H_0 .

Zu B - C: Die Wahrscheinlichkeit dafür, einen Fehler 1. Art zu begehen, hat bei obigem Test nur für $\mu = \mu_0$, also nur am Rande des Gültigkeitsbereichs von H_0 , den Wert $0,05$. Für alle anderen Werte μ , für die H_0 zutrifft (also für alle μ mit $\mu > \mu_0$), ist die Wahrscheinlichkeit dafür, einen Fehler 1. Art zu begehen, etwas kleiner als $0,05$ (vgl. die Abbildung zur Lösung von Aufgabe 15.2c in Kurs 33209, die sich ebenfalls auf einen linksseitigen Gauß-Test mit $\alpha = 0,05$ bezieht).

Zu D: vgl. hierzu ebenfalls die Abbildung zur Lösung von Aufgabe 15.2c in Kurs 33209.

Zu E: Die Ablehnung von H_0 beinhaltet hier eine korrekte Entscheidung. Ein Fehler 2. Art wäre im Falle $\mu < \mu_0$ die Nicht-Verwerfung von H_0 .

Lösung zu Aufgabe 16 (Korrelationsmessung; Regressionsanalyse) (5 Punkte)**B, D, E.**Zu A: Es gilt nach (16.6) und (16.7) bei Beachtung von $\bar{x} = 3$ und $\bar{y} = 5$

$$\hat{\beta} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{10} (x_i - \bar{x})^2} = \frac{2,2}{4,5} \approx 0,489$$

und damit $\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x} \approx 5 - 0,489 \cdot 3 = 3,533$. Die Regressionsgerade schneidet die y -Achse also etwa in 3,533.

Zu C: Nach (16.18) gilt für das Bestimmtheitsmaß

$$R^2 = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = \frac{2,2^2}{4,5 \cdot 5,8} = \frac{4,84}{26,1} \approx 0,185.$$

Zu D: Dass der Korrelationskoeffizient r positiv sein muss, folgt schon aus (16.18), wenn man dort die Wurzel zieht. Bei der resultierenden Gleichung $r = \frac{s_{xy}}{\sqrt{s_x^2} \cdot \sqrt{s_y^2}}$ ist ja hier der Zähler positiv, während der Nenner grundsätzlich keine negativen Werte annehmen kann. Man kann r natürlich auch explizit angeben:

$$r = \frac{2,2}{\sqrt{4,5} \cdot \sqrt{5,8}} \approx \frac{2,2}{5,109} \approx 0,431.$$

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

Anmerkung:

Zu dieser Aufgabe vgl. auch in Kurs 33209 die Lösungen zu Aufgaben 16.1 und 16.2.

Numerische Aufgaben

Lösung zu Aufgabe 41 (Phi-Koeffizient)

(3 Punkte)

0,135.

Herleitung: Wenn man die Vierfeldertafel um die beiden Randverteilungen ergänzt, erhält man

	b_1	b_2	Zeilensummen
a_1	18	15	33
a_2	6	9	15
Spaltensummen	24	24	48

Es folgt bei Anwendung von Formel (9.8) aus Kurs 33209

$$\Phi = \frac{18 \cdot 9 - 15 \cdot 6}{\sqrt{33 \cdot 15 \cdot 24 \cdot 24}} = \frac{72}{24 \cdot \sqrt{33 \cdot 15}} = \frac{3}{\sqrt{495}} \approx 0,135.$$

Anmerkung:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,133; 0,137]$ als richtig anerkannt.

Lösung zu Aufgabe 42 (Kombinatorik)

(3 Punkte)

2730.

Herleitung: In der Terminologie des Urnenmodells werden $n = 3$ Kugeln aus einer Urne mit $N = 15$ nummerierten Kugeln *ohne Zurücklegen* und wegen der unterschiedlichen Wertigkeit der Preise *mit Berücksichtigung der Anordnung* gezogen (vgl. Tabelle 10.1 in Kurs 33209 bzw. Gleichung (10.7)). Für die Anzahl der Möglichkeiten errechnet man

$$\frac{15!}{(15-3)!} = \frac{15!}{12!} = 13 \cdot 14 \cdot 15 = 2730.$$

Anmerkung:

Wenn man bei der Lösung der Aufgabe zwar Ziehen ohne Zurücklegen, aber fälschlich *ohne Berücksichtigung der Anordnung* zugrunde gelegt hatte, resultierte

$$\binom{15}{3} = \frac{15!}{(12)! \cdot 3!} = \frac{13 \cdot 14 \cdot 15}{1 \cdot 2 \cdot 3} = 455.$$

Für dieses Ergebnis wurde 1 Punkt gewährt.

Lösung zu Aufgabe 43 (Wahrscheinlichkeit bei einem Gesellschaftsspiel) (3 Punkte)**0,5787***Herleitung:*

Die Wahrscheinlichkeit dafür, bei *einmaligem* Würfeln mit einem fairen Würfel *keine* Sechs zu erhalten, beträgt $\frac{5}{6}$. Bei *dreimaligem* Würfeln errechnet sich die Wahrscheinlichkeit für das Auftreten von 0 Sechsen aufgrund der Unabhängigkeit der drei Ausgänge für jeden einzelnen Wurf als

$$P = \left(\frac{5}{6}\right)^3 = \frac{125}{216} \approx 0,5787.$$

Anmerkung:

Anstelle von 0,5787 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,573; 0,584]$ als richtig anerkannt.

Lösung zu Aufgabe 44 (Wahrscheinlichkeit bei einem Glücksspiel)**0,533**

Diese Aufgabe kann auf unterschiedliche Weise gelöst werden, nämlich allein anhand kombinatorischer Überlegungen oder aber unter Rückgriff auf die hypergeometrische Verteilung.

Kombinatorischer Lösungsansatz:

Seien 1, 2, 3 und 4 die Zahlen, die den vier Frauen zugeordnet werden und 5 resp. 6 die Codierungen für die beiden Männer. Es gibt 15 Möglichkeiten zwei verschiedene Zahlen auszuwählen, nämlich

(1; 2), (1; 3); (1; 4), (1; **5**), (1; **6**), (2; 3), (2; 4), (2; **5**), (2; **6**), (3; 4), (3; **5**), (3; **6**), (4; 5), (4; **6**), (5; 6)

Die Gesamtzahl der Möglichkeiten aus der Gruppe von 6 Personen 2 Personen auszuwählen (Ziehen ohne Zurücklegen und *ohne* Berücksichtigung einer Reihenfolge) lässt sich auch nach Tabelle 10.1 ermitteln:

$$\binom{6}{2} = \frac{6!}{4! \cdot 2!} = 15.$$

Unter diesen 15 Wertepaaren sind nur 8 Paare, bei denen nur genau eine der beiden Zahlen 5 und 6 vorkommt (s. Markierungen durch fette Schrift). Man errechnet also mit (10.5) für die gesuchte Wahrscheinlichkeit den Wert $\frac{8}{15} \approx 0,533$. Wenn man in Tabelle 10.1 fälschlich die Formel für den Fall „Ziehen ohne Zurücklegen und *mit* Berücksichtigung einer Reihenfolge“ verwendet, kommt man auf $6^2 = 36$ Möglichkeiten und erhält

dann für die gesuchte Wahrscheinlichkeit Wert $\frac{8}{36} = \frac{2}{9} \approx 0,222$.

Lösung anhand der hypergeometrischen Verteilung:

Die gesuchte Wahrscheinlichkeit lässt sich auch unter Verwendung der hypergeometrischen Verteilung mit Parametern $N = 6$, $M = 2$ und $n = 2$ bestimmen. Der Parameter M entspricht hier der Anzahl der Männer in der Grundgesamtheit, aus der eine Stichprobe gezogen wird. (Die Vorgehensweise ist analog zu der in Beispiel 11.3 von Kurs 33209 vorgenommenen Bestimmung der Wahrscheinlichkeit, beim Lotto „6 aus 49“ keinen einzigen Treffer zu erzielen.)

Man erhält für das Ereignis „eine Frau und ein Mann bilden das Gewinnerpaar“ nach (11.26) bei Einsetzen der genannten Parameter und mit $\binom{2}{1} = 2$ sowie $\binom{4}{1} = 4$

$$f(1) = \frac{\binom{2}{1} \cdot \binom{6-2}{2-1}}{\binom{6}{2}} = \frac{2 \cdot 4}{\binom{6}{2}} = \frac{8}{15} \approx 0,533.$$

Anmerkung:

Anstelle des korrekten Werts 0,533 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,525; 0,540]$ als richtig anerkannt.

Für den Wert 0,222 resp. für jeden Wert aus dem Intervall $[0,525; 0,540]$ wurde 1 P. vergeben.

Lösung zu Aufgabe 45 (Normalverteilung)

(3 Punkte)

1,7816.

Herleitung:

Der Wert x , an der die Verteilungsfunktion $F(x)$ einer Normalverteilung den Wert p annimmt ($0 < p < 1$) ist das p -Quantil der betreffenden Normalverteilung. Es ist also das 0,9-Quantil $x_{0,9}$ der Normalverteilung mit Erwartungswert $\mu = 0,5$ und Varianz $\sigma^2 = 1$ gesucht. Nach Formel (12.26) in Kurs 33209 gilt

$$x_{0,9} = 0,5 + z_{0,9} \cdot 1 = 0,5 + 1,2816 \cdot 1 = 0,5 + 1,2816 = 1,7816.$$

Wegen $F(x) = P(X \leq x)$ beinhaltet dies, dass eine Realisation von X den Wert 1,7816 mit Wahrscheinlichkeit 0,9 nicht überschreitet.

Anmerkung:

Anstelle des Werts 1,7816 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[1,768; 1,795]$ als richtig anerkannt.

Lösung zu Aufgabe 46 (Intelligenzmessung)

(3 Punkte)

0,4967.*Herleitung:*

Nach Formel (12.23) in Kurs 33209 gilt

$$P(94 \leq X \leq 115) = \Phi\left(\frac{115 - 100}{15}\right) - \Phi\left(\frac{94 - 100}{15}\right) = \Phi(1) - \Phi(-0,4) = \Phi(1) - [1 - \Phi(0,4)].$$

Mit Tabelle 19.2 des Kurses 33209 folgt

$$P(94 \leq X \leq 115) = 0,8413 - (1 - 0,6554) = 0,8413 - 0,3446 = 0,4967.$$

Die Wahrscheinlichkeit dafür, dass X zwischen 94 und 115 liegt, beträgt also 0,4967 (entspricht 49,67 %).

Anmerkungen:

Anstelle von 0,4967 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,492; 0,501]$ als richtig anerkannt.

Die Werte $\Phi(1) \approx 0,8413$ sowie $\Phi(0,4) \approx 0,6554$ kann man anhand eines auch auf mobilen Endgeräten lauffähigen statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und resp. 0,40 einstellen oder das interaktive Element direkt via nachstehendem QR-Code aufrufen).

**Lösung zu Aufgabe 47 (F-Verteilung)**

(3 Punkte)

2,58*Herleitung:*

Der gesuchte Wert $z = a$ hat die Eigenschaft, dass die Verteilungsfunktion $F(x)$ der in Rede stehenden F-Verteilung mit $m = 7$ und $n = 18$ Freiheitsgraden für $x = a$ den Wert 0,95 besitzt (vgl. auch Abbildung 12.7 in Kurs 33209, in der lediglich die Fläche rechts von $x = a$ jeweils betont ist). Zu bestimmen ist demnach das 0,95-Quantil $F_{7;18;0,95}$ der F-Verteilung mit $m = 7$ und $n = 18$ Freiheitsgraden. Dieses hat nach Tabelle 19.6 in

Kurs 33209 den Wert 2,58.

Anmerkungen:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[2,575; 2,585]$ als richtig anerkannt.

Der Wert 2,58 lässt sich auch anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und via Schieber die Freiheitsgrade $m = 7$ und $n = 18$ einstellen). Bei dem Experiment ist neben der Dichtefunktion $f(x)$ der gewählten F-Verteilung auch die zugehörige Verteilungsfunktion $F(x)$ zu sehen.

Lösung zu Aufgabe 48 (Gauß-Test)

(3 Punkte)

Die Wahrscheinlichkeit beträgt ca. **0,0485**.

Herleitung:

Die Wahrscheinlichkeit für die Verwerfung der Nullhypothese H_0 ist für jeden Wert des unbekanntes Parameters μ , auf den sich der Test bezieht, durch den Wert $G(\mu)$ der Gütefunktion (15.12) gegeben (Gütefunktion des Gauß-Tests im linksseitigen Fall). Man erhält nach Einsetzen von $\mu = 0,98$ sowie $\mu_0 = 1$, $\sigma = 0,2$ sowie $n = 9$ und $z_{0,975} \approx 1,96$ in (15.12):

$$G(0,98) \approx \Phi[-1,96 - (-0,1) \cdot 3] = \Phi(-1,66) = 1 - \Phi(1,66) \approx 0,0485.$$

Anmerkungen:

Die Lösung zu Aufgabe 48 ist analog zur Lösung von Aufgabe 15.2b in Kurs 33209. Die oben errechnete Wahrscheinlichkeit dafür, H_0 zu verwerfen, ist mit 0,0485 relativ gering, d. h. die Wahrscheinlichkeit dafür einen Fehler 2. Art zu begehen, ist hoch (0,9515). Dass dies auch plausibel ist, kann man im Kurs 33209 anhand der Grafik zur Lösung von Aufgabe 15.2c erkennen, die die Gütefunktionsverläufe für einen mit $\alpha = 0,05$ arbeitenden linksseitigen Gauß-Test im Falle $n = 5$ und $n = 10$ zeigt. Die bei der Klausuraufgabe in Rede stehende Gütefunktion für den mit $n = 9$ arbeitenden linksseitigen Gauß-Test verläuft etwas flacher als die gestrichelte Kurve aus L 15.2c und nimmt in $G(\mu_0)$ den Wert $\alpha = 0,025$ an. Der errechnete Gütefunktionswert 0,0485 entspricht dem Wert der für $n = 9$ resultierenden Gütefunktion an der Stelle $\frac{\mu - \mu_0}{\sigma} = 0,1$. Man erkennt aus der Grafik in L 15.2c, dass die Gütefunktion an dieser Stelle noch nicht deutlich über dem Signifikanzniveau α liegt.

Die Aussage $\Phi(1,66) \approx 0,9515$ lässt sich leicht anhand eines auch auf mobilen Endgeräten lauffähigen statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und im Menüfeld für z den Wert 1,66 einstellen. Alternativ kann das interaktive Element auch direkt von einem Smartphone oder Tablet über den QR-Code aufgerufen werden, der am Ende der Lösung zu Aufgabe 46 wiedergegeben ist – es ist dann im Menüfeld für z der Wert 1,66 einzustellen.

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,048; 0,049]$ als richtig anerkannt.

Falls anstelle von $1 - \Phi(1,66) \approx 0,0485$ fälschlich $\Phi(1,66) \approx 0,9515$ als Ergebnis eingetragen wurde, gab es 1 Punkt. Dies gilt auch für jeden Wert aus dem Intervall $[0,943; 0,961]$.

Wenn für $\Phi[-1,96 - (-0,1) \cdot 3] = \Phi(-1,66)$ fälschlich $\Phi[-1,96 - (-0,1) \cdot 3] = \Phi(-2,26)$ gerechnet wurde (fehlerhafte Verarbeitung des zweifachen Minuszeichens), resultierte

$$G(0,98) = \Phi(-2,26) = 1 - \Phi(2,26) \approx 1 - 0,9881 = 0,0119.$$

Wenn dieser Wert oder ein Wert aus dem Intervall $[0,0118; 0,0120]$ als Lösung angegeben wurde, gab es 1 Punkt.

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

$$G(0,98) \approx 1 - \Phi[1,96 - (-0,1) \cdot 3] = 1 - \Phi(2,26) \approx 1 - 0,9881 = 0,0119.$$

Auch hierfür bzw. für jeden Wert aus dem Intervall $[0,0118; 0,0120]$ gab es 1 Punkt.