FernUniversität in Hagen

LOTSE

Musterlösungen zur Klausur zum

Modul 2.1 im BA-Studiengang „Politik- und Verwaltungswissenschaft“

und zum

Modul 3 im BA-Studiengang „Soziologie“

Termin: 1. März 2011, 14.00 - 18.00 Uhr

Prüfer: apl. Prof. Dr. H.-J. Mittag

Multiple-Choice-Aufgaben

Aufgabe 1 (Aussagenlogik) (5 Punkte)

In der nachstehenden Wahrheitstabelle stehen im Tabellenkopf der ersten beiden Spalten zwei Aussagen a und b, die wahr (w) oder falsch (f) sind. In den beiden ersten Spalten sind alle möglichen Fälle bezüglich des Wahrheitsgehalts der beiden Aussagen angegeben (beide Aussagen „w“, nur eine Aussage „w“, beide Aussagen „f“).

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

Bei den letzten drei Spalten ist nur der Tabellenkopf ausgefüllt. Hier sind zwei Prämissen $P1$ und $P2$ angegeben. Ob diese erfüllt sind, hängt davon ab, ob die Aussagen a und b wahr oder falsch sind. Im Kopf der letzten Spalte steht ein aus den Prämissen abgeleiteter logischer Schluss K, der allerdings noch auf seinen Wahrheitsgehalt zu überprüfen ist. Das Zeichen $\neg$ bedeutet die Negation einer Aussage, $\wedge$ (Konjunktion) beinhaltet ein logisches „und“ (zwei Aussagen gelten gleichzeitig), während das Zeichen $\vee$ (Disjunktion) ein logisches „oder“ darstellt (von zwei Aussagen gilt mindestens eine).

Welche der folgenden Aussagen sind richtig ? (x aus 5)

A) Wenn die Aussagen a und b beide wahr sind, sind auch die Prämissen $P1$ und $P2$ erfüllt, also beide wahr.

B) Wenn die Aussage a falsch und b wahr ist, ist $P1$ erfüllt, nicht aber $P2$.

C) Wenn die Aussage a wahr und b falsch ist, sind weder $P1$ noch $P2$ erfüllt.

D) Wenn die Aussagen a und b beide falsch sind, ist $P2$ erfüllt, nicht aber $P1$.

E) Wenn $P1$ und $P2$ wahr sind, ist auch K wahr.

Hinweis:

Ergänzen Sie am besten zunächst die 12 fehlenden Werte der obigen Tabelle, d. h., setzen sie jeweils w oder f ein. Die ausgefüllte Tabelle geht zwar nicht in die maschinelle Bewertung ein, erleichtert es Ihnen aber, die zutreffenden der fünf vorgegebenen Antwortalternativen zu finden.

Lösung: B, E.

Zu A - D: In den vier Tabellenzeilen sind die drei Werte in den letzten drei Spalten wie folgt zu ergänzen: Zeilen 1 und 2: f, w, f; Zeile 3: w, f, f; Zeile 4: w, w, w (vgl. auch Kurs 33210; Kapitel 5). Wenn man diese Ergänzungen vorgenommen hat, kann man aus den Zeilen 1, 2 und 4 der Tabelle sofort erkennen, dass die Aussagen aus Aufgabenteil A, C und D nicht zutreffen und aus Zeile 3, dass die Aussage aus Aufgabenteil B zutrifft.

Zu E: Wenn $P1$ und $P2$ beide wahr sind, muss bei $P2$ die hinter dem Zeichen $\vee$ stehende Aussage $\neg b$ zutreffen – die Gültigkeit der vor dem Zeichen $\vee$ stehenden Aussage a ist ja wegen der vorausgesetzten Gültigkeit von $P1$ ausgeschlossen. Mit $P1$ und $P2$ gilt also sowohl $\neg a$ als auch $\neg b$ und dies beinhaltet zusammen die Gültigkeit von K.

Aufgabe 2 (Datengewinnung) (5 Punkte)

Welche der folgenden Aussagen sind richtig? Bei den Aufgabenteilen A und D geht es jeweils um die Beurteilung der Richtigkeit des letzten Satzes. (x aus 5)

A) Das Statistische Bundesamt (DeStatis) erfasst automatisiert die Nutzungshäufigkeit von DeStatis-Datenangeboten zu einzelnen Themenbereichen, um das Verhalten der Datennutzer zu verfolgen und auszuwerten. Diese Art der Gewinnung von Daten zum Nutzerverhalten ist ein Beispiel für nicht-reaktive Datenerhebung.

B) Wenn man allgemeine Bevölkerungsumfragen als offene Online-Befragungen organisiert, ist mit systematisch verzerrten Ergebnissen zu rechnen.

C) Das Random-Route-Verfahren ist ein Verfahren, mit dem sich Zufallsbevölkerungsstichproben ohne Rückgriff auf Namens- oder Adressdateien gewinnen lassen.

D) Bei Experimenten mit Personen, z. B. in der Psychologie, werden i. a. zwei Gruppen von Teilnehmern gebildet – eine Versuchsgruppe und eine Kontrollgruppe. Anschließend werden Einflussgrößen in beiden Gruppen planmäßig verändert.

E) „Auswahleinheit“ und „Erhebungseinheit“ sind Begriffe, die inhaltlich dasselbe bedeuten, also synonym verwendet werden können.

Lösung: A, B, C.

Zu A: Diekmann, Kapitel XIII, Abschnitt 1, oder Kurs 33209, Abschnitt 3.1.

zu B: Diekmann, Kapitel X, Abschnitt 11.

Zu C: vgl. Kromrey, Abschnitt 6.5.3.

Zu D: Das „Treatment“ erfolgt nur in der Versuchsgruppe – vgl. Kromrey, Abschnitt 2.4.3 und Kurs 33209, Abschnitt 3.1.

Zu E: vgl. Kromrey, Abschnitt 6.1.

Aufgabe 3 (Merkmalsklassifikationen) (5 Punkte)

Welche der folgenden Aussagen sind richtig? Bei den Aufgabenteilen B und D geht es jeweils um die Beurteilung der Richtigkeit des letzten Satzes. Der aus zwei Teilaussagen bestehende letzte Satz in Aufgabenteil D gilt als richtig, wenn jede Teilaussage zutrifft. Der hier und in Aufgabenteil E verwendete Begriff „metrische Skala“ ist als Oberbegriff für „Intervallskala“, „Verhältnisskala“ und „Absolutskala“ zu verstehen. (x aus 5)

A) Operationen, die für ordinalskalierte Daten zulässig sind, sind auch für nominalskalierte Daten zulässig.

B) Bei einer Einkommenserhebung wird u. a. der Bildungsstand von Arbeitnehmern erfasst und zwar anhand des höchsten erreichten Bildungsabschlusses (Ausprägungen des Merkmals „Bildungsstand“: ohne Schulabschluss, Hauptschule, mittlere Reife, Fachhochschulreife, Abitur, akademischer Abschluss). Das Merkmal „Bildungsstand“ ist ordinalskaliert.

C) Das in Aufgabenteil B spezifizierte Merkmal „Bildungsstand“ ist diskret.

D) Die Mitarbeiter einer Firma werden gebeten, die Entfernung zwischen Wohnung und Firma anzugeben sowie das für die Fahrt zur Arbeit überwiegend benutzte Transportmedium (z. B. „PKW“ oder „Fahrrad“). Das Merkmal „Entfernung“ ist metrisch skaliert, das Merkmal „Transportmedium“ nominalskaliert.

E) Metrisch skalierte Merkmale können sowohl qualitativ als auch quantitativ sein.

Lösung: B, C, D.

Zu A: Die Ausprägungen ordinalskalierter Merkmale lassen sich in eine Rangfolge bringen, bei nominalskalierten Merkmalen gilt dies nicht – vgl. Kromrey, Abschnitt 5.4.3, oder Schnell / Hill /Esser, Abschnitt 4.3 oder Kurs 33209, Abschnitt 2.2.

Zu B - D: vgl. Kromrey, Abschnitt 5.4.3, oder Schnell / Hill /Esser, Abschnitt 4.3 oder Kurs 33209, Abschnitt 2.2.

Zu E: Metrisch skalierte Merkmale sind immer quantitativ.

Aufgabe 4 (Messen / Stichprobenverfahren) (5 Punkte)

Welche der folgenden Aussagen sind richtig? Bei Aussage D geht es um die Beurteilung des Wahrheitsgehalts des letzten Satzes. (x aus 5)

A) Die Reliabilität eines Messinstruments charakterisiert, inwieweit ein Messinstrument bei wiederholter Messung die gleichen Messwerte liefert.

B) Wenn eine Messung dem Gütekriterium „Reliabilität“ genügt, genügt sie auch dem Gütekriterium der „Validität“.

C) Die Klumpenauswahl ist eine zweistufige Auswahlprozedur, bei der in der ersten Stufe Teilmengen der Grundgesamtheit (sog. Klumpen) zufällig ausgewählt werden, deren Elemente dann – in der zweiten Verfahrensstufe – alle für die Datenerhebung herangezogen werden.

D) Eine Grundgesamtheit von $N = 10.000$ (10 Tausend) Personen wird bezüglich des Merkmals „Höchster erreichter Bildungsabschluss“ in drei Teilpopulationen zerlegt, wobei letztere $N_1 = 2.600$, $N_2 = 5.200$ und $N_3 = 2.200$ Personen umfassen. Aus den Teilpopulationen werden dann in der zweiten Verfahrensstufe Zufallsstichproben des Umfangs $n_1 = 78$, $n_2 = 156$ resp. $n_3 = 44$ gezogen. Das damit praktizierte Auswahlverfahren repräsentiert eine geschichtete Stichprobenauswahl mit proportionaler Schichtung.

E) Das Quotenauswahlverfahren ist ein nicht-zufallsgesteuertes Verfahren zur Gewinnung einer Stichprobe, das z. B. bei Befragungen in der Markt- und Meinungsforschung Anwendung findet.

Lösung: A, C, E.

Zu A: vgl. Kromrey, Abschnitt 5.7, und Schnell / Hill / Esser, Abschnitt 4.3.

Zu B: 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 C: vgl. Kromrey, Abschnitt 6.5.2, oder Kurs 33209, Abschnitt 3.2.

zu D: Der in der zweiten Stufe angewandte Auswahlsatz $\frac{n_i}{N_i}$ $(i = 1, 2, 3)$ ist nicht fest (dies wäre die Voraussetzung für eine proportionale Schichtung) – er ist bei der dritten Teilpopulation niedriger als bei den beiden ersten Teilpopulationen – vgl. Kromrey, Abschnitt 6.5.2 oder Kurs 33209, Abschnitt 3.2 (dort insbesondere Abbildung 3.4).

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

Aufgabe 5 (Definitionen) (5 Punkte)

Welche der folgenden Aussagen sind zutreffend? Bei Aussage E geht es um die Beurteilung des Wahrheitsgehalts des zweiten Satzes. (x aus 5)

A) Eine Realdefinition beinhaltet eine Aussage über Eigenschaften eines Gegenstands oder Sachverhalts.

B) Mit einer Nominaldefinition wird ein Gegenstand oder Sachverhalt (das Definiendum) durch einen anderen Gegenstand oder Sachverhalt (das Definiens) erklärt.

C) Eine Realdefinition umfasst alle Eigenschaften des Definiendums.

D) Sowohl Nominaldefinitionen als auch Realdefinitionen sind entweder richtig oder falsch.

E) Wenn man theoretische Konstrukte (z. B. die latente Variable „Leistungsmotivation“) messen will, muss man zuerst eine Operationalisierung vornehmen. Letztere beinhaltet die Festlegung von Handlungsanweisungen, mit denen sich den betreffenden Variablen Ausprägungen beobachtbarer Variablen zuordnen lassen.

Lösung: A, B, E.

Zu A: Kromrey, Abschnitt 3.5.4.

Zu B: Kromrey, Abschnitt 3.5.1;

Zu C: Kromrey, Abschnitt 3.5.1;

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

Zu E: Kromrey, Abschnitt 4.3.

Aufgabe 6 (Univariate Häufigkeitsverteilungen / Kenngrößen) (5 Punkte)

Gegeben sei der folgende Datensatz für ein Merkmal X:
$5,2 \quad 6,4 \quad 4,2 \quad 4,6 \quad 4,8 \quad 3,9 \quad 6,1 \quad 7,1 \quad 4,2 \quad 7,6 \quad 6,5.$

Welche der folgenden Aussagen, die alle von diesem Datensatz ausgehen, sind richtig? Beachten Sie, dass eine aus mehreren Teilaussagen bestehende Aussage nur dann als richtig zu bewerten ist, wenn jede Teilaussage zutrifft. (x aus 5)

A) Der obige Datensatz hat einen eindeutig bestimmten Modalwert.

B) Wenn man bei obigem Datensatz den letzten Wert $(6,5)$ der Urliste um $0,6$ erhöht und gleichzeitig den ersten Wert $(5,2)$ um $0,5$ vermindert, hat dies zur Folge, dass der Mittelwert $\overline{x}$ des Datensatzes größer und der Median $\widetilde{x}$ kleiner wird.

C) Mit der in Aufgabenteil B spezifizierten Veränderung des ersten und letzten Wertes der originären Urliste verändert sich die Spannweite des Datensatzes.

D) Wenn man bei obigem Datensatz den letzten Wert $(6,5)$ streicht, werden sowohl der Mittelwert $\overline{x}$ als auch der Median $\widetilde{x}$ des Datensatzes kleiner.

E) Wenn man alle Werte des obigen Datensatzes halbiert, geht die empirische Standardabweichung s des Datensatzes auf die Hälfte ihres Ausgangswertes zurück.

Lösung: A, B, D, E.

Zu A: Der Datensatz hat den eindeutig bestimmten Modalwert $4,2$.

Zu B: Der Median des Ausgangsdatensatzes ist $\widetilde{x} = 5,2$ (sechster Wert der nach aufsteigender Größe geordneten Urliste des Umfangs $n = 11$), während für den Mittelwert $\overline{x} \approx 5,51$ gilt. Wenn man den letzten Wert $(6,5)$ der Urliste um $0,6$ erhöht und gleichzeitig den ersten Wert $(5,2)$ um $0,5$ vermindert, hat der sechste Wert des geordneten neuen Datensatzes den Wert $4,8$, d. h. der Median wird kleiner. Der Mittelwert vergrößert sich dagegen, weil sich die Summe der Merkmalswerte vergrößert und n gleich bleibt.

Zu C: Da sich die Extremwerte des Datensatzes $(3,9$ und $7,6)$ nicht verändern, bleibt die Spannweite unverändert $(R = 3,7)$.

Zu D: Nach Streichung des Wertes $6,5$ hat man einen Datensatz des Umfangs $n = 10$. Der Median ist dann der Mittelwert aus dem fünften und sechsten Wert der geordneten Liste, also $\widetilde{x} = \frac{1}{2}(4,8 + 5,2) = 5,0$. Auch der Mittelwert wird kleiner; für ihn gilt nach Streichung von $6,5$ nun $\overline{x} = 5,41$.

Zu E: Bei Halbierung aller Merkmalswerte geht die empirische Varianz (unkorrigierte oder korrigierte Version) auf ein Viertel des Ausgangswertes zurück, die empirische Standardabweichung folglich auf die Hälfte ihres Ausgangswertes.

Aufgabe 7 (absolute und relative Häufigkeiten) (5 Punkte)

Das Statistische Amt eines EU-Staates teilt mit, dass die Gesamtbevölkerung des betreffenden Landes Anfang 2010 bei $36,0$ Millionen gelegen habe, wobei $49,5$ % der Gesamtbevölkerung (Kinder eingerechnet) Männer seien. Ferner wird bekannt gegeben, dass der Anteil der als erwerbstätig registrierten Männer sich auf $58,0$ % belaufe, während bei den Frauen nur $44,5$ % als erwerbstätig gemeldet seien.

Welche der folgenden Aussagen sind richtig: (x aus 5)

A) Die Anzahl der erwerbstätigen Frauen liegt unter $8,2$ Millionen.

B) Die Anzahl der erwerbslosen Männer liegt zwischen $7,4$ und $7,5$ Millionen.

C) Der Anteil der erwerbstätigen Männer an der Gesamtbevölkerung beträgt mehr als $29,5$ %.

D) Die Gesamtzahl der Erwerbstätigen liegt über $18,7$ Millionen.

E) Wenn man die Anzahl der Erwerbstätigen und die der Erwerbslosen zunächst für die Personen männlichen Geschlechts und danach auch für die Personen weiblichen Geschechts addiert, erhält man die absolute Randverteilung für das binäre Merkmal „Geschlecht“.

Lösung: A, B, E - vgl. auch Beispiel 8.2 in Kurs 33209. Die Lösungen findet man am einfachsten anhand eines Baumdiagramms (nachstehend wiedergegeben).

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

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. Baumdiagramm oder die zweite der nachstehenden Vierfeldertafeln). Dies entspricht $28,71$ %. Die Aussage ist also nicht zutreffend.

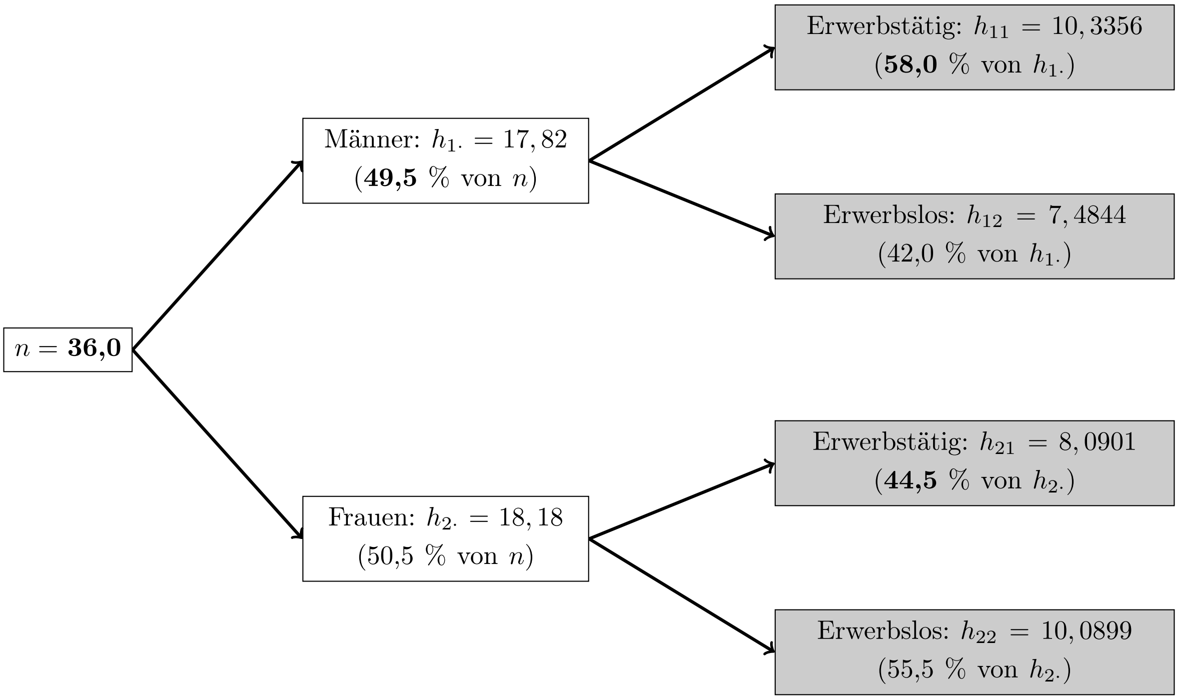
Zu D: Die Aussage bezieht sich auf die (in Millionen ausgewiesene) Randhäufigkeit $h_{\cdot 1} = h_{11} + h_{21}$, also auf $h_{\cdot 1} = 10,3356 + 8,091 = 18,4257$, die unterhalb von $18,7$ liegt.

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

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

	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



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

Aufgabe 8 (Konzentrationsmessung) (5 Punkte)

Die folgende Aufgabe ist adaptiert aus Toutenburg / Schomaker / Wißmann (2006), Arbeitsbuch zur deskriptiven und induktiven Statistik, Springer Verlag, Heidelberg:

An einem Gymnasium werden fünf Schüler einer Mittelstufenklasse nach ihrem monatlichen Taschengeld befragt. Dabei ergaben sich folgende Werte in Euro:

$x_1 = 20$, $x_2 = 40$, $x_3 = 50$, $x_4 = 65$ und $x_5 = 80$.

Welche der folgenden Aussagen sind richtig ? (x aus 5)

A) Wenn man auf der Basis des obigen Datensatzes die Lorenzkurve zeichnet – also einen Polygonzug, der den Nullpunkt mit den Punkten $(0,2; v_1)$, $(0,4; v_2)$, $(0,6; v_3)$, $(0,8; v_4)$ und $(1; 1)$ verbindet – so nimmt diese Kurve an der Stelle $0,4$ einen Wert an, der zwischen $0,21$ und $0,22$ liegt.

B) Der Wert, den man für den Ordinatenwert v_2 errechnet, gibt an, welcher Anteil des gesamten monatlichen Taschengeldes aller 5 Schüler auf die beiden Schüler mit dem geringsten monatlichen Taschengeld entfällt.

C) Der unnormierte Gini-Koeffizient hat bei obigem Datensatz einen Wert G, der unter $0,23$ liegt.

D) Der normierte Gini-Koeffizient hat bei obigem Datensatz einen Wert G^*, der zwischen $0,26$ und $0,27$ liegt.

E) Wenn man bei dem Datensatz $x_1 = 20$, $x_2 = 40$, $x_3 = 50$, $x_4 = 65$ und $x_5 = 80$ alle Werte verdoppelt, bleibt der Wert des normierten und auch des unnormierten Gini-Koffizienten unverändert.

Lösung: B, C, 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 Merkmalssumme).

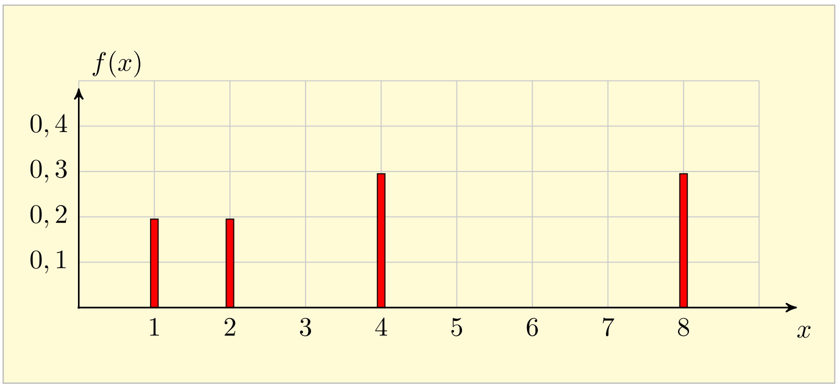
Zu C: Die gewichtete Merkmalsssumme q_5 aus Formel (6.4) in Kurs 33209 beträgt 910. Nach (6.5) errechnet sich dann für unnormierten Gini-Koeffizienten bei Rundung auf drei Dezimalstellen der Wert $G = \frac{2 \cdot 910}{5 \cdot 255} - \frac{6}{5} \approx 0,227$.

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 Verdoppelung der Merkmalswerte führt zu einer Verdoppelung sowohl der Merkmalsumme p_5 als auch der gewichteten Merkmalssumme (6.4). Man erkennt aus der Formel (6.5) in Kurs 33209, dass dies Verdoppelung keinen Effekt auf G und damit auch keinen Effekt auf G^* hat.

Aufgabe 9 (Wahrscheinlichkeits- und Verteilungsfunktion) (5 Punkte)

Die nachstehende Abbildung zeigt die Wahrscheinlichkeitsfunktion $f(x)$ einer diskreten Zufallsvariablen X, die vier Ausprägungen $x_1 = 1$, $x_2 = 2$, $x_3 = 4$ und $x_4 = 8$ aufweist. Die Ausprägungen x_1 und x_2 weisen jeweils die Eintrittswahrscheinlichkeit $0,2$ auf, x_3 und x_4 je die Eintrittswahrscheinlichkeit $0,3$.



Welche der folgenden Aussagen sind richtig? (x aus 5)

A) Die Verteilungsfunktion $F(x)$ der diskreten Zufallsvariablen X nimmt für $x = 2$ den Wert $0,4$ an.

B) Die Verteilungsfunktion $F(x)$ von X nimmt für $x = 3$ den Wert $0,4$ an.

C) Die Verteilungsfunktion $F(x)$ von X ist nur bis $x = 8$ definiert.

D) Der Erwartungswert $E(X)$ der Zufallsvariablen X ist größer als $3,9$, aber kleiner als $4,1$.

E) Wenn man die Zufallsvariable X der Lineartransformation $Y = X + 1$ unterzieht, so ist die Varianz der transformierten Variablen Y identisch mit der der ursprünglichen Variablen X.

Lösung: A, B, E.

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

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

- 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 $F(2) = 0,4$ und $F(3) = 0,4$. Die Verteilungsfunktion

$F(x)$ ist auch für $x > 8$ definiert. Sie erreicht an der Stelle $x = 8$ den Wert 1, den sie dann für $x > 8$ beibehält.

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

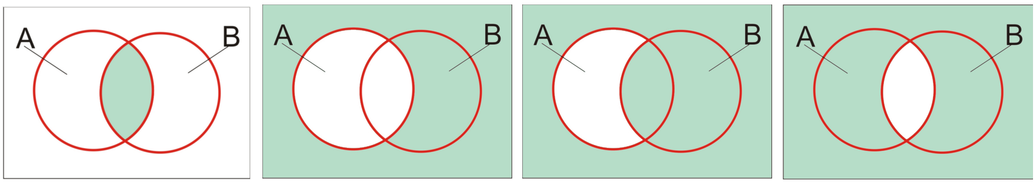
$$\mu = E(X) = 1 \cdot 0,2 + 2 \cdot 0,2 + 4 \cdot 0,3 + 8 \cdot 0,3 = 4,2.$$

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

Aufgabe 10 (Venn-Diagramme) (5 Punkte)

Zur Veranschaulichung von Ereignissen oder von Mengen lassen sich Venn-Diagramme heranziehen. Diese bestehen aus einem Rechteck, in dem die Ereignisse als Kreise oder Ellipsen dargestellt sind. Das Rechteck repräsentiert eine Grundgesamtheit, von der die eingezeichneten Mengen Teilmengen sind. Es bezeichnen $\overline{A}$ und $\overline{B}$ die Komplementärmengen von A und B, $A \cap B$ deren Schnittmenge und $A \cup B$ die Vereinigungsmenge von A und B.

Nachstehend sind vier Venn-Diagramme abgebildet, die sich auf die Verknüpfung zweier Ereignisse oder Mengen A und B beziehen.



Welche der folgenden Aussagen, die sich auf die obigen Diagramme beziehen, sind richtig? (x aus 5)

A) Das erste Venn-Diagramm (von links gezählt, also in der üblichen Leserichtung) veranschaulicht anhand der dunkler gefärbten Fläche die Schnittmenge von A und B, also $A \cap B$.

B) Die dunkel gefärbte Fläche im zweiten Venn-Diagramm stellt die Komplementärmenge $\overline{A}$ von A dar.

C) Im dritten Venn-Diagramm ist durch die dunkler gefärbte Fläche die Vereinigungsmenge aus $\overline{A}$ und B dargestellt, also $\overline{A} \cup B$.

D) Bildet man aus den beiden im zweiten und dritten Venn-Diagramm dargestellten Mengen die Schnittmenge, so resultiert die Menge, die im zweiten Venn-Diagramm dargestellt ist.

E) Im vierten Venn-Digramm ist durch die dunkler gefärbte Fläche die Schnittmenge der Komplementärmengen von A und B dargestellt, also $\overline{A} \cap \overline{B}$.

Lösung: A, B, C, D.

Zu A und C: vgl. hierzu Kurs 33210, Kapitel 5, und Kurs 33209, dort Abschnitt 10.1 und Aufgabe 10.1.

Zu D: Die Mengen, die im zweiten und dritten Venn-Diagramm dargestellt sind, unterscheiden sich offenbar nur um den Überlappungsbereich von A und B. Die Schnittmenge dieser beiden Mengen ist demnach die Menge, die im zweiten Venn-Diagramm visualisiert ist (hier ist ja der Überlappungsbereich gerade ausgespart), d. h. die Schnittmenge der im zweiten und dritten Venn-Diagramm dargestellten Mengen ist also durch $\overline{A}$ gegeben.

Zu E: Die dunkler gefärbte Fläche stellt die *Vereinigungsmenge* der Komplementärmengen von A und B dar, also $\overline{A} \cup \overline{B}$.

Aufgabe 11 (Wahrscheinlichkeiten, bedingte Wahrscheinlichkeiten) (5 Punkte)

An einer gemeinsamen Statistikklausur, die nur für Studierende der BA-Studiengänge „Politik- und Verwaltungswissenschaft“ (POL) und „Soziologie“ (SOZ) angeboten wird, haben in Berlin 80 Personen teilgenommen, davon 45 mit Studienfach POL. Nach der Klausurauswertung stellte sich heraus, dass 56 Studierende bestanden haben. Von den erfolgreichen Klausurteilnehmern waren 32 Studierende des Fachs POL.

Welche der folgenden Aussagen sind richtig? (x aus 5)

A) Die Wahrscheinlichkeit dafür, dass eine nach Abschluss der Auswertungen aus dem Stapel mit allen 80 Klausuren zufällig herausgegriffene Klausur mit „nicht bestanden“ bewertet wurde, ist kleiner als $0,3$.

B) Die Wahrscheinlichkeit dafür, dass eine zufällig herausgegriffene Klausur mit „bestanden“ bewertet und einem Studierenden des Fachs POL zuzuordnen ist, ist größer als $0,38$.

C) Die bedingte Wahrscheinlichkeit dafür, dass eine der mit dem Ergebnis „bestanden“ bewerteten Klausuren die Klausur eines POL-Studierenden ist, ist größer als $0,55$.

D) Wenn man für die beiden Merkmale „Studienfach“ (Ausprägungen „POL“ und „SOZ“) und „Klausurleistung“ (Ausprägungen „bestanden“ und „nicht bestanden“) eine Vierfeldertafel für absolute Häufigkeiten mit den beiden Randverteilungen anlegt, so besteht die Randverteilung jedes Merkmals aus zwei Elementen, deren Summe jeweils 80 ergibt.

E) Wenn man für die beiden in Aufgabenteil D genannten Merkmale „Studienfach“ und „Klausurleistung“ eine Vierfeldertafel für relative Häufigkeiten anlegt, wieder mit den beiden Randverteilungen, so besteht die Randverteilung des Merkmals „Studienfach“ aus zwei Werten, von denen einer der Wert $0,4375$ ist.

	Klausur bestanden (B)	Klausur nicht bestanden ($\overline{B}$)	Zeilensummen
POL(A)	32	*13*	45
SOZ ($\overline{A}$)	*24*	*11*	*35*
Spaltensummen	*56*	24	80

	Klausur bestanden (B)	Klausur nicht bestanden ($\overline{B}$)	Zeilensummen
POL(A)	$0,4$	*0,1625*	$0,5625$
SOZ ($\overline{A}$)	*0,3*	*0,1375*	*0,4375*
Spaltensummen	*0,7*	$0,3$	1

Lösung: B, C, D, E.

Zu A: Die Wahrscheinlichkeit $P(\overline{B})$ dafür, dass von den 80 Klausuren bei Zufallsauswahl einer Klausur eine der insgesamt 24 Klausuren mit dem Ausgang „nicht bestanden" gewählt wird, beträgt nach Formel (10.5) aus Kurs 33209 offenbar $P(\overline{B}) = \frac{24}{80} = 0,3$.

Zu B: Die Wahrscheinlichkeit $P(A \cap B)$ dafür, dass von den 80 Klausuren eine Klausur gewählt wird, die mit „bestanden" bewertet und einem Studierenden der POL zuzuordnen ist, beträgt $P(A \cap B) = \frac{32}{80} = 0,4$.

Zu C: Für die Wahrscheinlichkeit $P(A|B)$ gilt, dass $P(A|B) = \frac{32}{56} \approx 0,571$.

Zu D: vgl. hierzu die Tabelle 8.4 in Kurs 33209.

Zu E: Wenn man bei der oben auf dieser Seite wiedergegebenen Vierfeldertafel für absolute Häufigkeiten alle Werte im Inneren der Tafel und auch die Randverteilungen durch $n = 80$ dividiert, erhält man die darunter stehende Vierfeldertafel für relative Häufigkeiten (mit Randverteilungen). Man erkennt aus dieser z. B., dass die resultierende Randverteilung für das Merkmal „Studienfach" aus den Elementen $0,5625$ und $0,4375$ besteht, Aussage E also zutrifft.

Anmerkung:

Es gibt für diese Aufgabe mehrere Lösungsmöglichkeiten. Ein besonders einfacher Ansatz – ohne Heranziehung des Satzes von Bayes – beinhaltet, dass man zunächst alle Informationen in einer Vierfeldertafel für absolute oder relative Häufigkeiten zusammenfasst (vgl. auch Beispiel 10.4 in Kurs 33209) – siehe ganz oben auf dieser Seite. Es sind dort die *kursiv* gesetzten Zahlen solche, die nur indirekt gegeben waren. Es werden die Codierungen POL= A, SOZ = $\overline{A}$, bestanden = B, nicht bestanden = $\overline{B}$ verwendet.

Aufgabe 12 (Stetige Verteilungen) (5 Punkte)

Welche der folgenden Aussagen sind richtig? (x aus 5)

A) Für eine standardnormalverteilte Zufallsvariable Z gilt, dass eine Realisation von Z mit Wahrscheinlichkeit $0,05$ außerhalb des Intervalls $[-1,96; 1,96]$ liegt.

B) Ist X eine mit Erwartungswert 1 und Varianz 4 normalverteilte Zufallsvariable, so besitzt die Wahrscheinlichkeit $P(0 \leq X \leq 2)$ dafür, dass X zwischen 0 und 2 liegt, einen Wert, der größer als $0,4$ ist.

C) Bezeichnet $[-a; a]$ ein Intervall, in das die Ausprägung einer standardnormalverteilten Zufallsvariablen Z mit Wahrscheinlichkeit $0,9$ fällt, so ist die Wahrscheinlichkeit, dass eine mit 4 Freiheitsgraden t-verteilte Zufallsvariable X in das Intervall $[-a; a]$ fällt, größer als $0,9$.

D) Die Wahrscheinlichkeit dafür, dass eine standardnormalverteilte oder t-verteilte Zufallsvariable einen positiven Wert annimmt, beträgt $0,5$.

E) Die Wahrscheinlichkeit dafür, dass eine mit $m = 10$ und $n = 15$ Freiheitsgraden F-verteilte Zufallsvariable X eine Ausprägung annimmt, die oberhalb von $2,7$ liegt, ist kleiner als $0,05$.

Lösung: A, D, E.

Zu A: Eine standardnormalverteilte Zufallsvariable Z liegt mit Wahrscheinlichkeit $0,95$ in dem durch das $0,05$-Quantil $z_{0,025}$ und das $0,975$-Quantil $z_{0,975}$ begrenzten Intervall $[z_{0,025}; z_{0,975}] = [-1,96; 1,96]$ – vgl. auch in Kurs 33209 die Abbildung 12.4 (mit $\alpha = 0,05$) und Beispiel 12.2.

Zu B: Mit Gleichung (12.23) aus Kurs 33209 verifiziert man, dass für die $N(1; 2^2)$-verteilte Zufallsvariable X gilt

$$P(0 \leq X \leq 2) = \Phi\left(\frac{2-1}{2}\right) - \Phi\left(\frac{0-1}{2}\right) = \Phi(0,5) - \Phi(-0,5).$$

Mit (12.20) und Tabelle 19.2 erhält man

$$P(0 \leq X \leq 2) = \Phi(0,5) - [1 - \Phi(0,5)] \approx 2 \cdot 0,6915 - 1 = 0,383.$$

Zu C: Die Dichte der Standardnormalverteilung verläuft etwas flacher als die einer t-verteilten Zufallsvariablen, wobei der Unterschied bei kleiner Anzahl von Freiheitsgraden (z. B. bei nur 4 Freiheitsgraden) ausgeprägter ist – vgl. hierzu Abbildung 12.6 in Kurs 33209. Bezeichnet $[-a; a]$ ein Intervall, in das die Ausprägung einer standardnormalverteilten Zufallsvariablen Z mit Wahrscheinlichkeit $0,9$ fällt, so ist die Wahrscheinlichkeit, dass eine mit 4 Freiheitsgraden t-verteilte Zufallsvariable X in das Intervall $[-a; a]$ fällt, *kleiner* als $0,9$. Die letztgenannte Aussage geht auch direkt aus dem oberen Teil von Abbildung 12.6 hervor, wenn man dort $\alpha = 0,05$ wählt.

Man kann übrigens mit Tabelle 19.3 und Tabelle 19.5 die Intervalle angeben, in das Z resp. die mit 4 Freiheitsgraden t-verteilte Zufallsvariable X mit Wahrscheinlichkeit $0,9$ fallen:

$$P(Z \in [-z_{0,95}; z_{0,95}]) = P(Z \in [-1,6449; 1,6449]) = 0,9;$$
$$P(X \in [-t_{4;0,95}; t_{4;0,95}]) = P(X \in [-2,132; 2,132]) = 0,9.$$

Auch hieraus erschließt sich, dass Aussage C nicht zutrifft.

Zu D: Sowohl die Dichtekurve einer standardnormalverteilten Zufallsvariablen wie auch die Dichtekurve jeder t-verteilten Zufallsvariablen ist symmetrisch bezüglich des Erwartungswertes. Letzterer hat – bei der t-Verteilung unabhängig von der Anzahl der Freiheitsgrade – den Wert 0 (vgl. erneut Abbildung 12.6 in Kurs 33209). Hieraus folgt in Verbindung mit der genannten Symmetrieeigenschaft, dass die Aussage zutrifft.

Zu E: Aus Tabelle 19.6 kann man ablesen, dass das $0,95$-Quantil $F_{10;15;0,95}$ einer mit $m = 10$ und $n = 15$ Freiheitsgraden F-verteilten Zufallsvariablen X den Wert $2,54$ hat, d. h. es gilt $P(X \leq 2,54) = 0,95$ und $P(X > 2,54) = 0,05$ (vgl. auch in Kurs 33209 den linken Teil von Abbildung 12.7). Die Wahrscheinlichkeit $P(X > 2,7)$ dafür, dass X einen Wert oberhalb von $2,7$ annimmt, ist dann kleiner als $0,05$.

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

Bei einem statistischen Experiment mit n unabhängigen Wiederholungen wird jedesmal die Ausprägung einer Variablen X festgestellt (z. B. die Augenzahl beim n-fachen Wurf eines Würfels). Man will den Erwartungswert $\mu = E(X)$ und die Varianz $\sigma^2 = V(X)$ von X unter Heranziehung der beobachteten Werte x_1, x_2, ... , x_n schätzen. Letztere lassen sich als Realisationen unabhängiger Zufallsvariablen X_1, X_2, ... ,X_n interpretieren (auch Stichprobenvariablen genannt). Aus den n Stichprobenvariablen lässt sich der Stichprobenmittelwert $\overline{X}$ bilden.

Welche der folgenden Aussagen sind richtig? Bei Aussage E geht es um die Beurteilung des Wahrheitsgehalts des zweiten Satzes. (x aus 5)

A) Der Stichprobenmittelwert $\overline{X}$ repräsentiert eine unverzerrte Schätzung für den Erwartungswert μ.

B) Falls die obige Aussage A zutrifft, gilt auch, dass der mittlere quadratische Fehler des Stichprobenmittelwerts $\overline{X}$ und die Varianz von $\overline{X}$ übereinstimmen.

C) Wenn man die Summe der quadrierten Abweichungen $(X_1 - \overline{X})^2$, $(X_2 - \overline{X})^2$, ... , $(X_n - \overline{X})^2$ bildet und diese durch n dividiert, hat man eine unverzerrte Schätzung für die Varianz σ^2 des Merkmals X.

D) Die Varianz von $\overline{X}$ geht auf ein Viertel des Ausgangswertes zurück, wenn man n verdoppelt.

E) Man kann den Erwartungswert μ auch durch Angabe eines Konfidenzintervalls schätzen. Letzteres ist ein Intervall, das stets so groß gewählt wird, dass es den unbekannten Parameter μ enthält.

Lösung: A, B.

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

Zu B: Die Übereinstimmung von Varianz und mittlerem quadratischen Fehler ist bei einer Schätzfunktion immer dann gegeben, wenn die Schätzfunktion unverzerrt ist.

Zu C: Eine 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: Aus (14.7) in Kurs 33209 geht hervor, dass eine Verdoppelung von n zur Folge hat, dass die Varianz von $\overline{X}$ auf die Hälfte des Ausgangswertes zurückgeht.

Zu E: Ein Konfidenzintervall enthält den unbekannten Parameter μ nicht immer – vgl. z. B. in Kurs 33209 die Abbildungen 14.2 - 14.3, bei denen die dunkel markierten Konfidenzintervalle Gegenbeispiele darstellen.

Aufgabe 14 (Testen, Fehler beim Testen) (5 Punkte)

Es seien n Beobachtungen für ein Merkmal gegeben. Die Werte werden als Realisationen unabhängig identisch normalverteilter Stichprobenvariablen $X_1, ..., X_n$ aufgefasst (Normalverteilung mit unbekanntem Erwartungswert μ und Varianz σ^2). Getestet werden soll

$$H_0: \ \mu \geq \mu_0 \qquad \text{gegen} \qquad H_1: \ \mu < \mu_0$$

und zwar zum Signifikanzniveau $\alpha = 0,05$. Wenn man die Varianz σ^2 als bekannt voraussetzt, kann man den standardisierten Stichprobenmittelwert $Z = (\overline{X} - \mu_0)/\sigma_{\overline{X}}$ als Prüfgröße für den Test heranziehen (Gauß-Test).

Welche der folgenden Aussagen sind richtig? (x aus 5)

A) Die Nullhypothese wird bei obigem Test verworfen, wenn die Prüfgröße den Wert des $0,95$-Quantils der Standardnormalverteilung unterschreitet.

B) Die Wahrscheinlichkeit dafür, die Nullhypothese zu verwerfen, wenn $\mu = \mu_0$ gilt, beträgt hier $0,05$.

C) Die Wahrscheinlichkeit dafür, die Nullhypothese zu verwerfen, wenn $\mu > \mu_0$ gilt, liegt unterhalb von $0,05$.

D) Die Wahrscheinlichkeit für den Eintritt eines Fehlers 2. Art wird immer kleiner, je stärker μ den Wert μ_0 unterschreitet.

E) Die Wahrscheinlichkeit dafür, die Nullhypothese zu verwerfen, wenn $\mu = \mu_0$ gilt, ist identisch mit dem Wert der Gütefunktion des Tests an der Stelle $\mu = \mu_0$.

Lösung: B, C, D, E.

Zu A: Die Ablehnung der Nullhypothese erfolgt bei dem hier zugrunde gelegten linksseitigen Test, wenn die Prüfgröße den Wert $z_{0,05} = -z_{0,95}$ unterschreitet (nach Tabelle 19.3 in Kurs 33209 also bei Unterschreitung von $-1,6449$). Dies verifiziert man anhand von (15.7) in Kurs 33209.

Zu B: Im Falle $\mu = \mu_0$, also wenn die Nullhypothese $H_0: \ \mu \geq \mu_0$ des betrachteten linksseitigen Tests gerade noch gilt, wird die vom Testdesign vorgegebene Fehlerwahrscheinlichkeit $\alpha = 0,05$ erreicht (Fehler 1. Art).

Zu C: Wenn $\mu > \mu_0$ gilt, gilt zwar ebenfalls immer noch die Nullhypothese $H_0: \ \mu \geq \mu_0$, die Wahrscheinlichkeit für den Fehler 1. Art ist nun aber kleiner als $0,05$. Der Wert $\alpha = 0,05$ repräsentiert eine obere Schranke für die Wahrscheinlichkeit eines Fehlers 1. Art, die nur im Falle $\mu = \mu_0$ erreicht wird.

Zu D: Ein Fehler 2. Art, also die fälschliche Nicht-Verwerfung einer Nullhypothese, kann hier nur eintreten, wenn H_0 nicht zutrifft, also nur im Falle $\mu < \mu_0$. Je weiter μ sich nach unten von μ_0 entfernt, desto kleiner wird die Wahrscheinlichkeit für den Eintritt eines Fehlers 2. Art (vgl. hierzu Abbildung 15.3 in Kurs 33209, die sich allerdings auf einen rechtsseitigen Test bezieht).

Zu E: vgl. hierzu in Kurs 33209 die Lösung zu Aufgaben 15.2, Teil c.

Aufgabe 15 (Korrelationsmessung, Regressions- und Varianzanalyse) (5 P.)

Welche der folgenden Aussagen sind richtig? Bei den Aussagen A und B geht es darum, den Wahrheitsgehalt des jeweils letzten Satzes zu bewerten. (x aus 5)

A) Für einen acht Beobachtungspaare $(x_1; y_1)$, $(x_2; y_2)$, ... ,$(x_8; y_8)$ umfassenden Datensatz wurde $\overline{x} = 3$ und $\overline{y} = 7$ errechnet sowie

$$\sum_{i=1}^{8}(x_i - \overline{x})^2 = 4,1; \qquad \sum_{i=1}^{8}(y_i - \overline{y})^2 = 12; \qquad \sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y}) = 2,3.$$

Wenn man unterstellt, dass zwischen x_i und y_i $(i = 1, 2, ..., 8)$ ein linearer Zusammenhang besteht, kann man diesen durch das Regressionsmodell $y_i = \alpha + \beta x_i + u_i$ beschreiben und die Regressionskoeffizienten nach der Methode der kleinsten Quadrate schätzen. Die resultierende Regressionsgerade $\widehat{y} = \widehat{\alpha} + \widehat{\beta}x$ schneidet die y-Achse innerhalb des Intervalls $[5,0; 5,1]$.

B) Die Güte der Anpassung der mit der Kleinst-Quadrat-Methode erhaltenen Regressionsgeraden an den Datensatz lässt sich anhand des Bestimmtheitsmaßes bewerten. Für dieses errechnet man mit den Angaben aus Aufgabenteil A einen Wert, der im Intervall $[0,10; 0,12]$ liegt.

C) Der Korrelationskoeffizient für den Datensatz aus Aufgabenteil A ist positiv.

D) Bei Anwendung der Methode der kleinsten Quadrate wird die Regressionsgerade so bestimmt, dass die Summe der quadrierten Residuen Null ist.

E) Bei der einfaktoriellen Varianzanalyse wird die abhängige Variable als diskret modelliert.

Lösung: B, C.

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

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

und damit $\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} \approx 7 - 0,561 \cdot 3 = 5,317$. Die Regressionsgerade schneidet die y-Achse also etwa in $5,32$.

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

$$r = \frac{2,3}{\sqrt{4,1} \cdot \sqrt{12}} \approx \frac{2,3}{7,01} \approx 0,33.$$

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 (Responsevariable).

Numerische Aufgaben

Aufgabe 41 (Kombinatorik) (3 Punkte)

Mit einer „fairen“ Münze, also einer Münze mit gleichen Eintrittswahrscheinlichkeiten für „Zahl“ und „Kopf“, wird 8-mal nacheinander geworfen. Wie groß ist die Wahrscheinlichkeit mindestens 5-mal „Zahl“ zu erhalten? Tragen Sie Ihr Ergebnis, also ein Wert aus dem Intervall $[0; 1]$, auf *vier Stellen nach dem Dezimalkomma* genau rechtsbündig in das Antwortfeld ein. Das **Dezimalkomma** belegt ein **eigenes Feld**. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch)

Lösung : $0,3633$.

Bezeichnet man die Anzahl der Ausgänge mit „Zahl“ beim 8-maligem Münzwurf mit X, so ist die Wahrscheinlichkeit $P(X \geq 5)$ dafür, mindestens 5-mal „Zahl“ beim 10-maligen Münzwurf zu erhalten, die Komplementärwahrscheinlichkeit von $P(X \leq 4)$. Offenbar ist $P(X \leq 4)$ durch den Wert der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 8$ und $p = 0,5$ an der Stelle 4 gegeben, nach Tabelle 19.1 also durch $F(4) = 0,6367$. Es gilt also $P(X \geq 5) = 1 - 0,6367 \approx 0,3633$.

Anmerkung:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,360; 0,367]$ als richtig anerkannt.

Aufgabe 42 (Kombinatorik) (3 Punkte)

Den Mitarbeitern einer Firma wird als Intranet-Paßwort eine nur aus Großbuchstaben bestehende Buchstabenfolge zugewiesen, wobei Buchstaben auch mehrfach auftreten dürfen. Wieviele Mitarbeiter könnte man maximal anhand solcher Buchstabenfolgen unterscheiden, wenn für jeden Mitarbeiter genau 4 Großbuchstaben aus der Teilmenge $\{A, B, C, D, E, F\}$ des Alphabets verwendet werden, also z. B. $BAFB$, $AFBB$ oder $AECD$?

Tragen Sie die von Ihnen errechnete Anzahl, also ein ganzzahliges Ergebnis, rechtsbündig in das Antwortfeld ein. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch)

Lösung: $6^4 = 1296$.

Herleitung:

Die gesuchte Anzahl ist durch (10.8) in Kurs 33209 gegeben mit $N = 6$ und $n = 4$ (Ziehen von 4 Elementen mit Zurücklegen und mit Berücksichtigung der Anordnung aus einer Grundgesamtheit mit 6 Elementen; s. auch Tabelle 10.1 in Kurs 33209).

Aufgabe 43 (Lotto in Schweden) (3 Punkte)

Beim deutschen Lotto „6 aus 49" werden 6 Kugeln aus einer Trommel gezogen, die 49 fortlaufend nummerierte Kugeln enthält (Ziehen ohne Zurücklegen). In Schweden wird „7 aus 35" gespielt, also nur 35 Kugeln verwendet, von denen dann 7 gezogen werden. Bezeichne X die Anzahl der Richtigen bei der schwedischen Lottovariante. Berechnen Sie den Erwartungswert $\mu = E(X)$ der Zufallsvariablen X.

Tragen Sie Ihr Ergebnis rechtsbündig und auf *drei* Nachkommastellen genau in das Antwortfeld ein. Das **Dezimalkomma** belegt ein **eigenes Feld**. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch) $\mu =$ | | | | | | |

Lösung: $1,4$ (bzw. $1,400$)

Herleitung:

Der Erwartungswert bestimmt sich nach Formel (11.24) des Kurses 33209, wobei dort $n = 7$ sowie $M = 7$ und $N = 35$ einzusetzen ist:

$$\mu = 7 \cdot \frac{7}{35} = \frac{7}{5} = 1,4.$$

Da man den Erwartungswert auch wesentlich umständlicher – nämlich unter Heranziehung der Darstellungen (11.6) und (11.26) – berechnen kann und dabei Rundungen ins Spiel kommen, wurde bei der maschinellen Auswertung anstelle von $1,4$ auch jeder Wert aus dem Intervall $[1,39; 1,41]$ als richtig anerkannt.

Aufgabe 44 (Gewinnwahrscheinlichkeit beim „Glücksrad") (4 Punkte)

Bei einem Stadtteilfest ist eine als „Glücksrad" angesprochene drehbare Scheibe installiert, die in fünf gleich große Teile mit den Farben Schwarz, Rot, Weiß, Gelb und Grün unterteilt ist. Jeder Besucher des Festes darf das Rad 3x nacheinander drehen und bei jedem Drehversuch wird die Farbe festgestellt, die am Ende oben steht. Wenn bei allen drei Versuchen „Grün" resultiert, gibt es ein Freiabonnement für das Stadttheater für die nächste Saison.

Wie groß ist die Wahrscheinlichkeit, dass ein an diesem Spiel teilnehmender Festbesucher nach seinen drei Versuchen ein Freiabonnement erhält? Tragen Sie Ihr Ergebnis, also ein Wert aus dem Intervall $[0; 1]$, rechtsbündig und auf *vier* Nachkommastellen genau in das Antwortfeld ein. Das **Dezimalkomma** belegt auch hier wieder ein **eigenes Feld**. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch)

Lösung : $0,008$ (bzw. $0,0080$).

Herleitung:

Das Drehen des Glücksrades entspricht einem Bernoulli-Experiment (mögliche Ausgänge: eine bestimmte Farbe tritt auf / tritt nicht auf). Die Anzahl X des Auftretens von „Grün" ist binomialverteilt mit $p = 0,2$ und $n = 3$, weil es fünf Farben gibt (jede mit Eintrittswahrscheinlichkeit $p = 0,2$) und die Bernoulli-Kette $n = 3$ Experimente umfasst.

Die Wahrscheinlichkeit $P(X = 3)$ *genau dreimal* die Farbe „Grün" zu erreichen, errechnet sich als Differenz der Wahrscheinlichkeit $P(X \leq 3)$ *höchstens dreimal* die Farbe „Grün" zu erhalten und der Wahrscheinlichkeit $P(X \leq 2)$ dafür, *höchstens zweimal* die Farbe „Grün" zu erzielen. Beide Werte lassen sich aus Tabelle 19.1 aus Kurs 33209 als Werte $F(3)$ resp. $F(2)$ der Verteilungsfunktion genannten Binomialverteilung ablesen. Man erhält $F(3) - F(2) = 1 - 0,992 = 0,008$.

Anmerkung:

Der Wert $F(3) = 1$ versteht sich natürlich von selbst – bei dreimaligem Drehen des Glücksrades kann natürlich nur maximal dreimal „Grün" erzielt werden. Auch den Wert $F(2) = 0,992$ kann man sich, wenn auch etwas umständlicher, ohne Heranziehung von Tabelle 19.1 ableiten. „Höchstens zweimal Grün" (bzw. $X \leq 2$) umfasst ja alle Ausgänge außer „dreimal Grün" (d. h. außer $X = 3$). Da es bei 5 Farben und 3 Versuchen insgesamt $5^3 = 125$ Ausgänge gibt und nur einer dieser Ausgänge „dreimal Grün" ist, lassen sich 124 von 125 Ausgängen unter „Höchstens zweimal Grün" einordnen. Nach (10.5) erhält man dann für $P(X \leq 2) = \frac{124}{125} = 0,992$.

Bei der maschinellen Auswertung wurde anstelle von $0,0080$ auch jeder Wert aus dem Intervall $[0,0079; 0,0081]$ als richtig anerkannt.

Aufgabe 45 (Verteilungsfunktion einer Normalverteilung) (2 Punkte)

Eine Zufallsvariable X sei normalverteilt mit Erwartungswert $\mu = 5$ und Varianz $\sigma^2 = 1$. Die Verteilung von X kann man sowohl durch die Dichtefunktion $f(x)$ als auch durch die Verteilungsfunktion $F(x)$ charakterisieren. Welchen Wert nimmt $F(x)$ an der Stelle $x = 5$ an?

Geben Sie das Ergebnis auf *vier Stellen nach dem Dezimalkomma* genau an. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**.

(numerisch) $F(5) =$ | | | | | | | |

Lösung: $0,5$ (bzw. $0,5000$)

Herleitung: Für jede Normalverteilung gilt, dass ihre Dichtefunktion in $x = \mu$ das Maximum annimmt. Links und rechts vom Punkt $x = \mu$ hat der Flächeninhalt unter der Dichte jeweils den Inhalt $0,5$. Da jeder Wert $F(x)$ der Verteilungsfunktion durch den Flächeninhalt unter der Dichte bis zum Punkt x zu interpretieren ist, hat die Verteilungsfunktion an der Stelle $x = \mu$ (hier: $\mu = 5$) den Wert $0,5$ – vgl. auch Abbildung 12.2 in Kurs 33209.

Aufgabe 46 (Wahrscheinlichkeiten bei Normalverteilung) (3 Punkte)

Es sei erneut eine Zufallsvariable X betrachtet, die normalverteilt ist mit Erwartungswert $\mu = 5$ und Varianz $\sigma^2 = 1$. Bestimmen Sie einen Wert a, für den $P(X \leq a) = 0,1$ gilt, also einen Wert a, der mit Wahrscheinlichkeit $0,1$ nicht überschritten wird.

Geben Sie das Ergebnis auf *vier Stellen nach dem Dezimalkomma* genau an. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**. Falls Sie also z. B. 0,2471 errechnen, tragen Sie in die letzten sechs Felder $0,2471$ ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) $a =$ | | | | | | | |

Lösung: $3,7184$

Herleitung: Offenbar ist das $0,1$-Quantil $x_{0,1}$ der Normalverteilung mit Erwartungswert $\mu = 5$ und Varianz $\sigma^2 = 1$ gesucht. Nach Formel (12.26) in Kurs 33209 gilt

$$x_{0,1} = 5 + z_{0,1} \cdot 1 = 5 - 1,2816 \cdot 1 = 5 - 1,2816 = 3,7184.$$

Die Wahrscheinlichkeit dafür, dass eine Realisation von X den Wert $3,7184$ nicht überschreitet, beträgt also $0,1$.

Anmerkungen:

Man kann den gesuchten Wert a auch – umständlicher und mit geringerer Genauigkeit – anhand der Tabelle mit den Werten $\Phi(z)$ der Verteilungsfunktion der Standardnormalverteilung bestimmen (obere Tabelle auf S. 41 der Formelsammlung). Man kann nämlich aus der Tabelle zumindest näherungsweise einen Wert z bestimmen, für den $\Phi(z) = 0,9$ und damit nach (12.20) $\Phi(-z) = 0,1$ gilt. Aus der Tabelle entnimmt man, dass $\Phi(1,28) = 0,8997$ ist und $\Phi(1,29) = 0,9015$, also $\Phi(-1,28) = 0,1003$ und $\Phi(-1,29) = 0,0985$. Man würde bei Interpolation $\Phi(-1,281) \approx 0,1$ erhalten. Da mit $P(X \leq a) = 0,1$ und $X \sim N(5;1^2)$ ja nach Formel (12.21) $\Phi(\frac{a-5}{1}) = \Phi(a-5) = 0,1$ gilt, muß dann $a - 5 = -1,281$ gelten, also $a = 3,719$. Würde man nicht interpolieren und von $z = -1,28$ oder von $z = -1,29$ ausgehen, resultierte $a = 3,72$ resp. $a = 3,71$.

Anstelle des Werts $3,7184$, der sich beim zuerst skizzierten Lösungsgang ergibt, wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[3,710; 3,725]$ als richtig anerkannt.

Einige wenige Klausurteilnehmer sind von der Angabe $\Phi(3,9) = 0,1$ ausgegangen, die in der letzten Zeile der Tabelle für $\Phi(z)$ zu finden ist, und leiteten hieraus für das $0,1$-Quantil der $N(5;0,1^2)$-Verteilung den Wert $8,9$ ab. Dass dieser Wert nicht zutreffen kann, folgt schon daraus, das $\mu = 5$ das $0,5$-Quantil der Verteilung darstellt und das $0,1$-Quantil somit auf jeden Fall kleiner als 5 sein muss. Für die unzutreffende Lösung $8,9$ wurde hier dennoch aus formalen Gründen die volle Punktzahl vergeben, weil der Angabe $\Phi(3,9) = 0,1$ in der Tabelle für $\Phi(z)$ ein – allerdings klar erkennbarer – Druckfehler zugrunde liegt; richtig ist natürlich $\Phi(3,9) \approx 1,0$. Aus $\Phi(3,9) = 0,1$ wurde geschlossen, dass $a - 5 = 3,9$ sein muss und damit $a = 8,9$, d. h. $P(X \leq 8,9) = 0,1$.

Aufgabe 47 (lineares Regressionsmodell / KQ-Schätzung) (4 Punkte)

Für fünf Besucher eines Restaurants wurden der am Ende zu zahlende Rechnungsbetrag X und das gezahlte Trinkgeld Y erfasst. Dies führte zu folgenden Beobachtungsdaten $(x_1; y_1)$, $(x_2; y_2)$, ... ,$(x_5; y_5)$ für die beiden Merkmale X und Y:

$$(12; 1); \quad (15; 0); \quad (16; 1); \quad (19; 1); \quad (23; 2).$$

Wenn man davon ausgeht, dass zwischen dem Rechnungsbetrag und dem gezahlten Trinkgeld ein linearer Zusammenhang $y = \alpha + \beta x$ besteht, kann man auf der Basis des obigen kleinen Datensatzes unter Heranziehung der KQ-Schätzmethode eine Regressionsgerade $\widehat{y} = \widehat{\alpha} + \widehat{\beta} x$ bestimmen. Ermitteln Sie die Steigung $\widehat{\beta}$ dieser Regressionsgeraden und tragen Sie Ihr Ergebnis auf *vier* Stellen nach dem Dezimalkomma genau rechtsbündig in das Antwortfeld ein. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) $\widehat{\beta} =$ | | | | | | | |

Lösung: $0,1143$

Herleitung:

Aus den Daten errechnet man $\overline{x} = 17$ und $\overline{y} = 1$ und daraus unschwer

$$\sum_{i=1}^{5} (x_i - \overline{x})^2 = 70; \qquad \sum_{i=1}^{5} (x_i - \overline{x})(y_i - \overline{y}) = 8.$$

Es folgt dann nach Formel (16.6) aus Kurs 33209 $\widehat{\beta} = \frac{8}{70} = \frac{4}{35} \approx 0,1143$.

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von $0,1143$ auch jeder Wert aus dem Intervall $[0,113; 0,116]$ als richtig anerkannt.

Aufgabe 48 (Gauß-Test) (3 Punkte)

Es seien n Beobachtungen für ein Merkmal gegeben. Die Werte werden als Realisationen unabhängig identisch normalverteilter Stichprobenvariablen $X_1, ..., X_n$ aufgefasst (Normalverteilung mit Erwartungswert μ und Varianz σ^2). Getestet werden soll

$$H_0: \ \mu \leq \mu_0 \qquad \text{gegen} \qquad H_1: \ \mu > \mu_0$$

und zwar zum Signifikanzniveau $\alpha = 0,01$. Wenn man die Varianz σ^2 als bekannt voraussetzt, kann man den standardisierten Stichprobenmittelwert $Z = (\overline{X} - \mu_0)/\sigma_{\overline{X}}$ als Prüfgröße für den Test heranziehen (Gauß-Test). Man wird die Nullhypothese verwerfen, wenn der aus den Daten zu errechnende Wert der Prüfgröße Z einen bestimmten kritischen Wert überschreitet. Bestimmen Sie diesen kritischen Wert.

Tragen Sie Ihr Ergebnis auf *vier* Stellen nach dem Dezimalkomma genau rechtsbündig in das Antwortfeld ein. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch)

Lösung: $2,3263$

Herleitung:

Die Ablehnung der Nullhypothese erfolgt nach Formel (15.6) aus Kurs 33209, wenn die Prüfgröße den Wert $z_{0,99}$ überschreitet. Letzterer hat nach Tabelle 19.3 in Kurs 33209 den Wert 2,3263.

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von $2,3263$ auch jeder Wert aus dem Intervall $[2,30; 2,35]$ als richtig anerkannt.