FernUniversität in Hagen

LOTSE

Vorläufige Musterlösungen zur Klausur zum

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

und zum

Modul 3 im BA-Studiengang „Soziologie“

Termin: 6. März 2012, 14.00 - 18.00 Uhr

Prüfer: apl. Prof. Dr. H.-J. Mittag

Stand: 27. 3. 2012

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: $a \vee \neg b$	P2: $\neg a$	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, die sich aus a und b ableiten. 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 $P1$ und $P2$ wahr sind, ist auch K wahr, d. h. die Konklusion ist korrekt.

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

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

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

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, C, D.

Kommentar:

In den vier Tabellenzeilen sind die drei Werte in den letzten drei Spalten wie folgt zu ergänzen: Zeilen 1 und 2: w, f, f; Zeile 3: f, w, f; Zeile 4: w, w, w. Zu dieser Aufgabe vgl. auch Kurs 33210; Kapitel 5.

Aufgabe 2 (Merkmalsklassifikationen) (5 Punkte)

Welche der folgenden Aussagen sind richtig? Der Begriff „metrische Skala“ ist als Oberbegriff für „Intervallskala“ und „Verhältnisskala“ (letztere einschließlich des Sonderfalls „Absolutskala“) zu verstehen. (x aus 5)

A) Bei einer Nominal- und auch bei einer Ordinalskala ist die Bildung von Differenzen nicht zulässig.

B) Operationen, die für ordinalskalierte Daten zulässig sind, sind ebenso für metrisch skalierte Daten zulässig.

C) Das Merkmal „Art der Heizung“ (mit den Ausprägungen „Gas“, „Elektro“, „Öl“ und „Solar“) ist ein ordinalskaliertes Merkmal.

D) Das Merkmal „Art der Heizung“ (mit den o. g. Ausprägungen) ist ein diskretes Merkmal.

E) Die bei einem Volksmarathonlauf erreichten Plätze sind Ausprägungen eines metrisch skalierten Merkmals.

Lösung: A, B, D

Zu A und B: vgl. Kromrey, Abschnitt 5.4.4, oder Schnell / Hill /Esser, Abschnitt 4.3.1, oder Kurs 33209, Tabelle 2.1.

Zu C: Das Merkmal „Art der Heizung“ ist nominalskaliert – es gibt hier keine natürliche Rangordnung.

Zu D: Vgl. Kurs 33209, Ende von Abschnitt 2.2.

Zu E: Die Platzierungen sind Ausprägungen eines ordinalskalierten Merkmals.

Aufgabe 3 (Messung, Informationsgewinnung) (5 Punkte)

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

A) In der empirischen Sozialforschung versteht man unter „Operationalisierung" eine Handlungsanweisung, die es ermöglicht, nicht direkt beobachtbare theoretische Konstrukte einer Messung zugänglich zu machen.

B) Undercoverage ist ein Selektionsfehler, der bei stichprobenbasierten Datenerhebungen auftreten kann. Er entsteht, wenn nicht alle Elemente der Population, aus der eine Stichprobe gezogen werden soll, tatsächlich bei der Stichprobenziehung berücksichtigt werden.

C) Ein Gütekriterium für Messungen ist die Validität. Diese charakterisiert, inwieweit wirklich das gemessen wird, was gemessen werden soll.

D) Aus der Reliabilität einer Messung folgt stets auch deren Validität.

E) Das Random-Route-Verfahren ist ein bei der Befragung von Haushalten durch Interviewer angewendetes mehrstufiges Auswahlverfahren. Bei diesem wird der Startpunkt des Laufwegs des Interviewers zufällig ausgewählt, alle weiteren Verfahrensschritte sind dem Interviewer aber vorgegeben.

Lösung: A, B, C, E.

zu A: vgl. Kromrey, Abschnitt 4.3, oder Schnell / Hill / Esser, Abschnitt 4.2, oder Kurs 33209, Abschnitt 2.3.

zu B: vgl. Diekmann, Kapitel X, Abschnitte 9 und 11, oder Kurs 33209, Abschnitt 3.2.

zu C: vgl. Schnell / Hill / Esser, Abschnitt 4.3, oder Kurs 33209, Abschnitt 2.3.

Zu D: Aus der Validität einer Messung folgt deren Reliabilität – der Umkehrschluss gilt aber nicht (vgl. Kromrey, Ende von Abschnitt 5.7).

Zu E: vgl. Kromrey, Abschnitt 6.5.

Aufgabe 4 (Experimente, Stichprobenverfahren) (5 Punkte)

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

A) Ein Ansatz zur empirischen Überprüfung von Forschungshypothesen ist die Durchführung von Experimenten. Bei Experimenten mit Menschen, z. B. in der Medizin oder der Psychologie, werden i. a. zwei Gruppen von Teilnehmern gebildet (manchmal auch mehr als zwei Gruppen). In beiden genannten Gruppen werden dann Einflussgrößen planmäßig verändert.

B) Die Klumpenauswahl ist eine zweistufige Auswahlprozedur, bei der in der ersten Stufe Teilmengen der Grundgesamtheit (sog. Klumpen oder Cluster) zufällig ausgewählt werden.

C) Die Klumpenauswahl ist eine zweistufige Auswahlprozedur, bei der in der zweiten Verfahrensstufe aus den ausgewählten Teilmengen der Grundgesamtheit Zufallsstichproben gezogen 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. $n_3 = 66$ gezogen. Das damit praktizierte Auswahlverfahren repräsentiert eine geschichtete Stichprobenauswahl mit proportionaler Schichtung.

E) Bei interviewbasierten Datenerhebungen in der Markt- und Meinungsforschung wird oft das Quotenauswahlverfahren bei der Bestimmung der zu interviewenden Personen herangezogen. Bei diesem Verfahren wird dem Interviewer genau vorgegeben, welche Personen zu interviewen sind.

Lösung: B, D.

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

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

Zu C: In der zweiten Verfahrensstufe werden alle Elemente eines Klumpens herangezogen, nicht nur eine Zufallsauswahl der Klumpenelemente.

Zu D: Der in der zweiten Stufe angewandte Auswahlsatz $\frac{n_i}{N_i}$ $(i = 1, 2, 3)$ ist hier fest – er liegt einheitlich für alle Schichten bei 3 % (vgl. auch 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 (Univariate Häufigkeitsverteilungen / Kenngrößen) (5 Punkte)

Gegeben sei der folgende Datensatz für ein Merkmal X:

$4,9 \quad 6,2 \quad 4,2 \quad 4,9 \quad 3,9 \quad 4,9 \quad 4,2 \quad 6,0.$

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) Der Median $\widetilde{x}$ des obigen Datensatzes ist kleiner als dessen Mittelwert $\overline{x}$.

C) Wenn man bei obigem Datensatz den letzten Wert $(6,0)$ der Urliste um $0,4$ erhöht, hat dies zur Folge, dass sowohl der Mittelwert $\overline{x}$ als auch der Median $\widetilde{x}$ des Datensatzes größer werden.

D) Wenn man bei dem obigem Datensatz den ersten Wert $(4,9)$ streicht, bleibt der Median $\widetilde{x}$ unverändert, nicht aber der Mittelwert $\overline{x}$.

E) Mit der in Aufgabenteil C spezifizierten Veränderung des letzten Wertes der eingangs aufgeführten Urliste (Erhöhung dieses Wertes von $6,0$ auf $6,4$) verändert sich die Spannweite des Datensatzes.

Lösung: A, E.

Zu A: Der Datensatz hat den eindeutig bestimmten Modalwert $4,9$ - nur dieser Wert tritt 3-mal auf.

Zu B: Der Median des 8 Werte umfassenden Datensatzes ist $\widetilde{x} = 4,9$ (Mittel aus dem vierten und fünften Wert der nach aufsteigender Größe geordneten Urliste). Auch der Mittelwert $\overline{x}$ hat diesen Wert, d. h. es gilt $\overline{x} = 4,9$ und somit $\overline{x} = \widetilde{x}$.

Zu C: Wenn man den letzten Wert $(6,0)$ der Urliste um $0,4$ erhöht, vergrößert sich die Summe der Merkmalswerte und damit auch der Mittelwert. Er beträgt dann nicht mehr $\overline{x} = 4,90$, sondern $\overline{x} = 4,95$. Der Median bleibt hingegen unverändert.

Zu D: Nach Streichung des Wertes $4,9$ hat man einen Datensatz des Umfangs $n = 7$. Der Median ist dann der vierten Wert der geordneten Liste, also weiterhin $\widetilde{x} = 4,9$. Auch der Mittelwert bleibt unverändert, d. h. es ist auch $\overline{x} = 4,9$.

Zu E: Da sich einer der beiden Extremwerte des Datensatzes ($3,9$ und $6,2$), aus denen sich die Spannweite errechnet, mit Erhöhung des Wertes $6,0$ auf $6,4$ verändert (die Extremwerte sind dann durch $3,9$ und $6,4$ gegeben), erhöht sich die Spannweite um $0,2$.

Aufgabe 6 (Visualisierung empirischer Verteilungen) (5 Punkte)

Welche der folgenden Aussagen sind richtig? In Aufgabenteil A geht es um die Beurteilung des Wahrheitsgehalts des letzten Satzes. (x aus 5)

A) Die Lottozahlen, die in Deutschland im Jahr 2011 beim Lotto „6 aus 49“ gezogen wurden, definieren einen Datensatz, der Zahlen im Bereich von 1 bis 49 umfasst. Die durch den Datensatz definierte empirische Verteilung lässt sich anhand eines Stabdiagramms für absolute Häufigkeiten oder relative Häufigkeiten visualisieren. Beim Übergang von der einen zur anderen Darstellungsform ändert sich nur die Skalierung der Achse, an der die Häufigkeiten abgetragen werden (Ordinatenachse).

B) Die empirische Verteilung des Datensatzes „Gezogene Lottozahlen im Jahr 2011“ lässt sich ohne Informationsverlust auch anhand eines Boxplots veranschaulichen.

C) Bei dem Boxplot für den genannten Datensatz liegt der Median nicht notwendigerweise in der Mitte der Box.

D) Bei einem Boxplot sind Anfang und Ende der Box durch das untere resp. das obere Quartil des Datensatzes bestimmt.

E) Wenn man die Lottozahlen in 7-er Gruppen einteilt (Gruppen $1-7$, $8-14$, ... , $43-49$) und die Besetzungshäufigkeiten für die Gruppen visualisiert, resultiert ein Histogramm.

Lösung: A, C, D, E

Zu A: vgl. hierzu Abbildung 4.4 in Kurs 33209.

Zu B: Die Aggregation von $6 \cdot 52 = 312$ Werten der Urliste auf nur wenige Kenngrößen (5 Kenngrößen bei der einfachsten Boxplot-Variante) ist natürlich mit Informationsverlust verbunden.

Zu C: Der Median ist nur bei einer symmetrischen empirischen Verteilung in der Mitte der Box (vgl. auch Abbildung 5.4 in Kurs 33209).

Zu D: vgl. Abbildung 5.3 in Kurs 33209.

Zu E: vgl. auch Abbildung 4.5, die sich aber auf einen anderen Datensatz bezieht.

Aufgabe 7 (Konzentrationsmessung) (5 Punkte)

In der EU-27 teilen sich 5 Unternehmen, die hier mit U1, U2, ..., U5 bezeichnet seien, ein bestimmtes Marktsegment im Bereich der beruflichen Weiterbildung. Das Umsatzvolumen aller 5 Unternehmen im Geschäftsjahr 2011 belief sich auf insgesamt 40 Milliarden Euro. Davon entfallen auf U1 nur 10 %, auf U2 hingegen 30 %, während U3, U4 und U5 je 20 % des Gesamtumsatzes realisieren.

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 $0,28$ und $0,29$ liegt.

B) Der Wert, den man für den Ordinatenwert v_2 errechnet, gibt an, welcher Anteil des gesamten Aktienpakets auf diejenigen 2 Unternehmen entfällt, die die kleinsten Anteile am gesamten Umsatzvolumen in Höhe von 40 Milliarden Euro halten.

C) Allgemein gilt, dass der unnormierte Gini-Koeffizient G für einen Datensatz des Umfangs $n = 5$ den Wert $0,8$ nicht überschreiten kann.

D) Der normierte Gini-Koeffizient hat bei obigem Datensatz einen Wert G^*, der zwischen $0,19$ und $0,21$ liegt.

E) Wenn das Unternehmen U5 aus dem Markt ausschiede und der Umsatzanteil von U5 sich gleichmäßig auf die vier im Markt verbleibenden Unternehmen verteilte (Umsatzzuwachs von je 5 % für U1, ... , U4), würde der Wert des dann resultierenden normierten Gini-Koeffizienten G^* oberhalb von $0,21$ liegen.

Lösung: B, C, D.

Zu A: Die Umsatzvolumina der Unternehmen U1, ..., U5 betragen $x_1 = 4$ (10 % von 40), $x_2 = 12$ (30 % von 40) und $x_3 = x_4 = x_5 = 8$ (je 20 % von 40) - alle Werte in Milliarden Euro. Die geordnete Liste ist dann $x_{(1)} = 4$, $x_{(2)} = x_{(3)} = x_{(4)} = 8$, $x_{(5)} = 12$. Mit (6.2) – (6.3) aus Kurs 33209 folgt mit $p_5 = 40$ dann $v_2 = \frac{4+8}{40} = \frac{12}{40} = 0,3$.

Zu C: Nach (6.6) ist für den unnormierten Gini-Koeffizienten G im Falle $n = 5$ mit $G_{max} = \frac{5-1}{5} = 0,8$ eine obere Schranke gegeben.

Zu D: Die gewichtete Merkmalssumme q_5 aus Formel (6.4) in Kurs 33209 beträgt 136. Nach (6.5) errechnet sich dann für den unnormierten Gini-Koeffizienten der Wert

$$G = \frac{2 \cdot 136}{5 \cdot 40} - \frac{6}{5} = 0,16.$$

Für den normierten Gini-Koeffizienten G^* gilt somit

$$G^* = \frac{G}{G_{max}} = \frac{0,16}{0,8} = 0,2.$$

E: Die Umsatzvolumina der vier verbleibenden Unternehmen U1, ..., U4 betragen $x_1 = 6$ (15 % von 40), $x_2 = 14$ (35 % von 40) und $x_3 = x_4 = 10$ (je 25 % von 40) - alle Werte wieder in Milliarden Euro. Die geordnete Liste ist nun $x_{(1)} = 6$, $x_{(2)} = x_{(3)} = 10$, $x_{(4)} = 14$. Die gewichtete Merkmalssumme q_4 beträgt dann 112 und nach (6.5) errechnet sich für den unnormierten Gini-Koeffizienten der Wert

$$G = \frac{2 \cdot 112}{4 \cdot 40} - \frac{5}{4} = 0,15.$$

Ferner ist jetzt $G_{max} = \frac{3}{4} = 0,75$ und folglich

$$G^* = \frac{G}{G_{max}} = \frac{0,15}{0,75} = 0,2.$$

Der Wert G^* bleibt also trotz der geschrumpften Unternehmensanzahl unverändert.

Anmerkung:

Bei der Formulierung von Aufgabenteil B war ein offensichtlicher kleiner Schreibfehler unterlaufen. Es ging hier natürlich um einen Anteil am Gesamtumsatz, nicht um einen Anteil am Wert eines Aktienpakets. (Dass es sich um einen Schreibfehler handelte und Gesamtumsatz gemient war, wurde eigentlich schon durch den letzten Teil des Textes bei Aufgabenteil B deutlich - hier war vom Gesamtumsatz die Rede.) Dennoch wurde aufgrund des Schreibfehlers bei Aufgabenteil B auch Nicht-Ankreuzen als korrekte Lösung gewertet.

Aufgabe 8 (Zusammenhangsmessung) (5 Punkte)

Welche der folgenden Aussagen sind richtig? Die Aussagen A - D beziehen sich auf empirische Zusammenhangsmaße, sind also aus Beobachtungsdaten errechenbar. Aussage E bezieht sich hingegen auf ein theoretisches Zusammenhangsmaß, also auf ein Zusammenhangsmaß für Zufallsvariablen. In Aufgabenteil D geht es um die Beurteilung des Wahrheitsgehalts des letzten Satzes. (x aus 5)

A) Der Korrelationskoeffizient r nach Bravais-Pearson misst die Stärke eines linearen oder nicht-linearen Zusammenhangs zwischen zwei Merkmalen X und Y.

B) Wenn $r = 1$ ist, bedeutet dies, dass die Datenpaare $(x_1, y_1), ..., (x_n, y_n)$ alle auf einer steigenden Geraden liegen.

C) Der Rangkorrelationskoeffizient r_{SP} lässt sich auch auf nominalskalierte Merkmale anwenden.

D) Bei einer Studie wurden für alle Mitglieder einer Gruppe von Senioren jeweils zwei Merkmale X und Y erfasst. Das Merkmal X weise die Ausprägungen a_1 und a_2, das Merkmal Y die Ausprägungen b_1 und b_2 auf. Die beobachteten Häufigkeiten für die einzelnen Ausprägungen sind in der nachstehenden Tabelle zusammengefasst:

	b_1	b_2
a_1	20	25
a_2	40	15

Mit diesen Werten errechnet sich für den χ^2-Koeffizienten ein Wert, der unterhalb von $8,5$ liegt.

E) Die Kovarianz zweier Zufallsvariablen X und Y ist ein Zusammenhangsmaß, das nur Werte zwischen -1 und $+1$ annehmen kann.

Lösung: B, D.

Zu A: 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 B: Wenn alle Punkte auf einer steigenden Geraden liegen, gilt $r = 1$. Liegen Sie auf einer fallenden Geraden liegen, gilt $r = -1$ (vgl. auch Abbildung 9.2 im Kurs 33209).

Zu C: Wenn man die Vierfeldertafel um die Randverteilungen ergänzt hat, erhält man für den χ^2-Koeffizienten nach (9.8)

$$\chi^2 = \frac{100 \cdot (20 \cdot 15 - 40 \cdot 25)^2}{45 \cdot 55 \cdot 60 \cdot 40} = \frac{490000}{59400} \approx 8,249.$$

Zu dieser Rechnung vgl. auch die Lösungen zu den Aufgaben 8.1 und 9.1 in Kurs 33209.

Zu D: vgl. Kurs 33209, Gleichung (16.18).

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

Aufgabe 9 (Randverteilungen, bedingte Wahrscheinlichkeiten) (5 Punkte)

An einer Hochschule wurde für Studierende zweier Bachelorstudiengänge, die hier mit POL und SOZ abgekürzt seien, eine gemeinsame Methodenklausur angeboten. An der Klausur nahmen insgesamt 236 Studierende teil, darunter 184 aus dem Studiengang POL. Unter den 236 Klausurteilnehmern waren 120 Frauen. Von den weiblichen Klausurnehmern waren 24 im Studiengang SOZ eingeschrieben.

Welche der folgenden Aussagen sind richtig? Gehen Sie davon aus, dass kein Klausurteilnehmer in beiden Studiengängen gleichzeitig eingeschrieben ist. (x aus 5)

A) Weniger als ein Viertel aus der Gruppe aller männlichen Klausurteilnehmer war im Studiengang SOZ eingeschrieben.

B) Wählt man aus aus der Gesamtpopulation aller Klausurteilnehmer eine Person zufällig aus, so liegt die Wahrscheinlichkeit, dass diese SOZ studiert, über $0,23$.

C) Wählt man aus der Gesamtpopulation aller Klausurteilnehmer eine Person zufällig aus, so liegt die Wahrscheinlichkeit, dass eine im Studiengang POL eingeschriebene Frau ausgewählt wird, zwischen $0,39$ und $0,41$.

D) Wählt man aus der Teilpopulation der Studierenden mit Studienfach POL eine Person zufällig aus, so beträgt die Wahrscheinlichkeit dafür, dass die Zufallsauswahl auf einen Mann fällt, unter $0,46$.

E) Die Quote der weiblichen Klausurteilnehmer lag im Fach POL niedriger als beim Fach SOZ.

Lösung: A, C.

Es ist zweckmäßig die im Text enthaltenen Informationen entweder anhand eines Baumdiagramms zu visualisieren (vgl. in Kurs 33209 z. B. die Abbildung 8.1 oder die Lösung zu Teil a von Aufgabe 10.6) oder aber sie in einer Kontingenztafel mit Randverteilungen zusammenzufassen (vgl. in Kurs 33209 die Tabelle 10.2 oder die Lösung zu Teil e von Aufgabe 10.5). Man erhält im letztgenannten Fall – ähnlich wie beim SODUKO-Spiel – folgende Vierfeldertafel, bei der die Vorgaben dieser Aufgabe kursiv gesetzt sind (Codierung: POL $= A$, SOZ $= \overline{A}$, weiblich $= B$, männlich $= \overline{B}$):

	weiblich (B)	männlich ($\overline{B}$)	Zeilensummen
POL (A)	96	88	*184*
SOZ ($\overline{A}$)	*24*	28	52
Spaltensummen	*120*	116	*236*

Aus der Vierfeldertafel folgt dann:

Zu A: Von den an der Klausur teilnehmenden 116 Männern waren 28 für SOZ eingeschrieben, d. h. ca. $24,1$ % – somit weniger als ein Viertel der männlichen Teilnehmer.

Zu B: Die Wahrscheinlichkeit $\overline{A}$ dafür, dass von den 236 Klausurteilnehmern per Zufallsauswahl eine der 52 Personen gewählt wird, die SOZ studieren, beträgt nach der Formel (10.5) aus Kurs 33209 $P(\overline{A}) = \frac{52}{236} \approx 0,2203$, also ca. $22,0$ % .

Zu C: Die Wahrscheinlichkeit $P(A \cap B)$ dafür, dass per Zufallsauswahl aus der Population aller Klausurteilnehmer eine Frau gewählt wird, die POL studiert, beträgt $P(A \cap B) = \frac{96}{236} \approx 0,4068$, also etwa $40,7$ %.

Zu D: Die Wahrscheinlichkeit $P(\overline{B}|A)$ dafür, dass bei Zufallsauswahl einer Person aus der Teilpopulation der 184 Studierenden mit Studienfach *POL*, die Wahl auf einen Mann fällt, beträgt $\frac{88}{184} \approx 0,4783$, also ca. $47,8$ %.

Zu E: Die Quote der weiblichen Klausurteilnehmer betrug beim Fach POL $\frac{96}{184} \approx 0,5217$, also ca. $52,2$ %. Die entsprechende Quote für SOZ belief sich auf $\frac{24}{52} \approx 0,4625$, also auf ca. $46,2$ %.

Aufgabe 10 (Wahrscheinlichkeiten beim Roulettespiel) (5 Punkte)

Beim Roulettespiel ist der Ausgang durch eine diskrete Zufallsvariable X mit 37 Ausprägungen gegeben, nämlich den Ausprägungen $x_1 = 0, x_2 = 1, \ldots, x_{37} = 36$. Alle 37 Ausprägungen besitzen dieselbe Eintrittswahrscheinlichkeit.

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

A) Die Verteilungsfunktion $F(x)$ von X nimmt für $x = 0$ den Wert 0 an.

B) Für den Wert $F(4)$, den die Verteilungsfunktion $F(x)$ für $x = 4$ annimmt, gilt $F(4) > 0,12$.

C) Die Verteilungsfunktion $F(x)$ nimmt an der Stelle $x = 36$ den Wert 1 an.

D) Die Wahrscheinlichkeit dafür, dass der Ausgang beim einmaligen Roulettespiel eine zweistellige Zahl ist, liegt zwischen $0,72$ und $0,74$.

E) Die Wahrscheinlichkeit dafür, beim zweimaligen Roulettespiel in beiden Spielen die Zahl 10 zu realisieren, ist genau das Doppelte der Wahrscheinlichkeit dafür, beim einmaligen Roulettespiel eine 10 zu erzielen.

Lösung: B, C, D – vgl. zu dieser Aufgabe auch den Abschnitt 11.1 in Kurs 33209, insbesondere die Abbildung 11.1, die sich allerdings auf eine diskrete Gleichverteilung mit den Ausprägungen $x_1 = 1, x_2 = 2, \ldots, x_6 = 6$ und $p = \frac{1}{6}$ bezieht.

Kommentar:

Die Zufallsvariable X folgt einer diskreten Gleichverteilung mit $p = \frac{1}{37}$. Die Wahrscheinlichkeitsfunktion $f(x)$ ist durch (11.1) und die Verteilungsfunktion durch (11.3) definiert mit $p = \frac{1}{37}$ und $k = 37$. Wenn man die beiden Funktionen grafisch darstellt, resultiert eine Abbildung analog zu Abbildung 11.1. Die Balken im oberen Teil von Abbildung 11.1 – dort ist die Wahrscheinlichkeitsfunktion $f(x)$ dargestellt – setzen nun in den Punkten $x_1 = 0, x_2 = 1, \ldots, x_{37} = 36$ an und haben je die Länge $\frac{1}{37} \approx 0,0270$. Insbesondere gilt also auch $f(0) \approx 0,0270 > 0$.

Zu A: Die Verteilungsfunktion $F(x)$ hat in $x_1 = 0$ den Wert $F(0) = f(0) = \frac{1}{37} \approx 0,0270$.

Zu B: Die Funktion $F(x)$ weist dann jeweils Sprünge in $x_2 = 1, x_3 = 2, x_4 = 3, \ldots, x_{37} = 36$ auf. Die Sprunghöhe an den Sprungstellen beträgt einheitlich $\frac{1}{37} \approx 0,0270$

– s. auch den unteren Teil der analogen Abbildung 11.1. Es gilt also insbesondere $F(4) = f(0) + f(1) + ... + f(4) = \frac{5}{37} \approx 0,1351$, somit also insbesondere $F(4) > 0,12$.

Zu C: Die Treppenfunktion $F(x)$ erreicht an der letzten Sprungstelle $x_{37} = 36$ den Wert $F(36) = 1$ und bleibt dann auf dem Niveau 1.

Zu D: Von den 37 möglichen Ausgängen sind 27 zweistellig, nämlich die Ausgänge 10,..., 36. Mit dem Laplace-Ansatz (10.5) folgt dann $P(X \geq 10) = \frac{27}{37} \approx 0,7297$.

Zu E: Ist a_1 der Ausgang beim ersten Roullettespiel und a_2 der beim zweiten Spiel, so gibt es insgesamt $37^2 = 1369$ mögliche Ausgänge (a_1, a_2) bei zweimaligen Roulettespiel (vgl. auch Tabelle 10.1, dort links oben). Nur einer dieser 1369 Ausgänge, die alle dieselbe Eintrittswahrscheinlichkeit aufweisen, ist der Ausgang $(10; 10)$. Nach dem Laplace-Ansatz (10.5) ist die gesuchte Wahrscheinlichkeit also durch $\frac{1}{1369} \approx 0,00073$ gegeben, nicht aber durch $\frac{2}{37} \approx 0,0541$.

Aufgabe 11 (Zufallsvariablen) (5 Punkte)

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

A) Die Summe zweier unabhängiger und standardnormalverteilter Zufallsvariablen ist nicht standardnormalverteilt.

B) Das 0,9-Quantil einer mit 4 Freiheitsgraden t-verteilten Zufallsvariablen ist größer als das 0,9-Quantil einer standardnormalverteilten Zufallsvariablen.

C) Es sei X eine mit $m = 4$ und $n = 15$ Freiheitsgraden F-verteilte Zufallsvariable. Das 0,95-Quantil dieser F-Verteilung ist kleiner als 2,9.

D) Es bezeichne $F(x)$ die Verteilungsfunktion einer normalverteilten Zufallsvariablen. Die Wahrscheinlichkeit $P(X \leq 1)$ dafür, dass X eine Ausprägung annimmt, die nicht größer als 1 ist, ist dann durch $F(1)$ gegeben, also durch den Wert der Verteilungsfunktion an der Stelle $x = 1$.

E) Eine mit 8 Freiheitsgraden χ^2-verteilte Zufallsvariable nimmt mit Wahrscheinlichkeit 0,1 eine Ausprägung an, die oberhalb von 13,362 liegt.

Lösung: A, B, D, E.

Zu A: Für die Summe $X = Z_1 + Z_2$ zweier standardnormalverteilter Zufallsvariablen gilt nach (12.17) in Kurs 33209, dass sie zwar normalverteilt ist mit Erwartungswert $\mu = E(X) = 0$, aber mit Varianz $\sigma^2 = V(X) = 2$.

Zu B: Die Dichtekurven von t-verteilten und die einer standardnormalverteilten Zufallsvariablen sind symmetrisch bezüglich des für alle genannten Verteilungen geltenden Erwartungswertes $\mu = 0$. Die Dichtekurven von t-Verteilungen verlaufen allerdings flacher, wobei sich die Formunterschiede bei zunehmender Anzahl von Freiheitsgraden nivellieren (vgl. Abb. 12.6). Der flachere Verlauf bedingt, dass die $0,9$-Quantile der t-Verteilung größer sind als das $0,9$-Quantil $z_{0,9}$ der Standardnormalverteilung, insbesondere bei kleiner Anzahl von Freiheitsgraden. Man entnimmt den Tabelle 19.3 resp. der Tabelle 19.5, dass $z_{0,9} = 1,286$ und $t_{4;0,9} = 1,533$. (Der letztgenannte Wert entspricht ca. $1,192 \cdot z_{0,9}$, d. h. er übersteigt $z_{0,9}$ um etwa $19,2$ %).

Zu C: Nach Tabelle 19.6 in Kurs 33209 hat das $0,95$-Quantil einer mit $m = 4$ und $n = 15$ Freiheitsgraden F-verteilten Zufallsvariablen den Wert $3,06$.

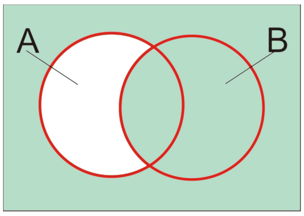
Zu D: Die Wahrscheinlichkeit $P(X \leq 1)$ dafür, dass X eine Ausprägung annimmt, die nicht größer als 1 ist, ist durch $F(1)$ gegeben – vgl. auch Abbildung 12.3 in Kurs 33209.

Zu E: Aus Tabelle 19.4 kann man ablesen, dass das $0,9$-Quantil $\chi^2_{8;0,9}$ einer mit 8 Freiheitsgraden χ^2-verteilten Zufallsvariablen X den Wert $13,362$ hat. Der Wert $\chi^2_{8;0,9} = 13,362$ hat nach der allgemeinen Definition (11.17) für Quantile die Eigenschaft, dass eine Ausprägung von X den Wert $13,362$ mit Wahrscheinlichkeit $0,9$ nicht überschreitet, mit Wahrscheinlichkeit $0,1$ also überschreitet.

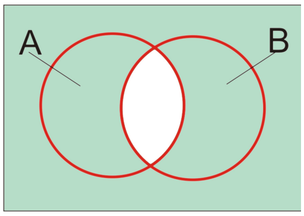
Aufgabe 12 (Venn-Diagramme / Kombinatorik / diskrete Verteilungen) (5 Punkte)

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

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



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



C) Mit einem fairen Würfel, also einem Würfel mit gleichen Eintrittswahrscheinlichkeiten für alle Augenzahlen, werde 8-mal in Folge gewürfelt. Bezeichne X die Anzahl der Ausgänge mit einer Augenzahl, die kleiner als 3 ist. Der Erwartungswert von X liegt zwischen $2,6$ und $2,7$.

D) Beim Würfeln mit zwei fairen Würfeln, liegt die Wahrscheinlichkeit dafür, eine Augensumme zu erhalten, die nicht größer als 10 ist, zwischen $0,85$ und $0,90$.

E) Auf einem Stadtteilfest wird ein Glücksrad mit 5 Feldern (Kreissektoren) gleicher Größe eingesetzt. Interessierte Besucher dürfen das Rad dreimal drehen (drei Spiele) und können bei jedem Spiel einen Preis erhalten. Ob ein Preis zugesprochen wird, hängt davon ab, welcher Sektor nach einem Spiel jeweils am Ende oben steht. Nur bei 2 der 5 Sektoren ist ein Preis zu bekommen. Die Wahrscheinlichkeit dafür, nach 3 Spielen keinen Gewinn zu erhalten, ist kleiner als $0,23$.

Lösung: A, C, D, E.

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

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

Zu C: Es ist der Erwartungswert $\mu = E(X)$ einer Binomialverteilung mit $n = 8$ und $p = \frac{1}{3}$ zu bestimmen. Der Parameter p bezeichnet die Wahrscheinlichkeit dafür, dass bei einem Wurf des Würfels eine Augenzahl erscheint, die kleiner als 3 ist. Für μ errechnet man nach Formel (11.20) aus Kurs 33209 den Wert $\mu = 8 \cdot \frac{1}{3} \approx 2,667$.

Zu D: 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 oberhalb von 10 liegt. Bei 33 Ausgängen ist die Augensumme also nicht größer als 10. Die Wahrscheinlichkeit P für die Erzielung einer Augensumme, die nicht größer als 10 ist, hat hier somit den Wert $P = \frac{33}{36} = \frac{11}{12} \approx 0,917$ (entspricht ca. $91,7$ %).

Zur vorstehenden Teilaufgabe vgl. auch die letzten Zeilen von Beispiel 10.1 in Kurs 33209. Die Lösung lässt sich zudem anhand eines interaktiven Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren).

Zu E: Die gesuchte Wahrscheinlichkeit ist hier durch den Wert $F(0)$ der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 3$ und $p = \frac{2}{5} = 0,4$ gegeben, nach Tabelle 19.1 also durch $0,2160$.

Aufgabe 13 (Schätzen von Modellparametern) (5 Punkte)

Gegeben seien n Stichprobenwerte, die als Ausprägungen x_1, x_2, ... ,x_n unabhängiger $N(\mu;\sigma^2)$-verteilter Zufallsvariablen X_1, X_2, ... ,X_n interpretiert werden. Aus den Zufallsvariablen X_1, X_2, ... ,X_n leiten sich der Stichprobenmittelwert $\overline{X}$ und die empirische Varianz S^2 ab – letztere definiert als Summe der quadrierten Mittelwertabweichungen $(X_i - \overline{X})^2$, dividiert durch n.

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

A) Wenn man den Erwartungswert μ anhand des Stichprobenmittelwerts $\overline{X}$ schätzt, hat man mit $\overline{X}$ eine Schätzfunktion, deren Varianz auf ein Viertel des ursprünglichen Werts zurückgeht, wenn man den Stichprobenumfang verdoppelt.

B) Die Stichprobenfunktion S^2 liefert eine verzerrte Schätzung für die Varianz σ^2.

C) Der mittlere quadratische Fehler (MSE) und die Varianz der Schätzfunktion $\overline{X}$ stimmen überein.

D) Man kann den Erwartungswert μ eines normalverteilten Merkmals auch durch Angabe eines Konfidenzintervalls schätzen. Ein solches Intervall ist so konstruiert, dass es μ einschließt.

E) Die Grenzen von Konfidenzintervallen sind zufallsabhängig.

Lösung: B, C, E.

Zu A: Die Varianz der Schätzfunktion $\overline{X}$ ist durch (14.7) gegeben, halbiert sich also bei Verdoppelung von n.

Zu B: Vgl. hierzu (14.8) in Kurs 33209 und den anschließenden Text.

Zu C: Da der Stichprobenmittelwert $\overline{X}$ nach (14.6) unverzerrt ist, ist der Verzerrungsterm in (14.5) Null (setze dort $\widehat{\Theta} := \overline{X}$).

Zu D: Die Abbildungen 14.2 und 14.3 zeigen Gegenbeispiele. Das Intervall ist so konstruiert, dass es den unbekannten Parameter mit einer hohen Wahrscheinlichkeit $1-\alpha$ einschließt (mit vorgegebener Nicht-Überdeckungswahrscheinlichkeit α).

Zu E: Die Grenzen von Konfidenzintervallen sind stochastisch, weil sie durch Zufallsvariablen definiert sind – vgl. auch Formel (14.14) und die Abbildungen 14.2 und 14.3 für das Beispiel des Gauß-Tests.

Aufgabe 14 (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 $\overline{y} = 5$ errechnet sowie

$$\sum_{i=1}^{8}(x_i - \overline{x})^2 = 4,5; \qquad \sum_{i=1}^{8}(y_i - \overline{y})^2 = 5,8; \qquad \sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y}) = 2,4.$$

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 $[3,3; 3,5]$.

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 $[0,23; 0,25]$ liegt.

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

D) Würde man für das Bestimmtheitsmaß den Wert $0,48$ errechnen, würde dies beinhalten, dass 48 % der Gesamtvariation des Datensatzes durch das verwendete Regressionsmodell erklärt ist.

E) Bei der einfaktoriellen Varianzanalyse werden sowohl die unabhängige als auch die abhängige Variable als diskret modelliert.

Lösung: A, C, D.

Zu A: Es gilt nach (16.6) und (16.7) bei Beachtung von $\overline{x} = 3$ und $\overline{y} = 5$

$$\widehat{\beta} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{8}(x_i - \overline{x})^2} = \frac{2,4}{4,5} \approx 0,533$$

und damit $\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} \approx 5 - 0,533 \cdot 3 = 3,401$. Die Regressionsgerade schneidet die y-Achse also etwa in $3,4$.

Zu B: Nach (16.18) gilt für das Bestimmtheitsmaß

$$R^2 = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = \frac{2,4^2}{4,5 \cdot 5,8} = \frac{5,76}{26,1} \approx 0,221.$$

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:

$$r = \frac{2,4}{\sqrt{4,5} \cdot \sqrt{5,8}} \approx \frac{2,4}{5,109} \approx 0,47.$$

Zu D: vgl. hierzu die Aufgabe 16.2 in Kurs 33209.

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).

Aufgabe 15 (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

$$H_0: \ \mu = \mu_0 \qquad \text{gegen} \qquad H_1: \ \mu \neq \mu_0$$

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) Wird bei obigem Test für die Prüfgröße ein Wert ermittelt, der den Wert des 0,95-Quantils der Standardnormalverteilung überschreitet, wird die Nullhypothese verworfen.

B) Wird bei obigem Test für die Prüfgröße ein Wert ermittelt, der kleiner als $-1,96$ ist, wird die Nullhypothese verworfen.

C) Die Wahrscheinlichkeit dafür, einen Fehler 1. Art zu begehen, beträgt bei obigem Test 0,05.

D) Wenn $\mu > \mu_0$ gilt, ist ein Fehler 2. Art möglich, nicht aber ein Fehler 1. Art.

E) Die Gütefunktion $G(\mu)$ des Tests hat an der Stelle $\mu = \mu_0$ den Wert 0,05.

Lösung: B, C, D, E.

Zu A und B: Die Ablehnung der Nullhypothese erfolgt bei dem hier zugrunde gelegten zweiseitigen Test, wenn die Realisation der Prüfgröße außerhalb des Intervalls $[z_{0,025}; z_{0,975}] = [-z_{0,975}; z_{0,975}]$ liegt, also außerhalb des Intervalls $[-1,96; 196]$ – vgl. auch Abbildung 15.1 in Kurs 33209 (mit $\alpha = 0,05$). Die Überschreitung des $0,95$-Quantils $z_{0,95} = 1,6449$ ist noch nicht hinreichend für die Verwerfung von H_0.

Zu C, D und E: vgl. Abbildung 15.5 im Kurs (mit $\alpha = 0,05$).

Numerische Aufgaben

Bei den folgenden Aufgaben sind die Antworten jeweils rechtsbündig mit der angegebenen Genauigkeit in die Lösungsfelder einzutragen. Bei negativen Ergebnissen ist das Vorzeichen (Minuszeichen) in das Feld vor der ersten Ziffer einzutragen.

Aufgabe 41 (Rangkorrelationskoeffizient) (3 Punkte)

Es sei angenommen, dass - in Konkurrenz zu den etablierten Ratingagenturen - zwei neue europäische Ratingagenturen A und B ihre Arbeit aufgenommen haben und u. a. das Ausfallrisiko für Staatsanleihen bewerten. Die beiden neuen Ratingagenturen beurteilen unabhängig voneinander das kurzfristige Ausfallrisiko von Staatsanleihen für vier Länder der Eurozone. Die Risikobewertung wird anhand einer 12-stufigen Ratingskala vorgenommen. Die Stufen seien hier mit 1, .. ,12 codiert, wobei die Punktzahl 12 die schlechteste Bewertung darstelle (höchstes Ausfallrisiko) und 1 die beste Bewertung. Die Ergebnisse der Bewertungen sind nachstehend ausgewiesen.

Land i	Agentur A Bewertung x_i	Agentur B Bewertung y_i
1	3	4
2	7	9
3	11	10
4	9	8

Untersuchen Sie anhand des Rangkorrelationskoeffizienten r_{SP} von Spearman, ob zwischen den Bewertungen der beiden Ratingagenturen ein Zusammenhang besteht. Tragen Sie Ihr Ergebnis auf *zwei* Stellen nach dem Dezimalkomma genau rechtsbündig in das Antwortfeld ein. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch) $r_{SP} =$ | | | | | | |

Lösung: $0,8$ (vgl. zu dieser Aufgabe auch Beispiel 9.4 in Kurs 33209)

Herleitung:

Wenn man den Beurteilungen der Agenturen A und 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:

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 Bewertung sowohl bei beiden Agenturen nicht mit der Mehr-

	Bank A		Bank B		
Unternehmen i	Bewertung x_i	$rg(x_i)$	Bewertung y_i	$rg(y_i)$	d_i
1	3	1	4	1	0
2	7	2	9	3	-1
3	11	4	10	4	0
4	9	3	8	2	1

fachbelegung eines Rangplatzes verbunden ist. Mit (9.16) resultiert

$$r_{SP} = 1 - \frac{6 \cdot [0^2 + (-1)^2 + 0^2 + 1^2]}{4 \cdot (16 - 1)} = 1 - 0,2 = 0,8.$$

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

Aufgabe 42 (mehrfacher Münzwurf) (3 Punkte)

Mit einer „fairen“ Münze, also einer Münze mit gleichen Eintrittswahrscheinlichkeiten für „Zahl“ und „Kopf“, wird 6-mal nacheinander geworfen. Wie groß ist die Wahrscheinlichkeit mindestens 4-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 : $0,3437$.

Bezeichnet man die Anzahl der Ausgänge mit „Zahl“ beim 6-maligem Münzwurf mit X, so ist die Wahrscheinlichkeit $P(X \geq 4)$ dafür, mindestens 4-mal „Zahl“ zu erhalten, die Komplementärwahrscheinlichkeit von $P(X \leq 3)$. Dabei ist $P(X \leq 3)$ durch den Wert der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 6$ und $p = 0,5$ an der Stelle 3 gegeben, nach Tabelle 19.1 also durch $F(3) = 0,6563$. Es gilt somit $P(X \geq 4) = 1 - 0,6563 \approx 0,3437$.

Der Wert $F(3) = P(X \leq 3) = 0,6563$ lässt sich auch online anhand eines interaktiven Experiments bestimmen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und beim Applet oben den Wert $n = 6$ und unten $x = 3$ und $p = 0,50$ einstellen).

Anmerkung:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,340; 0,347]$ als richtig anerkannt.

Aufgabe 43 (Kombinatorik) (3 Punkte)

Im Rahmen eines Betriebsjubiläums bei einem kleinen Unternehmen werden unter den 20 Mitarbeitern drei Preise unterschiedlicher Wertigkeit verlost (1. Preis: 4-tägige Reise nach Hamburg mit Übernachtungen in einem 4-Sterne-Hotel, Halbpension und Musicalbesuch; 2. Preis / 3. Preis: ein Hotelgutschein für zwei Übernachtungen / eine Übernachtung im gleichen 4-Sterne-Hotel). Die Verlosung ist so organisiert, dass in eine Schachtel 20 Lose gegeben werden, die mit $1, 2, .., 20$ nummeriert sind. Jeder Mitarbeiter ist durch genau eine der Nummern repräsentiert. Aus der Schachtel werden dann nacheinander 3 Lose gezogen – zuerst der Hauptgewinn, dann der 2. Preis und am Ende der 3. Preis. Um auszuschließen, dass jemand mehr als einen Preis gewinnt, wird eine gezogene Nummer vor dem Ziehen der nächsten Nummer nicht zurückgelegt.

Wieviele Möglichkeiten gibt es, die drei Preise innerhalb der 20 Personen umfassenden Belegschaft bei Anwendung dieses Losverfahrens zu verteilen? Tragen Sie Ihr (ganzzahliges) Ergebnis rechtsbündig in das Antwortfeld ein. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch)

Lösung: 6840.

Herleitung: In der Terminologie des Urnenmodells werden $n = 3$ Kugeln aus einer Urne mit $N = 20$ 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

$$\frac{20!}{(20-3)!} = \frac{20!}{17!} = 18 \cdot 19 \cdot 20 = 6840.$$

Anmerkung:

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

$$\binom{20}{3} = \frac{20!}{(17)! \cdot 3!} = 1140.$$

Für dieses Ergebnis wurde 1 Punkt gewährt.

Aufgabe 44 (Lotto in Litauen) (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 Litauen wird „6 aus 30“ gespielt, also nur 30 Kugeln verwendet, von denen wie in Deutschland 6 gezogen werden. Bezeichne X die Anzahl der Richtigen bei der litauischen 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: $1,2$ (bzw. $1,200$)

Herleitung:

Der Erwartungswert bestimmt sich nach Formel (11.24) des Kurses 33209, wobei dort $n = 6$ sowie $M = 6$ und $N = 30$ einzusetzen ist:

$$\mu = 6 \cdot \frac{6}{30} = \frac{6}{5} = 1,2.$$

Bei der maschinellen Auswertung wird anstelle von $1,2$ auch jeder Wert aus dem Intervall $[1,19; 1,21]$ als richtig anerkannt.

Aufgabe 45 (Wahrscheinlichkeiten bei normalverteiltem Merkmal) (3 Punkte)

Für Personen einer bestimmten Grundgesamtheit soll die Ausprägung der Variablen „Intelligenz X“ bestimmt werden. Dabei wird angenommen, dass X sich als eine Zufallsvariable X modellieren lässt, die normalverteilt ist mit Erwartungswert $\mu = 100$ und Varianz $\sigma^2 = 225$, also $\sigma = 15$.

Bestimmen Sie die Wahrscheinlichkeit $P(91 \leq X \leq 109)$ dafür, dass für die Variable X bei einer zufällig aus der betrachteten Grundgesamtheit ausgewählten Person ein Wert gemessen wird, der im Intervall $[91; 109]$ liegt.

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. $0,654$ errechnen, tragen Sie in die letzten sechs Felder $0,6540$ ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch)

Lösung: $0,4515$.

Herleitung:

Die Berechnung ist analog zur Lösung von Aufgabe 12.3b in Kurs 33209. Man erhält mit (12.23) und (12.20) für die $N(100; 15^2)$-verteilte Zufallsvariable X zunächst

$$\begin{aligned} P &= \Phi\left(\frac{109-100}{15}\right) - \Phi\left(\frac{91-100}{15}\right) \\ &= \Phi(0,6) - \Phi(-0,6) = \Phi(0,6) - [1-\Phi(0,6)] \end{aligned}$$

und daraus mit Tabelle 19.2

$$P = 0,7257 - [1 - 0,7257] = 0,7257 - 0,2743 = 0,4514.$$

Das Ergebnis lässt sich auch online anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren). Die Bedingung, dass eine $X \sim N(100; 15^2)$-verteilte Variable eine Ausprägung im Intervall $[95; 105]$ besitzt, ist ja äquivalent dazu, dass eine standardnormalverteilte Variable Z im Intervall $[-0,6; 0,6]$ liegt. Für die Wahrscheinlichkeit $P(-0,6 \leq Z \leq 0,6)$ ermittelt man mit dem interaktiven Experiment bei Einstellung von $a = -0,6$ und $B = 0,6$ den Wert $0,4515$ (minimale rundungsbedingte Abweichung gegenüber dem mit Tabelle 19.2 errechneten Ergebnis).

Anmerkung:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,446; 0,457]$ als richtig gewertet.

Aufgabe 46 (Wahrscheinlichkeiten bei Normalverteilung) (4 Punkte)

Bestimmen Sie einen Wert, der von einer standardnormalverteilten Zufallsvariablen mit Wahrscheinlichkeit $0,2$ nicht überschritten wird. Geben Sie das Ergebnis auf *vier* Stellen nach dem Dezimalkomma genau an. Verwenden Sie für das **Dezimalkomma** wieder ein **eigenes Feld**.

(numerisch)

Lösung: $-0,8416$

Herleitung:

Gesucht ist offenbar – vgl. (12.24) in Kurs 33209 – das $0,2$-Quantil der Standardnormalverteilung, das mit $z_{0,2}$ abgekürzt ist. Aufgrund der Symmetrie der Dichte der Standardnormalverteilung bezüglich des Nullpunktes gilt $z_{0,2} = -z_{0,2}$ Der Tabelle 19.2 ist zu entnehmen, dass $z_{0,8} = 0,8416$ ist. Es gilt also $z_{0,2} = -0,8416$.

Anmerkungen:

Anstelle von $-0,8416$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[-0,833; -0,850]$ als richtig anerkannt. Wenn das Vorzeichen fehlte, gab es 3 statt 4 Punkte.

Das Ergebnis lässt sich auch online anhand eines interaktiven Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und beim Applet im Menüfeld den Wert $p = 0,2$ wählen). Das Experiment liefert auch eine anschauliche Interpretation des Ergebnisses.

Aufgabe 47 (lineares Regressionsmodell / KQ-Schätzung) (3 Punkte)

Auf der Basis eines 20 Wertepaare umfassenden Datensatzes $(x_1, y_1), (x_2, y_2), ..., (x_{20}, y_{20})$, für dessen Schwerpunkt $(\overline{x}, \overline{y})$ sich $\overline{x} = 5,24$ und $\overline{y} = 2,28$ errechnet, wurde eine Regressionsgerade nach der Methode der kleinsten Quadrate (KQ-Methode) geschätzt.

Welchen Wert nimmt die KQ-Regressionsgerade $\widehat{y} = \widehat{\alpha} + \widehat{\beta}x$ an der Stelle $x = 1,5$ an? Gehen Sie davon aus, dass die KQ-Methode für die Steigung der Regressionsgeraden den Wert $\widehat{\beta} = -0,25$ lieferte. Tragen Sie das Ergebnis auf *drei 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: $3,215$

Herleitung:

Zunächst errechnet man für die KQ-Schätzung $\widehat{\alpha}$ den Wert

$$\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} = 2,28 - (-0,25) \cdot 5,24 = 2,28 + 1,31 = 3,59.$$

Einsetzen von $x = 1,5$ in die Regressionsgerade ergibt

$$\overline{y} = 3,59 - 0,25 \cdot 1,5 = 3,590 - 0,375 = 3,215.$$

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von 3,215 auch jeder Wert aus dem Intervall [3,185; 3,245] als richtig anerkannt.

Aufgabe 48 (KQ-Schätzung / Residuen) (3 Punkte)

Auf der Basis eines Datensatzes $(x_1, y_1), (x_2, y_2), ..., (x_{10}, y_{10})$ seien für die Koeffizienten α und β des linearen Regressionsmodells nach der Methode der kleinsten Quadrate die Schätzungen $\widehat{\beta} = 0,58$ und $\widehat{\alpha} = 0,35$ bestimmt worden. Der Datenpunkt (x_1, y_1) sei durch die Koordinaten $x_1 = 2,40$ und $y_1 = 1,96$ definiert.

Berechnen Sie das Residuum $\widehat{u}_1$, also die Abweichung zwischen dem Ordinatenwert y_1 des Datenpunktes (x_1, y_1) und dem Wert der Regressionsgeraden an der Stelle x_1. Tragen Sie das Ergebnis auf *drei 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: $0,218$

Herleitung:

Nach (16.4) in Kurs 33209 ergibt sich

$$\widehat{u}_1 = y_1 - \widehat{y}_1 = y_1 - \widehat{\alpha} - \widehat{\beta} \cdot x_1 = 1,96 - 0,35 - 0,58 \cdot 2,4 = 1,96 - 0,35 - 1,392 = 0,218.$$

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von 0,218 auch jeder Wert aus dem Intervall [0,216; 0,220] als richtig anerkannt.

Wenn bei der Bestimmung von $\widehat{\alpha}$ fälschlich $2,28 - 1,31$ statt $2,28 + 1,31$ gerechnet wurde (Vorzeichenfehler), resultierte $\widehat{\alpha} = 0,97$ und damit $\overline{y} = 0,595$. Für dieses Ergebnis bzw. für jedes Ergebnis aus dem Intervall [0,59; 0,60] wurde 1 Punkt vergeben.