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: **9. März 2009, 14.00 - 18.00 Uhr**

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

Aufgabe 1 (Aussagenlogik) (5 Punkte)

In der nachstehenden Wahrheitstabelle stehen im Tabellenkopf der ersten beiden Spalten zwei Ausagen 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 die Aussage a wahr und b falsch ist, ist $P1$ erfüllt, nicht aber $P2$.

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

D) Wenn die Aussagen a und b beide falsch sind, sind sowohl $P1$ als auch $P2$ erfüllt.

E) Wenn $P1$ und $P2$ wahr sind, ist auch K wahr, d. h. die Konklusion ist korrekt.

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.

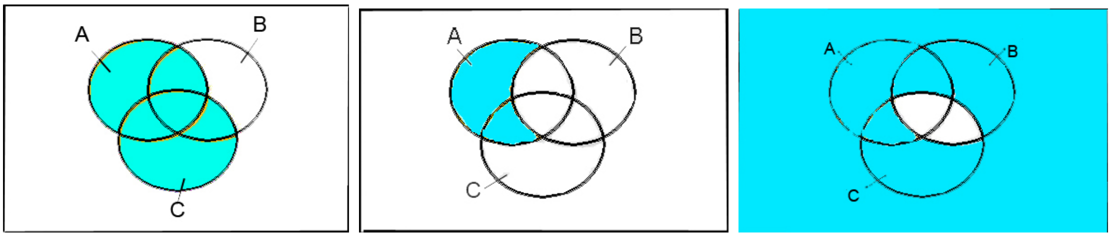
Lösung: B, C, D, E.

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.

Aufgabe 2 (Operationen mit Mengen) (5 Punkte)

Zur Veranschaulichung von Mengenoperationen lassen sich Venn-Diagramme heranziehen. Diese bestehen aus einem Rechteck, in dem Mengen A, B, C, ... als Kreise oder Ellipsen dargestellt sind. Das Rechteck repräsentiert eine Grundgesamtheit, von der die eingezeichneten Mengen Teilmengen sind.

Nachstehend sind drei Venn-Diagramme abgebildet, die sich auf die Verknüpfung von drei Mengen A, B und C beziehen. Die Verknüpfungen erfolgen über die Symbole $\cup$ (Vereinigung von Mengen) und $\cap$ (Schnittmengenbildung) oder $\setminus$ (Differenzmengenbildung). Die Komplementärmengen von A, B und C werden mit $\overline{A}$, $\overline{B}$ und $\overline{C}$ bezeichnet.



Welche der folgenden Aussagen, die sich alle auf die dunkel markierten Flächen beziehen, sind richtig? (x aus 5)

A) Im ersten Venn-Diagramm ist die Vereinigungsmenge der beiden Mengen A und C dargestellt, also $A \cup C$.

B) Im zweiten Venn-Diagramm ist die Schnittmenge aus A und der Komplementärmenge von B dargestellt, also $A \cap \overline{B}$.

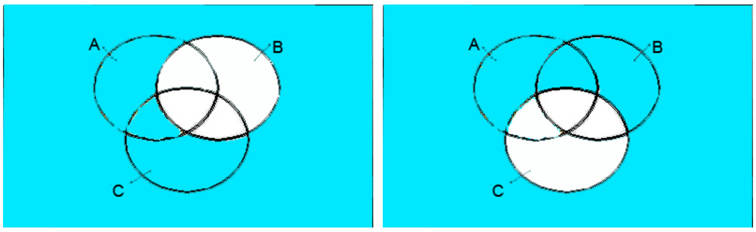
C) Im zweiten Venn-Diagramm ist die Differenzmenge von A und B dargestellt, also $A \setminus B$.

D) Im dritten Venn-Diagramm ist die Vereinigungsmenge von $\overline{B}$ und $\overline{C}$ dargestellt, also $\overline{B} \cup \overline{C}$.

E) Im dritten Venn-Diagramm ist die Komplementärmenge von $B \cap C$ dargestellt.

Lösung: A, B, C, D, E

Zu D: Man kann diese Aussage leichter auf ihren Wahrheitsgehalt hin beurteilen, wenn man zunächst $\overline{B}$ und $\overline{C}$ separat visualisiert:



Die Vereinigungsmenge der beiden dunklen Flächen, die $\overline{B}$ resp. $\overline{C}$ darstellen, ergibt offenbar die Fläche, die im dritten Venn-Diagramm der Aufgabe dunkel dargestellt ist.

Aufgabe 3 (Nominaldefinition) (5 Punkte)

Die folgenden Aussagen beziehen sich auf die Nominaldefinition, also auf die in den Erfahrungswissenschaften übliche Form des Definierens.

Welche Aussagen sind zutreffend? (x aus 5)

A) Eine Nominaldefinition legt fest, dass ein bestimmter Begriff (das Definiendum) mit einem anderen Begriff (dem Definiens) gleichbedeutend ist.

B) Eine Nominaldefinition kann niemals falsch oder richtig sein.

C) Eine Nominaldefinition hat keinen empirischen Informationsgehalt.

D) Eine Nominaldefinition berücksichtigt alle Merkmale des zu definierenden Objekts oder Sachverhalts.

E) Eine Nominaldefinition kann dem Zweck dienen, Begriffe auf die Bedürfnisse eines Forschungsprojekts zuzuschneiden und eine intersubjektive Kontrolle des Forschungsprozesses zu ermöglichen.

Lösung: A, B, C, E.

Zu A: Kromrey, Abschnitt 3.5; 116; Behnke / Behnke, Kapitel 3;
zu B: Kromrey, Abschnitt 3.5;
zu C: Kromrey, Abschnitt 3.5;
zu D: Kromrey, Abschnitt 3.5;
zu E: Kromrey, Abschnitt 3.5.

Aufgabe 4 (Messen / Verfahren der Datenerhebung) (5 Punkte)

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

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

B) Wenn man allgemeine Bevölkerungsumfragen als offene Online-Befragungen organisiert, ist mit erheblichen Verzerrungen zu rechnen.

C) Die Randomized Response Technik ist ein Verfahren der Datenerhebung, bei dem in Interviews vollständige Anonymität der Befragten gewährleistet werden kann.

D) Die „Total-Design-Methode“ ist ein Ansatz der Datenerhebung, der vor allem darauf abzielt, die Rücklaufquote bei schriftlichen Befragungen zu erhöhen.

E) Das Quotenauswahlverfahren ist ein Verfahren der Zufallsauswahl, das Kostenvorteile gegenüber anderen Verfahren der Datenerhebung bietet.

Lösung: B, C, D.

Zu A: Schnell / Hill / Esser, Abschnitt 4.3. Wenn man in der Aussage „Validität (Gültigkeit)“ durch „Reliabilität (Zuverlässigkeit“ ersetzte, wäre sie richtig - vgl. auch Kromrey, Abschnitt 5.7.

zu B: Diekmann, Kapitel X, Abschnitt 11;
zu C: Diekmann, Kapitel X, Abschnitt 7;
zu D: Diekmann, Kapitel X, Abschnitt 10;
zu E: Kromrey, Abschnitte 6.3 - 6.4.

Anmerkung:

Aussage E besteht aus den Teilausaussagen „Das Quotenauswahlverfahren ist ein Verfahren der Zufallsauswahl“ und „Das Quotenauswahlverfahren bietet gegenüber anderen Datenerhebungsverfahren Kostenvorteile“. Die Gesamtaussage ist nur dann richtig, wenn beide Teilaussagen gleichzeitig zutreffen. Die erste Teilaussage ist aber unzutreffend (s. die Übersicht am Ende von Abschnitt 6.3 im Kurs von Kromrey), so dass der Wahrheitsgehalt der Gesamtaussage eindeutig beurteilbar ist. Dennoch wurde bei dieser Aussage generell ein Punkt vergeben, weil der Wahrheitsgehalt der zweiten Teilaussage aus der Pflichtlektüre nur schwer zu erschließen ist.

Aufgabe 5 (Befragungen) (5 Punkte)

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

A) Bei einem Telefoninterview ist der Einfluss von Einflüssen, die in der Person des Interviewers liegen, geringer als bei einem persönlichen Interview (Face-to-Face-Interview).

B) Bei Befragungen wird manchmal die Methode der Klumpenauswahl herangezogen. Diese stellt ein zweistufiges Auswahlverfahren dar, bei dem auf der ersten Stufe eine Zufallsauswahl von Teilmengen einer Grundgesamtheit erfolgt und auf der zweiten Stufe eine Untersuchung aller Elemente der zufällig ausgewählten Teilmengen.

C) Das narrative Interview ist ein Beispiel für eine strukturierte Befragung.

D) Bei Interviews sind Antwortverzerrungen möglich, die z. B. in der Art der Frageformulierung begründet sein können.

E) Non-Response, also Antwortausfälle durch Antwortverweigerung oder Nichterreichbarkeit bei Befragungen, kann die Repräsentativität einer Stichprobe stark beeinträchtigen, also zu verzerrten Stichproben führen.

Lösung: A, B, D, E.

Zu A: Diekmann, Kapitel X, Abschnitt 9;
zu B: Kurs 33209, Abschnitt 3.2, und Kromrey, Abschnitt 6.3.2;
zu C: Diekmann, Kapitel X, Abschnitt 12;
zu D: Diekmann, Kapitel X, Abschnitt 4;
zu E: Diekmann, Kapitel X, Abschnitt 11.

Aufgabe 6 (Merkmalsklassifikationen) (5 Punkte)

Ein Konzern will im Rahmen von Neubauplanungen auch Informationen über den Bedarf an Parkplätzen und Kinderbetreuungsplätzen für seine Mitarbeiter gewinnen. Hierfür werden alle Beschäftigten gebeten, Angaben zu folgenden Merkmalen zu machen:

1 Anzahl der im Haushalt lebenden Kinder

2 Alter der im Haushalt lebenden Kinder

3 Verkehrsmittel, das für die Fahrt zur Arbeitsstätte überwiegend genutzt wird (z. B. eigener PKW, Fahrrad, Bus, ...)

4 Entfernung zwischen Wohnung und Arbeitsstätte

5 Einschätzung des vom Arbeitgeber aktuell vorgehaltenen Kinderbetreuungsangebots (1 = sehr gut, ... , 6 = völlig unzureichend).

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

A) Die Merkmale 1, 2 und 4 sind metrisch skaliert.

B) Das Merkmal 3 ist nominalskaliert, Merkmal 5 hingegen ordinalskaliert.

C) Metrisch skalierte Merkmale lassen sich – unter Informationsverlust – auch auf einer Ordinalskala messen.

D) Bei ordinalskalierten Merkmalen ist die Differenzenbildung zulässig, nicht aber bei nominalskalierten Merkmalen.

E) Merkmal 1 ist ein Beispiel für ein diskretes Merkmal, Merkmal 4 für ein stetiges Merkmal.

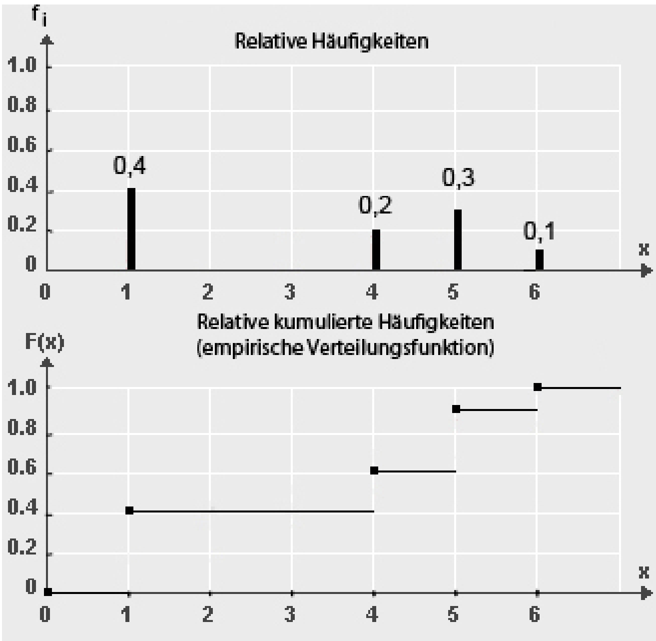
Hinweis: Der Begriff „metrische Skala“ wird als Oberbegriff für „Intervallskala“, „Verhältnisskala“ und „Absolutskala“ verwendet.

Lösung: A, B, C, E.

Zu D: Vgl. Kurs 33209, Tabelle 2.1.

Aufgabe 7 (Häufigkeitsverteilungen) (5 Punkte)

Die nachstehende Abbildung zeigt im oberen Teil die relative Häufigkeitsverteilung anhand eines Stabdiagramms und im unteren Teil die relative kumulierte Häufigkeitsverteilung für einen 20 Werte umfassenden Datensatz, der durch ein Würfelexperiment zustande kam (20-maliges Würfeln mit einem Würfel). Die relativen Häufigkeiten sind im Stabdiagramm auch numerisch ausgewiesen.



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

A) Bei den 20 Würfen trat 6-mal die Augenzahl 5 auf.

B) Von den 20 Würfen führten 12 zu Augenzahlen, die unter 5 lagen.

C) Von den 20 Würfen führten 14 zu Augenzahlen, die unter 5 lagen.

D) Durch Angabe der relativen Häufigkeiten oder der kumulierten relativen Häufigkeiten ist die empirische Verteilung des Merkmals „Augenzahl X“ eindeutig bestimmt.

E) Die relativen Häufigkeiten gehen aus den absoluten Häufigkeiten hervor, indem man letztere durch den Umfang n des Datensatzes dividiert.

Lösung: A, B, D, E.

Zu E: vgl. Kurs 33209, Formel (4.2).

Aufgabe 8 (Kenngrößen empirischer Verteilungen) (5 Punkte)

Bei 20 Arbeitnehmern eines mittelständischen Betriebes mit 150 Beschäftigten wurde die Zeit gemessen, die sie täglich am Bildschirm verbringen (Angaben in vollen Stunden). Es ergab sich der nachstehende Datensatz:

$$3, 4, 1, 2, 2, 6, 5, 4, 3, 5, 2, 2, 5, 2, 5, 4, 1, 2, 5, 5.$$

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

A) Der Datensatz besitzt zwei Modalwerte.

B) Der Mittelwert des Datensatzes ist durch $3,4$ gegeben.

C) Der Median des Datensatzes hat den Wert $3,0$.

D) Streicht man bei obigem Datensatz den ersten Wert (also die erste 3), bleibt der Median unverändert.

E) Die Spannweite des Datensatzes hat den Wert 6.

Lösung: A, B.

Zu A: Sowohl 2 als auch 5 treten je 6-mal auf (beides Modalwerte).
Zu B: Der Mittelwert hat den Wert 3,4 ($n = 20$).
Zu C und D: Der Median hat im Falle $n = 20$ den Wert 3,5, im Falle $n = 19$ den Wert 4,0.
Zu E: Die Spannweite hat den Wert $R = 6 - 1 = 5$.

Aufgabe 9 (Visualisierung empirischer Verteilungen) (5 Punkte)

Es sei erneut der 20 Werte umfassende Datensatz aus Aufgabe 8 betrachtet. Welche der folgenden Aussagen sind richtig ? (x aus 5)

A) Die durch obigen Datensatz definierte empirische Verteilung lässt sich z. B. anhand eines Stabdiagramms für absolute Häufigkeiten oder relative Häufigkeiten visualisieren. Beim Übergang von der einen zur anderen Darstellungsform ändert sich nur die Skalierung der Ordinatenachse.

B) Die empirische Verteilung des Datensatzes lässt sich ohne Informationsverlust auch anhand eines Boxplots veranschaulichen.

C) Ein Boxplot (einfachste Variante) veranschaulicht 5 Charakteristika des Datensatzes. Der Median ist genau in der Mitte der Box eingezeichnet.

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

E) Wenn man die Zeitangaben für alle 150 Beschäftigten hätte und zwar mit höherer Genauigkeit (Angaben in Minuten statt in vollen Stunden), könnte man die Daten zu Klassen zusammenfassen (z. B. Gruppierung der Daten zu 1-Stunden-Bereichen) und die Klassenbesetzungshäufigkeiten anhand eines Histogramms darstellen.

Lösung: A, D, E

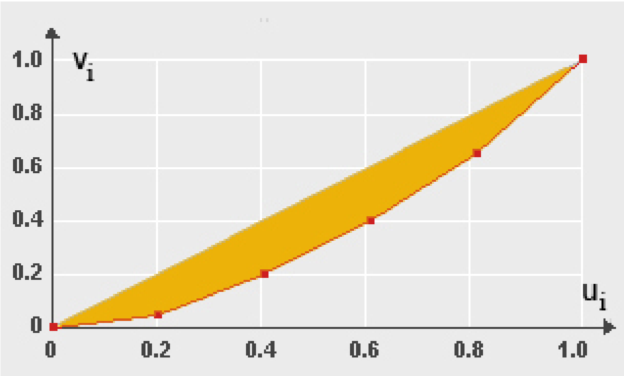
Zu B: Die Aggregation von 20 Werten der Urliste auf nur 5 Kenngrößen ist natürlich mit Informationsverlust verbunden.
Zu C: Der Median ist nur bei einer symmetrischen empirischen Verteilung in der Mitte der Box (vgl. auch Abbildung 5.4 in Kurs 33209).

Anmerkung:

Aussage C war als allgemeine Aussage konzipiert, deren Wahrheitsgehalt unabhängig von einem konkreten Datensatz beurteilt werden sollte. Die Aussage konnte aber auch so interpretiert werden, dass sie in Verbindung mit dem Datensatz aus Aufgabe 8 bewertet werden sollte – bei der Visualisierung dieses Datensatzes anhand eines Boxplots liegt der Median genau in der Mitte der Box. Aufgrund dieser Unschärfe wurde für Aufgabenteil C generell ein Punkt vergeben.

Aufgabe 10 (Konzentrationsmessung) (5 Punkte)

In einer Region konkurrieren fünf Energieversorgungsunternehmen. Es seien $x_1 = 10$, $x_2 = 30$, $x_3 = 40$, $x_4 = 50$ und $x_5 = 70$ die Umsätze dieser Firmen im letzten Geschäftsjahr (Umsätze jeweils in Millionen Euro). Die nachstehende Abbildung zeigt die auf der Basis dieser Umsatzdaten errechnete Lorenzkurve (Polygonzug). Die Stützpunkte (u_i, v_i) der Lorenzkurve sind auf der Lorenzkurve betont. In der Tabelle neben der Grafik sind die Abszissenwerte u_i der Lorenzkurve schon eingetragen.



i	u_i	v_i
0	0	0
1	0,2	v_1
2	0,4	v_2
3	0,6	v_3
4	0,8	v_4
5	1	1

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

A) Der Ordinatenwert v_4 des Stützpunkts $(0,8; v_4)$ der Lorenzkurve hat den Wert $v_4 = 0,65$.

B) Der Gini-Koeffizient liefert Ausagen des Typs „x % der Merkmalsträger teilen sich y % der Merkmalssumme“ (Messung relativer Merkmalskonzentration).

C) Der unnormierte Gini-Koeffizient kann im Falle $n = 5$ nicht größer als $0,8$ werden.

D) Der normierte Gini-Koeffizient hat bei obigem Datensatz den Wert $0,30$.

E) Der Wert, den man für den Ordinatenwert v_2 errechnet, gibt an, welcher Anteil des Gesamtumsatzes aller 5 Energieversorger auf die beiden umsatzschwächsten Unternehmen entfällt.

Beachten Sie auch Aufgabe 41, die an die vorstehende Aufgabe direkt anknüpft.

Lösung: A, B, C, E.

Zu A: Es ist $v_4 = \frac{130}{200} = 0,65$.
Zu C: Der unnormierte Gini-Koeffizient errechnet sich zu $G = 0,28$. Er ist im Falle $n = 5$ durch $G_{max} = \frac{4}{5} = 0,8$ nach oben begrenzt.
Zu D: Für den normierten Gini-Koeffizienten G^* gilt $G^* = \frac{G}{G_{max}} = 0,35$.

Aufgabe 11 (Bivariate empirische Verteilungen) (5 Punkte)

In der nachstehenden Kontingenztabelle für absolute Häufigkeiten sind Ergebnisse der Befragung einer Wählerstichprobe wiedergegeben (Daten des ZDF-Politbarometers vom Februar 1998). Die Häufigkeiten beziehen sich auf $n = 931$ befragte Personen, die ihre Parteipräferenz für den fiktiven Fall einer am nächsten Sonntag bevorstehenden Bundestagswahl geäußert hatten. Die Tabelle zeigt die Ergebnisse differenziert nach Geschlecht und jeweils mit Angabe der Randverteilungen für die Merkmale „Parteipräferenz X“ und „Geschlecht Y“.

Ausprägungen von X		Ausprägungen von Y: b_1	b_2	Zeilensummen (Randverteilung von X)
CDU CSU	a_1	142	203	345
SPD	a_2	194	160	354
GRÜNE	a_3	40	53	93
PDS	a_4	35	20	55
F.D.P. Die Liberalen	a_5	14	23	37
sonstige	a_6	20	27	47
Spaltensummen → (Randverteilung von Y)		445	486	931

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

A) Die Randverteilung von X ist nichts anderes als die Häufigkeitsverteilung dieses Merkmals, die sich bei Verzicht auf die Differenzierung nach Geschlecht ergibt.

B) Aus der Tabelle ergibt sich, dass von den 931 Befragten 445 Männer waren.

C) Der Tabelle entnimmt man, dass von den 931 Befragten 354 Personen die SPD präferierten.

D) Aus den Werten in der Tabelle lassen sich auch bedingte relative Häufigkeiten ermitteln, z. B. die bedingte relative Häufigkeit $f_Y(b_2|a_2)$. Letztere gibt an, welcher Anteil der befragten Personen, die sich für die SPD entschieden hatten, Frauen waren.

E) Für die in Aufgabenteil D genannte bedingte Häufigkeit errechnet man bei Rundung auf 3 Dezimalstellen den Wert $0,352$, d. h. $35,2\%$.

Lösung: A, B, C, D.

Zu A: Vgl. hierzu Beispiel 8.1 in Kurs 33209.
Zu B: Vgl. die Summe der Elemente der ersten Spalte der in der Aufgabe wiedergegebenenen Häufigkeitstabelle.
Zu C: Vgl. die Summe der beiden Elemente der zweiten Zeile der in der Aufgabe wiedergegebenenen Häufigkeitstabelle.
Zu D: Vgl. hierzu Aufgabe 8.1 des Kurses 33209.
Zu E: Man errechnet den Wert $0,452$, also $45,2\%$ (s. hierzu Beispiel 8.2 in Kurs 33209).

Aufgabe 12 (Zusammenhangsmessung) (5 Punkte)

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

A) Der Korrelationskoeffizient r nach Bravais-Pearson misst die Stärke eines 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 oder fallenden Geraden liegen.

C) Im Falle $r = 0$ ist noch nicht ausgeschlossen, dass zwischen den Merkmalen X und Y ein nicht-linearer Zusammenhang besteht.

D) Der Rangkorrelationskoeffizient r_{SP} nach Spearman ist ein für ordinalskalierte Merkmale anwendbares Zusammenhangsmaß, das – wie der Korrelationskoeffizient r nach Bravais-Pearson – stets Werte zwischen -1 und $+1$ annimmt.

E) Der Rangkorrelationskoeffizient r_{SP} lässt sich auch auf metrisch skalierte Merkmale anwenden. Die in den Daten enthaltene Information wird dann aber nicht ausgeschöpft, weil bei der Berechnung von r_{SP} nur die Rangpositionen der Werte für X resp. für Y verarbeitet werden.

Lösung: A, C, D, E.

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

Aufgabe 13 (Kombinatorik) (5 Punkte)

Eine „faire" Münze, also eine Münze mit gleichen Eintrittswahrscheinlichkeiten für „Kopf" und „Zahl", wird n-mal geworfen und die Anzahl X der Ausgänge mit „Zahl" festgestellt. Welche der folgenden Aussagen sind richtig ? (x aus 5)

A) Die Wahrscheinlichkeit dafür, im Falle $n = 3$ höchstens einmal „Zahl" zu erhalten, ist $0,5$.

B) Die Wahrscheinlichkeit dafür, im Falle $n = 3$ genau einmal „Zahl" zu erhalten, ist $0,375$.

C) Die Wahrscheinlichkeit dafür, im Falle $n = 4$ mindestens einmal „Zahl" zu erhalten, ist kleiner als $0,9$.

D) Die Wahrscheinlichkeit dafür, im Falle $n = 4$ mindestens einmal „Zahl" zu erhalten, ist genau so groß wie die Wahrscheinlichkeit höchstens dreimal „Zahl" zu erzielen.

E) Die Anzahl X der Ausgänge mit „Zahl" lässt sich durch eine Binomialverteilung mit den Parametern n und $p = 0,5$ modellieren.

Lösung: A, B, D, E.

Zu A: Die Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 3$ und $p = 0,5$ nimmt für $x = 1$ den Wert $F(1) = 0,500$ an (s. Tabelle 19.1).
Zu B: Die Differenz der Werte $F(1)$ und $F(0)$ der Verteilungsfunktion der Binomialverteilung mit $n = 3$ und $p = 0,5$ ist gegeben durch $F(1) - F(0) = 0,500 - 0,125 = 0,375$ (vgl. wieder Tabelle 19.1).
Zu C: Die Wahrscheinlichkeit bei 4 Würfen 0-mal „Zahl" zu erhalten, ist durch den Wert $F(0) = 0,0625$ der Verteilungsfunktion der Binomialverteilung mit $n = 4$ und $p = 0,5$ gegeben. Gesucht ist hier die Gegenwahrscheinlichkeit $1 - F(0) = 1 - 0,0625 = 0,9375$.

Aufgabe 14 (Konfidenzintervalle) (5 Punkte)

Welche der folgenden Aussagen, die sich auf Konfidenzintervalle für den Erwartungswert $\mu = E(X)$ eines normalverteilten Merkmals X mit bekannter Varianz σ^2 beziehen, sind richtig ? (x aus 5)

A) Das Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha$ wird schmaler, wenn der Stichprobenumfang n erhöht wird.

B) Das Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha$ wird schmaler, wenn $1-\alpha$ erhöht, also die Irrtumswahrscheinlichkeit α verkleinert wird.

C) Der unbekannte Parameter μ kann auch außerhalb des Konfidenzintervalls liegen.

D) Die Grenzen eines Konfidenzintervalls lassen sich als Ausprägungen von Stichprobenfunktionen interpretieren, also als Realisationen von Zufallsvariablen.

E) Keine der vorstehenden Aussagen ist richtig.

Lösung: A, C, D.

Zu A: Vgl. hierzu in Kurs 33209 die Formel (14.15) sowie die Abbildungen 14.2 und 14.3.
Zu B: Wenn in Formel (14.15) der Wert α verkleinert wird, bedeutet dies, dass $p := 1 - \frac{\alpha}{2}$ etwas größer wird und damit auch das Quantil z_p (vgl. Tabelle 19.3 in Kurs 33209).
Zu C und D: Die Abbildungen 14.2 und 14.3 des Kurses 33209 liefern Beispiele für die Korrektheit der Aussage.

Aufgabe 15 (Normalverteilung; Standardnormalverteilung) (5 Punkte)

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

A) Die Wahrscheinlichkeit dafür, dass eine normalverteilte Zufallsvariable eine Ausprägung x mit $x \leq 0$ annimmt, beträgt $0,5$.

B) Jede normalverteilte Zufallsvariable kann anhand einer Lineartransformation in eine standardnormalverteilte Zufallsvariable überführt werden.

C) Der Wert der Verteilungsfunktion einer standardnormalverteilten Zufallsvariablen an der Stelle 1 stimmt mit dem Wert der Verteilungsfunktion an der Stelle -1 überein.

D) Die Wahrscheinlichkeit $P(X \leq a)$ dafür, dass eine normalverteilte Variable X eine Ausprägung hat, die nicht größer als a ist, entspricht dem Wert der Verteilungsfunktion $F(x)$ der betreffenden Normalverteilung an der Stelle $x = a$.

E) Ist Z eine standardnormalverteilte Zufallsvariable, so ist für diese die Bedingung $P(Z \leq a) = 0,975$ erfüllt, wenn man für a den Wert $a = 1,96$ wählt.

Lösung: B, D, E.

Zu A: Die Aussage gilt nur für die Standardnormalverteilung.
Zu B: Vgl. in Kurs 33209 den Text vor der Tabelle 19.2.
Zu C: Die Aussage wäre für die Dichtefunktion zutreffend, nicht aber für die Verteilungsfunktion $\Phi(z)$ (vgl. die obere Hälfte von Abbildung 12.2 in Kurs 33209). Für $\Phi(z)$ gilt gemäß (12.20) vielmehr $\Phi(1) = 1 - \Phi(-1)$.
Zu E: Der Wert a ist das $0,975$-Quantil der Standardnormalverteilung (s. Tabelle 19.2).

Aufgabe 16 (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 α (zweiseitiger Test).

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

A) Wenn man die Varianz σ^2 als bekannt annimmt, kann man den standardisierten Stichprobenmittelwert $Z = (\overline{X} - \mu_0)/\sigma_{\overline{X}}$ als Prüfgröße für den Test heranziehen (Gauß-Test). Die Dichte dieser Prüfgröße ist symmetrisch bezüglich des Nullpunkts.

B) Die Nullhypothese wird verworfen, wenn die Prüfgröße einen Wert annimmt, der innerhalb des Ablehnungsbereichs liegt. Die Grenzen des Ablehnungsbereichs hängen vom Signifikanzniveau α ab.

C) Die fälschliche Verwerfung der Nullhypothese H_0 wird als Fehler 1. Art bezeichnet. Die Wahrscheinlichkeit für den Eintritt eines Fehlers 1. Art hat beim hier betrachteten Test den Wert α.

D) Wenn die Prüfgröße im Annahmebereich liegt, kann die Nullhypothese als statistisch „bewiesen“ angesehen werden, in dem Sinne, dass ihre Gültigkeit mit einer Irrtumswahrscheinlichkeit von α als gesichert angenommen werden kann.

E) Wenn man bei obigem Test die Varianz σ^2 resp. die Standardabweichung σ nicht als bekannt voraussetzen kann und eine Schätzung $\widehat{\sigma}$ heranzieht (Schätzung von σ durch die korrigierte Stichprobenstandardabweichung), ist die resultierende Prüfgröße t-verteilt. Die Grenzen des Ablehnungsbereichs für die Nullhypothese liegen dann – bei unverändertem Signifikanzniveau α – enger zusammen.

Beachten Sie auch Aufgabe 43, die an die vorstehende Aufgabe direkt anknüpft.

Lösung: A, B, C.

Zu C: Vgl. auch Tabelle 15.1 in Kurs 33209.
Zu D: Wenn die Prüfgröße im Ablehnungsbereich liegt, kann die Alternativhypothese H_1 (nicht aber die Nullhypothese) als statistisch „bewiesen“ angesehen werden, in dem Sinne, dass ihre Gültigkeit mit einer Irrtumswahrscheinlichkeit von α als gesichert angenommen werden kann.
Zu E: Die Grenzen des Ablehnungsbereichs rücken etwas weiter auseinander (s. Abbildung 15.2 in Kurs 33209).

Aufgabe 41 (Konzentrationsmessung) (4 Punkte)

Geben Sie den Flächeninhalt A an, der in Aufgabe 10 in der Abbildung dunkel markiert ist. Gemeint ist also der Inhalt der Fläche, die von der Lorenzkurve und der Strecke gebildet wird, die den Nullpunkt $(0;0)$ mit dem Punkt $(1;1)$ verbindet. Geben Sie Ihre Antwort mit drei Nachkommastellen rechtsbündig in das Antwortfeld ein. Verwenden Sie für das **Dezimalkomma** unbedingt ein **eigenes Feld**. Falls Sie also z. B. „0,36“ errechnen, tragen Sie in die letzten fünf Felder „0,360“ ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) $A =$ ☐☐☐☐☐☐

Lösung: Der unnormierte Gini-Koeffizient errechnet sich zu $G = 0,28$ – vgl. auch Formel (6.5) in Kurs 33209. Die veranschaulichte Fläche entspricht $\frac{G}{2} = 0,14$.

Anmerkung: Wenn anstelle von $0,14$ der Wert $0,28$, also G statt $\frac{G}{2}$, eingetragen wurde, wurde nur ein Punkt abgezogen (Vergabe von 3 Punkten). Wenn der Wert $0,56$ angegeben wurde, der Wert G also versehentlich verdoppelt statt halbiert wurde, wurden 2 Punkte vergeben.

Aufgabe 42 (Kombinatorik) (4 Punkte)

Bei der Deutschen Meisterschaft in der Disziplin 100-m-Lauf (Herren) treten im Endlauf sechs Kandidaten $K1, K2, ...K6$ gegeneinander an. Wieviele Möglichkeiten für die Verteilung der ersten drei Plätze gibt es? Tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch) ☐☐☐☐☐☐

Lösung:

$$\frac{6!}{(6-3)!} = \frac{6 \cdot 4 \cdot ... \cdot 1}{3 \cdot 2 \cdot 1} = 6 \cdot 5 \cdot 4 = 120.$$

Kommentar: In der Terminologie des Urnenmodells werden $n = 3$ Kugeln aus einer Urne mit $N = 6$ nummerierten Kugeln *ohne Zurücklegen* gezogen und *mit Berücksichtigung der Anordnung*. Es ist also die erste der in der ersten Zeile von Tabelle 10.1 des Kurs 33209 wiedergegebenen Formeln mit $n = 3$ und $N = 6$ anzuwenden.

Anmerkung: Falls fälschlich die zweite Formel in der ersten Zeile von Tabelle 10.1 mit $n = 3$ und $N = 6$ angewendet wurde (das Ergebnis lautet dann $6^3 = 216$), die sich auf den Fall des Ziehens *mit Zurücklegen* gezogen und *mit Berücksichtigung der Anordnung* bezieht, wurde ein Punkt vergeben, weil hier zumindest der Fall "Ziehen mit Zurücklegen" erkannt wurde. Wenn versehentlich die erste Formel aus der zweiten Zeile von Tabelle 10.1 angewendet wurde, die beim Ziehen *ohne Zurücklegen* und *ohne Berücksichtigung der Anordnung* zutrifft, gab es nur deswegen die volle Punktzahl, weil hier – zufällig – ebenfalls der Wert 120 als Ergebnis resultiert.

Aufgabe 43 (Gauß-Test) (4 Punkte)

Berechnen Sie für den Gauß-Test aus Aufgabe 16 (s. dort die Aufgabenteile A-B) für den Fall $\alpha = 0,05$ zunächst die Grenzen des Annahmebereichs für die Nullhypothese und tragen Sie dann in das Antwortfeld ein, wie weit diese beiden Werte auseinander liegen. Es ist also die Länge des Annahmebereichs gesucht. Geben Sie ihre Antwort auf vier Dezimalstellen nach dem Komma genau an. Falls Sie für den Annahmebereich das Intervall $[a; b]$ errechnen mit irgendwelchen Werten a und b und für die Länge $b - a$ des Intervalls z. B. $4,523$, wäre $4,5230$ in die sechs Felder einzutragen. Das **Dezimalkomma** muss also auch hier unbedingt ein **eigenes Feld** belegen.

(numerisch)

Lösung: Die Grenzen des Annahmebereichs sind durch die Quantile $z_{\alpha/2} = -z_{1-\alpha/2}$ und $z_{1-\alpha/2}$ der Standardnormalverteilung gegeben (s. Abbildung 15.1), also durch $-z_{0,975}$ und $z_{0,975}$. Nach Tabelle 19.2 ist $z_{0,975} = 1,96$. Der Annahmebereich $[-1,96; 1,96]$ ist also ein Intervall der Länge $3,92$.

Anmerkung: Für die Standardisierung des Stichprobenmittelwerts und die mit ihr einhergehende Berechnung einer Realisation der Prüfstatistik Z (Formel 15.2 im Kurs) werden noch Informationen benötigt, die in der obigen Aufgabe fehlten – z. B. die Angabe des Stichprobenumfangs n. Erst mit Kenntnis des Werts der Prüfgröße kann eine Testentscheidung getroffen werden. Hier war allerdings nur die Berechnung der Länge des Annahmebereichs verlangt, für deren Berechnung bei diesem Test die Angabe eines Werts für die Irrtumswahrscheinlichkeit α genügt.

Für diese Aufgabe wurde generell die volle Punktzahl gegeben (Ausnahmeregelung), weil es an einem Klausurort mit dieser Aufgabe ein Problem gab.

Aufgabe 44 (Wahrscheinlichkeiten bei Normalverteilung) (4 Punkte)

Bei der Serienfertigung eines bestimmten Schraubentyps ist aus der Auswertung eines Produktionsvorlaufs bekannt, dass der Durchmesser X des Schraubenkopfs normalverteilt ist mit Erwartungswert $\mu = 6,0$ (Zielwert) und einer Standardabweichung $\sigma = 0,01$ (Angaben in mm), also $X \sim N(6; 0,01^2)$. Weicht der Schraubenkopfdurchmesser bei einer Schraube um mehr als $0,02$ (mm) nach oben oder nach unten vom Zielwert ab, d. h. liegt eine beobachtete Realisation x von X nicht im Intervall $[5,98; 6,02]$, so ist die Schraube für den vorgesehenen Einsatzzweck nicht brauchbar. Sie gilt dann als Ausschuss.

Wie groß ist die Wahrscheinlichkeit P dafür, dass der Durchmesser einer Schraube im Intervall $[5,98; 6,02]$ liegt, die Schraube also kein Ausschuss ist? 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. „0,8934“ errechnen, tragen Sie in die letzten sechs Felder „0,8934“ ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) $P =$ | | | | | | |

Lösung: $0,9544$

Herleitung:

$$P(5,98 \leq X \leq 6,02) = P(-0,02 \leq X - 6 \leq 0,02) = P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2).$$

Nach Tabelle 19.2 ist $\Phi(2) = 0,9772$. Ferner gilt $\Phi(-2) = 1 - 0,9772 = 0,0228$. Damit ist $\Phi(2) - \Phi(-2) = 0,9544$.

Anmerkung:

Bei der maschinellen Auswertung wurden Werte von $0,95$ bis $0,96$ als richtig anerkannt. Bei Eintragung des Wertes $0,9772$, also $\Phi(2)$ statt $\Phi(2) - \Phi(-2)$, wurde ein Punkt vergeben, weil hier zumindest der Wert $\Phi(2)$ korrekt ermittelt wurde.

Aufgabe 45 (Kleinst-Quadrat-Schätzung) (4 Punkte)

Für zwei Merkmale X und Y stehen 6 Beobachtungswerte $(x_1, y_1), ..., (x_6, y_6)$ zur Verfügung, nämlich die Werte $(6,0; 3,0)$, $(5,0; 3,2)$, $(7,0; 2,5)$, $(7,0; 2,3)$, $(8,0; 2,0)$ und $(9,0; 2,0)$. Für diesen Datensatz wurde nach der Methode der kleinsten Quadrate eine Regressionsgerade

$$\widehat{y} = \widehat{\alpha} + \widehat{\beta} x.$$

bestimmt, wobei sich für $\widehat{\beta}$ aus den Daten der Wert $\widehat{\beta} = -0,34$ ergab. Errechnen Sie auch den Wert $\widehat{\alpha}$, der sich nach der Kleinst-Quadrat-Methode ergibt. Tragen Sie Ihr Ergebnis rechtsbündig und auf zwei Stellen nach dem Dezimalstellen genau in das Antwortfeld ein. Sie würden also vier Felder benötigen, wenn etwa „5,43“ Ihre Lösung wäre, denn das **Dezimalkomma** muss auch hier wieder ein **eigenes Feld** belegen. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

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

Lösung: Es ist $\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x}$. Aus den Daten erhält man $\overline{x} = 7$ und $\overline{y} = 2,5$ und folglich $\widehat{\alpha} = 2,5 + 0,34 \cdot 7 = 4,88$.

Lösung: Es ist $\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x}$. Aus den Daten erhält man $\overline{x} = 7$ und $\overline{y} = 2,5$ und folglich $\widehat{\alpha} = 2,5 - (-0,34) \cdot 7 = 2,5 + 2,38 = 4,88$.

Anmerkung:

Bei der maschinellen Auswertung wurden Werte von $4,87$ bis $4,89$ als richtig anerkannt. Wenn allerdings anstelle von $4,88$ der Wert $0,12$ als Lösung eingetragen wurde, wurde nur ein Punkt abgezogen (Vergabe von 3 Punkten), weil hier offenbar nur ein Vorzeichen übersehen wurde und fälschlich $\widehat{\alpha} = 2,5 - 0,34 \cdot 7 = 2,5 - 2,38$ gerechnet wurde.