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: 13. September 2010, 14.00 - 18.00 Uhr

Prüfer: apl. Prof. Dr. H.-J. Mittag

Erstellt: 27. Oktober 2010

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

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

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, 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, 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. Zu dieser Aufgabe vgl. auch Kurs 33210; Kapitel 5.

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) Ein Ansatz zur empirischen Überprüfung von Forschungshypothesen ist die Durchführung von Experimenten. Bei Experimenten, z. B. in der Medizin oder der Psychologie, werden i. a. zwei Gruppen von Teilnehmern gebildet (manchmal auch mehr als zwei Gruppen) – eine Versuchsgruppe und eine Kontrollgruppe. Wenn die Zuordnung der Teilnehmer zu den beiden Gruppen nicht zufällig erfolgt, spricht man von einem Quasi-Experiment.

B) Für beide in Aufgabenteil A genannten Gruppen – also sowohl für die Versuchs- als auch für die Kontrollgruppe – gilt, dass Einflussgrößen planmäßig verändert werden.

C) Die verdeckte Erfassung und Auswertung der Verweildauer von Kunden in einzelnen Abteilungen von Supermärkten ist ein Beispiel für ein Verfahren der nichtreaktiven Datengewinnung.

D) Undercoverage ist ein Fehler, der bei stichprobenbasierten Datenerhebungen auftreten kann. Er entsteht, wenn nicht alle Elemente der Population, aus der eine Stichprobe gezogen wird, bei der Stichprobenziehung berücksichtigt werden.

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

Lösung: A, C, D, E.

Zu A: Kromrey, Abschnitt 2.4.3 und Kurs 33209, Abschnitt 3.1.

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

zu C: Diekmann, Kapitel XIII, Abschnitt 1, oder Kurs 33209, Abschnitt 3.1.

zu D: Diekmann, Kapitel X, Abschnitt 11, oder Kurs 33209, Abschnitt 3.2.

zu E: Diekmann, Kapitel X, Abschnitt 11.

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 Bildungssabschlusses (Ausprägungen des Merkmals “Bildungsstand“: ohne Abschluss, Hauptschule, mittlere Reife, Fachhochschulreife, Abitur, akademischer Abschluss). Das Merkmal „Bildungsstand“ ist ordinalskaliert.

C) Das in Aufgabenteil B genannte Merkmal „Bildungsstand“ ist ein diskretes Merkmal.

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 (Nominal- und Realdefinitionen) (5 Punkte)

Welche der folgenden Aussagen, die sich auf Nominal - und Realdefinitionen beziehen, sind zutreffend? (x aus 5)

A) Nominaldefinitionen sind Worterklärungen, die festlegen, dass ein bestimmter Begriff (das Definiendum) mit einem anderen Begriff (dem Definiens) gleichbedeutend ist.

B) Eine Realdefinition umfasst alle Eigenschaften des Definiendums.

C) Realdefinitionen sind Worterklärungen, die – im Gegensatz zu Nominaldefinitionen – das Definiendum durch bestimmte Eigenschaften charakterisieren.

D) Eine Nominaldefinition ist entweder richtig oder falsch.

E) Eine Nominaldefiniton hat keinen empirischen Informationsgehalt.

Lösung: A, C, E.

Zu A: Kromrey, Abschnitt 3.5.1 oder Kurs 33210, Kapitel 3;

zu B: Kromrey, Abschnitt 3.5.4;

zu C: Kromrey, Abschnitt 3.5.4;

Zu D: Kromrey, Abschnitt 3.5.3;

zu E: Kromrey, Abschnitt 3.5.1.

Aufgabe 5 (Messen / Stichprobenverfahren) (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) Die Klumpenauswahl ist eine zweistufige Auswahlprozedur, bei der in der ersten Stufe Teilmengen der Grundgesamtheit (als Klumpen oder Cluster bezeichnet) zufällig ausgewählt werden.

C) Die Klumpenauswahl ist eine zweistufige Auswahlprozedur, bei der aus den Klumpen (in der zweiten Verfahrensstufe) Zufallsstichproben gezogen werden.

D) Bei der Realisierung einer geschichteten Stichprobe wird die Grundgesamtheit bezüglich eines Merkmals – z. B. Familienstand oder Berufsgruppe – in Teilmengen zerlegt (erste Verfahrensstufe) und dann aus jeder Teilmenge eine Zufallsstichprobe gezogen (zweite Verfahrensstufe).

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

Lösung: B, D, E.

Zu A: Wenn man in der Aussage „Validität“ durch „Reliablität“ ersetzte, wäre sie richtig. Die Validität (Gültigkeit) ist das Gütekriterium für Messungen, mit dem beschrieben wird, inwieweit wirklich das gemessen wird, was gemessen werden soll; vgl. Kromrey, Abschnitt 5.7, und Schnell / Hill / Esser, Abschnitt 4.3.

zu B: vgl. Kromrey, Abschnitt 6.5, oder Kurs 33209, Abschnitt 3.2.

zu C: In der zweiten Stufe werden alle Elemente eines Klumpens herangezogen, nicht nur eine Zufallsauswahl – vgl. Kromrey, Abschnitt 6.5.2 oder Kurs 33209, Abschnitt 3.2.

zu D: vgl. Kromrey, Abschnitt 6.5.2 oder Kurs 33209, Abschnitt 3.2.

zu E: vgl. Kromrey, Abschnitt 6.5.3.

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 letzten 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)

In einer Notiz des Statistischen Bundesamtes hieß es, dass $80,9$ % der Gesamtbevölkerung Deutschlands (D-Gesamt) in den alten Bundesländern lebt. Ferner wurde mitgeteilt, dass der Anteil der ausländischen Mitbürger in den alten Bundesländern (D-West) bei $10,3$ % liegt, in den neuen Bundesländern (D-Ost) hingegen bei nur $2,4$ %.

Bei der Bevölkerung in D-Gesamt, D-West und D-Ost seien, leicht vereinfachend, jeweils nur zwei Gruppen unterschieden, nämlich ausländische Mitbürger – im Folgenden kurz als Ausländer bezeichnet – und Deutsche. (Es werden also Sonderfälle wie doppelte oder fehlende Staatsbürgerschaften ausgeschlossen.)

Welche der folgenden Aussagen sind richtig, wenn man die Gesamtbevölkerung Deutschlands mit $82,0$ Millionen veranschlagt: (x aus 5)

A) Die Anzahl der in D-West lebenden Deutschen liegt zwischen 59 und 60 Millionen.

B) Die Anzahl der in D-Ost lebenden Ausländer liegt unter $0,4$ Millionen.

C) Der Anteil der in D-West lebenden Ausländer an der Gesamtbevölkerung Deutschlands beträgt mehr als $8,7$ %.

D) Die Gesamtzahl der Ausländer in D-Gesamt liegt über $7,4$ Millionen.

E) Wenn man die Anzahl der Ausländer und die Anzahl der Deutschen zunächst für D-West und danach auch für D-Ost addiert, erhält man die absolute Randverteilung für das hier binär spezifizierte Merkmal „Wohnort“ (Ausprägungen „West“ und „Ost“).

Lösung: A, B, E - vgl. auch die Beispiele 8.1 und 8.2 in Kurs 33209. Die Lösungen findet man am einfachsten anhand eines Baumdiagramms oder einer Vierfeldertafel (nachstehend wiedergegeben).

Zu A: Die Aussage bezieht sich bei dem folgenden Baumdiagramm auf die absolute Häufigkeit $h_{12} = 59,505$, trifft also zu.

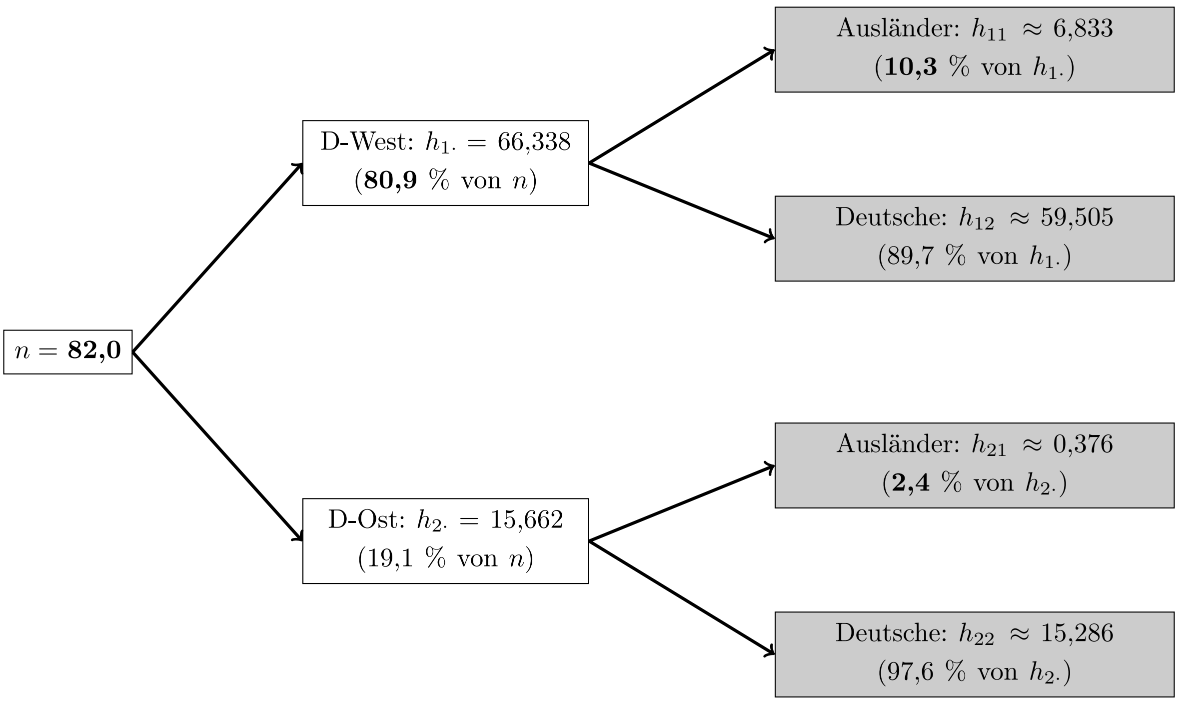
Zu B: Die Aussage bezieht sich auf die absolute Häufigkeit $h_{21} = 0,376$, trifft also zu (s. Baumdiagramm oder die erste der nachstehenden Vierfeldertafeln).

Zu C: Die Aussage bezieht sich auf die relative Häufigkeit $f_{11} = 0,083$ (s. Baumdiagramm oder die zweite der nachstehenden Vierfeldertafeln). Dies entspricht $8,3$ %. 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 $h_{\cdot 1} = 6,833 + 0,376 = 7,209$, die unterhalb von $7,4$ liegt.

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

Die der Pressenotiz zu entnehmenden Angaben lassen sich in Form eines Baumdiagramms darstellen (vgl. auch Beispiel 8.2 in Kurs 33209). Ein solches Baumdiagramm ist nachstehend wiedergegeben, wobei die Angaben aus der Pressenotiz durch Fettdruck betont sind:



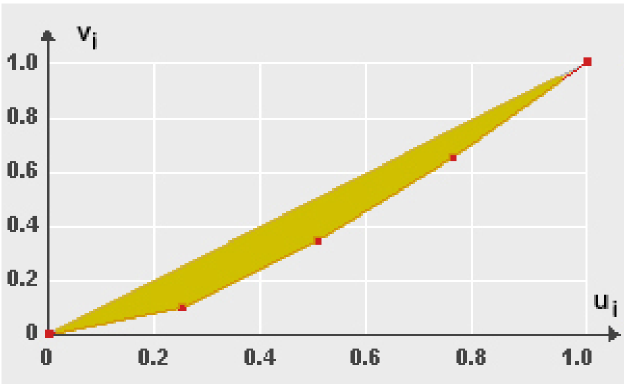
Hieraus erhält man folgende Vierfeldertafel für absolute Häufigkeiten bzw. – nach Division aller Werte durch n – für relative Häufigkeiten (alle Werte auf drei Nachkommastellen gerundet):

	Ausländer	Deutsche	Zeilensummen
D-West	$6,833$	$59,505$	$66,338$
D-Ost	$0,376$	$15,286$	$15,662$
Spaltensummen	$7,209$	$74,791$	$82,0$

	Ausländer	Deutsche	Zeilensummen
D-West	$0,083$	$0,726$	$0,809$
D-Ost	$0,005$	$0,186$	$0,191$
Spaltensummen	$0,088$	$0,912$	1

Aufgabe 8 (Konzentrationsmessung) (5 Punkte)

In einer Region konkurrieren vier Energieversorgungsunternehmen. Es seien $x_1 = 20$, $x_2 = 50$, $x_3 = 50$ und $x_4 = 80$ 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	0,25	v_1
2	0,50	v_2
3	0,75	v_3
5	1	1

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

A) Der Ordinatenwert v_2 des Stützpunkts $(0,5; v_2)$ der Lorenzkurve hat den Wert $v_2 = 0,35$.

B) Der Wert, den man für den Ordinatenwert v_3 errechnet, gibt an, welcher Anteil des Gesamtumsatzes aller 4 Energieversorger auf die drei umsatzschwächsten Unternehmen entfällt.

C) Allgemein gilt, dass der unnormierte Gini-Koeffizient im Falle $n = 4$ einen Wert bis maximal $0,8$ annehmen kann.

D) Der normierte Gini-Koeffizient hat bei obigem Datensatz den Wert $0,3$.

E) Wenn man den Herfindahl-Index H für die Konzentrationsmessung heranzieht, erhält man hier einen Wert, der zwischen $0,27$ und $0,29$ liegt.

Lösung: A, B, D.

Zu A: Es ist $v_2 = \frac{70}{200} = 0,35$.

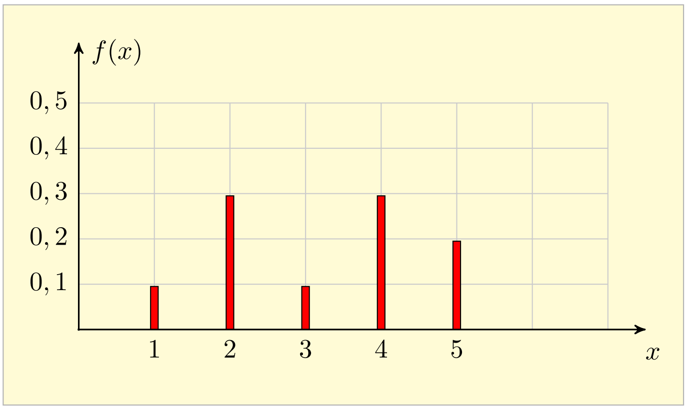
Zu C: Der unnormierte Gini-Koeffizient errechnet sich zu $G = 0,225$. Er ist im Falle $n = 4$ durch $G_{max} = \frac{3}{4} = 0,75$ nach oben begrenzt.

Zu D: Für den normierten Gini-Koeffizienten G^* gilt $G^* = \frac{G}{G_{max}} = \frac{0{,}225}{0{,}75} = 0,30$.

Zu E: Nach Formel (6.9) in Kurs 33209 ist $H = \frac{1}{200^2} \cdot (20^2 + 50^2 + 50^2 + 80^2) = 0,295$.

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

Die nachstehende Abbildung zeigt die Wahrscheinlichkeitsfunktion $f(x)$ einer diskreten Zufallsvariablen X, die fünf Ausprägungen $x_1 = 1, x_2 = 2, \ldots, x_5 = 5$ aufweist. Die Ausprägungen x_1 und x_3 weisen jeweils die Eintrittswahrscheinlichkeit $0,1$ auf, x_2 und x_4 je die Eintrittswahrscheinlichkeit $0,3$ und x_5 wird mit Wahrscheinlichkeit $0,2$ realisiert.



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 $x = 2,5$ den Wert $0,4$ an.

C) Die Verteilungsfunktion $F(x)$ von X ist nur bis $x = 5$ definiert.

D) Der Erwartungswert $E(X)$ der Zufallsvariablen X ist größer als $2,9$, aber kleiner als $3,1$.

E) Wenn man die Zufallsvariable X gemäß $Y = X^2$ transformiert und den Erwartungswert $E(X)$ der ursprünglichen Variablen X mit μ bezeichnet, so ist der Erwartungswert $E(Y)$ der neuen Variablen durch μ^2 gegeben.

Lösung: A, B.

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

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

- 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 $F(2,5) = 0,4$. Die Verteilungsfunktion $F(x)$ ist auch für $x > 5$ definiert. Sie erreicht an der Stelle $x = 5$ den Wert 1, den sie dann für $x > 5$ beibehält.

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

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

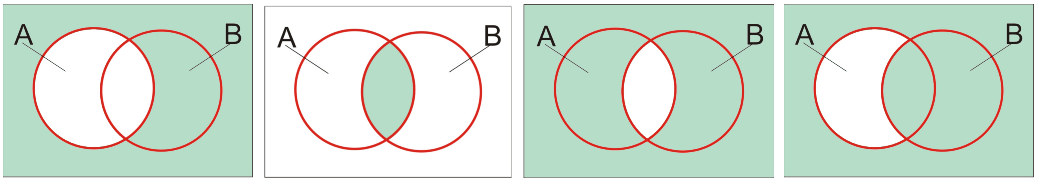
Zu E: Auch der Erwartungswert $\mu_Y := E(Y)$ von $Y = X^2$ errechnet sich nach (11.6) als Summe der mit den Eintrittswahrscheinlichkeiten gewichteten Ausprägungen. Allerdings sind die Ausprägungen nun zu quadrieren:

$$\mu_Y = E(Y) = 1 \cdot 0,1 + 4 \cdot 0,3 + 9 \cdot 0,1 + 16 \cdot 0,3 + 25 \cdot 0,2 = 12,0 \quad (\neq \mu^2 = 10,24).$$

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 Komplementärmenge $\overline{A}$ von A.

B) Die dunkel gefärbte Fläche im zweiten Venn-Diagramm stellt die Vereinigungsmenge von $\overline{A}$ und $\overline{B}$ dar, also $\overline{A} \cup \overline{B}$.

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

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

E) Bildet man die Schnittmengen $A \cap B$ und $\overline{A} \cap B$, so sind die beiden resultierenden Mengen disjunkt.

Lösung: A, D, E.

Zu A und D: 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 Schnittmenge von A und B dar, also $A \cap B$.

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

Zu E: Die erste der genannten Schnittmengen ist in der zweiten Abbildung dargestelt; die zweite Schnittmenge ist B ohne den Überlappungsbereich mit A. Diese beiden Mengen sind disjunkt (ihre Vereinigung ergibt B).

Aufgabe 11 (Kombinatorik / diskrete Verteilungen) (5 Punkte)

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

A) Aus einer Lostrommel, in der jedes fünfte Los einen Gewinn repräsentiert, werden nacheinander 3 Lose *mit* Zurücklegen gezogen. Die Wahrscheinlichkeit dafür, nach den 3 Ziehungen keinen Gewinn gezogen zu haben, ist nicht größer als $0,5$.

B) Wenn man eine „faire“ Münze, also eine Münze mit gleichen Eintrittswahrscheinlichkeiten für „Kopf“ und „Zahl“, 6-mal wirft und die Anzahl X der Ausgänge mit „Zahl“ feststellt, liegt die Wahrscheinlichkeit dafür, *mindestens* dreimal „Zahl“ zu erhalten, zwischen $0,60$ und $0,62$.

C) Die Wahrscheinlichkeit dafür, bei dem $6-$maligem Münzwurf aus Aufgabenteil B *genau* dreimal „Zahl“ zu erhalten, ist größer als $0,3$.

D) Beim Würfeln mit zwei fairen Würfeln, also Würfeln mit gleichen Eintrittswahrscheinlichkeiten für alle Augenzahlen, liegt die Wahrscheinlichkeit dafür, eine Augensumme zu erhalten, die nicht kleiner als 9 ist, zwischen $0,30$ und $0,32$.

E) Mit einem fairen Würfel werde 9-mal in Folge gewürfelt. Bezeichne X die Anzahl der Ausgänge mit einer Augenzahl, die kleiner als 3 ist. Der Erwartungswert von X hat den Wert 3.

Lösung: C, E.

Zu A: Die gesuchte Wahrscheinlichkeit ist hier durch den Wert $F(0)$ der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 3$ und $p = 0,2$ gegeben, nach Tabelle 19.1 also durch $0,5120$.

Zu B: Die Wahrscheinlichkeit $P(X \geq 3)$ dafür, mindestens dreimal „Zahl“ beim 6-maligen Münzwurf zu erhalten, ist die Komplementärwahrscheinlichkeit von $P(X \leq 2)$. Letztere ist durch den Wert der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 6$ und $p = 0,5$ an der Stelle 2 gegeben, nach Tabelle 19.1 also durch $F(2) = 0,3438$. Es gilt also $P(X \geq 3) = 1 - 0,3438 = 0,652$.

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

Zu D: Vgl. hierzu die letzten Zeilen von Beispiel 10.1 in Kurs 33209. Von den 36 Elementarereignissen, die den Ereignisraum Ω beim Würfeln mit zwei Würfeln definieren, führen 10 Elementarereignisse, nämlich die Ausgänge $(3;6)$, $(6;3)$; $(4;5)$, $(5;4)$, $(4;6)$, $(6;4)$; $(5;5)$, $(5;6)$, $(6;5)$ und $(6;6)$, zu einer Augensumme, die nicht kleiner als 9 ist.

Die Wahrscheinlichkeit P für die Erzielung einer Augensumme, die mindestens 9 ist, hat somit den Wert $P = \frac{10}{36} \approx 0,278$.

Zu E: Es ist der Erwartungswert $\mu = E(X)$ einer Binomialverteilung mit $n = 9$ 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 $\mu = 9 \cdot \frac{1}{3} = 3$.

Aufgabe 12 (Zufallsvariablen) (5 Punkte)

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

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

B) Ist X eine normalverteilte Zufallsvariable mit Erwartungswert $\mu = 1$ und Varianz $\sigma^2 = 4$, so ist die Aussage $P(X \leq a) = 0,5$ zutreffend, wenn man für a den Wert $a = 1$ wählt.

C) Es sei X eine mit $m = 5$ und $n = 20$ Freiheitsgraden F-verteilte Zufallsvariable. Das $0,95$-Quantil dieser F-Verteilung liegt zwischen $2,70$ und $2,72$.

D) Es sei X eine t-verteilte Zufallsvariable. Die Wahrscheinlichkeit dafür, dass X eine Ausprägung annimmt, die den Wert 0 nicht überschreitet, ist $0,5$.

E) Eine mit 10 Freiheitsgraden χ^2-verteilte Zufallsvariable nimmt mit Wahrscheinlichkeit $0,05$ eine Ausprägung an, die oberhalb von $3,940$ liegt.

Lösung: B, C, D.

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 $\sigma = V(X) = 2$.

Zu B: Die Dichtekurven von normalverteilten Zufallsvariablen X sind symmetrisch bezüglich des Erwartungswertes μ. Daraus folgt, dass $P(X \leq \mu) = 0,5$ ist. Da hier $\mu = 1$ ist, ist die Aussage $P(X \leq 1) = 0,5$ also zutreffend.

Man kann natürlich auch anhand von (12.21) herleiten, dass für die Zufallsvariable X mit $X \sim N(1; 2^2)$ die Gleichung $P(X \leq 1) = 0,5$ gilt. Hiernach gilt ja $P(X \leq 1) = \Phi(\frac{1-1}{2}) = \Phi(0) = 0,5$.

Zu C: Nach Tabelle 19.6 in Kurs 33209 hat das $0,95$-Quantil einer mit $m = 5$ und $n = 20$ Freiheitsgraden F-verteilten Zufallsvariablen den Wert $2,71$.

Zu D: Die Dichtekurven von t-verteilten Zufallsvariablen X sind symmetrisch bezüglich des Erwartungswertes. Letzerer hat – unabhängig von der Anzahl der Freiheitsgrade – den Wert 0. Daraus folgt, dass $P(X \leq 0) = 0,5$ ist.

Zu E: Aus Tabelle 19.4 kann man ablesen, dass das $0,95$-Quantil $\chi^2_{10;0,95}$ einer mit 10 Freiheitsgraden χ^2-verteilten Zufallsvariablen X den Wert $18,307$ hat und das $0,05$-Quantil $\chi^2_{10;0,05}$ den Wert $3,940$. Das letztgenannte Quantil hat die Eigenschaft, dass X mit Wahrscheinlichkeit $0,05$ eine Ausprägung annimmt, die $\chi^2_{10;0,05} = 3,940$ nicht überschreitet, also $\chi^2_{10;0,05} = 3,940$ mit Wahrscheinlichkeit $0,95$ überschreitet. Die Aussage E wäre z. B. korrekt, wenn man in der Aufgabe „oberhalb“ durch „nicht oberhalb“ ersetzte oder aber den Wert $3,940$ durch $18,307$.

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? (x aus 5)

A) Die Stichprobenfunktion S^2 liefert eine unverzerrte Schätzung für die Varianz σ^2.

B) Wenn man den Erwartungswert μ anhand des Stichprobenmittelwerts $\overline{X}$ schätzt und die Qualität dieses Schätzers anhand des mittleren quadratischen Fehlers (MSE) beurteilt, hat man mit $\overline{X}$ eine Schätzfunktion, deren MSE und Varianz übereinstimmen.

C) Man kann den Erwartungswert μ eines normalverteilten Merkmals auch durch Angabe eines Konfidenzintervalls schätzen. Die Grenzen einer solchen Intervallschätzung sind nicht zufallsabhängig.

D) Ein Konfidenzintervall zum Konfidenzniveau $0,95$ für den Erwartungswert μ ist ein Intervall, das μ mit einer Wahrscheinlichkeit von $0,95$ enthält – den zu schätzenden Wert μ mit einer Wahrscheinlichkeit von $0,05$ also nicht enthält.

E) Die Länge eines Konfidenzintervalls für μ nimmt ab, wenn die Irrtumswahrscheinlichkeit α reduziert wird. (Mit „Länge“ ist hier der Abstand zwischen den beiden Intervallgrenzen gemeint.)

Lösung: B, D.

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

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

Zu C: Formel (14.14) zeigt die Zufallsabhängigkeit der Grenzen der Konfidenzintervalle (die Abbildungen 14.2 und 14.3 veranschaulichen dies).

Zu D: Vgl. hierzu Abbildung 14.2.

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

Aufgabe 14 (Korrelationsmessung, lineares Regressionsmodell) (5 Punkte)

In der nachstehende Tabelle sind für zwei Merkmale X und Y Beobachtungsdaten $(x_i; y_i)$ wiedergegeben $(i = 1, 2, ..., 6)$.

i	x_i	y_i
1	10	3,2
2	12	3,0
3	14	2,5
4	14	2,3
5	16	2,0
6	18	2,0

Aus diesen Daten errechnet man u. a.

$$\sum_{i=1}^{6} (x_i - \overline{x})^2 = 40; \qquad \sum_{i=1}^{6} (y_i - \overline{y})^2 = 1,28; \qquad \sum_{i=1}^{6} (x_i - \overline{x})(y_i - \overline{y}) = -6,8.$$

Diese Werte sind hier ungeprüft zu übernehmen.

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

A) Wenn man für die in der obigen Tabelle wiedergegebenen Beobachtungspaare $(x_i; y_i)$ unterstellt, dass zwischen x_i und y_i $(i = 1, 2, ..., 6)$ 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 $\widehat{y} = \widehat{\alpha} + \widehat{\beta} x$ ist fallend (negative Steigung).

B) Der nach der Kleinst-Quadrat-Methode errechnete Wert $\widehat{\alpha}$ kennzeichnet den Schnittpunkt der Regressionsgeraden mit der y-Achse.

C) Mit dem den in der Tabelle wiedergegebenen Datensatz errechnet man für den Korrelationskoeffizienten nach Bravais-Pearson einen Wert zwischen $-0,8$ und $-0,7$.

D) Die Güte der Anpassung der mit der Kleinst-Quadrat-Methode erhaltenen Regressionsgeraden an den Datensatz lässt sich anhand des Bestimmtheitsmaßes bewerten. Dieses stimmt im Falle des einfachen Regressionsmodells mit dem Korrelationskoeffizienten nach Bravais-Pearson überein.

E) Würde man für das Bestimmtheitsmaß den Wert $0,45$ errechnen, würde dies beinhalten, dass 45 % der Gesamtvariation des Datensatzes durch das verwendete Regressionsmodell erklärt ist.

Lösung: A, B, E.

Zu A - B: Es gilt nach (16.6) und (16.7) bei Beachtung von $\overline{x} = 14$ und $\overline{y} = 2,5$

$$\widehat{\beta} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{-6,8}{40} = -0,17.$$
$$\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} = 2,5 - (-0,17) \cdot 14 = 0,12.$$

Setzt man in die berechnete KQ-Regressionsgerade $\widehat{y} = 0,12 - 0,17 \cdot x$ für x den Wert $x = 0$ ein, folgt $\widehat{y} = 0,12$, also der für $\widehat{\alpha}$ berechnete Wert.

Zu C: Nach (9.11) und (9.12) in Kurs 33209 gilt für den Korrelationskoeffizienten r nach Bravais-Pearson

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{-6,8}{\sqrt{40} \cdot \sqrt{1,28}} = \frac{-6,8}{\sqrt{12,8}} \approx -0,95.$$

Zu D: Es gilt $R^2 = r^2$. Der Korrelationskoeffizient r kann zudem, anders als das Bestimmtheitsmaß, auch negative Werte annehmen.

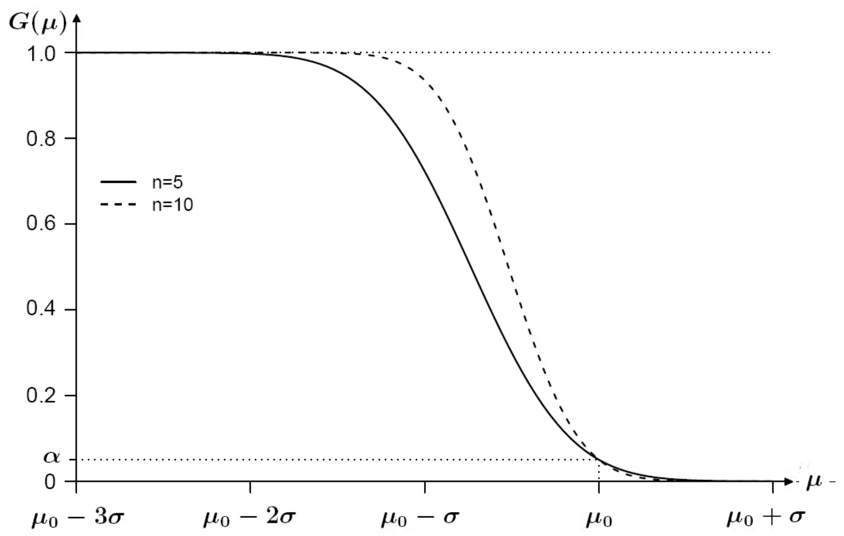
Zu E: vgl Aufgabe 16.2 in Kurs 33209.

Aufgabe 15 (Gütefunktion beim Gauß-Test, Fehler beim Testen) (5 Punkte)

Es seien n unabhängige Beobachtungen für ein normalverteiltes Merkmal gegeben (Normalverteilung mit unbekanntem Erwartungswert μ und bekannter 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 $\alpha = 0,05$ (Gauß-Test). Die nachstehende Grafik zeigt die Gütefunktion des Tests, die die Ablehnwahrscheinlichkeit für die Nullhypothese als Funktion des Erwartungswerts μ darstellt, für $n = 5$ und zusätzlich auch für $n = 10$:



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

A) Die Werte $G(\mu)$, die beide Gütefunktionen für $\mu \geq \mu_0$ annehmen, sind als Wahrscheinlichkeiten für den Eintritt eines Fehlers 1. Art zu interpretieren.

B) Die Werte $G(\mu)$, die beide Gütefunktionen für $\mu < \mu_0$ annehmen, sind Wahrscheinlichkeiten für den Eintritt eines Fehlers 2. Art.

C) Bei obigem Test wird die Nullhypothese im Falle $\mu = \mu_0$ mit Wahrscheinlichkeit $0,95$ nicht verworfen.

D) Der Wert α, den die beiden in der Grafik wiedergegebenen Gütefunktionen an der Stelle $\mu = \mu_0$ annehmen, repräsentiert bei dem hier betrachteten Test die maximale Wahrscheinlichkeit für den Eintritt eines Fehlers 1. Art.

E) Die Erhöhung des Stichprobenumfangs führt für alle Werte μ mit $\mu \neq \mu_0$ zur Senkung der Wahrscheinlichkeiten für den Eintritt beider Testrisiken (Fehler 1. und Fehler 2. Art).

Lösung: A, C, D, E.

Zu A: Die Werte $G(\mu)$ spezifizieren die Wahrscheinlichkeit für die Ablehnung der Nullhypothese $H_0: \ \mu \geq \mu_0$. Wenn für den zu testenden (unbekannten) Parameter μ die Bedingung $\mu \geq \mu_0$ erfüllt ist, ist die Ablehnung von H_0 eine Fehlentscheidung (Fehler 1. Art).

Zu B: Zu B: Wenn der tatsächliche (aber unbekannte) Wert des Parameters μ in der Grundgesamtheit kleiner als μ_0 ist, ist die Ablehnung von H_0 eine korrekte Entscheidung. Eine solche korrekte Testentscheidung tritt mit Wahrscheinlichkeit $G(\mu)$ ein. Die Wahrscheinlichkeit für den Eintritt eines Fehlers 2. Art (Nicht-Ablehnung von H_0) ist durch $1 - G(\mu)$ gegeben, nicht durch $G(\mu)$.

Zu C: Im Falle $\mu = \mu_0$ wird H_0 mit Wahrscheinlichkeit $0,05$ verworfen (Fehler 1. Art), also mit Wahrscheinlichkeit $0,95$ nicht verworfen (korrekte Entscheidung).

Zu D: Die Nullhypothese $H_0: \ \mu \geq \mu_0$ wird für $\mu = \mu_0$ mit der Wahrscheinlichkeit $G(\mu_0) = 0,05$ fälschlich verworfen (Fehler 1. Art), im Falle $\mu < \mu_0$ hingegen mit einer Fehlerwahrscheinlichkeit $G(\mu)$, die kleiner als $0,05$ ist. Die Eintrittswahrscheinlichkeit für einen Fehler 1. Art ist also für $\mu = \mu_0$ am größten.

Zu E: Vgl. hierzu in Kurs 33209 den auf Abbildung 15.4 folgenden Absatz.

Numerische Aufgaben

Aufgabe 41 (bedingte Häufigkeiten) (3 Punkte)

Nachstehend sind Ergebnisse der Befragung einer Wählerstichprobe wiedergegeben. Die Daten beziehen sich auf $n = 1021$ befragte Personen, die ihre Parteipräferenz für den Fall einer am nächsten Sonntag bevorstehenden Bundestagswahl geäußert hatten. Die Tabelle zeigt absolute Häufigkeiten, differenziert nach Geschlecht und jeweils mit Angabe der Randverteilungen für die Merkmale „Parteipräferenz X“ und „Geschlecht Y“. Beim Merkmal X werden hier die Ausprägungen a_1 = CDU /CSU, a_2 = SPD, a_3 = FDP, a_4 = Linke, a_5 = Grüne, a_6 = Sonstige unterschieden (bei $a_1 - a_5$ jeweils mit Ausweis des Parteienlogos). Die Ausprägungen des Merkmals Y sind mit b_1 (= männlich, mit Marssymbol) und b_2 (= weiblich, mit Venussymbol) codiert.

			Ausprägungen von Y		
			♂ b_1	♀ b_2	
Ausprägungen von X		a_1	179	204	383
	SPD	a_2	100	117	217
	FDP	a_3	80	59	139
		a_4	67	50	117
		a_5	54	62	116
	Sonstige	a_6	21	28	49
			501	520	1021
			Randverteilung von Y		Randverteilung von X

Bestimmen Sie anhand der Tabelle den Anteil der Männer in der Stichprobe, die bei der Sonntagsfrage die SPD favorisiert hatten. Runden Sie das Ergebnis auf *drei Stellen nach dem Dezimalkomma* und tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Wenn Sie z. B. den Wert $0,24$ errechnen ($24,0$ %), wäre also $0,240$ einzutragen. Das **Dezimalkomma** belegt wieder ein **eigenes Feld**. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) | | | | | | |

Lösung: $0,200$.

Herleitung:

Wenn man – wie in Beispiel 8.3 in Kurs 33209 – die Merkmale „Parteipräferenz“ und „Geschlecht“ mit X resp. Y bezeichnet und auch die Merkmalsausprägungen wie in Beispiel 8.3 bezeichnet, ist der gesuchte Anteil die bedingte relative Häufigkeit von X unter der Bedingung $Y = b_1$:

$$f_X(a_2|b_1) = \frac{h_{21}}{h_{\cdot 1}} = \frac{100}{501} \approx 0,200.$$

Anmerkungen:

Bei der maschinellen Auswertung dieser Aufgabe, bei der das Antwortformat im Text eindeutig vorgegeben war $(0, xxx)$, wurde anstelle von $0,200$ auch jeder Wert aus dem Intervall $[0,195; 0,205]$ als richtig anerkannt.

Da die Aufgabe u. U. auch so interpretiert werden konnte, dass der Anteil der Frauen mit Wahlpräferenz „SPD“ an der Gesamtstichprobe gemeint war – die Bezugsbasis wäre hier nicht die Gesamtzahl 501 der Männer in der Stichprobe, sondern die Gesamtzahl 1021 aller Befragten – oder auch der Anteil der Männer an der (Teil-) Stichprobe der Befragten mit SPD-Präferenz – die Bezugsbasis wäre dann die Gesamtzahl 217 aller Befragten mit SPD-Präferenz – wurden auch die bei diesen Interpretationen resultierenden Antworten $\frac{100}{1021} \approx 0,098$ resp. $\frac{100}{217} \approx 0,461$ anerkannt, wobei anstelle von $0,098$ auch jeder Wert aus dem Intervall $[0,097; 0,099]$ und anstelle von $0,461$ jeder Wert aus dem Intervall $[0,455; 0,465]$ als richtig galt.

In der Klausur für „Psychologie“ war in dieser Aufgabe nach dem Anteil der *Frauen* in der Wählerstichprobe gefragt.

Aufgabe 42 (Rangkorrelationskoeffizient) (3 Punkte)

Zwei Restauranttester bewerten in 7 Frankfurter Sterne-Restaurants die Qualität eines Desserts, das von allen beteiligten Restaurants angeboten wird. Die Bewertung erfolgt anhand eines 10-stufigen Schemas (1 = sehr schlecht, 10 = hervorragend). Die Ergebnisse der Bewertungen sind nachstehend ausgewiesen.

Untersuchen Sie anhand des Rangkorrelationskoeffizienten r_{SP} von Spearman, ob zwischen den Bewertungen der beiden Tester ein Zusammenhang besteht. Tragen Sie Ihr Ergebnis auf *drei Stellen* nach dem Dezimalkomma genau rechtsbündig in das Antwortfeld ein. Wenn Sie also z. B. den Wert $0,4673$ errechnen, wäre $0,467$ einzutragen. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

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

Restaurant i	Tester 1 Bewertung x_i	Tester 2 Bewertung y_i
1	7	8
2	8	9
3	6	3
4	4	5
5	10	7
6	9	10
7	5	6

Lösung: $0,679$.

Herleitung:

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

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

Bei der Zusammenhangsmessung anhand des Rangkorrelationskoeffizienten r_{SP} von Spearman kann anstelle von (9.14) aus Kurs 33209 die vereinfachte Formel (9.16) angewendet werden (S. 9 unten in der Formelsammlung), weil die Bewertungen nicht mit der Mehrfachbelegung eines Rangplatzes verbunden ist. Mit (9.16) resultiert

$$r_{SP} = 1 - \frac{6 \cdot [1^2 + 1^2 + (-2)^2 + 1^2 + (-3)^2 + 1^2 + 1^2]}{7 \cdot (49 - 1)} = 1 - \frac{18}{56} \approx 1 - 0,321 = 0,679.$$

Anmerkungen:

Bei der maschinellen Auswertung wurde anstelle von $0,679$ auch jeder Wert aus dem Intervall $[0,675; 0,685]$ als richtig anerkannt.

Wenn anstelle von $r_{SP} \approx 1 - 0,321 = 0,679$ fälschlich der Wert $0,321$ als Lösung angegeben wurde, wurden für diesen Wert bzw. für jeden Wert aus dem Intervall $[0,315; 0,325]$ 2 P. gewährt.

Aufgabe 43 (Kombinatorik) (3 Punkte)

Bei der Fußball-WM in Südafrika gab es zunächst Vorrundenspiele. Es gab 4 Gruppen mit je 4 Mannschaften. In jeder Gruppe wurden alle möglichen Mannschaftspaarungen realisiert, was zu 6 Spielen pro Gruppe führte.

Hätte man nur 2 Gruppen mit je 8 Mannschaften gebildet, hätte man in jeder Gruppe weitaus mehr Spiele durchführen müssen (die Vorrunden hätten also mehr Zeit beansprucht), wenn jede Mannschaft einmal gegen jede Mannschaft derselben Gruppe spielen sollte. Bestimmen Sie die Anzahl der Spiele, die es pro Gruppe dann gegeben hätte. Tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch) | | | | | | |

Lösung: 28.

Herleitung:

Bei 8 Mannschaften errechnet sich die Anzahl der möglichen Mannschaftskombinationen nach Formel (10.9) in Kurs 33209 mit $N = 8$ und $n = 2$ (Ziehen ohne Zurücklegen und ohne Berücksichtigung der Anordnung) von 2 Elementen aus einer Grundgesamtheit von $N = 8$ Elementen. Man erhält

$$\binom{8}{2} = \frac{8!}{6! \cdot 2!} = \frac{8 \cdot 7}{2 \cdot 1} = 28.$$

Aufgabe 44 (Wahrscheinlichkeiten bei Normalverteilung) (3 Punkte)

Es sei angenommen, dass sich in einer größeren Population ein Persönlichkeitsmerkmal (z. B. „Intelligenz“) anhand einer Variablen X modellieren lässt, die normalverteilt ist mit Erwartungswert $\mu = 100$ und Standardabweichung $\sigma = 15$.

Wie groß ist bei Gültigkeit dieser Annahme die Wahrscheinlichkeit dafür, dass die Ausprägung des Merkmals für eine zufällig ausgewählte Person mindestens 70, aber nicht mehr als 100 beträgt? 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. 0,3276 errechnen, tragen Sie in die letzten sechs Felder 0,3276 ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

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

Lösung: $0,4772$

Herleitung: Nach Formel (12.23) in Kurs 33209 gilt

$$P(70 \leq X \leq 100) = \Phi\left(\frac{100-100}{15}\right) - \Phi\left(\frac{70-100}{15}\right) = \Phi(0) - \Phi(-2) = \Phi(0) - [1 - \Phi(2)]\,.$$

Mit Tabelle 19.2 des Kurses 33209 folgt

$$P(70 \leq X \leq 100) = 0,5 - (1 - 0,9772) = 0,5 - 0,0228 = 0,4772.$$

Die Wahrscheinlichkeit dafür, dass X zwischen 70 und 100 liegt, beträgt also $0,4772$ (entspricht $47,72$ %).

Anmerkungen:

Anstelle von $0,4772$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,475; 0,480]$ als richtig anerkannt.

Aufgabe 45 (Mini-Lotto) (3 Punkte)

Beim deutschen Lotto „6 aus 49“ ist die Wahrscheinlichkeit alle bei der Ziehung zum Zuge kommenden Zahlen gewählt zu haben, äußerst gering. Wenn man die Anzahl der Richtigen mit X bezeichnet, erhält man für die Wahrscheinlichkeit $P(X = 6)$ eines „Volltreffers“ den Wert

$$P(X = 6) = \frac{1}{13983816} \approx 0,0000000715 = 7,15 \cdot 10^{-8}.$$

Bei der Eröffnung eines Freizeitzentrums will man den am Eröffnungstag erscheinenden Kunden kostenlos ein Lotto „6 aus 9“ anbieten, eine Art „Mini-Lotto“, bei dem die Gewinnchance – die Aussicht auf einen 20-Euro-Einkaufgutschein – wesentlich höher liegt. Es ist also hier nur eine Gewinnklasse vorgesehen, anders als beim normalen Lotto, bei dem man auch mit weniger als 6 Richtigen gewinnen kann.

Das Mini-Lottospiel ist so organisiert, dass man den Kunden beim Betreten des Gebäudes einen Tippschein überreicht, auf dem die Zahlen $1, 2, \ldots, 8, 9$ stehen, von denen sechs Zahlen anzukreuzen sind. Am Ende des Tages werden dann aus einer Trommel, in der sich neun mit $1, 2, \ldots, 8, 9$ nummerierte Kugeln befinden, sechs verschiedene Kugeln wie bei der Lottoziehung ausgewählt.

Wie groß ist für jeden teilnehmenden Kunden die Wahrscheinlichkeit $P(X = 6)$, einen 20-Euro-Einkaufsgutschein zu gewinnen? Tragen Sie Ihr Ergebnis 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: $0,0119$

Herleitung:

In der Terminologie des Urnenmodells werden $n = 6$ Kugeln aus einer Urne mit $N = 9$ nummerierten Kugeln *ohne Zurücklegen* und *ohne Berücksichtigung der Anordnung* gezogen (vgl. Tabelle 10.1 in Kurs 33209). Die Anzahl der Möglichkeiten errechnet sich also gemäß

$$\binom{9}{6} = \frac{9!}{3! \cdot 6!} = \frac{9 \cdot 8 \cdot 7}{2 \cdot 3} = 84.$$

Für die Wahrscheinlichkeit $P(X = 6)$ folgt dann nach Formel (10.5) in Kurs 33209

$$P(X = 6) = \frac{1}{84} \approx 0,0119.$$

Dies entspricht immerhin ca. $1,19$ %.

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von $0,0119$ auch jeder Wert aus dem Intervall $[0,0115; 0,0125]$ als richtig anerkannt.

Einige Klausurteilnehmer haben die gesuchte Wahrscheinlichkeit, die einen Wert zwischen 0 und 1 haben muss, in Prozentwerte umgerechnet und entsprechend $0,0119 \cdot 100 = 1,1900$ in den Lotsebogen eingetragen. Es wurde jeder Wert aus dem Intervall $[1,15; 1,25]$ berücksichtigt und 2 P. vergeben.

Wenn anstelle von $\frac{1}{84} \approx 0,0119$ fälschlich der Wert 84 als Lösung angegeben wurde, also die Invertierung vergessen wurde, wurde noch 1 P. gewährt.

Aufgabe 46 (Erwartungswert beim Mini-Lotto) (3 Punkte)

Berechnen Sie den Erwartungswert $\mu = E(X)$ der Zufallsvariablen X aus Aufgabe 45, also den Erwartungswert für die Anzahl der Richtigen beim Mini-Lotto „6 aus 9".

Tragen Sie Ihr Ergebnis rechtsbündig und auf *eine* Nachkommastelle 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) $\mu =$ ☐☐☐☐☐☐☐

Lösung : 4,0.

Herleitung:

Der Erwartungswert bestimmt sich nach Formel (11.24) des Kurses 33209, wobei dort $n = 6$ sowie $M = 6$ und $N = 9$ einzusetzen ist:

$$\mu = 6 \cdot \frac{6}{9} = 4.$$

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

Auf der Basis eines Datensatzes $(x_1, y_1), (x_2, y_2), ..., (x_{15}, y_{15})$, für dessen Schwerpunkt $(\overline{x}, \overline{y})$ sich $\overline{x} = 9,5$ und $\overline{y} = 3,6$ 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 $\widehat{\beta} = -0,16$ lieferte.

Tragen Sie das Ergebnis auf *zwei Stellen nach dem Dezimalkomma* genau rechtsbündig in das Antwortfeld ein. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**. Falls Sie also z. B. 3,126 errechnen, tragen Sie in die letzten vier Felder 3,13 ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) ☐☐☐☐☐☐

Lösung: $4,88$

Herleitung:

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

$$\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} = 3,6 - (-0,16) \cdot 9,5 = 3,6 + 1,52 = 5,12.$$

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

$$\overline{y} = 5,12 - 0,16 \cdot 1,5 = 5,12 - 0,24 = 4,88.$$

Anmerkung:

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von $2,48$ auch jeder Wert aus dem Intervall $[4,87;4,89]$ als richtig anerkannt. Wenn das Zwischengebnis $5,12$ auf dem Lotsebogen eingetragen bzw. ein Wert aus dem Intervall $[5,13;5,14]$ wurde 1 P. gewährt.

Aufgabe 48 (Gauß-Test) (3 Punkte)

Es sei noch einmal der Gauß-Test betrachtet (vgl. Aufgabe 15), der sich auf n unabhängige Beobachtungen für ein normalverteiltes Merkmal (Normalverteilung mit unbekanntem Erwartungswert μ und bekannter Varianz σ^2 und die Hypothesen

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

bezog. Es sei $n = 5$, $\mu_0 = 50$ und $\sigma^2 = 25$, also $\sigma = 5$. Für das Signifikanzniveau des Tests gelte $\alpha = 0,05$. Wie groß ist die Wahrscheinlichkeit dafür, dass die Nullhypothese verworfen wird, wenn $\mu = 45$ ist?

Tragen Sie das Ergebnis auf *drei Stellen nach dem Dezimalkomma* genau rechtsbündig in das Antwortfeld ein. Verwenden Sie für das **Dezimalkomma** ein **eigenes Feld**. Falls Sie also z. B. 0,5282 errechnen, tragen Sie in die letzten fünf Felder 0,528 ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

Hinweis zu Aufgabe 48:

Wenn Sie bei der Berechnung der gesuchten Wahrscheinlichkeit eine Tabelle der Formelsammlung heranziehen und diese nicht die gewünschte Genauigkeit liefert, können Sie den nächstgelegenen Tabellenwert verwenden oder interpolieren. Beispiel: Sollten Sie etwa den Wert der Verteilungsfunktion $\Phi(z)$ an der Stelle $z = 0,347$ suchen, könnten Sie den tabellierten Wert $\Phi(0,35)$ verwenden oder aber $\Phi(0,34)$ und $\Phi(0,35)$ ablesen und interpolieren.

(numerisch)

Lösung: $0,722$

Herleitung:

Die Gütefunktion des hier betrachteten linksseitigen Gauß-Tests ist durch

$$G(\mu) = \Phi\left(-z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \cdot \sqrt{n}\right)$$

gegeben, für $\mu = 45$ und mit $\mu_0 = 50$, $\sigma = 5$, $n = 5$ sowie $\alpha = 0,05$ also durch

$$G(\mu) = \Phi\left(-z_{0,95} + 1 \cdot \sqrt{5}\right) \approx \Phi(-1,6449 + 2,236) \approx \Phi(0,591) \approx 0,722.$$

Die Größenordnung der Lösung kann man bereits der Grafik zu Aufgabe 15 entnehmen. Die durchgezogene Kurve dort zeigte die Gütefunktion des Tests im Falle $n = 5$. Zu berechnen war der Wert $G(\mu)$ dieser Gütefunktion an der Stelle $\mu = 45$, wegen $\mu_0 = 50$ und $\sigma = 5$, also für $\mu = \mu_0 - \sigma$. Man erkennt schon grob, dass $G(\mu)$ hier etwas größer als $0,7$ ist, d. h. die Wahrscheinlichkeit für den Eintritt eines Fehlers 2. Art ist etwas kleiner als $0,3$ (genauer: $0,278$ oder $27,8$ %).

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von $0,722$ auch jeder Wert aus dem Intervall $[0,715; 0,728]$ als richtig anerkannt. Dieses Intervall enthält das Ergebnis $0,721$, das man bei Verwendung des in der Formelsammlung an einer Stelle (S. 41) fälschlich angegebenen Wertes $z_{0,95} = 1,6499$ erhielte.

Studierende, die anstelle von $\Phi(0,591)$ das Zwischenergebnis $0,591$ auf dem Lotsebogen bzw. einen Wert auf dem Intervall $[0,584; 0,593]$ eingetragen haben, erhielten 1 P.