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: **8. März 2010, 14.00 - 18.00 Uhr**

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

Aufgabe 1 (Aussagenlogik) (5 Punkte)

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

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

Bei den anderen 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, sind sowohl $P1$ als auch $P2$ erfüllt.

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

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

Lösung: B, C, E – vgl. zu dieser Aufgabe auch den Abschnitt 5.1.3 in Kurs 33210, insbesondere die letzte der dort wiedergegebenen beiden Wahrheitstabellen.

Kommentar: In den vier Tabellenzeilen sind die drei Werte in den letzten drei Spalten wie folgt zu ergänzen: Zeile 1: w, f, f; Zeile 2: w, f, f;Zeile 3: w, w, w; Zeile 4: f, w, f

Aufgabe 2 (Messen / Stichprobenverfahren) (5 Punkte)

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

A) Ein Gütekriterium für Messungen ist die Reliabilität (Zuverlässigkeit). Diese charakterisiert, inwieweit wirklich das gemessen wird, was gemessen werden soll.

B) Um theoretische Konstrukte (z. B. „Berufserfolg“ oder „Zufriedenheit“) messen zu können, muss man sie mit beobachtbaren Konstrukten verknüpfen, d. h. es gilt Handlungsanweisungen für die Gewinnung von Daten zu spezifizieren. Diesen Schritt bezeichnet man als Operationalisierung.

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

D) Die Klumpenauswahl ist eine zufallsgesteuerte Auswahlprozedur, bei der sich die Zufallsauswahl auf Teilmengen einer Grundgesamtheit bezieht, nicht auf die Untersuchungseinheiten selbst.

E) Bei einer proportional geschichteten Stichprobe wird eine Grundgesamtheit zunächst in Teilmengen zerlegt. Danach wird aus jeder Teilmenge ein fester Anteil von Untersuchungseinheiten zufällig ausgewählt.

Lösung: B, C, D, E.

Zu A: Die Validität (Gültigkeit) – nicht die Reliabilität (Zuverlässigkeit) – ist das Gütekriterium für Messungen, mit der 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 2.2 oder Schnell / Hill /Esser, Abschnitt 4.2 oder Kurs 33209, Abschnitt 2.3.

zu C: vgl. Kromrey, Abschnitt 6.4.3 oder Kurs 33209, Abschnitt 3.2.

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

zu E: Kromrey, Abschnitt 6.5.2 oder Kurs 33209, Abschnitt 3.2.

Aufgabe 3 (Datenerhebung) (5 Punkte)

Welche der folgenden Aussagen sind richtig? (Bei Aussage C geht es nur darum, den Wahrheitsgehalt des letzten Satzes zu bewerten.) (x aus 5)

A) Die Auswertung digitaler Verhaltensspuren, z. B. Logfileanalysen, repräsentiert ein nicht-reaktives Verfahren der Datengewinnung.

B) Das problemzentrierte und auch das narrative Interview zählen zu den qualitativen Methoden der Befragung.

C) In einer stichprobenbasierten Untersuchung sollen politische Einstellungen von Schülern an Hauptschulen in der Bundesrepublik Deutschland anhand eines standardisierten Interviews erhoben werden. Die Stichprobe wird dadurch erzeugt, dass aus allen Hauptschulen in Deutschland nach einem Zufallsverfahren 120 Schulen ausgewählt werden, in denen dann alle Schüler befragt werden. Bei dieser Vorgehensweise sind die Schüler die Auswahleinheiten.

D) Wenn man Umfragen mit freiwilligen Teilnehmern durchführt, ist mit Ergebnissen zu rechnen, die systematisch verzerrt sind, also für die zu untersuchende Grundgesamtheit nicht repräsentativ sind.

E) Overcoverage 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.

Lösung: A, B, D.

zu A: Diekmann, Kapitel XIII, Abschnitt 3, oder Kurs 33209, Abschnitt 3.1.

Zu B: Diekmann, Kapitel X, Abschnitt 2.

zu C: Auswahleinheiten sind hier die Schulen - vgl. Kromrey, Abschnitt 6.1.

zu D: Diekmann, Kapitel X, Abschnitt 4.

zu E: Der beschriebene Fehler heißt Undercoverage; vgl. auch Diekmann, Kapitel X, Abschnitt 11, oder Kurs 33209, Abschnitt 3.2.

Aufgabe 4 (Nominal- und Realdefinition; Versuchsanordnungen) (5 Punkte)

Welche der folgenden Aussagen 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 Nominaldefinition hat stets einen empirischen Informationsgehalt.

C) Eine Nominaldefinition kann falsch sein.

D) Eine Realdefinition kann falsch oder unvollständig sein.

E) Bei einem Quasi-Experiment mit Personen erfolgt die Zuordnung der Teilnehmer zu einer Versuchs- und einer Kontrollgruppe nicht auf der Basis einer Zufallsauswahl.

Lösung: A, D, E.

zu A: Kromrey, Abschnitt 3.5.1 oder Kurs 33210,Kapitel 3.

Zu B - C: Kromrey, Abschnitt 3.5.3;

zu D: Kromrey, Abschnitt 3.5.4;

zu E: Kromrey, Abschnitt 2.4.3 oder Kurs 33209, Abschnitt 3.1

Aufgabe 5 (Merkmalsklassifikationen) (5 Punkte)

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

A) Das Gewicht einer Person ist ein metrisch skaliertes Merkmal.

B) Das Gewicht einer Person ist ein stetiges Merkmal.

C) Das Merkmal „Bildungsstand“ (mit den Ausprägungen „ohne Schulabschluss“, „Hauptschulabschluss“, „mittlere Reife“, „Abitur“, „Hochschulabschluss“) ist nominalskaliert.

D) Das Merkmal „Bildungsstand“ (mit den unter C genannten Ausprägungen) ist ein diskretes Merkmal.

E) Qualitative Merkmale sind stets nominalskaliert.

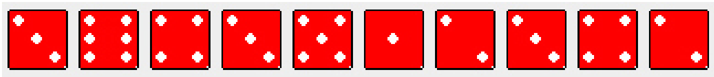
Lösung: A, B, D.

Zu C: Das Merkmal „Bildungsstand“ ist ordinalskaliert, weil sich die Ausprägungen des Merkmals in eine Rangfolge bringen lassen – vgl. Schnell / Hill / Esser (Abschnitt 4.3.1) oder Kurs 33209 (Abschnitt 2.2).

Zu E: Qualitative Merkmale können sowohl nominal- als auch ordinalskaliert sein – vgl. Kurs 33209 Ende von Abschnitt 2.2).

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

Mit einem Würfel wurde 10-mal gewürfelt. Dabei wurden die folgenden Ausprägungen für das Merkmal „Augenzahl" beobachtet, die einen Datensatz des Umfangs $n = 10$ definieren:



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

A) Die relative Häufigkeit für die Augenzahl 4 beträgt $0,2$.

B) Median und Mittelwert des Datensatzes stimmen hier überein.

C) Die Spannweite des Datensatzes hat den Wert 6.

D) Wenn noch einmal gewürfelt wird und dabei die Augenzahl 6 erzielt wird, bleibt der Wert des Medians unverändert.

E) Der Erwartungswert für die Zufallsvariable „Augenzahl" beträgt $3,5$, wenn man einen fairen Würfel voraussetzt, d. h. einen Würfel, bei dem jede Augenzahl mit derselben Wahrscheinlichkeit auftritt.

Lösung: A, D, E

Zu A: Durch die Grafik ist der Datensatz $1, 2, 2, 3, 3, 3, 4, 4, 5, 6$ definiert. Die absolute Häufigkeit für die Augenzahl 4 beträgt $h_4 = 2$; die relative Häufigkeit ist dann $f_4 = \frac{h_4}{10} = 0,2$.

Zu B: Der Median ist hier ($n = 10$) das arithmetische Mittel aus dem fünften und sechsten Wert des obigen geordneten Datensatzes, die beide den Wert 3 haben. Der Median hat also den Wert $\widetilde{x} = 3$. Für den Mittelwert errechnet man aus den 10 Werten $3,3$.

Zu C: Die Spannweite ist durch $6 - 1 = 5$ gegeben.

Zu D: Wenn an den obigen geordneten Datensatz noch das Element 6 angehängt wird, ist der Median der sechste Wert des erweiterten Datensatzes. Dieser ist 3, d. h. es gilt weiterhin $\widetilde{x} = 3$.

Zu E: Der Erwartungswert der diskreten Zufallsvariablen „Augenzahl" ist durch die Gleichung (11.6) in Kurs 33209 gegeben, wobei dort $k = 6$ und $p_1 = ... = p_6 = \frac{1}{6}$ ist und $x_1 = 1, ..., x_6 = 6$. Man erhält $E(X) = \frac{21}{6} = 3,5$.

Aufgabe 7 (Univariate Häufigkeitsverteilungen / Kenngrößen) (5 Punkte)

Gegeben sei der folgende Datensatz mit 15 Werten für ein Merkmal X:

$1,8$	$2,4$	$3,3$	$3,9$	$4,2$	$4,2$	$4,6$	$5,0$
$5,4$	$5,8$	$6,1$	$6,1$	$6,4$	$6,5$	$7,6.$	

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

A) Der Datensatz hat einen eindeutig bestimmten Modalwert.

B) Die Bestimmung des Modalwertes für einen Datensatz für ein Merkmal X ist auch möglich, wenn das Merkmal nur nominalskaliert ist.

C) Wenn man bei obigem Datensatz in der oberen Datenzeile den letzten Wert $(5,0)$ um $0,2$ erhöhte und gleichzeitig in der zweiten Datenzeile den ersten Wert $(5,4)$ um $0,4$ verminderte, resultierte für den Datensatz ein größerer Median $\widetilde{x}$, aber dafür ein kleinerer Mittelwert $\overline{x}$.

D) Wenn man den obigen Datensatz anhand eines Boxplots visualisierte (Rückführung des Datensatzes auf nur fünf Charakteristika), hätte die Gesamtlänge des Boxplots den Wert $5,8$.

E) Wenn man alle Werte des obigen Datensatzes verdoppelte, würde sich auch die empirische Varianz s^2 des Datensatzes verdoppeln.

Lösung: B, C, D.

Zu A: Der Datensatz hat zwei Modalwerte, nämlich $4,2$ und $6,1$.

Zu C: Der Median ist hier ($n = 15$) der achte Wert des nach Größe geordneten Datensatzes, d. h. es ist also $\widetilde{x} = 5,0$.

Wenn man den Wert von $5,0$ auf $5,2$ vergrößert und gleichzeitig den Wert $5,4$ auf $5,0$ absenkt und den neuen Datensatz wieder nach Größe ordnet, hat der achte Wert des neuen geordneten Datensatzes weiterhin den Wert $5,0$, d. h. der Median bleibt unverändert. Der Mittelwert, der aus der Merkmalssumme nach Division durch n hervorgeht, verkleinert sich dagegen. Dies gilt, weil sich die Summe der Merkmalswerte verkleinert, wenn man ein Element um $0,2$ vergrößert und ein anderes Element um $0,4$ verkleinert. Die Gesamtaussage, dass sich der Median vergrößert *und* der Mittelwert verkleinert, ist also nicht korrekt.

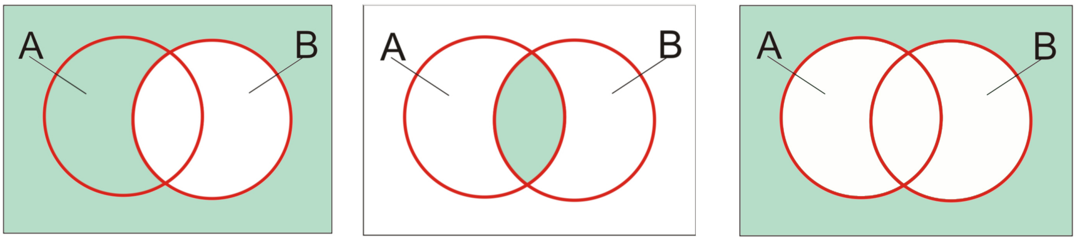
Zu D: Die Gesamtlänge des Boxplots entspricht der Spannweite des Datensatzes (vgl. Abbildung 5.3 im Kurs 33209).

Zu E: Bei Verdoppelung aller Merkmalswerte vervierfacht sich die empirische Varianz (die empirische Standardabweichung verdoppelt sich).

Aufgabe 8 (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 von A und B. Zwei Mengen, deren Darstellungen in Venn-Diagrammen sich nicht überschneiden, werden als disjunkt bezeichnet.

Nachstehend sind drei Venn-Diagramme abgebildet, die sich auf die Verknüpfung zweier – hier durch Kreise dargestellten – Ereignisse oder Mengen A und B beziehen.



Welche der folgenden Aussagen, die sich z. T. 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 $\overline{B}$ von B.

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

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

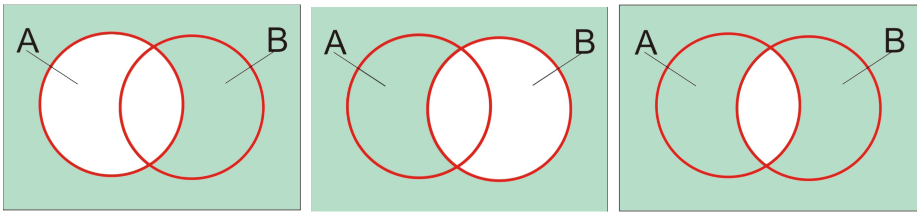
D) Die Vereinigung der beiden Schnittmengen $A \cap B$ und $\overline{A} \cap B$ liefert B.

E) Die beiden Schnittmengen $A \cap B$ und $A \cap \overline{B}$ sind disjunkt.

Lösung: A, D, E – vgl. hierzu auch Aufgabe 10.1 in Kurs 33209.

Zu B: Die dunkler gefärbte Fläche stellt die Schnittmenge von A und B dar, also $A \cap B$.

Zu C: Die Lösung lässt sich leichter nachvollziehen, wenn man zunächst je ein Venn-Diagramm für $\overline{A}$ und $\overline{B}$ und erst danach die Vereinigungsmenge $\overline{A} \cup \overline{B}$ von $\overline{A}$ und $\overline{B}$ zeichnet. Man erhält dann die folgenden drei Diagramme:



Das letzte Venn-Diagramm stimmt offenbar nicht mit dem dritten Venn-Diagramm aus der Aufgabe überein.

Anmerkung:

Teil C der Aufgabe 8 ist identisch mit Teil C der Zusatzaufgabe 7, die im Februar 2010 in Moodle eingestellt wurde.

Aufgabe 9 (Kontingenztafeln; Randverteilungen) (5 Punkte)

In der nachstehenden Kontingenztabelle für absolute Häufigkeiten sind Ergebnisse der Befragung einer Wählerstichprobe wiedergegeben (Daten des ZDF-Politbarometers vom 16. Oktober 2009). Die Häufigkeiten beziehen sich auf $n = 1021$ befragte Personen, die ihre Parteipräferenz für den 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". 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 beiden 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	**Randverteilung von X**
Ausprägungen von X	CDU/CSU	a_1	179	204	383
	SPD	a_2	100	117	217
	FDP	a_3	80	59	139
	DIE LINKE.	a_4	67	50	117
		a_5	54	62	116
	Sonstige	a_6	21	28	49
			501	520	1021
			Randverteilung von Y		

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

A) Der Tabelle entnimmt man, dass von den 1021 Befragten der Stichprobe 501 Männer waren.

B) Von den befragten Frauen präferierten mehr als ca. $8,0$ %, aber weniger als $9,0$ % die Linken.

C) Wenn man die in der obigen Tabelle wiedergegebenen absoluten Häufigkeiten und Randhäufigkeiten durch n dividiert, hier also durch 1021, resultieren relative Häufigkeiten bzw. relative Randhäufigkeiten.

D) Aus den Werten in der obigen Tabelle (Kontingenztabelle mit Randverteilungen) lässt sich u. a. ableiten, dass von den Befragten, die sich für die SPD entschieden hatten (Bedingung $X = a_2$), weniger als 52,0 % Frauen waren.

E) Ferner lässt sich aus der obigen Tabelle ableiten, dass von den an der Befragung beteiligten Personen, die männlichen Geschlechts waren (Bedingung $Y = b_1$), mehr als $34,0$ %, aber weniger als $37,0$ % die CDU/CSU favorisierten.

Lösung: A, C, E - vgl. auch die Beispiele 8.1 und 8.2 in Kurs 33209.

Zu A: Der Wert 501 ist das erste Element der Randverteilung des Merkmals Y.

Zu B: Es ist $h_{42} = 50$ und $f_{42} = \frac{50}{520} \approx 0,096$. Dies entspricht $9,6$ %.

Zu D: Gesucht ist die bedingte relative Häufigkeit $f_Y(b_2|a_2)$. Man erhält $f_Y(b_2|a_2) = \frac{117}{217} \approx 0,539$. Dies entspricht $53,9$ %.

Zu E: Gesucht ist hier die bedingte relative Häufigkeit $f_X(a_1|b_1)$ als $f_X(a_1|b_1) = \frac{179}{501} \approx 0,357$. Dies entspricht $35,7$ %.

Aufgabe 10 (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	6	3,0
2	5	3,2
3	7	2,5
4	7	2,3
5	8	2,0
6	9	2,0

Aus diesen Daten errechnet man u. a.

$$\sum_{i=1}^{6}(x_i - \overline{x})^2 = 10; \qquad \sum_{i=1}^{6}(y_i - \overline{y})^2 = 1,28; \qquad \sum_{i=1}^{6}(x_i - \overline{x})(y_i - \overline{y}) = -3,4.$$

Diese Werte sind hier ungeprüft zu übernehmen.

Welche der folgenden Aussagen sind richtig? (Bei den Aussagen A, D und E geht es nur 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 $\widehat{y} = \widehat{\alpha} + \widehat{\beta}x$ verläuft durch den Punkt $(7,0; 2,5)$.

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

C) Für den in der Tabelle wiedergegebenen Datensatz errechnet man für den Korrelationskoeffizienten nach Bravais-Pearson einen Wert zwischen $-0,9$ 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) Es seien wieder sechs Datenpaare $(x_1; y_1), ..., (x_6; y_6)$ gegeben, wobei $(x_1; y_1) = (6,0; 3,0)$ aus der vorstehenden Tabelle stammt und $(x_2; y_2), ..., (x_6; y_6)$ andere, hier nicht wiedergegebene Datenpaare seien. Auf der Basis dieser sechs Datenpaare seien für die Koeffizienten α und β des linearen Regressionsmodells nach der Methode der kleinsten Quadrate die Schätzungen $\widehat{\beta} = 0,38$ und $\widehat{\alpha} = 0,46$ bestimmt worden. Für das Residuum $\widehat{u}_1 = y_1 - \widehat{y}_1$ errechnet sich dann ein Wert, der zwischen 0 und $0,5$ liegt.

Lösung: A,B,E.

Zu A - B: Es gilt nach (16.6) und (16.7) (S. 30 unten der Formelsammlung) bei Beachtung von $\overline{x} = 7$ und $\overline{y} = 2,5$

$$\widehat{\beta} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{-3,4}{10} = -0,34.$$
$$\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} = 2,5 - (-0,34) \cdot 7 = 4,88.$$

Setzt man in die berechnete KQ-Regressionsgerade $\widehat{y} = 4,88 - 0,34 \cdot x$ für x den Wert $x = 7$ ein, folgt $\widehat{y} = 4,88 - 2,38 = 2,5$, d. h. die Gerade geht durch den Punkt $(7,0; 2,5)$. Letzteres ist nicht überraschend, weil der Punkt $(7,0; 2,5)$ gerade der Schwerpunkt $(\overline{x}; \overline{y})$ des Datensatzes ist und jede KQ-Gerade durch $(\overline{x}; \overline{y})$ geht.

Setzt man in die berechnete KQ-Regressionsgerade für x den Wert $x = 0$ ein, folgt $\widehat{y} = 4,88$, also der für $\widehat{\alpha}$ berechnete Wert.

Zu C: Nach (9.11) und (9.12) im Kurs (S. 9 der Formelsammlung) gilt für den Korrelationskoeffizienten r nach Bravais-Pearson

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{-3,4}{\sqrt{10} \cdot \sqrt{1,28}} = \frac{-3,4}{\sqrt{12,8}} \approx -0,95.$$

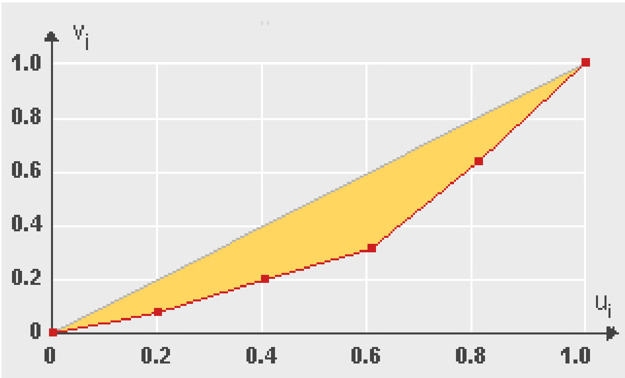
Zu D: Es gilt $R^2 = r^2$. Der Korrelationskoeffizient r kann zudem, anders als das Bestimmtheitsmaß, auch negative Werte annehmen.

Zu E:

$$\widehat{u}_1 = y_1 - \widehat{y}_1 = y_1 - (\widehat{\alpha} + \widehat{\beta} \cdot x_1) = 3,0 - (0,46 + 0,38 \cdot 6) = 0,26.$$

Aufgabe 11 (Konzentrationsmessung) (5 Punkte)

Im Gebiet der EU-27 gibt es fünf Hersteller von Windkraftwerken, die auf dem europäischen Markt miteinander konkurrieren. Es seien $x_1 = 20$, $x_2 = 30$, $x_3 = 30$, $x_4 = 80$ und $x_5 = 90$ die Umsätze dieser Firmen im letzten Geschäftsjahr (Umsätze jeweils in Millionen Euro). Die nachstehende Skizze 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_3 des Stützpunkts $(0,6; v_3)$ der Lorenzkurve hat den Wert $v_3 = 0,35$.

B) Der Wert, den man für den Ordinatenwert v_2 errechnet, gibt an, welcher Anteil des Gesamtumsatzes aller 5 Hersteller von Windkraftanlagen auf die beiden umsatzstärksten Unternehmen entfällt.

C) Der Gini-Koeffizient liefert Ausagen des Typs „x % der Merkmalsträger teilen sich y % der Merkmalssumme“.

D) Der Wert des unnormierten Gini-Koeffizienten G, der – anders als der normierte Gini-Koeffizient G^*– eine vom Umfang n des Datensatzes abhängige Obergrenze hat, ist genau halb so groß wie der Inhalt der Fläche, die in der obigen Abbildung betont ist (Fläche zwischen Lorenzkurve und und der Strecke, die den Nullpunkt mit dem Punkt $(1; 1)$ verbindet).

E) Der normierte Gini-Koeffizient G^* hat bei obigem Datensatz einen Wert, der zwischen $0,37$ und $0,39$ liegt.

Lösung: C, E.

Zu A: Es ist $v_3 = \frac{80}{250} = 0,32$.

Zu B: Der Wert v_2 gibt an, welcher Anteil des Gesamtumsatzes auf die beiden umsatzsschwächsten Unternehmen entfällt.

Zu D: G ist doppelt so groß wie der genannte Flächeninhalt.

Zu E: Der unnormierte Gini-Koeffizient errechnet sich nach Formel (6.5) in Kurs 33209 zu

$$G = \frac{2 \cdot 940}{5 \cdot 250} - \frac{5+1}{5} = \frac{188}{125} - \frac{150}{125} = \frac{38}{125} = 0,304.$$

Hieraus folgt dann für den normierten Gini-Koeffizienten G^*, wenn G_{max} die Obergrenze des unnormierten Gini-Koeffizienten G bezeichnet:

$$G^* = \frac{G}{G_{max}} = \frac{0,304}{0,8} = 0,38.$$

Aufgabe 12 (Kombinatorik / diskrete Verteilungen) (5 Punkte)

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

A) Wenn man eine „faire" Münze, also eine Münze mit gleichen Eintrittswahrscheinlichkeiten für „Kopf" und „Zahl", 5-mal wirft und die Anzahl X der Ausgänge mit „Zahl" feststellt, hat die Wahrscheinlichkeit dafür, *höchstens* zweimal „Zahl" zu erhalten, den Wert $0,5$.

B) Die Wahrscheinlichkeit dafür, bei dem $5-$maligem Münzwurf aus Aufgabenteil A *genau* dreimal „Zahl" zu erhalten, ist kleiner als $0,35$.

C) 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 10 ist, zwischen $0,10$ und $0,15$.

D) Aus einer Lostrommel, in der jedes vierte Los einen Gewinn repräsentiert, werden nacheinander 5 Lose *mit* Zurücklegen gezogen. Die Wahrscheinlichkeit dafür, nach den 5 Ziehungen keinen Gewinn gezogen zu haben, ist größer als $0,2$.

E) Mit einem fairen Würfel werde 12-mal in Folge gewürfelt. Bezeichne X die Anzahl der Ausgänge mit einer Augenzahl, die nicht größer als 4 ist. Der Erwartungswert von X hat den Wert 9.

Lösung: A, B, D.

Zu A: Die gesuchte Wahrscheinlichkeit ist durch den Wert der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 5$ und $p = 0,5$ an der Stelle $x = 2$ gegeben, nach Tabelle 19.1 also durch $F(2) = 0,5$.

Zu B: Die Wahrscheinlichkeit dafür, genau dreimal „Zahl" zu erhalten, ergibt sich als Differenz der Werte $F(3)$ und $F(2)$ der Verteilungsfunktion der Binomialverteilung mit $n = 5$ und $p = 0,5$, also nach Tabelle 19.1 als $F(3) - F(2) = 0,8125 - 0,5 = 0,3125$.

Zu C: Vgl. hierzu die letzten Zeilen von Beispiel 10.1 in Kurs 33209. Von den 36 Elementarereignissen, die den Ereignisraum Ω beim Würfeln mit zwei Würfeln definieren, führen nur 6 Elementarereignisse, nämlich die Ausgänge $(4;6)$, $(6;4)$; $(5;5)$, $(5;6)$, $(6;5)$ und $(6;6)$, zu einer Augensumme, die nicht kleiner als 10 ist. Die Wahrscheinlichkeit P für die Erzielung einer Augensumme, die mindestens 10 ist, hat somit den Wert $P = \frac{1}{6} \approx 0,167$.

Zu D: Die gesuchte Wahrscheinlichkeit ist hier durch den Wert $F(0)$ der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 5$ und $p = 0,25$ gegeben, nach Tabelle 19.1 also durch $0,2373$.

Zu E: Es ist der Erwartungswert $\mu = E(X)$ einer Binomialverteilung mit $n = 12$ und $p = \frac{2}{3}$ zu bestimmen. Der Parameter p bezeichnet die Wahrscheinlichkeit dafür, dass bei einem Wurf des Würfels keine der Augenzahlen 5 und 6 erscheint. Für μ errechnet man nach Formel (11.20) aus Kurs 33209 den Wert $\mu = 12 \cdot \frac{2}{3} = 8$.

Aufgabe 13 (Stetige Verteilungen) (5 Punkte)

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

A) Ist Z eine standardnormalverteilte Zufallsvariable , so ist für diese die Aussage $P(Z > a) = 0,05$, zutreffend, wenn für a der Wert $a = 1,96$ gewählt wird.

B) Es sei X eine stetige Zufallsvariable mit der Verteilungsfunktion $F(x)$. Es bezeichne $x_{0,05}$ das $0,05$-Quantil der Verteilung. Die Wahrscheinlichkeit dafür, dass X eine Ausprägung x mit $x > x_{0,05}$ annimmt, beträgt $0,95$.

C) Es sei X eine χ^2-verteilte Zufallsvariable. Wenn man die Dichtefunktion $f(x)$ grafisch darstellt und auf der x-Achse das $0,05$-Quantil $x_{0,05}$ und das $0,95$-Quantil $x_{0,95}$ der Verteilung markiert, so hat der vom Punkt $x_{0,05}$ bis zum Punkt $x_{0,95}$ gerechnete Flächeninhalt unter der Dichtekurve den Wert $0,90$ (Flächeninhalt zwischen Dichtekurve und x-Achse).

D) Die Dichtekurven von χ^2-verteilten und auch von t-verteilten Zufallsvariablen sind symmetrisch.

E) Eine mit 15 Freiheitsgraden t-verteilte Zufallsvariable nimmt mit Wahrscheinlichkeit $0,95$ eine Ausprägung an, die im Intervall $[-1,753; 1,753]$ liegt.

Lösung: B, C.

Zu A: Es ist $P(Z > 1,96) = 1 - \Phi(1,96) = 1 - 0,975 = 0,025$.

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

Zu D: Nur die Dichtekurven von t-verteilten Zufallsvariablen sind symmetrisch, nicht aber die von χ^2-verteilten Zufallsvariablen.

Zu E: Ein symmetrisches Intervall, in dem eine mit 15 Freiheitsgraden t-verteilte Zufallsvariable mit Wahrscheinlichkeit $0,95$ liegt, ist durch $[t_{15;0,025}; t_{15;0,975}] = [-t_{15;0,975}; t_{15;0,975}]$ gegeben. Das Quantil $t_{15;0,975}$ hat aber den Wert $2,131$, nicht $1,753$. Aussage E wäre korrekt gewesen, wenn die vorgegebene Wahrscheinlichkeit $0,90$ betragen hätte (anstelle von $0,95$).

Aufgabe 14 (Zufallsvariablen; Schätzung von Parametern) (5 Punkte)

Bei einem n Münzwürfe umfassenden Münzwurfexperiment lässt sich der Ausgang jedes einzelnen Münzwurfs (Bernoulli-Experiment) anhand einer Zufallsvariablen X_i mit den Ausprägungen 1 (= "Kopf")' und 0 (= „Zahl") beschreiben ($i = 1, 2, .., n$). Es bezeichne $X = \sum_{i=1}^{n} X_i$ die Anzahl der Ausgänge mit 'Kopf" und p die Eintrittswahrscheinlichkeit für das Ereignis „Kopf".

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

A) Die Variable X ist binomialverteilt mit Parametern n und p, also $X \sim B(n, p)$.

B) Die Varianz der Zufallsvariablen X hat den Wert $V(X) = \frac{n}{2}$, wenn man von einer fairen Münze ausgehen kann, also einer Münze mit gleichen Eintrittswahrscheinlichkeiten für „Kopf" und „Zahl".

C) Die Eintrittswahrscheinlichkeit p für „Kopf"" kann man anhand der Ausprägung des Stichprobenmittelwerts $\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$ erwartungstreu schätzen.

D) Die Varianz der Schätzfunktion aus Aufgabenteil C hat im Falle $n = 25$ den Wert $0,01$.

E) Zur Schätzung von p kann man anstelle einer Punktschätzung auch ein Konfidenzintervall heranziehen. Ein solches Konfidenzintervall muss den unbekannten Parameter p nicht notwendigerweise enthalten.

Lösung: A, C, D, E.

Zu A: Vgl hierzu (11.19) in Kurs 33209 und den anschließenden Text.

Zu B: Die Varianz von X hat im Falle $p = 0,5$ gemäß Formel (11.21) den Wert $V(X) = \frac{n}{4}$.

Zu C: Vgl hierzu Formel (14.10).

Zu D: Nach Formel (14.11) gilt $V(\widehat{p}) = \frac{p(1-p)}{n} = \frac{0,5 \cdot 0,5}{25} = 0,01$.

Zu E: Wenn man die Abbildungen 14.2 - 14.3 in Kurs 33209 betrachtet (dort die dunkel markierten Konfidenzintervalle), erkennt man, dass ein Konfidenzintervall nicht zwingend den zu schätzenden Parameter überdeckt.

Anmerkung:

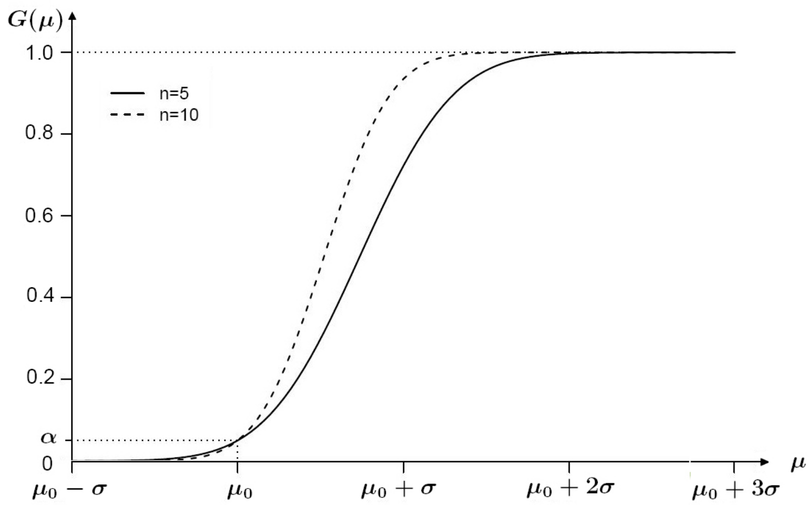
Bei der maschinellen Auswertung wurde Aufgabenteil D generell als richtig gewertet, weil sich der Wert $0,01$ nur ergibt, wenn $p = 0,5$ ist (fairer Münzwurf) und diese Information bei der Formulierung von Aufgabenteil D nicht explizit ausgewiesen war.

Aufgabe 15 (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

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

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

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

C) Der mit $n = 5$ durchgeführte Test besitzt im Vergleich zu dem Test mit $n = 10$ die größere Trennschärfe.

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

E) Die Werte, die beide Gütefunktionen für $\mu > \mu_0$ annehmen, repräsentieren Wahrscheinlichkeiten für den Eintritt eines Fehlers 2. Art.

Lösung: A, B, D.

Zu C: Die Trennschärfe ist für $n = 10$ höher.

Zu E: Jeder Wert $G(\mu)$ der Gütefunktion repräsentiert die Wahrscheinlichkeit für die Ablehnung der Nullhypothese, wenn μ der Parameterwert ist. Eine solche Ablehnwahrscheinlichkeit kann nicht die Wahrscheinlichkeit für den Eintritt eines Fehlers 2. Art sein, weil ein Fehler 2. Art ja beinhaltet, dass H_0 nicht abgelehnt wird, wenn eine Ablehnung die richtige Entscheidung darstellt. Bei dem dem hier betrachteten rechtsseitigen Test tritt ein Fehler 2. Art, wenn H_0 abgelehnt wird, obwohl $\mu > \mu_0$ gilt. Für $\mu > \mu_0$ repräsentieren also nicht die Werte $G(\mu)$, sondern die Werte $1 - G(\mu)$ Wahrscheinlichkeiten für den Eintritt eines Fehlers 2. Art.

Aufgabe 41 (Rangkorrelationskoeffizient) (3 Punkte)

Zwei Ratingagenturen A und B beurteilen unabhängig voneinander das kurzfristige Ausfallrisiko von Staatskrediten für fünf außereuropäische Länder. Die Risikobewertung wird anhand einer 8-stufigen Ratingskala vorgenommen. Die Stufen seien hier mit $1, ..., 8$ codiert, wobei die Punktzahl 8 die schlechteste und 1 die beste Bewertung darstellt (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	8
3	4	3
4	8	6
5	5	7

Untersuchen Sie anhand des Rangkorrelationskoeffizienten r_{SP} von Spearman, ob zwischen den Bewertungen der beiden Agenturen 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: $0,6$.

Herleitung: Wenn man den Beurteilungen der Agenturen A und B 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:

	Agentur A		Agentur B		
Land i	Bewertung x_i	$rg(x_i)$	Bewertung y_i	$rg(y_i)$	d_i
1	3	1	4	2	-1
2	7	4	8	5	-1
3	4	2	3	1	1
4	8	5	6	3	2
5	5	3	7	4	-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 Bewertung bei beiden Agenturen nicht mit der Mehrfachbelegung eines Rangplatzes verbunden ist. Mit (9.16) resultiert

$$r_{SP} = 1 - \frac{6 \cdot [(-1)^2 + (-1)^2 + 1^2 + 2^2 + (-1)^2]}{5 \cdot (25-1)} = 1 - \frac{6 \cdot 8}{5 \cdot 24} = 1 - \frac{2}{5} = 0,6.$$

Anmerkung:

Wenn bei der obigen Tabelle für einen der vier Rangplätze mit $d_i = 1$ oder $d_i = -1$ fälschlich der Wert $d_i = 2$ resp. $d_i = -2$ errechnet wurde, führte dies zum Ergebnis

$$r_{SP} = 1 - \frac{6 \cdot 11}{5 \cdot 24} = 1 - \frac{11}{20} = 0,45.$$

Für dieses (nicht zutreffende) Ergebnis wurde 1 Punkt vergeben.

Aufgabe 42 (Phi-Koeffizient) (4 Punkte)

Bei einer Klausur mit $n = 96$ Teilnehmern wurden für die Merkmale „Klausurerfolg X“ mit den Ausprägungen a_1 und a_2 (bestanden / nicht bestanden) und „Geschlecht Y“ mit den Ausprägungen b_1 und b_2 (männlich / weiblich) die absoluten Häufigkeiten erfasst. Die Ergebnisse sind in der nachstehenden Kontingenztafel zusammengefasst:

	b_1	b_2
a_1	36	30
a_2	12	18

Berechnen Sie den Phi-Koeffizienten auf der Basis der obigen (2 x 2)-Kontingenztabelle. Runden Sie das Ergebnis auf *drei Stellen nach dem Dezimalkomma* und tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Das **Dezimalkomma** belegt wieder ein **eigenes Feld**. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) $\Phi =$ | | | | | | |

Lösung : 0,135.

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

	b_1	b_2	Zeilensummen
a_1	36	30	66
a_2	12	18	30
Spaltensummen	48	48	96

Es folgt bei Anwendung von Formel (9.8) aus Kurs 33209 (S. 8 unten in der Formelsammlung)

$$\Phi = \frac{36 \cdot 18 - 30 \cdot 12}{\sqrt{66 \cdot 30 \cdot 48 \cdot 48}} = \frac{1}{\sqrt{55}} \approx 0,135.$$

Anmerkung:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,13; 0,14]$ als richtig anerkannt.

Aufgabe 43 (Kombinatorik) (3 Punkte)

Auf einer Betriebsfeier mit 15 Teilnehmern werden 2 Theaterkarten verlost. Die Verlosung ist so organisiert, dass in einen Schuhkarton 15 Lose gegeben werden, die mit $1, 2, .., 15$ nummeriert sind. Jeder Teilnehmer ist durch genau eine der Nummern repräsentiert. Aus dem Karton werden dann nacheinander 2 Lose gezogen. Um auszuschließen, dass jemand zweimal gewinnt, wird die erste gezogene Nummer vor dem Ziehen der zweiten Nummer nicht zurückgelegt.

Wieviele Möglichkeiten der Zufallsauswahl gibt es? Gehen Sie davon aus, dass die Theaterkarten gleichwertig sind, die Reihenfolge, in der die 2 Personen bestimmt werden, also unerheblich ist.

Tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch)

Lösung : 105

Herleitung: In der Terminologie des Urnenmodells werden $n = 2$ Kugeln aus einer Urne mit $N = 15$ 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

$$\binom{15}{2} = \frac{15!}{13! \cdot 2!} = \frac{15 \cdot 14}{2 \cdot 1} = 105.$$

Anmerkung:

Wenn man bei dieser Aufgabe anstelle der Formel (10.9) hier fälschlich Formel (10.7) anwendet, errechnet man den Wert 210. Für dieses (nicht zutreffende) Ergebnis wurde 1 Punkt gewährt.

Aufgabe 44 (Erwartungswert beim Lottospiel „6 aus 49“) (3 Punkte)

Es bezeichne X die Anzahl der einstelligen Lottozahlen (Zahlen $1, 2, ..., 9$) , die bei einer Lottoziehung gezogen werden (deutsches Lotto „6 aus 49“, ohne Zusatzzahl). Berechnen Sie den Erwartungswert $\mu = E(X)$ der Zufallsvariablen X.

Geben Sie Ihre Lösung auf *drei Nachkommastellen* genau rechtsbündig in das Antwortfeld ein. Verwenden Sie für das **Dezimalkomma** unbedingt ein **eigenes Feld**. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch) $\mu =$ □□□□□□

Lösung : $1,102$.

Herleitung: Da bei einer Lottoziehung stets *ohne Zurücklegen* gezogen wird, ist X hypergeometrisch verteilt mit den Parametern $n = 6$ (Anzahl der Ziehungen von jeweils einer Kugel), $M = 9$ (Anzahl der Kugeln mit einstelligen Zahlen in der Lostrommel) und $N = 49$ (Gesamtzahl der Kugeln in der Lostrommel): $X \sim H(6; 9; 49)$. Für den Erwartungswert $\mu = E(X)$ gilt nach (11.24)

$$E(X) = 6 \cdot \frac{9}{49} = \frac{54}{49} \approx 1,102.$$

Anmerkung: Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[1,05; 1,15]$ als richtig anerkannt.

Aufgabe 45 (KQ-Schätzung) (3 Punkte)

Für einen hier nicht wiedergegebenen Datensatz $(x_1, y_1), (x_2, y_2), ..., (x_{10}, y_{10})$ wurde $\overline{x} = -8,64$ und $\overline{y} = 5,32$ errechnet. Zu den 10 Datenpunkten wird nach der Kleinst-Quadrat-Methode auch eine Regressionsgerade $\widehat{y} = \widehat{\alpha} + \widehat{\beta}x$ bestimmt. Welcher Wert resultiert für $\widehat{\alpha}$, wenn sich für $\widehat{\beta}$ die Schätzung $\widehat{\beta} = 0,25$ ergeben hat?

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) $\widehat{\alpha} =$ □□□□□□

Lösung: $7,48$

Herleitung:

Die gesuchte KQ-Schätzung $\widehat{\alpha}$ errechnet sich nach (16.7) (vgl. Formelsammlung, S. 30 unten)

$$\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} = 5,32 - 0,25 \cdot (-8,64) = 5,32 + 2,16 = 7,48.$$

Anmerkungen:

Wenn am Ende der letzten Gleichung der Wert $2,16$ versehentlich subtrahiert statt addiert wurde, also

$$\widehat{\alpha} = 5,32 - 2,16 = 3,16$$

gerechnet wurde, gab es 2 Punkte.

Anstelle von $7,48$ resp. $3,16$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[1,05; 1,15]$ als richtig anerkannt (3 P.) resp. jeder Wert aus dem Intervall $[3,15; 3,17]$ (2 P.).

Aufgabe 46 (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

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

bezog. Es sei $n = 25$, $\mu_0 = 1000$ 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 $\mu = 1002,5$ 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. „0,427“ errechnen, tragen Sie in die letzten fünf Felder „0,427“ ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch)

Lösung: $0,802$

Herleitung: Die Gütefunktion des hier betrachteten rechtsseitigen Gauß-Tests ist (s. Formelsammlung, S. 28 oben) durch

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

gegeben, also durch

$$G(\mu) = 1 - \Phi\left(z_{0,95} - 0,5 \cdot 5\right) \approx 1 - \Phi\left(-0,8501\right) = \Phi\left(0,8501\right) \approx 0,802.$$

Zumindest die Größenordnung der Lösung kann man bereits der Grafik zu Aufgabe 15 entnehmen. Diese zeigte die Gütefunktionen des Tests im Falle $n = 5$ und $n = 10$. Für $n = 10$ entnimmt der Grafik, dass $G(\mu)$ an der Stelle $\mu = \mu_0 + 0,5 \cdot \sigma$ knapp unter $0,5$ liegt. Da die Gütefunktion im Falle $n = 25$ deutlich höher verläuft (höhere Trennschärfe des Tests), kann man schon erkennen, dass der gesuchte Wert deutlich oberhalb von $0,5$ liegen muss.

Anmerkungen:

Wenn anstelle von $0,802$ der Wert $1 - 0,802 = 0,198$ als Ergebnis eingetragen wurde, wurde 1 P. gegeben. Bei der maschinellen Auswertung wurde anstelle von $0,802$ auch jeder Wert aus dem Intervall $[0,795; 0,810]$ anerkannt (3 P.) bzw. jeder Wert aus dem Intervall $[0,195; 0,200]$ (1 P.).

Aufgabe 47 (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 $\sigma = 10$.

Wie groß ist bei Gültigkeit dieser Annahme die Wahrscheinlichkeit dafür, dass die Ausprägung des Merkmals für eine zufällig ausgewählte Person größer als 105 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. „0,4256“ errechnen, tragen Sie in die letzten sechs Felder „0,4256“ ein. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

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

Lösung: $0,3085$

Herleitung: Nach Formel (12.22) in Kurs 33209 gilt

$$P(X > 105) = 1 - P(X \leq 105) = 1 - \Phi\left(\frac{105 - 100}{10}\right) = 1 - \Phi(0,5) \approx 1 - 0,6915 = 0,3085.$$

Die Wahrscheinlichkeit dafür, dass X größer als 105 ist, beträgt also $0,3085$ (entspricht $30,85$ %).

Anmerkungen:

Wenn $0,6915$ anstelle von $1 - 0,6915 = 0,3085$ als Lösung eingetragen wurde, gab es 1 Punkt.

Anstelle von $0,3085$ resp. $0,6915$ wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,305; 0,312]$ (3 P.) resp. jeder Wert aus dem Intervall $[0,690; 0,693]$ (1 P.) als richtig anerkannt.

Aufgabe 48 (Wahrscheinlichkeiten bei Normalverteilung) (3 Punkte)

In der vorigen Aufgabe wurde eine Normalverteilung mit Erwartungswert $\mu = 100$ und Standardabweichung $\sigma = 10$ zur Modellierung einer latenten Variablen herangezogen. Wie groß ist das – gelegentlich auch als Dezil D9 – bezeichnete Quantil $x_{0,9}$ einer Zufallsvariablen X, die dieser Normalverteilung folgt, für die also $X \sim N(100; 10^2)$ gilt?

Tragen Sie Ihr Ergebnis rechtsbündig und auf *drei Stellen nach dem Dezimalkomma* genau in das Antwortfeld ein. Tragen Sei Ihr Ergebnis so ein, dass das **Dezimalkomma** auch hier wieder ein **eigenes Feld** belegt. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

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

Lösung: $112,816$

Herleitung:

Eine $N(\mu; \sigma^2)$-verteilte Zufallsvariable X wird im Falle $\mu = 100$ und $\sigma = 10$ gemäß Formel (12.11) in Kurs 33209 durch die Lineartransformation $Z = \frac{X-100}{10}$ standardisiert. Der Zusammenhang gilt auch für die Ausprägungen x und z von X resp. Z und insbesondere für die Quantile. Für $x_{0,9}$ gilt also $z_{0,9} = \frac{x_{0,9}-100}{10}$. Mit $z_{0,9} \approx 1,2816$ nach Tabelle 19.2 hat man

$$x_{0,9} \approx 100 + z_{0,9} \cdot 10 = 100 + 10 \cdot 1,2816 = 112,816.$$

Anmerkung:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall von $[112,80; 112,83]$ als richtig anerkannt.