

Musterlösungen

zur Klausur zu Modul 2.1 im BA-Studiengang

„Politikwissenschaft, Verwaltungswissenschaft, Soziologie“

und zum Modul 3 im alten BA-Studiengang

„Soziologie“

Termin: 4. September 2012, 14.00 - 18.00 Uhr

Erstprüfer: apl. Prof. Dr. H.-J. Mittag

Zweitprüfer: Dr. H.-G. Sonnenberg

Rückfragen zur Klausur und zur Musterlösung
sind stets **an den Erstprüfer** zu richten.

Multiple-Choice-Aufgaben

Aufgabe 1 (Aussagenlogik)

(5 Punkte)

In der nachstehenden Wahrheitstabelle stehen oberhalb der ersten beiden Spalten im Tabellenkopf zwei Aussagen 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: $\neg a$	P2: $a \vee b$	K: $(\neg a) \wedge (a \vee b)$
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 $P2$ erfüllt, nicht aber $P1$.
- 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 zu Aufgabe 1:

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: f, w, f ; Zeile 2: f, w, f ; Zeile 3: w, w, w ; Zeile 4: w, f, f .

Aufgabe 2 (Merkmalsklassifikationen)

(5 Punkte)

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

- A) Auf einem Fragebogen für die Beschäftigten eines größeren Betriebs wird erhoben, mit welchem Transportmittel der Weg zur Arbeitsstätte zurückgelegt wird. Das Merkmal „Transportmittel“, für das auf dem Formular acht Ausprägungen (u. a. „Bus“, „Bahn“, „PKW“ sowie eine Restkategorie „Sonstiges Transportmittel“) unterschieden werden, ist ein diskretes Merkmal.
- B) Das Merkmal „Transportmittel“ aus Aufgabenteil A ist ein ordinalskaliertes Merkmal.
- C) Die Bildung von Differenzen ist nicht bei nominalskalierten, wohl aber bei ordinalskalierten Daten zulässig.
- D) Operationen, die für nominalskalierte Daten zulässig sind, sind ebenso für ordinalskalierte Daten zulässig.
- E) Ein qualitatives Merkmal ist stets nominalskaliert.

Anmerkung zu dieser Aufgabe:

Der Begriff „metrische Skala“ ist als Oberbegriff für „Intervallskala“ und „Verhältnisskala“ (letztere einschließlich des Sonderfalls „Absolutskala“) zu verstehen.

Lösung zu Aufgabe 2: A, D.

Zu A und B: vgl. Kurs 33209, Abschnitt 2.2 2.1.

Zu B: Das Merkmal „Transportmittel“ ist nominalskaliert, denn die Ausprägungen sind nur Kategorien, die sich nicht in eine natürliche Rangfolge bringen lassen. Vgl. auch Kromrey, Abschnitt 5.4.4 oder Schnell / Hill / Esser, Abschnitt 4.3.1 sowie Kurs 33209, Tabelle 2.1.

Zu C: Die Bildung von Differenzen ist weder bei nominalskalierten, noch bei ordinalskalierten Daten zulässig. Vgl. Kromrey, Abschnitt 5.4.4, oder Kurs 33209, Tabelle 2.1.

Zu D: Die Operationen, die für Daten eines bestimmten Skalenniveaus zulässig sind, sind stets auch auf Daten anwendbar, die auf einer „höheren“ Skala erfasst werden. Vgl. auch Kromrey, Abschnitt 5.4.4, oder Kurs 33209, Abschnitt 2.2.

Zu E: Qualitative Merkmale sind entweder nominalskaliert oder auch ordinalskaliert. Güteklassen bei Gemüse und Obst sind Beispiele für qualitative Merkmale, die ordinalskaliert sind.

Aufgabe 3 (Datengewinnung, Messung)

(5 Punkte)

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

- A) Aus der Validität einer Messung folgt stets auch deren Reliabilität.
- B) Die Validität charakterisiert, inwieweit ein Messinstrument bei wiederholter Messung die gleichen Messwerte liefert.
- C) Eine Grundgesamtheit von $N = 5.000$ (5 Tausend) Personen wird bezüglich eines sozioökonomischen Merkmals in vier Teilpopulationen zerlegt (Schichtung der Grundgesamtheit). Die Teilgesamtheiten umfassen $N_1 = 1.200$, $N_2 = 1.300$, $N_3 = 1.600$ und $N_4 = 900$ Personen. Aus den Teilpopulationen werden dann in der zweiten Verfahrensstufe Zufallsstichproben des Umfangs $n_1 = 36$, $n_2 = 39$, $n_3 = 56$ resp. $n_4 = 27$ gezogen. Das damit praktizierte Auswahlverfahren repräsentiert eine geschichtete Stichprobenauswahl mit proportionaler Schichtung.
- D) Die Klumpenauswahl ist eine zufallsgesteuerte Auswahlprozedur, bei der sich die Zufallsauswahl auf Teilmengen einer Grundgesamtheit bezieht, nicht auf die Untersuchungseinheiten selbst.
- 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 zu Aufgabe 3: A, D, E.

Zu A und B: Die Validität (Gültigkeit) ist das Gütekriterium für Messungen, mit dem beschrieben wird, inwieweit wirklich das gemessen wird, was gemessen werden soll. Aus der Validität einer Messung folgt auch deren Reliabilität. Letztere sagt aus, inwieweit ein Messinstrument bei wiederholter Messung die gleichen Messwerte liefert (Zuverlässigkeit im technischen Sinne). Eine Messung, die nicht-valide ist, kann durchaus reliabel sein (vgl. Kromrey, Ende von Abschnitt 5.7, oder Schnell / Hill / Esser, Anfang von Abschnitt 4.3.2.2 oder Kurs 33209, Abschnitt 2.3).

Zu C: Der in der zweiten Stufe angewandte Auswahlatz $\frac{n_i}{N_i}$ ($i = 1, \dots, 4$) ist hier nicht konstant – er liegt bei drei der Teilpopulationen bei 3 %, bei einer Schicht (der mit $N_3 = 1.600$) aber bei 3,5 % (vgl. auch Kromrey, Abschnitt 6.5.2 oder Kurs 33209, Abschnitt 3.2 – dort insbesondere Abbildung 3.4).

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

Zu E: Kromrey, Abschnitt 2.4.3 oder Kurs 33209, Abschnitt 3.1.

Aufgabe 4 (Nominal- und Realdefinitionen)

(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 ist entweder richtig oder falsch.
- C) Eine Nominaldefinition hat keinen empirischen Informationsgehalt.
- D) Realdefinitionen sind Worterklärungen, die – im Gegensatz zu Nominaldefinitionen – das Definiendum durch bestimmte Eigenschaften charakterisieren.
- E) Eine Realdefinition umfasst alle Eigenschaften des Definiendums.

Lösung zu Aufgabe 4: A, C, D.

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

Zu B: Nominaldefinitionen können niemals falsch sein – vgl. Kromrey, Abschnitt 3.5.3;

Zu C: Kromrey, Abschnitt 3.5.1.

Zu D: Kromrey, Abschnitt 3.5.4.

Zu E: Realdefinitionen sind niemals vollständig, sie heben vielmehr nur auf das Wesentliche des Definiendums ab (vgl. Kromrey, Abschnitt 3.5.4).

Aufgabe 5 (Univariate Häufigkeitsverteilungen / Kenngrößen)

(5 Punkte)

Gegeben sei der folgende Datensatz für ein stetiges Merkmal X :

5,5 6,2 5,9 4,5 4,2 3,9 4,0 3,0 6,5, 6,9 5,5 6,3.

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) Bei obigem Datensatz ist der Median \tilde{x} kleiner als der Mittelwert \bar{x} .
- C) Wenn man den Datensatz anhand eines Boxplots visualisiert, liegen der sechste bis zehnte Wert (also die Werte 3,9 bis 6,9) und der letzte Wert (6,3) der obigen ungeordneten Urliste außerhalb der Box.

- D) Wenn man den ersten Wert (5, 5) der oben wiedergegebenen Urliste um 0, 1 erhöht und den letzten Wert (6, 3) um 0, 1 senkt, bleiben sowohl der Median \tilde{x} als auch der Mittelwert \bar{x} des Datensatzes unverändert.
- E) Mit der in Aufgabenteil D spezifizierten Veränderung des ersten und letzten Wertes der Urliste – also der Werte 5, 5 und 6, 3 – ist auch eine Veränderung der Spannweite des Datensatzes verbunden.

Lösung zu Aufgabe 5: A, C.

Zu A: Der Wert 5, 5 (= Modalwert) tritt doppelt auf, alle anderen Werte nur einmal.

Zu B: Der Median des Datensatzes des Umfangs $n = 12$ ist $\tilde{x} = 5, 5$ (Mittelwert aus dem sechsten Wert 5, 5 und siebten Wert 5, 5 der nach aufsteigender Größe geordneten Urliste), während für den Mittelwert $\bar{x} = 5, 2$ gilt.

Zu C: Die Begrenzungen der Box sind durch die in den Interquartilsabstand eingehenden beiden Quartile definiert. Innerhalb der Box oder auf den Begrenzungslinien der Box liegen hier die „mittleren 50 %“ des nach Größe geordneten Datensatzes. Bei dem nach aufsteigender Größe geordneten Datensatz sind dies die Werte 4, 2, 4, 5, ..., 6, 2. Dies sind die ersten fünf und der vorletzte Wert des Ausgangsdatsatzes. Außerhalb der Box liegen also genau die in der Aufgabe genannten Werte.

Zu D: Wenn man den ersten Wert (5, 5) der Urliste um 0, 1 erhöht und den letzten Wert um 0, 1 senkt, bleibt die Merkmalssumme konstant und damit auch der Mittelwert. Der Median, der sich wieder ($n = 12$) als Mittelwert aus dem sechsten Wert 5, 5 und dem – nun auf 5, 6 erhöhten – siebten Wert bestimmt, steigt nun auf 5, 55.

Zu E: Da sich keiner der beiden Extremwerte des Datensatzes (3, 0 und 6, 9), aus denen sich die Spannweite errechnet, verändert, bleibt die Spannweite gleich.

Aufgabe 6 (absolute und relative Häufigkeiten; Vierfeldertafeln) (5 Punkte)

Ägypten hatte Anfang 2011 insgesamt 82, 0 Millionen Einwohner. Davon waren 50, 7 % männlichen Geschlechts. Anders als in Deutschland ist der Anteil der jungen Bevölkerung sehr hoch – Anfang 2011 betrug der Anteil der ägyptischen Bevölkerung unter 15 Jahren 32, 7 %. Innerhalb der Altersgruppe der unter 15-jährigen waren 51, 1 % männlichen Geschlechts.

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

- A) Die Anzahl der weiblichen Einwohner Ägyptens im Alter von 15 oder mehr Jahren (Altersklasse 15+) betrug unter 27, 0 Millionen.
- B) Die Anzahl der männlichen Einwohner Ägyptens im Alter von 0 – 14 Jahren lag über 12, 9 Millionen.

- C) Der Anteil der Ägypter männlichen Geschlechts im Alter von 0 – 14 Jahren an der Gesamtbevölkerung betrug weniger als 17,0 %.
- D) Die Gesamtzahl der weiblichen Einwohner Ägyptens lag unter 40,6 Millionen.
- E) Wenn man die Anzahl der Personen in der Altersklasse 0 – 14 und der Altersklasse 15+ jeweils für die Personen männlichen Geschlechts und danach auch für die Personen weiblichen Geschlechts addiert, erhält man die absolute Randverteilung für das binäre Merkmal „Geschlecht“.

Lösung zu Aufgabe 6: B, C, D, E - vgl. auch Beispiel 8.2 in Kurs 33209.

Die Angaben lassen sich anhand einer Vierfeldertafel für absolute Häufigkeiten (Angaben in Millionen) bzw. – nach Division aller Werte durch $n = 82,0$ – für relative Häufigkeiten darstellen.

	0 – 14	15+	Zeilensummen
männlich	13,702	27,872	41,574
weiblich	13,112	27,314	40,426
Spaltensummen	26,814	55,186	82,0

	0 – 14	15+	Zeilensummen
männlich	0,167	0,340	0,507
weiblich	0,160	0,333	0,493
Spaltensummen	0,327	0,673	1

Anmerkung: Wenn man mit den auf die Gesamtbevölkerung bezogenen Angaben „50,7 % männlich“ und „32,7 % jünger als 15 Jahre“ die Randverteilungen der beiden Merkmale bestimmt hat, ermöglicht die Zusatzinformation „Innerhalb der Altersgruppe der unter 15-jährigen waren 51,1 % männlichen Geschlechts“ die Bestimmung von h_{11} gemäß $h_{11} = 0,511 \cdot h_{.1} \approx 13,702$. Nach Division durch $n = 82,0$ erhält man dann die relative Häufigkeit $f_{11} \approx 0,167$ (Ägypter männlichen Geschlechts unter 15 Jahren).

Anhand der beiden Tafeln kann man dann den Wahrheitsgehalt der fünf Aussagen sofort bewerten:

Zu A: Die Aussage bezieht sich auf die absolute Häufigkeit $h_{21} = 27,314$, trifft also nicht zu (s. die erste der obigen Vierfeldertafeln).

Zu B: Die Aussage bezieht sich auf die absolute Häufigkeit $h_{12} = 13,112$, trifft also zu.

Zu C: Die Aussage bezieht sich auf die relative Häufigkeit $f_{11} = 0,167$ (s. die zweite der obigen Vierfeldertafeln). Dies entspricht 16,7 %. Die Aussage ist also zutreffend.

Zu D: Die Aussage bezieht sich auf die (in Millionen ausgewiesene) Randhäufigkeit $h_{2.} = h_{21} + h_{22}$, also auf $h_{2.} = 40,426$.

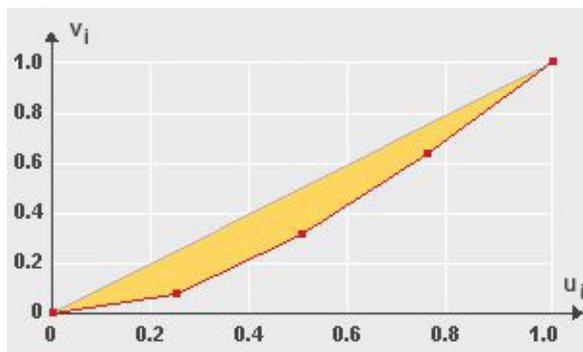
Zu E: Vgl. hierzu Tabelle 8.2 oder Tabelle 8.3 in Kurs 33209.

Aufgabe 7 (Konzentrationsmessung)

(5 Punkte)

In einer Region konkurrieren vier Energieversorgungsunternehmen E_1, E_2, E_3 und E_4 . Es seien $x_1 = 90, x_2 = 20, x_3 = 60$ und $x_4 = 80$ die Umsätze dieser Firmen im letzten Geschäftsjahr (Umsätze jeweils in Millionen Euro). Es hat also z. B. E_3 den Umsatz 60.

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
4	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 einen Wert, der im Intervall $[0,30; 0,31]$ liegt (Intervallgrenzen sind hier eingeschlossen).
- B) Der Wert, den man für den Ordinatenwert v_2 errechnet, gibt an, welcher Anteil des Gesamtumsatzes aller 4 Energieversorger auf die beiden umsatzschwächsten Unternehmen E_2 und E_3 entfällt.
- C) Allgemein gilt, dass der unnormierte Gini-Koeffizient G hier (Datensatz des Umfangs $n = 4$) einen Wert bis maximal 0,75 annehmen kann.
- D) Der normierte Gini-Koeffizient hat bei obigem Datensatz einen Wert G^* , der unterhalb von 0,3 liegt.
- E) Wenn ein weiteres Unternehmen E_5 im Energieversorgungsmarkt der Region aktiv wäre und zwar mit einem Umsatz $x_5 = 70$, würde man auf der Basis des vergrößerten Datensatzes ($n = 5$ anstelle von $n = 4$) für G^* einen Wert errechnen, der unterhalb von 0,24 liegt.

Lösung zu Aufgabe 7: B, C.

Zu A: Ordnet man die Umsatzwerte x_1, \dots, x_4 nach aufsteigender Größe und bezeichnet man die geordnete Folge der Umsatzwerte mit $x_{(1)}, \dots, x_{(4)}$, so gilt nach (6.3) aus Kurs 33209 die Gleichung $v_2 = \frac{80}{250} = 0,32$.

Zu C: Der unnormierte Gini-Koeffizient G ist im Falle $n = 4$ gemäß (6.6) durch $G_{max} = \frac{3}{4} = 0,75$ nach oben begrenzt.

Zu D: Für die gewichtete Merkmalssumme (6.4) errechnet man zunächst $q_4 = 740$ und damit für den unnormierten Gini-Koeffizienten mit (6.5) den Wert $G = 0,23$. Mit (6.7) folgt dann für den normierten Gini-Koeffizienten bei Rundung auf vier Dezimalstellen $G^* = \frac{G}{G_{max}} = \frac{0,23}{0,75} \approx 0,3067$.

Zu E: Für die ungewichtete und die gewichtete Merkmalssumme errechnet man nun $p_5 = 320$ bzw. $q_5 = 1120$. Für den den unnormierten Gini-Koeffizienten folgt $G = 0,2$ und für die normierte Variante G^* wegen der neuen Obergrenze $G_{max} = \frac{4}{5} = 0,8$ dann $G^* = \frac{0,2}{0,8} = 0,25$.

Anmerkung zur Lösung zu Teil D:

Anstelle von 0,3067 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,304; 0,0,309]$ als richtig anerkannt.

Aufgabe 8 (Zusammenhangsmessung)

(5 Punkte)

Welche der folgenden Aussagen sind richtig?

(x aus 5)

- A) Wenn man auf der Basis eines Datensatzes $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$ für zwei Merkmale X und Y für den Korrelationskoeffizienten r nach Bravais-Pearson den Wert 0 errechnet, beinhaltet dies, dass zwischen den beiden Merkmalen kein Zusammenhang vorliegt.
- B) Wenn $r = -1$ ist, bedeutet dies, dass die Datenpaare $(x_1, y_1), \dots, (x_n, y_n)$ alle auf einer fallenden Geraden liegen.
- C) Wenn für den Datensatz $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$ für X und Y ein Wert r errechnet wird, dessen Absolutbetrag nahe bei 1 liegt (z. B. zwischen 0,9 und 1,0) bedeutet dies, dass zwischen den beiden Merkmalen eine ausgeprägte sachlogische Verbindung besteht.
- D) Der Rangkorrelationskoeffizient r_{SP} lässt sich nicht auf nominalskalierte Merkmale anwenden.

- E) Quadriert man den Korrelationskoeffizient r nach Bravais-Pearson, erhält man ein Maß für die Anpassungsgüte einer Regressionsgeraden an einen Datensatz $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$.

Lösung zu Aufgabe 8: B, D, E.

Zu A: Es kann durchaus ein nicht-linearer Zusammenhang vorliegen; vgl. Kurs 33209, Abschnitt 9.2 (Abbildung 9.2).

Zu B: Vgl. Kurs 33209, Abschnitt 9.2.

Zu C: Vgl. Kurs 33209, Abschnitt 9.2. Im Text, der sich direkt an Beispiel 9.2 anschließt, wird explizit festgestellt, dass ein auf der Basis eines Datensatzes für zwei Merkmale X und Y errechneter hoher Absolutbetrag für r nicht zwingend das Vorliegen eines Kausalzusammenhangs zwischen X und Y impliziert, also nicht notwendigerweise eine sachlogische Verbindung zwischen den Variablen vorliegt.

Zu D: Vgl. Kurs 33209, Abschnitt 9.3 (Die Anwendung des für ordinalskalierte Daten konzipierten Zusammenhangsmaßes ist auch bei metrisch skalierten Daten möglich, allerdings unter Informationsverlust).

Zu E: Das Quadrat von r ist nach Formel (16.18) aus Kurs 33209 identisch mit dem Bestimmtheitsmaß R^2 . Letzteres ist ein Anpassungsgütemaß für eine Regressionsgerade an einen bivariaten Datensatz.

Aufgabe 9 (Venn-Diagramme; Kombinatorik)

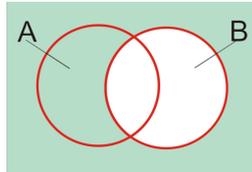
(5 Punkte)

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 \bar{A} und \bar{B} die Komplementärmenge von A und B , $A \cap B$ deren Schnittmenge und $A \cup B$ die Vereinigungsmenge von A und B .

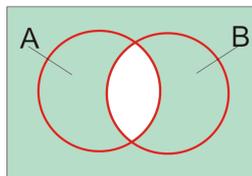
Von den nachstehenden fünf Aussagen beziehen sich nur die ersten drei auf Venn-Diagramme, die beiden anderen sind der Kombinatorik gewidmet.

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

- A) Das nachstehende Venn-Diagramm veranschaulicht anhand der dunkler gefärbten Fläche die Komplementärmenge \bar{B} von B .



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



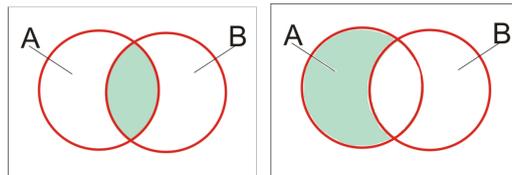
- C) Bezeichnet man die Menge $A \cap B$ (Schnittmenge aus A und B) mit C und die Menge $A \cap \bar{B}$ (Schnittmenge aus A und \bar{B}) mit D , so gilt $C \cap D = \emptyset$ (leere Menge), d. h. C und D sind disjunkt.
- D) Wenn man eine faire Münze, also eine Münze mit gleichen Eintrittswahrscheinlichkeiten für „Kopf“ und „Zahl“, 7-mal wirft und die Anzahl X der Ausgänge mit „Zahl“ feststellt, liegt die Wahrscheinlichkeit dafür, *mindestens* fünfmal „Zahl“ zu erhalten, oberhalb von 0,24.
- E) Die Wahrscheinlichkeit dafür, bei dem 7-maligem Münzwurf aus Aufgabenteil A *genau* fünfmal „Zahl“ zu erhalten, ist größer als 0,15.

Lösung zu Aufgabe 9: A, C, E

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

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

Zu C: Die beiden in Aufgabenteil C angesprochenen Mengen $C := A \cap B$ und $D := A \cap \bar{B}$ sind nachfolgend wiedergegeben:



Die Schnittmenge aus den in den beiden Venn-Diagrammen wiedergegebenen Mengen C und D ist offenbar die leere Menge, d. h. C und D sind in der Tat disjunkt.

Zu D: Die Wahrscheinlichkeit $P(X \geq 5)$ dafür, mindestens fünfmal „Zahl“ beim 7-maligen Münzwurf zu erhalten, ist die Komplementärwahrscheinlichkeit von $P(X \leq 4)$. Letztere ist durch den Wert der Verteilungsfunktion $F(x)$ der Binomialverteilung mit $n = 7$ und $p = 0,5$ an der Stelle $x = 4$ gegeben, nach Tabelle 19.1 also durch $F(4) = 0,7734$. Es gilt also $P(X \geq 5) = 1 - 0,7734 = 0,2266$.

Zu E: Die Wahrscheinlichkeit dafür, genau fünfmal „Zahl“ zu erhalten, ergibt sich als Differenz der Werte $F(5)$ und $F(4)$ der Verteilungsfunktion der Binomialverteilung mit $n = 7$ und $p = 0,5$, nach Tabelle 19.1 somit als $F(5) - F(4) = 0,9375 - 0,7734 = 0,1641$.

Aufgabe 10 (Randverteilungen, bedingte Wahrscheinlichkeiten) (5 Punkte)

Ein Unternehmen will vor der Markteinführung eines neuen Produkts zunächst zwei Varianten P_1 und P_2 des geplanten Produkts potenziellen Kunden vorstellen und beauftragt ein Marktforschungsinstitut damit, die Kundenpräferenzen bezüglich der beiden Produktvarianten zu ermitteln. Das Marktforschungsinstitut befragt insgesamt 196 Personen, von denen sich 112 für die Produktvariante P_1 entschieden. Von den befragten 196 Personen waren 108 Männer. Es präferierten 60 Männer die Produktvariante P_2 .

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

- A) Von den weiblichen Befragten favorisierten über 70 % die Produktvariante P_1 .
- B) Wählt man aus der Population aller befragten Personen eine Person zufällig aus, so liegt die Wahrscheinlichkeit, dass diese die Produktvariante P_2 bevorzugt, über 0,44.
- C) Wählt man aus der Population aller befragten Personen eine Person zufällig aus, so liegt die Wahrscheinlichkeit, dass eine Frau mit Präferenz für P_2 ausgewählt wird, oberhalb von 0,13 – in Prozentwerten ausgedrückt also oberhalb von 13 %.
- D) Von den befragten Personen mit Präferenz für die Produktvariante P_2 waren weniger als 30 % weiblich.
- E) Wenn man bei den befragten Frauen und bei den befragten Männern jeweils den Anteil der Personen mit Präferenz für die Produktvariante P_2 ermittelt, stellt man fest, dass der Anteil bei den Männern mehr als doppelt so groß ist.

Lösung zu Aufgabe 10: A, D, E.

Es ist zweckmäßig die im Text enthaltenen Informationen entweder anhand eines Baumdiagramms zu visualisieren (vgl. in Kurs 33209 z. B. die Abbildung 8.1 oder die Lösung zu Teil a von Aufgabe 10.6) oder aber sie in einer Kontingenztafel mit Randverteilungen zusammenzufassen (vgl. in Kurs 33209 die Tabelle 10.2 oder die Lösung zu Teil e von Aufgabe 10.5). Man erhält im letztgenannten Fall folgende Vierfeldertafel, bei der die Vorgaben dieser Aufgabe kursiv gesetzt sind (Codierung: Präferenz von Produktvariante $P_1 = A$, Präferenz von Produktvariante $P_2 = \bar{A}$, weiblich = B , männlich = \bar{B}):

	weiblich (B)	männlich (\bar{B})	Zeilensummen
Präferenz von P_1 (= A)	64	48	112
Präferenz von P_2 (= \bar{A})	24	60	84
Spaltensummen	88	108	196

Aus der Vierfeldertafel folgt dann:

Zu A: Von den an der Befragung beteiligten 88 Frauen bevorzugten 64 die Produktvariante P_1 , also ca. 72,7 %.

Zu B: Die Wahrscheinlichkeit $P(A)$ dafür, dass von den 196 befragten Personen per Zufallsauswahl eine der 84 Personen gewählt wird, die die Produktvariante P_2 bevorzugten, beträgt nach der Formel (10.5) aus Kurs 33209 $P(A) = \frac{84}{196} = 0,429$, also ca. 42,9 %.

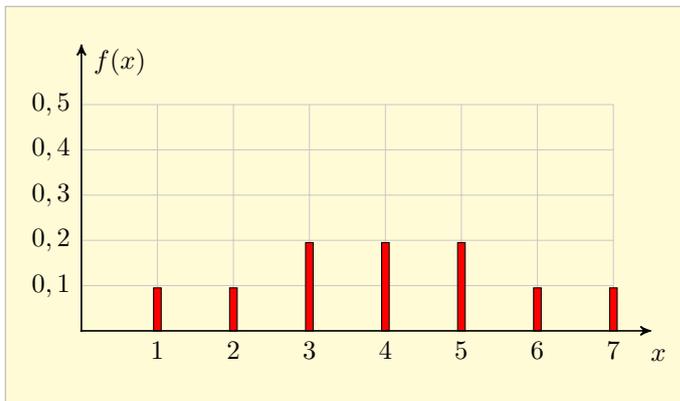
Zu C: Die Wahrscheinlichkeit $P(\bar{A} \cap B)$ dafür, dass per Zufallsauswahl eine der 24 Frauen gewählt wird, die P_2 bevorzugten, beträgt $P(\bar{A} \cap B) = \frac{24}{196} = 0,122$.

Zu D: von den befragten 84 Personen mit Präferenz für P_2 waren 24 Personen weiblichen Geschlechts, also ca. 28,6 %.

Zu E: Von den 88 weiblichen Befragten bevorzugten 24 Personen P_2 , also ca. 27,3 %, während bei den 108 männlichen Befragten 60 die Variante P_2 bevorzugten, also ca. 55,6 %. Der letztgenannte Prozentsatz ist im Vergleich zum ersten Prozentsatz mehr als doppelt so groß.

Aufgabe 11 (Wahrscheinlichkeits- und Verteilungsfunktion) (5 Punkte)

Die nachstehende Abbildung zeigt die Wahrscheinlichkeitsfunktion $f(x)$ einer diskreten Zufallsvariablen X , die sieben Ausprägungen $x_1 = 1, x_2 = 2, \dots, x_6 = 6, x_7 = 7$ aufweist. Die Ausprägungen x_1, x_2, x_6 und x_7 weisen jeweils die Eintrittswahrscheinlichkeit 0,1 auf, während x_3, x_4 und x_5 je mit Wahrscheinlichkeit 0,2 realisiert werden.



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

- A) Die Verteilungsfunktion $F(x)$ der diskreten Zufallsvariablen X nimmt für $x = 3, 0$ den Wert 0,4 an.
- B) Die Verteilungsfunktion $F(x)$ von X nimmt für $x = 3, 5$ den Wert 0,4 an.
- C) Die Verteilungsfunktion $F(x)$ von X ist nur bis $x = 7$ definiert.
- D) Der Erwartungswert $E(X)$ der Zufallsvariablen X hat den Wert 4,0.
- 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 zu Aufgabe 11: A, B, D.

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

$$F(x) = \begin{cases} 0 & \text{für } x < 1 \\ 0,1 & \text{für } 1 \leq x < 2 \\ 0,1 + 0,1 = 0,2 & \text{für } 2 \leq x < 3 \\ 0,2 + 0,2 = 0,4 & \text{für } 3 \leq x < 4 \\ 0,4 + 0,2 = 0,6 & \text{für } 4 \leq x < 5 \\ 0,6 + 0,2 = 0,8 & \text{für } 5 \leq x < 6 \\ 0,8 + 0,1 = 0,9 & \text{für } 6 \leq x < 7 \\ 1 & \text{für } x \geq 7 \end{cases}$$

- 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 $F(3) = 0,4$ und $F(3,5) = 0,4$. Die Verteilungsfunktion $F(x)$ ist auch für $x > 7$ definiert. Sie erreicht an der Stelle $x = 7$ den Wert 1, den sie dann für $x > 7$ beibehält.

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

$$\mu = E(X) = 1 \cdot 0,1 + 2 \cdot 0,1 + 3 \cdot 0,2 + 4 \cdot 0,2 + 5 \cdot 0,2 + 6 \cdot 0,1 + 7 \cdot 0,1 = 4,0.$$

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:

$$\mu_Y = 1 \cdot 0,1 + 4 \cdot 0,1 + 9 \cdot 0,2 + 16 \cdot 0,2 + 25 \cdot 0,2 + 36 \cdot 0,1 + 49 \cdot 0,1 = 19,0 \neq \mu^2 = 16,0.$$

Aufgabe 12 (Stetige Verteilungen)

(5 Punkte)

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

- A) Es sei Z eine standardnormalverteilte Zufallsvariable. Wenn man die Dichtefunktion $\phi(z)$ von Z grafisch darstellt und auf der z -Achse das 0,05-Quantil $z_{0,05}$ und das 0,95-Quantil $z_{0,95}$ der Verteilung markiert, so hat der vom Punkt $z_{0,05}$ bis zum Punkt $z_{0,95}$ gerechnete Flächeninhalt unter der Dichtekurve den Wert 0,90 (Flächeninhalt zwischen Dichtekurve und z -Achse).
- B) Die Dichtefunktion $\phi(z)$ einer standardnormalverteilten Zufallsvariablen nimmt an der Stelle $z = 0$ den Wert 0,5 an.

- C) Es sei X eine stetige Zufallsvariable mit der Verteilungsfunktion $F(x)$. Es bezeichne $x_{0,05}$ das 0,05-Quantil der Verteilung. Die Verteilungsfunktion an der Stelle $x_{0,05}$ hat den Wert 0,05.
- D) Die Dichtekurven von χ^2 -verteilten Zufallsvariablen sind symmetrisch bezüglich des Erwartungswerts.
- E) Die Standardabweichung einer χ^2 -verteilten Zufallsvariablen nimmt mit zunehmender Anzahl von Freiheitsgraden ab.

Lösung zu Aufgabe 12: A, C.

Zu A: Vgl. Kurs 33209, Abschnitt 12.3, Abb. 12.4 (diese mit $\alpha = 0,05$).

Zu B: Die Verteilungsfunktion $\Phi(z)$ einer standardnormalverteilten Zufallsvariablen nimmt an der Stelle $z = 0$ den Wert 0,5 an, nicht aber die Dichtefunktion. Diese Aussage lässt sich auch anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren).

Es wäre überraschend, wenn auch die Dichtefunktion an der Stelle $z = 0$ den Wert 0,5 annähme. Dass die Dichtefunktion $\phi(\cdot)$ der Standardnormalverteilung, die gemäß Seite 16 der Formelsammlung durch

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

erklärt ist, an der Stelle $z = 0$ nicht den Wert 0,5 hat, sondern unterhalb dieses Wertes liegt, kann man auch sofort durch Einsetzen von $z = 0$ in die Dichtedarstellung erkennen. Es ist ja $\phi(0) = \frac{1}{\sqrt{2\pi}}$ und der Nenner $\sqrt{2\pi}$ ist größer als $\sqrt{4}$. Es gilt also

$$\phi(0) = \frac{1}{\sqrt{2\pi}} < \frac{1}{\sqrt{4}} = 0,5.$$

Zu C: vgl. hierzu (11.17).

Zu D: Symmetrie bezüglich des Erwartungswertes ist bei normalverteilten und auch bei t -verteilten Zufallsvariablen gegeben, nicht aber bei χ^2 -verteilten Zufallsvariablen (vgl. hierzu Abbildung 12.5 in Kurs 33209 oder das bei der Lösung zu Teil E genannte Experiment).

Zu E: Die Varianz σ^2 einer mit n Freiheitsgraden χ^2 -verteilten Zufallsvariablen hat den Wert $2n$. Dies impliziert, dass die Standardabweichung den Wert $\sqrt{2n}$ besitzt. Es ist also genau anders herum als in Aussage E formuliert – vgl. auch etwa Abbildung 12.5 in Kurs 33209, aus der man ersieht, dass sich die Dichtekurve einer χ^2 -verteilten Zufallsvariablen mit wachsender Anzahl von Freiheitsgraden abflacht (Zunahme der Streuung). Letzteres lässt sich auch anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren).

Aufgabe 13 (Punkt- und Intervallschätzungen)

(5 Punkte)

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

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

- A) Der Stichprobenmittelwert \bar{X} repräsentiert eine unverzerrte Schätzung für den Erwartungswert μ .
- B) Falls die obige Aussage A zutrifft, gilt auch, dass der mittlere quadratische Fehler des Stichprobenmittelwerts \bar{X} und die Varianz von \bar{X} übereinstimmen.
- C) Wenn man die quadrierten Abweichungen $(X_1 - \bar{X})^2, (X_2 - \bar{X})^2, \dots, (X_n - \bar{X})^2$ aufsummiert und die resultierende Summe durch $n - 1$ dividiert, hat man eine unverzerrte Schätzung für die Varianz σ^2 des Merkmals X .
- D) Die Standardabweichung von \bar{X} geht auf die Hälfte des Ausgangswertes zurück, wenn man n verdoppelt.
- E) Man kann den Erwartungswert μ auch durch Angabe eines Konfidenzintervalls schätzen. Letzteres ist ein Intervall, das stets so groß gewählt wird, dass es den unbekannt Parameter μ enthält.

Lösung zu Aufgabe 13: A, B, C.

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

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

Zu C: Vgl. (14.9) in Kurs 33209 in Verbindung mit (13.5).

Zu D: Aus (14.7) in Kurs 33209 geht hervor, dass eine Verdoppelung von n zur Folge hat, dass die Varianz von \bar{X} auf die Hälfte des Ausgangswertes zurückgeht, d. h. sie ändert sich um den Faktor $\frac{1}{2}$. Die Standardabweichung ändert sich daher bei Verdoppelung von n um den Faktor $\frac{1}{\sqrt{2}}$.

Zu E: Ein Konfidenzintervall enthält den unbekannt Parameter μ nicht immer – vgl. z. B. in Kurs 33209 die Abbildung 14.3, bei der die dunkel markierten Konfidenzintervalle Gegenbeispiele darstellen.

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

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

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.

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

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 ($i = 1, 2, \dots, 6$) 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. Man erhält so eine Regressionsgerade $\hat{y} = \hat{\alpha} + \hat{\beta}x$, deren Parameter $\hat{\alpha}$ und $\hat{\beta}$ sich aus den Beobachtungswerten errechnen. Für $\hat{\alpha}$ errechnet man einen Wert, der größer als 0,13 ist.
- B) Der nach der Kleinst-Quadrat-Methode errechnete Wert $\hat{\alpha}$ kennzeichnet den Schnittpunkt der Regressionsgeraden mit der y -Achse.
- C) Mit dem in der Tabelle wiedergegebenen Datensatz errechnet man für den Korrelationskoeffizienten nach Bravais-Pearson einen Wert zwischen $-0,85$ und $-0,75$.
- D) Die Güte der Anpassung der mit der Kleinst-Quadrat-Methode erhaltenen Regressionsgeraden an den Datensatz lässt sich anhand des Bestimmtheitsmaßes bewerten. Würde man für das Bestimmtheitsmaß den Wert 0,9 errechnen, würde dies beinhalten, dass 90 % der Gesamtvariation des Datensatzes durch das verwendete Regressionsmodell erklärt ist.
- E) Bei Anwendung der Methode der kleinsten Quadrate wird die Regressionsgerade so bestimmt, dass die Summe der quadrierten Residuen Null ist.

Lösung zu Aufgabe 14: A, B, D.

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

$$\hat{\beta} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{-6,8}{40} = -0,17.$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x} = 2,5 - (-0,17) \cdot 14 = 2,5 + 2,38 = 4,88.$$

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

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{-6,8}{\sqrt{40} \cdot \sqrt{1,28}} = \frac{-6,8}{\sqrt{51,2}} \approx -0,95.$$

Zu D: vgl Aufgabe 16.2 in Kurs 33209.

Zu E: Bei der KQ-Methode wird die Regressionsgerade dadurch bestimmt, dass die Summe der quadrierten Residuen minimiert wird. Diese Summe hat aber i. a. nicht den Wert Null — sie ist nur dann Null, wenn alle Punkte des Datensatzes exakt auf einer Geraden liegen.

Aufgabe 15 (Vermischte Aussagen zum Thema „Wahrscheinlichkeit“) (5 Punkte)

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

- A) Die Wahrscheinlichkeit dafür beim Wurf zweier fairer Würfel die Augensumme 6 zu erzielen, ist genauso groß wie die die Wahrscheinlichkeit beim Doppelwurf als Augensumme den Wert 8 zu erhalten.
- B) Zwei Spieler werfen zwei faire Würfel (Doppelwurf). Spieler 1 macht dies einmal, Spieler 2 insgesamt viermal. Es bezeichne p_1 die Wahrscheinlichkeit dafür, dass Spieler 1 die Augensumme 6 erzielt. Ferner bezeichne p_2 die Wahrscheinlichkeit dafür, dass Spieler 2 in seinem letzten Doppelwurf die Augensumme 6 erzielt, nachdem er vorher schon dreimal beim Doppelwurf die Augensumme 6 erhalten hatte. Dann gilt $p_2 < p_1$.
- C) Wenn eine Zufallsvariable t-verteilt ist, so ist die Wahrscheinlichkeit dafür, dass sie einen positiven Wert annimmt, genauso so groß wie die Wahrscheinlichkeit, dass sie einen negativen Wert annimmt.
- D) Getestet werden sollen zu einem vorgegebenen Signifikanzniveau die Hypothesen

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

die sich auf den Erwartungswert μ eines normalverteilten Merkmals beziehen. Anhand des aus Stichprobendaten errechneten Wertes einer geeigneten Prüfgröße kommt man zu einer Testentscheidung, die entweder richtig ist oder falsch ist. Im Falle $\mu = \mu_0$ kann bei dem Test, wenn er zu einer Fehlentscheidung führt, der Fehler kein Fehler 2. Art sein.

- E) Bezeichnet man den Wert der Dichtefunktion einer standardnormalverteilten Zufallsvariablen Z an der Stelle $z = a$ mit p , so lässt sich p als Wahrscheinlichkeit dafür interpretieren, dass Z eine Ausprägung hat, die a nicht übersteigt.

Anmerkung:

Der in den Aufgabenteilen A-B angesprochene „faire“ Würfel bezeichnet einen Würfel mit gleichen Eintrittswahrscheinlichkeiten für alle 6 Augenzahlen.

Lösung zu Aufgabe 15: A, C, D.

Zu A: Von den 36 Elementarereignissen, die den Ereignisraum

$$\Omega = \{(1; 1), (1; 2), \dots, (1; 6), (2; 1), (2; 2), \dots, (2; 6), \dots, (6; 1), (6; 2), \dots, (6; 6)\}$$

beim Würfeln mit zwei Würfeln definieren, führen 5 Elementarereignisse, nämlich die Ausgänge (3; 3), (2; 4); (4; 2), (1; 5), (5; 1) zur Augensumme 6 und ebenfalls 5 Elementarereignisse, nämlich die Ausgänge (4; 4), (3; 5); (5; 3), (2; 6), (6; 2) zur Augensumme 8. Die Wahrscheinlichkeit P für die Erzielung der Augensumme 6 und die zur Erzielung der Augensumme 8 hat somit nach (10.5) in Kurs 33209 denselben Wert $p_1 = \frac{5}{36} \approx 0,139$.

Die theoretische Verteilung des Merkmals „Augensumme“ kann man (neben der empirischen Verteilung) auch anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren – nach Aufruf des Experiments die Option „zugehöriges Modell wählen“ anklicken).

Zu B: Die Augensummen, die Spieler 2 bei den vier Doppelwürfen erzielt, lassen sich durch Zufallsvariablen X_1, X_2, X_3 und X_4 modellieren. Die Eintrittswahrscheinlichkeiten für die Realisation 6 bleibt jedem Doppelwurf konstant, hat also für jede der Variablen X_1, X_2, X_3 und X_4 den festen Wert $p_2 = p_1 = \frac{5}{36} \approx 0,139$. Die Vorgeschichte beeinflusst nicht die Eintrittswahrscheinlichkeit für den nächsten Doppelwurf (Unabhängigkeit der Zufallsvariablen). Man kennt dieses Phänomen der „Gedächtnislosigkeit“ auch von anderen Glücksspielen. Wenn jemand beim Roulette eine lange Serie von Ausgängen mit „Schwarz“ beobachtet hat, ist die Eintrittswahrscheinlichkeit von „Schwarz“ beim nächsten Roulettespiel noch genau wie beim ersten Spiel (vgl. auch den ersten Absatz von Abschnitt 13.1 in Kurs 33209).

Zu C: Dies folgt daraus, dass jede t-Verteilung – wie auch die Standardnormalverteilung – symmetrisch bezüglich des Nullpunkts ist.

Zu D: Im Falle $\mu = \mu_0$ ist die Nullhypothese richtig. Die einzig mögliche Fehlentscheidung ist die Verwerfung der Nullhypothese und diese beinhaltet einen Fehler 1. Art (vgl. auch Tabelle 15.1 in Kurs 33209).

Zu E: Bezeichnet man den Wert der Verteilungsfunktion $\Phi(z)$ (nicht aber der Dichtefunktion) einer standardnormalverteilten Zufallsvariablen Z an der Stelle $z = a$ mit p , so ist $p = P(Z \leq a)$ die Wahrscheinlichkeit dafür, dass Z eine Ausprägung hat, die a nicht übersteigt. Den Sachverhalt kann man aus Abbildung 12.3 in Kurs ansehen (dort wurde $a = 1,38$ gewählt) oder auch anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren).

Aufgabe 16 (Testen, Fehler beim Testen)

(5 Punkte)

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

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

und zwar zum Signifikanzniveau $\alpha = 0,01$. Wenn man die Varianz σ^2 als bekannt voraussetzt, kann man den standardisierten Stichprobenmittelwert $Z = (\bar{X} - \mu_0)/\sigma_{\bar{X}}$ als Prüfgröße für den Test heranziehen (Gauß-Test).

Welche der folgenden Aussagen sind richtig?

(x aus 5)

- A) Wird bei obigem Test für die Prüfgröße der Wert 2,31 ermittelt, erfolgt keine Ablehnung der Nullhypothese.
- B) Die Wahrscheinlichkeit dafür, einen Fehler 1. Art zu begehen, beträgt bei obigem Test für alle μ mit $\mu \leq \mu_0$ stets 0,01.
- C) Wenn $\mu > \mu_0$ gilt, ist ein Fehler 2. Art möglich, nicht aber ein Fehler 1. Art.
- D) Wenn man den Stichprobenumfang n erhöht, wird für alle Werte μ , an denen ein Fehler 2. Art auftreten kann, die Eintrittswahrscheinlichkeit für den Fehler 2. Art reduziert.
- E) Wenn für den unbekannt Parameter μ die Ungleichung $\mu > \mu_0$ gilt, beinhaltet die Ablehnung von H_0 den Eintritt eines Fehlers 2. Art.

Lösung zu Aufgabe 16: A, C, D.

Zu A: Die Ablehnung der Nullhypothese erfolgt bei dem hier zugrunde gelegten rechtsseitigen Test, wenn für die Realisation z der Prüfgröße Z die Bedingung $z > z_{1-\alpha}$ gilt, wegen $\alpha = 0,01$ also bei Überschreitung von $z_{0,99} \approx 2,3263$.

Zu B: Die Wahrscheinlichkeit dafür, einen Fehler 1. Art zu begehen, hat bei obigem Test nur für $\mu = \mu_0$, also nur am Rande des Gültigkeitsbereichs von H_0 , den Wert $0,01$. Für alle anderen Werte μ , für die H_0 zutrifft (also für alle μ mit $\mu < \mu_0$), ist die Wahrscheinlichkeit dafür, einen Fehler 1. Art zu begehen, kleiner als $0,01$ (vgl. Abbildung 15.5 in Kurs 33209).

Zu C und D: vgl. hierzu ebenfalls Abbildung 15.5 in Kurs 33209 (hier ist $\alpha = 0,01$ zu wählen).

Zu E: Die Ablehnung von H_0 beinhaltet hier eine korrekte Entscheidung. Ein Fehler 2. Art wäre im Falle $\mu > \mu_0$ die Nicht-Verwerfung von H_0 .

Numerische Aufgaben

Aufgabe 41 (Kombinatorik)

(3 Punkte)

Den Mitarbeitern einer Firma wird als Intranet-Code eine genau vier Ziffern umfassende Ziffernfolge zugewiesen. Jede Ziffer darf auch mehrfach innerhalb des Codes vorkommen. Zugelassen sind nur die Ziffern von 1 bis 9 (Ausschluss der 0).

Wieviele Mitarbeiter könnte man maximal anhand solcher Ziffernfolgen unterscheiden, d. h. wieviele unterschiedliche Ziffernkombinationen sind realisierbar?

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

(numerisch)

--	--	--	--	--	--

Lösung zu Aufgabe 41: $9^4 = 6561$.

Herleitung:

Die gesuchte Anzahl ist durch (10.8) in Kurs 33209 gegeben mit $N = 9$ und $n = 4$ (Ziehen von 4 Elementen mit Zurücklegen und mit Berücksichtigung der Anordnung aus einer Grundgesamtheit mit 9 Elementen; s. auch Tabelle 10.1 in Kurs 33209).

Aufgabe 42 (Zusammenhangsmessung)

(3 Punkte)

Bei einer schriftlichen Prüfung mit $n = 38$ Teilnehmern wurden für die Merkmale „Prüfungserfolg X “ mit den Ausprägungen a_1 und a_2 (bestanden / nicht bestanden) und „Geschlecht Y “ mit den Ausprägungen b_1 und b_2 (weiblich / männlich) die absoluten Häufigkeiten erfasst. Die Ergebnisse sind in der nachstehenden Kontingenztafel zusammengefasst: Um einen empirischen Zusammenhang zwischen den beiden Merkmalen zu

	b_1	b_2
a_1	20	6
a_2	8	4

quantifizieren, kann man das Zusammenhangsmaß V von Cramér heranziehen (Cramér's V). Berechnen Sie dieses Maß auf der Basis der obigen (2 x 2)-Kontingenztafel. Runden Sie das Ergebnis auf *vier Stellen nach dem Dezimal komma* und tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Das **Dezimal komma** belegt wieder ein **eigenes Feld**. Vergessen Sie nicht, Ihre Antwort rechtzeitig vor dem Ende der Klausur auf den Markierungsbogen zu übertragen.

(numerisch)

$V =$

Lösung zu Aufgabe 42: 0,1083.

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

	b_1	b_2	Zeilensummen
a_1	20	6	26
a_2	8	4	12
Spaltensummen	28	10	38

Es folgt bei Anwendung von Formel (9.8) aus Kurs 33209

$$V = \frac{20 \cdot 4 - 8 \cdot 6}{\sqrt{26 \cdot 12 \cdot 28 \cdot 10}} = \frac{32}{\sqrt{87360}} \approx 0,1083.$$

Anmerkung:

Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0, 1070; 0, 1095]$ als richtig anerkannt.

Aufgabe 43 (Roulette)

(3 Punkte)

Ein Hotelrestaurant in Norddeutschland bietet an jedem Samstag ein „Roulette-Menü“ an. Jeder Gast, der daran teilnimmt, erhält zunächst ein zweigängiges Tagesmenü. Danach gehen alle teilnehmenden Gäste an einen Roulettetisch. Die Hotelleitung lässt einmal unter den Blicken der Gäste die Kugel rollen. Das Ergebnis des einmaligen Roulettespiels bestimmt den Preis X , den jeder Gast für das Menü zu zahlen hat. Der Rechnungsbetrag X ist genau die Zahl (in Euro), bei der die Kugel liegen bleibt. Wenn also das Ergebnis des Roulettespiels 0 lautet, haben alle Gäste auf Kosten des Hauses gespeist (Rechnungsbetrag: 0,00 Euro). Damit das Menü für die Gäste nicht zu teuer und damit unattraktiv wird, wird eine Rouletteversion verwendet, bei der nur die Ausgänge 0, 1, ...,18 auftreten können (mit identischen Eintrittswahrscheinlichkeiten für alle denkbaren Ausgänge). Im – aus der Sicht der Gäste – ungünstigsten Fall hat also jeder Gast 18,00 Euro zu zahlen.

Berechnen Sie den Erwartungswert $E(X)$ des beim Roulette-Menü am Ende von jedem Gast zu zahlenden Betrags. Geben Sie Ihr Ergebnis auf 2 Stellen nach dem Dezimalkomma genau an, also in Euro und Cent. Tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch)

Lösung zu Aufgabe 43:

Der Erwartungswert von X beträgt $\mu = E(X) = 9,00$ (Euro).

Herleitung:

Für den Ausgang beim Roulette gibt es bei der hier verwendeten Roulettevariante 19 Möglichkeiten (vgl. auch Abbildung 4.10 in Kurs 33209). Alle denkbaren Ausgänge haben die gleiche Eintrittswahrscheinlichkeit $p = \frac{1}{19}$. Man erhält somit nach (11.2) für den Erwartungswert $\mu = E(X)$ der diskreten Zufallsvariablen X (diskrete Gleichverteilung):

$$\mu = 0 \cdot \frac{1}{19} + 1 \cdot \frac{1}{19} + \dots + 18 \cdot \frac{1}{19} = \frac{1}{19} \cdot \sum_{i=0}^{18} i.$$

Es gilt also zunächst die Zahlen 0, 1, .. 18 zu addieren. Man kann die Zahlen umständlich einzeln aufsummieren oder aber ausnutzen, dass die 19 Zahlen im Mittel den Wert $\frac{0+18}{2} = 9$ haben, die Summe sich also als $19 \cdot 9 = 171$ ergibt. Für den Erwartungswert von X folgt also

$$\mu = \frac{171}{19} = 9.$$

Das Ergebnis ist auch ohne ausführlichere Rechnung sofort einleuchtend: Die Realisation 9 liegt bei der hier betrachteten diskreten Gleichverteilung mit 19 Realisationen (ungerade Anzahl) genau in der Mitte — je 9 Ausgänge liegen unterhalb von $\mu = 9$ (nämlich 0, 1, ..., 8) und oberhalb dieses Werts (oberhalb: 10, 11, ..., 18). Wenn man die Wahrscheinlichkeitsfunktion von X anhand eines Balkendiagramms zeichnet (19 an den Stellen 0, 1, ... , 18 ansetzende Balken mit je der Länge $\frac{1}{19}$), ist der Balken an der Stelle $\mu = 9$ der genau in der Mitte stehende Balken (Zentrum der Verteilung).

In der analogen Abbildung 11.1 (oberer Teil) aus Kurs 33209, die sich auf eine diskrete Gleichverteilung mit 6 Realisationen bezog (gerade Anzahl), wird die Mitte der Verteilung nicht angenommen. Der in Beispiel 11.1 errechnete Erwartungswert liegt hier bei 3,5 (je 3 Ausgänge liegen unterhalb und oberhalb von $\mu = 3,5$).

Anmerkungen:

Anstelle von 9,00 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[8,95; 9,05]$ als richtig anerkannt.

Aufgabe 44 (Gewinnwahrscheinlichkeit beim Roulette)

(3 Punkte)

Das in Aufgabe 43 beschriebene „Roulette-Menü“ (Roulettespiel mit den Ausgängen 0, 1, ..., 18) wird von dem norddeutschen Hotelrestaurant an jedem Samstag wahlweise auch in einer zweiten Version angeboten, die risikoscheuere Gäste ansprechen soll. Die Hotelleitung lässt auch hier einmal unter den Blicken der Gäste die Kugel rollen und das Ergebnis 0 bedeutet auch hier, dass der Gast nichts zu zahlen hat. Bei den Ergebnissfeldern von 1 bis 18 wird der Preis X , den jeder Gast bei Wahl dieser Variante für das Menü zu zahlen hat, von der Farbe des Felds bestimmt, auf das die Kugel fällt. Der Rechnungsbetrag X beträgt 7,50 Euro, wenn die Kugel auf einem der 9 roten Felder liegen bleibt und 12,50 Euro, wenn es eines der 9 schwarzen Felder ist. Im – wieder aus der Sicht der Gäste – ungünstigsten Fall hat jeder Gast hier also nur 12,50 je Euro zu zahlen.

Ein Gast nimmt an an zwei Samstagen am Roulette-Menü teil und wählt jeweils die Spielvariante, bei der der Rechnungsbetrag im Falle eines von 0 verschiedenen Ergebnisses von der Farbe des Ergebnisfelds bestimmt wird. Wie groß ist die Wahrscheinlichkeit, dass er nach den beiden Terminen insgesamt 15 Euro für die beiden Menüs bezahlt hat?

Geben Sie die Wahrscheinlichkeit, also einen zwischen 0 und 1 liegenden Wert (Ergebnis nicht in Prozenten ausdrücken), auf 4 Stellen nach dem Dezimalkomma genau an. Tragen Sie Ihr Ergebnis rechtsbündig in das Antwortfeld ein. Übertragen Sie Ihr Ergebnis rechtzeitig vor Ende der Klausur auf den Markierungsbogen.

(numerisch)

--	--	--	--	--	--	--	--

Lösung zu Aufgabe 44: Die Wahrscheinlichkeit beträgt 0,2244.

Herleitung:

Die Wahrscheinlichkeit für den Ausgang *Rot* beträgt bei jedem Spiel $\frac{9}{19} \approx 0,4737$. Die Wahrscheinlichkeit dafür, bei zwei Spielen zweimal den Ausgang *Rot* zu bekommen, ist aufgrund der Unabhängigkeit der beiden Einzelereignisse nach (10.16) gegeben durch

$$p = \frac{9}{19} \cdot \frac{9}{19} = \frac{81}{361} \approx 0,2244.$$

Anmerkung:

Anstelle von 0,2244 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,222; 0,227]$ als richtig anerkannt.

Aufgabe 45 (Schwellenwertbestimmung bei einer Normalverteilung) (3 Punkte)

Es sei eine Zufallsvariable X betrachtet, die normalverteilt ist mit Erwartungswert $\mu = 1$ und Varianz $\sigma^2 = 0,01$. Bestimmen Sie einen Wert a , für den $P(X > a) = 0,95$ gilt, also einen Wert a , der mit Wahrscheinlichkeit 0,95 überschritten wird.

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

(numerisch)

 $a =$

--	--	--	--	--	--	--

Lösung zu Aufgabe 45: 0,8355.

Herleitung:

Gesucht ist ein Wert a , für den $P(X > a) = 0,95$ und damit $P(X \leq a) = 0,05$ gilt. Dies ist offenbar das 0,05-Quantil $x_{0,05}$ der Normalverteilung mit Erwartungswert $\mu = 1$ und Varianz $\sigma^2 = 0,01 = 0,1^2$. Für dieses gilt nach Formel (12.26)

$$x_{0,05} = 1 + z_{0,05} \cdot 0,1.$$

Nach (12.25) gilt $z_{0,05} = -z_{0,95}$. Man hat daher mit Tabelle 19.3 aus Kurs 33209

$$x_{0,05} = 1 - z_{0,95} \cdot 0,1 = 1 - 1,6449 \cdot 0,1 = 1 - 0,16449 \approx 0,8355.$$

Die Wahrscheinlichkeit dafür, dass eine Realisation von X den Wert 0,8355 überschreitet, beträgt also 0,95.

Anmerkungen:

Anstelle von 0,8355 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,827; 0,843]$ als richtig anerkannt.

Wenn anstelle des 0,05-Quantils $x_{0,05} = 0,8355$ das 0,95-Quantil $x_{0,95} = 1 + 0,16449 \approx 1,1645$ als Ergebnis eingetragen wurde, gab es 2 Punkte. Letzteres galt auch für jeden Wert aus dem Intervall $[1,153; 1,176]$.

Aufgabe 46 (Normalverteilung und Standardnormalverteilung) (3 Punkte)

Es bezeichne X eine mit Erwartungswert $\mu = 3$ und Varianz $\sigma^2 = 4$ normalverteilte Zufallsvariable. Berechnen Sie die Wahrscheinlichkeit $P(X > 5)$ dafür, dass eine Realisation von X den Wert 5 überschreitet.

Geben Sie die Wahrscheinlichkeit auf 4 Stellen nach dem Dezimalkomma genau an. Tragen Sie Ihr Ergebnis 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)

--	--	--	--	--	--	--	--

Lösung zu Aufgabe 46: Die Wahrscheinlichkeit beträgt 0,1587.

Herleitung:

Nach Formel (12.22) in Kurs 33209 gilt

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \Phi\left(\frac{5-3}{2}\right) = 1 - \Phi(1) \approx 1 - 0,8413 = 0,1587.$$

Anmerkungen:

Der Wert $\Phi(1) \approx 0,8413$ lässt sich auch – zumindest in sehr guter Näherung – anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und Schieber in etwa auf $z = 1,00$ einstellen.) Anstelle von 0,1587 wurde bei der maschinellen Auswertung jeder Wert aus dem Intervall $[0,157; 0,160]$ als richtig anerkannt.

Wenn 0,8413 anstelle von $1 - 0,8413 = 0,1587$ als Lösung eingetragen wurde, wurden 2 P. vergeben. Die Vergabe von 2 P. wurde auf jeden Wert aus dem Intervall $[0,833; 0,850]$ angewendet.

Aufgabe 47 (Gauß-Test)

(3 Punkte)

Es sei nochmals – wie schon bei Aufgabe 16 – ein Gauß-Test zum Signifikanzniveau $\alpha = 0,01$ betrachtet, der sich die Hypothesen

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

bezieht. Dabei bezeichnet μ den Erwartungswert eines als normalverteilt spezifizierten Merkmals, dessen Varianz σ^2 als bekannt vorausgesetzt wird. Als Prüfgröße des Tests wird der standardisierte Stichprobenmittelwert $Z = (\bar{X} - \mu_0)/\sigma_{\bar{X}}$ herangezogen. Es sei $\mu_0 = 1$, d. h. die Nullhypothese H_0 trifft für jedes μ mit $\mu \leq 1$ zu. Ferner gelte $\sigma = 1$.

Wenn der Parameter μ , auf den sich der Test bezieht, oberhalb von $\mu_0 = 1$ liegt, sollte der Test möglichst zu dem Ergebnis führen, dass die Nullhypothese verworfen wird. Wie groß ist bei dem beschriebenen Test die Wahrscheinlichkeit einer Verwerfung von H_0 , wenn der (unbekannte) Parameter μ den $\mu = 1,5$ hat und der Test mit einer Stichprobe des Umfangs $n = 9$ arbeitet?

Geben Sie die Wahrscheinlichkeit auf 3 Stellen nach dem Dezimalkomma genau an. Wenn Sie bei der Lösung dieser Aufgabe eine Tabelle aus der Formelsammlung verwenden und die Tabelle den gesuchten Wert nicht genau ausweist, können Sie den nächstgelegenen Tabellenwert heranziehen. Tragen Sie Ihr Ergebnis 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)

--	--	--	--	--	--	--

Lösung zu Aufgabe 47: Die Wahrscheinlichkeit beträgt ca. 0,2033.

Die Wahrscheinlichkeit für die Verwerfung der Nullhypothese H_0 ist für jeden Wert des unbekanntem Parameters μ , auf den sich der Test bezieht, durch den Wert $G(\mu)$ der Gütefunktion (15.11) gegeben (Gütefunktion für den rechtsseitigen Fall). Man erhält nach Einsetzen von $\mu_0 = 1$, $\sigma = 1$, $n = 9$ und $z_{0,99} \approx 2,3263$ in (15.11), wenn man den nächstgelegenen Wert aus Tabelle 19.2 heranzieht (dies ist der unter $z = 0,83$ tabellierte Wert 0,7967):

$$G(1, 5) \approx 1 - \Phi(2,3263 - 0,5 \cdot 3) = 1 - \Phi(0,8263) \approx 1 - 0,7967 = 0,2033.$$

Wenn man den unter $z = 0,82$ tabellierten Wert 0,7939 verwenden würde, erhielte man

$$G(1, 5) \approx 1 - 0,7939 = 0,2067.$$

Bei Interpolation würde man einen Wert erhalten (0,2057), der zwischen 0,2033 und 0,2067 liegt.

Der Wert, der hier die Wahrscheinlichkeit einer korrekten Testentscheidung beinhaltet, lässt sich anhand von Abbildung 15.5 in Kurs 33209 zumindest näherungsweise ablesen – vgl. dort den Wert der gestrichelten Kurve an der Stelle 0,5 (obere Abszissenachse) – die gestrichelte Kurve basiert allerdings auf $n = 10$ (für $n = 9$ verläuft sie geringfügig flacher).

Anmerkungen:

Die Aussage $\Phi(0,8263) \approx 0,7967$ lässt sich auch – zumindest in sehr guter Näherung – anhand eines statistischen Experiments nachvollziehen (grün umrahmten Link bei der Online-Fassung dieses Dokuments aktivieren und Schieber auf $z = 0,83$ einstellen. Bei der maschinellen Auswertung wurde jeder Wert aus dem Intervall $[0,201; 0,209]$ als richtig anerkannt.

Wenn 0,7967 anstelle von $1 - 0,7967 = 0,2033$ bzw. 0,7939 statt $1 - 0,7939 = 0,2067$ als Lösung eingetragen wurde, wurden 2 P. vergeben. Die Vergabe von 2 P. wurde auf jeden Wert aus dem Intervall $[0,786; 0,805]$ angewendet.

Wurde anstelle der Gütefunktion (15.11) (Gütefunktion für den *rechtsseitigen* Fall) fälschlich die Gütefunktion (15.12) angewendet (Gütefunktion für den *linksseitigen* Fall), resultierte

$$G(1, 5) \approx \Phi(-2,3263 - 0,5 \cdot 3) = 1 - \Phi(3,8263) \approx 1 - 0,9999 = 0,0001.$$

Wenn ein Wert aus dem Intervall $[0; 0,0001]$ als Lösung angegeben wurde, gab es 1 Punkt.

Aufgabe 48 (KQ-Schätzung)

(3 Punkte)

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

$$(12; 0); \quad (24; 1); \quad (21; 1); \quad (36; 2, 5); \quad (42; 3).$$

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

(numerisch)

$$\hat{\beta} = \boxed{} \boxed{} \boxed{} \boxed{} \boxed{} \boxed{} \boxed{} \boxed{}$$

Lösung zu Aufgabe 48: 0,1016**Herleitung:**

Aus den Daten errechnet man $\bar{x} = 27$ und $\bar{y} = 1,5$ und daraus unschwer

$$\sum_{i=1}^5 (x_i - \bar{x})^2 = 576; \quad \sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = 58,5.$$

Nach Formel (16.6) aus Kurs 33209 folgt dann $\hat{\beta} = \frac{58,5}{576} \approx 0,1016$.

Anmerkung:

Bei der maschinellen Auswertung wurde anstelle von 0,1016 auch jeder Wert aus dem Intervall $[0,100; 0,103]$ als richtig anerkannt.