

## Musterlösungen zur Übungsklausur zum Modul 2.1 "Methoden und Analyseverfahren" im Bachelor-Studiengang "Politik und Organisation"

Klausur zur Modulfassung des WS 2007/08 mit den Kursen 03607, 33207 und 33208.

Die zu dieser Modulfassung folgende "echte" **Klausur am 10. März 2008** wird  
bezüglich Aufbau und Schwierigkeitsgrad vergleichbar sein.

**Name:**
**Matrikel-Nr. :**

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

**Unterschrift:**

<b>Block 1 (Kurs 03607 „Empirische Sozialforschung“)</b>						
<b>Aufgabe</b>	1-1	1-2	1-3	1-4	1-5	<b>Σ</b>
<b>maximale Punktzahl</b>	<b>6</b>	<b>6</b>	<b>6</b>	<b>4</b>	<b>3</b>	<b>25</b>
<b>erreichte Punktzahl</b>						

<b>Block 2 (zu den Kursen 33207 „Statistik und wissenschaftstheoretische Grundlagen“ und 33208 „SPSS“)</b>										
<b>Aufgabe</b>	2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9	<b>Σ</b>
<b>maximale Punktzahl</b>	<b>6</b>	<b>10</b>	<b>3</b>	<b>13</b>	<b>7</b>	<b>9</b>	<b>13</b>	<b>10</b>	<b>4</b>	<b>75</b>
<b>erreichte Punktzahl</b>										

**Gesamtpunktzahl:**
**Note:**
**Prüfer:**
**Unterschrift des Prüfers:**

## **Bearbeitungshinweise und Hinweise zu dieser Übungsklausur:**

Es sind *beide* Aufgabenblöcke zu bearbeiten. Für Berechnungen in dieser Klausur können eine separat verteilte **Formelsammlung** sowie ein **Taschenrechner** herangezogen werden.

Bitte beantworten Sie alle Fragen stets kurz, nicht in Form von Aufsätzen.

### **Hinweise zu dieser Übungsklausur:**

**Block 1:** Da der Kurs 03607 bei der Änderung der Modulstruktur zum WS 2007/08 nicht verändert wurde (abgesehen von einer Streichung der Kapitel 8-9), sind die Aufgaben zu Block 1 bei der Musterklausur durchweg Aufgaben aus alten Klausuren. Auch das Gewicht des Blocks bleibt mit 25 P. von 100 P. unverändert.

**Block 2:** Dieser Block hat mit 75 P. dasselbe Gewicht, das bei früheren Klausuren insgesamt den Aufgabenblöcken zu den ehemaligen Kursen 03607 und 33208 zukam (25 P. und 50 P. waren hier bisher veranschlagt). Diese beiden Kurse bilden ja nun den Kurs 33207. Neu ist aber, dass der Block zum früheren Statistikkurs 33208 ein Wahlblock war und diese Wahloption nun nicht mehr besteht, weil der frühere Kurs 33211 jetzt im Rahmen des Moduls 2.2 geprüft wird. Die Klausur *im März 2008* wird in Block 2 – wie diese Übungsklausur – noch *keine Fragen* enthalten, die sich auf SPSS beziehen.

**Noch ein Tipp:** Nutzen Sie die bei der Prüfungsvorbereitung auch die in der Lernumgebung „Moodle“ bereitgestellten Arbeitsmaterialien (optional), die Ihnen den Zugang zu einzelnen Themenbereichen der Statistik erleichtern können (Experimente und Animationen mit Kommentaren) – den Zugang zu „Moodle“ finden Sie im Studienportal zum Modul 2.1.

**Aufgabe 1-1 (Hypothesenbildung)****6 P.**

In der Wahlforschung wird üblicherweise davon ausgegangen, dass die Parteiidentifikation, die Kandidatensympathie und die Einstellung zu Sachthemen des Wahlkampfes die Parteiwahlentscheidung beeinflussen. Formulieren Sie zum Zusammenhang zwischen den hier benannten Variablen eine Hypothese, benennen Sie die unabhängigen und die abhängigen Variablen und stellen Sie den Zusammenhang in Form eines Hempel-Oppenheim-Schemas dar.

Lösung:

Zur Lösung diese Aufgabe vgl. Abschnitt 2.4.1 von Kurs 03607 (Kromrey-Kurs):

Unabhängige Variablen: PID, Kandidatensympathie, Issue-Orientierung

Abhängige Variable: Parteiwahl

Zusammenhangshypothese:

Wenn ein Wähler eine Parteiidentifikation für die SPD hat, wählt er auch die SPD.

Alternativ: Die Entscheidung, eine Partei zu wählen, hängt davon ab, welcher Spitzenkandidat einem Wähler am sympathischsten ist.

Hempel-Oppenheim-Schema:

allg. Gesetz: Ein Wähler gibt seine Stimme für die Partei ab, für die er eine Parteiidentifikation aufweist.

Randbedingung: Ein Wähler hat eine PID für die SPD  
(beides zusammen bildet das Explanans)

---

Schlussfolgerung (explanandum): Der Wähler wählt die SPD

**Aufgabe 1-2 (experimentelles Design)****6 P.**

Welche Merkmale kennzeichnen ein klassisches experimentelles Design, das in sozialwissenschaftlichen Experimenten typischerweise mit Menschen durchgeführt wird?

Lösung:

Zur Lösung diese Aufgabe vgl. Abschnitt 2.4.3 von Kurs 03607 (Kromrey-Kurs):

Ziel eines experimentellen Designs:

Ein experimentelles Design zeichnet sich durch das Bemühen aus, alle Faktoren in der Untersuchungssituation auszuschalten, die nicht auf das vermutete Ursache-Wirkungs-Prinzip zurückzuführen sind. Alle Randbedingungen sind also möglichst konstant zu halten. Zugleich sollte das Ausmaß der kausalen Wirkung messbar sein. Hierfür benötigt man eine Kontrollgruppe.

Merkmale eines sozialwissenschaftlichen Experiments:

Die Merkmale umfassen somit:

- eine Experimentalgruppe, die tatsächlich dem Treatment ausgesetzt ist
- eine Kontrollgruppe, die dem Treatment nicht ausgesetzt ist, ansonsten aber der Experimentalgruppe möglichst gleich
- in beiden Gruppen werden die Ausprägungen der abh. Variablen vor und nach dem Treatment gemessen und verglichen
- Unterschiede zwischen den Gruppen, die nach dem Treatment auftreten, werden als Treatment-Effekt im Sinne des Experiments interpretiert.

### Aufgabe 1-3 (Auswahlverfahren)

6 P.

Definieren Sie kurz die Begriffe *Grundgesamtheit*, *Vollerhebung*, *Auswahleinheit* und *Erhebungseinheit* und erläutern sie den unterschiedlichen Bedeutungsumfang an einem Beispiel

Lösung:

Zur Lösung dieser Aufgabe vgl. Abschnitt Kap. 6.1 in Kurs 03607 (Kromrey-Kurs):

*Grundgesamtheit:*

diejenige Menge von Individuen, Fällen oder Ereignissen, auf die sich die Aussagen der Untersuchung beziehen sollen und die im Hinblick auf die Fragestellung und die Operationalisierung vorher eindeutig abgegrenzt werden muss.

*Vollerhebung:*

alle Elemente der Grundgesamtheit gehen in die Erhebung ein

*Auswahleinheit:*

diejenige Einheit, auf die sich der Auswahl(Stichproben-)plan konkret bezieht

*Erhebungseinheit:*

diejenige Einheit, die repräsentativ in der Stichprobe vertreten sein soll

Beispiel zur Illustration aller vier Begriffe:

Angestrebt ist eine Erhebung über die politischen Einstellungen von Erstwählern. Diese soll mit Hilfe einer Face-to-Face Befragung durchgeführt werden, wobei der Auswahlplan mit zufälliger Auswahl von Sample Points und Random-Route-

Verfahren Haushalte (Adressen) wählt. Grundgesamtheit ist dann die Menge aller Personen, die zwischen dem vergangenen und dem kommenden Termin für die Bundestagswahlen das 18. Lebensjahr vollendet haben und damit für die nächste Bundestagswahl erstmals wahlberechtigt sind oder werden. Eine Vollerhebung würde man durchführen, wenn man alle diese Personen befragen würde. Die Auswahleinheit ist (aufgrund des Stichprobenverfahrens) ein Haushalt, bei dem geklingelt und gefragt wird. Die Erhebungseinheit sind die befragten Personen (Jungwähler).

#### **Aufgabe 1-4 (Datenerhebung)**

**4 P.**

Beschreiben Sie, was ein reaktives oder um ein nicht-reaktives Verfahren der Datenerhebung kennzeichnet. Wie sind Befragung und Beobachtung in diesem Zusammenhang zu bewerten? Geben Sie nur kurze Antworten.

Lösung:

*Zur Lösung diese Aufgabe Abschnitt XI.1 und XIII.1 bei Diekmann*

Charakterisierung von reaktiven bzw. nicht-reaktiven Verfahren der Datenerhebung:

*Reaktiv ist ein Messverfahren immer dann, wenn sich durch den Akt der Messung das zu Messende verändert. Nicht-reaktive Verfahren der Datenerhebung sind daher typischerweise die Inhalts- und Dokumenten- oder die Sekundäranalyse. Befragung hingegen ist immer reaktiv (wenn ich keine Frage gestellt hätte, hätte ich vermutlich keine Antwort erhalten, allein hier liegt bereits die Veränderung des zu Messenden).*

Einordnung der Erhebungsverfahren *Befragung* sowie *Beobachtung* :

*Bei der Beobachtung können zwei Fälle unterschieden werden, die offene und die verdeckte Beobachtung. Die verdeckte Beobachtung ist eigentlich ein nicht-reaktives Verfahren, da die Beobachteten ja gar nicht wissen, dass sie beobachtet werden. Eine offene Beobachtung hingegen ist immer reaktiv. Zwar können die Beobachteten vergessen, dass sie beobachtet werden, aber eine gewisse Unnatürlichkeit der Situation bleibt in den meisten Fällen bestehen.*

#### **Aufgabe 1-5 (Fragebogen)**

**3 P.**

Was versteht man bei der Mikroplanung (Gestaltung einzelner Themenblöcke) des Fragebogens unter Kontexteffekten? Was kann man tun, um diese zu herauszufinden?

Lösung:

*Zur Lösung diese Aufgabe Abschnitt X.6 bei Diekmann*

*Kontexteffekt:*

*Antworten auf vorhergehende Fragen können das Antwortverhalten bei nachfolgenden Fragen beeinflussen.*

*Identifikation von Kontexteffekten:*

*Man kann die Effekte feststellen, indem man die Reihenfolge der Fragen in einem Frageblock variiert.*

## Block 2: Statistik und wissenschaftstheoretische Grundlagen 75 P.

### Aufgabe 2-1 (Skalenarten)

6 P.

In etlichen Städten werden Mietspiegel erstellt, die für Mieter und Vermieter eine Marktübersicht zu Miethöhen liefern. In solchen Mietspiegeln werden u. a. die nachstehenden 6 Merkmale erfasst. Geben Sie bei jedem Merkmal an, welcher Skalentyp zutrifft, indem Sie in jeder der 6 Tabellenzeilen genau ein Kreuz (x) setzen. Der Begriff „Metrische Skala“ wird als Oberbegriff für die Skalentypen Intervallskala, Verhältnisskala und Absolutskala verwendet. (In der letzten Spalte ist also genau dann ein Kreuz zu setzen, wenn das betreffende Merkmal einem der drei genannten Skalentypen zuzuordnen ist.)

Merkmal	Zutreffender Skalentyp		
	Nominalskala	Ordinalskala	Metrische Skala
Größe einer Wohnung in m <sup>2</sup>			x
Baujahr des Gebäudes			x
Anzahl der Zimmer			x
Art der Heizung (Gas, Elektro, Öl)	x		
Nettomiete für die Wohnung in €			x
Lage der Wohnung (6 Kategorien von „Spitzenlage = 1“ bis „Problemlage = 6“)		x	

Hinweis zur Lösung von Aufgabe 2-1: s. Kurs 33207, Abschnitte 4.1 – 4.5

### Aufgabe 2-2 (Aussagenlogik)

10 P.

- a) Nachstehend sind jeweils zwei Prämissen (P1 und P2) und ein aus diesen abgeleiteter logischer Schluss (Konklusion K) aufgelistet, die zusammen jeweils ein Argument darstellen. Nicht alle Schlüsse K sind unbedingt korrekt. Tragen Sie unter „Antworten“ das Kürzel „w“ für korrekte Konklusionen und „f“ bei Fehlschlüssen ein. (Bevor Sie sich für die Gültigkeit oder Ungültigkeit beim letzten der vier Argumente entscheiden geben, bearbeiten Sie am besten den Aufgabenteil b.) (3 P.)

P1	$a \rightarrow \neg b$	$\neg a \rightarrow \neg b$	$b \rightarrow \neg a$
P2	b	a	b
	_____	_____	_____
K	$\neg a$	b	$\neg a$

Antworten: ( w ) ( f ) ( w )

- b) Verfahren Sie ebenso bei den nachstehend angegebenen Prämissen P1 und P2 und der angegebenen Konklusion K, d. h. tragen Sie auch hier „w“ oder „f“ ein. Füllen Sie zuvor die vorgegebene Wahrheitstabelle aus, aus der sich die korrekte Antwort („w“ oder „f“) ergibt:

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

Antwort: ( w )

(7 P.)

Bepunktung bei Aufgabenteil a: Je 1 P. pro zutreffender Antwort. Bei unzutreffenden Antworten wird zur Vermeidung von Raststrategien je ein Punkt abgezogen. Die Gesamtpunktzahl für Aufgabenteil a darf dabei aber 0 P. nicht unterschreiten.

Bepunktung bei Aufgabenteil b: Für jede korrekt ergänzte Spalte der Wahrheitstabelle 2 P., also maximal 6 P. (keine Abzüge bei Fehlantworten). Für die korrekte Antwort („w“ oder „f“) gibt es nur dann einen weiteren Punkt, wenn die Antwort nicht im Widerspruch zu den Inhalten der letzten Spalte der Wahrheitstabelle steht.

**Hinweis zur Lösung von Aufgabe 2-2: s. Kurs 33207, Kapitel 5. Der erste Fall in aufgabenteil a ist der Modus Tollens, der dritte Fall der Modus Ponens. Der zweite Fall ist im Kurs als Fehlschluss der Verneinung de Antezedens angesprochen.**

**In Aufgabenteil b sind nur in der letzten Zeile beide Prämissen und gleichzeitig auch die Konklusion wahr. Hieraus ergibt sich für die Antwort „w“.**

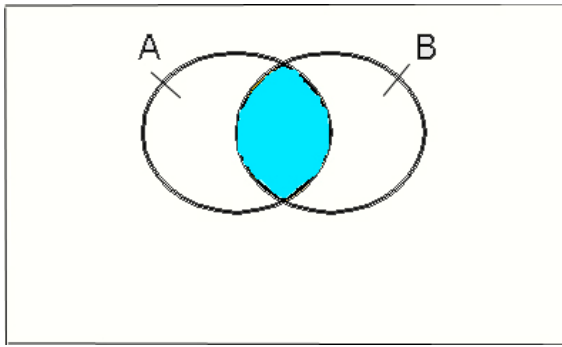
### Aufgabe 2-3 (Venn-Diagramme / Mengenoperationen)

3 P.

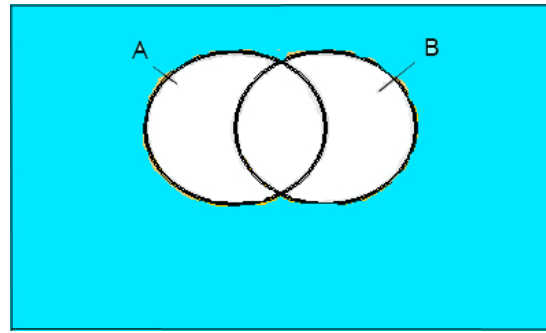
Venn-Diagramme werden zur Veranschaulichung von Mengenoperationen verwendet. Diese bestehen aus einem Rechteck, in dem Mengen A, B, C, ... als Kreise oder Ellipsen dargestellt sind. Das Rechteck repräsentiert eine Grundgesamtheit, von der die eingezeichneten Mengen Teilmengen sind.

Nachstehend sind vier Venn-Diagramme abgebildet, die sich auf die Verknüpfung von zwei Mengen A und B oder drei Mengen A, B und C beziehen. Die Verknüpfungen erfolgen über die Symbole „ $\cup$ “ (Vereinigung von Mengen) und „ $\cap$ “ (Schnittmengenbildung) und „ $\neg$ “ (Komplementärmenge einer Menge, auch durch das Zeichen „ $\bar{\phantom{x}}$ “ über der betreffenden Menge darstellbar).

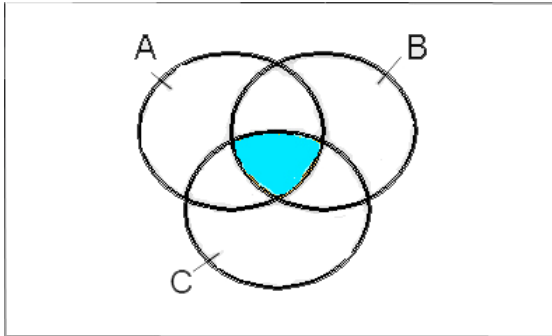
Geben Sie unter Verwendung der genannten Symbole wieder, was in den letzten drei Venn-Diagrammen durch die markierten Flächen dargestellt ist. Beim ersten Venn-Diagramm ist die Lösung schon vorgegeben.



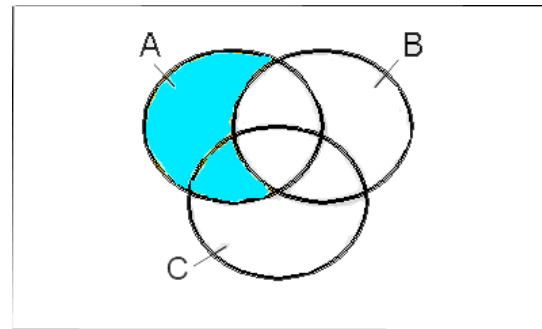
Lösung:  $A \cap B$



Lösung:  $\neg(A \cup B)$



Lösung:  $A \cap B \cap C$



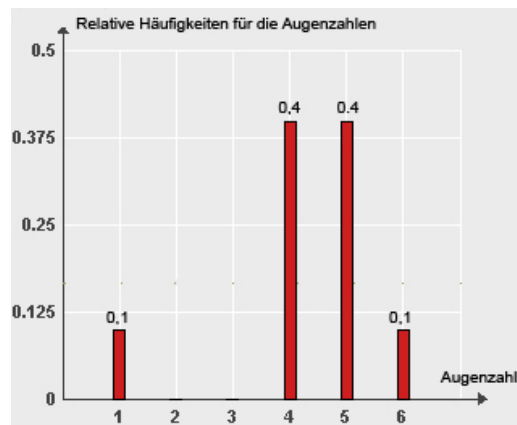
Lösung:  $A \cap \neg B$

Hinweis zur Lösung von Aufgabe 2-3: s. Kurs 33207, Abschnitt 5.2.

**Aufgabe 2-4** (Lage- und Streuungsparameter)

**13 P.**

Nachstehend ist das Ergebnis eines Würfelexperimentes visualisiert, bei dem 10 Mal nacheinander mit einem Würfel gewürfelt wurde. Dargestellt sind die relativen Häufigkeiten für die sechs möglichen Ausgänge eines Wurfs.



a) Vervollständigen Sie die mittlere Zeile in der folgenden Tabelle:

(1 P.)



	Beobachtete Augenzahl					
	1	2	3	4	5	6
Absolute Häufigkeit	1	0	0	4	4	1
Relative Häufigkeit	0,1	0	0	0,4	0,4	0,1

b) Geben Sie auch die alle 10 Beobachtungen umfassende Urliste an, wobei die Werte der Urliste nach zunehmender Größe zu ordnen sind: (1 P.)

Geordnete Urliste:

1	4	4	4	4	5	5	5	5	6
---	---	---	---	---	---	---	---	---	---

c) Berechnen Sie dann auf der Basis der beim Experiment beobachteten Werte den Median  $x_{med}$  und den Mittelwert  $\bar{x}$ . (2 P.)

Median:

$$x_{med} = \frac{1}{2} (4 + 5) = 4,5 \quad (n = 10 \text{ ist gerade})$$

Mittelwert:

$$\bar{x} = \frac{1}{10} (1 \cdot 1 + 4 \cdot 4 + 4 \cdot 5 + 1 \cdot 6) = \frac{1}{10} (1 + 16 + 20 + 6) = \frac{1}{10} \cdot 43 = 4,3$$

d) Berechnen Sie für den per Experiment generierten Datensatz auch die Spannweite und die Standardabweichung. (3 P.)

Spannweite:

$$r = 6 - 1 = 5$$

Standardabweichung:

$$s^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - 4,3)^2 = \frac{1}{10} [(-3,3)^2 + 4 \cdot (-0,3)^2 + 4 \cdot 0,7^2]$$

$$= \frac{1}{10} (10,89 + 0,36 + 1,96) = 1,321$$

e) Die Augenzahl beim einmaligen Wurf mit einem Würfel ist eine Zufallsvariable X. Berechnen Sie den Erwartungswert  $E(X)$  und die Varianz  $Var(X)$  dieser Zufallsvariablen.

(4 P.)

Erwartungswert:

Die Variable  $X$  hat 6 Ausprägungen  $x_i = i$  ( $i = 1, 2, \dots, 6$ ). Die Eintrittswahrscheinlichkeit  $p_i$  für das Ereignis  $x_i = i$  ist  $\frac{1}{6}$ . Es gilt daher

$$E(X) = \sum_{i=1}^6 p_i x_i = \frac{1}{6} \sum_{i=1}^6 x_i = \frac{1}{6} \sum_{i=1}^6 i = \frac{1}{6} \cdot 21 = 3,5$$

Varianz:

$$\text{Var}(X) = \sum_{i=1}^6 p_i (x_i - \mu)^2 = \frac{1}{6} \sum_{i=1}^6 (i - 3,5)^2 = \frac{1}{6} [ (-2,5)^2 + (-1,5)^2 + (-0,5)^2 + 0,5^2 + 1,5^2 + 2,5^2 ]$$

$$= \frac{1}{6} (0,5^2 + 1,5^2 + 2,5^2) = \frac{35}{12} \approx 2,917$$

f) Die Abbildung zu Beginn dieser Aufgabe zeigte das Ergebnis eines Würfelexperimentes, das aus  $n = 10$  aufeinanderfolgenden Würfeln eines Würfels bestand. Wie würde sich der aus den  $n$  beobachteten Augenzahlen errechnete Mittelwert  $\bar{x}$  verhalten, wenn man den Wert  $n$  bei diesem Experiment immer weiter erhöht? (2 P.)

Antwort (nur ein Satz!):

Der Mittelwert strebt dann gegen den Erwartungswert  $\mu = E(X)$  des Merkmals „Augenzahl  $X$  beim einmaligen Wurf mit einem Würfel“, also gegen 3,5.

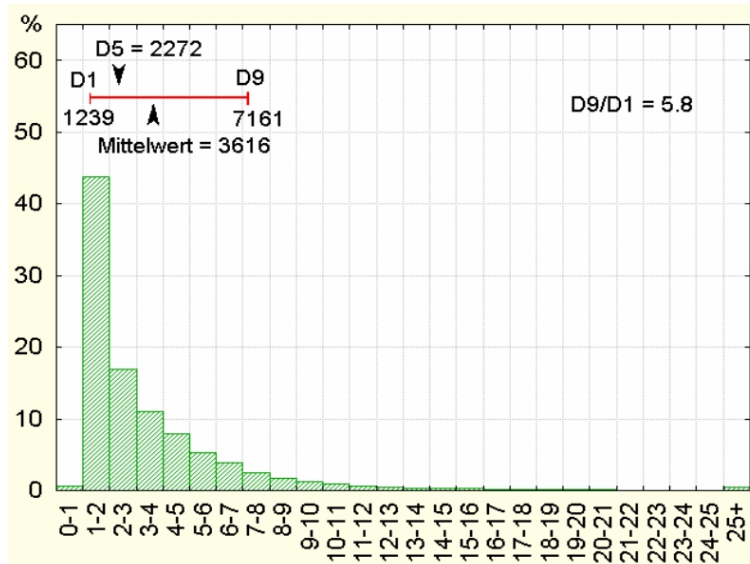
Hinweis zur Lösung von Aufgabe 2-4: s. Kurs 33207, Abschnitte 13.1 und 13.2 (Aufgabenteile a –d) sowie Abschnitte 18.1 und Kap. 20 (Aufgabenteile e – f).

### Aufgabe 2-5 (Darstellung und Beschreibung von Verteilungen)

7 P.

Die folgende Abbildung zeigt für das Referenzjahr 2002 Jahresbruttoverdienste von Arbeitnehmern in Lettland (über eine Million Arbeitnehmer im Bereich „Industrie und Dienstleistungen, offizielle Daten von Eurostat). Auf der horizontalen Achse sind Einkommensklassen wiedergegeben (jeweils in 1000 €), auf der vertikalen Achse die Besetzungshäufigkeiten für die einzelnen Einkommensklassen (relative Häufigkeiten in Prozent). Man erkennt z. B., dass fast 45 % der Arbeitnehmer Jahresbruttoverdienste hatten, die zwischen 1000 € und 2000 € lagen.

In der Abbildung sind auch drei aus den Originaldaten errechneten Quantile wiedergegeben, nämlich das 10%-Quantil (hier abgekürzt mit D1), das 90%-Quantil (mit D9 abgekürzt) und das 50%-Quantil, also der Median (hier mit D5 angesprochen). Außerdem ist der - ebenfalls aus den Originaldaten errechnete - Mittelwert  $\bar{x}$  ausgewiesen.



a) Wie heißt das hier zur Visualisierung der Einkommensverteilung verwendete grafische Instrument? (Die über der Grafik wiedergegebenen numerischen Informationen können Sie bei der Beantwortung dieser Frage ignorieren.) (1 P.)

Antwort (nur ein Wort oder ein Satz!): **Histogramm**

b) Charakterisieren Sie die Form der Einkommensverteilung mit einem geeigneten Begriff.

(1 P.)

Antwort: **Deutlich linkssteile (oder rechtsschiefe) Verteilung.**

c) Das durch das 10% - und 90%-Quantil definierte Intervall gibt zusammen mit dem Median bereits wesentliche Charakteristika des umfangreichen Originaldatensatzes wieder. Man könnte anstelle dieses Intervalls auch einen Boxplot zur Beschreibung wesentlicher Eigenschaften des Datensatzes verwenden. Welche Größen werden benötigt, um einen Boxplot zeichnen zu können? (2 P.)

Antwort (nur Stichworte):

**Neben dem Median werden für die Konstruktion eines Boxplots die 25%- und 75%-Quantile und die Extremwerte (Minimum, Maximum) des Datensatzes benötigt.**

d) Grundsätzlich könnte man Jahresbruttoverdienste auch auf Euro und Cent genau ausweisen (ohne jede Rundung oder Zusammenfassung von Daten zu Einkommensklassen) und die Daten anhand eines Stabdiagramms darstellen. Welcher Nachteil wäre damit im Vergleich zur oben gewählten Visualisierungsform verbunden? (1 P.)

Antwort (nur ein Satz):

**Ein Stabdiagramm wäre viel zu unübersichtlich. (Für den Bereich von 0 – 25 könnten bis zu 2500 Stäbe auftreten!)**

e) Neben dem Median oder Mittelwert könnten man zur Beschreibung von Einkommensdaten auch den Modus verwenden. Wie ist dieser definiert? Ist der Modus auch im Falle ordinal- und nominalskalierter Merkmale erklärt? (2 P.)

Definition des Begriffs „Modus“ (nur ein Satz!):

Am häufigsten vorkommender Wert eines Datensatzes.

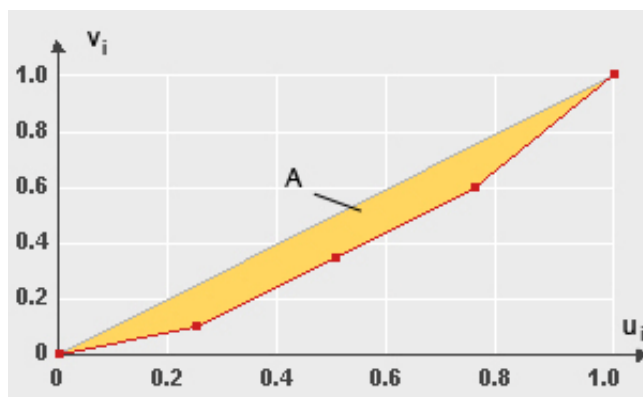
Modus ist auch für ordinal- und nominalskalierte Merkmale erklärt: ja (  ) nein (  )

Hinweis zur Lösung von Aufgabe 2-5: s. Kurs 33207, Abschnitt 13.2.

### Aufgabe 2-6 (Konzentrationsmessung)

9 P.

In einem EU-Staat konkurrieren vier Firmen im Bereich der Telekommunikation. Es seien  $x_1 = 20$ ,  $x_2 = 50$ ,  $x_3 = 50$  und  $x_4 = 80$  die Umsätze dieser Firmen im letzten Geschäftsjahr (Umsätze jeweils in Millionen €). 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 und ihre Koordinaten neben der Grafik wiedergegeben.



$i$	$u_i$	$v_i$
0	0	0
1	0,25	0,10
2	0,50	0,35
3	0,75	0,60
4	1	1

a) Berechnen Sie den Gini-Koeffizienten und den normierten Gini-Koeffizienten. Bezeichnen Sie ersteren mit  $G$  und letzteren mit  $G^*$ . (5 P.)

Berechnung von  $G$ :

$$G = \sum_{i=1}^4 v_i u_{i-1} - \sum_{i=1}^4 v_{i-1} u_i \text{ mit } u_0 = v_0 = 0 \text{ und } u_4 = v_4 = 1$$

$$G = (0,35 \cdot 0,25 + 0,6 \cdot 0,5 + 0,75) - (0,1 \cdot 0,5 + 0,35 \cdot 0,75 + 0,6)$$

$$= 1,1375 - 0,9125 = 0,225$$

Berechnung von  $G^*$ :

$$G^* = \frac{4}{4-1} \cdot G = \frac{4}{3} \cdot 0,225 = 0,3$$

- b) In der Grafik ist neben der Lorenzkurve noch eine durchgezogene Kurve eingezeichnet, die die Punkte (0;0) und (1;1) direkt verbindet. Wie hängt die mit A bezeichnete Fläche zwischen Lorenzkurve und durchgezogener Kurve mit dem Gini-Koeffizienten zusammen?

(1 P.)

Antwort (nur ein kurzer Satz oder eine Gleichung):

Es gilt  $G = 2A$ .

- c) Wie müsste man die Umsätze  $x_1$ ,  $x_2$ ,  $x_3$  und  $x_4$  der vier Firmen definieren, wenn man erreichen will, dass die Lorenzkurve mit der durchgezogenen Linie übereinstimmt?

(2 P.)

Antwort (nur ein Satz) :

Die Umsätze müssten  $x_1$ ,  $x_2$ ,  $x_3$  und  $x_4$  identisch sein (fehlende Konzentration).

- c) Welchen Wert nimmt der Gini-Koeffizient bei der Aufgabenteil c beschriebenen Situation an? (1P.)

Antwort (nur ein Satz oder eine Gleichung):

Der Gini-Koeffizient (unnormiert oder normiert) ist dann 0.

Hinweis zur Lösung von Aufgabe 2-7: s. Kurs 33207, Abschnitt 13.4.

**Aufgabe 2-7** (Zusammenhangsmessung bei nominalskalierten Merkmalen)

**13 P.**

Bei einer Kommunalwahl in einer Großstadt geht es um die Besetzung des Oberbürgermeisteramts. Zwei Kandidaten A und B stellen sich zur Wahl. Vor der Wahl werden  $n = 642$  wahlberechtigte Männer und Frauen nach ihrer Präferenz befragt. Die Ergebnisse sind in der folgenden Vierfeldertafel zusammengefasst:

	Kandidat A	Kandidat B	Zeilensumme
Männer	144	153	297
Frauen	200	145	345
Spaltensumme	344	298	642

- a) Berechnen Sie auf 3 Dezimalstellen genau den  $\chi^2$ -Koeffizienten ( $Ch^2$ -Koeffizienten).

Lösung:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(h_{ij} - e_{ij})^2}{e_{ij}} = \frac{(h_{11} - e_{11})^2}{e_{11}} + \frac{(h_{12} - e_{12})^2}{e_{12}} + \frac{(h_{21} - e_{21})^2}{e_{21}} + \frac{(h_{22} - e_{22})^2}{e_{22}}$$

$$\text{mit } e_{ij} = \frac{1}{642} h_i \cdot h_j, \text{ also } e_{11} = \frac{297 \cdot 344}{642} = 159,14, e_{12} = \frac{297 \cdot 298}{642} = 137,86,$$

$$e_{21} = \frac{345 \cdot 344}{642} = 184,14, e_{22} = \frac{345 \cdot 298}{642} = 160,14$$

$$\chi^2 = \frac{(144 - 159,14)^2}{159,14} + \frac{(153 - 137,86)^2}{137,86} + \frac{(200 - 184,14)^2}{184,14} + \frac{(145 - 160,14)^2}{160,14}$$

$$= 1,440 + 1,663 + 1,240 + 1,431 = 5,774$$

b) Berechnen Sie dann auch auf 3 Dezimalstellen genau den Kontingenzkoeffizienten  $C$  sowie den korrigierten Kontingenzkoeffizienten  $C_{\text{korr}}$ . (4 P.)

Berechnung von  $C$ :

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{5,774}{5,774 + 642}} = 0,094;$$

Berechnung von  $C_{\text{korr}}$ :

$$C_{\text{korr}} = \frac{C}{C_{\text{max}}} = \frac{0,094}{0,707} \approx 0,133 \quad \text{weil } C_{\text{max}} = \sqrt{\frac{q-1}{q}} = \sqrt{0,5} \approx 0,707$$

c) Welche Unter- und Obergrenze gibt es für den korrigierten Kontingenzkoeffizienten, d. h. welche Schranken kann  $C_{\text{korr}}$  generell nicht unter- bzw. überschreiten? (2 P.)

Antwort: Die Untergrenze von  $C_{\text{korr}}$  ist gegeben durch **0**, die Obergrenze ist durch **1**

**(kein Zusammenhang:  $C_{\text{korr}} = 0$ ; maximaler Zusammenhang:  $C_{\text{korr}} = 1$ )**

d) Was lässt sich aus dem in Aufgabenteil b errechneten Ergebnis zum Zusammenhang zwischen den Merkmalen „Geschlecht der befragten Person“ und „Kandidatenpräferenz“ aussagen? (1 P.)

Antwort: **Der Zusammenhang ist nur schwach ausgeprägt.**

**Hinweis zur Lösung von Aufgabe 2-7: s. Kurs 33207, Abschnitt 14.2.**

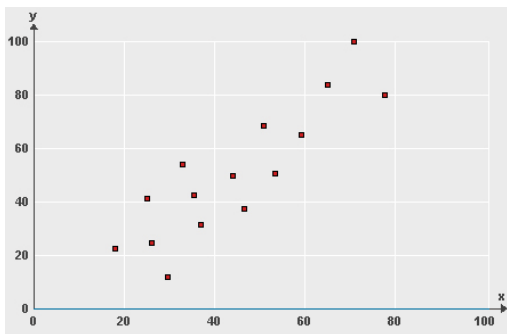
**Aufgabe 2-8** (Zusammenhangsmessung bei ordinal- und intervallskalierten Merkmalen)

**10 P.**

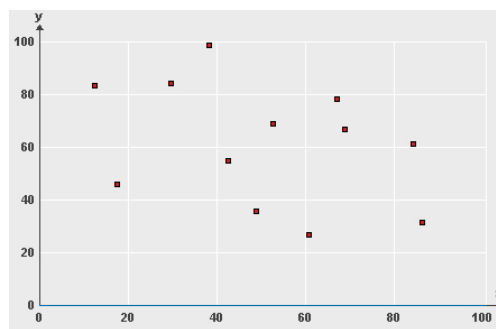
a) Für zwei intervallskalierte Variablen X und Y liegen Beobachtungen in Form von Wertepaaren  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , vor, die in den folgenden zwei Grafiken anhand je eines Streudiagramms veranschaulicht seien. Geben Sie an, welche Größenordnung der Korrelationskoeffizient r nach Bravais-Pearson in den beiden Abbildungen hat. Verwenden Sie dabei Codes für die folgenden Bereiche:

- |   |   |
|---|---|
| 1 = Korrelationskoeffizient r ist genau - 1 | 4 = r liegt im Bereich von 0 bis 0,5      |
| 2 = r liegt im Bereich von -1 bis -0,5      | 5 = r ist größer als 0,5                  |
| 3 = r liegt im Bereich von -0,5 bis 0       | 6 = Korrelationskoeffizient r ist genau 1 |

(4 P.)



Code: **5**



Code: **3**

b) Welche der drei Aussagen sind wahr, welche falsch. Tragen Sie „w“ bzw. „f“ ein. (3 P.)

Der Korrelationskoeffizient r nach Bravais-Pearson misst bei zwei intervallskalierten Merkmalen die Stärke eines linearen Zusammenhangs. ( **f** )

Wenn r für zwei intervallskalierte Merkmalen den Wert Null annimmt, kann durchaus ein nicht-linearer Zusammenhang vorliegen. ( **f** )

Wenn r nahe bei 1 liegt, ist zwischen den Merkmalen stets ein starker kausaler Zusammenhang gegeben. ( **f** )

Bepunktung bei Aufgabenteil b: Je 1 P. pro zutreffender Antwort. Bei unzutreffenden Antworten wird zur Vermeidung von Ratesstrategien je ein Punkt abgezogen. Die Gesamtpunktzahl für Aufgabenteil b darf dabei aber 0 P. nicht unterschreiten.

d) Es sei angenommen, dass zwei unabhängige Kreditsachbearbeiter die Kreditwürdigkeit von fünf Sparkassenkunden – hier mit 1 bis 5 nummeriert - anhand einer 10-stufigen Ratingskala bewerten, bei der die Punktzahl 1 sehr schlechte und die Punktzahl 10 sehr gute Bonität bezeichne. Die Ergebnisse sind nachstehend ausgewiesen:

Nr des Kunden	Beurteilung durch Sachbearbeiter A	Beurteilung durch Sachbearbeiter B
1	5	6
2	8	9
3	9	7
4	2	4
5	6	5

Berechnen Sie den Rangkorrelationskoeffizienten  $r_s$  nach Spearman.

(3 P.)

Lösung: Lösung: Zunächst werden Rangplätze und daraus Rangdifferenzen ermittelt:

Nr i des Kunden	Beurteilung durch Sachbearbeiter A	Rang des Kunden i (Sachbearbeiter A)	Beurteilung durch Sachbearbeiter B	Rang des Kunden i (Sachbearbeiter B)	Rangdifferenz $d_i$
1	5	4	6	3	1
2	8	2	9	1	1
3	9	1	7	2	-1
4	2	5	4	5	0
5	6	3	5	4	-1

Hieraus folgt dann für das Zusammenhangsmaß  $r_s$  :

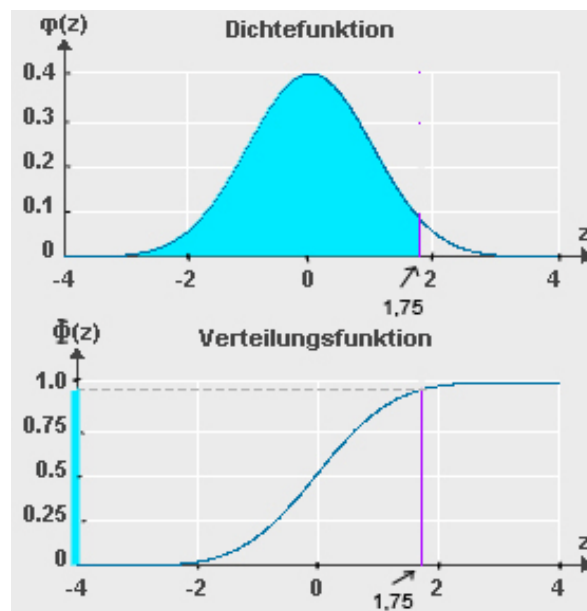
$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^5 d_i^2}{5(5^2 - 1)} = 1 - \frac{6 \cdot 4}{120} = 0,8$$

Hinweis zur Lösung von Aufgabe 2-9: s. Kurs 33207, Abschnitte 14.3 – 14.4.

#### Aufgabe 2-9 (Dichte- und Verteilungsfunktion der Standardnormalverteilung)

4 P.

Nachstehend sind die Dichtefunktion der Standardnormalverteilung und die Verteilungsfunktion der Standardnormalverteilung wiedergegeben. Erstere ist hier mit  $\varphi(z)$ , letztere mit  $\Phi(z)$  bezeichnet. Auf der z-Achse ist jeweils der Punkt  $z = 1,75$  markiert. Einer hier nicht wiedergegebenen Tabelle der Standardnormalverteilung kann man entnehmen, dass  $\Phi(1,75) = 0,9599$  gilt. (Dieser Wert entspricht im oberen Teil der Abbildung dem Inhalt der markierten Fläche unter der Dichtekurve und im unteren Teil der Abbildung der Länge des auf der vertikalen Achse markierten Abschnitts, der von 0 bis zur gestrichelten Linie reicht.)



a) Welchen Wert nimmt die Verteilungsfunktion  $\Phi(z)$  an der Stelle  $z = -1,75$  an?



(2 P.)

Lösung:

Da der Fläche unter der gesamten Dichtekurve den Inhalt 1 hat und gilt bei Beachtung der Symmetrie der Dichtekurve bezüglich  $z = 0$

$$\Phi(-1,75) = 1 - \Phi(1,75) = 1 - 0,9599 = 0,0401.$$

- b) Berechnen Sie die Wahrscheinlichkeit dafür, dass eine standardnormalverteilte Variable  $Z$  Werte zwischen  $z = -1,75$  und  $z = 1,75$  annimmt. (2 P.)

Lösung:

$$P(-1,75 \leq Z \leq 1,75) = \Phi(1,75) - \Phi(-1,75) = 0,9599 - 0,0401 = 0,9198.$$

Hinweis zur Lösung von Aufgabe 2-9: s. Kurs 33207, Abschnitt 19.2.