

Alle Zitate (Seitenzahlen in Klammern) sind entnommen aus: Mittag, H.-J. (2011). *Statistik. Eine interdisziplinäre Einführung. Kurseinheit 1: Beschreibende Statistik*. Studienbrief 33209. Hagen: FernUniversität.

## Kurseinheit 1: Beschreibende Statistik

### Kap. 1: Statistik, Daten und statistische Methoden

- 1A) Harmonisierte Daten sind Zahlen im positiven Bereich.
- B) Statistik umfasst die Bereiche deskriptive und deduktive Statistik.
- C) Evidence Based Decision Making bedeutet datengestützte Entscheidungsfindung.
- D) Die explorative Datenanalyse ging aus der beschreibenden Statistik hervor.
- E) Ziel der beschreibenden Statistik ist die Reduktion von statistischen Informationen durch Aggregation. Dabei sollen relevante Informationen möglichst erhalten bleiben.

Lösung: C (S. 7), D (S. 6), E (S.6)

zu A) Harmonisierte Daten sind (z. B. über Ländergrenzen) vergleichbare Daten. (S. 4)

zu B) Statistik unterscheidet die Bereiche deskriptive und **induktive** Statistik. (S. 5f)

### Kap. 2: Grundbegriffe der Statistik

- 2A) Merkmalsträger sind die Objekte, auf die sich die statistische Untersuchung bezieht.
- B) Statistische Einheiten sind Merkmalsträger.
- C) Die Grundgesamtheit wird auch Teilpopulation genannt.
- D) Teilmengen von Grundgesamtheiten nennt man auch Population.
- E) Die Grundgesamtheit muss klar abgegrenzt werden.

Lösung: A (12), B (12), E (12)

zu C) Die **Grundgesamtheit** wird **Population** genannt. (12)

Zu D) **Teilmengen** von Grundgesamtheiten nennt man auch **Teilpopulation**. (12)

- 3A) Merkmale sind Variablen.
- B) Variablen sind Eigenschaften statistischer Einheiten.
- C) Merkmalsausprägungen sind die möglichen Werte, die ein Merkmal annehmen kann.
- D) Der Familienstand einer Person ist eine Merkmalsausprägung.
- E) Eine Stichprobe ist eine Teilmenge der Grundgesamtheit, die nach einem bestimmten Verfahren ausgewählt wurde.

Lösung: A (12), B (12), C (12), E (12)

zu D) Der Familienstand einer Person bezeichnet ein **Merkmal**. Ausprägungen können hier z. B. sein „ledig“, „verheiratet“, „verwitwet“, „geschieden“ usw. (13/Bsp. 2.2)

- 4A) Rohdaten sind die Werte, die man für statistische Einheiten beobachtet.
- B) Eine Urliste ist eine Liste, die alle Urwerte zusammenfasst.
- C) Primärdaten dürfen nicht in der Urliste zusammengefasst werden.
- D) In der Urliste dürfen Merkmalswerte nicht mehrfach auftreten.
- E) Merkmale lassen sich in Typen einteilen.

Lösung: B (12), E (13)

zu A) Rohdaten sind die Werte, die man für ein **Merkmal** beobachtet. (12)

Zu C) Primärdaten sind Urwerte und gehören in die Urliste. (12)

Zu D) In der Urliste können Merkmalswerte mehrfach auftreten. (12)

5A) Ein diskretes Merkmal ist dadurch gekennzeichnet, dass die Ausprägungen ein Intervall bilden.

B) Ein diskretes Merkmal kann nur endlich viele Ausprägungen annehmen.

C) Körper- bzw. Wohnungsgröße sind stetige bzw. quasi-stetige Merkmale.

D) Nicht jedes stetige Merkmal kann durch Rundung oder Gruppierung in diskrete Variablen überführt werden.

E) Durch Gruppierung von Daten gehen Informationen verloren.

Lösung: B (14), C(14/Fußnote 2), E (14)

zu A) Ein **stetiges** Merkmal ist dadurch gekennzeichnet, dass die Ausprägungen ein Intervall bilden. (14)

zu D) **Jedes** stetige Merkmal kann durch Rundung oder Gruppierung in diskrete Variablen überführt werden. (14)

6A) Klassenbildung erhöht bei größeren Datensätzen mit stetigen Merkmalen die Übersichtlichkeit.

B) Klassenbildung erhöht die Differenzierung innerhalb der Klassen.

C) Alter und Familienstand einer Person sind stetige Merkmale.

D) Zeitangaben, Längen und Gewichte sind stetige Merkmale.

E) Das Körpergewicht (von Frauen) ist ein diskretes Merkmal.

Lösung: A (14), D (14/Bsp 2.3)

zu B) Klassenbildung **vermindert** die Differenzierung innerhalb der Klassen. (14)

zu C) Alter und Familienstand einer Person sind **diskrete** Merkmale. (14/Bsp 2.3)

zu E) Das Körpergewicht einer Person ist ein **stetiges** Merkmal. (14/Bsp 2.3)

7A) Studienfächer (wie Biwi, POL/SOZ und Psych.) werden auf einer Ordinalskala erfasst.

B) Schulnoten werden auf der Rangskala erfasst.

C) Das Merkmal „Anzahl der Fachsemester“ wird auf der Absolutskala erfasst.

D) Das Merkmal „Geburtsjahr“ ist ordinalskaliert.

E) Das Merkmal „Lebensalter“ ist verhältnisskaliert.

Lösung: B (15), C (15), E (16/Bsp 2.4)

zu A) Studienfächer werden auf einer **Nominalskala** erfasst. (14)

zu D) Das Merkmal „Geburtsjahr“ ist **intervallskaliert**. (16/Bsp 2.4)

8A) Qualitative Merkmale können nominal- oder ordinalskaliert sein.

B) Für qualitative Merkmale können nur Begriffe verwendet werden.

C) Für quantitative Merkmale werden nur „echte“ Zahlen verwendet.

D) Die Güteklassen von Lebensmitteln (1, 2 und 3) sind quantitative Merkmale.

E) Metrisch skalierte Merkmale sind stets quantitativ.

Lösung: A (16), C (16), E (16)

zu B) Für qualitative Merkmale können **Begriffe und Zahlen** verwendet werden. Mit den Zahlen kann nicht im üblichen Sinne gerechnet werden. (16)

zu D) Die Güteklassen von Lebensmitteln (1, 2 und 3) sind **qualitative** Merkmale. (16)

9A) Die Festlegung von Messanweisungen bezeichnet man als Operationalisierung.

B) Latente Variablen sind nicht direkt beobachtbar.

C) Intelligenz und Leistungsmotivation sind direkt beobachtbare Variablen.

D) Durch Operationalisierung soll ein Messverfahren festgelegt werden, mit dem sich Merkmale qualifizieren lassen.

E) Für hypothetische Konstrukte gibt es jeweils nur eine Möglichkeit der Operationalisierung.

Lösung: A (16), B (16)

zu C) Intelligenz und Leistungsmotivation sind **nicht** direkt beobachtbare Variablen. (16)

Zu D) Durch Operationalisierung soll ein Messverfahren festgelegt werden, mit dem sich Merkmale **quantifizieren** lassen. (16)

zu E) Für hypothetische Konstrukte gibt es **mehrere** Möglichkeiten der Operationalisierung. (16)

10A) Die Qualität von Messverfahren wird anhand der Kriterien Objektivität, Rentabilität und Validität beurteilt.

B) Objektivität gewährleistet die intersubjektive Nachvollziehbarkeit des Verfahrens.

C) Reliabilität (= Messgenauigkeit) bezieht sich auf die technische Ebene des Messinstrumentes.

D) Validität (= Gültigkeit des Messverfahrens) bezieht sich auf den inhaltlichen Aspekt der Messung.

E) Ein reliables Messverfahren ist immer valide.

Lösung: B (16), C (17), D (17)

zu A) Die Qualität von Messverfahren wird anhand der Kriterien Objektivität, **Reliabilität** und Validität beurteilt. (16)

zu E) Ein reliables Messverfahren muss nicht valide sein! (17)

Zwischenauswertung: mögliche Punktzahl: 50

erreichte Punkte: .....

11A) Proxyvariablen sind eine näherungsweise verwendbare beobachtbare Variable.

B) Das Bruttoeinkommen eines Arbeitnehmers ist eine nicht direkt beobachtbare Variable.

C) Bei Personenbefragungen oder Verhaltensbeobachtungen können die Ergebnisse nur qualitativ ausgewertet werden.

D) Unter Meta-Daten werden besonders große Datensätze verstanden.

E) EU-Verordnung zur Harmonisierung von Daten regeln, welche Komponenten zu einer Variablen gehören.

Lösung: A (17/Bsp 2.5), E (17)

zu B) Das Bruttoeinkommen eines Arbeitnehmers ist eine **direkt** beobachtbare (= manifeste) Variable. (17)

Zu C) Bei Personenbefragungen oder Verhaltensbeobachtungen gilt es, die Ergebnisse zu quantifizieren. (17/Bsp 2.5)

zu D) Meta-Daten legen den methodischen Hintergrund und eventuelle Besonderheiten der Datenerfassung offen. (18/Bsp2.6)

### Kap. 3: Datengewinnung und Auswahlverfahren

- 12A) Unter Erhebungsdesign wird ein optisch gelungener Fragebogen verstanden.
- B) Die Gewinnung von Daten bezeichnet man als Datenerhebung.
- C) In der Psychologie werden in den meisten Fällen Sekundärerhebungen durchgeführt.
- D) Bei Tertiärerhebungen werden statistische Informationen aus vorhandenen Quellen in aggregierter Form geschöpft.
- E) In den Sozialwissenschaften sind Primärerhebungen von besonderer Bedeutung.

Lösung: B (19), D (19), E (20)

- zu A) Unter Erhebungsdesign wird die **Planung** der Datengewinnung verstanden. (19)
- zu C) In der Psychologie werden in den meisten Fällen **Primärerhebungen** durchgeführt. (19/Fußnote 1)

- 13A) Die Befragungen ist das dominierende Instrument der naturwissenschaftlichen Forschung.
- B) Unstrukturierte Befragungen können ohne Fragebogen realisiert werden.
- C) CAPI und CATI sind computergestützte Interviewverfahren.
- D) Nimmt der Befragte die Antworteingabe am PC selbst vor, spricht man von administrierter computergestützten Befragung.
- E) Die Form der Befragung ist unabhängig von der Größe des zu gewinnenden Datensatzes.

Lösung: B (20), C (20f)

- zu A) Die Befragungen ist das dominierende Instrument sozialwissenschaftlichen Forschung. (20)
- zu D) Nimmt der Befragte die Antworteingabe am PC selbst vor, spricht man von **selbst-administrierter computergestützten Befragung**. (21)
- Zu E) Die Form der Befragung ist **abhängig u. a.** von der Größe des zu gewinnenden Datensatzes. (21)

- 14A) Die Beobachtung ist ein kaum verbreitetes Verfahren der Datenerhebung.
- B) Die Beobachtung als Verfahren kann sich nur auf das Verhalten von Personen beziehen.
- C) In den Sozialwissenschaften geht es um die Beobachtung von Einzelpersonen oder Gruppen.
- D) In der empirischen Sozialforschung können Beobachtungen seltener Ereignisse spontan durchgeführt werden.
- E) Logfile-Analysen sind nicht-reaktive Erhebungsverfahren zur Untersuchung des Verhaltens von Internetnutzern.

Lösung: C (22), E (22)

- zu A) Die Beobachtung ist ein **sehr** verbreitetes Verfahren der Datenerhebung. (22)
- Zu B) Die Beobachtung als Verfahren kann sich auf ganz unterschiedliche Objekte beziehen. (22)
- zu C) In der empirischen Sozialforschung werden Beobachtungen systematisch **geplant**. (22)

15A) Experimente dienen der empirischen Überprüfung von Hypothesen über kausale Zusammenhänge zwischen Merkmalen.

- B) In den Sozialwissenschaften sind Experimente kaum verbreitet.
- C) Bei einem Experiment werden die abhängigen Variablen variiert.
- D) Aus einem beobachteten Zusammenhang zwischen zwei Merkmalen kann in der sozialwissenschaftlichen Forschung nicht zwingend auf einen Kausalzusammenhang geschlossen werden.
- E) Bei einem Quasi-Experiment werden die Personen den Gruppen zufällig zugeordnet.

Lösung: A (23), B (23), D (23)

zu C) Bei einem Experiment werden die **unabhängigen** Variablen (= Einflussfaktoren) variiert. (23)

Zu E) Bei einem Quasi-Experiment erfolgt die Zuordnung **nicht**-randomisiert. (24)

16A) Querschnittsstudien werden an verschiedenen Merkmalsträgern zu einem festen Zeitpunkt erhoben.

B) Die Beobachtung von Merkmalen an verschiedenen statistischen Einheiten im Zeitverlauf nennt man Zeitreihe.

C) Ein Panel kombiniert Querschnittsstudien und Querschnittsreihen.

D) Bei einer Vollerhebung werden alle Elemente einer Grundgesamtheit in die Erhebung einbezogen.

E) ALLBUS basiert auf einem Querschnittsdesign.

Lösung: A (25), D (25), E (26/Bsp 3.3)

zu B) Die Beobachtung von Merkmalen an **einer** statistischen Einheit im Zeitverlauf nennt man Zeitreihe. (25)

zu C) Ein Panel kombiniert Querschnitts- und Zeitreihenanalysen. (25)

17A) Die Auswahlpopulation ist die Population, aus der eine Stichprobe gezogen wird.

B) Enthält die Auswahlpopulation Elemente, die nicht zur definierten Grundgesamtheit gehören, wird das Undercoverage genannt.

C) Zufallsstichproben sollen ein repräsentatives Abbild der Grundgesamtheit darstellen.

D) Als Inferenzschluss wird der Rückschluss von Eigenschaften der Grundgesamtheit auf Eigenschaften der Stichprobe bezeichnet.

E) Bei einer einfachen Zufallsstichprobe besitzt jede Teilmenge der Grundgesamtheit dieselbe Auswahlwahrscheinlichkeit.

Lösung: A (27), C (27), E (27)

zu B) Enthält die Auswahlpopulation Elemente, die nicht zur definierten Grundgesamtheit gehören, wird das **Overcoverage** genannt. (27)

zu D) Als Inferenzschluss wird der Rückschluss **von** Eigenschaften der **Stichprobe auf** Eigenschaften der **Grundgesamtheit** bezeichnet. (27)

18A) Das Verfahren der einfachen Zufallsstichprobe entspricht dem Urnenmodell.

B) Disjunkte Teilgesamtheiten überlappen sich.

C) Die Schichten einer Stichprobe sollen in Bezug auf das zu untersuchende Merkmal in sich möglichst heterogen sein.

D) Die Schichten einer Stichprobe sollen in Bezug auf das zu untersuchende Merkmal untereinander möglichst heterogen sein.

E) Bei zweistufigen Verfahren wird nur aus der zweiten Schicht eine Zufallsstichprobe gezogen.

Lösung: A (27), D (28)

zu B) Disjunkte Teilgesamtheiten überlappen sich **nicht**. (27)

Zu C) Die Schichten einer Stichprobe sollen in Bezug auf das zu untersuchende Merkmal in sich möglichst **homogen** sein. (27)

Zu E) Bei zweistufigen Verfahren wird aus **jeder** Schicht eine Zufallsstichprobe gezogen. (28)

19A) Disproportionale Schichtung wird oft angewendet, wenn Schichten sehr stark besetzt sind.

B) Eine Schichtungsvariable ist ein Hilfsmerkmal zur Zerlegung der Grundgesamtheit in Schichten.

C) Zerfällt eine Grundgesamtheit in „natürliche“ disjunkte Teilgesamtheiten, spricht man von Klumpen.

D) Bei einer Klumpenstichprobe werden im ersten Schritt alle Elemente der ausgewählten Klumpen untersucht.

E) Quotenauswahlen stellen hinsichtlich eines Merkmals eine Art verkleinertes Abbild der Grundgesamtheit dar.

Lösung: B (28), C (29), E (29)

zu A) Disproportionale Schichtung wird oft angewendet, wenn Schichten sehr **dünn** besetzt sind. (28)

zu D) Bei einer Klumpenstichprobe werden im **zweiten** Schritt alle Elemente der ausgewählten Klumpen untersucht. (29)

#### **Kap. 4: Univariate Häufigkeitsverteilungen**

20A) Werden Daten für ein Merkmal ausgewertet, spricht man von multivariater Datenanalyse.

B) Bei stetigen Merkmalen ist das wiederholte Auftreten von Merkmalswerten um so seltener, je ungenauer gemessen wird.

C) Urlisten werden mit sich wiederholenden Merkmalswerten übersichtlicher.

D) Die absolute Häufigkeit bezeichnet die Anzahl der jeweiligen Elemente in der Urliste.

E) Absolute Häufigkeiten haben den Nachteil, dass sie von der Länge  $n$  der Urliste abhängen.

Lösung: D (34), E (34)

zu A) Werden Daten für ein Merkmal ausgewertet, spricht man von **univariater** Datenanalyse. (33)

zu B) Bei stetigen Merkmalen ist das wiederholte Auftreten von Merkmalswerten um so seltener, je **genauer** gemessen wird. (33)

Zu C) Urlisten werden mit sich wiederholenden Merkmalswerten rasch **unübersichtlicher**. (33)

Zwischenauswertung: mögliche Punktzahl: 50

erreichte Punktzahl: .....

- 21A) Werden absolute Häufigkeiten durch den Umfang  $n$  der Beobachtungsreihe geteilt, resultieren relative Häufigkeiten.  
B) Relative Häufigkeiten werden benötigt, um Datensätze unterschiedlichen Umfangs direkt vergleichbar zu machen.  
C) Absolute Häufigkeiten werden oft in Form von Prozentwerten angegeben.  
D) Unter „empirischer Verteilung“ wird die Häufigkeitsverteilung für ein Merkmal  $X$  verstanden.  
E) Die absoluten Häufigkeiten eines Datensatzes ergeben in der Summe 1.

Lösung: A (34), B (34), D (34)

zu C) **Relative** Häufigkeiten werden oft in Form von Prozentwerten angegeben. (34)

Zu E) Es addieren sich die nur die **relativen** Häufigkeiten zu 1. (34)

- 22A) Ein Säulendiagramm wird auch Stabdiagramm genannt.  
B) Bei Kreisdiagrammen hat die dritte Dimension rein dekorativen Charakter.  
C) Säulendiagramme können manipulativ wirken, wenn die vertikale Achse nicht auf dem Niveau 0 beginnt.  
D) Bei klassierten Daten bezieht sich die Häufigkeitsverteilung auf Klassenbesetzungshäufigkeiten.  
E) Ein Histogramm entsteht, wenn die (durch Rechtecke repräsentierten) Besetzungshäufigkeiten direkt aneinander anschließen.

Lösung: B (35), C (35), D (39), E (39)

zu A) Ein Säulendiagramm wird auch **Balkendiagramm** genannt. (34)

- 23A) Die Länge der Rechtecke im Histogramm entspricht (bei gleicher Breite) den Klassenbesetzungshäufigkeiten.  
B) Beim Kreisdiagramm lassen sich Anteile ähnlicher Größe nicht so gut unterscheiden wie beim Stab- oder Säulendiagramm.  
C) Ein Vorteil von Histogrammen ist, dass der optische Eindruck wesentlich von der Klasseneinteilung abhängt.  
D) Kerndichteschätzer können als Verallgemeinerung des Konzepts der Histogramme angesehen werden.  
E) Die Treppenfunktion am oberen Rand eines Histogramms wird bei Kerndichteschätzern durch eine diskrete Variable ersetzt.

Lösung: A (39), B (38), D (40)

zu C) Ein **Nachteil** von Histogrammen ist, dass der optische Eindruck wesentlich von der Klasseneinteilung abhängt. (40)

zu E) Die Treppenfunktion am oberen Rand eines Histogramms wird bei Kerndichteschätzern durch eine **stetige Funktion** ersetzt. (40)

- 24A) Mehrere multivariate Verteilungen können in einer Grafik dargestellt werden.  
B) Das Übereinanderstapeln von Häufigkeiten bei kann auch bei gruppierten Daten als Alternative zum Histogramm angewendet werden.  
C) Bei einer Zerlegung in mehr als drei Komponenten wird ein gestapeltes Säulendiagramm unübersichtlich.  
D) 3D-Säulendiagramme sollte man eher zurückhaltend verwenden.  
E) Für stetige Merkmale kann man die Werte einer Urliste zu  $k$  Klassen zusammenfassen.

Lösung: B (41), D (44), E (45)

zu A) Mehrere **univariate** Verteilungen können in einer Grafik dargestellt werden. (41)

zu C) Bei einer Zerlegung in mehr als **zwei** Komponenten wird ein gestapeltes Säulendiagramm unübersichtlich. (42)

25A) Schwellenwerte finden bei Merkmalswerten Anwendung, die mindestens ordinalskaliert sind.

B) Anhand von kumulierten Häufigkeitsverteilungen lässt sich erkennen, wie viele Werte unter- oder oberhalb des Schwellenwertes liegen.

C) Die relative kumulierte Häufigkeitsverteilung erhält man, indem man die absolute kumulierte Häufigkeitsverteilung durch 100 dividiert.

D) Dividiert man die absolute kumulierte Häufigkeitsverteilung durch den Umfang  $n$  des Datensatzes, erhält man eine empirische Verteilungsfunktion.

E) Kumulierte Häufigkeiten stellen eine monoton fallende Treppenfunktion dar.

Lösung: A (45), B (45), D (46)

zu C) Die relative kumulierte Häufigkeitsverteilung erhält man, indem man die absolute kumulierte Häufigkeitsverteilung **durch den Umfang  $n$  des Datensatzes** dividiert. (45)

zu E) Kumulierte Häufigkeiten stellen eine monoton **steigende** Treppenfunktion dar. (45)

26A) Zwischen zwei benachbarten Ausprägungen von  $X$  ändert sich an der Summe der Häufigkeiten nichts. Die empirische Verteilungsfunktion bleibt dabei ebenfalls auf konstantem Niveau.

B) Absolute und relative Häufigkeitsverteilungen unterscheiden sich nur hinsichtlich der Skalierung der Ordinatenachse.

C) Beim mehrmaligen Würfeln mit einem Würfel kann die Verteilungsfunktion sehr viele Sprünge aufweisen.

D) Bei einem Datensatz für ein diskretes Merkmal mit einer sehr großen Anzahl von Ausprägungen ist die Verteilungsfunktion immer als Sprungfunktion wahrzunehmen.

E) Bei Datensätzen für stetige Merkmale wird die Visualisierung der Häufigkeiten der Daten zweckmäßigerweise in Form von Histogrammen vorgenommen, die die Klassenbesetzungshäufigkeiten darstellen.

Lösung: A (47+Abb. 4.9 [S.46!]), B (47), E (48)

zu C) Beim mehrmaligen Würfeln mit einem Würfel kann die Verteilungsfunktion **nur wenige** Sprünge aufweisen, weil nur wenige verschiedene Ausprägungen beobachtet werden können. (48)

zu D) Bei einem Datensatz für ein diskretes Merkmal mit einer sehr großen Anzahl von Ausprägungen ist die Verteilungsfunktion **kaum noch** als Sprungfunktion wahrzunehmen. Sie erscheint als relativ glatter Kurvenzug.(48)

## **Kap. 5: Kenngrößen univariater empirischer Verteilungen**

27A) Kenngrößen sollen Informationen aus Datensätzen verdichten.

B) Informationsverdichtung ist nicht mit Informationsverlust verbunden.

C) Kenngrößen liefern für einen Datensatz wertvolle zusätzliche Informationen.

D) Zur Charakterisierung des „Zentrums“ einer Verteilung werden Lageparameter herangezogen.



E) Modus ist ein Synonym für Modalwert.

Lösung: A (49), C (49), D (49), E (49)

zu B) Informationsverdichtung ist **grundsätzlich** mit Informationsverlust verbunden. (49)

28A) Der Modalwert ist ein besonders leicht zu bestimmender Lageparameter.

B) Modalwerte lassen sich nur bei quantitativen Merkmalen anwenden.

C) Der Modus ist definiert als Merkmalsausprägung mit der wahrscheinlichsten Häufigkeit.

D) Der Modus ist nur dann eindeutig erklärt, wenn die Häufigkeitsverteilung ein eindeutig bestimmtes Maximum aufweist.

E) In einem Datensatz kann es nur einen Modalwert geben.

Lösung: A (49), D (49)

zu B) Modalwerte lassen sich **immer**, also **auch** bei **qualitativen** Merkmalen anwenden. (49)

Zu C) Der Modus ist definiert als Merkmalsausprägung mit der **größten** Häufigkeit. (49)

zu E) In einem Datensatz kann es **mehrere** Modalwerte geben. (49/Bsp 5.1)

29A) Median ist ein Synonym für Modus.

B) Der Median kann auch als Zentralwert bezeichnet werden.

C) Der Median ist nur bei mindestens ordinalskalierten Merkmalen anwendbar.

D) Der Median ist der „mittlere“ Wert eines geordneten Datensatzes.

E) Bei metrisch skalierten Merkmalen wird aus den beiden zentralen Werten der Mittelwert gebildet, um den Median zu erhalten.

Lösung: B (50), C (50), D (50), E (50)

zu A) Der Median ist ein weiterer Lageparameter. (50)

30A) Der Mittelwert wird auch Median genannt.

B) Der Mittelwert ist auch bei metrisch skalierten Merkmalen anwendbar.

C) Der Mittelwert berücksichtigt alle Werte eines Datensatzes mit gleichem Gewicht.

D) In die Berechnung des Medians gehen nur ein oder zwei zentrale Elemente eines Datensatzes ein.

E) Vergrößert man deutlich den größten Wert eines Datensatzes, hat dies nur auf den Median einen Effekt.

Lösung: C (50), D (50)

zu A) Der Mittelwert wird auch **arithmetisches Mittel** genannt. (50)

zu B) Der Mittelwert ist **nur** bei metrisch skalierten Merkmalen anwendbar. (50)

zu E) Vergrößert man deutlich den größten Wert eines Datensatzes, hat dies nur auf den **Mittelwert** einen Effekt. (50)

Zwischenauswertung: mögliche Punktzahl: 50

erreichte Punktzahl: .....

- 31A) Der Mittelwert reagiert empfindlicher als der Median auf extreme Werte.  
B) Ausreißer sind Studierende, die in ihren Klausurunterlagen die Formelsammlung abtrennen.  
C) Der Mittelwert hat eine geringe Sensitivität gegenüber Ausreißern.  
D) Das arithmetische Mittel hat eine höhere Robustheit gegenüber Ausreißern.  
E) Der Mittelwert lässt sich als Schwerpunkt eines Datensatzes interpretieren.

Lösung: A (50), E (50)

zu B) Ausreißer sind auffällig große oder kleine Beobachtungswerte. (50)

Zu C) Der Mittelwert hat eine **höhere** Sensitivität gegenüber Ausreißern. (50)

Zu D) Das arithmetische Mittel hat eine **geringere** Robustheit gegenüber Ausreißern. (50)

32A) Der Mittelwert lässt sich nicht so einfach berechnen, wenn Merkmalswerte mehrfach auftreten.

B) Anstelle der Urliste kann man für die Berechnung des Mittelwertes auch die relative Häufigkeitsverteilung verwenden.

C) Auch bei gruppierten Daten kann der Mittelwert berechnet werden.

D) Bei metrisch skalierten Daten wird oft nur ein Lageparameter berechnet.

E) Das gewichtete arithmetische Mittel, das getrimmte arithmetische Mittel und das geometrische Mittel sind weitere Lageparameter für metrisch skalierte Merkmale.

Lösung: B (52), C (52), E (54/Experte)

zu A) Der Mittelwert lässt sich **etwas einfacher** berechnen, wenn Merkmalswerte mehrfach auftreten. (51)

zu D) Bei metrisch skalierten Daten wird oft **nicht** nur ein Lageparameter berechnet. (52)

33A) Ein Datensatz definiert eine theoretische Verteilung.

B) Zwei Datensätze können in den Lageparametern übereinstimmen.

C) Zwei Datensätze können sich bezüglich der Variablen nicht unterscheiden.

D) Zur Charakterisierung von quantitativen Merkmalen muss man noch Kenngrößen heranziehen, die die Streuung innerhalb des Datensatzes messen.

E) Die Spannweite R eines Datensatzes ist ein besonders einfaches Streuungsmaß für metrisch skalierte Merkmale.

Lösung: B (55), D (55), E (55)

zu A) Ein Datensatz definiert eine **empirische** Verteilung. (55)

zu C) Zwei Datensätze können sich bezüglich der Variablen **deutlich** unterscheiden. (55)

34A) Für die Berechnung der Spannweite und des Medians ordnet man den Datensatz zunächst nach aufsteigender Größe.

B) Die Spannweite ergibt sich aus der Summe des größten und kleinsten Wertes des Datensatzes.

C) Die Spannweite hat den Nachteil, dass sie eine hohe Empfindlichkeit gegenüber Ausreißern besitzt.

D) Varianz, Stichprobenvarianz und empirische Varianz sind synonyme Begriffe.

E) Die Varianz bildet den Mittelwert aus den quadrierten Abweichungen und ist damit ein quadratisches Streuungsmaß.

Lösung: A (55), C (55), D (55), E (55f)

zu B) Die Spannweite ergibt sich **als Differenz** aus dem größten und kleinsten Wert des Datensatzes. (55)

35A) Die Varianz geht in ein lineares Streuungsmaß über, wenn man die Wurzel zieht.

B) Die empirische Standardabweichung ist ein quadratisches Streuungsmaß.

C) Durch Wurzelziehen erhält man aus der korrigierten Stichprobenvarianz die korrigierte Standardabweichung.

D) Die Unterschiede zwischen Varianz und korrigierter Varianz verschwinden mit zunehmendem  $n$ .

E) Bei kleinem  $n$  können die Unterschiede zwischen beiden Größen durchaus ins Gewicht fallen.

Lösung: A (56), C (56), D (56), E (56)

zu B) Die empirische Standardabweichung ist ein **lineares** Streuungsmaß. (56)

36A) Treten Merkmalswerte mehrfach auf, kann man bei der Ermittlung der Varianz auf absolute Häufigkeiten zurückgreifen.

B) Bei gruppierten Daten kann die Varianz nicht berechnet werden.

C) Die Standardisierung von Datensätzen ist erforderlich, wenn sich die Datensätze auf Messungen in unterschiedlichen Grundgesamtheiten beziehen.

D) Bei der Verwendung unterschiedlicher Messinstrumente ist es nicht möglich, Datensätze direkt zu vergleichen.

E) Median, Mittelwert und Standardabweichung sind vom Maßstab abhängig.

Lösung: C (58), E (59/Exkurs)

zu A) Treten Merkmalswerte mehrfach auf, kann man bei der Ermittlung der Varianz auf **relative** Häufigkeiten zurückgreifen. (58)

B) Auch bei gruppierten Daten kann die Varianz berechnet werden. (58)

D) Bei der Verwendung unterschiedlicher Messinstrumente ist es **durch Transformation** möglich, Datensätze direkt zu vergleichen. (58)

37A) Der Median markiert die „Mitte“ eines Datensatzes, d. h. 50% der Elemente sind kleiner und 50% größer als der Median.

B) Das  $p$ -Quantil setzt ein zumindest ordinalskaliertes Merkmal voraus.

C) Der Median ist ein spezielles Quantil, nämlich das 0,25-Quantil.

D) Das 0,25-Quantil wird auch unteres Quartil genannt.

E) Das obere Quartil trennt die unteren 0,25% der Elemente von den oberen 75%.

Lösung: A (59), B (59), D (60)

zu C) Der Median ist ein spezielles Quantil, nämlich das **0,5-Quantil**. (60)

Zu E) Das obere Quartil (= 0,75-Quantil) trennt die unteren **0,75%** der Elemente von den oberen 25%. (61/Abb. 5.2)

38A) Die Summe der beiden Quartile (0,25 und 0,75) wird Quartilsabstand genannt.

B) Der Median stimmt mit dem Dezil  $D_5$  überein.

C) Der Boxplot fasst in seiner einfachsten Form fünf Charakteristika eines Datensatzes zusammen.

D) Die beiden Quartile definieren die Länge einer Box.

E) Die Länge der Box entspricht der Spannweite  $R$ .

Lösung: B (60), C (61), D (61)

zu A) Die **Differenz** der beiden Quartile (0,25 und 0,75) wird Quartilsabstand genannt. (60)

zu E) Die Länge der Box entspricht dem **Quartilsabstand  $Q$** . (61)

39A) Innerhalb der Box liegen etwa 50% der Daten.

B) Unterhalb und oberhalb der Box liegen insgesamt ca. 25% der Daten.

C) Bei einer symmetrischen Verteilung liegt der Median genau in der Mitte der Box.

D) Extremwerte, die eine Länge von 1,5 IQR überschreiten, werden nicht mehr eingezeichnet.

E) Der maximale Wert des Datensatzes definiert die obere Begrenzung des Boxplots.

Lösung: A (61), C (61), E (62/Bsp 5.7)

zu B) Unterhalb und oberhalb der Box liegen **jeweils** ca. 25% der Daten. (61)

Zu D) Extremwerte, die eine Länge von 1,5 IQR überschreiten, werden **separat** eingezeichnet. (62)

Zwischenauswertung: mögliche Punkte: 50

erreichte Punkte: .....

## Kap. 6: Konzentration von Merkmalswerten

40A) Konzentration bezüglich eines Merkmals liegt vor, wenn sie die Merkmalssumme gleichmäßig auf die statistischen Einheiten verteilt.

B) Bei der Konzentrationsmessung geht es darum, die Lage und und Streuung eines Datensatzes zu charakterisieren.

C) Die Lorenzkurve ist ein wichtiges Instrument für die grafische Beurteilung von Konzentrationsphänomenen.

D) Die Lorenzkurve visualisiert, wie sich die Summe aller Merkmalswerte innerhalb der Grundgesamtheit verteilt.

E) Je stärker die Lorenzkurve von der Diagonalen abweicht, desto größer ist die Konzentration.

Lösung: C (65), D (66), E (66)

zu A) Konzentration bezüglich eines Merkmals liegt vor, wenn sie die Merkmalssumme **ungleichmäßig** auf die statistischen Einheiten verteilt. (65)

zu B) Bei der Konzentrationsmessung geht es **nicht** darum, die Lage und und Streuung eines Datensatzes zu charakterisieren. (65/Bsp. 6.1)

41A) Die Stützpunkte der Lorenzkurve bleiben unverändert, wenn man die Werte der Urliste mit einem positiven Faktor multipliziert.

B) Die Berechnung von Lorenzkurven ist auch bei gruppierten Daten möglich.

C) Die Lorenzkurve repräsentiert ein Maß für die Stärke von Konzentration.

D) Der Gini-Koeffizient  $G$  ist ein Konzentrationsmaß.

E) Die Berechnung von  $G$  setzt die Kenntnis der Stützpunkte der Lorenzkurve voraus.

Lösung: A (66), B (66), D (67)

zu C) Die Lorenzkurve repräsentiert **kein** Maß für die Stärke von Konzentration. (67)

Zu E) Die Berechnung von  $G$  setzt **nicht** die Kenntnis der Stützpunkte der Lorenzkurve voraus. (69)

42A) Bei fehlender Konzentration nimmt der Gini-Koeffizient sein Minimum an.

B) Die obere Schranke von  $G$  hängt von der Länge  $n$  der Urliste ab.

C) Der normierte Gini-Koeffizient  $G^*$  gleicht den Nachteil der Aussage B aus.

D) Unterschiedliche Urlisten der Länge  $n$  führen zu unterschiedlichen Gini-Koeffizienten.

E) Der Gini-Koeffizient zeigt volle Konzentration an, wenn alle Merkmalsträger einer Urliste übereinstimmen.

Lösung: A (69), B (69), C (69f)

zu D) Unterschiedliche Urlisten der Länge  $n$  **können** zum **gleichen** Gini-Koeffizienten führen. (70)

Zu E) Der Gini-Koeffizient zeigt **fehlende** Konzentration an, wenn alle Merkmalsträger einer Urliste übereinstimmen. (70)

43A) Lorenzkurve und Gini-Koeffizient liefern Aussagen des Typs „ $x$  Merkmalsträger sind für  $y\%$  der Merkmalssumme verantwortlich“.

B) Formuliert man Aussagen für einzelne Merkmalsträger, wird die relative Konzentration bewertet.

C) Formuliert man Aussagen für Anteile in der Grundgesamtheit, wird die absolute Konzentration bewertet.

D) Der Gini-Koeffizient misst die relative Konzentration.

E) Bei einem Wert  $G = 0$  kann von fehlender Konzentration gesprochen werden.

Lösung: D (70)

zu A) Lorenzkurve und Gini-Koeffizient liefern Aussagen des Typs „ $x\%$  der Merkmalsträger **teilen sich**  $y\%$  der Merkmalssumme“. (70)

Zu B) Formuliert man Aussagen für einzelne Merkmalsträger, wird die **absolute** Konzentration bewertet. (70)

zu C) Formuliert man Aussagen für Anteile in der Grundgesamtheit, wird die **relative** Konzentration bewertet. (70)

zu E) Bei einem Wert  $G = 0$  kann **nicht** von fehlender Konzentration gesprochen werden. (70)

44A) Der Herfindahl-Index ist ein Maß zur Erfassung der absoluten Merkmalskonzentration.

B) Bei der Berechnung des Herfindahl-Indexes ist es – genau wie beim Gini-Koeffizienten – nicht erforderlich, die Daten der Urliste der Größe nach zu ordnen.

C) Bei vollständiger Konzentration nimmt der Herfindahl-Index den Wert 1 an.

D) Der Herfindahl-Index besitzt, anders als der Gini-Koeffizient, eine positive untere Schranke.

E) Die untere Schranke des Herfindahl-Indexes wird mit abnehmender Länge  $n$  der Urliste kleiner.

Lösung: A (71), C (71), D (71)

zu B) Bei der Berechnung des Herfindahl-Indexes ist es – **anders als** beim Gini-Koeffizienten – nicht erforderlich, die Daten der Urliste der Größe nach zu ordnen. (71)

zu E) Die untere Schranke des Herfindahl-Indexes wird mit abnehmender Länge  $n$  der Urliste **größer**. (71)

**Kap. 7: Index- und Verhältniszahlen**

45A) Zahlen, die einen Sachverhalt quantifizieren, nennt man Maßzahlen.

B) Wenn man zwei Maßzahlen durch Summenbildung miteinander verknüpft, spricht man von einer Verhältniszahl.

C) Verhältniszahlen sollen die Vergleichbarkeit statistischer Informationen für unterschiedliche Regionen oder Zeitpunkte ermöglichen.

D) Relative Häufigkeiten sind Verhältniszahlen.

E) Verhältniszahlen verknüpfen durch Anteilsbildung eine Teilgesamtheit mit einer Grundgesamtheit.

Lösung: A (73), C (73), D (73), E (73)

zu B) Wenn man zwei Maßzahlen durch **Quotientenbildung** miteinander verknüpft, spricht man von einer Verhältniszahl. (73)

46A) Gliederungszahlen sind Verhältniszahlen, bei denen eine Grundgesamtheit durch Anteilsbildung bezüglich zweier Merkmale strukturiert wird.

B) Gliederungszahlen sind dimensionslos.

C) Beziehungszahlen sind Verhältniszahlen, die durch Quotientenbildung eine Verbindung zwischen zwei gleichen Merkmalen herstellen.

D) Die Verknüpfung der beiden Merkmale muss inhaltlich Sinn ergeben.

E) Bei Zeitreihen werden die Daten der aktuellen Periode durch die Werte einer Referenz- oder Basisperiode geteilt. Dadurch werden Veränderungen besser sichtbar.

Lösung: B (73), D (73), E (73)

zu A) Gliederungszahlen sind Verhältniszahlen, bei denen eine Grundgesamtheit durch Anteilsbildung bezüglich **eines** Merkmals strukturiert wird. (73)

zu C) Beziehungszahlen sind Verhältniszahlen, die durch Quotientenbildung eine Verbindung zwischen zwei **unterschiedlichen** Merkmalen herstellen. (73)

47A) Einfache Indexzahlen sind Verhältniszahlen, die die Werte für ein Merkmal von zwei Zeitpunkten verknüpfen.

B) Einfache Indexzahlen können sich auch auf zwei verschiedene Merkmale beziehen.

C) Maß- und Verhältniszahlen werden oft als Indikatoren herangezogen.

D) Indikatoren ermöglichen Vergleiche zwischen Regionen.

E) Die Frage nach der sachadäquaten Operationalisierung eines Merkmals muss auf der Basis sachlogischer Überlegungen geklärt werden.

Lösung: A (74), C (74), D (74), E (76)

zu B) Einfache Indexzahlen beziehen sich nur auf **ein einziges** Merkmal. (74)

48A) Ranglisten von Universitäten oder Ergebnisse der PISA-Studien entsprechen zusammengesetzten Indexzahlen.

B) Zusammengesetzte Indikatoren sind aus mehreren Indikatoren gebildete Aggregate einer

einzigsten Maßzahl.

- C) Aggregierte Maß- und Verhältniszahlen dürfen nicht unterschiedlich gewichtet sein.
- D) Zusammengesetzte Indexzahlen repräsentieren additive Verknüpfungen eines Sets von Maß- und Verhältniszahlen.
- E) Zusammengesetzte Indexzahlen brechen umfassende Indikatorensysteme auf mehrere Variablen herunter.

Lösung: A (76), B (76), D (76)

zu C) Aggregierte Maß- und Verhältniszahlen **können** unterschiedlich gewichtet sein. (76)

zu E) Zusammengesetzte Indexzahlen brechen umfassende Indikatorensysteme auf **eine einzige** Variable herunter. (79)

49A) Zusammengesetzte Indexzahlen ermöglichen eine mehrdimensionale Betrachtung eindimensionaler Phänome.

- B) Zusammengesetzte Indexzahlen gestatten einen direkten Ländervergleich.
- C) Zusammengesetzte Indexzahlen haben oft nur eine begrenzte Aussagekraft.
- D) Die Werte zusammengesetzter Indexzahlen sind unabhängig von den Gewichten der einfließenden Maß- und Verhältniszahlen.
- E) Zusammengesetzte Indexzahlen haben den Nachteil, dass die in sie eingehenden Einzelindikatoren im Zeitverlauf nicht selten geändert werden.

Lösung: B (79), C (79), E (79)

zu A) Zusammengesetzte Indexzahlen ermöglichen eine **eindimensionale** Betrachtung **mehrdimensionaler** Phänome. (79)

zu D) Die Werte zusammengesetzter Indexzahlen sind **abhängig** von den Gewichten der einfließenden Maß- und Verhältniszahlen. (79)

50A) Eurostat ist das Europäische Amt für Statistik.

- B) Die Verwendung zusammengesetzter Indexzahlen dient lediglich als grobe erste Orientierung.
- C) Zusammengesetzte Indexzahlen erlauben eine differenzierte Bewertung komplexer Sachverhalte.
- D) Zusammengesetzte Indexzahlen werden auch zur Beschreibung von Entwicklungen im ökonomischen Bereich herangezogen.
- E) Der Deutsche Aktienindex (DAX) und der amtliche Verbraucherindex sind Beispiele für zusammengesetzte Indexzahlen.

Lösung: A (79), B (79), D (76), E (76)

zu C) **Nur** die in zusammengesetzte Indexzahlen eingehenden **Einzelindikatoren** erlauben eine differenzierte Bewertung komplexer Sachverhalte. (79)

Zwischenauswertung: mögliche Punkte: 50

erreichte Punkte: .....

**Kap. 8: Bivariate Häufigkeitsverteilungen**

- 51A) Bivariate Datenanalysen erfassen Zusammenhänge zwischen zwei Merkmalen.  
B) Die Datenaufbereitung und Verwendung des Zusammenhangsmaßes hängt vom Zusammenhang ab.  
C) Merkmalspaare dürfen nicht mehrfach auftreten.  
D) Kontingenztabelle = Kontingenztafel = Kreuztabelle.  
E) Die unter D) genannten Synonyme definieren die gemeinsame empirische Verteilung der beiden Merkmale.

Lösung: A (81), D (81), E (81f)

zu B) Die Datenaufbereitung und Verwendung des Zusammenhangsmaßes hängt von der **Merkmalskalierung** ab. (81)

zu C) Merkmalspaare **können** mehrfach auftreten. (81)

- 52A) Die zusätzliche Spalte einer Tabelle für absolute Häufigkeiten weist die Spaltensummen aus.  
B) Die Randverteilungen sind die Häufigkeitsverteilungen der Einzelmerkmale.  
C) Durch die Randverteilungen wird eine Verbindung zwischen univariaten und bivariaten Häufigkeitsverteilungen hergestellt.  
D) Aus zwei gegebenen Randverteilungen kann man auf die gemeinsamen Häufigkeiten zurück schließen.  
E) Summenbildung beinhaltet Verdichtung von Information.

Lösung: B (83), C (83), E (83)

zu A) Die zusätzliche Spalte einer Tabelle für absolute Häufigkeiten weist die **Zeilensummen** aus. (82)

zu D) Aus zwei gegebenen Randverteilungen kann man **i. a. nicht eindeutig** auf die gemeinsamen Häufigkeiten zurück schließen. (83)

- 53A) Die Vierfeldertafel ist ein Spezialfall einer Kontingenztabelle.  
B) Vierfeldertafeln eignen sich für die Untersuchung von Zusammenhängen zwischen zwei Merkmalen, die je nur eine Ausprägung aufweisen.  
C) Binäre Merkmale sind Merkmale mit zwei Ausprägungen.  
D) Dichotome Merkmale sind Merkmale mit mehr als zwei Ausprägungen.  
E) Durch das Baumdiagramm wird eine hierarchische Struktur dargestellt.

Lösung: A (84), C (84f), E (85)

zu B) Vierfeldertafeln eignen sich für die Untersuchung von Zusammenhängen zwischen zwei Merkmalen, die je nur **zwei** Ausprägungen aufweisen. (84)

zu D) Dichotome (= binäre) Merkmale sind Merkmale mit **zwei** Ausprägungen. (85)

- 54A) Aus gemeinsamen Häufigkeiten für zwei Merkmale kann man Aussagen über Zusammenhänge zwischen den Merkmalen ableiten.  
B) Eine geeignete Verknüpfung der gemeinsamen Häufigkeiten für zwei diskrete Merkmale mit den Randhäufigkeiten führt zu bedingten relativen Häufigkeiten.  
C) Dividiert man jedes der Elemente durch die Randhäufigkeit, so erhält man absolute



Häufigkeiten.

D) Bedingte Häufigkeitsverteilungen sind univariate Verteilungen, weil nur die Merkmalsausprägungen für ein einziges Merkmal variieren.

E) Von einem fehlenden Zusammenhang zwischen zwei Merkmalen spricht man bei Unabhängigkeit der Merkmale.

Lösung: B (87), D (89), E (91)

zu A) Aus gemeinsamen Häufigkeiten für zwei Merkmale kann man noch **nicht direkt** Aussagen über Zusammenhänge zwischen den Merkmalen ableiten. (87)

zu C) Dividiert man jedes der Elemente durch die Randhäufigkeit, so erhält man **bedingte relative** Häufigkeiten. (87)

55A) Empirische Unabhängigkeitsaussagen von Merkmalen leiten sich aus Wahrscheinlichkeitsmodellen ab.

B) Bei Unabhängigkeit zweier Merkmale steckt die gesamte Information über die gemeinsame Häufigkeitsverteilung bereits in den Randverteilungen.

C) Bei stetigen Merkmalen wird man selten beobachten, dass Merkmalspaare mehrfach auftreten.

D) Ein Datensatz für zwei stetige Merkmale wird in einem zweidimensionalen Koordinatensystem dargestellt.

E) Ein Streudiagramm liefert einen visuellen Anhaltspunkt über das Bestehen oder Fehlen von Zusammenhängen bei stetigen Merkmalen.

Lösung: B (92), C (94), D (95), E (95)

zu A) Empirische Unabhängigkeitsaussagen von Merkmalen leiten sich aus **Daten** ab. (92)

## Kap. 9: Zusammenhangsmaße

56A) Um ein Zusammenhangsmaß zu quantifizieren, verwendet man die Summe der quadrierten Differenzen.

B) Der Chi-Quadrat-Koeffizient ist ein Zusammenhangsmaß für zwei nominalskalierte Merkmale.

C) Der Chi-Quadrat-Koeffizient wird in der induktiven Statistik u. a. verwendet, um Hypothesen über Merkmalszusammenhänge zu testen.

D) Der Chi-Quadrat-Koeffizient ist ein Zusammenhangsmaß, das beliebig groß werden kann.

E) Wenn der Chi-Quadrat-Koeffizient den maximalen Wert annimmt, spricht man von vollständiger Abhängigkeit.

Lösung: A (97), B (97), C (97/Fußnote 1), E (98)

zu D) Der Chi-Quadrat-Koeffizient ist ein Zusammenhangsmaß, das **nicht** beliebig groß werden kann. Es ist **nach oben beschränkt**. (97)

57A) Der Phi-Koeffizient ist ein Zusammenhangsmaß, dessen Wert vom Umfang  $n$  abhängt.

B) Der Phi-Koeffizient ist ein negatives Maß und nimmt bei schwachem Merkmalszusammenhang kleine Werte an.

C) Der maximale Wert, den der Phi-Koeffizient bei vollständiger Abhängigkeit der beiden Merkmale annimmt, hängt von der Dimension der Kontingenztafel ab.

- D) Das Zusammenhangsmaß von Cramer nimmt stets Werte zwischen 0 und 1 an.  
 E) Cramer's V ist ein normiertes Zusammenhangsmaß.

Lösung: C (98), D (98), E (98)

- zu A) Der Phi-Koeffizient ist ein Zusammenhangsmaß, dessen Wert **nicht** vom Umfang  $n$  abhängt. (98)  
 zu B) Der Phi-Koeffizient ist ein **nicht**-negatives Maß und nimmt bei schwachem Merkmalszusammenhang kleine Werte an. (98)

- 58A) Mit dem Zusammenhangsmaß von Cramer lässt sich die Stärke von Merkmalszusammenhängen bei Kontingenztafeln beliebiger Dimension direkt vergleichen.  
 B) Im Falle  $M = 2$  (Vierfeldertafel) stimmen der Phi-Koeffizient und Cramer's V stets überein.  
 C) Die empirische Kovarianz ist ein normiertes Zusammenhangsmaß für den Zusammenhang zwischen zwei metrischen Merkmalen.  
 D) Je mehr Datenpunkte im ersten und dritten Quadranten liegen, desto größer wird die Kovarianz.  
 E) Wenn alle Punkte auf einer fallenden Geraden liegen, liefert jeder Punkt einen nicht-negativen Beitrag zur Kovarianz.

Lösung: A (100), B (100), D (102)

- zu C) Die empirische Kovarianz ist ein **nicht**-normiertes Zusammenhangsmaß für den Zusammenhang zwischen zwei metrischen Merkmalen. (101)  
 zu E) Wenn alle Punkte auf einer **steigenden** Geraden liegen, liefert jeder Punkt einen nicht-negativen Beitrag zur Kovarianz. (102)

- 59A) Eine positive Kovarianz bedeutet, dass die Ausprägungen der Merkmale X und Y eine gleichgerichtete Tendenz haben.  
 B) Die Kovarianz ist maßstabsunabhängig.  
 C) Die Kovarianz ist dimensionslos.  
 D) Der Korrelationskoeffizient nach Bravais-Pearson ist ebenfalls maßstabsunabhängig und dimensionslos.  
 E) Das Vorzeichen des Korrelationskoeffizienten ist stets mit dem Vorzeichen der Kovarianz identisch.

Lösung: A (102), D (102f), E (103)

- zu B) Die Kovarianz ist **maßstabsabhängig**. (102)  
 zu C) Die Kovarianz ist **nicht** dimensionslos. (102)

- 60A) Liegen alle Datenpunkte auf einer steigenden oder fallenden Geraden, besteht lineare Abhängigkeit zwischen den Merkmalen.  
 B) Im Falle von  $r = 0$  besteht Unkorreliertheit zwischen beiden Merkmalen. Es kann durchaus ein nicht-linearer Zusammenhang vorliegen.  
 C) Der Wert  $r = 1$  ist in der Realität nur unter bestimmten Voraussetzungen erreichbar.  
 D) Ein anhand eines großen Wertes  $|r|$  festgestellter Zusammenhang bedeutet, dass zwischen den Merkmalen ein Kausalzusammenhang besteht, also eine sachlogische Verbindung.  
 E) Scheinkorrelation bedeutet, dass ein direkter sachlogischer Zusammenhang fehlt.

Lösung: A (103), B (103f), E (107)

- zu C) Der Wert  $r = 1$  ist in der Realität **nie** erreichbar. (105/Bsp 9.2)

zu D) Ein anhand eines großen Wertes  $|r|$  festgestellter Zusammenhang muss **nicht** zwingend bedeuten, dass zwischen den Merkmalen ein Kausalzusammenhang besteht, also eine sachlogische Verbindung. (106)

Zwischenauswertung: mögliche Punkte: 50

erreichte Punkte: .....

61A) Der Korrelationskoeffizient nach Bravais-Pearson ist auch für ordinalskalierte Merkmale anwendbar.

B) Der Rangkorrelationskoeffizient nach Spearman misst nur einen linearen Zusammenhang zwischen den Rangplätzen der Merkmalswerte.

C) Bezogen auf die Merkmalswerte selbst misst der Rangkorrelationskoeffizient nach Spearman lediglich, ob ein gleichsinniger oder ein gegensinniger monotoner Zusammenhang vorliegt.

D) Der Rangkorrelationskoeffizient nach Spearman ist auch für metrische Merkmale anwendbar.

E) Bei der Berechnung des Rangkorrelationskoeffizienten nach Spearman können Rangplätze mehrfach besetzt sein.

Lösung: B (109), C (109), D (109)

zu A) Der Korrelationskoeffizient nach Bravais-Pearson ist **nicht** für ordinalskalierte Merkmale anwendbar. (109)

zu E) Bei der Berechnung des Rangkorrelationskoeffizienten nach Spearman wird vorausgesetzt, das **kein** Rangplatz mehrfach besetzt ist. (109)

Gesamtauswertung: mögliche Punktzahl: 305

erreichte Punktzahl: .....