Statistik-
Skript

Studienort: Potsdam

Verfasser:
Dipl.-Kfm. Thomas Rochow

Verwaltungs- und Wirtschafts-Akademie Potsdam e. V.

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Anmerkungen zu Grundbegriffen

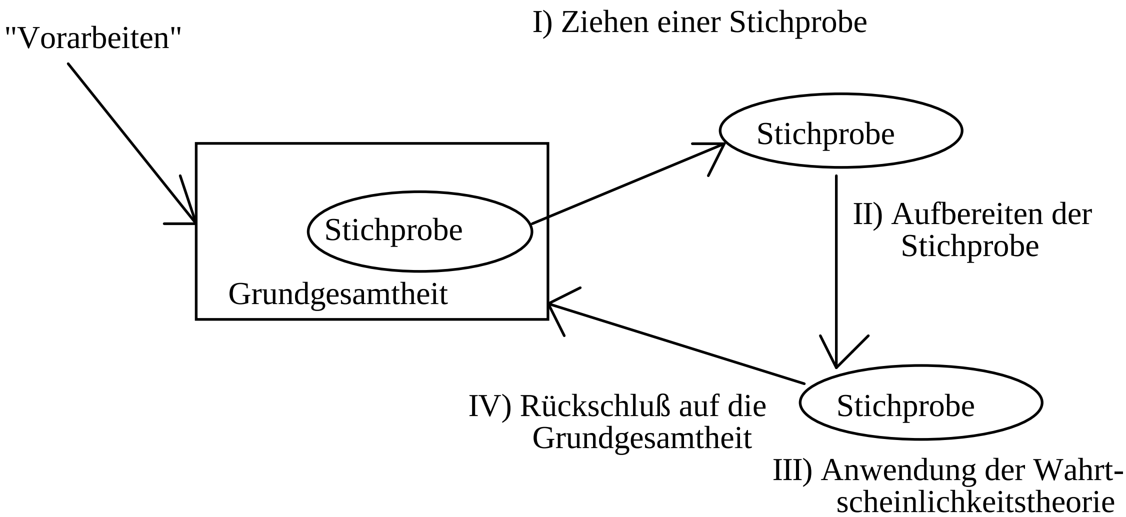
Der Begriff ***Statistik*** hat mehrere Bedeutungen:

1. Ansammlung von Zahlen
   (Statistik der Verkehrsunfälle, Statistik der erwerbsfähigen Bevölkerung nach Alter, Statistik der Rentenzugänge, ...)

2. Methoden zur Analyse von Massenerscheinungen und ihre Anwendung
   (Durchführung und Auswertung von Versuchen, Meinungsforschung, Volkszählung, Mikrozensus, ...)

3. Wissenschaft von den Methoden und ihren mathematischen Grundlagen
   (deskriptive Statistik, Stochastik (Wahrscheinlichkeitsrechnung), Entscheidungstheorie, Versuchsplanung, induktive Statistik, Indexrechnung)

Vorgehensweise bei einer statistischen Analyse



Bemerkungen:

zu I) Probleme, die hier oder davor ("Vorarbeiten") liegen, werden nicht behandelt

zu II) - IV) Wichtige Grundbegriffe:

a) **Grundgesamtheit:** Die für eine Statistische Untersuchung relevanten Merkmalsträger werden zur Grundgesamtheit zusammengefasst.

b) **Merkmalsträger:** Personen, Sachen, an denen bestimmte Merkmale erhoben werden.

c) **Merkmal:** die bei einer statistischen Erhebung interessierende Eigenschaft eines Merkmalträgers z. B. Einkommen, höchster Schulabschluss, Größe einer Person; Farbe, Höhe, Arbeitsfläche eines Tisches

Vollerhebung und Teilerhebung

Werden **alle Merkmalsträger** der Grundgesamtheit bezüglich der interessierenden Größe befragt, liegt eine **Totalerhebung (Vollerhebung)** vor, wird hingegen nur **ein Teil der Merkmalsträger** untersucht, liegt eine **Teilerhebung (Stichprobenerhebung, Partial-erhebung)** vor.

Probleme der Vollerhebung:

- nicht immer sinnvoll / nicht immer möglich
  (bei zerstörender Prüfung, z.B. bei Überprüfung von Lebensmitteln oder technischen Geräten (Qualitätskontrolle); Meinungsforschung, Marktforschung)

- zu langwierig
  (oft ist man an schnellen Ergebnissen interessiert, z. B. Meinungsforschung bei Wahlen)

- zu teuer
  (insb. wenn die Grundgesamtheit sehr groß ist; Verhältnismäßigkeit der Mittel)

- zu ungenau
  (bei großer Grundgesamtheit ist die Richtigkeit der Angaben kaum feststellbar (Plausibilitätskontrolle) und überprüfbar; und die Ergebnisse der Untersuchung kommen i. d. R. zu spät)

Probleme der Teilerhebung

- Repräsentanz

  Eine Auswahl heißt repräsentativ, wenn sie alle für die Grundgesamtheit typischen und charakteristischen Erhebungsmerkmale nebst deren verschiedenen Kombinationen genau entsprechend ihrer relativen Häufigkeit in der Grundgesamtheit enthält und somit ein getreues Miniaturbild der Grundgesamtheit, sozusagen ihr verkleinertes Modell, darstellt.

Primärerhebung und Sekundärerhebung

Bei der **Primärerhebung (field research)** wird der Untersuchungsträger eigens tätig, d. h. es handelt sich um die Gewinnung originärer Daten:
positiv: aktuell, auf das Problem zugeschnitten
negativ: hoher Zeitaufwand, hohe Kosten

Bei der **Sekundärerhebung (desk research)** greift der Untersuchungsträger auf bereits vorhandenes Datenmaterial zurück.
positiv: geringer Zeitaufwand, geringe Kosten
negativ: Datenmaterial ist alt, passt nicht exakt auf das zu untersuchende Problem

Allgemeine Vorgehensweise einer statistischen Untersuchung

a) Was ist das Problem?
(Formulierung der Aufgabenstellung, dabei ist auf Präzision zu achten)

b) Konkretisierung
(wie ist das, was man wissen will, messbar? Aufstellen eines geeigneten Modells)

c) Versuchsplanung/ Stichprobenplanung incl. Methoden der Auswertung
(Versuchsplanung: Sollen verschiedene Versuchsbedingungen berücksichtigt werden?
Wie soll die Versuchsanordnung aussehen?;
Stichprobenplanung: Vollerhebung oder Teilerhebung?
Mit welchen Methoden sollen die gewonnenen Daten ausgewertet werden?

Darüber muss schon vor der Erhebung entschieden werden.)

d) Erhebungs- bzw. Messtechnik festlegen
(**Befragung** (Fragebogen, Interview, telefonische Befragung; Panel),
Beobachtung,
Experiment)

e) Durchführung des Versuchs/Erhebung

f) Auswertung der gewonnenen Daten mithilfe der unter c) festgelegten Methoden

g) Interpretation der Ergebnisse

Erhebungstechniken

Befragung:
Befragung ist die zielgerichtete Veranlassung von Personen, Aussagen über bestimmte Sachverhalte zu treffen.

Beobachtung:
Beobachtung ist die planmäßige Wahrnehmung oder Anschauung mit dem Ziel einer möglichst umgehenden und umfassenden Kenntnisgewinnung über einen Gegenstand oder Sachverhalt

Experiment:
Ziel eines Experimentes ist es, durch eine entsprechende Anordnung die Wirkung der Veränderung einer Größe auf eine oder mehrere andere festzustellen.

Arten der Befragung

schriftlich über einen Erhebungsbogen
mündlich über Interviewer
telefonisch
Panel, laufend wiederkehrende Befragung eines gleichbleibenden Personenkreises über einen gleichbleibenden Erhebungstatbestand

Vorteile und Nachteile der schriftlichen Befragung

Vorteile:	
• Anonymität kann gewahrt werden	• Rücklauf der Erhebungsunterlagen
• räumliche Entfernung kein Problem	• Repräsentanz ist gestört
• Kosten der Durchführung sind gering	• Fragebogenumfang eingeschränkt
• keine Interviewerkosten	• Befragungstaktik ist stark eingeschränkt

Vorteile und Nachteile der mündlichen Befragung

Vorteile	Nachteile:
• Fragethematik im Grundsatz unbeschränkt	• Feldorganisation erforderlich (Gewinnung von Interviewern)
• befragungstaktisches Instrumentarium vollständig anwendbar	• hohe Interviewerkosten
• Befragungssituation kontrolliert	• Interviewereinfluss (Interviewerbias)
• ergänzende Beobachtungen sind möglich	

Vorteile und Nachteile der telefonischen Befragung

Vorteile:	Nachteile:
• rasche Durchführbarkeit	• Kreis der Auskunftspersonen eingeschränkt
• geringer Erhebungsaufwand	• dadurch eingeschränkte Repräsentanz
	• Fragethematik ist eingeschränkt

Vorteile und Nachteile der Panel – Erhebung (hier Haushaltspanel)

über einen längeren Zeitraum wird ein fester Personenkreis (Haushalt) über einen im Wesentlichen gleich bleibenden Themenkomplex (getätigte Einkäufe) befragt.

Vorteile:	Nachteile:
• rasches Erkennen von Veränderungen	• „Panelsterblichkeit“ Mitglieder scheiden aus dem Panel aus
	• „Paneleffekt“ Mitglieder verändern durch die Fragebogen – Kontrolle ihr Kaufverhalten
	• overreporting nicht – getätigte Käufe werden angegeben
	• underreporting getätigte Käufe werden nicht angegeben

Arten der Beobachtung:

teilnehmende und nicht-teilnehmende Beobachtung
Feld- und Laboratoriumsbeobachtung

Bei der **teilnehmenden Beobachtung** nimmt der Beobachter auf der gleichen Ebene wie der Beobachtete am Ablauf des Geschehens teilt.

Die **nicht-teilnehmende Beobachtung** wird z. B. auf fototechnischem Wege durchgeführt.

Eine **Laboratoriumsbeobachtung** liegt vor, wenn sich das Beobachtungsobjekt nicht in der "normalen" Umwelt befindet, sondern unter künstlich geschaffenen Bedingungen.

In einer **Feldbeobachtung** bewegt sich das Beobachtungsobjekt unter "normalen", d. h. bekannten, vertrauten Bedingungen.

Arten des Experiments:

Befragungs- und Beobachtungsexperiment
Feld- und Laboratoriumsexperiment

Der Unterschied zwischen **Befragungsexperiment** und **Beobachtungsexperiment** liegt in der Position des Beobachtungsträgers begründet, der zwischen **Feldexperiment** und **Laboratoriumsexperiment** in der Situation, der der Beobachtete ausgesetzt wird.

Theorie zu Auswertungsmethoden für eindimensionales Datenmaterial

Messskalen und ihre Eigenschaften

- **Nominalskala**

Eine Skala, deren Skalenwerte nur nach dem Kriterium **gleich** oder **ungleich** geordnet werden können, heißt **Nominalskala**. Ein Merkmal, dessen Werte nur auf einer Nominalskala gemessen werden können, heißt **nominalskaliert.**
Bsp.: Farbe, Geschlecht, Beruf, Konfession

- **Ordinalskala**

Eine Skala, deren Skalenwerte nicht nur nach dem Kriterium **gleich** oder **ungleich** sondern außerdem in einer **natürlichen Reihenfolge** geordnet werden können, heißt **Ordinalskala**. Ein Merkmal, dessen Werte auf einer Ordinalskala gemessen werden können, heißt **ordinalskaliert.**
Bsp.: Zensuren, Güteklassen von Hotels

- **Kardinalskala**

Eine Skala, deren Skalenwerte reelle Zahlen sind und die alle Ordnungseigenschaften der reellen Zahlen besitzt, heißt **Kardinalskala**. Ein Merkmal, dessen Werte auf einer Kardinalskala gemessen werden können, heißt **kardinalskaliert.**
Bsp.: Entfernungen (in km), Flächen (in cm^2), Volumina (in m^3), Gewichte (in kg)

Notation:

Allgemeine Notation:

X, Y: statistisches Merkmal, statistische Größe

N: Umfang der Grundgesamtheit

n: Stichprobenumfang

e_i: statistische Einheit i, Merkmalsträger i

x_i: Beobachtungswert Nummer i bezüglich Merkmal X ohne Beachtung des Merkmalsträgers

Notation für eindimensionales ungruppiertes (unklassiertes) Datenmaterial:

a_j: Werte, die die betrachtete statistische Größe in unserer Stichprobe für die statistische Größe konkret annimmt

$h(a_j)$: absolute Häufigkeit der Ausprägung a_j

$f(a_j)$: relative Häufigkeit der Ausprägung a_j

$H(x)$: absolute kumulierte Häufigkeit

$F(x)$: relative kumulierte Häufigkeit

Kumulierte Häufigkeiten geben Auskunft über höchstens (also kleiner/gleich) Fragestellungen

Notation für eindimensionales gruppiertes (klassiertes) Datenmaterial:

j: Nummer der Gruppe (Klasse)

h_j: absolute Häufigkeit der Gruppe j

f_j: relative Häufigkeit der Gruppe j

$H(x)$: absolute kumulierte Häufigkeit

$F(x)$: relative kumulierte Häufigkeit

Kumulierte Häufigkeiten geben Auskunft über höchstens (also kleiner/gleich) Fragestellungen

b_j: Breite der Gruppe j

$$b_j = Obergrenze\ \ Gruppe\ j\ -\ Untergrenze\ der\ Gruppe\ j$$

m_j: Mitte der Gruppe j

$$m_j = \frac{Untergrenze\ \ Gruppe\ j\ + Obergrenze\ der\ Gruppe\ j}{2}$$

Darstellung der absoluten bzw. relativen Häufigkeiten in Graphiken

• Anforderungen an Grafiken:

Grafiken sollen anschaulich sein;
übersichtlich sein;
einfach sein;
das Wesentliche zum Ausdruck bringen.

Formen grafischer Darstellungen der absoluten bzw. relativen Häufigkeiten

• **Stabdiagramm**

- Anforderung an die Skalierungsart:
  Die statistische Größe sollte kardinalskaliert sein.
- Maß der absoluten/ relativen Häufigkeit:
  Höhe des Stabes ist proportional zur absoluten/ relativen Häufigkeit

• **Blockdiagramm**

- Besonderheit: es kann das Ergebnis der Erhebung von zwei statistischen Größen dargestellt werden
- Anforderungen an die Skalierungsart
  mindestens eine statistische Größe sollte kardinalskaliert sein.
- Maß der absoluten/ relativen Häufigkeit:
  Höhe (manchmal findet sich in der Literatur auch die Fläche) des Blockes der einen statistischen Größe ist proportional zur absoluten/ relativen Häufigkeit

• **Kreissektorendiagramm**

- Anforderungen an die Skalierungsart

  Die statistische Größe kann auch nominalskaliert sein.

- Maß der absoluten/ relativen Häufigkeit:

  Fläche des Kreissektors ist proportional zur absoluten/ relativen Häufigkeit, oder alternativ Winkel des Kreissektors ist ein Maß für die absolute bzw. relative Häufigkeit

Formen graphischer Darstellungen der absoluten kumulierten bzw. der relativen kumulierten Häufigkeiten

• **Empirische Verteilungsfunktion**

- Anforderung an die Skalierungsart:

  Die statistische Größe muss kardinalskaliert sein.

- Es entsteht im Falle ungruppierter Daten eine Treppenfunktion.

Gruppieren, Klassieren von Datenmaterial:

Wann gruppiert man?

- viele verschiedene Ausprägungen der betrachteten statistischen Größe,
- absolute bzw. relative Häufigkeiten sind ungefähr gleich und damit in der Regel klein

Warum gruppiert man?

- weil die Darstellung der ungruppierten Graphiken und Tabellen schnell unübersichtlich wird

- Formen graphischer Darstellungen der absoluten bzw. relativen Häufigkeiten:
  - **Histogramm** (Prinzip der Flächentreue)
    Die Fläche eines Blockes über einer Klasse entspricht der absoluten/ relativen Häufigkeit dieser Klasse.

- Formen graphischer Darstellungen der absoluten kumulierten bzw. der relativen kumulierten Häufigkeiten:
  - **Empirische Verteilungsfunktion**

- Vorteile/ Nachteile/ Gefahren des Klassierens
  - Vorteil: Die Darstellungen werden übersichtlicher und können das Wesentliche besser herausstellen.
  - Nachteil: In jedem Fall entsteht aber ein Informationsverlust.
  - Gefahr: Manipulationsgefahr durch die selbst wählbaren Klassengrenzen

- Weiteres Problem sind **offene Randklassen**. Diese entstehen, wenn im oberen (unteren) Bereich der Merkmalsausprägungen sehr große Intervalle nur sehr schwach besetzt sind. In Tabellen finden sich dann Bezeichnungen wie: "mehr als", "über", "weniger als", "unter".
  Folge ist, dass das Histogramm nicht mehr zeichenbar ist, auch Maßzahlen sind nicht mehr berechenbar.

  Schließen offener Randklassen:
  - nächster Wert, der sich bei gleicher Klassenbreite ergibt;
  - geschätzter, vermuteter Wert, der die Realität am besten abbildet
  - falls bekannt: der tatsächliche Wert

Lage- und Streuungsparameter

- **Lageparameter**
  - arithmetisches Mittel $\overline{x}$
  - Median x_{med}
  - Modus x_{mod}
  - spezielle Mittelwerte (x_{geo}, x_{harm})
  - Minimum und Maximum $(Min \;\; und \;\; Max)$
  - unteres und oberes Quartil $(x_{0,25} \;\; und \;\; x_{0,75})$

- **Streuungsparameter**
  - (empirische) Varianz s^2
  - Standardabweichung s
  - absolute durchschnittliche Abweichung $\overline{s}$
  - Spannweite SP
  - Quartilsabstand QA
  - Variationskoeffizient V

Achtung: Mit der Verdichtung der Daten zu Lage- und Streuungsparametern ist stets ein Informationsverlust verbunden!

Beispiel:

In einer psychologischen Untersuchung haben 20 Personen eine anschauungsgebundene Denksportaufgabe zu lösen. Folgende Lösungszeiten (in min) ergaben sich:

$x_1 = 6$; $x_2 = 5$; $x_3 = 5$; $x_4 = 8$; $x_5 = 4$;
$x_6 = 6$; $x_7 = 5$; $x_8 = 6$; $x_9 = 7$; $x_{10} = 3$;
$x_{11} = 2$; $x_{12} = 5$; $x_{13} = 6$; $x_{14} = 5$; $x_{15} = 7$;
$x_{16} = 6$; $x_{17} = 5$; $x_{18} = 6$; $x_{19} = 6$; $x_{20} = 7$

Aus der Urliste ergibt sich.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{20} * 110 = 5{,}5 \text{ min}$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 1{,}85 \text{min}^2 \, bzw. s = \sqrt{s^2} = 1{,}36 \text{min}$$

Aus der Häufigkeitsfunktion ergibt sich:

a_j	$h(a_j)$	$a_j h(a_j)$	$a_j \frac{h(a_j)}{n}$	$a_j - \bar{x}$	$(a_j - \bar{x})^2$	$(a_j - \bar{x})^2 \frac{h(a_j)}{n}$	a_j^2	$a_j^2 \frac{h(a_j)}{n}$
2	1	2	0,10	-3,5	12,25	0,6125	4	0,20
3	1	3	0,15	-2,5	6,25	0,3125	9	0,45
4	1	4	0,20	-1,5	2,25	0,1125	16	0,80
5	6	30	1,50	-0,5	0,25	0,0750	25	7,50
6	7	42	2,10	+0,5	0,25	0,0875	36	12,60
7	3	21	1,05	+1,5	2,25	0,3375	49	7,35
8	1	8	0,40	+2,5	6,25	0,3125	64	3,20
	20	110	5,50			1,8500		32,10

$$\bar{x} = \frac{1}{n}\sum_j a_j h(a_j) = \frac{1}{20} * 110 = 5{,}5 \text{min}$$

$$s^2 = \frac{1}{n}\sum_j \left(a_j - \bar{x}\right)^2 h(a_j) = 1{,}85 \text{ min}^2 \, bzw. s = \sqrt{s^2} = 1{,}36 \text{min}$$

gruppiertes Datenmaterial:

Da die $\overline{x}_i$ bei gruppiertem Material nicht berechnet werden können, da die einzelnen Realisationen nicht bekannt sind, werden $\overline{x}_i$ durch ihre Gruppenmitten m_i geschätzt

von ... bis unter	h_j	m_j	$m_j h_j$	$m_j - \overline{x}$	$(m_j - \overline{x})^2$	$(m_j - \overline{x})^2 \frac{h_j}{n}$	m_j^2	$m_j^2 \frac{h_j}{n}$
2 - 4	2	3	6	-3	9	0,90	9	0,90
4 - 6	7	5	35	-1	1	0,35	25	8,75
6 - 8	10	7	70	+1	1	0,50	49	24,50
8 – 10	1	9	9	+3	9	0,45	81	4,05
	20		120			2,20		38,20

$$\overline{x} = \frac{1}{n}\sum_j m_j h_j = \frac{1}{20} * 120 = 6,0 \text{ min}$$

$$s^2 = \frac{1}{n}\sum_j \left(m_j - \overline{x}\right)^2 h_j = 2{,}20 \text{min}^2 \, bzw. s = \sqrt{s^2} = 1{,}48 \text{min}$$

Anmerkungen zum arithmetischen Mittel:

Das arithmetische Mittel wird durch Ausreißer beeinflusst.

Das arithmetische Mittel lässt keinen Aufschluss über die Form der Verteilung zu.

Median x_{med}: Median ist derjenige Wert x, für den gilt:

Urliste: Gegeben seien n geordnete Beobachtungswerte:

falls n eine ungerade Zahl, so hat der mittlere Wert die Ordnungsnummer $x_{(n+1)/2}$ und es gilt:

$$x_{med} = x_{(n+1)/2}$$

falls n eine gerade Zahl, so kommen alle Werte zwischen $x_{n/2}$ und $x_{(n/2+1)}$ als Median in Frage.

Im Beispielsfalle ergibt sich nach einer Sortierung der Werte aus der Urliste: $x_{med} = 6$ min

Durchschnittliche Abweichung: Streuungsmaß um den Median

a_j	$h(a_j)$	$\lvert a_j - x_{med} \rvert$	$\lvert a_j - x_{med} \rvert \frac{h(a_j)}{n}$
2	1	4	0,20
3	1	3	0,15
4	1	2	0,10
5	6	1	0,30
6	7	0	0,00
7	3	1	0,15
8	1	2	0,10
	20		1,00

$$\bar{s} = \frac{1}{n} \sum_j \lvert a_j - x_{med} \rvert h(a_j) = 1 \text{ min}$$

Klassiertes Datenmaterial: Median ist derjenige Wert, für den gilt: $F(x) = 0{,}5$

von ... bis unter	h_j	$H(x)$	m_j	$\lvert m_j - x_{med} \rvert$	$\lvert m_j - x_{med} \rvert \frac{h_j}{n}$
2 - 4	2	2	3	3,2	0,32
4 - 6	7	9	5	1,2	0,42
6 - 8	10	19	7	0,8	0,40
8 – 10	1	20	9	2,8	0,14
	20				1,28

Der Median wird grafisch über die Verteilungsfunktion bestimmt. Im Beispiel ergibt sich: $x_{med} = 6{,}2$ min

$$\bar{s} = \frac{1}{n} \sum \lvert m_j - x_{med} \rvert h_j = 1{,}28 \text{min}$$

Anmerkungen zum Median:
Der Median wird nicht durch Ausreißer beeinflusst.

Modus und Spannweite

Modus x_{mod}: Modus ist derjenige Wert x,der am häufigsten realisiert ist.

In unserem Beispiel ergibt sich für den ungruppierten Fall: $x_{\text{mod}} = 6$ min

Für den gruppierten Fall folgt: Der Modus ist die dritte Gruppe. Dieser Parameter ist für gruppiertes Datenmaterial aber eher untypisch.

Spannweite SP: $SP = |x_{\max} - x_{\min}| = 8 - 2 = 6$ min

Spezielle Mittelwerte

1. **Geometrisches Mittel:**
   Hat man es mit zeitlich aufeinanderfolgenden Zuwächsen, Wachstumsraten oder ähnlichen zu tun, so ist das geometrische Mittel der geeignete Durchschnittswert. Dabei wird das geometrische Mittel nicht aus den gegebenen Zuwachsraten, Zinssätzen oder Wachstumsraten berechnet, sondern aus den daraus abzuleitenden Zuwachsfaktoren, Zinsfaktoren oder Wachstumsfaktoren.

   Bsp.: Zinssatz: $p(\%)$; Zinsfaktor: $1+\frac{p}{100}$

Formelinstrumentarium:
Gegeben seien n Beobachtungswerte $x_i(i=1,2,...,n)$ eines Merkmals X. Für das **ungewogene geometrische Mittel** gilt:

$$x_{geo} = \sqrt[n]{x_1 * x_2 * \ldots * x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Gegeben sei die Häufigkeitsverteilung des Merkmals X, dessen Ausprägungen $a_j(j=1,2,...,m)$ mit absoluten bzw. relativen Häufigkeiten $h(a_j)$ bzw. $f(a_j)$ auftreten. Für das **gewogene geometrische Mittel** gilt:

$$x_{geo} = \sqrt[n]{a_1^{h(a_1)} * a_2^{h(a_2)} * \ldots * a_m^{h(a_m)}} = \sqrt[n]{\prod_{j=1}^{m} a_j^{h(a_j)}} = \prod_{j=1}^{m} a_j^{f(a_j)}$$

2. **Harmonisches Mittel:**
   Wenn ein Durchschnittswert aus Verhältniszahlen (z. B. km/h) gebildet werden soll, so muss dies mit dem **harmonischen Mittel** (arithmetischen Mittel) geschehen, wenn die im **Zähler** (Nenner)der Verhältniszahl stehende Größe vorgegeben ist.

   Bsp.: Ein PKW legt eine Gesamtstrecke in vier Teilstrecken zurück. Will man die Durchschnittsgeschwindigkeit für die Gesamtstrecke berechnen, so muss man das **harmonische Mittel** (arithmetische Mittel) verwenden, wenn **die jeweiligen Längen der Teilstrecken** (die für die Teilstrecken jeweils benötigte Zeit) vorgegeben ist.

Formelinstrumentarium:
Gegeben seien n Beobachtungswerte $x_i (i = 1,2,...,n)$ eines Merkmals X. Für das **ungewogene harmonische Mittel** gilt:

$$x_{harm} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Gegeben sei die Häufigkeitsverteilung des Merkmals X, dessen Ausprägungen $a_j (j = 1,2,...,m)$ mit absoluten bzw. relativen Häufigkeiten $h(a_j)$ bzw. $f(a_j)$ auftreten. Für das **gewogene harmonische Mittel** gilt:

$$x_{harm} = \frac{n}{\sum_{j=1}^{m} \frac{h(a_j)}{a_j}} = \frac{1}{\sum_{j=1}^{m} \frac{f(a_j)}{a_j}}$$

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Aufgaben zu Auswertungsmethoden für eindimensionales Datenmaterial

Aufgabe 1

Um einen Überblick über die Studiendauer im Fach Betriebswirtschaftslehre an der Technischen Universität zu bekommen, wurden die 200 Universitätsabgänger des letzten Semesters nach ihrer Studiendauer befragt.

Nachstehende Tabelle zeigt das Ergebnis der Erhebung in Form relativer Häufigkeiten.

Semesterzahl	10	11	12	13	14	15
$f(a_j)$	0,1	0,1	0,4	0,2	0,15	0,05

a) Wie heißt das statistische Merkmal und wie ist es skaliert?
b) Bestimmen Sie die absoluten Häufigkeiten!
c) Bestimmen Sie die absoluten kumulierten Häufigkeiten!
d) Stellen Sie die absoluten und die absoluten kumulierten Häufigkeiten graphisch dar

Aufgabe 2

Bei einer Untersuchung wurde die Körpergröße (gemessen in cm) von 50 weiblichen Personen festgestellt. Nachfolgend ist Ihnen das Ergebnis dieser Untersuchung in Form der Urliste gegeben:

145; 177; 152; 159; 160; 160; 163; 167; 160; 160; 164; 174; 160; 177;
157; 164; 173; 179; 156; 150; 165; 167; 153; 177; 171; 175; 149; 163;
148; 162; 156; 154; 171; 172; 153; 168; 158; 170; 158; 167; 162; 164;
166; 177; 166; 152; 175; 152; 155; 165

a) Wie heißt das statistische Merkmal und wie ist es skaliert?

b) Bestimmen Sie die absoluten und die relativen Häufigkeiten sowie die absoluten und die relativen kumulierten Häufigkeiten!

c) Stellen Sie die absoluten Häufigkeiten und die relativen kumulierten Häufigkeiten graphisch dar!

d) Bestimmen Sie das arithmetische Mittel und den Median der Häufigkeitsverteilung!

e) Bestimmen Sie die empirische Varianz, die Standardabweichung und die Spannweite der Häufigkeitsverteilung!

f) Gruppieren Sie das vorliegende Datenmaterial wie folgt (jeweils: von ... bis unter):

140-150; 150-160; 160-170; 170-180

und bestimmen Sie erneut die absoluten und die relativen sowie die absoluten kumulierten und die relativen kumulierten Häufigkeiten. Berechnen Sie abschließend das arithmetische Mittel und den Median der Häufigkeitsverteilung? Was stellen Sie fest?

Aufgabe 3

Barbara fährt mit dem Bus zur Schule. Um auf dem Schulweg beschäftigt zu sein, notiert sich die zahlenbegeisterte Barbara ihre Wartezeiten auf den Bus (in Minuten).

Barbara, die noch keine Statistik hat, bittet nun Sie, einige statistische Analysen vorzunehmen und überreicht Ihnen dazu die Liste der Wartezeiten des vergangenen Monats:

$$
\begin{array}{lllll}
x_1 = 4; & x_2 = 5; & x_3 = 5; & x_4 = 9; & x_5 = 3 \\
x_6 = 4; & x_7 = 9; & x_8 = 6; & x_9 = 7; & x_{10} = 3 \\
x_{11} = 8; & x_{12} = 2; & x_{13} = 8; & x_{14} = 3; & x_{15} = 7 \\
x_{16} = 3; & x_{17} = 3; & x_{18} = 2; & x_{19} = 6; & x_{20} = 3
\end{array}
$$

1. Wie lautet das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie die absoluten, die relativen sowie die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die relativen Häufigkeiten graphisch dar!
4. Zeichnen Sie die (empirische) Verteilungsfunktion!
5. Bestimmen Sie das arithmetische Mittel, den Median und den Modalwert der vorliegenden Häufigkeitsverteilung!
6. Bestimmen Sie die empirische Varianz, die durchschnittliche Abweichung und die Spannweite der vorliegenden Häufigkeitsverteilung!
7. An wie viel Tagen wartete Barbara mehr als 5 Minuten auf den Bus?
8. Diskutieren Sie - auch anhand selbst gewählter Beispiele - die grundsätzliche Problematik statistischer Maßzahlen (Lage- und Streuungsparameter)!

Aufgabe 4

Am Quartalsende hat die statistische Abteilung einer Sparkasse den Wertpapierbesitz von 100 ihrer Kunden in folgender Tabelle zusammengestellt:

Höhe des Wertpapierbesitzes (in Tsd. €) von... bis unter	**Anzahl der Wertpapierbesitzer** h_j
0 – 10	20
10 – 30	30
30 – 50	30
50 – 70	10
70 –110	10

1. Wie heißt das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die absoluten Häufigkeiten graphisch dar!
4. Zeichnen Sie die empirische Verteilungsfunktion!
5. Berechnen Sie das arithmetische Mittel der Häufigkeitsverteilung!
6. Wie groß ist nach dem vorliegenden Datenmaterial das Gesamtvermögen der Wertpapierbesitzer?

Der Vorstand der Sparkasse ist mit den Ergebnissen nicht einverstanden und gruppiert das Datenmaterial erneut.

Höhe des Wertpapierbesitzes (in Tsd. €) von... bis unter	Anzahl der Wertpapierbesitzer h_j
0 - 20	30
20 - 40	30
40-120	40

7. Wie groß ist nach der Analyse des Vorstandes das Gesamtvermögen der Wertpapierbesitzer?

Aufgabe 5
Der Absatz der Waren A und B (Angaben jeweils in Stück) wächst von 2013 bis 2017 wie folgt:

Jahr	**2013**	**2014**	**2015**	**2016**	**2017**
Absatz der Ware A	20.000	35.000	42.000	52.500	63.000
Absatz der Ware B	20.000	15.000	22.000	36.000	28.000

Berechnen Sie jeweils die durchschnittliche Zuwachsrate im Absatz für die Waren A und B.

Aufgabe 6
Ein PKW legt vier Teilstrecken (die Länge jeder Teilstrecke enthält nachstehende Tabelle) einer Gesamtstrecke mit folgenden Geschwindigkeiten zurück:

Teilstrecke	**1**	**2**	**3**	**4**
Länge der Teilstrecke in km	30	10	40	20
Geschwindigkeit auf Teilstrecke in km/h	40	50	80	100

Mit welcher Durchschnittsgeschwindigkeit hätte der PKW die Gesamtstrecke in der gleichen Zeit bewältigt?

Aufgabe 7
Ein PKW legt vier Teilstrecken (die für jede Teilstrecke benötigte Zeit ist in nachstehender Tabelle wiedergegeben) einer Gesamtstrecke mit folgenden Geschwindigkeiten zurück:

Teilstrecke	**1**	**2**	**3**	**4**
für die Teilstrecke benötigte Zeit in h	0,75	0,2	0,5	0,2
Geschwindigkeit auf Teilstrecke in km/h	40	50	80	100

Mit welcher Durchschnittsgeschwindigkeit hätte der PKW die Gesamtstrecke in der gleichen Zeit bewältigt?

Aufgabe 8

Während der Fußball-WM gibt es in einigen Haushalten, die weder ein Videogerät noch ein zweites Fernsehgerät besitzen, Unstimmigkeiten wegen des Fernsehprogramms. Um festzustellen, wie viele Stunden pro Spieltag ein Fußballfan die Berichterstattung im Fernsehen verfolgt, werden 20 Fußballfans nach ihrem Fernsehkonsum während einer Fußball-WM befragt. Die Befragung erbrachte folgendes Ergebnis:

Fernsehkonsum	2	3	4	5	6
absolute Häufigkeit	1	2	8	4	5

1. Wie heißt das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die relativen Häufigkeiten sowie die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die absoluten Häufigkeiten grafisch dar!
4. Zeichnen Sie die empirische Verteilungsfunktion!
5. Bestimmen Sie das arithmetische Mittel, den Median und den Modus aus den vorliegenden Daten!
6. Bestimmen Sie die Spannweite aus den vorliegenden Daten!
7. Bestimmen Sie die durchschnittliche absolute Abweichung und die empirische Varianz!
8. Geben Sie an, wie viele Stunden 85% der Befragten mindestens fernsehen!
9. Welche Skalierungsart muss gegeben sein, um die empirische Verteilungsfunktion sinnvoll zeichnen zu können! Begründen Sie bitte!

Aufgabe 9

Herr Maier besitzt einen Großhandel. Am Ende des Geschäftsjahres möchte er einen Überblick über die Geschäftslage erhalten und stellt deshalb die Informationen über die innerhalb des letzten Jahres eingegangenen Aufträge zu nachstehender Tabelle zusammen:

Auftragshöhe (in 1 000 €) von ... bis unter	**Anzahl der Aufträge**
1 - 4	10
4 - 10	30
10 - 12	40
12 – 16	20

1. Wie heißt die statistische Größe und wie ist diese skaliert?
2. Stellen Sie die relativen, die absoluten kumulierten sowie die relativen kumulierten Häufigkeiten tabellarisch dar!
3. Stellen Sie die absoluten Häufigkeiten grafisch dar! Wie heißt die von Ihnen gewählte Darstellungsform? Welches Prinzip haben Sie bei der Darstellung berücksichtigt?
4. Zeichnen Sie die empirische Verteilungsfunktion! Welche Annahme treffen Sie dabei für jede Gruppe?
5. Welche Annahme treffen Sie, wenn Sie aus gruppiertem Datenmaterial das arithmetische Mittel bestimmen?
6. Berechnen Sie das arithmetische Mittel und schätzen Sie den Median grafisch aus der Verteilungsfunktion!
7. Berechnen Sie die durchschnittliche absolute Abweichung und die empirische Varianz, welche Annahme treffen Sie dabei für jede Gruppe?

Aufgabe 10

Aus der Kriminalstatistik des Monats Juni wurden folgende Daten bezüglich der Steuerkriminalität entnommen (Schadenshöhe in 10 000 €):

Höhe des Schadens (in 10 000 €) von ... bis unter	**Anzahl**
1,5 – 2,5	1
2,5 – 3,5	1
3,5 – 4,5	5
4,5 – 7,5	3

1. Wie heißt das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die relativen sowie die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Bestimmen Sie aus den vorliegenden Daten das arithmetische Mittel! Welche Annahme treffen Sie bei der Berechnung des arithmetischen Mittels?
4. Stellen Sie die absoluten Häufigkeiten grafisch dar! Wie heißt die von Ihnen gewählte Darstellungsform? Welches Prinzip haben Sie bei der Darstellung hoffentlich beachtet?
5. Zeichnen Sie die empirische Verteilungsfunktion! Welche Annahme treffen Sie hier?
6. Schätzen Sie grafisch den Median der vorliegenden Häufigkeitsverteilung!
7. Welche Skalierungsart muss gegeben sein, um die kumulierten Häufigkeiten sinnvoll tabellarisch bestimmen zu können? Begründen Sie bitte!

Aufgabe 11

Aus der Kriminalstatistik des Monats April wurden folgende Daten bezüglich der Steuerkriminalität entnommen (Schadenshöhe in 10 000 €):

3 ; 7 ; 6 ; 5 ; 6 ; 4,5 ; 4,5 ; 5 ; 5 ; 4

Hinweis: Lassen Sie sich nicht durch die beiden „Kommawerte“ irritieren!

1. Wie lautet das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die absoluten, die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die absoluten Häufigkeiten grafisch dar!
4. Zeichnen Sie die empirische Verteilungsfunktion!
5. Berechnen Sie das arithmetische Mittel, und bestimmen Sie den Median und den Modus der vorliegenden Häufigkeitsverteilung!
6. Bestimmen Sie die Spannweite der vorliegenden Häufigkeitsverteilung!
7. Welche Skalierungsart muss mindestens gegeben sein, um die empirische Verteilungsfunktion sinnvoll **zeichnen** zu können! Begründen Sie bitte!

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Aufgabe 1

Bei Piloten kommt es darauf an, dass sie möglichst schnell auf optische Signale reagieren. Um die Reaktionszeit zu verbessern, wurde eine Trainingsmethode entwickelt. Eine Untersuchung an 10 Personen ergab folgende Reaktionszeiten vor Anwendung und nach Anwendung der Trainingsmethode (in Sekunden):

Person	1	2	3	4	5	6	7	8	9	10
Reaktionszeit vor der Maßnahme	14	9	57	15	19	38	43	29	16	17
Reaktionszeit nach der Maßnahme	10	11	20	10	15	30	32	30	16	17

Prüfen Sie mit Hilfe von Boxplots, ob die Trainingsmethode eine positive Wirkung hat! Begründen Sie Ihre Aussage.
Boxplot: Minimum, Maximum, unteres Quartil (25%-Wert), Median (50%-Wert), oberes Quartil (75%-Wert), Spannweite, Quartilsabstand

Aufgabe 2

Ein Pharmaunternehmen möchte wissen, ob sein neu entwickeltes Medikament gegen Bluthochdruck besser wirkt als ein älteres Medikament. Dazu wurde eine Gruppe von 21 Probanden untersucht, davon bekamen 10 das ältere und die restlichen 11 Probanden das neue Medikament. Der Blutdruck wurde jeweils zu Beginn des Tests und zwei Wochen später gemessen. Erfasst wurde jeweils die Höhe der Blutdrucksenkung:

Mit dem alten Medikament: 7; 15; 11; 15; 25; 4; 19; 10; 8; 31
Mit dem neuen Medikament: 11; 16; 25; 9; 36; 40; 48; 4; 26; 36; 14

Prüfen Sie mit Hilfe von Boxplots, ob das neue Medikament eine bessere Wirkung hat als das alte Medikament. Begründen Sie Ihre Aussage.
Boxplot: Minimum, Maximum, unteres Quartil (25%-Wert), Median (50%-Wert), oberes Quartil (75%-Wert), Spannweite, Quartilsabstand

Aufgabe 3

Die folgende Messreihe zeigt uns die Eigenkapitalrentabilität von den 20 Unternehmen einer Branche:

5	**16**	**11**	**19**	**14**
7	**14**	**6**	**6**	**28**
13	**8**	**9**	**11**	**7**
14	**7**	**13**	**26**	**9**

1. Fertigen Sie ein Standard-Box-Whisker-Plot (Ausreißer werden zunächst nicht beachtet)!

   Boxplot: Minimum, Maximum, unteres Quartil (25%-Wert), Median (50%-Wert), oberes Quartil (75%-Wert), Spannweite, Interquartilsabstand (IQR)

2. Fertigen Sie ein Box-Whisker-Plot unter Kennzeichnung von Ausreißern, wobei die Länge eines Whisker auf 1,5*IQR beschränkt ist (das ist eine oft zitierte Definition eines Ausreißers nach Tukey) oder anders ausgedrückt: Ausreißer sind Werte, die kleiner sind als $x_{0,25} - 1{,}5 \cdot IQA$ oder größer sind als $x_{0,75} + 1{,}5 \cdot IQA$.

Aufgabe 4

Die folgende Messreihe zeigt uns den Bruttostunden-Verdienst von 20 Mitarbeitern unseres Unternehmens:

30	**28**	**27**	**29**	**27**
29	**23**	**30**	**29**	**28**
28	**27**	**30**	**25**	**29**
21	**29**	**26**	**30**	**30**

Fertigen Sie ein Standard-Box-Whisker-Plot…

Boxplot: Minimum, Maximum, unteres Quartil (25%-Wert), Median (50%-Wert), oberes Quartil (75%-Wert), Spannweite, Interquartilsabstand (IQA)

… unter Kennzeichnung von Ausreißern, wobei die Länge eines Whisker auf $1{,}5 \cdot$ IQA beschränkt ist (das ist eine oft zitierte Definition eines Ausreißers nach Tukey) oder anders ausgedrückt: Ausreißer sind Werte, die kleiner sind als $x_{0,25} - 1{,}5 \cdot IQA$ oder größer sind als $x_{0,75} + 1{,}5 \cdot IQA$.

Hinweis: Es ist nur **EIN** Box-Whisker-Plot zu zeichnen!

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Klausurtests zu Auswertungsmethoden für eindimensionales Datenmaterial

Klausurtest 1 – Deskriptive Statistik

Aufgabe 1

Student Alois besitzt eine Erdbeerplantage. Da das Ernteergebnis je nach Qualität des Sommers verschieden ist, notierte sich Alois, wie viel Stunden die Sonne in der diesjährigen Saison pro Tag auf seine Beeren einwirkte:

Sonnenstunden pro Tag von ... bis unter	Anzahl der Tage
0 - 2	20
2 - 3	15
3 - 5	20
5 - 8	35
8 - 12	10

1. Wie lautet das statistische Merkmal und wie ist es skaliert?
2. Stellen Sie die absoluten Häufigkeiten grafisch dar? Wie heißt die von Ihnen gewählte Darstellungsform? Welches Prinzip haben Sie dabei beachtet?
3. Zeichnen Sie die empirische Verteilungsfunktion! Welche Annahme treffen Sie beim Zeichnen der Verteilungsfunktion?
4. Schätzen Sie grafisch den Median.
5. Berechnen Sie das arithmetische Mittel der Häufigkeitsverteilung! Welche Annahme treffen Sie jetzt?
6. An wie viel Tagen der Saison schien die Sonne mindestens 4 Stunden?
7. Wie lange höchstens schien die Sonne an den 40 sonnenärmsten Tagen?

Aufgabe 2
In einem Schachverein haben sich 20 Mitglieder an einem Schachrätsel probiert. Dabei ergaben sich folgende Lösungszeiten (in Minuten):

$$x_1 = 7 \quad x_2 = 9 \quad x_3 = 5 \quad x_4 = 6 \quad x_5 = 6 \quad x_6 = 4 \quad x_7 = 8 \quad x_8 = 6 \quad x_9 = 7 \quad x_{10} = 7$$
$$x_{11} = 6 \quad x_{12} = 7 \quad x_{13} = 8 \quad x_{14} = 7 \quad x_{15} = 3 \quad x_{16} = 8 \quad x_{17} = 6 \quad x_{18} = 6 \quad x_{19} = 7 \quad x_{20} = 7$$

1. Wie lautet das statistische Merkmal und wie ist es skaliert?

2. Bestimmen Sie die absoluten, die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!

3. Stellen Sie die absoluten Häufigkeiten grafisch dar!

4. Zeichnen Sie die empirische Verteilungsfunktion.

5. Berechnen Sie das arithmetische Mittel und bestimmen Sie den Median der Häufigkeitsverteilung!

6. Fassen Sie die erhaltenen Werte in folgende Gruppen zusammen:

von ... bis unter
0 - 4
4 - 6
6 - 8
8 - 14

   und bestimmen Sie für das gruppierte Material die absoluten, die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!

7. Berechnen Sie das arithmetische Mittel für das gruppierte Material!

Aufgabe 3

20 Beschäftigte wurden nach der Entfernung ihres Wohnsitzes zum Arbeitsplatz (in km) befragt. Das Ergebnis der Befragung ist in nachfolgender Tabelle wiedergegeben:

X: „Entfernung des Wohnsitzes zum Arbeitsplatz (in km)	2	3	4	5	6	7
absolute Häufigkeit	5	4	1	3	3	4

1. Wie lautet das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die absoluten Häufigkeiten grafisch dar!
4. Zeichnen Sie die empirische Verteilungsfunktion!
5. Berechnen Sie das arithmetische Mittel, und bestimmen Sie den Median der Häufigkeitsverteilung! Welchen Wert hat der Modus? (Denken Sie an die Angabe der Einheit!)
6. Bestimmen Sie die Spannweite der vorliegenden Häufigkeitsverteilung! (Denken Sie an die Angabe der Einheit!)
7. Berechnen Sie die empirische Varianz und die Standardabweichung! (Denken Sie an die Angabe der Einheit!)
8. Berechnen Sie die durchschnittliche absolute Abweichung! (Denken Sie an die Angabe der Einheit!)
9. Welche Skalierungsart muss mindestens gegeben sein, um die empirische Verteilungsfunktion sinnvoll zeichnen zu können? Begründen Sie bitte!
10. Welche Skalierungsart muss mindestens gegeben sein, um die kumulierten Häufigkeiten sinnvoll tabellarisch bestimmen zu können? Begründen Sie bitte!

Aufgabe 4
Um eine Übersicht über das Zahlungsverhalten seiner Kunden zu erhalten, lässt der Finanzvorstand eines Unternehmens die Daten von 100 Kunden erheben. Das Ergebnis ist in nachstehender Tabelle zusammengefasst und soll nun von Ihnen analysiert werden:

Wochen bis zur Begleichung der Rechnung von ... bis unter	**Anzahl der Kunden (absolute Häufigkeit)**
2 - 4	10
4 - 10	30
10 - 12	40
12 - 16	20

1. Wie heißt die statistische Größe und wie ist diese skaliert?
2. Stellen Sie die relativen, die absoluten kumulierten sowie die relativen kumulierten Häufigkeiten tabellarisch dar!
3. Berechnen Sie das arithmetische Mittel der vorliegenden statistischen Größe! (Denken Sie an die Angabe der Einheit!)
4. Stellen Sie die absoluten Häufigkeiten grafisch dar! Wie heißt die von Ihnen gewählte Darstellungsform? Welches Prinzip haben Sie bei der Darstellung berücksichtigt?
5. Zeichnen Sie die empirische Verteilungsfunktion und schätzen Sie anschließend grafisch den Median aus der Verteilungsfunktion! (Denken Sie an die Angabe der Einheit!)
6. Berechnen Sie den Median mit Hilfe der Formel: $x_{med} = x_e^u + \frac{0{,}5 - F\left(x_e^u\right)}{f\left(x_e\right)} \cdot \left(x_e^o - x_e^u\right)$
7. Ermitteln Sie rechnerisch den Interquartilsabstand!
   Hinweis: $x_{0,25}$ und $x_{0,75}$ bestimmen Sie bitte analog zu x_{med} **!**

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Klausurtest 2 - Mittelwerte

Aufgabe 1

Ein in seiner Höhe nicht näher spezifiziertes Kapital werde in den ersten beiden Jahren mit 4,5%, im dritten bis sechsten Jahr mit jeweils 8%, im siebenten bis zwölften Jahr mit jeweils 11% verzinst. Bestimmen Sie den durchschnittlichen Zinssatz für den angegebenen Zeitraum!

Aufgabe 2

Studentin Carola hat an den sieben Tagen der letzten Woche folgende Mengen Milch getrunken (Angaben in Liter):

2,4; 1,0; 0,5; 2,0; 1,5; 1,2; 1,9

Berechnen Sie den durchschnittlichen Milchkonsum von Carola!

Aufgabe 3

Ein PKW-Fahrer legt von einer Gesamtstrecke 20% mit einer Geschwindigkeit von 100 km/h, 30% mit einer Geschwindigkeit von 90 km/h und 50% mit einer Geschwindigkeit von 50 km/h zurück. Mit welcher (konstanten) Durchschnittsgeschwindigkeit würde er die gesamte Strecke in der gleichen Zeit bewältigen?

Musteraufgabe: Geometrisches Mittel

Ein in seiner Höhe nicht näher spezifiziertes Kapital werde im ersten Jahr mit 5%, im zweiten bis fünften Jahr mit jeweils 7%, im sechsten bis zehnten Jahr mit jeweils 10% und im elften bis fünfzehnten Jahr mit 12% verzinst. Bestimmen Sie den durchschnittlichen Zinssatz für den angegebenen Zeitraum!

$$\begin{array}{lll} p_1 = 5 & q_1 = 1{,}05 & n_1 = 1 \\ p_2 = 7 & q_2 = 1{,}07 & n_2 = 4 \\ p_3 = 10 & q_3 = 1{,}10 & n_3 = 5 \\ p_4 = 12 & q_4 = 1{,}12 & n_4 = 5 \end{array}$$

Und damit ergibt sich:

$$q_d = \sqrt[15]{1{,}05^1 * 1{,}07^4 * 1{,}1^5 * 1{,}12^5} = 1{,}0951$$

$$p_d = (1{,}0951 - 1) * 100 = 9{,}51$$

Der Durchschnittszinssatz beträgt also 9,51%.

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Theorie zu Konzentrationsrechnung

<u>Hinweis:</u>

Zur relativen Konzentration mit Hilfe der Lorenzkurve und dem Ginikoeffizient werden keine gesonderten theoretischen Hinweise gegeben.

Absolute Konzentration mit Hilfe des Herfindahl-Index

Ein Maß für die Messung der absoluten Konzentration ist der Herfindahl-Index H. Dieser ist definiert als: $H = \sum_{i=1}^{n} p_i^2$, wobei p_i den Anteil darstellt, den die i-te statistische Einheit auf sich vereinigt.

Rechnerisch: $$p_i = \frac{Beitrag\ \ der\ \ i-ten\ \ statistischen\ \ Einheit}{Summe\ \ der\ Beitäge\ \ aller\ \ statistischen\ \ Einheiten}$$

Der kleinste Wert, den H annehmen kann, hängt ab von der Anzahl n der Marktteilnehmer (Unternehmen), denn es gilt: $\frac{1}{n} \leq H \leq 1$. Die Konzentration ist dabei umso höher, je höher der Wert von H ist.

Der Herfindahlindex H wird zu $\frac{1}{n}$, wenn alle n Marktteilnehmer (Unternehmen) den gleichen Beitrag aufweisen. H wird 1, wenn es nur noch ein Unternehmen gibt.

Fusionieren zwei Unternehmen mit den Anteilen p_1 *und* p_2, so erhöht sich der Wert von H, d. h. die Konzentration nimmt zu; bleiben zudem die Anteile der anderen Marktteilnehmer / Unternehmen unverändert, so erhöht sich der Herfindahl-Index um $\Delta H = 2 p_1 \, p_2$, wobei ΔH die Veränderung von H bedeutet. (Warum ist die Veränderung so?)

Beispiel für die Berechnung von H:

Gegeben sei die statistische Größe X: „Umsatz auf dem Markt für ...“. Die vier beteiligten Unternehmen gaben an: $x_1 = 100;\ x_2 = 400;\ x_3 = 300;\ x_4 = 200$.

Damit ergeben sich folgende Marktanteile:

$$p_i = \frac{x_i}{\sum x_i} \quad und \quad damit$$

$$p_1 = \frac{100}{1000} = 0{,}1; \quad p_2 = \frac{400}{1000} = 0{,}4; \quad p_3 = \frac{300}{1000} = 0{,}3; \quad p_4 = \frac{200}{1000} = 0{,}2$$

Und damit wird H zu:

$$H = 0{,}1^2 + 0{,}4^2 + 0{,}3^2 + 0{,}2^2 = 0{,}3$$

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Aufgaben zu Konzentrationsrechnung

Aufgabe 1
Folgende Tabelle enthält die Spareinlage von 100 Kleinsparern einer Gemeinde.

Sparguthaben in € von ... bis unter	**Anzahl der Sparer** h_j
0 - 1 000	400
1 000 - 2 000	100
2 000 - 4 000	150
4 000 - 6 000	200
6 000 - 10 000	150

1. Bestimmen Sie aufgrund der vorliegenden Daten die Lorenz´sche Konzentrationsverteilung! Von welcher Annahme gehen Sie dabei innerhalb einer Klasse aus?
2. Schätzen Sie aufgrund der vorliegenden Daten, wie viel die 50% "kleinsten" Kleinsparer gespart haben!
3. Schätzen Sie aufgrund der vorliegenden Daten, wie viel Prozent der Einlagen auf die 50% "reichsten" Kleinsparer entfällt!
4. Wie viel haben die Kleinsparer aufgrund der vorliegenden Daten insgesamt gespart? Welche Annahme haben Sie dabei getroffen?
5. Berechnen Sie das arithmetische Mittel der vorliegenden Häufigkeitsverteilung! Interpretieren Sie den dann erhaltenen Wert inhaltlich!
6. Auf einem Markt hat das Unternehmen A einen Umsatz von 10 Mill. €, die Unternehmen B und C weisen jeweils einen Umsatz von 20 Mill. € auf. Berechnen Sie aufgrund dieser Daten den Herfindahl-Index!
7. Diskutieren Sie kurz die Problematik der relativen Konzentrationsmessung mithilfe der Lorenzkurve und dem Ginikoeffizient!

Aufgabe 2
Die Unternehmensberatungsfirma ABC führte im letzten Jahr 100 Beratungen durch. Je nach Schwere des Falles war die Beratungsdauer unterschiedlich lang.

Folgende Tabelle zeigt die Verteilung der Beratungsdauer in Stunden:

Zeitbedarf in Wochen von ... bis unter	**Anzahl der Beratungsfälle** h_j
0 - 10	20
10 - 30	20
30 - 50	30
50 - 170	30

1. Bestimmen Sie aufgrund der Daten die Lorenz´sche Konzentrationsverteilung und zeichnen Sie die Lorenzkurve!

2. Welche Verteilungsannahme innerhalb der Gruppen liegt der Konzentrationsbetrachtung nach Lorenz zugrunde?

3. Schätzen Sie aufgrund der vorliegenden Daten, wie viel Prozent der Beratungszeit auf die 50% zeitintensivsten Fälle entfiel!

4. Berechnen Sie das arithmetische Mittel der Häufigkeitsverteilung und interpretieren Sie den erhaltenen Wert inhaltlich!

5. Wie viel Beratungszeit hat die Firma ABC aufgrund der vorliegenden Daten insgesamt geleistet?

Neben der Firma ABC (100 Beratungsfälle) gibt es auf dem Unternehmensberatungsmarkt noch die Firmen XYZ und GfUB. Die XYZ führte 50 Beratungen durch, die GfUB 150 Beratungen.

6. Ermitteln Sie für den betrachteten Markt Unternehmensberatungen den Herfindahl-Index!

7. Um welchen Betrag ändert sich der Herfindahl-Index, wenn die XYZ mit der GfUB fusionieren und sich ihre Marktanteile dabei addieren?

Aufgabe 3
Folgende Tabelle enthält die Umsatzzahlen (in Mio. €) für die Unternehmen auf den Märkten A, B und C:

	Umsatz (in Mio. €)
Markt A	100; 100; 600; 600; 600
Markt B	50; 50; 50; 50; 50; 50; 50; 50; 300; 300
Markt C	50; 50; 50; 50; 1 800

1. Berechnen Sie für jeden Markt die Lorenz´sche Konzentrationsverteilung und zeichnen Sie dann alle Lorenzkurven in ein Diagramm! Zeichnen Sie bitte genügend groß, damit Sie über-sichtlich bleiben!

2. Welche Märkte sind bezüglich ihrer Lorenzkurven nicht vergleichbar? Begründen Sie bitte Ihre Antwort!

3. Ordnen Sie die Märkte hinsichtlich der Stärke der Konzentration, indem Sie als Vergleichskriterium die Größe des Ginikoeffizient betrachten, ohne die einzelnen Ginikoeffizient konkret auszurechnen!

4. Wie groß ist zahlenmäßig der Ginikoeffizient auf dem Markt D, auf dem die 100 beteiligten Unternehmen jeweils einen Umsatz von 20 Mio. € aufweisen?

5. Wie groß ist zahlenmäßig der Ginikoeffizient auf dem Markt E, auf dem die fünf beteiligten Unternehmen jeweils einen Umsatz von 200 Mio. € aufweisen?

6. Berechnen Sie für die Märkte A, B und C jeweils den Herfindahl-Index!

7. Auf dem Markt A fusionieren die beiden Unternehmen mit Umsatz 100 Mio. € - es gibt jetzt also ein Unternehmen mit dem Umsatz 200 Mio. €. Um welchen Betrag ändert sich der Herfindahl-Index?

8. Diskutieren Sie kurz die Problematik der relativen Konzentrationsmessung mithilfe der Lorenzkurve und dem Ginikoeffizient!

Aufgabe 4
Eine Unternehmensberatung im Land A führte im letzten Jahr 100 Beratungsgespräche durch. Da je nach Problemstellung die Gesamtberatungsdauer unterschiedlich lang war, soll für die Personalplanung die Verteilung der Beratungsdauer (in Stunden) analysiert werden:

Beratungsdauer von ... bis unter	**Anzahl der Gespräche**
0 - 10	20
10 - 30	20
30 - 50	30
50 - 170	30

1. Zeichnen Sie für den vorliegenden Sachverhalt die Lorenzkurve!
2. Wie viel Prozent der Beratungszeit entfällt auf die 30% zeitintensivsten Fälle?
3. Wie viel Prozent der Beratungszeit fallen auf die hinsichtlich der Beratungsdauer mittleren 50% der Fälle?
4. Gehen Sie davon aus, dass jeder Angestellte 1840 Stunden im Jahr arbeitet und die Gesamt-Arbeitsbelastung im nächsten Jahr gleich bleibt. Wie viele Angestellte sind zu beschäftigen?

Im Land A gebe es insgesamt fünf Beratungsunternehmen. Das obige hält 100 Kunden, zwei weitere beraten jeweils 50 Kunden sowie die restlichen zwei jeweils 400 Kunden.
1. Errechnen Sie den Herfindahl-Index!
2. Die beiden großen Unternehmen im Land A fusionieren, Wie groß ist nun der Herfindahl-Index, wenn Sie davon ausgehen können, dass sich deren Marktanteile addieren und die anderen Marktanteile konstant bleiben?

Aufgabe 5
Gegeben seien die Jahresumsätze (in Mio. €) von den 10 Unternehmen einer Branche:

$$x_1 = 20; \quad x_2 = 30; \quad x_3 = 20; \quad x_4 = 15; \quad x_5 = 30$$
$$x_6 = 15; \quad x_7 = 40; \quad x_8 = 15; \quad x_9 = 15; \quad x_{10} = 50$$

1. Berechnen Sie die Lorenz´sche Konzentrationsverteilung und zeichnen Sie die Lorenzkurve! Ist der Gini-Koeffizient gleich =0?
2. Welcher Anteil am Gesamtumsatz entfällt auf die 80% umsatzschwächsten Unternehmen?
3. Wie groß ist der Anteil am Gesamtumsatz, der auf die „mittleren“ 60% der Unternehmen entfällt?
4. Berechnen Sie die Konzentrationsraten CR1, CR3, CR5 und CR8!
5. Berechnen Sie den Herfindahl-Index!
6. Um welchen Betrag ändert sich – unter sonst gleichen Bedingungen (d. h. nicht ändernden Umsätzen) – der Herfindahl-Index, wenn die beiden Unternehmen mit Umsatz 30 Mio. € fusionieren, wobei sich deren Umsätze addieren?

Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.
Dipl.-Kfm. Thomas Rochow

Klausurtests zu Konzentrationsrechnung

Aufgabe 1

Sie wollen auf dem Markt, auf dem das Unternehmen ABC tätig ist, eine Konkurrenzanalyse vornehmen, das Unternehmen ABC hat vier Konkurrenten, die Umsätze des letzten Jahres sind Ihnen sortiert in nachstehender Auflistung gegeben, das Unternehmen ABC ist das umsatzstärkste Unternehmen (alle Angaben in Mio. €):

$$x_1 = 3; \quad x_2 = 3 \quad x_3 = 4 \quad x_4 = 4; \quad x_5 = 11$$

1. Berechnen Sie die Lorenzsche Konzentrationsverteilung und zeichnen Sie die Lorenzkurve!
2. Welcher Anteil am Gesamtumsatz entfällt auf die 40% umsatzschwächsten Unternehmen? Bestimmen Sie diesen Anteil mit Hilfe der erstellten Tabelle!
3. Berechnen Sie den Gini-Koeffizient $(Gini\ oder\ \ G)$ und den normierten Gini-Koeffizient $(Gini^*\ \ oder\ \ G^*)$!

$$Gini = G \quad = \frac{2\sum_{i=1}^{n} i \cdot x_i}{n \cdot \sum_{i=1}^{n} x_i} - \frac{n+1}{n}$$

$$Gini^* = \quad G^* \quad = \frac{n}{n-1} \cdot Gini$$

Aufgabe 2
Gegeben seien die Jahresumsätze (in Mio. €) der 10 Unternehmen eines kleinen Landes:

$$x_1 = 8; \quad x_2 = 12; \quad x_3 = 8; \quad x_4 = 12; \quad x_5 = 10;$$
$$x_6 = 8; \quad x_7 = 12; \quad x_8 = 8; \quad x_9 = 4; \quad x_{10} = 18$$

1. Berechnen und interpretieren Sie die Konzentrationsraten CR1, CR3, CR5 und CR9!

2. Berechnen Sie den Herfindahl-Index!

3. Die Unternehmen, die einen Umsatz mit 4 Mio.€ und 18 Mio.€ aufweisen, fusionieren nach der Fusion gibt es also folgende Marktaufteilung:

$$x_1 = 8; \quad x_2 = 12; \quad x_3 = 8; \quad x_4 = 12; \quad x_5 = 10;$$
$$x_6 = 8; \quad x_7 = 12; \quad x_8 = 8; \quad x_9 = 22$$

   Berechnen Sie den Herfindahl-Index erneut! Vergleichen Sie diesen mit der Situation unter 2.! Um welchen Wert hat sich der Herfindahl-Index geändert?