KLAUSURTRAINER

Deskriptive Statistik 1

Musteraufgaben mit Musterlösungen

SILVIO GERLACH, BORIS FÜHRER

STUDEO

Probeauszug

Vorwort

Wie und für wen dieser Klausurtrainer entstand

Statistik ist ein Grundlagenfach in Studiengängen wie BWL, VWL, Psychologie, Medizin, technische und Ingenieurwissenschaften sowie Sozialwissenschaften. Und es ist bei den meisten Prüfungskandidaten wegen seiner formalen Struktur nicht sehr beliebt.

Wir von Studeo® bereiten schon seit über 15 Jahren auf Prüfungen in Statistik vor. Dafür haben wir unser eigenes Trainingsmaterial entwickelt. Dieser Klausurtrainer ist der erste von zwei Klausurtrainern in Statistik und für jene gedacht, die sich zielgerichtet auf Statistik-Klausuren an Universitäten, Fachhochschulen und Fachschulen vorbereiten wollen.

Es gibt bereits sehr viele Aufgaben- und Übungsbücher für Statistik. Wieso also noch ein neues Buch?

Was ist neu an diesem Klausurtrainer?

In folgenden Punkten unterscheidet sich dieses Buch von den meisten anderen Büchern in Statistik:

Fachliche Inhalte von der Klausur her entwickelt und dargestellt!
Dieser Klausurtrainer ist kein Lehr- und auch kein Übungsbuch im herkömmlichen Sinne, sondern ein Klausurtrainer mit einem neuen didaktischen Ansatz. Basierend auf einer sorgfältigen Analyse und Systematisierung typischer Statistik-Klausuren wird der Lernstoff konsequent von der Klausur her dargestellt. Den Kern des Klausurtrainers bilden ausführliche Schritt-für-Schritt-Lösungsanleitungen für typische Aufgabenstellungen. Zahlreiche Übersichten wie Mindmaps, Formelsammlungen, Symbollisten und Glossare erleichtern den Einstieg.

Systematische Entwicklung von Aufgaben
Wir haben uns bemüht, eine Standardaufgabenstellung von möglichst vielen Seiten zu beleuchten und so die Zusammenhänge deutlich zu machen. Zwar tauchen in keiner Klausur jemals Aufgaben mit 32 Unterfragen auf. Doch indem wir so viele mögliche Varianten abarbeiten, ist die Wahrscheinlichkeit hoch, dass die drei oder vier Aufgaben der Klausur in unserem Katalog enthalten sind. So ist man auf der sicheren Seite.

Effektives Lernen beim Lernen lernen
Der Klausurtrainer verbindet Lerninhalte mit Lernorganisation. Der Aufbau des Lernstoffs in Form von Checklisten und Profilen hilft, den lehrstuhlabhängigen, relevanten Stoff selbständig zu ordnen und das Lernen selbst effektiver zu organisieren. Das spart Zeit bei der Vorbereitung und ermöglicht bei Befolgung der Hinweise Antworten oder wenigstens Teilantworten auf die stets bewegende Frage: "Was kommt dran?"

Personalisierung des Lernstoffes möglich
Da es Unterschiede in der Stoffauswahl, -darstellung und Aufgabenstellung zwischen Lehrstühlen gibt, wenn auch im Vordiplom noch nicht so große, müssen die Vorbereitungsunterlagen auf die Anforderungen des jeweiligen Lehrstuhls zugeschnitten werden. Mit unserem Klausurtrainer ist das möglich. Es kann und soll stets geprüft werden, ob die dargestellten Inhalte relevant und ob sie eventuell zu ergänzen sind. Bei konsequenter Überprüfung der Inhalte und Benutzung der Checklisten entsteht ein persönlicher Klausurtrainer als Kompass und Grundlage für die Klausurvorbereitung. (Siehe auch das Handbuch Klausur – für professionelle Klausurvorbereitung.)

Ziele dieser Klausurtraineres

Das Hauptziel dieser Klausurtraineres ist: **KLAUSURERFOLG!**

Das Buch soll Prüfungskandidaten im Fach Deskriptive Statistik in die Lage versetzen:

- Aufgabenstellungen und vor allem –varianten besser und schneller zu verstehen,
- Begriffe, Symbole, Formeln und Fragen richtig zuzuordnen,
- Den richtigen Lösungsansatz zu finden,
- Formeln und Rechenregeln sicher anzuwenden,
- Graphiken zu skizzieren,
- Ergebnisse richtig zu interpretieren und weiterzuverarbeiten und
- Inhaltliche Fragen richtig zu beantworten.

Da wir schon seit Jahren erfolgreich nach den Methoden dieses Buches auf Klausuren vorbereiten, sind wir überzeugt, dass sich der Erfolg bei konsequenter Vorbereitung mit dem Klausurtrainer einstellt.
Wir empfehlen zur Vorbereitung auch unser Handbuch Klausur – für professionelle Klausurvorbereitung (Infos auf www.studeo.de).

Inhalte und Methodik dieses Klausurtrainers

Wichtig: An den meisten Hochschulen wird Statistik in zwei Veranstaltungen gehalten, Statistik I und II. Mitunter entspricht das der Unterteilung in Deskriptive und Induktive Statistik, oft aber auch nicht. Wir haben zwei Klausurtrainer, einen für Deskriptive und einen zweiten Klausurtrainer für Induktive Statistik und Wahrscheinlichkeitsrechnung.

Dieser Klausurtrainer konzentriert sich auf die Standard-Themenbereiche der Deskriptiven Statistik an Hochschulen im Vordiplom sowie in wichtigen Lehrbüchern: Ein- und zweidimenionale Häufigkeitsverteilung, Konzentrationsrechnung, Indexrechnung und Zeitreihenanalyse.
Typische Aufgabenstellungen aus diesen Bereichen werden übersichtlich aufgelistet und ausführlich gelöst. Selbstverständlich kann der Klausurtrainer nicht den Anspruch erheben, alle relevanten Bereiche des jeweiligen Lehrstuhls abzudecken. Es ist daher äußerst wichtig, sich genau zu informieren, welche Anforderungen der betreffende Lehrstuhl stellt, welche Materialien relevant sind, sich diese zu "organisieren" und bei der Vorbereitung zu nutzen.

Hier sind einige Innovationen hinsichtlich der Inhaltsdarstellung:

- **Mindmaps als Kompass durch den Stoff.**
  Sie erlauben die übersichtliche Darstellung selbst komplexester Inhalte, ein wahrhaft intelligentes Instrument.

- **Systematik der Aufgabenvarianten zu den Themenbereichen.**
  Eine solche Systematik machen Dozenten, die eine Klausur stellen müssen, allerdings nur für sich im "stillen Kämmerlein". Wir machen solche Analysen in diesem Buch unseres Wissens erstmals öffentlich zugänglich.

- **Aufgabenstellungen eines Themenbereichs durch Unterfragen von vielen möglichen Seiten betrachten.**
  Das ist die Fortsetzung bzw. Umsetzung der Aufgabensystematik in den Musteraufgaben. Es ermöglicht, ein breites Aufgabenspektrum zur Auswahl der für die spezielle Klausur relevanten Fragen.

- **Systematik der verwendeten Symbole.**
  Solch eine Liste ist sogar in großen Lehrbüchern nicht immer enthalten.

- **Eine Formelsammlung der typischen Formeln.**
  Diese Sammlung ist für die Inhalte entwickelt worden. Wichtig ist hier, die Schreibweise an die des eigenen Lehrstuhls anzupassen oder sich wenigsten die Nummer aus der eigenen Formelsammlung dazu zu schreiben

- **Detaillierte Lösungen der Musteraufgaben – Schritt-für-Schritt**
  Wir versuchen, die Lösungen so elementar wie möglich zu halten und so viel wie nötig zu erklären.

Einige dieser Instrumente und Ansätze sind natürlich auch in anderen Büchern zu finden. Die Stärke unseres Klausurtrainers ist jedoch die Abstimmung der Elemente aufeinander und ihre methodische Ausrichtung auf das Ziel Klausurerfolg.

Wie man mit diesem Klausurtrainer arbeiten sollte

In der Einleitung findet sich eine Anleitung zum Arbeiten mit diesem Buch. Wir empfehlen auch dringend, sich in den "Niederungen des Rechnens" wieder fit zu machen, mit unserem Rechentrainer "Schlag auf Schlag – Rechnen bis ich's mag" (www.rechentrainer.de). Denn Termumformungen sind eine Hautfehlerquelle in Klausuren, nicht nur in Statistik.

Wir hoffen sehr, dass Ihnen unsere Anstrengungen helfen, dass Sie Ihnen bei der Klausurvorbereitung Zeit sparen und dass Sie die Klausur letztlich erfolgreich zu bestehen.

Danksagung

Wir danken unseren Kursteilnehmern, die uns durch unzählige Fragen zu diesem Buch inspiriert haben und ihm eine Down-to-earth-Richtung gaben.

Wir haben uns um größtmögliche Sorgfalt bemüht. Für alle verbleibenden Fehler und Unzulänglichkeiten sind wir allein verantwortlich (wir sind über selbige zwar betrübt, freuen uns aber, wenn Sie uns diese mitteilen, per Email an verlag@studeo.de).

Wir wünschen viel Erfolg beim Arbeiten mit diesem Buch und vor allem eine erfolgreiche Klausur!

Berlin im September 2011

Silvio Gerlach
Boris Führer

Inhaltsverzeichnis

Abbildungsverzeichnis

Einleitung – Wie Sie mit diesem Klausurtrainer arbeiten sollten

Um den größten Nutzen für Ihre Klausurvorbereitung aus diesem Klausurtrainer zu ziehen, sollten Sie die folgenden Hinweise und Tipps beachten.

Bevor Sie überhaupt anfangen, für die Klausur zu lernen, müssen Sie wissen, was relevant ist. Besorgen Sie sich deshalb zu Beginn des Semesters die folgenden Materialien von Ihrem Lehrstuhl oder der Fachschaft:

- Vorlesungsgliederung,
- Literaturempfehlungen,
- Vorlesungsskript oder –mitschrift (falls vorhanden),
- Aufgabensammlung zur Vorlesung und Übung,
- Alte Klausuren des Lehrstuhls oder wenigstens Probeklausuren.

Die alten Klausuren sind sehr wichtig. Analysieren Sie diese sorgfältig. Bezeichnen Sie die Aufgaben anhand der Vorlesungsgliederung nach Themenbereichen oder Hauptkonzepten und erstellen Sie dann eine **Klausurinhaltsmatrix** wie in diesem Beispiel:

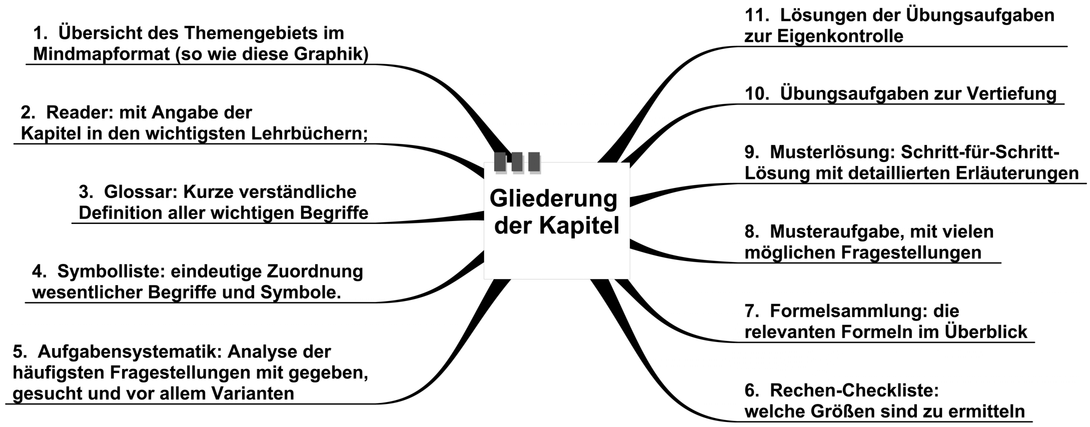
Thema	WS 06/07	SS 20007	WS 07/08	SS 20008	WS 08/09	SS 2009
Verteilungsfunktion			x		x	
Unabhängigkeit zweier Merkmale	x					x
Index		x		x		
Konzentration			x		x	
Zeitreihen	x	x		x		x
Regression	x	x		x		x
Saisonanalyse		x	x	x	x	
Gini-Koeffizient	x		x		x	x
Rangkorrelation						

Erstellen Sie dann eine Tabelle nach diesem Muster (Beispiel):

Aufgabeninhalte	Schwer (Ja / Nein)	Lösung vorhanden?	Selbst gelöst	Ü 1	Ü 2	Ü 3	OK
WS 2008/09							
1.Quantile, Median, Arith. Mittel							
2.Regressionsrechnung							
3. Indexrechnung							
4. Concentration Ratio							
SS 2009							
1. Box-Plot							
2. Zeitreihe Saisonanalyse							
3. Gini-Koeffizient							
4. Rangkorellation							
etc.							

Jetzt sehen Sie klarer, auf welche Themen es besonders ankommt und können effektiver mit dem Klausurtrainer arbeiten.
Wenn Sie die alten Klausuren nicht haben, weil der Lehrstuhl keine herausgibt, dann müssen Sie von den Dozenten und vor allem aus der Übung erfahren, welche Themen und vor allem wie sie für die Klausur relevant sind. Mit etwas Hartnäckigkeit erhalten Sie Antworten auf die Fragen nach den Aufgabentypen und –inhalten. (Mehr dazu, wie man diese Informationen bekommt und analysiert im Handbuch Klausur – für professionelle Klausurvorbereitung unter www.studeo.de)

Jedes Kapitel unseres Klausurtrainers ist nach diesem Schema gegliedert:



Gehen Sie beim Arbeiten idealerweise so vor:

Lesen Sie die Literaturquellen zu den relevanten Konzepten, die Sie aus der Vorlesungsgliederung und den alten Klausuren als relevant erkennen können. Fassen Sie die Inhalte zusammen (Karteikarten).
Vergleichen Sie die Klausuraufgabenstellungen mit der Aufgabensystematik und der Rechencheckliste im Klausurtrainer hinsichtlich der Relevanz. Kreuzen Sie die entsprechenden Stellen an und machen Sie sich weitere Notizen.
Überprüfen Sie die verwendeten Symbole. Kennzeichnen Sie die Symbole in der Liste, die so wie an Ihrem Lehrstuhl verwendet werden und schreiben Sie diejenigen dazu, die anders bezeichnet werden.
Machen Sie sich die relevanten Begriffe anhand des Glossars klar.
Vergleichen Sie die Aufgabensammlung im Klausurtrainer mit den Fragestellungen Ihrer Übungsaufgaben und alten Klausuren und kennzeichnen Sie die besonders wichtigen. Lassen Sie sich von den nicht relevanten Fragen nicht beeindrucken. (Wir haben versucht, eine möglichst große Bandbreite abzudecken.) Wahrscheinlich finden Sie manche Aufgabenstellungen auch (noch) nicht in unserer Sammlung. Dann schreiben Sie uns eine Email (verlag@studeo.de) und wir nehmen sie vielleicht in die nächste Auflage mit auf.
Arbeiten Sie jetzt die Musterlösungen aller für Sie relevanten Fragestellungen gründlich durch und versuchen Sie, die Rechenabläufe eigenständig nachzuvollziehen.
Wenn Sie sich sicher fühlen, sollten Sie sich an den Übungsaufgaben versuchen.

Statistik ist kein leichtes Fach. Es gibt viele Aufgabenvarianten und es erfordert viel Ausdauer. Manche Aufgabenstellungen und vor allem Lösungswege schrecken Sie vielleicht aufgrund der Länge und Komplexität etwas ab. Versuchen Sie das positiv zu sehen: Solche Rechenaufgaben können Sie trainieren. Bei anderen, sachlichen und vor allem offenen Fragen ist das nur begrenzt möglich!

Wenn Sie Fragen oder Anregungen zu diesem Klausurtrainer haben oder Fehler entdeckt haben (über gefundene Fehler informieren wir im Internet!), schreiben Sie uns bitte eine Email an: verlag@studeo.de.

Viel Spaß und Erfolg bei der Arbeit mit diesem Buch.

1.2 Reader zur eindimensionalen Häufigkeitsverteilung

Mit Hilfe dieses Readers können Sie die wichtigsten Konzepte aus der relevanten Literatur zum Thema erarbeiten. Diese Liste ist natürlich nicht erschöpfend, sondern der Übersichtlichkeit wegen auf die Standard-Lehrbücher und ausgewählte Spezialquellen beschränkt. Wir haben nicht immer die aktuellsten Auflagen verwendet. Der Grund dafür ist, dass die Bibliotheken meist nicht die neuesten Auflagen in der Lehrbuchsammlung haben, wohl aber frühere Auflagen. So können Sie zuerst in diesen Büchern nachschlagen, bevor Sie entscheiden welches Sie kaufen wollen. Eine Übersicht, welche Auflagen verwendet wurden, finden Sie im Literaturverzeichnis
<u>Arbeiten Sie mit diesem Reader, indem Sie</u> weitere Quellen einfügen, die Ihr Lehrstuhl empfiehlt. Vor allem aber lesen Sie die Quellen

Konzept	Quelle	Seiten	Gelesen, ja / nein	Kommentar
Bestimmung des Merkmals	Schwarze	32-43		
	Fahrmeir	13-19		
	Bamberg/Baur	6-8		
	Bleymüller	3-4		
Grafische Darstellung	Schwarze	43-63		
	Fahrmeir	29-50, 62-67		
	Bamberg/Baur	11-16		
	Bleymüller	7-12, 23-24		
Lageparameter	Schwarze	63-83		
	Fahrmeir	51-67		
	Bamberg/Baur	16-20		
	Bleymüller	13-17		
Streuungsparameter	Schwarze	83-97		
	Fahrmeir	67-72		
	Bamberg/Baur	20-24		
	Bleymüller	19-24		
Formparameter	Schwarze	97-102		
	Fahrmeir	72-74		

1.3 Glossar zur eindimensionalen Häufigkeitsverteilung

Das folgende Glossar enthält die wichtigsten Begriffe zum Teilgebiet. Die Definitionen sind bewusst einfach gehalten, um das Lernen und Erinnern zu erleichtern. Weitergehende Darstellungen und Erläuterungen finden Sie in den Literaturquellen im Reader.
<u>Arbeiten Sie mit diesem Glossar, indem Sie</u> die rechten Spalten der Tabelle ausfüllen und prüfen, ob für Ihre Klausur noch weitere Begriffe relevant sind. Versuchen Sie diese zu definieren und ergänzen Sie die Tabelle. Bei einigen Begriffen sind noch Zusatzdefinitionen bzw. Fallunterscheidungen angegeben. Prüfen Sie genau, welche Variante an Ihrem Lehrstuhl bevorzugt wird.

Begriff	Definition	Symbol?	Relevant	Kann ich	noch lernen
Absolute Häufigkeit	Anzahl der beobachteten Werte einer bestimmten Merkmalsausprägung.				
Arithmetischer Mittelwert = Durchschnitt, arithmetisches Mittel	Wert, der sich aus der gleichmäßigen Aufteilung der Gesamtsumme aller beobachteten Merkmalsausprägungen auf alle einbezogenen statistischen Einheiten ergibt. Lagemaß, welches mindestens eine Kardinalskala voraussetzt.				
Balkendiagramm	Häufigkeiten werden als vertikale Balken mit Zwischenräumen grafisch dargestellt. Die Merkmalsausprägungen werden auf der Abszisse abgetragen.				
Binäres Merkmal	Ein Merkmal, das nur zwei sich gegenseitig ausschließende (disjunkte) Ausprägungen aufweisen kann. Beispiel: Geschlecht				
Bowley-Kenngröße = Quartilskoeffizient der Schiefe	Ein Schiefemaß, welches sich rein auf Quantile bezieht. Sein Vorteil ist, dass es robuster ist als bspw. die Schiefe nach dem 3.ten zentralen Moment. Formparameter, wobei mindestens eine Kardinalskala vorliegen muss.				
Box-Plot = Box-and-Whisker Plot	Graphische Darstellung spezieller, ermittelter Parameter der Häufigkeitsverteilung mindestens ordinalskalierter Merkmale. Aufgrund der Anschaulichkeit kann man Verteilungen gut vergleichen.				
Deskriptive Statistik	Gesamtheit der Verfahren zur Erhebung, Aufbereitung und Auswertung von empirischen (=beobachtbaren) Daten.				

Begriff	**Definition**	**Symbol?**	**Relevant**	**Kann ich**	**noch lernen**
Dezil	Quantile, die eine geordnete Reihe in 10 gleiche Teilgruppen zerlegt. (→ Quantil) Lagemaß, welches mindestens eine Ordinalskala voraussetzt.				
Diskretes Merkmal	Kardinalskaliertes Merkmal, das nur abzählbar viele Werte annehmen kann. Beispiel: Studentenzahl				
Einfallsklasse	Die Einfallsklasse eines p-Quantils bezeichnet die Klasse, in der der Wert der Verteilungsfunktion gerade den Wert p überschreitet. wird bei der Berechnung von Quantilen bei klassierten Merkmalen verwendet wird.				
Entropie	Maßzahl für die Homogenität bzw. Heterogenität der Merkmalswerte. Streuungsmaß, welches mindestens eine Nominalskala voraussetzt.				
Externe Varianz	Die Streuung zwischen den Klassen bei einer klassierten Merkmal = aus den Klassenmittelwerten berechneten Varianz gewichtet mit den Klassenhäufigkeiten Streuungsmaß, welches mindestens eine Kardinalskala voraussetzt.				
Flächendiagramm	Häufigkeiten werden als Flächen grafisch dargestellt. Die Merkmalsausprägungen werden auf der Abszisse abgetragen.				
Formparameter	Maßzahl zur Darstellung der Form bzw. des Aussehens der Merkmalswerte, bspw. Schiefe, Symmetrie oder Wölbung der Verteilung.				
Geometrischer Mittelwert = geometrisches Mittel	Durchschnittswert bei multiplikativ verknüpften Merkmalswerten. Der Logarithmus des geometrischen Mittels entspricht dem arithmetischen Mittel der logarithmierten Beobachtungswerte. Lagemaß, welches mindestens eine Kardinalskala voraussetzt.				
Gesamtvarianz	Die Summe aus interner und externer Varianz (Streuungszerlegung) bei einem klassierten Merkmal. Ist die Urliste gegeben, dann entspricht die Varianz der unklassierten Einzelwerte gleich der Gesamtvarianz bei klassiertem Merkmal. Ist sie nicht vorhanden, besteht sie nur aus der externen Varianz, weil die interne Varianz nicht bestimmt werden kann. Streuungsmaß, welches mindestens eine Kardinalskala voraussetzt.				
Grundgesamtheit = statistische Masse	Menge aller statistischen Einheiten, die in die Untersuchung bestimmter Merkmale einbezogen werden. Beispiel: alle Studenten eines Jahrgangs				
Harmonischer Mittelwert = harmonisches Mittel	Spezialfall des arithmetischen Mittels für verhältnisskalierte Merkmale (Entfernungen, Flächen, Gewichte etc.). Für Durchschnitt aus Verhältniszahlen zu verwenden. Notwendig sind Zusatzinformationen (g_i), die sich inhaltlich auf den Zähler des Verhältnisses x_i beziehen. Lagemaß, welches mindestens eine Kardinalskala voraussetzt.				
Häufbares Merkmal	Merkmal, von dem an einer statistischen Einheit mehr als eine Ausprägung festgestellt werden kann. Beispiel: Studienfach eines Studenten, BWL *und* Jura				
Häufigkeitsdichtefunktion	Funktion, die das Verhältnis der absoluten bzw. relativen Häufigkeiten einer Klasse in Beziehung zur Klassenbreite dieser Klasse setzt. (→ Histogramm)				
Häufigkeitsverteilung = Häufigkeitstabelle	Darstellung der geordneten Merkmalsausprägungen und die Angabe der zugehörigen absoluten bzw. relativen Häufigkeiten.				
Histogramm	Grafische Darstellung der Häufigkeiten eines stetigen bzw. quasistetigen klassierten Merkmals durch rechteckige Flächen in Höhe der Häufigkeitsdichten der jeweiligen Klassen. **Abszisse:** Klassengrenzen abtragen **Ordinate:** Häufigkeitsdichte abtragen (bei gleicher Klassenbreite gleich der Häufigkeit der Klasse)				
Identifikationsmerkmal	Merkmal, das darüber entscheidet, ob ein Element zu einer statistischen Gesamtheit gehört und damit in die Untersuchung einbezogen werden soll. Beispiel: Alle Studenten				

1.4 Aufgabensystematik zur eindimensionalen Häufigkeitsverteilung

Die folgende Übersicht stellt die am häufigsten in Klausuren verwendeten Aufgabentypen und -stellungen dieses Themenbereichs dar. Die genaue Aufgabenstellung in Klausuren Ihres Lehrstuhls kann davon abweichen.
Arbeiten sie mit dieser Übersicht, indem Sie die Inhalte der alten Klausuraufgaben Ihres Lehrstuhls anhand dieses Schemas sorgfältig überprüfen und systematisieren. Passen Sie die Übersicht gegebenenfalls an oder ergänzen Sie sie.

Aufgabensystematik Eindimensionale Häufigkeitsverteilung Teil 1

Abb. 1-3: Mindmap Aufgabensystematik eindimensionale Häufigkeitsverteilung Teil 1

1.5 Rechencheckliste zur eindimensionalen Häufigkeitsverteilung

Diese Liste stellt die in Standard-Klausuren zu errechnenden Größen des Themenbereiches dar. **Arbeiten Sie mit dieser Rechencheckliste, indem Sie** sorgfältig prüfen, nach welchen Größen in den alten Klausuren Ihres Lehrstuhls gefragt wurde und passen Sie die Tabelle entsprechend an. Füllen Sie dann die rechten Spalten aus. Ordnen Sie vor allem auch die Aufgaben aus Ihrer Übung / Ihrem Tutorium entsprechend zu. Prüfen Sie immer wieder, welche der Aufgabentypen Sie noch üben müssen. (Ü1, Ü2, Ü3 bezeichnen Ihre "Trainingsdurchgänge".)

Zu errechnende Größe	Ihr Symbol	Musterlös. Nr.	Relevant ja / nein	Schwierig? ja / nein	Aufgaben aus Übung/ Tutor.?	Kann ich	Ü1	Ü2	Ü3
Absolute Häufigkeit		A 1.4,A 2.1, A 3.1							
Arithmetischer Mittelwert		A 1.13,A 5.5							
Durchschnittliche absolute Abweichung vom Median		A 1.16							
Durchschnittliche absolute Abweichung von einem Wert		A 1.17							
Entropie		A 6.6							
Externe Varianz		A 2.15							
Geometrischer Mittelwert		A 5.2							
Gesamtmittelwert (klassiert)		A 2.10							
Gesamtvarianz (klassiert)		A 2.12							
Harmonischer Mittelwert		A 5.3,A 5.4							
Interne Varianz		A 2.15							
Kenngröße von Bowley		A 1.25							
Klassenanzahl		A 2.2							
Klassenmittelwert		A 2.9							
Klassenvarianz		A 2.11							
Kumulierte absolute Häufigkeit		A 1.5,A 2.4							
Kumulierte relative Häufigkeit		A 1.5,A 2.4							
Median		A 1.8,A 2.6, A 5.1,A 6.5							
Median der absoluten Abweichung vom Median		A 1.18							
Mittlere quadratische Abweichung vom einem Wert		A 1.19							
Modalklasse		A 2.5							
Modus		A 1.7,A 6.4							
Normierte Entropie		A 6.7							
Normierte Summenhäufigkeitsentropie		A 6.9							
Quantile		A 2.19							
Quartile		A 1.9,A 2.7							
Quartilsabstand		A 1.10							
Quartilsdispersionskoeffizient		A 1.21							
Relative Häufigkeit		A 1.4,A 2.3, A 3.1,A 6.2							
r-tes Moment		A 1.27							

1.6 Symbolliste zur eindimensionalen Häufigkeitsverteilung

Diese Tabelle enthält die gängigen Größen des Themenbereiches und die dafür am häufigsten verwendeten Symbole. **Arbeiten Sie mit dieser Checkliste, indem Sie** sorgfältig prüfen, welche Symbole in Ihren Vorlesungen / Übungen und Unterlagen verwendet werden. Tragen Sie diese Symbole ein oder haken Sie die Symbole ab, die mit Ihren übereinstimmen. Ergänzen Sie die Liste falls nötig.

Studeo-Version	Wie bei uns? j/n	Varianten	Bedeutung	Meine Symbole
n_i		n_j	Absolute Häufigkeit	
$n(x_i)$		$n(x_j)$, $n(a_j)$	Absolute Häufigkeit der Merkmalsausprägung x_i	
$\bar{x}$		μ	Arithmetischer Mittelwert	
$D(x_{0.5})$			Durchschnittliche Abweichung vom Median	
$D(a)$		d, $\bar{s}$, $MQ(M)$	Durchschnittliche Abweichung vom Wert a	
r_m			Durchschnittliche Wachstumsrate	
H			Entropie	
$s^2_{ext.}$			Externe Varianz	
$\bar{x}_g$		$\bar{x}_G$, $\bar{x}_{geom}$, G	Geometrischer Mittelwert	
n		N	Gesamtanzahl der Merkmalswerte	
$\hat{\bar{x}}$			Gesamtmittelwert einer klassierten Verteilung bei Unkenntnis der Originaldaten ($\Rightarrow$ Schätzung)	
g_i		a_j	Gewichte bei der Berechnung des harmonischen Mittelwertes	
$\bar{x}_h$		$\bar{x}_H$, $\bar{x}_{har}$	Harmonischer Mittelwert	
l_i		h_i^*	Höhe des Balkens im Histogramm	
$s^2_{int.}$			Interne Varianz	
x_{ij}			j-ter Merkmalswert der i-ten Klasse	
α		g_p	Kenngröße von Bowley	
b_i		d_j, Δx_i	Klassenbreite	
$\bar{x}_i$		$\bar{x}_j$, x_i'	Klassenmittelwert (der Klasse i) bei Kenntnis der Originaldaten	
$\hat{\bar{x}}_i$			Klassenmittelwert (der Klasse i) bei Unkenntnis der Originaldaten	
s_i^2			Klassenvarianz bei Kenntnis der Originaldaten	
S_i			Kreissektor	
$N(x)$		$H(x_j)$	Kumulierte absolute Häufigkeit	
$H(x)$		$F(x_j)$	Kumulierte relative Häufigkeit	
$x_{(k)}$			k-te Wert einer geordneten statistischen Reihe	
x_{max}		$x_{(n)}$	Maximaler Merkmalswert	
$x_{0.5}$		$\bar{x}_Z$, x_{med}, Me	Median	
$MAD(x_{0.5})$			Median der absoluten Abweichung vom Median	
$X, Y, \ldots$			Merkmal	
$x_i, y_j, \ldots$		x_j, a_j	Merkmalsausprägungen	
x_{min}		$x_{(1)}$	Minimaler Merkmalswert	
$MQA(a)$		$S(M)$	Mittlere quadratische Abweichung vom Wert a	
x_{mod}		$h(\bar{x}_D)$	Modus	
H^*			Normierte Entropie	
VD^*			Normierte Summenhäufigkeitsentropie	
x_o		x_j^*, c_{j-1}, x_i^o	Obere Grenze eines Intervalls	
$x_{i,o}$			Obere Grenze eines Intervalls der Klasse i	
x_p		$\bar{x}_{p/k}$	Quantil	

1.7 Formelsammlung zur eindimensionalen Häufigkeitsverteilung

Diese Formelsammlung enthält die wichtigsten Formeln zum Teilgebiet, geordnet nach inhaltlichen Gesichtspunkten, nicht alphabetisch! **Arbeiten Sie mit dieser Formelsammlung, indem Sie** sie mit den Formeln bzw. Aufgabenstellungen Ihrer Übung/ Ihres Tutoriums vergleichen**.** Stellen Sie fest, ob jede Formel für Sie relevant ist und tragen Sie Ihre Schreibweise ein. Ergänzen Sie die Sammlung gegebenenfalls um weitere Formeln. (Rel in der ersten Zeile heißt relevant!)

Hinweis: Da die Berechnung von Parametern/Maßen für klassierte/unklassierte Merkmale in den meisten Fällen lediglich eines zusätzlichen Index bedarf (meist ein Dach), werden hier nur die Formeln für unklassierte Merkmale angegeben. Auf klassierte Merkmale ist diese Schreibweise leicht übertragbar. Einzige Ausnahme ist die Bildung von Mittelwert und Varianz bei klassierten Merkmalen.

Bezeichnung	Formel	Nr.	Rel	Ihre Schreibweise
Absolute Häufigkeit	$n_i = n \cdot h_i$	(1.1)		
Arithmetische Mittelwert	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot x_i = \sum_{i=1}^{k} h_i \cdot x_i$	(1.2)		
Durchschnittliche absolute Abweichung vom Median	$D(x_{0.5}) = \frac{1}{n} \cdot \sum_{i=1}^{n} \lvert x_i - x_{0.5} \rvert = \frac{1}{n} \cdot \sum_{i=1}^{k} \lvert x_i - x_{0.5} \rvert \cdot n_i = \sum_{i=1}^{k} \lvert x_i - x_{0.5} \rvert \cdot h_i$	(1.3)		
Durchschnittliche absolute Abweichung von einem Wert a	$D(a) = \frac{1}{n} \cdot \sum_{i=1}^{n} \lvert x_i - a \rvert = \frac{1}{n} \cdot \sum_{i=1}^{k} \lvert x_i - a \rvert \cdot n_i = \sum_{i=1}^{k} \lvert x_i - a \rvert \cdot h_i$	(1.4)		
Durchschnittliche Wachstumsrate	$r_m = \sqrt[k]{\frac{\text{Endwert}}{\text{Anfangswert}}} - 1$	(1.5)		
Entropie	$H = -\frac{1}{\ln 2} \cdot \sum_{i=1}^{k} h_i \cdot \ln h_i$	(1.6)		
Externe Varianz	$s^2_{ext.} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot (\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{k} h_i \cdot (\bar{x}_i - \bar{x})^2$	(1.7)		
Geometrischer Mittelwert	$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$	(1.8)		
Gesamtmittelwert (klassiert; Unkenntnis der Originaldaten)	$\hat{\bar{x}} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot \hat{\bar{x}}_i = \sum_{i=1}^{k} h_i \cdot \hat{\bar{x}}_i$	(1.9)		
Gesamtvarianz (klassiert; Kenntnis der Originaldaten)	$s^2 = s^2_{int.} + s^2_{ext.}$	(1.10)		
Gesamtvarianz (klassiert; Unkenntnis der Originaldaten)	$\hat{s}^2 = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot (\hat{\bar{x}}_i - \hat{\bar{x}})^2 = \sum_{i=1}^{k} h_i \cdot (\hat{\bar{x}}_i - \hat{\bar{x}})^2$	(1.11)		
Harmonischer Mittelwert	$\bar{x}_h = \frac{1}{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{x_i}} = \frac{1}{\frac{1}{g} \cdot \sum_{i=1}^{k} \frac{g_i}{x_i}}$	(1.12)		
Höhe eines Balkens im Histogramm	$l_i = \frac{h_i}{b_i}$	(1.13)		
Interne Varianz	$s^2_{int.} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot s_i^2 = \sum_{i=1}^{k} h_i \cdot s_i^2$	(1.14)		
Kenngröße von Bowley	$\alpha = \frac{(x_{0.75} - x_{0.5}) - (x_{0.5} - x_{0.25})}{x_{0.75} - x_{0.25}}$	(1.15)		
Klassenbreite	$b_i = x_{i,o} - x_{i,u}$	(1.16)		

1.8 Musteraufgaben zur eindimensionalen Häufigkeitsverteilung

Diese Aufgaben sind beispielhaft für den Themenbereich. **Arbeiten Sie mit diesen Musteraufgaben, indem Sie** die einzelnen Fragen mit den Aufgabenstellungen Ihrer Übung / Ihres Tutoriums, vor allem aber mit denen der alten Klausuren Ihres Lehrstuhls vergleichen. Kreuzen Sie in den rechten Spalten die Fragestellungen an, die für Sie relevant sind und ergänzen Sie die Liste gegebenenfalls um weitere relevante Fragestellungen in diesem Themenbereich. Schicken Sie uns diese Fragestellungen per Email an verlag@studeo.de.

WICHTIG: Weitere Aufgaben finden Sie im Internet unter www.studeo.de.

1.8.1 Musteraufgabe 1 – Unklassiertes kardinales Merkmal

Die folgende Punkteverteilung in der Statistik-Klausur wurde bei einer Gruppe Studenten festgestellt:

31 66 12 39 19 91 23 18 27 39
77 42 77 91 31 77 54 77 12 54

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 1.1.	Ordnen Sie die Begriffe Merkmal, Merkmalsträger, Merkmalsausprägung und Häufigkeit zu.			
A 1.2.	Was für ein Merkmal wurde untersucht?			
A 1.3.	Erstellen Sie eine geordnete statistische Reihe.			
A 1.4.	Stellen Sie die absolute und relative Häufigkeitsverteilung tabellarisch und grafisch (in einem Stabdiagramm) dar.			
A 1.5.	Ermitteln Sie die kumulierten absoluten und relativen Häufigkeiten.			
A 1.6.	Zeichnen Sie die empirische Verteilungsfunktion.			
A 1.7.	Ermitteln und interpretieren Sie den Modus.			
A 1.8.	Ermitteln Sie den Median grafisch und rechnerisch und interpretieren Sie den Median.			
A 1.9.	Ermitteln und interpretieren Sie das 1. und 3. Quartil.			
A 1.10.	Ermitteln Sie den Quartilsabstand.			
A 1.11.	Ermitteln Sie H(31) und H(78) und interpretieren Sie diese Werte.			
A 1.12.	Errechnen und interpretieren Sie die Spannweite.			
A 1.13.	Errechnen und interpretieren Sie den arithmetischen Mittelwert.			
A 1.14.	Errechnen und interpretieren Sie die empirische Varianz.			
A 1.15.	Errechnen und interpretieren Sie die Standardabweichung.			
A 1.16.	Errechnen Sie die durchschnittliche absolute Abweichung vom Median.			
A 1.17.	Errechnen Sie die durchschnittliche absolute Abweichung vom Wert 50.			
A 1.18.	Errechnen Sie den Median der absoluten Abweichung vom Median.			
A 1.19.	Errechnen Sie die mittlere quadratische Abweichung vom Wert 40.			
A 1.20.	Errechnen Sie den Variationskoeffizienten und interpretieren Sie ihn.			
A 1.21.	Errechnen Sie den Quartilsdispersionskoeffizienten.			
A 1.22.	Bestimmen und interpretieren Sie die Schiefe nach Pearson.			
A 1.23.	Bestimmen und interpretieren Sie die Schiefe nach Yule-Pearson.			
A 1.24.	Bestimmen und interpretieren Sie die Schiefe nach dem 3. zentralen Moment.			
A 1.25.	Bestimmen und interpretieren Sie die Kenngröße von Bowley.			
A 1.26.	Bestimmen und interpretieren Sie die Wölbung.			
A 1.27.	Bestimmen Sie das 5. Moment.			
A 1.28.	Bestimmen Sie das 5. zentrale Moment.			
A 1.29.	Bestimmen Sie das 5. standardisierte Moment.			
A 1.30.	Stellen Sie die Verteilung in einem Box-Plot dar.			
A 1.31.	Bestimmen Sie die Verteilung des Merkmals Y, wenn folgender Zusammenhang besteht: $Y = 0{,}5X - 10$.			
A 1.32.	Bestimmen Sie Mittelwert und Varianz des Merkmals Y.			

1.8.2 Musteraufgabe 2 – Klassiertes kardinales Merkmal

Die folgende Punkteverteilung in der Statistik-Klausur wurde bei einer Gruppe Studenten festgestellt:

31 66 12 39 19 91 23 18 27 39
77 42 77 91 31 77 54 77 12 54

Alle Aufgaben außer A 2.1 und A 2.15 beziehen sich auf die klassierte Verteilung!

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 2.1.	Erstellen Sie eine klassierte absolute Häufigkeitsverteilung mit einer Klassenbreite von 20.			
A 2.2.	Welche Klassenanzahl, wäre nach der Formel von Sturges geeignet gewesen?			
A 2.3.	Erstellen Sie die relative Häufigkeitsverteilung.			
A 2.4.	Ermitteln Sie die kumulierten absoluten und relativen Häufigkeiten.			
A 2.5.	Ermitteln Sie die Modalklasse.			

A 2.6.	Ermitteln Sie den Median durch Interpolation und interpretieren Sie ihn.			
A 2.7.	Ermitteln Sie das 1. und 3. Quartil durch Interpolation.			
A 2.8.	Bestimmen Sie rechnerisch den Wert H(70) der empirischen Verteilungsfunktion.			
A 2.9.	Errechnen Sie den Mittelwert für jede einzelne Klasse.			
A 2.10.	Errechnen Sie den Mittelwert der Verteilung.			
A 2.11.	Errechnen Sie, wenn möglich, die Varianz für jede einzelne Klasse.			
A 2.12.	Errechnen Sie die Varianz der Verteilung.			
A 2.13.	Errechnen Sie die Standardabweichung der Verteilung.			
A 2.14.	Unter welchen Bedingungen kann eine Verbesserung der Schätzung der Varianz erreicht werden? Welchen Wert nimmt die Varianz dann an?			
A 2.15.	Bestimmen und interpretieren Sie die interne und externe Varianz.			
A 2.16.	Stellen Sie die klassierte Verteilung in einem Balkendiagramm dar.			
A 2.17.	Stellen Sie die klassierte Verteilung in einem Kreisdiagramm dar.			
A 2.18.	Stellen Sie die klassierte Verteilung in einem Histogramm dar.			
A 2.19.	Stellen Sie die empirische Verteilungsfunktion graphisch dar und bestimmen Sie anhand Ihrer Zeichnung das 80%-Quantil.			

1.8.3 Musteraufgabe 3 – Textaufgabe klassierte Häufigkeitsverteilung

Ein Bekannter von Ihnen, von Beruf Fahrlehrer in Berlin, lästert mal wieder über Fahrradfahrer. Er teilt Ihnen mit, dass er Fahrradfahrer in folgende Kategorien unterteilt:

Gruppe 1 setze sich seiner Meinung nach über alle geltenden Verkehrsregeln hinweg, er nennt sie die „Rücksichtslosen“. Die zweite Gruppe („ Schwarzfahrer“), die er ausmacht, fahre auch in der tiefsten Nacht ohne Licht. Schließlich gebe es noch diejenigen, die verkehrskonform fahren würden („Normalos“).
Im letzten Jahr sind ihm in Ausübung seines Berufes 37 „Normalos“ aufgefallen. Weiter waren 63% entweder „Schwarzfahrer“ oder „Rücksichtslose“. 27% der Radfahrer waren sogenannte „Schwarzfahrer“.

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 3.1.	Erstellen Sie die absolute und relative Häufigkeitsverteilung.			
A 3.2.	Erklären Sie, weshalb man keine kumulierten absoluten und relativen Häufigkeiten bestimmen kann.			
A 3.3.	Stellen Sie die Häufigkeitsverteilung in einem Balkendiagramm dar.			

1.8.4 Musteraufgabe 4 – Stem & Leaf Display

An der Scanner-Kasse eines Supermarktes wurde für 50 aufeinanderfolgende Kunden folgende Bedienungszeiten in Sekunden registriert:

40	20	22	15	18	51	37	42	31	58
33	39	49	22	23	62	42	53	43	44
19	49	39	36	37	38	22	24	32	29
41	40	39	38	27	51	52	54	28	22
64	19	50	40	18	68	51	41	48	57

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 4.1.	Erstellen Sie ein Stem and Leaf Display.			

1.8.5 Musteraufgabe 5 – Mittelwerte

Errechnen Sie jeweils die Mittelwerte und begründen Sie Ihre jeweilige Entscheidung:

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 5.1.	Bei Olympischen Winterspielen werden Skispringer von Punktrichtern in Noten bewertet. Für jeden Skispringer soll die durchschnittliche Einordnung bestimmt werden.			
A 5.2.	Die Preissteigerung in Deutschland betrug in den letzten Jahren 1,2%, 2,1%, 2,0%. Bestimmen Sie die durchschnittliche Steigerungsrate.			
A 5.3.	Ein Tischler schneidet 14 m^2 Holz mit 2 m^2 je Stunde und 8 m^2 Holz mit 4 m^2 je Stunde. Wie hoch war die Durchschnittsleistung?			
A 5.4.	Das Rennauto mit der Startnummer 17 legt bei dem 24h-Rennen in Le Mans (unter anderem) 4 Runden in unterschiedlichen Geschwindigkeiten (230 km/h, 210 km/h, 270 km/h und 110 km/h [infolge eines Ausfluges in die Prärie]) zurück. Welche Durchschnittsgeschwindigkeit hatte das Auto?			
A 5.5.	Ein vielreisender Urlauber benötigt für das erste Teilstück (ein Drittel) seiner Schiffsreise von Frankfurt nach New York 6 Stunden. Für das zweite Drittel 12 Stunden und für das letzte Drittel 4 Stunden. Welche Durchschnittsgeschwindigkeit ergibt sich somit für die 9.000 km lange Strecke?			

1.9 Musterlösungen zur eindimensionalen Häufigkeitsverteilung

Diese Musterlösungen sind beispielhaft. Wir haben uns bemüht, insbesondere die Rechenschritte ausführlicher darzustellen als in der Klausur eigentlich nötig. Erläuterungen stehen in der rechten Spalte. **Arbeiten Sie mit diesen Lösungen, indem Sie** den Weg eigenständig nachvollziehen und sich Bemerkungen am Rande machen. Sie haben bereits die Aufgabenstellungen mit den Aufgaben Ihrer Übung und der alten Klausuren verglichen. Jetzt müssen Sie dasselbe für die Lösungen machen. Vergleichen Sie die Lösungen Schritt für Schritt und machen Sie sich Notizen. Haken Sie die Lösungen ab, die Sie beherrschen. Lösen Sie die Aufgaben immer wieder, bis Sie sie ohne Nachzuschauen beherrschen. Üben Sie Termumformungen mit dem Studeo®-Rechentrainer (www.rechentrainer.de).

1.9.1 Musterlösung 1 – Unklassiertes kardinales Merkmal

Die folgende Punkteverteilung in der Statistik-Klausur wurde bei einer Gruppe Studenten festgestellt:

31 66 12 39 19 91 23 18 27 39
77 42 77 91 31 77 54 77 12 54

Lösung | **Erläuterungen / Notizen**

A 1.1. Ordnen Sie die Begriffe Merkmal, Merkmalsträger, Merkmalsausprägung und Häufigkeit zu.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Das Merkmal bezeichnet die zu untersuchende Eigenschaft. Hier ist das Merkmal X: Anzahl der Punkte eines Studenten in der Statistik-Klausur.

Lesen Sie dazu die Aufgabenstellung sorgfältig.

Der Merkmalsträger bezeichnet das einzelne Objekt/Einheit, anhand dessen die Eigenschaft feststellbar ist. Hier Studenten.
Unter der Merkmalsausprägung sind alle Werte zu verstehen, die das Merkmal am Objekt konkret annehmen kann, bspw. 31 Punkte für den ersten Studenten. Da eine obere Grenze nicht gegeben ist, kann man dies hier nicht genau angeben (von 0 bis ?).

Sinnvoll wäre wohl ein Wert von bspw. 100 Punkten.

Der Begriff Häufigkeiten bestimmt die Anzahl der Einheiten, bei denen das Merkmal X eine bestimmte Ausprägung x_i annimmt. So erreichten bspw. 2 Studenten jeweils 39 Punkte.

A 1.2. Was für ein Merkmal wurde untersucht?

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Es ist zu prüfen, welches Skalenniveau für das Merkmal X vorliegt.
In diesem Falle ist es eine Kardinalskala, genauer: eine Verhältnisskala, da sich die Verhältnisse der Merkmalsausprägungen sinnvoll interpretieren lassen. Grund dafür ist, dass es einen natürlichen Ursprung, nämlich Null (Punkte), gibt.

Man kann sagen:
Student A hat doppelt so viele Punkte in der Klausur erreicht wie Student B.

Alternative Möglichkeiten in Klausuren:

- Nominalskala: die Merkmalsausprägungen stehen gleichberechtigt nebeneinander und lassen sich nur durch unterschiedliche Namen differenzieren, nicht aber ordnen;
Beispiel: verschiedene Farben,
- Ordinalskala: Ausprägungen lassen sich in eine sinnvolle Reihenfolge bringen, z. B. ordnen nach der Größe; es ist aber keine Aussage über den absoluten Wert möglich Beispiel: Noten
- Intervallskala: Hier lassen sich Differenzen von Ausprägungen interpretieren, wobei es keinen absoluten Nullpunkt und keine natürliche Einheit gibt;
Beispiel: Längen- und Breitengrade.

Siehe Mindmap auf Seite 15.

Die Intervallskala wird mit der Verhältnisskala oft zur Kardinalskala zusammengefasst.

A 1.3. Erstellen Sie eine geordnete statistische Reihe.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Die Merkmalsträger sind nach ihrer Merkmalsausprägung aufsteigend zu sortieren, also:
12, 12, 18, 19, 23, 27, 31, 31, 39, 39, 42, 54, 54, 66, 77, 77, 77, 77, 91, 91.

A 1.4. Stellen Sie die absolute und relative Häufigkeitsverteilung tabellarisch und grafisch (in einem Stabdiagramm) dar.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Zuerst wird die absolute Häufigkeitsverteilung behandelt. Dazu müssen die Werte aus der Urliste abgezählt werden und mit der so gewonnenen Strichliste in Tabellenform gebracht werden:

Wenn man die geordnete statistische Reihe schon bestimmt hat, sollte man sie statt der Urliste benutzen.

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	12	18	19	23	27	31	39	42	54	66	77	91
n_i	2	1	1	1	1	2	2	1	2	1	4	2
h_i	0,1	0,05	0,05	0,05	0,05	0,1	0,1	0,05	0,1	0,05	0,2	0,1

Das i im Index bezeichnet den i-ten Wert in der Tabelle, nicht etwa den i-ten Wert der Urliste.

Mittels der Gleichung $h_i = \frac{n_i}{n}$ werden die relativen Häufigkeiten errechnet:

Aufgabe 2

Vorarbeit: Sortieren der Reihe

2 2 3 3 3 3 4 4 4 4 4 4 4 5 5 5 5 5 5 6 6 6 6 7 7 7 8 8 9 10

1. – 3.

Klasse	h_i	H_i	f_i	F(x)
[2,4]	13	13	0,433	0,433
[5,7]	13	26	0,433	0,867
[8,10]	4	30	0,133	1

4.	Klassen 1 und 2, also keine
5.	5,31
6.	3,15 und 6,46
7.	3,31
8.	8
9.	5,10
10.	4,29
11.	2,07
12.	0,41

Aufgabe 3

Vorarbeit: Sortieren der Reihe

4,5,5,6,6,11,12,14,16,18,21,23,29,29,31,36,36,41,43,43,43,43,43,44,44,46,47,51,52,57,59,63,69,77,80,81,112

1. – 3.

Klasse	h_i	H_i	f_i	F(x)
[0,30]	14	14	0,38	0,38
[30,60]	17	31	0,46	0,84
[60,90]	5	36	0,14	0,97
[90,120]	1	37	0,03	1,00

4.	Klasse 2
5.	37,94
6.	19,82 und 54,26
7.	34,44
8.	
9.	108
10.	39,32
11.	527,25
12.	22,96
13.	0,58

Aufgabe 4

Vorarbeit: Sortieren der Reihe

119 121 122 125 126 127 128 131 132 133 134 136 137 138 141 141 143 143 144 147

1. – 2.

Klasse	h_i	H_i	f_i	F(x)
[115,120]	1	**1**	**0,05**	**0,05**
[120,125]	2	**3**	**0,10**	**0,15**
[125,130]	4	**7**	**0,20**	**0,35**
[130,135]	4	**11**	**0,20**	**0,55**
[135,140]	3	**14**	**0,15**	**0,70**
[140,145]	5	**19**	**0,25**	**0,95**
[145,150]	1	**20**	**0,05**	**1,00**

3.	Klasse 6
4.	133,75
5.	127,50; 141,00
6.	11,67
7.	0,45 und 0,95
8.	28
9.	133,75
10.	64,69
11.	8,04
12.	0,06

2.8 Musteraufgaben zur zweidimensionalen Häufigkeitsverteilung

Diese Aufgaben sind beispielhaft für den Themenbereich. **Arbeiten Sie mit diesen Musteraufgaben, indem Sie** die einzelnen Fragen mit den Aufgabenstellungen Ihrer Übung / Ihres Tutoriums, vor allem aber mit denen der alten Klausuren Ihres Lehrstuhls vergleichen. Kreuzen Sie in den rechten Spalten die Fragestellungen an, die für Sie relevant sind und ergänzen Sie die Liste gegebenenfalls um weitere relevante Fragestellungen in diesem Themenbereich. Schicken Sie uns diese Fragestellungen per Email an verlag@studeo.de.

WICHTIG: Weitere Aufgaben finden Sie im Internet unter www.studeo.de.

2.8.1 Musteraufgabe 1 – Kardinale Merkmale (Korrelationsrechnung)

100 Personen wurden nach dem monatlichen Einkommen [in €] und dem Alter [in Jahren] befragt. Im Ergebnis ergab sich nachstehende Korrelationstabelle:

Einkommen X	Alter Y				
	[20;30)	[30;40)	[40;50)	[50;60)	[60;70)
[0;500)	4	2	2	4	2
[500;1.000)	2	8	6	4	2
[1.000;1.500)	6	14	10	8	6
[1.500;2.000)	2	6	6	4	2

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 1.1.	Welches Skalenniveau besitzen die beiden Merkmale X und Y. Begründen Sie Ihre Entscheidung.			
A 1.2.	Bestimmen Sie das Tableau der gemeinsamen relativen Häufigkeiten.			
A 1.3.	Bestimmen Sie die Randverteilungen der beiden Merkmale X und Y.			
A 1.4.	Bestimmen Sie die gemeinsamen kumulierten relativen Häufigkeiten.			
A 1.5.	Erstellen Sie für die Altersgruppe [50;60) die bedingte Verteilung für das Einkommen X.			
A 1.6.	Erstellen Sie für die Einkommensgruppe [1.500;2.000) die bedingte Verteilung für das Alter Y.			
A 1.7.	Wie groß ist der Anteil der 30-40jährigen, die ein Einkommen von 0-500 € haben?			
A 1.8.	Wie groß ist der Anteil der Personen, die ein Einkommen von 1.500-2.000 € besitzen?			
A 1.9.	Berechnen Sie das durchschnittliche Einkommen der 100 Personen.			
A 1.10.	Berechnen Sie das durchschnittliche Einkommen einer Person, die zwischen 50 und 60 Jahre alt ist.			
A 1.11.	Berechnen Sie das durchschnittliche Alter der 100 Personen.			
A 1.12.	Bestimmen Sie das Einkommen, welches die 100 Personen am häufigsten angaben.			
A 1.13.	Welches Einkommen hatten die 50% ärmsten Personen höchstens?			
A 1.14.	Berechnen Sie die Standardabweichung für das Merkmal X und das Merkmal Y.			
A 1.15.	Berechnen Sie die bedingte Varianz für das Einkommen einer Person, die zwischen 50 und 60 Jahre alt ist.			
A 1.16.	Berechnen Sie die empirische Kovarianz zwischen Einkommen und Alter.			
A 1.17.	Sind die Merkmale Einkommen und Alter unabhängig?			
A 1.18.	Berechnen und interpretieren Sie den Korrelationskoeffizienten nach Bravais-Pearson zwischen Alter und Einkommen.			
A 1.19.	Sind für kardinale Merkmale auch Korrelationsmaße berechenbar, die eigentlich nur für ordinale oder gar nominale Merkmale berechnet werden?			

2.8.2 Musteraufgabe 2 – Kardinale Merkmale (Regressionsrechnung)

Eine Unternehmen stellt ein Produkt her. Für die Kalkulation soll seine Kostenfunktion ermittelt werden. Dafür werden in 10 Perioden die Produktionsmengen X und die Gesamtkosten Y registriert:

Periode t	1	2	3	4	5	6	7	8	9	10
Output [in t]	9	12	14	12	12	13	10	11	12	15
Kosten [in 1.000 €]	1.216	1.300	1.356	1.288	1.276	1.292	1.260	1.244	1.288	1.360

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 2.1.	Welches Skalenniveau besitzen die beiden Merkmale? Begründen Sie Ihre Entscheidung.			
A 2.2.	Welche Variable ist die unabhängige, welche die abhängige Variable? Begründen Sie Ihre Entscheidung.			
A 2.3.	Zeichnen Sie die Merkmale X und Y in ein Streudiagramm ein.			
A 2.4.	Bestimmen Sie die Mittelwerte beider Merkmale.			
A 2.5.	Bestimmen Sie die Standardabweichungen beider Merkmale.			
A 2.6.	Bestimmen Sie die Kovarianz.			
A 2.7.	Bestimmen und interpretieren Sie die Regressionsfunktion mit deren Parametern.			
A 2.8.	Bestimmen Sie den Korrelationskoeffizienten nach Bravais-Pearson.			
A 2.9.	Wie ändert sich der Korrelationskoeffizient nach Bravais-Pearson, wenn die Daten der Kosten falsch waren und sie richtigerweise noch hätten quadriert werden müssen?			

A 2.10.	Wie ändert sich der Korrelationskoeffizient nach Bravais-Pearson, wenn man die Kosten mit einem Minuszeichen versehen hätte?			
A 2.11.	Berechnen und interpretieren Sie das Bestimmtheitsmaß.			
A 2.12.	Berechnen Sie die Streuung, die durch die Regressionsgerade erklärt wird.			
A 2.13.	Berechnen Sie die Streuung, die nicht durch die Regressionsgerade erklärt wird.			
A 2.14.	Bestimmen und interpretieren Sie den Residualstandardfehler.			
A 2.15.	In Periode 14 soll der durchschnittliche Output 18 Tonnen betragen. Wie hoch werden etwa die Kosten sein?			
A 2.16.	Kann mit der gleichen Güte vom Output auf die Kosten geschlossen werden, wie umgekehrt?			

2.8.3 Musteraufgabe 3 – Ordinale Merkmale

Die Unternehmensberatungsfirma „McLinsey“ trägt für 100 Unternehmen, die allesamt in der Konsumgüterbranche tätig sind, folgende Informationen für die Merkmale X: „Höhe des Forschungsaufwandes eines Unternehmens in Mio. €“ und Y: „Image des Unternehmens“ zusammen:

Y → X ↓	negativ	neutral	positiv	äußerst positiv
1	0,15	0,1	0,1	0
5	0,05	0,2	0,05	0,1
10	0	0,15	0,05	0,05

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 3.1.	Welches Skalenniveau besitzen die beiden Merkmale X und Y. Begründen Sie Ihre Entscheidung.			
A 3.2.	Stellen Sie die gemeinsame Verteilung grafisch dar.			
A 3.3.	Bestimmen Sie die gemeinsamen kumulierten relativen Häufigkeiten.			
A 3.4.	Bestimmen Sie die Verteilung des Merkmals Y bei den Unternehmen, die einen Forschungsaufwand von 5 Mio. € tätigten.			
A 3.5.	Bestimmen Sie den Median für das Merkmal Y.			
A 3.6.	Bestimmen Sie die Varianz des Merkmals X.			
A 3.7.	Bestimmen und interpretieren Sie das Assoziationsmaß nach Goodman-Kruskal.			
A 3.8.	Bestimmen und interpretieren Sie den Rangkorrelationskoeffizienten nach Spearman. Nehmen Sie dazu an, dass die Beratungsfirma „McLinsey“ das falsche Tableau benutzt hat und die (Zahlen-)Werte aus Musteraufgabe 2 die richtigen sind (Output = Forschungsaufwand, Kosten = Image).			
A 3.9.	Geben Sie zu folgenden Merkmalspaaren geeignete Zusammenhangsmaße an, die man berechnen *könnte*: Haarfarbe – Körpergewicht [in kg], Abiturnote – Wohnort, Schuhgröße [in cm] – Einkommen [in €], Kinderanzahl – Entfernung zur nächsten Schule [in km].			

2.8.4 Musteraufgabe 4 – Nominale Merkmale

Im Bundesland Hessen veröffentlichte die Forstverwaltung 1992 eine Übersicht (fiktiv), mit den Merkmalen X: Baumart und Y: Schadensart der Baumart. Es wurde dabei insgesamt 260 Bäume begutachtet.

Y → X ↓	geringer Schaden	hoher Schaden
Ahorn	35	5
Eiche	55	*a*
Buche	30	10
Fichte	65	25
Birke	6	4

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 4.1.	Welches Skalenniveau besitzen die beiden Merkmale X und Y. Begründen Sie Ihre Entscheidung.			
A 4.2.	Berechnen Sie den fehlenden Parameter *a*.			
A 4.3.	Bestimmen Sie das Tableau der relativen Häufigkeiten.			
A 4.4.	Stellen Sie das Tableau grafisch dar.			
A 4.5.	Berechnen und Interpretieren Sie $h(x_2, y_1), h(x_1 \mid y_2), h(y_1 \mid x_4), n(x_2)$.			
A 4.6.	Bestimmen und interpretieren Sie das Chi-Quadrat-Maß.			
A 4.7.	Bestimmen Sie den Kontingenzkoeffizienten nach Pearson.			
A 4.8.	Berstimmen und interpretieren Sie den normierten Kontingenzkoeffizienten nach Pearson.			
A 4.9.	Bestimmen Sie die mittlere quadratische Kontingenz.			
A 4.10.	Bestimmen und interpretieren Sie den Kontingenzkoeffizienten nach Cramer.			
A 4.11.	Bestimmen Sie das PRE-Maß, so dass man vom Merkmal X auf das Merkmal Y schließen kann.			

Das Bestimmtheitsmaß kann auch über folgenden Zusammenhang bestimmt werden:

$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$. Dabei bezeichnet s_y^2 die Gesamtvarianz der Variablen Y und $s_{\hat{y}}^2$ die erklärte Varianz der Variablen Y durch die Regressionsgerade. Somit wird der Anteil der erklärten Varianz an der Gesamtvarianz bestimmt (s. Interpretation).

Zur Berechnung der erklärten Varianz: A 2.12

A 2.12. Berechnen Sie die Streuung, die durch die Regressionsgerade erklärt wird.

Relev. ☐ Ü1 ☐ Ü2 ☐ Ü3 ☐ OK ☐

Bestimmt werden soll die Varianz, die die Regressionsgerade erklärt. Dazu benötigt man zuerst die Werte $\hat{y}_i$, die sich ergeben, wenn man nacheinander jeden Wert für x_i in die aus A 2.7 bekannte Regressionsfunktion gemäß $\hat{y}_i = \hat{a} + \hat{b} \cdot x_i$ einsetzt.
Somit ergibt sich für die Werte von $\hat{y}_i$ (für $i = 1,...,10$):

$\hat{y}_1 = 1.000 + 24 \cdot 9 = \underline{1.216}$, $\hat{y}_2 = 1.000 + 24 \cdot 12 = \underline{1.288}$, $\hat{y}_3 = 1.000 + 24 \cdot 14 = \underline{1.336}$,
... , $\hat{y}_9 = 1.000 + 24 \cdot 12 = \underline{1.288}$, $\hat{y}_{10} = 1.000 + 24 \cdot 15 = \underline{1.360}$

Aufgereiht erhält man:

1.216	1.288	1.336	1.288	1.288	1.312	1.240	1.264	1.288	1.360

Nun ist aus diesen (neuen) Werten $\hat{y}_i$ die Varianz gemäß $s_{\hat{y}}^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ zu bestimmen.

Da der Mittelwert $\bar{y}$ bereits in A 2.4 bestimmt wurde, ergibt sich:

$$s_{\hat{y}}^2 = \frac{1}{10} \cdot \left[(1.216 - 1.288)^2 + (1.288 - 1.288)^2 + ... + (1.360 - 1.288)^2\right] \Rightarrow \boxed{s_{\hat{y}}^2 = 1.612{,}8}$$

Die $\hat{y}_i$ sind also die Schätzungen für die Kosten, die sich bei Verwendung der Regressionsgleichung für jedes gegebene x_i (siehe Tabelle der Aufgabenstellung) ergeben hätten.

A 2.13. Berechnen Sie die Streuung, die nicht durch die Regressionsgerade erklärt wird.

Relev. ☐ Ü1 ☐ Ü2 ☐ Ü3 ☐ OK ☐

Hat man die Gesamtvarianz s_y^2 (A 2.5) und die erklärte Varianz (A 2.12) schon bestimmt, kann man die Varianz $s_{\hat{u}}^2$, die nicht durch die Regressionsgerade erklärt wird, durch den Zusammenhang $s_y^2 = s_{\hat{y}}^2 + s_{\hat{u}}^2$ bestimmen. Hier soll jedoch so getan werden, als hätte man dies noch nicht berechnet.
Doch was bezeichnet nun das u? Die Werte $\hat{u}_i$ geben die Differenz aus den beobachteten Werten für die Variable Y, also y_i, und den geschätzten Werten für die Variable Y, also $\hat{y}_i$, an. Man bezeichnet sie auch als Residuen bzw. Abweichungen. Da diese Abweichungen im Durchschnitt nicht vorhanden sein sollen, muss deren Mittelwert zwangsläufig Null sein. Die Varianz der Residuen bestimmt sich deshalb nur nach $s_{\hat{u}}^2 = \frac{1}{n} \cdot \hat{u}_i^2$

$s_{\hat{u}}^2$ wird auch als Varianz der Residuen (Abweichungen) bezeichnet bzw. als Streuung *um* die Regressionsgerade.

Dazu müssen zuerst die Residuen gemäß $\hat{u}_i = y_i - \hat{y}_i$ bestimmt werden, also:

$\hat{u}_1 = 1.216 - 1.216 = \underline{0}$, $\hat{u}_2 = 1.300 - 1.288 = \underline{12}$, ... , $\hat{u}_{10} = 1.360 - 1.360 = \underline{0}$

Aufgereiht erhält man:

0	12	20	0	-12	-20	20	-20	0	0

Somit ergibt sich für die Varianz der Residuen:

$$s_{\hat{u}}^2 = \frac{1}{10} \cdot (0^2 + 12^2 + ... + 0^2) \Rightarrow \boxed{s_{\hat{u}}^2 = 188{,}8}$$

Von den Ausgangswerten für Y (s. Aufgabenstellung) werden die geschätzten Werte für Y aus A 2.12 abgezogen.

Man sieht: In der Summe heben sich die Residuen gerade auf (⇒ Mittelwert der Residuen gleich Null)

A 2.14. Bestimmen und interpretieren Sie den Residualstandardfehler.

Relev. ☐ Ü1 ☐ Ü2 ☐ Ü3 ☐ OK ☐

Der (empirische) Residualstandardfehler ist wie das Bestimmtheitsmaß eine Maßzahl zur Beschreibung der Güte einer Regressionsfunktion. Allgemein berechnet man den Residualstandardfehler nach $s_{\hat{u}} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}$. Das Maß basiert auf der Varianz der Residuen, welche in A 2.13 bereits bestimmt wurde, so dass es sich hier leichter nach $s_{\hat{u}} = \sqrt{s_{\hat{u}}^2}$ berechnen lässt:

$$s_{\hat{u}} = \sqrt{188{,}08} \Rightarrow \boxed{s_{\hat{u}} = 13{,}71}$$

Interpretation: Im Durchschnitt weicht jeder aus der Regressionsgerade geschätzte Wert für die Kosten um 13.710 € vom Ausgangswert ab.
Alternative Möglichkeiten in Klausuren:

3.8 Musteraufgaben zur Konzentrationsrechnung

Diese Aufgaben sind beispielhaft für den Themenbereich. **Arbeiten Sie mit diesen Musteraufgaben, indem Sie** die einzelnen Fragen mit den Aufgabenstellungen Ihrer Übung / Ihres Tutoriums, vor allem aber mit denen der alten Klausuren Ihres Lehrstuhls vergleichen. Kreuzen Sie in den rechten Spalten die Fragestellungen an, die für Sie relevant sind und ergänzen Sie die Liste gegebenenfalls um weitere relevante Fragestellungen in diesem Themenbereich. Schicken Sie uns diese Fragestellungen per Email an verlag@studeo.de.

WICHTIG: Weitere Aufgaben finden Sie im Internet unter www.studeo.de.

3.8.1 Musteraufgabe 1 – Relative Konzentration

Auf einem Markt für elektrische Schreibmaschinen hat man 1976 eine Erhebung durchgeführt. In dieser Erhebung wurde die Anzahl der Marktteilnehmer und deren Umsatz festgehalten:

Klasse i	Umsatz [in Mio. €]	Anzahl der Marktteilnehmer	Umsatz je Klasse [in Mio. €]
1	[0 ; 5)	1.050	4.200
2	[5 ; 10)	700	6.300
3	[10 ; 15)	250	3.375
4	[15 ; 30)	50	1.250

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 1.1.	Bestimmen Sie die Anteile jeder Klasse an der Merkmalssumme.			
A 1.2.	Bestimmen Sie die kumulierten Anteile jeder Klasse an der Merkmalssumme.			
A 1.3.	Bestimmen Sie den Anteil der Merkmalsträger für jede Klasse.			
A 1.4.	Bestimmen Sie die kumulierten Anteile der Merkmalsträger für jede Klasse.			
A 1.5.	Zeichnen Sie die Lorenzkurve.			
A 1.6.	Was sagt der Punkt [0,85 ; 0,70] auf der Lorenzkurve aus?			
A 1.7.	Bestimmen Sie grafisch aus der Lorenzkurve die Konzentrationsrate $CR_{20\%}$.			
A 1.8.	Berechnen Sie den durchschnittlichen Umsatz eines Marktteilnehmers.			
A 1.9.	Berechnen und interpretieren Sie den Gini-Koeffizienten.			
A 1.10.	Berechnen Sie den normierten Gini-Koeffizienten.			
A 1.11.	Wie verändert sich der Gini-Koeffizient, wenn der Umsatz in jeder Klasse um 0,5 Millionen € ansteigt?			
A 1.12.	Wie verändert sich der Gini-Koeffizient, wenn bei der Aufstellung der Tabelle 300 Marktteilnehmer vergessen wurden, die aber alle einen Umsatz von 0 € erreicht haben?			
A 1.13.	Berechnen Sie den Rosenbluth-Koeffizienten.			

3.8.2 Musteraufgabe 2 – Absolute Konzentration

Auf einem Markt für elektrische Schreibmaschinen hat man 1996 eine Erhebung durchgeführt. In dieser Erhebung wurde für die einzelnen Marktteilnehmer der Umsatz festgehalten:

Umsatz [in Mio. €]	1, 1, 2, 2, 2, 3, 3, 5, 7, 24

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 2.1.	Bestimmen Sie die Anteile jedes Unternehmens an der Merkmalssumme, indem Sie die n Einheiten nach absteigenden Merkmalswerten sortieren.			
A 2.2.	Bestimmen Sie die Anzahl der Einheiten.			
A 2.3.	Bestimmen Sie die kumulierte Anzahl der Einheiten.			
A 2.4.	Bestimmen Sie die Konzentrationsrate für jede Einheit.			
A 2.5.	Was sagen die Werte der Konzentrationsrate CR_5 und CR_6 aus?			
A 2.6.	Zeichnen Sie die Konzentrationskurve.			
A 2.7.	Wie ändert sich die Konzentrationskurve, wenn der Umsatz des größten Unternehmens nicht 24 Mio. €, sondern 74 Mio. € betragen würde?			
A 2.8.	Berechnen und interpretieren Sie den Hirschmann-Herfindahl-Koeffizienten.			
A 2.9.	Wie ändert sich der Hirschmann-Herfindahl-Koeffizient, wenn statt in € in DM gemessen worden wäre?			
A 2.10.	Berechnen und interpretieren Sie den Rosenbluth-Koeffizienten.			
A 2.11.	Berechnen Sie den Gini-Koeffizienten.			
A 2.12.	Vergleichen Sie die Gini-Koeffizienten aus Musteraufgabe 1 und 2. In welchem Jahr war die Konzentration höher?			
A 2.13.	Worin besteht der Unterschied zwischen relativen und absoluten Konzentrationsmaßen?			

3.9 Musterlösungen zur Konzentrationsrechnung

Diese Musterlösungen sind beispielhaft. Wir haben uns bemüht, insbesondere die Rechenschritte ausführlicher darzustellen als in der Klausur eigentlich nötig. Erläuterungen stehen in der rechten Spalte. **Arbeiten Sie mit diesen Lösungen, indem Sie** den Weg eigenständig nachvollziehen und sich Bemerkungen am Rande machen. Sie haben bereits die Aufgabenstellungen mit den Aufgaben Ihrer Übung und der alten Klausuren verglichen. Jetzt müssen Sie dasselbe für die Lösungen machen. Vergleichen Sie die Lösungen Schritt für Schritt und machen Sie sich Notizen. Haken Sie die Lösungen ab, die Sie beherrschen. Lösen Sie die Aufgaben immer wieder, bis Sie sie ohne Nachzuschauen beherrschen. Üben Sie Termumformungen mit dem Studeo®-Rechentrainer (www.rechentrainer.de).

3.9.1 Musterlösung 1 – Relative Konzentration

Auf einem Markt für elektrische Schreibmaschinen hat man 1976 eine Erhebung durchgeführt. In dieser Erhebung wurde die Anzahl der Marktteilnehmer und deren Umsatz festgehalten:

Klasse i	Umsatz [in Mio. €]	Anzahl der Marktteilnehmer	Umsatz je Klasse [in Mio. €]
1	[0 ; 5)	1.050	4.200
2	[5 ; 10)	700	6.300
3	[10 ; 15)	250	3.375
4	[15 ; 30)	50	1.250

Lösung | **Erläuterungen / Notizen**

A 1.1. Bestimmen Sie die Anteile jeder Klasse an der Merkmalssumme. | Relev. ☐ Ü1 ☐ Ü2 ☐ Ü3 ☐ OK ☐

Bevor die Anteile jeder Klasse an der Merkmalssumme bestimmt werden, müssen einige Fragen geklärt werden. Zuerst ist festzuhalten, welches Merkmal betrachtet werden soll. Dies geht hier aus der Aufgabenstellung hervor. Das Merkmal X ist der Umsatz [in Mio. €] und besitzt die Ausprägungen [0 ; 5) , [5 ; 10) , [10 ; 15) und [15 ; 30). Es liegt also ein klassiertes Merkmal vor, welches kardinalskaliert ist. Das ist wichtig, weil für die Konzentrationsmessung mindestens ordinale Merkmale vorliegen müssen. Außerdem dürfen die Werte nicht negativ sein.
Die absoluten Häufigkeiten n_i (wobei i die i-te Klasse bezeichnet) sind in der dritten Spalte der Tabelle abgetragen. Die Marktteilnehmer sind die Merkmalsträger. n_i bezeichnet also die Anzahl der Marktteilnehmer mit einem Umsatz in Höhe der i-ten Klasse (bspw. von 5 bis 10 Mio. €).

Die Bezeichnung $n(x_i)$ ist hier nicht notwendig, da nur ein Merkmal betrachtet wird.

Es sollen hier relative Konzentrationsmaße bestimmt werden. Die Frage lautet demnach: wie viel Prozent der Unternehmen machen wie viel Prozent des Umsatzes. Daher müssen die Merkmalsträger nach aufsteigenden Merkmalswerten (=Merkmalsausprägungen) sortiert werden. Das ist in obiger Tabelle (in Spalte zwei) bereits der Fall.

Bei absoluten Konzentrationsmaßen werden sie absteigend sortiert!

Zum Begriff Merkmalssumme:
Unter der Merkmalssumme einer Klasse versteht man die Summe aller Merkmalswerte x_i einer Klasse. Da die Originaldaten nicht gegeben sind (klassiertes Merkmal), kann die genaue Merkmalssumme je Klasse nicht berechnet werden, es sei denn, man hat zusätzliche Informationen. Das ist der Fall. In der vierten Spalte findet sich die aus den Einzelwerten aggregierte Merkmalssumme.

Eine Schätzung ist aber möglich.

Bezogen auf die vorliegende Aufgabe versteht man unter der Merkmalssumme S_i der Klasse i, den Umsatz, den die einzelnen Unternehmen in der Klasse i zusammen gemacht haben.
Der Anteil q_i der Klasse i an der Merkmalssumme ergibt sich nun gemäß $q_i = \frac{S_i}{S}$, wobei S die Merkmalssumme aller Klassen bedeutet und sich damit nach $S = \sum_{i=1}^{k} S_i$ berechnen lässt. Daraus folgt:

Für $i = 1,\ldots,k$

$$S = \sum_{i=1}^{k} S_i = S_1 + S_2 + S_3 + S_4 = 4.200 + 6.300 + 3.375 + 1.250 = \underline{15.125}$$

$$\Rightarrow q_1 = \frac{S_1}{S} = \frac{4.200}{15.125} = \underline{0,28}, \qquad q_2 = \frac{S_2}{S} = \frac{6.300}{15.125} = \underline{0,42},$$

$$q_3 = \frac{S_3}{S} = \frac{3.375}{15.125} = \underline{0,22} \quad \text{und} \quad q_4 = \frac{S_4}{S} = \frac{1.250}{15.125} = \underline{0,08}$$

Dabei gilt stets: $\sum_{i=1}^{k} q_i = 1$

Alternative Möglichkeiten in Klausuren:
Natürlich kann es in Klausuren vorkommen, dass die Merkmalssummen nicht gegeben sind. Dann muss die Annahme unterstellt werden, dass sich innerhalb jeder Klasse die Werte gleich verteilen – und somit innerhalb der Klasse keine Konzentration herrscht. Die Merkmalssumme jeder Klasse muss nach $\hat{S}_i = n_i \cdot \hat{\bar{x}}_i$ geschätzt werden, wobei der Klassenmittelwert nach $\hat{\bar{x}}_i = 0,5 \cdot (x_u + x_o)$

Vgl. Sie dazu auch K.1/A. 2.9.

x_u (x_o) war die untere (obere) Grenze der betrachteten Klasse

4.7 Formelsammlung zur Indexrechnung

Diese Formelsammlung enthält die wichtigsten Formeln zum Teilgebiet, geordnet nach inhaltlichen Gesichtspunkten, nicht alphabetisch! **Arbeiten Sie mit dieser Formelsammlung, indem Sie** sie mit den Formeln bzw. Aufgabenstellungen Ihrer Übung/ Ihres Tutoriums vergleichen. Stellen Sie fest, ob jede Formel für Sie relevant ist und tragen Sie Ihre Schreibweise ein. Ergänzen Sie die Sammlung gegebenenfalls um weitere Formeln. (Rel in der ersten Zeile heißt relevant!)

Bezeichnung	Formel	Nr.	Rel	Ihre Schreibweise
Aggregation von Preismesszahlen	$P_{0,t} = \sum_{j=1}^{k} P_{0,t}^{j} \cdot \frac{u_j}{u_{gesamt}}$	(4.1)		
Gesamtumsatz	$u_{gesamt} = \sum_{j=1}^{k} u_j$	(4.2)		
Kaufkraftparitäten	$KKP = \frac{\sum_{i=1}^{n} p_{Bi} \cdot q_i}{\sum_{i=1}^{n} p_{Ai} \cdot q_i}$	(4.3)		
Mengenindex nach Fisher	$Q_{0,t}^{F} = \sqrt{Q_{0,t}^{L} \cdot Q_{0,t}^{P}}$	(4.4)		
Mengenindex nach Laspeyres	$Q_{0,t}^{L} = \frac{\sum_{i=1}^{n} p_{0i} \cdot q_{ti}}{\sum_{i=1}^{n} p_{0i} \cdot q_{0i}}$	(4.5)		
Mengenindex nach Lowe	$Q_{0,t}^{Lo} = \frac{\sum_{i=1}^{n} \overline{p}_i \cdot q_{ti}}{\sum_{i=1}^{n} \overline{p}_i \cdot q_{0i}}$	(4.6)		
Mengenindex nach Paasche	$Q_{0,t}^{P} = \frac{\sum_{i=1}^{n} p_{ti} \cdot q_{ti}}{\sum_{i=1}^{n} p_{ti} \cdot q_{0i}}$	(4.7)		
Mengenmittelwert eines Gutes	$\overline{q}_i = \frac{1}{T} \cdot \sum_{\tau=1}^{T} q_{\tau i}$	(4.8)		
Preisindex nach Drobisch	$P_{0,t}^{Dro} = \frac{\frac{\sum_{i=1}^{n} p_{ti} \cdot q_{ti}}{\sum_{i=1}^{n} q_{ti}}}{\frac{\sum_{i=1}^{n} p_{0i} \cdot q_{0i}}{\sum_{i=1}^{n} q_{0i}}}$	(4.9)		
Preisindex nach Fisher	$P_{0,t}^{F} = \sqrt{P_{0,t}^{L} \cdot P_{0,t}^{P}}$	(4.10)		
Preisindex nach Laspeyres	$P_{0,t}^{L} = \frac{\sum_{i=1}^{n} p_{ti} \cdot q_{0i}}{\sum_{i=1}^{n} p_{0i} \cdot q_{0i}}$	(4.11)		

4.8 Musteraufgaben zur Indexrechnung

Diese Aufgaben sind beispielhaft für den Themenbereich. **Arbeiten Sie mit diesen Musteraufgaben, indem Sie** die einzelnen Fragen mit den Aufgabenstellungen Ihrer Übung / Ihres Tutoriums, vor allem aber mit denen der alten Klausuren Ihres Lehrstuhls vergleichen. Kreuzen Sie in den rechten Spalten die Fragestellungen an, die für Sie relevant sind und ergänzen Sie die Liste gegebenenfalls um weitere relevante Fragestellungen in diesem Themenbereich. Schicken Sie uns diese Fragestellungen per Email an verlag@studeo.de.

WICHTIG: Weitere Aufgaben finden Sie im Internet unter www.studeo.de.

4.8.1 Musteraufgabe 1 – Indexberechnung

Der Partyservice „Juchhu" , der sich insbesondere im Liefern von kulinarischen Köstlichkeiten einen Namen gemacht hat, notierte in den Jahren 1998 und 2001 folgende Güterpreise pro 100g in DM:

	1998		2001	
	Preis	Menge	Preis	Menge
Schnitzel	4,00	100	7,05	50
Salat	3,00	150	6,55	125
Kartoffeln	3,50	125	4,70	175
Gurken	3,20	115	5,00	170
kl. Snacks	2,50	145	6,15	145

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 1.1.	Ermitteln und interpretieren Sie den Preisindex nach Laspeyres der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.2.	Ermitteln und interpretieren Sie den Preisindex nach Paasche der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.3.	Ermitteln und interpretieren Sie den Preisindex nach Marshall-Edgeworth der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.4.	Ermitteln und interpretieren Sie den Preisindex nach Fisher der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.5.	Ermitteln und interpretieren Sie den Preisindex nach Lowe der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.6.	Ermitteln und interpretieren Sie den Preisindex nach Drobisch der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.7.	Ermitteln und interpretieren Sie die Kaufkraftparitäten, indem Sie annehmen, dass die Spalten mit den Jahreszahlen nun den Ländern Belgien (entspricht 1998) und Deutschland (entspricht 2001) entsprechen.			
A 1.8.	Berechnen und interpretieren Sie den Mengenindex nach Laspeyres der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.9.	Berechnen und interpretieren Sie den Mengenindex nach Paasche der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.10.	Berechnen und interpretieren Sie den Mengenindex nach Fisher der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.11.	Berechnen und interpretieren Sie den Mengenindex nach Lowe der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.12.	Berechnen und interpretieren Sie den Umsatzindex der kulinarischen Köstlichkeiten von 1998 auf 2001.			
A 1.13.	Fassen Sie die kulinarischen Köstlichkeiten zu zwei Gruppen zusammen. Gruppe eins beinhaltet Schnitzel und Salat, Gruppe zwei den Rest. Bestimmen Sie nun die beiden Sub-Indizes nach Laspeyres und anschließend den Gesamtindex nach Laspeyres.			

4.8.2 Musteraufgabe 2 – Rechnen mit Indizes

In den letzten Jahren wurden die Preise für Videokassetten erfasst. Die Preismesszahlen zu den Basisjahren 1995 und 1998 sind in nachfolgender, unvollständiger Tabelle dargestellt:

Jahr	1995	1996	1997	1998	1999	2000	2001
$\frac{p_t}{p_{95}}$		1,13	1,18		1,10		
$\frac{p_t}{p_{98}}$					0,95	0,85	0,80

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 2.1.	Berechnen Sie die Preismesszahl im Berichtsjahr 1998 zum Basisjahr 1995.			
A 2.2.	Berechnen Sie alle Preismesszahlen zum Basisjahr 1995.			
A 2.3.	Berechnen Sie alle Preismesszahlen zum Basisjahr 1998.			
A 2.4.	Interpretieren Sie die Preismesszahl im Berichtsjahr 2001 zum Basisjahr 1998.			
A 2.5.	Basieren Sie alle Preismesszahlen des Basisjahres 95 auf das Basisjahr 2001 um.			

hen: $P^{P}_{98,01} = \dfrac{\sum_{i=1}^{n} p_{01i} \cdot q_{01i}}{\sum_{i=1}^{n} p_{98i} \cdot q_{01i}}$

$$P^{P}_{98,01} = \frac{7,05 \cdot 50 + 6,55 \cdot 125 + 4,70 \cdot 175 + 5,00 \cdot 170 + 6,15 \cdot 145}{4,00 \cdot 50 + 3,00 \cdot 125 + 3,50 \cdot 125 + 3,20 \cdot 115 + 2,50 \cdot 145} = \frac{3.735,5}{2.094} \quad \Rightarrow \boxed{P^{P}_{98,01} = 1,78}$$

Interpretation: Der Preisindex nach Paasche gibt an, dass sich der Preis der kulinarischen Köstlichkeiten von 1998 zu 2001 bei aktualisierten Verbrauchsmengen, d.h. den Verbrauchsmengen aus dem Jahr 2001, um 78% erhöht hat.

A 1.3. Ermitteln und interpretieren Sie den Preisindex nach Marshall-Edgeworth der kulinarischen Köstlichkeiten von 1998 auf 2001.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Man erhält den Preisindex nach Marshall-Edgeworth allgemein nach der Formel $P^{M/E}_{0,t} = \dfrac{\sum_{i=1}^{n} p_{ti} \cdot (q_{0i} + q_{ti})}{\sum_{i=1}^{n} p_{0i} \cdot (q_{0i} + q_{ti})}$. Somit ist $P^{M/E}_{98,01} = \dfrac{\sum_{i=1}^{n} p_{01i} \cdot (q_{98i} + q_{01i})}{\sum_{i=1}^{n} p_{98i} \cdot (q_{98i} + q_{01i})}$ zu berechnen:

Das M/E steht für Marshall-Edgeworth.

$$P^{M/E}_{98,01} = \frac{7,05 \cdot (100+50) + 6,55 \cdot (150+125) + \ldots + 6,15 \cdot (145+145)}{4,00 \cdot (100+50) + 3,00 \cdot (150+125) + \ldots + 2,50 \cdot (145+145)} = \frac{7.477,25}{4.112}$$

$$\Rightarrow \boxed{P^{M/E}_{98,01} = 1,82}$$

Interpretation: Der Preisindex nach Marshall-Edgeworth gibt an, dass sich der Preis der kulinarischen Köstlichkeiten von 1998 zu 2001, legt man die Verbrauchsmengen aus Basis- und Berichtsjahr zugrunde, um 82% erhöht hat.

A 1.4. Ermitteln und interpretieren Sie den Preisindex nach Fisher der kulinarischen Köstlichkeiten von 1998 auf 2001.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Der Preisindex nach Fisher lässt sich allgemein nach $P^{F}_{0,t} = \sqrt{P^{L}_{0,t} \cdot P^{P}_{0,t}}$ bestimmen. Auf die Aufgabenstellung bezogen muss also $P^{F}_{98,01} = \sqrt{P^{L}_{98,01} \cdot P^{P}_{98,01}}$ berechnet werden, wobei die Ergebnisse aus A 1.1 und A 1.2 Verwendung finden:

Das F steht für Fisher.

$$P^{F}_{98,01} = \sqrt{1,85 \cdot 1,78} \quad \Rightarrow \boxed{P^{F}_{98,01} = 1,81}$$

Interpretation: Der Preisindex nach Fisher gibt an, dass sich der Preis der kulinarischen Köstlichkeiten von 1998 zu 2001, legt man das geometrische Mittel der Preisindizes nach Laspeyres und Paasche zugrunde, um 81% erhöht hat.

A 1.5. Ermitteln und interpretieren Sie den Preisindex nach Lowe der kulinarischen Köstlichkeiten von 1998 auf 2001.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Der Preisindex nach Lowe berechnet sich allgemein nach $P^{Lo}_{0,t} = \dfrac{\sum_{i=1}^{n} p_{ti} \cdot \overline{q}_i}{\sum_{i=1}^{n} p_{0i} \cdot \overline{q}_i}$. Dabei bezeichnet $\overline{q}_i$ den Mittelwert aus allen Mengendaten des Gutes i, die zwischen Basis- und Berichtsperiode liegen. Berechnen kann man den Mittelwert mittels $\overline{q}_i = \dfrac{1}{T} \cdot \sum_{\tau=1}^{T} q_{\tau i}$. Für das Basisjahr 1998 und das Berichtsjahr 2001 ergibt sich somit:

Das Lo steht für Lowe.

τ (mit $\tau = 1,\ldots,T$) zählt die Anzahl der Mengendaten für ein Gut i. Da hier nur die Daten aus Basis- und Berichtsjahr vorhanden sind, muss man nur den Mittelwert aus 2 Werten berechnen.

$$\overline{q}_1 = \frac{1}{2} \cdot \sum_{\tau=1}^{2} q_{\tau 1} = \frac{1}{2} \cdot (q_{11} + q_{21}) = \frac{1}{2} \cdot (100+50) = \underline{75} \quad ,$$

$$\overline{q}_2 = \frac{1}{2} \cdot \sum_{\tau=1}^{2} q_{\tau 2} = \frac{1}{2} \cdot (q_{12} + q_{22}) = \frac{1}{2} \cdot (150+125) = \underline{137,5} \quad ,$$

Begriff	Definition	Symbol?	Relevant	Kann ich	noch lernen
Verfahren gleitender Durchschnitte	Ein Verfahren zur Glättung von Zeitreihenwerten. Die sich ergebenden neuen Zeitreihenwerte können als lokale arithmetische Mittelwerte der ursprünglichen Zeitreihenwerte aufgefaßt werden.				
Zeitreihenana-lyse	Bereich der Statistik, der sich mit der Beschreibung von Beobachtungswerten eines Merkmals für unterschiedliche Zeitpunkte und deren Untersuchung auf Gesetzmäßigkeiten hin beschäftigt.				

5.8 Musteraufgaben zur Zeitreihenanalyse

Diese Aufgaben sind beispielhaft für den Themenbereich. **Arbeiten Sie mit diesen Musteraufgaben, indem Sie** die einzelnen Fragen mit den Aufgabenstellungen Ihrer Übung / Ihres Tutoriums, vor allem aber mit denen der alten Klausuren Ihres Lehrstuhls vergleichen. Kreuzen Sie in den rechten Spalten die Fragestellungen an, die für Sie relevant sind und ergänzen Sie die Liste gegebenenfalls um weitere relevante Fragestellungen in diesem Themenbereich. Schicken Sie uns diese Fragestellungen per Email an verlag@studeo.de.

WICHTIG: Weitere Aufgaben finden Sie im Internet unter www.studeo.de.

5.8.1 Musteraufgabe 1 – Lokale Trendbestimmung

Das Unternehmen „4you Telecomunications" veröffentlicht in den Jahren 1999 bis 2001 folgende Quartalsumsatzzahlen (in 10^5 DM), wobei im folgenden stets ein additives Modell unterstellt wird:

Jahr	1999				2000				2001			
Index t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	50	62	64	78	65	71	69	92	79	84	84	101

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 1.1.	Zeichnen Sie die Zeitreihe der Quartalsumsatzzahlen.			
A 1.2.	Wie lautet das additive Modell zur Beschreibung von Zeitreihen? Erklären Sie zudem an einem Beispiel, was die einzelnen Komponenten des Modells bedeuten.			
A 1.3.	Bestimmen Sie den gleitenden Durchschnitt 3. Ordnung.			
A 1.4.	Bestimmen Sie den gleitenden Durchschnitt 4. Ordnung.			
A 1.5.	Bestimmen Sie den gleitenden Durchschnitt 5. Ordnung.			
A 1.6.	Bestimmen Sie den gleitenden Durchschnitt 6. Ordnung.			
A 1.7.	Welcher gleitende Durchschnitt ist besser geeignet, den Saisoneffekt zu eliminieren?			
A 1.8.	Bestimmen Sie die konstante Saisonkomponente.			
A 1.9.	Bestimmen Sie die saisonbereinigte Reihe.			
A 1.10.	Zeichnen sie die saisonbereinigte Reihe in Ihre bereits vorhandene Zeichnung mit ein.			
A 1.11.	Bestimmen Sie die exponentiell geglätteten Werte 1.Ordung mit $y_0^* = 48$ und $\alpha = 0{,}7$.			
A 1.12.	Bestimmen Sie die exponentiell geglätteten Werte 2.Ordung mit $y_0^{**} = 49$ und $\alpha = 0{,}7$.			
A 1.13.	Stellen sie die exponentiell geglätteten Werte beider Ordnungen in einer Grafik dar.			
A 1.14.	Prognostizieren Sie zum Zeitpunkt $t = 12$ den Wert des Umsatzes im nächsten Quartal unter Verwendung des Verfahrens der exponentiellen Glättung.			
A 1.15.	Unter welchen Bedingungen wird anstatt eines additiven Modells ein multiplikatives Modell verwendet?			

5.8.2 Musteraufgabe 2 – Globale Trendbestimmung

Der Sohn einer reichen Familie, Klein-Herbert, stellt folgende Übersicht über die Entwicklung seiner Weihnachtsgeschenke [in DM] dar:

Jahr	1994	1995	1996	1997	1998	1999	2000	2001
Wert der Geschenke	50	2.525	3.763	4.381	4.691	4.845	4.923	4.961

Ref.Nr.	Aufgabenstellung	Relevant	Klar	Üben
A 2.1.	Worin besteht der Unterschied zwischen der lokalen und der globalen Trendbestimmung?			
A 2.2.	Stellen Sie die Entwicklung der Weihnachtsgeschenke von Klein-Herbert grafisch dar.			
A 2.3.	Bestimmen Sie rechnerisch, ob zur Beschreibung der Zeitreihe ein linearer Trend geeignet erscheint.			
A 2.4.	Bestimmen Sie rechnerisch, ob zur Beschreibung der Zeitreihe ein Exponentialtrend geeignet erscheint.			
A 2.5.	Bestimmen Sie rechnerisch, ob zur Beschreibung der Zeitreihe ein modifizierter Exponentialtrend geeignet erscheint.			
A 2.6.	Bestimmen Sie rechnerisch, ob zur Beschreibung der Zeitreihe ein logistischer Trend geeignet erscheint.			
A 2.7.	Bestimmen Sie rechnerisch, ob zur Beschreibung der Zeitreihe ein Gompertztrend geeignet erscheint.			
A 2.8.	Bestimmen Sie für diese Zeitreihe die Parameter eines linearen Trends.			
A 2.9.	Bestimmen Sie für diese Zeitreihe die Parameter eines Exponentialtrends.			
A 2.10.	Bestimmen Sie für diese Zeitreihe die Parameter eines modifizierten Exponentialtrends und zeichnen Sie die Trendfunktion.			
A 2.11.	Bestimmen Sie für diese Zeitreihe die Parameter eines logistischen Trends.			
A 2.12.	Mit welchem Euro-Betrag kann Klein-Herbert an Weihnachten 2003 rechnen? (Ziehen Sie zur Berechnung den geeigneten Trend heran.)			
A 2.13.	In welchem Jahr überschreiten die Geschenke einen Betrag von 4.999,50 DM? (Runden Sie dazu auf zwei Nachkommastellen genau; Wahl des geeigneten Trends.)			
A 2.14.	Warum ist eine Interpretation des Ergebnisses aus Aufgabe 2.13 nicht unproblematisch?			

5.9 Musterlösungen zur Zeitreihenanalyse

Diese Musterlösungen sind beispielhaft. Wir haben uns bemüht, insbesondere die Rechenschritte ausführlicher darzustellen als in der Klausur eigentlich nötig. Erläuterungen stehen in der rechten Spalte. **Arbeiten Sie mit diesen Lösungen, indem Sie** den Weg eigenständig nachvollziehen und sich Bemerkungen am Rande machen. Sie haben bereits die Aufgabenstellungen mit den Aufgaben Ihrer Übung und der alten Klausuren verglichen. Jetzt müssen Sie dasselbe für die Lösungen machen. Vergleichen Sie die Lösungen Schritt für Schritt und machen Sie sich Notizen. Haken Sie die Lösungen ab, die Sie beherrschen. Lösen Sie die Aufgaben immer wieder, bis Sie sie ohne Nachzuschauen beherrschen. Üben Sie Termumformungen mit dem Studeo®-Rechentrainer (www.rechentrainer.de).

5.9.1 Musterlösung 1 – Lokale Trendbestimmung

Das Unternehmen „4you Telecomunications“ veröffentlicht in den Jahren 1999 bis 2001 folgende Quartalsumsatzzahlen (in 10^5 DM), wobei im folgenden stets ein additives Modell unterstellt wird:

Jahr	1999				2000				2001			
Index t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	50	62	64	78	65	71	69	92	79	84	84	101

Lösung | Erläuterungen / Notizen

A 1.1. Zeichnen Sie die Zeitreihe der Quartalsumsatzzahlen.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Dazu zeichnet man ein Koordinatensystem, wobei man auf die Abszisse den Index t und auf die Ordinate den Zeitreihenwert y_t abträgt. Meist verbindet man die Punkte noch geradlinig:

Die Werte sind aus der Aufgabenstellung zu entnehmen.

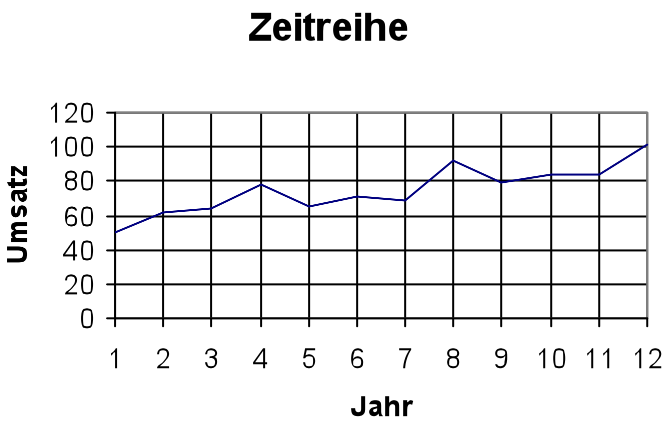


Abb. 5-3: Einfache Zeitreihe

A 1.2. Wie lautet das additive Modell zur Beschreibung von Zeitreihen? Erklären Sie zudem an einem Beispiel, was die einzelnen Komponenten des Modells bedeuten.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Die Zeitreihenanalyse beschäftigt sich mit der Beschreibung von Beobachtungswerten eines Merkmals für unterschiedliche Zeitpunkte und deren Untersuchung auf Gesetzmäßigkeiten. Dazu wird meist ein additives Modell der Form $y_t = m_t + c_t + s_t + u_t$ (⇒ Linearität) unterstellt, manchmal auch ein multiplikatives.

Für $t = 1,\ldots,n$.

Dabei stellt die Trendkomponente m_t eine längerfristige systematische Veränderung des Niveaus der Zeitreihe dar, welche durch längerfristig wirkende Ursachen hervorgerufen wird. Die Konjunkturkomponente c_t stellt eine mehrjährige, nicht unbedingt regelmäßige Schwankung der Zeitreihe dar, die in konjunkturellen Einflüssen ihren Ursprung hat. Die Saisonkomponente s_t beinhaltet Schwankungen, die abhängig von der Jahreszeit sind. Sie verläuft innerhalb eines Jahres wellenförmig und wiederholt sich jedes Jahr, wobei sich die Werte in der Summe (innerhalb eines Jahres) gegenseitig aufheben. Die Bewegungen, welche auf und ab verlaufen können, nennt man konstante Saisonfigur. Die irreguläre Komponente u_t , auch Restkomponente genannt, fasst die restlichen Einflüsse zusammen, welche noch nicht durch die zuvor genannten Komponenten erklärt wurden.

Manchmal auch zyklische Komponente genannt.

Die Saisonkomponente nimmt also zu denselben Zeitpunkten (bspw. 1.Quartal) in verschiedenen Jahren stets die gleichen Werte an.

Weil die Trennung von Trend und Konjunkturkomponente oftmals Schwierigkeiten bereitet, wird sie deswegen zur glatten Komponente g_t zusammengefasst. Das Modell vereinfacht sich dann zu $y_t = g_t + s_t + u_t$.

A 1.3. Bestimmen Sie den gleitenden Durchschnitt 3. Ordnung.

Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Um die Trend- und die Konjunkturkomponente aus den Werten der Zeitreihe zu isolieren, wird

Zeitpunkten $t = 11$ und $t = 12$ möglich, da keine zwei Werte nachfolgen. Daraus ergibt sich diese Rechnung:

$$\hat{g}_3^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot y_{3-2} + y_{3-1} + y_3 + y_{3+1} + \frac{1}{2} \cdot y_{3+2} \right) = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot y_1 + y_2 + y_3 + y_4 + \frac{1}{2} \cdot y_5 \right)$$

$$= \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 50 + 62 + 64 + 78 + \frac{1}{2} \cdot 65 \right) = \underline{65{,}38} \quad ,$$

$$\hat{g}_4^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot y_2 + y_3 + y_4 + y_5 + \frac{1}{2} \cdot y_6 \right) = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 62 + 64 + 78 + 65 + \frac{1}{2} \cdot 71 \right) = \underline{68{,}38} \quad ,$$

Die Werte „rutschen“ der Reihe nach von vorne nach hinten durch.

$$\hat{g}_5^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot y_3 + y_4 + y_5 + y_6 + \frac{1}{2} \cdot y_7 \right) = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 64 + 78 + 65 + 71 + \frac{1}{2} \cdot 69 \right) = \underline{70{,}13} \quad ,$$

$$\hat{g}_6^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 78 + 65 + 71 + 69 + \frac{1}{2} \cdot 92 \right) = \underline{72{,}5} \quad ,$$

$$\hat{g}_7^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 65 + 71 + 69 + 92 + \frac{1}{2} \cdot 79 \right) = \underline{76} \quad ,$$

$$\hat{g}_8^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 71 + 69 + 92 + 79 + \frac{1}{2} \cdot 84 \right) = \underline{79{,}38} \quad ,$$

$$\hat{g}_9^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 69 + 92 + 79 + 84 + \frac{1}{2} \cdot 84 \right) = \underline{82{,}88} \quad \text{und}$$

$$\hat{g}_{10}^{(4)} = \frac{1}{4} \cdot \left(\frac{1}{2} \cdot 92 + 79 + 84 + 84 + \frac{1}{2} \cdot 101 \right) = \underline{85{,}88}$$

A 1.5. Bestimmen Sie den gleitenden Durchschnitt 5. Ordnung. — Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Um den gleitenden Durchschnitt fünfter Ordnung bestimmen zu können, muss Gleichung $\hat{g}_t^{(k)} = \frac{1}{k} \cdot \sum_{i=-q}^{q} y_{t+i}$ mit $(t = q+1, \ldots, n-q)$ verwendet werden. Man muss folglich gemäß $\hat{g}_t^{(5)} = \frac{1}{5} \cdot \sum_{i=-2}^{2} y_{t+i} = \frac{1}{5} \cdot (y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2})$ aus $k = 5$ Werten symmetrisch um den dritten Wert herum den Mittelwert bilden. Für 4 der 12 Zeitreihenwerte können die gleitenden Durchschnitte fünfter Ordnung nicht gebildet werden, für die ersten beiden und für die letzten beiden nicht:

$k = 5$ ist ungerade. Wegen $k = 2q + 1$ muss $q = 2$ sein.

Vgl. Sie auch A 1.3.

$$\hat{g}_3^{(5)} = \frac{1}{5} \cdot (y_1 + y_2 + y_3 + y_4 + y_5) = \frac{1}{5} \cdot (50 + 62 + 64 + 78 + 65) = \underline{63{,}8} \quad ,$$

$$\hat{g}_4^{(5)} = \frac{1}{5} \cdot (y_2 + y_3 + y_4 + y_5 + y_6) = \frac{1}{5} \cdot (62 + 64 + 78 + 65 + 71) = \underline{85} \quad ,$$

$\hat{g}_5^{(5)} = \ldots = \underline{86{,}75}$ $\quad \hat{g}_6^{(5)} = \ldots = \underline{93{,}75}$ $\quad \hat{g}_7^{(5)} = \ldots = \underline{94}$

$\hat{g}_8^{(5)} = \ldots = \underline{98{,}75}$,

$\hat{g}_9^{(5)} = \ldots = \underline{102}$ und $\hat{g}_{10}^{(5)} = \ldots = \underline{110}$

Alternative Möglichkeiten in Klausuren:
Manchmal soll sogar der gleitende Durchschnitt einer höheren ungeraden Ordnung bestimmt werden. Dann ist darauf zu achten, dass auch für immer weniger Zeitreihenwerte der gleitende Durchschnitt berechnet werden kann, bzw. immer mehr Werte herausfallen. Wird der gleitende Durchschnitt $k = 2q + 1$ Ordnung bestimmt, können für $2q$ Zeitreihenwerte keine gleitenden Durchschnitte k-ter Ordnung bestimmt werden.

Genauer: Für die ersten und die letzten q Werte geht dies nicht.

A 1.6. Bestimmen Sie den gleitenden Durchschnitt 6. Ordnung. — Relev. □ Ü1 □ Ü2 □ Ü3 □ OK □

Vgl. Sie auch A 1.4.

Für den gleitenden Durchschnitt sechster Ordnung muss Gleichung $\hat{g}_t^{(k)} = \frac{1}{k} \cdot \left(\frac{1}{2} \cdot y_{t-q} + \sum_{i=-q+1}^{q-1} y_{t+i} + \frac{1}{2} \cdot y_{t+q} \right)$ mit $(t = q+1, \ldots, n-q)$ benutzt werden. Dementsprechend muss aus $k = 6$ Werten symmetrisch um den vierten Wert herum der Mittelwert bestimm werden, wobei hier der erste und siebte Wert nur zur Hälfte miteinbezogen werden, so dass insgesamt aus sechs „vollen“ Werten der Durchschnitt gebildet wird. Somit kann aus sechs Werten der gleitende Durchschnitt sechster Ordnung nicht bestimmt werden. Nach

Es sind dies die ersten und letzten drei.

5.10 Übungsaufgaben zur Zeitreihenanalyse

Diese Übungsaufgaben helfen Ihnen beim Training der behandelten Aufgabentypen. Sie sollten erst versuchen, die Aufgaben zu lösen, bevor Sie sich die Lösungen ansehen. Blättern Sie im Zweifelsfall zurück zur Aufgabenstellung der Musterlösung und den Algorithmen, um den Lösungsweg selbst zu finden. Weitere Übungsaufgaben finden Sie im Internet auf www.studeo.de.

Aufgabe 1

Für 1995 und 1996 liegen die folgenden Umsatzzahlen (in Mio. DM) vor:

Jahr	1995				1996			
Quartal	I	II	III	IV	I	II	III	IV
x_t	100	82	73	107	109	93	84	126

a) Zeichnen Sie die Quartalsumsatzzahlen!
b) Bestimmen Sie einen geeigneten gleitenden Durchschnitt!
c) Bestimmen Sie die saisonbereinigte Zeitreihe und zeichnen Sie sie in die Grafik aus Teilaufgabe a ein!
d) Bestimmen Sie die exponentiell geglätteten Werte 1.Ordnung mit $y_0^* = 95$ und $\alpha = 0{,}6$!
e) Bestimmen Sie die exponentiell geglätteten Werte 2.Ordnung mit $y_0^{**} = 90$ und $\alpha = 0{,}6$!
f) Stellen Sie die exponentiell geglätteten Werte beider Ordnungen in einer Grafik dar!
g) Prognostizieren Sie im vierten Quartal im Jahr 1996 den Umsatz für das erste Quartal 1997 unter Verwendung des Verfahrens der exponentiellen Glättung!

Aufgabe 2

Für die Mineralölimporte eines Industrielandes liege die folgende Zeitreihe y_t von Halbjahres- werten vor:

Jahr	1984		1985		1986		1987		1988	
Halbjahr	I	II	I	II	I	II	I	II	I	II
Umsatz in Mio. DM	50	40	56	46	58	48	54	48	60	52

Das additive Zeitreihenmodell mit konstanter Saisonfigur soll hier unterstellt werden.

a) Zeichnen Sie die Umsatzzahlen!
b) Bestimmen Sie einen geeigneten gleitenden Durchschnitt!
c) Bestimmen Sie die saisonbereinigte Zeitreihe und zeichnen Sie sie in die Grafik aus Teilaufgabe a ein!
d) Bestimmen Sie die exponentiell geglätteten Werte 1.Ordnung mit $y_0^* = 47$ und $\alpha = 0{,}4$!
e) Bestimmen Sie die exponentiell geglätteten Werte 2.Ordnung mit $y_0^{**} = 45$ und $\alpha = 0{,}4$!
f) Stellen Sie die exponentiell geglätteten Werte beider Ordnungen in einer Grafik dar!
g) Prognostizieren Sie im zweiten Halbjahr im Jahr 1988 den Umsatz für das erste Halbjahr 1989 unter Verwendung des Verfahrens der exponentiellen Glättung!

Aufgabe 3

Die Außenhandelsstatistik liefert folgende 4-Monats-Daten (Trimesterdaten) für die Einnahmen der Bundesrepublik aus dem Reiseverkehr mit dem Ausland (Merkmal X, gemessen in Mio. DM):

	Jahr →	1986	1987	1988	1989	1990
Trimester	I	3.537	3.445	3.653	4.260	4.487
	II	5.344	5.653	6.100	6.607	6.940
	III	4.615	4.773	5.136	5.412	5.834

a) Zeichnen Sie die Trimesterzahlen für die Einnahmen im Reiseverkehr!
b) Bestimmen Sie einen geeigneten gleitenden Durchschnitt!
c) Bestimmen Sie die saisonbereinigte Zeitreihe und zeichnen Sie sie in die Grafik aus Teilaufgabe a ein!
d) Bestimmen Sie die exponentiell geglätteten Werte 1.Ordnung mit $y_0^* = 3.500$ und $\alpha = 0{,}5$!
e) Bestimmen Sie die exponentiell geglätteten Werte 2.Ordnung mit $y_0^{**} = 3.450$ und $\alpha = 0{,}5$!
f) Stellen Sie die exponentiell geglätteten Werte beider Ordnungen in einer Grafik dar!
g) Prognostizieren Sie im dritten Trimester im Jahr 1990 die Einnahmen für das erste Trimester im Jahr 1991 unter Verwendung des Verfahrens der exponentiellen Glättung!

Aufgabe 4

Das Statistische Bundesamt veröffentlicht 1999 einen Bericht über die Entwicklung der Haushalte (in 10^3), die zu Hause Internet besitzen (fiktiv):

Zusatzaufgabensammlung

Aufgabe 1.

Das Unternehmen „Rhiel Logistik“ beschäftigt sich unter anderem mit dem Gütertransport. Dort setzt das Unternehmen immer noch auf die Straße, sprich auf Speditionsunternehmen. Um der Finanzbuchhaltung (Fibu) zu ermöglichen, die Kosten für die Speditionsunternehmen schon im voraus abschätzen zu können, stellt man aus alten Datenbeständen folgende Informationen zusammen, mit X: Anzahl der Paletten, die pro Tag an ein Speditionsunternehmen geliefert werden:

X	1	2	3	4	5	6	7	8	9	10
n(x)	2	5	10	15	35	30	20	25	8	2

a.) Bestimmen Sie die Werte H(5) und H(9) der empirischen Verteilungsfunktion des Merkmals X!

b.) Die Fibu teilt Ihnen nun mit, dass Ihr Unternehmen einen Rabatt ab einer Anzahl von 7 Paletten pro Tag erhält. Geben Sie mittels der Verteilungsfunktion an, in wie viel Prozent der Fälle dies eintritt!

Aufgabe 2.

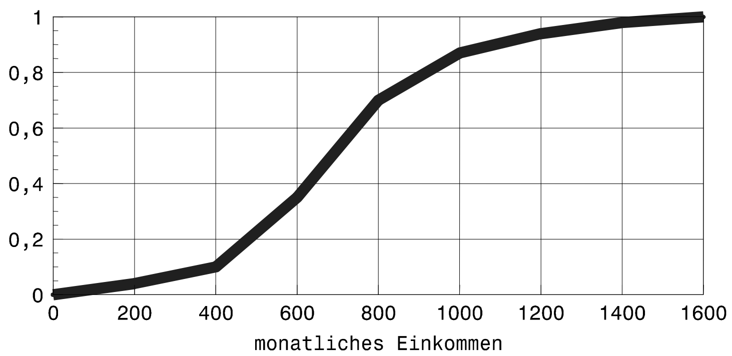
Der 18-jährige „G. Ibgas“ möchte sich unbedingt ein Auto kaufen. Da er über wenig Geld verfügt, aber seit neuesten über ein regelmäßiges Einkommen, überlegt er sich ein Auto zu „leasen“. Dabei liegen ihm folgende Angebote mit den Informationen PS (Merkmal X), auch Pferdestärken genannt, und der monatlichen Leasingrate in DM (Merkmal Y) vor:

Angebotsnr.	1	2	3	4	5	6	7
Merkmal X	56	61	89	101	47	64	71
Merkmal Y	350	385	420	495	355	375	400

a.) Bestimmen Sie ein geeignetes Zusammenhangsmaß für die beiden Merkmale!

b.) Berechnen und interpretieren Sie das in Teilaufgabe a gefundene Zusammenhangsmaß!

Aufgabe 3.
100 Studenten wurden anonym befragt, über welches Einkommen sie monatlich verfügen. Heraus kam folgende empirische Verteilungsfunktion:



a.) Wie viel Prozent der Studenten verfügen über ein monatliches Einkommen von mehr als 800 €?
b.) Wie viele Studenten liegen unter dem „studentischen Existenzminimum“ von 500 € ?
c.) Unter welcher (finanziellen) Grenze liegen 35% der Studenten, die am wenigsten Einkommen haben ?
d.) Bestimmen Sie den Median der Verteilung!

Aufgabe 4.

Aus 100 Beobachtungswerten ermittelte man folgende zweidimensionale Häufigkeitsverteilung:

	y_1	y_2	y_3
x_1	0,20		0,10
x_2		0,10	
x_3			0,10

Berechnen Sie die restlichen Werte unter Berücksichtigung der Informationen $h(x_2|y_3) = 0,5$, $n(y_1) = 40$, $h(y_2|x_3) = 0,25$ und der Aussage, dass die Kombination aus zweitem x-Wert und erstem y-Wert nicht vorkommt!