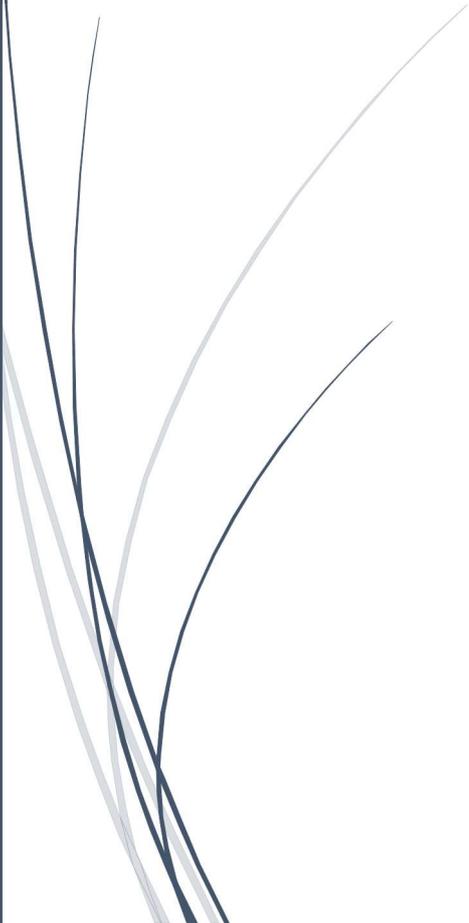


Herbst 2018

# Studieren in den Veranstaltungen zur Datenanalyse

Materialien zur Lernreflexion und  
Lernunterstützung



Manuela Pötschke

UNIVERSITÄT KASSEL, ARBEITSGEBIET ANGEWANDTE STATISTIK

## Vorwort

Im Laufe der letzten Jahre wurden die Lehrkonzepte in den grundlegenden Veranstaltungen zur Datenanalyse immer weiterentwickelt und neuen Medien- und Lerngewohnheiten angepasst. Dabei wird jede Innovation lerntheoretisch fundiert und begleitet. Die Erfahrungen aus diesen Konzepten und den durchgeführten Lehrveranstaltungen sind reichhaltig und zeigen, dass es sich lohnt, über akademisches Lernen und fachliche Auseinandersetzung immer wieder neu nachzudenken.

Gleichzeitig traten wiederkehrende Fragen auf, die einerseits einer gewissen Unsicherheit von Studienanfängern geschuldet sind. Andererseits müssen sie als Zeichen von fehlender Reflexionskompetenz über den Lernprozess und ein passendes akademisches Verhalten erfahrener Studierender verstanden werden. Dabei ist diese Beobachtung nicht vordergründig als Kritik an Studierenden, vor allem nicht an Studienanfängern, zu verstehen. Vielmehr beschreiben sie eine Bedingung des Lernprozesses, die auch in den Überlegungen der Lehrenden eine zu geringe Rolle spielt. Die vorliegenden Dokumente greifen basale, viel zu häufig nicht gestellte Fragen auf. Sie sollen einen Beitrag zur Selbstbefähigung der Studierenden leisten, sich mit empirischen Methoden auf vielfältige Art und Weise zu beschäftigen. Sich mit Methoden zu beschäftigen umfasst dabei

- eine kritische Auseinandersetzung mit der Methodennutzung anderer Forscher und
- die begründete Anwendung spezifischer Methoden für eigene Analysen sowie
- methodologische Überlegungen und die Untersuchung von Konsequenzen aus einer spezifischen Methodenverwendung (Methodenforschung).

Erste Fassungen dieser Dokumente wurden im Rahmen des von der Universität Kassel geförderten Projektes „Statistik lernen“ von Studierenden für Studierende entwickelt (vgl. [www.uni-kassel.de/go/statistik lernen](http://www.uni-kassel.de/go/statistik_lernen) ) erstellt. Für die Erstellung der vorliegenden Dokumentation wurden sie gründlich überarbeitet und angepasst. So hat sich das inhaltliche Angebot der Veranstaltungen mittlerweile erweitert, was sich auch in einem Bezeichnungswechsel von „Statistik“ hin zu „Methoden der Datenanalyse“ niederschlägt. Dabei sind die Veränderungen substantiell und nicht kosmetischer Natur. Die grundlegende Herangehensweise an die Auseinandersetzung mit Themen der Datenanalyse ist noch studierendenorientierter und bezieht neue Lerngewohnheiten systematisch und nachhaltig mit ein.

Die Dokumentation richtet sich in erster Linie an Studienanfänger, kann aber auch für „alte Hasen“ inspirierend sein. Hinweise und Kommentare werden gern entgegengenommen.

Dr. Manuela Pötschke  
Nora-Platiel-Str. 5, Raum 3119  
Tel.: 0561 804 3139  
E-Mail: [manuela.poetschke@uni-kassel.de](mailto:manuela.poetschke@uni-kassel.de)

## Inhaltsverzeichnis

FAQs an das Lehrgebiet „Angewandte Statistik“ .....	1
Fragen zur Veranstaltung.....	1
Fragen zum Selbststudium.....	2
Fragen zur Prüfung.....	1
Empirische Texte “richtig” lesen.....	2
Warum müssen wir immer so viel Lesen? .....	2
Wie sollen wir denn beginnen? .....	3
Wie kann der Text dann weiter (gedanklich) verarbeitet werden?.....	6
Übungsaufgaben leicht gelöst .....	8
Schlüssel zur Lösung .....	12
Handlungsanleitungen für das Lesen statistischer Formeln.....	14
Warum brauchen wir diese [...] Formeln? .....	14
Wie können wir Formeln lesen? .....	15
Gibt’s da mal ein Beispiel? .....	17

# FAQs an das Lehrgebiet „Angewandte Statistik“

## Fragen zur Veranstaltung

### **Wie sind die Grundlagenveranstaltungen zur Datenanalyse aufgebaut?**

Zur Konzeption und zum Aufbau der Veranstaltungen zur Datenanalyse I und II liegen in diesen Learning Materials zwei Konzeptpapiere vor, die sowohl den inhaltlichen Aufbau als auch das didaktische Vorgehen thematisieren. Kernelemente der Veranstaltungen sind Input- und Anwendungsphasen, die sich permanent abwechseln. Deshalb reicht es auch nicht, nur die Hälfte der vierstündigen Veranstaltung einzuplanen.

Ein wichtiges Element der Grundlagenausbildung sind die Tutorien, die in enger Abstimmung mit der Veranstaltung konzipiert sind.

### **Wo finde ich das Material und Informationen zu den Veranstaltungen?**

Auf der Plattform Moodle wird relevantes Material (z.B. die meisten Präsentationen zur Vorlesung) in semesterbezogenen Kursbereichen eingestellt. Die Benennung der Kurse folgt dabei der Logik „Jahr\_Datenanalyse I“ oder „Jahr\_Datenanalyse II“. Vor allem multimediale Hilfen finden sich auf der Homepage des Arbeitsgebietes ([www.uni-kassel/go/angst](http://www.uni-kassel/go/angst)). Organisatorische Informationen wie beispielsweise Termine für Nachklausuren und Zeiten und Räume usw. finden sich in Moodle und im Vorlesungsverzeichnis.

### **Wie kann ich die Tutorinnen und Tutoren kontaktieren?**

Die Tutorinnen und Tutoren sind häufig in der Veranstaltung anwesend und können dort angesprochen werden. Darüber hinaus können sie in ihren Tutorien erreicht werden. Der Tutorienplan ist in der Folienpräsentation zur ersten Veranstaltungssitzung enthalten.

Weitere Kontaktgelegenheiten besprechen die Tutorinnen und Tutoren mit ihren Tutanden individuell. Beachten Sie, dass die Tutorinnen und Tutoren nicht permanent verfügbar sein müssen und sich ihre Unterstützung vertragsgemäß auf die unmittelbaren Aktivitäten im Tutorium beschränkt.

### **Ich habe in einer Sitzung den roten Faden verloren – Was kann ich tun?**

Melden Sie sich sofort und bitten Sie um Wiederholung oder weitere Erläuterung.

Die Veranstaltung ist keine Vorlesung im herkömmlichen Sinne, dass niemand außer der Dozentin sprechen darf. Ganz im Gegenteil, wir sind bestrebt, auch in der großen Veranstaltung mit Ihnen über unsere Inhalte zu sprechen. Tragen Sie dazu bei! Ein Feedback über Ihre Gedanken beim Fortgang des Themas ist für die Dozentin darüber hinaus ein wichtiger Hinweis für die weitere didaktische Gestaltung der Veranstaltung.

Ist das Problem grundsätzlicher, helfen Gespräche mit Dozentin und Tutorinnen oder Tutoren.

### **Was mache ich, wenn ich in der Vorlesung bzw. in der Übungsphase überhaupt nicht mehr mitkomme?**

Bloß keine Panik! Stellen Sie sich zunächst die entscheidende Frage, wann genau der Verständnisabbruch passiert ist und melden Sie sich! Bitten Sie die Dozentin um eine Wiederholung des Sachverhalts oder darum, etwas langsamer fortzufahren. Suchen Sie auch das Gespräch mit den Kommilitoninnen und den Kommilitonen und mit den Tutorinnen und Tutoren.

### **Was mache ich, wenn ich die Aufgabenstellung nicht verstehe?**

Lesen Sie die Aufgabe genau durch. Formulieren Sie die Aufgabe in Ihren eigenen Worten. Überlegen Sie dazu, was Sie machen sollen (Ergebnis) und was Sie brauchen, um das Ergebnis zu erreichen (Weg der Bearbeitung). Achten Sie auf Schlüsselworte in der Aufgabenformulierung. Die meist verwendeten Schlüsselworte und ihre Bedeutung finden Sie in diesem Material im Abschnitt „Schlüssel zur Lösung“. Schreiben Sie die Aufgabe in eigenen Worten noch einmal auf.

Schlagen Sie Fachbegriffe oder Fremdwörter aus der Aufgabenstellung nach. Ordnen Sie die Aufgabe in das aktuell besprochene Thema ein.

Fragen Sie nach! Bei Kommilitoninnen, Kommilitonen, den Tutorinnen und Tutoren oder der Dozentin.

### **Ich traue mich nicht, direkt mit der Dozentin zu sprechen – Was kann ich tun?**

Zuerst einmal keine Panik! Es ist möglich, die Dozentin per Mail zu kontaktieren (vgl. nächste Frage). In die Sprechstunde können Sie sich von einem Kommilitonen begleiten lassen. Am schnellsten lassen sich manche Fragen aber im Anschluss an die Veranstaltung klären. Lassen Sie sich von einem guten Freund/ einer guten Freundin begleiten.

Bitte beachten Sie, dass die Tutorinnen und Tutoren nicht das Sprachrohr der Studierenden gegenüber der Dozentin sind. Sie geben aber Unterstützung in der Kommunikation, soweit das geht.

### **Was muss ich beachten, wenn ich die Dozentin per Mail kontaktiere?**

Es ist hilfreich, ein paar formale Anforderungen an die Mailkommunikation zu beachten. Aus (zum Teil bitteren) Erfahrungen sollen dazu einige Punkte kurz angesprochen werden.

- Die Dozentin möchte gern angesprochen, aber nicht mit „Liebste“ verabschiedet werden.
- Es ist hilfreich, das Anliegen in einem normalen Tonfall und deutlich zu formulieren.
- Wenden Sie sich nicht mit organisatorischen Fragen an die Dozentin, die Sie sich leicht mit dem Besuch der Veranstaltung oder mit einem Blick in Moodle bzw. das Vorlesungsverzeichnis selbst beantworten könnten (z.B. zum Klausurzeitpunkt, Ort, Inhalte der Klausur).
- Sie sollten auch beachten, dass die Dozentin nicht für Ihre Versäumnisse verantwortlich ist und deshalb auch die Konsequenzen daraus nicht trägt.

## Fragen zum Selbststudium

### **Welche lernunterstützenden Angebote kann ich nutzen?**

Die lernunterstützenden Angebote sind vielfältig. Sie sind entweder auf der Lernplattform Moodle zum jeweiligen Kurs oder allgemein auf der Homepage des Arbeitsgebietes ([www.uni-kassel.de/go/angst](http://www.uni-kassel.de/go/angst)) zu finden. Auf der Homepage sind unter dem Menüpunkt „von studies für studies“ beispielsweise vielfältige multimediale Lernbausteine enthalten. Darüber hinaus gibt es Literaturempfehlungen von Studierenden und Werbefilme zu Datenanalyse.

In Moodle sind themenbezogenen Handbuchbeiträgen angeführt, die sich sowohl zur Nach- als auch zur Vorbereitung einsetzen lassen. Ebenfalls in Moodle sind aktuelle Artikel aus Fachzeitschriften angeführt, die das abstrakte Verständnis beispielbezogen unterstützen. Zur Bearbeitung sind Leitfragen formuliert, die die Bearbeitung der Texte und das Verständnis des Inhaltes unterstützen sollen.

In der Veranstaltung werden interaktive Elemente eingesetzt, die ein Smartphone oder Tablet erfordern. Niemand muss diese Geräte einsetzen, eine Beteiligung ist aber nur mit ihrer Nutzung möglich.

Die Konzeption der Lehrveranstaltung insgesamt ist auf Lernbegleitung ausgerichtet. Es macht also Sinn, sich an den Übungsphasen aktiv zu beteiligen. Das lohnt sich vor allem deshalb, weil hier auch wenn es sich um eine Großveranstaltung handelt, unmittelbar individuelles Feedback gegeben wird.

Darüber hinaus hat sich die Teilnahme am Tutorium als hilfreich erwiesen. Empirische Ergebnisse aus Untersuchungen der letzten Jahre belegen, dass Tutoriumsteilnehmende erfolgreicher in der Klausur abschneiden als andere.

### **Wie kann ich die Inhalte selbstständig üben?**

Generell kann empfohlen werden, Lerngruppen zu bilden. In diesen Lerngruppen können neben dem Übungsangebot aus Veranstaltung und Tutorium weitere Anwendungsaufgaben analog generiert werden. Empirische Analysen haben gezeigt, dass Studierende in der Klausur besser abschneiden, wenn sie sich in Lerngruppen organisieren. Darüber hinaus zeigt sich, dass das Formulieren von Anwendungsaufgaben das fachliche Verständnis fördert.

Zu einigen Veranstaltungen gibt es auf Moodle Onlinetests mit Beispielaufgaben, die mehrfach gelöst werden können. Diese geben zum einen die Gelegenheit, sich auch durch Ausprobieren mit den Themen auseinanderzusetzen und zum anderen einen unmittelbaren Eindruck davon, wie der Lernstand ist.

Die Tutoriengestaltung ist flexibel. Studierende können Einfluss darauf nehmen, welche Aufgabenlösungen besprochen werden. Tutorinnen und Tutoren sowie die Dozentin geben individuelles Feedback zu selbst gestellten und gelösten Aufgaben (Aufgaben am besten mailen).

Darüber hinaus gibt es vielfältige Angebote zur Lernunterstützung mit Beispielaufgaben im Internet. Zu nennen sind beispielhaft der Methodenlehrebaukasten unter [www.methodenlehre-baukasten.de](http://www.methodenlehre-baukasten.de) (12 €) und die Lernprogramme LernStats unter [www.lernstats.de](http://www.lernstats.de). Darüber hinaus laden eine Vielzahl von YouTube-Tutorials zur Beschäftigung mit Themen der Datenanalyse ein.

### **Woher bekomme ich weiterführende bzw. grundlegende Literatur?**

Auf der Homepage finden sich Literaturempfehlungen ehemaliger Tutorinnen und Tutoren.

Außerdem kann ein Besuch in der Bereichsbibliothek 4 der Uni-Bibliothek am Holländischen Platz helfen. Hier sind Methodenliteraturen regalweise vor allem unter der Kennung '25 gew' vorgehalten. Die Begleittexte, die konkret in Moodle angeführt sind, sind in der Regel online über die Bibliothek verfügbar.

Informationen zu Datensätzen finden sich auf den Homepages der Forschungsdatenzentren. Hier finden sich zum Teil auch Analysen mit den Daten (vgl. z.B. [www.diw/soep.de](http://www.diw/soep.de) ; [www.gesis.org](http://www.gesis.org))

### **Was mache ich, wenn ich mehr lernen möchte?**

Wenn Sie sich über die Standard-Anforderungen hinaus mit Themen der empirischen Sozialforschung beschäftigen wollen, können Sie die weiterführenden Literaturempfehlungen auf der Homepage oder eine Recherche in der Bibliothek nutzen. Auch YouTube Tutorials erweitern den Horizont für Fragen zu Datenanalyse. Um ein individuelles weiterführendes Studienprogramm abzusprechen, wenden Sie sich bitte an die Dozentin.

## **Ich habe Angst vor Statistik – Was kann ich tun?**

Sind Themen oder Fächer mit negativen Gefühlen verbunden, kann man nicht gut lernen. Nur ein offener Geist kann sich unvoreingenommen mit Themen beschäftigen. Deshalb ist es wichtig, den Gründen für Ängste gegenüber dem Fach nachzugehen und einzuordnen. Sehr viele Studierende begründen ihre Ängste beispielsweise mit spezifischen Erfahrungen mit der Schulmathematik, mit Mathematiklehrern und vor allem mit der Wahrnehmung, nicht zu verstehen, wofür das Wissen gebraucht wird. Schauen wir uns die Gründe im Einzelnen an:

- Schulmathematik unterscheidet sich inhaltlich deutlich von den Inhalten im Studium der empirischen Sozialforschung. Das ist vor allem auf unterschiedliche Lernziele zurückzuführen. In der Grundausbildung zur Datenanalyse geht es darum, statistische Verfahren als Werkzeuge kennen zu lernen und anzuwenden, die uns helfen, Fragen über die Gesellschaft zu beantworten und Ideen, die wir über die Gesellschaft haben, zu überprüfen. Wissen über Datenanalyse ist also nie Selbstzweck, wir verwenden es als Instrument für inhaltliche Antworten.
- Die Dozentin engagiert sich besonders für Fragen des Lernens im Bereich der Datenanalyse. Dazu gehört auch eine intensive Auseinandersetzung mit Lerntheorien und didaktischen Empfehlungen zur Lernbegleitung.
- Unsere Inhalte sind zentrale Werkzeuge der empirischen Bearbeitung soziologischer und politikwissenschaftlicher Fragen. Das bedeutet, dass wir immer einen inhaltlichen Bezug herstellen (können und müssen), der sich aus Fachthemen unmittelbar ergibt. Auch die Beispieltexte haben ihren Schwerpunkt auf Anwendungen der Techniken zur Lösung inhaltlicher Fragen. Näher dran an der Praxis empirischer Forschung kann man nicht sein.

Wenn klarer ist, warum Sie Angst haben, wird es eventuell auch leichter zu erkennen, dass Sie es schaffen können, diese Ängste zu überwinden. Das gilt auch für die Aussage „Mathe konnte ich noch nie“ und die meist zugehörige Befürchtung, den Anforderungen nicht gerecht werden zu können. Vielleicht ist es besser, die gesamte Kraft darauf zu richten, das gegebene Angebot anzunehmen und sich neu auszuprobieren. Die Dozentin und die Tutorinnen und Tutoren sind bereit, Sie auf diesem Weg auch individuell zu unterstützen.

## **Was mache ich, wenn ich die Übungsaufgabe nicht lösen kann?**

Sie können Übungsaufgaben immer mit anderen besprechen. Dazu gehören Ihre Kommilitoninnen und Kommilitonen, die Tutorinnen und Tutoren und die Dozentin.

## **Was sind multimediale Szenarien?**

Multimediale Szenarien sind kleine Animationen, Filme, Comics und Graphiken, die Ihnen statistische Themen anschaulicher und beispielhafter verdeutlichen können. Sie sind auf der Homepage des Arbeitsgebietes im E-Learning-Angebot zu finden.

## **Welche sozialwissenschaftlichen Zeitschriften im deutschsprachigen Raum sind für Anwendungen zur Datenanalyse besonders wichtig?**

Die [KzfSS](#) → Kölner Zeitschrift für Soziologie und Sozialpsychologie

Die [MDA](#) → Methoden, Daten, Analysen

Die [Zpol](#) → Zeitschrift für Politikwissenschaft

Die [ZFS](#) → Zeitschrift für Soziologie

Die [PVS](#) → Politische Vierteljahrszeitschrift

## Fragen zur Prüfung

### **Was mache ich, wenn ich zur Prüfung nicht antreten kann?**

Bis zum Abmeldeschluss ist es jederzeit möglich, die Klausuranmeldung ohne Konsequenzen rückgängig zu machen. Der Abmeldeschluss wird in der Veranstaltung bekannt gegeben und findet sich sowohl in Moodle als auch im Vorlesungsverzeichnis.

Wer zum Klausurtermin krank ist, gibt das Attest bitte im Prüfungsamt ab. Die Kolleginnen und Kollegen dort tragen die Krankmeldung ein und dadurch wird die Anmeldung ebenfalls rückgängig gemacht. Wenn kein Krankenschein abgegeben wird, gilt die verpasste Klausur als Fehlversuch.

In besonderen persönlichen Situationen werden in Ausnahmen auch Einzelfallentscheidungen gefällt. Dazu kontaktieren Sie Manuela Pötschke ([manuela.poetschke@uni-kassel.de](mailto:manuela.poetschke@uni-kassel.de)) zeitnah zum Eintritt der Situation.

### **Ich habe Angst vor der Prüfung – Was kann ich tun?**

Sprechen Sie mit der Dozentin und den Tutorinnen und Tutoren! Häufig lassen sich Prüfungsängste durch individuelle Vorbereitungspläne und positives Feedback auf Vorbereitungsaufgaben mildern. Auch Entspannungsübungen und sportlicher Ausgleich kann Prüfungsängste bearbeiten lassen.

Für eine professionelle Bearbeitung von Prüfungsängsten ist die [Psychosoziale Beratungsstelle](#) beim Studentenwerk der Universität Kassel eine gute Anlaufstelle. Dieses Angebot sollten Sie auf jeden Fall nutzen, wenn Prüfungsängste fächerübergreifend und so gravierend auftreten, dass Sie Ihr Leistungsspektrum gar nicht abrufen können.

## Empirische Texte “richtig” lesen

### Warum müssen wir immer so viel Lesen?

Eine Klage, die von Studierenden in den Sozialwissenschaften häufiger zu hören ist. Das verwundert deshalb etwas, weil „der Text“ ein zentrales Medium der wissenschaftlichen Diskussion und Auseinandersetzung darstellt. Für Lehrende ist das so selbstverständlich, dass sie manchmal zu wenig mit Studierenden darüber reden. Das soll hier anders ein.

### **Sozialwissenschaftler kommunizieren miteinander (auch) über Texte.<sup>1</sup>**

Wenn wir also wissen wollen, **was** unsere fachlichen Kolleginnen und Kollegen über die Welt herausgefunden haben und **wie** sie das gemacht haben, können wir das z.B. in Texten nachlesen. Wir können die Funktionen, die Fachtexte für uns erfüllen, weiter differenzieren. Vor allem in einer sozialwissenschaftlichen Ausbildung geht es darum,

- dass Texte uns etwas Fachliches mitteilen,
- dass wir etwas über die Welt oder die Gesellschaft oder Teilaspekte lernen,
- dass Texte uns über empirische Ergebnisse informieren,
- dass wir besser verstehen, wie empirische Forschung durchgeführt werden kann,
- dass wir zu eigenen Überlegungen angeregt und inspiriert werden,
- dass wir das Privileg spüren, als Sozialwissenschaftler die Welt zu erkennen.

Vor dem Glücksgefühl, die Überlegungen anderer zu einem Thema nachvollziehen und eigene Überlegungen beitragen zu können, steht der (manchmal steinige) Weg des Textverstehens. Hierzu stehen uns einige Strategien als Unterstützung zur Verfügung. Welche Vorgehensweise die beste ist, hängt von

- Art und Inhalt des Textes,
- den persönlichen Vorkenntnissen, Vorlieben und Stärken und
- vor allem dem Ziel ab, mit dem wir den Text lesen.

Über diese Aspekte sollten Sie sich klar werden, bevor Sie beginnen, den Text durcharbeiten. Dieser Aspekt der Selbstreflexion spielt in vielen studienbezogenen Situationen eine zu geringe Rolle. Nur wenn Sie sich darüber klar sind, was Sie schon können, was Sie gerade lernen wollen und wie Sie persönlich das am besten machen, kann ein Studienprozess auch gelingen. Wenn Sie beispielsweise nichts über den Forschungsprozess wissen und darüber, warum Hypothesen aus Theorien abgeleitet werden, wird es für Sie schwer sein zu verstehen, warum andere Forscherinnen und Forscher gerade die vorgestellten Theorien ausgewählt haben und welche Konsequenzen das hat. Wenn Sie nicht wissen, welche Bedeutung ein Median hat, dann werden Sie im deskriptiven Teil eines empirischen Beitrages nicht verstehen, welche Besonderheiten die Verteilung der interessierenden Variable aufweist. Dann wäre es sicher besser, erst einmal in ein Lehrbuch oder die Unterlagen zu Datenanalyse I zu schauen.

---

<sup>1</sup> Andere verbreitete Kommunikationsformen sind Vorträge, Podiumsgespräche oder Poster. Deshalb wird auf die Entwicklung entsprechender Kompetenzen im Verlauf des Studiums insgesamt auch besonderer Wert gelegt.

## Wie sollen wir denn beginnen?

Helfen kann in jedem Fall, die Struktur eines Textes zu durchschauen. Diese Struktur kann mit zwei Perspektiven untersucht werden: eine formal-funktionale und eine inhaltliche Struktur.

Empirische Texte folgen in den meisten Fällen einer klassischen formalen Struktur. Auf die

- Zusammenfassung und die
- Einleitung folgen der
- Forschungsstand zur Fragestellung, die
- Präsentation methodischer Überlegungen, die
- empirischen Ergebnisse und eine
- Zusammenfassung oder ein
- Fazit.

Der Grund für diese relative Übereinstimmung der grundlegenden formalen Struktur empirischer Texte liegt darin, dass die zu unterscheidenden Abschnitte spezifische Funktionen für den Gesamttext und das Leseverständnis erfüllen. Wird ein Abschnitt vernachlässigt oder vergessen, können auch die oben beschriebenen grundsätzliche Ziele des Lesens und Verstehens von Fachtexten nicht erfüllt werden.

Die Abschnitte und ihre spezifischen Funktionen sind in der folgenden Tabelle aufgeführt.

Tabelle: Formale und inhaltliche Struktur empirischer Texte

<b>Abschnitt</b>	<b>formale Funktion</b>	<b>inhaltliche Funktion</b>
Zusammenfassung	<ul style="list-style-type: none"> <li>• schnellen Eindruck über Gesamtbeitrag ermöglichen</li> <li>• Lesende entscheiden auf dieser Basis, ob sie den gesamten Text lesen wollen</li> </ul>	<ul style="list-style-type: none"> <li>• Forschungsfrage, Methode und Ergebnis „auf einen Blick“ zu erfassen</li> </ul>
Abstract	<ul style="list-style-type: none"> <li>• Reichweite der Forschung erhöhen: englischsprechende Menschen für den Text interessieren</li> </ul>	
Einleitung	<ul style="list-style-type: none"> <li>• Lesende für das Thema interessieren</li> <li>• Inhaltliches Problem deutlich machen</li> <li>• Beschreibung des „roten Fadens der Argumentation“ (Gliederung des Beitrages)</li> </ul>	<ul style="list-style-type: none"> <li>• Präsentation der Forschungsfrage</li> <li>• Begründung der sozialwissenschaftlichen Relevanz der Forschungsfrage</li> </ul>
Forschungsstand	<ul style="list-style-type: none"> <li>• Darstellung des Forschungsstandes insgesamt</li> <li>• Verdeutlichung der Forschungslücke</li> </ul>	<ul style="list-style-type: none"> <li>• Basis bestehend aus Begriffen und theoretischen Aussagen vorstellen</li> <li>• Überlegungen zu plausiblen (Wirk-)Mechanismen ausführen</li> <li>• Bisherige empirische Befunde darstellen</li> <li>• Forschungslücken könnten in der notwendigen Testung theoretischer Annahmen oder in fehlenden empirischen Analysen oder in der Anwendung neuer Modellspezifikationen bestehen</li> </ul>
Hypothesen	<ul style="list-style-type: none"> <li>• Scharnier zwischen Theorie und Empirie bilden</li> </ul>	<ul style="list-style-type: none"> <li>• Hypothesen können in einem eigenständigen Abschnitt oder in das Kapitel zum Forschungsstand integriert präsentiert werden</li> <li>• Hypothesen sind Aussagen darüber, wie die vorgestellten Theorien die Forschungsfrage beantworten</li> <li>• Hypothesen können sich aber anstelle auf Theorien auch auf empirische Befunde anderer Forschenden stützen</li> </ul>

Methoden	<ul style="list-style-type: none"> <li>• Transparenz des empirischen Vorgehens</li> <li>• Voraussetzung für Replikation schaffen</li> </ul>	<ul style="list-style-type: none"> <li>• Einschätzung der Datengüte durch Beurteilung von: <ul style="list-style-type: none"> <li>○ Spezifischen Erhebungsformen</li> <li>○ Wirkungen von Auftraggebern</li> <li>○ Stichprobenziehung</li> <li>○ Fallzahl</li> </ul> </li> <li>• Beurteilung der Operationalisierung und der Angemessenheit der verwendeten Indikatoren ermöglichen</li> <li>• Skalierung bekannt machen, um Analysepotential und Reichweite der Analyse beurteilen zu können</li> <li>• Angemessenheit, Potentiale und Grenzen der Analysemethode zeigen</li> <li>• Begründung für die Auswahl bestimmter Maßzahlen</li> </ul>
Ergebnisse	<ul style="list-style-type: none"> <li>• Präsentation der Ergebnisse der eigenen empirischen Forschung</li> </ul>	<ul style="list-style-type: none"> <li>• Deskription von relevanten Verteilungen</li> <li>• Präsentation der Schätzergebnisse aus Erklärungsmodelle</li> <li>• Interpretation der Schätzergebnisse formulieren</li> <li>• umfassende Beantwortung der Forschungsfrage</li> </ul>
Zusammenfassung	<ul style="list-style-type: none"> <li>• Verdichtung der Antworten auf die Forschungsfrage</li> <li>• Ausblick auf weitere Forschungen</li> </ul>	<ul style="list-style-type: none"> <li>• Ergebnisse „auf einen Blick“ darstellen</li> <li>• Deutlich machen, welche Hypothesen bestätigt oder nicht bestätigt wurden</li> <li>• Offene Fragen nennen</li> </ul>
Fazit	<ul style="list-style-type: none"> <li>• Empfehlungen geben</li> </ul>	<ul style="list-style-type: none"> <li>• Konsequenzen aus den Ergebnissen ableiten und begründen</li> </ul>

Aus den zwei Perspektiven auf die Struktur eines Textes können bereits die ersten beiden Strategien zur Bearbeitung unbekannter Texte abgeleitet werden.

### **1. Fragen zur formalen Struktur beantworten**

- In welche Abschnitte ist der Text gegliedert?
- Welche abstrakten Funktionen übernehmen diese Abschnitte?
- Welche Funktion für das Textverständnis wird nicht oder unzureichend erfüllt?

### **2. Fragen zur inhaltlichen Struktur beantworten**

- In welchem Forschungsfeld ist der Text verortet?
- Was soll untersucht werden? Wie heißt die Forschungsfrage?
- Welche Theorie wird herangezogen?
- Welche Argumente stehen sich gegenüber?
- Welche Hypothesen werden formuliert?
- Welche Operationalisierung wird für abstrakte Begriffe gefunden?
- Welche Daten werden verwendet?
- Welche Ergebnisse werden berichtet?
- Welche Schlüsse werden aus den empirischen Ergebnissen gezogen?

Wie kann der Text dann weiter (gedanklich) verarbeitet werden?

Wenn durch diese Fragen der Text inhaltlich erschlossen ist, kann eine kritische Würdigung erfolgen. Auch hier hilft die Fragetechnik weiter. Dabei sind nicht immer alle Fragen relevant, aber sie geben so etwas wie ein Spektrum an reflektierenden Überlegungen wieder, die auf jeden Text anzuwenden sind.

### **3. Fragen zur kritischen Würdigung beantworten**

- Wird die Forschungsfrage verständlich und präzise formuliert?
- Beinhalten die vorgestellten Theorien theoretische Antworten auf die Forschungsfrage?
- Lassen sich die Hypothesen nachvollziehbar aus den theoretischen Überlegungen oder anderen empirischen Befunden ableiten?
- Ist die Operationalisierung gelungen? Gibt es eine Überprüfung der Validität von Messinstrumenten?
- Welche Einschränkungen bringen die spezifischen Operationalisierungen mit sich?
- Sind die Daten geeignet, um der Forschungsfrage nachzugehen?
- Beziehen sich die empirischen Ergebnisse auf die Hypothesen?
- Sind die Interpretationen richtig und nachvollziehbar präsentiert?
- Sind die Schlussfolgerungen ausreichend und nachvollziehbar?

#### **4. Selbst inhaltliche Fragen stellen**

Um einen generellen Eindruck über den Text zusammenzufassen, kann man gut mit eigenen Fragen an den Text herantreten. Diese können dann durchaus auch über den Text hinausgehen und bereits eigene Argumente mit einbeziehen. Zum Beispiel kann man sich fragen:

- Was habe ich über die Forschungsfrage gelernt?
- Warum wurden gerade die genutzten Daten verwendet?
- Welche weiter gehenden Schlüsse können aus der Studie gezogen werden?

#### **5. Tiefenverständnis entwickeln**

Die vorangegangenen Techniken beziehen sich darauf, ein generelles Verständnis über den Text insgesamt zu entwickeln. Darüber hinaus sollten aber auch spezifische Passagen und einzelne Aussagen nach der Textbearbeitung klarer sein. Dabei haben Fachbegriffe und Fremdworte eine große Bedeutung. Sie können markiert und sollen dann zuerst übersetzt werden. Oft wird es deutlich einfacher einzelne Sätze und Argumente zu verstehen, wenn man als Lesende den übersetzten Text an Stelle des unbekanntes Fachbegriffs oder Fremdwortes setzt. Je öfter die entsprechende Umschreibung genutzt wird, umso vertrauter werden diese Begriffe und gehen dann auch schnell in den aktiven Wortschatz ein.

#### **6. Text aufbereiten**

Um die wesentlichen Argumente und Ergebnissen eines empirischen Textes und seine Bearbeitung auch für eine spätere Verwendung nutzbar zu machen, eignen sich Zusammenfassungen und Visualisierungen. Besonders die Übertragung des Textes in eine andere Darstellungsform wie ein Mindmap, einen Metaplan, eine Skizze, ein Schau- oder Pfeilbild, ein Strukturdiagramm oder eine Tabelle kann das aktuelle Verständnis erweitern und vorhalten (vgl. für ein Beispiel das Schaubild im Abschnitt „Übungsaufgaben leicht gelöst“).

Aber auch die klassische Strategie, den Text in eigenen Worten zusammenzufassen, ist eine gute Möglichkeit, auch später schnell wieder in den Text hineinzufinden, ohne ihn wiederholt aufbereiten zu müssen. Die Lesefragen helfen bei der Komprimierung zentraler Aussagen, Schlüsselwörter sollten in Zusammenfassungen unbedingt enthalten sein.

#### **7. Verständnis erweitern**

Wenn es nicht nur um den einen Text geht, sondern darum, sich ein ganzes Themenfeld zu erschließen, ist es hilfreich, auch andere Texte mit einzubeziehen. Beginnen kann man hier sehr gut mit Texten, auf die sich der Autor oder die Autorin „unseres“ Textes bezieht. Außerdem können Texte einbezogen werden, in denen wiederum „unser“ Text zitiert wird. Es kann helfen zu lesen, wie andere den Text verstanden haben. Durch dieses Vorgehen entsteht ein verwobenes Netz an Bezügen zwischen Einzelargumenten, die einen Überblick über ein Forschungsfeld geben können.

## Übungsaufgaben leicht gelöst

Übungsaufgaben spielen eine große Rolle in den Veranstaltungen zur Datenanalyse. Wir unterscheiden dabei verschiedene Schwierigkeitsstufen. Im einfachsten Fall geht es darum, gerade eingeführte Überlegungen wieder zu geben. Etwas komplizierter ist bereits die Anwendung auf ein sehr ähnliches Beispiel. Richtig gefordert ist man mit Aufgaben, die einen Transfer zwischen dem Bekannten und Unbekanntem erfordern. Für ein gutes Ergebnis ist es unerlässlich, genau zu verstehen, was Sie machen sollen und welche Anforderung mit der Übungsaufgabe verbunden sind. Dabei helfen die folgenden Schritte.

### 1. Schritt

Lese Sie sich die Aufgabenstellung vollständig und aufmerksam mehrmals durch.

### 2. Schritt

Überlegen Sie, worum es in der Aufgabe inhaltlich geht? Formulieren Sie die inhaltliche Aufgabe in eigenen Worten. Unterstreichen Sie unbekannte Begriffe und klären Sie sie, bevor Sie mit der Aufgabe beginnen.

### 3. Schritt

Überlegen Sie, was genau von Ihnen verlangt wird. Das Ergebnis Ihrer Aufgabenbearbeitung kann beispielsweise eine Aufzählung, oder eine geordnete Liste oder eine Zuordnung oder ein Fließtext oder ein Bild sein. Was genau Ihr Werk sein soll erkennen Sie an sogenannten Operatoren. Operatoren sind Verben, die signalisieren, was genau erwartet wird. Danach richtet sich dann auch die Bearbeitungsstrategie für die Aufgabe. Außerdem geben manche Operatoren einen Hinweis auf die Lösung selbst. Wenn in einer Aufgabe beispielsweise eine Häufigkeitstabelle gegeben und als Auftrag formuliert ist: „Geben Sie die Werte für ein geeignetes Lagemaß an.“, dann geht es hier nicht um eine Berechnung. Es kann sich also nur um ein Maß handeln, dessen Wert Sie direkt aus der Tabelle ablesen können (also Median oder Modus). Ist dagegen gefordert „Berechnen Sie ein angemessenes Streuungsmaß“, dann können Sie den Taschenrechner getrost einmal auspacken und sich an die Arbeitstabelle für die Varianz machen.

Die Operatoren signalisieren neben dem konkreten Ergebnis auch, welche Schwierigkeit die Aufgabe hat. Oft wird dabei eine gewisse Hierarchie unterstellt. Um komplexe und sehr schwierige Aufgaben zu lösen, gehen die Lösungen auf der Basis geringerer Operatoren mit ein. Eine genaue Beschreibung der Bedeutung vieler relevanter Operatoren findet sich im nachfolgenden Dokument „Schlüssel zur Lösung“.

#### Schwierigkeit



Quelle: Utheß, Sabine  
2008: Lesen fördern  
und trainieren 3.  
Stuttgart: Ernst Klett  
Verlag, S. 36

An einem Beispiel sollen die Konsequenzen unterschiedlicher Operatoren verdeutlicht werden. Gegeben ist in jedem Fall die folgende Datentabelle mit Informationen zu den 9 Klassen, die in Kassel/Wehlheiden in einer Sitzfußballliga spielen.

Klasse	erreichte Punkte	Schule	Häufigkeit des Trainings	Trainingszeit pro Woche in Stunden
1a	23	1	(1) Täglich	0,75
3b	4	1	(2) mehrmals in der Woche	2,25
1c	5	1	(1) Täglich	0,75
2b	3	2	(2) Mehrmals in der Woche	1,5
3b	4	2	(1) Täglich	2
1a	2	2	(3) Einmal pro Woche	1,2
1a	3	3	(2) Mehrmals in der Woche	1,8
2b	6	3	(2) Mehrmals in der Woche	1
3b	4	3	(3) Einmal pro Woche	1,5

In der folgenden Tabelle werden mögliche Aufgaben zu dieser Datentabelle beispielhaft vorgestellt und angegeben, was genau erwartet wird, wie schwierig die Aufgabe eingeschätzt wird und wie eine mögliche Musterlösung aussehen kann. Die entscheidenden Hinweise auf die Lösung und den Lösungsweg geben dabei die Operatoren.

### Aufgabenstellungen und Lösungen in Abhängigkeit von den verwendeten Operatoren

Aufgabe	Was wird erwartet?	Schwierigkeit	Musterlösung
Nennen Sie drei Lagemaße.	Liste	Gering	<ul style="list-style-type: none"> <li>• Modus</li> <li>• Median</li> <li>• Arithmetisches Mittel</li> </ul>
Nennen Sie für jedes Skalenniveau das informationshaltigste Lagemaß?	Liste mit Zuordnung	Etwas höher	<ul style="list-style-type: none"> <li>• nominal → Modus</li> <li>• Intervall/ Ratio → arithmetisches Mittel</li> <li>• ordinal → Median</li> </ul>
Nennen Sie für jede Variable das passende Lagemaß und begründen Sie Ihre Entscheidung.	Liste mit Zuordnung und Begründungssatz	Noch etwas höher	<ul style="list-style-type: none"> <li>• Klasse/ Schule → Modus, weil nominales Skalenniveau vorliegt</li> <li>• Erreichte Punkte → Median, weil zwar metrisches Skalenniveau vorliegt und deshalb das arithmetische Mittel möglich wäre, aber Median ist besser wegen des Ausreißers</li> <li>• Häufigkeit des Trainings → Median, weil ordinales Skalenniveau vorliegt</li> <li>• Trainingszeit → arithmetisches Mittel, weil Ratioskalenniveau vorliegt</li> </ul>
Sie wollen den Zusammenhang zwischen erreichten Punkten und Trainingszeit untersuchen. Welches Analyseverfahren verwenden Sie? Welche Maßzahl interpretieren Sie? Begründen Sie Ihre Entscheidung.	zwei Nennungen und einen Begründungssatz	Noch höher	<ul style="list-style-type: none"> <li>• Korrelationsanalyse</li> <li>• <math>r</math></li> <li>• Weil beide Variablen metrisches Skalenniveau aufweisen.</li> </ul>
Berechnen Sie den Zusammenhang zwischen erreichten Punkten und Trainingszeit.	Rechenergebnis	mittel (Formel gegeben)	<ul style="list-style-type: none"> <li>• <math>r=-0,51</math></li> </ul>

Aufgabe	Was wird erwartet?	Schwierigkeit	Musterlösung
<p>Der Zusammenhang zwischen den erreichten Punkten und der Trainingszeit beträgt <math>r=-0,51</math> (<math>p=0,16</math>).</p> <p>Interpretieren Sie den erhaltenen Wert für die Stichprobe und begründen Sie die Interpretation.</p>	<p>Zusammenhängender kurzer Text.</p>	<p>Mittel bis schwierig</p>	<ul style="list-style-type: none"> <li>• Es handelt sich hier um einen mittleren negativen Zusammenhang. Das bedeutet, dass überdurchschnittliche Werte in der Punktzahl systematisch mit einer unterdurchschnittlichen Trainingszeit einhergehen und umgekehrt. Es handelt sich um einen mittelstarken Zusammenhang, da der Wertebereich für <math>r</math> zwischen <math>-1</math> und <math>+1</math> liegt. Für eine negative Korrelation sind die Werte <math>0</math> (kein Zusammenhang) und <math>-1</math> (perfekter negativer Zusammenhang) relevant. <math>-0,51</math> liegt fast genau in der Mitte dieses Bereiches und drückt deshalb einen mittleren Zusammenhang aus.</li> </ul>
<p>In einer repräsentativen Studie mit insgesamt 150 Vereinen wurde eine Korrelation zwischen den Punkten und der Trainingszeit von <math>r=-0,12</math> (<math>p=0,004</math>) errechnet. Diskutieren Sie Gründe für diese abweichenden Ergebnisse und Konsequenzen daraus.</p>	<p>Längerer Text</p>	<p>Sehr komplex und schwierig</p>	<ul style="list-style-type: none"> <li>• Die Unterschiede sind darauf zurück zu führen, dass es sich bei den Ausgangsdaten um eine Vollerhebung einer geringen Anzahl von regional verorteten Akteuren handelt. In der zweiten Studie handelt es sich um eine Zufallsauswahl. Deshalb ist es auf der Basis dieser Ergebnisse möglich, auf die unbeobachtete Grundgesamtheit zu schließen. Aus dem <math>p</math>-Wert, der kleiner als die Irrtumswahrscheinlichkeit von fünf Prozent ist, können wir erkennen, dass auch ein der Grundgesamtheit ein Zusammenhang besteht. Die Nullhypothese (<math>Rho=0</math>) wurde verworfen. Im ersten Beispiel konnte die Signifikanz nicht sinnvoll interpretiert werden, weil keine Zufallsstichprobe gezogen wurde. Das Stichprobenergebnis entspricht hier dem Ergebnis der Grundgesamtheit. Offensichtlich ist der Zusammenhang in den Wehlheider Vereinen deutlich stärker ausgeprägt als insgesamt. Vielleicht liegt das daran, dass die Trainer in Kassel dazu neigen, Punktgewinne mit Trainingsausfall zu belohnen. Sie hatte letztes Jahr nämlich eine Motivationsschulung und haben dort gelernt, dass positive Bestärkung durch Lob und weniger Druck besser für die Leistungsentwicklung sind als andere Strategien. Um dieser These nachzugehen, müssen weitere empirische Studien unternommen werden. ...</li> </ul>

## Schlüssel zur Lösung

Hinter dem *Schlüssel zur Lösung* verbirgt sich eine Sammlung von Schlüsselvokabeln typischer Aufgabenstellungen. Im folgenden finden Sie eine Auflistung, Beschreibung und Erläuterung der sogenannten "Operatoren". Operatoren geben an, welche kognitive Leistung erwartet wird.

### **abschätzen**

durch begründete Überlegungen Größenordnungen angeben

### **analysieren/ untersuchen**

unter einer gegebenen Fragestellung wichtige Bestandteile oder Eigenschaften herausarbeiten (Untersuchen beinhaltet unter Umständen zusätzliche praktische Anteile.)

### **auswerten**

Daten, Einzelergebnisse oder sonstige Elemente in einen Zusammenhang stellen und ggf. zu einer Gesamtaussage zusammenführen

### **begründen**

Sachverhalte auf Regeln, Gesetzmäßigkeiten bzw. kausale Zusammenhänge zurückführen

### **berechnen**

den Wert einer statistischen Größe auf der Basis von Formeln ausrechnen

### **bestimmen**

statistische Werte aus Darstellungen (z.B. einer Tabelle) ablesen

### **beschreiben**

Strukturen, Sachverhalte oder Zusammenhänge strukturiert und fachsprachlich richtig mit eigenen Worten wiedergeben

### **bestätigen**

die Gültigkeit einer Hypothese, Modellvorstellung, Gesetzmäßigkeit durch eine Berechnung /ein Experiment verifizieren

### **beurteilen**

zu einem Sachverhalt ein selbstständiges Urteil unter Verwendung von Fachwissen und Fachmethoden formulieren und begründen

### **bewerten**

Sachverhalte, Gegenstände, Methoden, Ergebnisse etc. an Beurteilungskriterien oder Normen und Werten messen

### **darstellen**

Sachverhalte, Zusammenhänge, Methoden, Bezüge in angemessenen Kommunikationsformen strukturiert wiedergeben

### **diskutieren / erörtern**

in Zusammenhang mit Sachverhalten, Aussagen oder Thesen unterschiedliche Positionen bzw. Pro- und Kontra-Argumente einander gegenüberstellen und abwägen

### **dokumentieren**

alle notwendigen Erklärungen, Herleitungen und Skizzen darstellen

**entwerfen / planen**

zu einem vorgegebenen Problem einen Lösungsweg / Skizze erfinden

**entwickeln / aufstellen**

Sachverhalte und Methoden zielgerichtet miteinander verknüpfen (eine Idee, eine Hypothese, ein Modell oder eine Theorie schrittweise weiterführen und ausbauen)

**erklären**

einen Sachverhalt nachvollziehbar und verständlich machen

**erläutern**

einen Sachverhalt durch zusätzliche Informationen veranschaulichen und verständlich machen

**ermitteln**

einen Zusammenhang oder eine Lösung finden und das Ergebnis formulieren

**interpretieren / deuten**

Zusammenhänge in Hinblick auf Erklärungsmöglichkeiten untersuchen und abwägend herausstellen

**nennen / angeben**

Elemente, Sachverhalte, Begriffe, Daten oder Erläuterungen aufzählen

**skizzieren**

Sachverhalte, Strukturen und Ergebnisse auf das Wesentliche reduziert übersichtlich darstellen

**strukturieren / ordnen**

vorliegende Objekte kategorisieren und hierarchisieren

**überprüfen / testen / prüfen**

Sachverhalte oder Aussagen an Fakten, Wahrscheinlichkeiten oder innerer Logik messen und evtl. Widersprüche aufdecken

**vergleichen**

Gemeinsamkeiten, Ähnlichkeiten und Unterschiede zwischen mindestens zwei Sachverhalten beschreiben

## Handlungsanleitungen für das Lesen statistischer Formeln

Vielen Menschen ist die Formelschreibweise fremd und sie ignorieren sie gern. Das gilt auch für viele Studierende in den Sozialwissenschaften. Wir wollen hier zeigen, warum eine Formelschreibweise sehr hilfreich sein kann und welche Techniken Sie anwenden können, um Formeln besser zu verstehen. Denn auch hier gilt: die Übung macht den Meister und das Potential der Schreibweise wird mit zunehmendem Verständnis immer deutlicher.

### Warum brauchen wird diese [...] Formeln?

Statistische Formeln sind die **kürzeste Form**, um zu beschreiben, was sich hinter einem bestimmten statistischen Konzept verbirgt. In diesem Sinne sind sie eine besondere Form der Definition, welche jedoch nicht in Textform sondern in Zeichenform wiedergibt, was in einem bestimmten Wert drinsteckt. Wer schreibfaul ist, findet mit Formeln also einen guten Weg, trotzdem Sachverhalte darstellen zu können.

Schauen wir uns zur Illustration ein Beispiel an:

Formelschreibweise bedeutet Darstellung „mit Hilfe von Zeichen“	Verbale Schreibweise bedeutet Darstellung „mit Hilfe der Sprache“
$r = \frac{s_{xy}}{s_x \cdot s_y}$	Die Korrelation drückt das Verhältnis zwischen der Kovarianz zweier metrischer Variablen und dem Produkt ihrer Standardabweichungen aus.

Was sieht übersichtlicher und kürzer aus?

Nun gibt es immer wieder Studierende, die gegen die Formelschreibweise einwenden, dass sie die Zeichen nicht verstehen. Wenn Sie selbst dazu gehören, dann fragen Sie sich jetzt bitte, ob Sie wirklich das Kürzel  $s_{xy}$  nicht verstehen, oder ob Ihnen nicht vielmehr unklar ist, was eine Kovarianz ist. Wenn das Konzept der Kovarianz unklar ist, dann nützt Ihnen auch die verbale Umschreibung nicht. In beiden Fällen müssen Sie sich in Ihre Unterlagen aus Datenanalyse I oder in ein Lehrbuch vertiefen, um die Bedeutung von Kovarianz richtig zu verstehen.

Ein weiterer Aspekt ist der, dass einige Studierende mit **mathematischen Ausdrücken** wie Verhältnis, Differenz oder Summe nicht viel anfangen können. Der Bruchstrich, das Minus oder Plus dagegen werden eher erinnert, auch wenn nicht ganz klar ist, wie ihre Beschreibungen heißen. Die meisten Studierenden können daraus ableiten, dass sie teilen, subtrahieren oder addieren müssen.

Gilt es, selbst eine statistische Maßzahl zu berechnen, hat die Formel außerdem den großen Vorteil, als **Handlungsanleitung für die entsprechenden Rechenschritte** zu dienen. Das ist insbesondere bei komplexen Konzepten sehr hilfreich. Wer sich hier ein paar grundlegende Regeln zur Gewohnheit macht, kann Fehler leicht vermeiden. Und, um das deutlich zu sagen, Formeln zu verwenden, hat nichts mit (fehlendem) mathematischen Verstand zu tun. Das kann in der Tat jeder, so lange Sie bereit sind, sich auf Übersetzungsarbeit einzulassen. Wie das geht, steht im Mittelpunkt der folgenden Darstellung.

## Wie können wir Formeln lesen?

Wir verdeutlichen ein mögliches Vorgehen am Beispiel des Korrelationskoeffizienten  $r$ . Er gibt über die Art und die Stärke eines Zusammenhangs zwischen zwei metrischen Variablen Auskunft. In der vereinfachten Schreibweise haben wir die Formel oben bereits gesehen. Nun wollen wir die vereinfachte Schreibweise um eine ausführliche Schreibweise ergänzen, die die Konzepte Kovarianz und Standardabweichung ebenfalls in ihren Bestandteilen zeigt.

$$r = \frac{s_{xy}}{s_x \times s_y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

↑ vereinfachte Formel                      ↑ ausführliche Formel

**Fragen Sie sich** - Aus welchen Bestandteilen ist die Formel zusammengesetzt *oder anders*: Was steckt eigentlich alles drin in der Formel? Wollen wir die Formel verstehen, schauen wir in die ausführliche Darstellung.

### So geht's:

1. Es hilft sehr, sich zuerst einmal klar zu machen, **für was die Buchstaben in der Formel stehen**. Wir wissen, dass die Korrelation den Zusammenhang zwischen zwei metrischen Variablen angibt. Diese Variablen heißen X und Y (Wir könnten andere Namen für die Variablen verwenden, aber dann würden wir sie anders bezeichnen als alle anderen Sozialforscher. Warum sollten wir das tun?). Nun hat jede Person, die wir befragt haben, in den beiden Variablen X und Y bestimmte Werte (nämlich die Merkmalsausprägung für die Person). Wenn wir nun nicht die Variable insgesamt meinen, sondern die einzelne Merkmalsausprägung, verwenden wir die kleinen Buchstaben x und y.

Außerdem finden wir noch ein n vor. Dieser Buchstabe steht für den Stichprobenumfang, also die Anzahl aller Personen, die wir befragt haben.

Weiter finden wir einen griechischen Buchstaben vor, nämlich das große Sigma ( $\Sigma$ ). Dieser Buchstabe steht für eine spezifische Rechenoperation, wir kommen deshalb unter 3. darauf zurück.

2. An manchen Buchstaben stehen tiefgestellte „i“s. Solche Zeichen heißen Subskript. Ein Subskript ist eine tiefgestellte Nummerierung aller Personen. Das heißt,  $x_{12}$  steht für den Wert der Person Nr. 12 in der Variable X (und wieder wird deutlich, wie vorteilhaft die Formelschreibweise ist, denn die verbale Umschreibung ist deutlich länger). Meinen wir die konkreten Ausprägungen aller Personen vergeben wir als allgemeines Subskript i (i steht für „jede Person im Datensatz“).

Manche  $x$  und  $y$  tragen einen Balken. Dann stehen diese Zeichen nicht für einen (variablen) Wert, sondern für den Durchschnitt (arithmetisches Mittel) aus allen Beobachtungen in den Variablen  $X$  und  $Y$ .

3. Und dann gibt es noch Zeichen, die angeben, was gerechnet werden muss. Beginnen wir mit „-“. Minus heißt, wir sollen die Differenz aus zwei Werten berechnen, z.B. aus dem Wert für die Person 12 in der Variable  $X$  ( $x_{12}$ ) und dem arithmetischen Mittel über die Werte aller Personen in der Variable  $X$ . Wir sollen das für alle Personen machen. Für die Standardabweichung unterhalb des Bruchstriches quadrieren wir diesen Wert. Wenn wir für alle Beobachtungen je eine Differenz quadriert haben, summieren wir die erhaltenen Werte auf. Dafür steht das Sigma. Dann multiplizieren wir das Ergebnis mit  $1/n$  (wir können auch einfach durch  $n$  teilen) und ziehen die Wurzel. Schon ist die Standardabweichung von  $X$  fertig. Analog gehen wir für die Standardabweichung von  $Y$  vor. Damit haben wir die beiden Elemente unterhalb des Bruchstriches gebildet. Jetzt müssen wir beide Werte nur noch multiplizieren.

Schauen wir, was oberhalb des Bruchstrichs zu tun ist. Den ersten Schritt haben wir schon, nämlich die Differenzen zwischen den empirischen Beobachtungen und dem jeweiligen Mittelwert. Jetzt wird danach gleich das Produkt aus beiden Werten gebildet und zwar für jeden Fall einzeln. Erst danach wird wieder summiert (Sigma), und zwar vom ersten ( $i=1$ ) bis zum letzten Fall ( $n$ ). Abschließend teilen wir durch  $n$ , wodurch wieder ein durchschnittlicher Wert erzeugt wird.

Der letzte Schritt besteht nun darin, das Verhältnis aus den beiden berechneten Werten zu bilden. Wir erhalten einen Wert zwischen  $-1$  und  $+1$ . Wenn Ihr Ergebnis außerhalb liegt, müssen Sie sich Ihre Rechnung nochmal anschauen.

Oft sind Studierende unsicher, was sie in welcher Reihenfolge machen sollen. Dabei helfen drei grundlegende Regeln:

- Die Rechenschritte arbeiten die Formel von rechts nach links ab.
- Die Rechenschritte gehen von innerhalb der Klammern nach außen.
- Punktrechnung geht vor Strichrechnung (und ja, man muss das ab und zu mal wieder sagen).

## Gibt's da mal ein Beispiel?

Ja, hier ist es. Wir halten es klein, damit der Rechenaufwand überschaubar bleibt und betrachten für vier Personen ihr Einkommen und ihre Betriebszugehörigkeit. Wir fragen uns, ob sich Loyalität mit dem Betrieb lohnt und sich in höherem Einkommen niederschlägt.

**Unsere Hypothese lautet: Je länger eine Person im Betrieb ist desto mehr Einkommen hat sie auch und umgekehrt.**

Wir haben zwei metrische Variablen, das passende Maß zur Beschreibung des symmetrischen Zusammenhangs ist die Pearson-Korrelation  $r$ . Wir haben die Erwartung, dass wir einen positiven Wert für  $r$  erhalten, der nicht zu nah bei 0 liegt.

Hier sind die Ausgangsdaten:

Name *	i	Einkommen in € (X)	Betriebszugehörig- keit in Jahren (Y)
Andreas Müller	1	2000	25
Inge Seidel	2	1300	17
Sebastian Premm	3	1700	7
Stefan Kurz	4	1900	11

\* Zum Kennenlernen haben wir hier die Namen der Probanden noch stehen. In den nachfolgenden Berechnungen brauchen wir sie nicht mehr und werden darauf verzichten.

### 1. Schnell mal die arithmetischen Mittel ausrechnen

$$\bar{x} = \frac{2000+1300+1700+1900}{4} = 1725$$

$$\bar{y} = \frac{25+17+7+11}{4} = 15$$

### 2. Differenzen zwischen jeder Beobachtung und dem jeweiligen arithmetischen Mittel berechnen

Unsere Tabelle sieht so aus, nachdem wir die Bezeichnungen gekürzt, die erste Spalte gelöscht und die ersten beiden Berechnungsspalten angehängt haben:

i	X	(Y)	$x_i - \bar{x}$	$y_i - \bar{y}$
1	2000	25	2000-1725=275	25-15=10
2	1300	17	1300-1725=-425	17-15=2
3	1700	7	1700-1725=-25	7-15=-8
4	1900	11	1900-1725=175	11-15=-4

In der Formel sieht das so aus:

$$r_{xy} = \frac{\frac{275 \cdot 10 - 425 \cdot 2 + 25 \cdot 8 - 175 \cdot 4}{4}}{\sqrt{\frac{275^2 - 425^2 - 25^2 + 175^2}{4}} \cdot \sqrt{\frac{10^2 + 2^2 - 8^2 - 4^2}{4}}}$$

**3. Produkte aus den Differenzen zu beiden Variablen (oberhalb des Bruchstrichs) und quadrierte Werte (unterhalb des Bruchstrichs) bilden**

i	X	(Y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	2000	25	2000-1725=275	25-15=10	275 <sup>2</sup> =75625	10 <sup>2</sup> =100	275·10=2750
2	1300	17	1300-1725=-425	17-15=2	-425 <sup>2</sup> =180625	2 <sup>2</sup> =4	-425·2=-850
3	1700	7	1700-1725=-25	7-15=-8	-25 <sup>2</sup> =625	-8 <sup>2</sup> =64	-25·-8=200
4	1900	11	1900-1725=175	11-15=-4	175 <sup>2</sup> =30625	-4 <sup>2</sup> =16	175·-4=-700

In der Formel sieht das so aus:

$$r_{xy} = \frac{\frac{2750-850+200-700}{4}}{\sqrt{\frac{75625+180625+625+3062}{4}} \cdot \sqrt{\frac{100+4+64+16}{4}}}$$

Kleine Tipps zur Kontrolle:

- Klammern werden zuerst berechnet.
- Quadrierte Zahlen sind immer positiv.
- Punktrechnung geht vor Strichrechnung, deshalb brauchen wir in der allgemeinen Formel die Klammern für die Kovarianz, denn dort sollen ja gerade zuerst die Differenzen berechnet werden.

**4. Aufsummieren, wo es Sinn macht**

i	X	(Y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	2000	25	2000-1725=275	25-15=10	275 <sup>2</sup> =75625	10 <sup>2</sup> =100	275·10=2750
2	1300	17	1300-1725=-425	17-15=2	-425 <sup>2</sup> =180625	2 <sup>2</sup> =4	-425·2=-850
3	1700	7	1700-1725=-25	7-15=-8	-25 <sup>2</sup> =625	-8 <sup>2</sup> =64	-25·-8=200
4	1900	11	1900-1725=175	11-15=-4	175 <sup>2</sup> =30625	-4 <sup>2</sup> =16	175·-4=-700
					$\Sigma=287500$	$\Sigma=184$	$\Sigma=1400$

In der Formel sieht das so aus:

$$r_{xy} = \frac{\frac{1400}{4}}{\sqrt{\frac{287500}{4}} \cdot \sqrt{\frac{184}{4}}}$$

**5. Durch n (hier 4) teilen, wo wir es brauchen**

i	X	(Y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	2000	25	2000-1725=275	25-15=10	275 <sup>2</sup> =75625	10 <sup>2</sup> =100	275·10=2750
2	1300	17	1300-1725=-425	17-15=2	-425 <sup>2</sup> =180625	2 <sup>2</sup> =4	-425·2=-850
3	1700	7	1700-1725=-25	7-15=-8	-25 <sup>2</sup> =625	-8 <sup>2</sup> =64	-25·-8=200
4	1900	11	1900-1725=175	11-15=-4	175 <sup>2</sup> =30625	-4 <sup>2</sup> =16	175·-4=-700
					Σ=287500	Σ=184	Σ=1400
					287500/4=71875	184/4=46	1400/4=350

In der Formel sieht das so aus:

$$r_{xy} = \frac{350}{\sqrt{71875} \cdot \sqrt{46}}$$

**6. Jetzt noch schnell die Wurzel ziehen**

i	X	(Y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	2000	25	2000-1725=275	25-15=10	275 <sup>2</sup> =75625	10 <sup>2</sup> =100	275·10=2750
2	1300	17	1300-1725=-425	17-15=2	-425 <sup>2</sup> =180625	2 <sup>2</sup> =4	-425·2=-850
3	1700	7	1700-1725=-25	7-15=-8	-25 <sup>2</sup> =625	-8 <sup>2</sup> =64	-25·-8=200
4	1900	11	1900-1725=175	11-15=-4	175 <sup>2</sup> =30625	-4 <sup>2</sup> =16	175·-4=-700
					Σ=287500	Σ=184	Σ=1400
					287500/4=71875	184/4=46	1400/4=350
					√=268,1	√=6,78	

In der Formel sieht das so aus:

$$r_{xy} = \frac{350}{268,1 \cdot 6,78}$$

**7. Der Rest ist kinderleicht und erinnert uns an die erste Bruchrechnung in der 5. Klasse (!!!)**

Das Ergebnis sieht dann so aus:

$$r_{xy} = 0,19$$

Und das heißt: Wie erwartet besteht zwischen Einkommen und Betriebszugehörigkeit ein positiver Zusammenhang, denn das gefundene Vorzeichen ist positiv. Außerdem ist der Zusammenhang mäßig ausgeprägt, denn der Wert unterscheidet sich zwar deutlich von 0 (kein Zusammenhang), ist ihr aber auch näher als der +1 (perfekter Zusammenhang).

Es ist vollbracht. Das ging doch leicht. Wenn man sich einfach merkt:

Durchschnitt und dann Differenz,  
das ist schnell gemacht.  
Unten hoch zwei und oben mal,  
das hat uns weitergebracht.  
Aufsummieren, wo es passt  
und dann schnell durch n geteilt,  
unten noch die Wurzel ziehen  
schon ist's fast geschafft.  
Ach ne, da ist der Bruch noch da,  
den haben wir damals schon gehasst.  
Doch heute ist's ein Kinderspiel:  
erst mal dann durch ...  
wir sind am Ziel.

Natürlich machen wir das alles in der Forschungspraxis nicht selbst, sondern lassen von Computern und Software rechnen. Die Algorithmen machen aber nichts anderes als wir hier gerade. Und es ist nicht übel zu verstehen, was Maschinen uns ausrechnen.

Und außerdem konnten wir zeigen, dass in so einer kleinen Formel doch eine ganze Menge an Informationen stecken. Das wollen wir nicht immer in einem Text ausführen. Also:

**Formeln nicht mehr überlesen.**

**Formeln selbst schreiben erleichtert uns das Leben und spart Ressourcen.**