

Springer-Lehrbuch

Einführung in die Statistik mit EXCEL und SPSS

Ein anwendungsorientiertes Lehr- und Arbeitsbuch

Bearbeitet von
Christine Duller

3., überarbeitete Auflage 2013. Buch. XII, 283 S. Kartoniert

ISBN 978 3 642 37858 4

Format (B x L): 16,8 x 24 cm

Gewicht: 499 g

[Wirtschaft > Wirtschaftswissenschaftliche Nachbardisziplinen > Statistik, Mathematik, Ökonometrie](#)

Zu [Inhaltsverzeichnis](#)

schnell und portofrei erhältlich bei


DIE FACHBUCHHANDLUNG

Die Online-Fachbuchhandlung beck-shop.de ist spezialisiert auf Fachbücher, insbesondere Recht, Steuern und Wirtschaft. Im Sortiment finden Sie alle Medien (Bücher, Zeitschriften, CDs, eBooks, etc.) aller Verlage. Ergänzt wird das Programm durch Services wie Neuerscheinungsdienst oder Zusammenstellungen von Büchern zu Sonderpreisen. Der Shop führt mehr als 8 Millionen Produkte.

Ablauf einer statistischen Analyse

Dieses Kapitel skizziert die Schritte, die vor bzw. nach der eigentlichen statistischen Auswertung notwendig sind. NewcomerInnen in der Statistik kennen zwar die Methoden zur Datenauswertung, in der Praxis liegen aber die Probleme oft in den Phasen vor bzw. nach der statistischen Auswertung.

2.1 Planung

Die Planung ist ein sehr wichtiger Teil einer statistischen Erhebung. Erst eine gute und durchdachte Planung ermöglicht den Erfolg.

In einem ersten Schritt sollte das meist mehr oder weniger diffus vorliegende Forschungsziel konkretisiert werden. Dazu gehört die möglichst exakte Zielformulierung und die Abgrenzung der Grundgesamtheit (vgl. Kapitel 1.3), weiters sind die zu erhebenden Merkmale festzulegen. Der nächste Schritt liegt in der Wahl der Methode zur Datengewinnung und die Datengewinnung selbst. Nach einer Probeerhebung (Pretest) werden notwendige Verbesserungen beim Erhebungsinstrument und bei den Merkmalen durchgeführt, dann kann mit der eigentlichen Datenerhebung begonnen werden. Danach erfolgt die Auswertung der Daten mit den Methoden der Statistik. Den Abschluss bildet ein schriftlicher Bericht. Die einzelnen Schritte sollten im Vorfeld terminiert werden und alle Arbeiten sind ausreichend zu dokumentieren.

Praxistipp:

Bei der Erstellung des Zeitplanes sollte die veranschlagte Zeit immer großzügig bemessen sein. Dokumentieren ist eine lästige Arbeit, aber in der Praxis unumgänglich. Eine unzureichende Planung und Vorbereitung kann später nicht mehr korrigiert werden.

Ablauf einer statistischen Analyse

1. Vorbereitung
 - Zielformulierung
 - Zeitplan erstellen
 - Festlegen von Grundgesamtheit und Merkmalen
 - Methode der Datengewinnung festlegen
 - Auswertungsmethode festlegen
 - Probeerhebung und Durchführung von Korrekturen
2. Datengewinnung
3. Datenerfassung und -aufbereitung
4. Statistische Auswertung
5. Bericht

Eine ausreichende Dokumentation ist unumgänglich.

2.2 Merkmale und Merkmalstypen

Bereits in der Planungsphase muss man sich über Eigenschaften von Merkmalen im Klaren sein. Merkmale lassen sich im Wesentlichen nach zwei Prinzipien einteilen: Einerseits gibt es verschiedene Skalenniveaus, andererseits unterscheidet man diskrete und stetige Merkmale.

2.2.1 Skalenniveaus von Merkmalen

Hinsichtlich des Skalenniveaus differenziert man metrische, ordinale und nominale Merkmale.

Ein Merkmal heißt **metrisch** (= quantitativ, kardinalskaliert), wenn seine Ausprägungen Vielfache einer Einheit sind (z.B. Länge, Einkommen). Die Ausprägungen sind voneinander verschieden, haben eine eindeutige Anordnung und einen eindeutig definierten Abstand. Bei metrischen Merkmalen kann man zwischen verhältnisskalierten und intervallskalierten Merkmalen unterscheiden.

Bei **verhältnisskalierten** Merkmalen gibt es einen natürlichen Nullpunkt (z.B. Preis) und das Verhältnis zweier Ausprägungen lässt sich sinnvoll interpretieren (Produkt A ist doppelt so teuer wie Produkt B).

Intervallskalierte Merkmale haben keinen natürlichen Nullpunkt, daher können auch Verhältnisse nicht sinnvoll interpretiert werden (z.B. Temperatur in Grad Celsius).

Ein Merkmal heißt **ordinal** (= Rangmerkmal), wenn die Ausprägungen nur in einer Ordnungsbeziehung wie größer, kleiner, besser oder schlechter zueinander stehen (z.B. Schulnoten, Güteklassen). Die Ausprägungen sind voneinander verschieden und haben eine eindeutige Anordnung, der Abstand zweier Merkmalsausprägungen ist hingegen nicht klar definiert und daher auch nicht interpretierbar. Es ist unumstritten, dass im österreichischen Schulnotensystem die Note 1 besser ist als die Note 2, der Abstand zwischen diesen beiden Noten lässt sich aber weder interpretieren noch mit dem Abstand zwischen den Noten 4 und 5 gleichsetzen.

Ein Merkmal heißt **nominal**, wenn seine Ausprägungen nicht in eindeutiger Weise geordnet werden können, sondern nur durch ihre Bezeichnungen unterschieden sind (z.B. Geschlecht, Familienstand, Beruf). Die Ausprägungen sind von einander verschieden, es gibt keine eindeutige Anordnung, der Abstand zweier Merkmalsausprägungen ist nicht definiert. Diese Merkmale werden auch als qualitative oder kategoriale Merkmale bezeichnet.

Diese Skalenniveaus sind hierarchisch aufgebaut, an unterster Stelle stehen die nominalen Merkmale, dann folgen die ordinalen Merkmale und an der Spitze stehen die metrischen Merkmale.

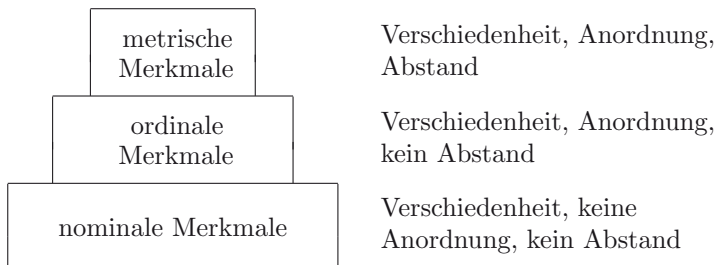


Abb. 2.1. Hierarchie der Skalenniveaus

Daraus ergeben sich zwei Konsequenzen:

- Jedes Merkmal aus einer höheren Hierarchiestufe kann durch Zusammenfassen und Umbenennen von Merkmalsausprägungen in ein Merkmal der niedrigeren Stufe umgewandelt werden, allerdings entsteht dadurch ein Informationsverlust. Das Merkmal Körpergröße gemessen in cm kann auch angegeben werden mit den Ausprägungen klein - mittel - groß, die Berechnung von Größenunterschieden ist aber nach dieser Transformation nicht mehr möglich.
- Alle Verfahren, die für ein Merkmal aus einer bestimmten Stufe zulässig sind, sind auch zulässig für Merkmale aus darüber liegenden Stufen.

Das Skalenniveau eines Merkmals bestimmt, welche Verfahren und Berechnungen im Umgang mit dem Merkmal gestattet sind. Tabelle 2.1 gibt einen ersten Überblick über sinnvolle und zulässige Verfahren für verschiedene Skalenniveaus.

Tabelle 2.1. Zulässige Verfahren für Skalenniveaus

Skalenniveau	Auszählen	Ordnen	Summen Differenzen	Quotienten
verhältnisskaliert	ja	ja	ja	ja
intervallskaliert	ja	ja	ja	nein
ordinal	ja	ja	nein	nein
nominal	ja	nein	nein	nein

2.2.2 Stetige und diskrete Merkmale

Ein Merkmal heißt **stetig**, wenn seine Ausprägungen beliebige Zahlenwerte aus einem Intervall annehmen können (z.B. Länge, Gewicht).

Ein Merkmal heißt **diskret**, wenn seine Ausprägungen bei geeigneter Skalierung (bzw. Kodierung) nur ganzzahlige Werte annehmen können (z.B. Fehlerzahlen, Schulnoten, Geschlecht). Diskrete Merkmale haben abzählbar viele Ausprägungen.

Dichotome Merkmale sind eine Sonderform von diskreten Merkmalen und besitzen nur zwei Ausprägungen.

Von **quasistetigen** Merkmalen spricht man bei Merkmalen, die aufgrund der Definition diskret sind, gleichzeitig aber über eine so feine Abstufung verfügen, dass man sie als stetige Merkmale behandeln kann. Insbesondere zählen hierzu alle monetären Merkmale (Preis, Kredithöhe, Miete, ...).

Die Bezeichnung **diskretisierte** Merkmale wird verwendet, wenn stetige Merkmale nur in diskreter Form erfasst werden, beispielsweise die Frage nach dem Alter in ganzen Jahren. Die Zusammenfassung von Ausprägungen eines Merkmals in Gruppen wird als **Gruppieren** bezeichnet.

Merkmale	
metrisch:	Abstand, Anordnung, Verschiedenheit der Ausprägungen intervall- oder verhältnisskaliert intervallskaliert: kein absoluter Nullpunkt verhältnisskaliert: absoluter Nullpunkt
ordinal:	kein Abstand, aber Anordnung und Verschiedenheit
nominal:	Ausprägungen sind Namen kein Abstand, keine Anordnung, nur Verschiedenheit
stetig:	Ausprägungen sind beliebige Werte aus einem Intervall
diskret:	abzählbar viele Ausprägungen
dichotom:	zwei Ausprägungen

2.3 Methode der Datengewinnung

Die Datengewinnung ist ein wesentlicher Teil der statistischen Analyse, der oft in den Überlegungen etwas zu kurz kommt. Dies ist aber ein fataler Fehler, denn wenn die Daten schlecht oder unzureichend erhoben wurden, so kann diese Unzulänglichkeit auch mit den Methoden der Statistik nicht behoben werden. Die Daten sind die Basis der Analyse und dessen sollte man sich immer bewusst sein.

Werden die Daten speziell für einen bestimmten Zweck erhoben, spricht man von primärstatistischen Erhebungen. Sekundärstatistische Erhebungen greifen hingegen auf bereits vorhandenes Zahlenmaterial zurück. Im technischnaturwissenschaftlichen Bereich werden Daten meist im Rahmen von Experimenten gewonnen, in den Sozialwissenschaften stammen die Daten im Normalfall aus einer Befragung oder einer Beobachtung. Eine Befragung wird meist mit einem Fragebogen durchgeführt, daher werden die wesentlichen Aspekte der Fragebogengestaltung kurz dargestellt.

Generell ist bei der Erstellung eines Fragebogens darauf zu achten, dass die Komplexität und die Wortwahl dem Zielpublikum angepasst sind. Der Fragebogen beginnt mit einfachen und harmlosen Fragen, z.B. nach dem Geschlecht oder dem Familienstand. Nach dieser Aufwärmphase kommen die Fragen zum Thema, die einen sinnvollen Aufbau haben sollten. Besonders heikle Fragen (z.B. nach dem Einkommen) sollten möglichst am Ende des Fragebogens platziert werden, um einen vorzeitigen Abbruch der Befragung zu vermeiden. Ein Fragebogen sollte so lang wie notwendig und so kurz wie möglich sein, ein ansprechendes Layout erhöht die Motivation der Befragten. Jeder Fragebogen endet höflicherweise mit einem Dank für die Mitarbeit.

Bei **geschlossenen Fragen** werden die möglichen Antwortalternativen im Fragebogen angeführt. Dabei ist darauf zu achten, dass sich einerseits die Alternativen nicht überschneiden (ausschließend), andererseits aber alle möglichen Alternativen angeführt werden (erschöpfend). Im Zweifel kann die Vollständigkeit immer mit der Kategorie *Sonstiges* erreicht werden. Ein freies Textfeld neben der Kategorie *Sonstiges* ermöglicht den Befragten eine nähere Spezifizierung.

Geschlossene Fragen werden manchmal dazu verwendet, besonders heikle Fragen zu entschärfen. Als Beispiel sei die Frage nach dem Einkommen erwähnt, die eher beantwortet wird, wenn die Befragten nur eine Einkommensklasse ankreuzen müssen anstatt das exakte Einkommen anzugeben. Bei der Erstellung der Antwortalternativen ist Fingerspitzengefühl notwendig. Die Einkommensklassen sollen einerseits möglichst klein sein, weil man an detaillierten Informationen interessiert ist, andererseits führen zu kleine Klassen wieder zu vermehrter Verweigerung. Auf jeden Fall sollten die Klassen möglichst so angelegt werden, dass niemand die niedrigste oder die höchste Kategorie auswählen muss, weil dies für die Befragten unangenehm wäre.

Eine besondere Stellung nehmen die **Bewertungsfragen** ein, bei denen die Befragten beispielsweise die Qualität von Produkten bewerten müssen. Hier ist es meist sinnvoll ein Bewertungsschema zu verwenden, das dem Schulnotensystem angepasst ist (1 = Sehr Gut bis 5 = Nicht Genügend). Das Bewertungsschema muss jedenfalls so aufgebaut sein, dass die Befragten in ihrer Meinungsäußerung nicht beeinflusst oder eingeschränkt werden.

Beispiel 2.1. Bewertungsfrage Mensa

Die Mensa an der Universität Linz stellte im Rahmen einer Befragung im Jahr 2003 folgende Frage: „Sagt Ihnen unser Suppenangebot zu?“ Als mögliche Antwort waren lediglich die Alternativen sehr gut, gut und befriedigend vorgesehen. Die Beeinflussung der Befragten durch diese Art der Antwortvorgabe ist offensichtlich

Eine fünfteilige Bewertungsskala ist in den meisten Fällen ein gutes Mittelmaß, zudem ist die Assoziation mit den Schulnoten für viele Befragte eine Erleichterung. Die Anzahl der Kategorien rechts und links von der Mittelkategorie muss jedenfalls gleich sein und auch die verbale Formulierung sollte möglichst symmetrisch sein. Bei einer ungeraden Anzahl von Kategorien kann die mittlere Kategorie von manchen Befragten als Fluchtkategorie verwendet werden, manchmal spiegelt aber die Mittelkategorie auch eine echte Einstellung wieder.

Bei **offenen Fragen** werden keine Antwortalternativen vorgegeben. Diese Fragestruktur sollte nur sehr eingeschränkt verwendet werden, da einerseits die Auswertung von offenen Fragen sehr aufwändig ist und andererseits ge-

geschlossene Fragen den Befragten weniger Mühe bereiten. Offene Fragestellungen sind zu bevorzugen, wenn man über den Forschungsgegenstand nicht gut genug Bescheid weiß, um alle Antwortalternativen einer geschlossenen Frage angeben zu können. In manchen Fällen wird bewusst eine offene Frage gestellt, weil deren Beantwortung andere Fähigkeiten von Befragten verlangt als die einer geschlossenen Frage. Fragt man nach Fakten, so müssen sich Befragte bei offenen Fragen erinnern, bei geschlossenen Fragen reicht das Wiedererkennen der richtigen Antwort. Offene Fragen sind auch dazu geeignet, um bei Mehrfachantworten die spontane Reihenfolge zu erheben, z.B. bei der Frage „Welche PolitikerInnen der Gegenwart sind Ihnen am sympathischsten?“. Auch für sehr einfache Fragen mit vielen Antwortmöglichkeiten kann eine offene Frage besser sein als eine geschlossene, z.B. die Frage nach dem Alter.

Sind **Mehrfachantworten** möglich, so sollte dies auch eindeutig aus der Fragestellung hervorgehen bzw. direkt nach der Frage vermerkt werden.

Die Erstellung eines guten Fragebogens erfordert Erfahrung und spezielles Wissen über die Fragebogengestaltung. Viele unterliegen dem Trugschluss, dass sie ohne Hilfe einen Fragebogen erstellen können, weil Fragen zu stellen eine einfache Sache ist. Gerade bei der erstmaligen Durchführung einer Befragung sollte man jedoch unbedingt auf professionelle Hilfe zurückgreifen, weil ansonsten die Gefahr zu groß ist, dass die erhobenen Daten unbrauchbar sind.

Datengewinnung	
Primärstatistik	Neue Daten werden z.B. mittels Befragung oder Experiment erhoben
Sekundärstatistik	Zurückgreifen auf bereits veröffentlichtes Datenmaterial
Fragetypen	
geschlossene Fragen	Antwortalternativen sind vorgegeben erschöpfende und ausschließende Alternativen
Bewertungsfragen	am besten dem Schulnotensystem entsprechend symmetrische Kategorien
offene Fragen	keine Antwortalternativen vorgegeben
Mehrfachantworten	Hinweis im Fragebogen

2.4 Datenerfassung und -aufbereitung

Die Informationen müssen aus den ausgefüllten Fragebögen entnommen und aufbereitet werden. Im Normalfall wird die Auswertung mit Hilfe eines Computers erfolgen, im einfachsten Fall unter Zuhilfenahme des Tabellenkalkulationsprogramms EXCEL. Wir wollen daher die Daten in einem ersten Schritt für EXCEL aufbereiten, die Grundprinzipien der Aufbereitung sind ohnehin für alle gängigen Programmpakete gleich.

Der Fragebogen in Abbildung 2.2 dient als Beispiel zur Datenerfassung und -aufbereitung.

Vor der Dateneingabe werden die Fragebögen durchnummeriert, danach wird der **Kodeplan** erstellt. Der **Kodeplan** ist eine Anleitung für die Übersetzung von Informationen aus dem Fragebogen in Zahlen. Die Informationen des Fragebogens werden in Tabellenform erfasst. Jede Zeile entspricht einem Fragebogen und damit einer Erhebungseinheit und jede Spalte der Tabelle enthält ein Merkmal, was aber nicht immer gleichzusetzen ist mit einer Frage.

Fragen, bei denen nur eine Antwort möglich ist lassen sich durch eine Variable (= ein Merkmal) darstellen. Metrische Merkmale werden als Zahlenwerte übertragen, wobei Maße immer in der gleichen Maßeinheit und mit der gleichen Anzahl an Dezimalstellen erfasst werden sollten. Ordinale Merkmale werden durch natürliche Zahlen kodiert, die Reihenfolge der Ordnung muss dabei erhalten werden. Üblich und auch sinnvoll ist die Verwendung von aufeinander folgenden natürlichen Zahlen, mit 1 beginnend. Nominale Merkmale werden ebenfalls mit natürlichen Zahlen kodiert. Welche Ausprägung welche Kodierung erhält ist unwesentlich, aber auch hier werden üblicherweise von 1 ansteigende ganze Zahlen verwendet.

Bei Fragen mit Mehrfachantworten muss jede Antwortalternative als eigenes dichotomes Merkmal mit den Ausprägungen „ausgewählt“ und „nicht ausgewählt“ kodiert werden (vgl. Tabelle 2.2 auf Seite 20).

Datenerfassung

- Informationen werden nach Möglichkeit als Zahlen erfasst
- Erste Zeile der Datentabelle enthält die Variablennamen
- Erste Spalte enthält eine laufende Nummer
- Jede Zeile entspricht einer Erhebungseinheit
- Jede Spalte entspricht einer Variable (Fragen mit Mehrfachantworten benötigen für jede Antwortalternative eine Variable)

Fragebogen	
Einführung in die Methodenlehre	
Geschlecht:	<input type="checkbox"/> weiblich <input type="checkbox"/> männlich
Alter:	_____ Jahre
Höchste abgeschlossene Schulbildung:	<input type="checkbox"/> Pflichtschule <input type="checkbox"/> Höhere Schule ohne Matura <input type="checkbox"/> Höhere Schule mit Matura <input type="checkbox"/> Universität, Fachhochschule
Art der Studienzulassung:	<input type="checkbox"/> AHS <input type="checkbox"/> HAK <input type="checkbox"/> Studienberechtigung <input type="checkbox"/> HBLA <input type="checkbox"/> HTL <input type="checkbox"/> Sonstige: _____
Studienrichtung: (Mehrfachnennungen möglich)	<input type="checkbox"/> Wirtschaftspädagogik <input type="checkbox"/> Statistik <input type="checkbox"/> Wirtschaftswissenschaften <input type="checkbox"/> Soziologie <input type="checkbox"/> Sozialwirtschaft <input type="checkbox"/> Sonstiges: _____
Maturanote (bzw. Abschlussnote) in Mathematik	_____
Maturanote (bzw. Abschlussnote) in Deutsch	_____
Körpergröße:	_____ cm Körpergewicht: _____ kg
Vielen Dank für Ihre Mitarbeit!	

Abb. 2.2. Fragebogen

Tabelle 2.2. Auszug aus dem Kodeplan

Merkmal	Merkmalstyp	Ausprägungen	Kodierung
Geschlecht	nominal, diskret dichotom	weiblich	1
		männlich	2
Alter	metrisch, stetig verhältnisskaliert	Messwerte	nicht notwendig
Schulbildung	ordinal, diskret	Pflichtschule	1
		Höhere Schule ohne Matura	2
		Höhere Schule mit Matura	3
		Universität, Fachhochschule	4
Studien- zulassung	nominal, diskret	AHS	1
		HBLA	2
		HAK	3
		HTL	4
		Studienberechtigung	5
		Sonstige	6
Wirtschafts- pädagogik	nominal, dichotom	nicht ausgewählt	0
		ausgewählt	1

Nun kann mit der Dateneingabe begonnen werden, fehlende Antworten werden im einfachsten Fall mit leeren Zellen (kein Leerzeichen, sondern eine Zelle ohne Inhalt) kodiert. Empfehlenswerter ist es, auch fehlende Werte mit einer Zahl zu kodieren. Die Kodierung für einen fehlenden Wert darf natürlich nicht als sinnvolle Ausprägung möglich sein (fehlende Werte für das Merkmal Alter würde man beispielsweise mit 999 kodieren). Bei Fragen mit Mehrfachnennungen werden nicht ausgewählte Alternativen mit 0 und ausgewählte Alternativen mit 1 kodiert. In der ersten Zeile der Tabelle werden die Variablenamen vermerkt. Das Ergebnis der Datenerfassung ist eine Datenmatrix wie in Abbildung 2.3 ersichtlich.

Die nun folgende Datenauswertung wird in diesem Buch ausführlich beschrieben. Daher wird bei der Beschreibung des Ablaufes auf diesen Teil verzichtet und gleich der Abschlussbericht als letzter Teil näher beleuchtet.

2.5 Abschlussbericht

Beim Verfassen des Abschlussberichtes ist darauf zu achten, dass nicht alle LeserInnen ausgebildete StatistikerInnen sind. In einer Einleitung wird kurz das Problem geschildert, weiters werden die Grundgesamtheit und die Untersuchungsmethode beschrieben. Der Auswertungsteil geht von der einfachen zur



<http://www.springer.com/978-3-642-37858-4>

Einführung in die Statistik mit EXCEL und SPSS
Ein anwendungsorientiertes Lehr- und Arbeitsbuch
Duller, C.
2013, XII, 283 S., Softcover
ISBN: 978-3-642-37858-4