

Vahlen kompakt

Marketing-Forschung

Mit Fallbeispiel

Bearbeitet von

Von Prof. Dr. Stefan Hoffmann, Dr. Anja Franck, Dr. Uta Schwarz, Prof. Dr. Katja Soye, und Dr. Stefan Wünschmann

1. Auflage 2018. Buch. IX, 183 S. Kartoniert

ISBN 978 3 8006 5646 2

Format (B x L): 14,1 x 22,4 cm

[Wirtschaft > Spezielle Betriebswirtschaft > Marketing, Werbung, Marktforschung](#)

Zu [Inhalts-](#) und [Sachverzeichnis](#)

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.

würde die Skala vermutlich keiner Normalverteilung folgen, sondern der Wert 0 würde besonders häufig auftreten.

Es sollte noch darauf hingewiesen werden, dass sich mehrere in diesem Buch besprochenen Verfahren (t-Test, ANOVA, Regressionsanalyse) aus derselben grundlegenden Logik ableiten lassen und dass sie deshalb als Spezialfälle eines sog. **Allgemeinen Linearen Modells** (ALM bzw. Generalized Linear Model, GLM) verstanden werden können (mehr dazu z.B. bei Sedlmeier/Renkewitz 2013, S. 609 ff.; Fox 2015; Bortz/Schuster 2010, S. 363 ff.; Cohen et al. 2003). Dieses Modell basiert wie die Regressionsanalyse auf Linearkombinationen. Gruppenunterschiede (wie beim t-Test oder der ANOVA) werden durch dichotome Prädiktoren modelliert. Aus diesem Grund ist die oben besprochene Dummy-Kodierung hier von besonderem Interesse.

Die Regressionsanalyse kann auch zur Theorieentwicklung eingesetzt werden. Dabei interessieren häufig komplexere Zusammenhänge als nur der Einfluss einer (oder mehrerer) unabhängigen Variablen auf eine abhängige Variable. Insbesondere relevant sind dabei zwei Formen von Drittvariablen:

- **Mediation:** Mediatorvariablen vermitteln den Einfluss von einer unabhängigen Variablen auf die abhängige Variable. Sie erklären damit, wie die UV auf die AV wirkt. Marta Formosa könnte bspw. annehmen, dass die UV Spaß am Einkaufen die AV Einstellung gegenüber Mixgetränken beeinflusst, weil der Spaß (UV) die Gemütslage (= Mediator) verbessert, was wiederum zu einer positiveren Einstellung führt. Ein klassisches Verfahren zur Prüfung von Mediationseffekten haben Baron/Kenny (1986) vorgeschlagen. Es besteht aus einer Abfolge von drei Regressionsanalysen, bei denen (etwas verkürzt ausgedrückt) der Einfluss der unabhängigen auf die abhängige Variable mit und ohne Mediator verglichen wird. Der Sobel-Test bietet einen Signifikanz-Test des indirekten Pfades von der UV über die Mediatorvariablen auf die AV (Sobel 1982). Eine kritische Betrachtung dieser ursprünglichen Methoden und mehr Informationen zur Mediationsanalyse findet sich u. a. bei Pieters (2017) und Zhao et al. (2010).
- **Moderation:** Eine Moderatorvariable beeinflusst die Stärke und/oder die Form des Zusammenhangs zwischen der UV und der AV. Möglicherweise ist der Einfluss der Innovationsfreudigkeit (UV) auf die Einstellung gegenüber Mixgetränken (AV) abhängig vom Einkommen des Konsumenten (Moderatorvariable). Im Rahmen einer Regressionsanalyse prüft man Moderatoreffekte dadurch, dass neben dem Prädiktor (unabhängige Variable) auch die Moderatorvariable (als zweite unabhängige Variable) und das Produkt aus Prädiktor und Moderatorvariable als Interaktionsterm in das

Modell aufgenommen werden. Viele Autoren empfehlen die unabhängigen Variablen zu standardisieren oder zu zentrieren (d. h. von den einzelnen Werten den Mittelwert abzuziehen, sodass der neue, zentrierte Mittelwert 0 entsteht), damit der Interaktionsterm besser interpretierbar ist (Cohen et al. 2003).

Zur Prüfung des Einflusses von Mediator- und Moderatorvariablen verbreitet sich das frei-zugängliche Makro „**Process**“ von Andrew Hayes, das für gängige Statistiksoftware wie SPSS und SAS verfügbar ist (www.processmacro.org/) und mit dessen Hilfe auch komplexere moderierte Mediationen und mediierte Moderationen berechnet werden können (Hayes 2013). Eine weitere Möglichkeit, komplexe Zusammenhänge zu modellieren, besteht in der Anwendung der **Strukturgleichungsmodellierung**. Strukturgleichungsmodelle setzen sich aus einem Strukturmodell, welches die Beziehungen zwischen den latenten Variablen beschreibt (ähnlich wie die Regressionsanalyse), und Messmodellen, welche die latenten Variablen durch manifeste Indikatoren operationalisieren (ähnlich wie die Faktorenanalyse), zusammen (z. B. Weiber/Mühlhaus 2014; Jöreskog/Sörbom 1982; Anderson/Gerbing 1988). Übliche Software-Pakete hierfür sind LISREL, AMOS, EQS oder MPLUS. In den letzten Jahren hat mit dem **PLS-Ansatz** ein zweiter Ansatz der Strukturgleichungsmodellierung weite Verbreitung gefunden, der (im Gegensatz zum oben genannten „kovarianzbasierten“, „harten“ Ansatz) als „varianzbasiert“ oder „weich“ bezeichnet wird (Wold 1982; Hair et al. 2011, 2016). Beim PLS-Ansatz werden zur Beurteilung einzelner Wirkungspfade die Standardfehler der Modellparameter mit Hilfe von Resample-Methoden (d. h. durch die wiederholte Ziehung von Stichproben aus dem Datensatz) berechnet (z. B. Bootstrapping; Lohmöller 1989). Es existieren verschiedene Softwarepakete; am häufigsten findet derzeit die Software SmartPLS Anwendung (<https://www.smartpls.com/>).

Schließlich werden in der Marketing-Forschung auch immer häufiger **Mehrebenen-Regressionen** durchgeführt (Raudenbush/Bryk 2002; Wieseke et al. 2008). Sie kommen zum Einsatz, wenn die untersuchten Fälle eine hierarchische bzw. genestete Struktur aufweisen. Beispielsweise könnten Konsumenten aufgrund ihrer Herkunft geordnet werden. Ihre Loyalität gegenüber einer Marke könnte von individuellen Einflüssen (pro Person gemessen), aber auch von der kulturellen Zugehörigkeit abhängen (pro Land gemessen; Krautz/Hoffmann 2017). Auch für die Mehrebenen-Regressionen existieren verschiedene Softwarepakete; weit verbreitet ist HLM (Hierarchical Linear Modelling, <http://www.ssicentral.com/hlm/>).

4.8 Literaturempfehlungen

- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2015). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (14. Aufl.). Berlin: Springer, Kap. 1, S. 63-134.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied Multiple Regression /Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models* (3rd ed.). Thousand Oaks: Sage.
- Hair, J. F. Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th ed.). New Jersey: Pearson, Kap. 4, S. 151-230.
- Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach* (6th ed.). Mason: Cengage Learning.

4.9 Lernhilfen

4.9.1 Kontrollfragen

1. Ergänzen Sie die Lückentexte.
 Mit Hilfe von lässt sich die Stärke des Einflusses zweier oder mehrerer unabhängiger Variablen überprüfen. Hohe absolute Beta-Werte sprechen für einen Einfluss eines Prädiktors auf die abhängige Variable. Ob eine weitere aufgenommene unabhängige Variable auch einen zusätzlichen Erklärungsbeitrag leistet, kann man weiterhin aus dem Anstieg des ablesen.
2. Nach welcher Maßgabe wird die Regressionsanalyse geschätzt? Welches Minimierungsproblem soll dabei gelöst werden?
3. Was versteht man unter Autokorrelation? Bei welchen Daten ist es wichtig, auf Autokorrelation zu prüfen?
4. Welche der folgenden Aussagen sind richtig?
 - Der Regressionskoeffizient (b) entspricht bei der einfachen linearen Regressionsanalyse dem Wert der Steigung der linearen Regressionsgeraden.
 - Der t-Test prüft die lokale Güte einer Regressionsanalyse. Er testet für alle unabhängigen Variablen gemeinsam, ob sich die Regressionskoeffizienten von 0 unterscheiden.
 - Damit der F- und der t-Test interpretiert werden dürfen, müssen die Residuen bei einer multiplen Regressionsanalyse der Normalverteilung folgen.

4.9.2 Vernetzungsaufgaben

Sie sollen nun eine multiple Regressionsanalyse rechnen. Die abhängige Variable ist „mix1“ („Ich mag es, Mixgetränke zu trinken.“). Als unabhängige Variablen fließen die folgenden Variablen ein: fak1_reg_hedo: Spaß am Einkauf/Hedonismus, fak2_reg_mb: Markenbewusstsein, fak3_reg_qb: Qualitätsbewusstsein, fak4_reg_inno: Innovationsakzeptanz, fak5_reg_mt: Markentreue, fak6_reg_kf: Kontaktfreudigkeit.

1. Berechnen Sie die Regressionskoeffizienten und stellen Sie die Regressionsgleichung auf. Verwenden Sie dabei die „Schrittweise Methode“. Wie beurteilen Sie Richtung und Stärke des Einflusses, den die einzelnen UV auf die AV ausüben?
2. Betrachten Sie nun das Bestimmtheitsmaß und das korrigierte Bestimmtheitsmaß. Welche Aussagen lassen sich aus beiden Werten ableiten?
3. Prüfen Sie die globale und die lokale Güte des berechneten Regressionsmodells.

Den Datensatz für eigene Berechnungen finden Sie auf www.wahlen.de bzw. dem QR-Code:



5.1 Ziel

Fallbeispiel Limonatik

Bei ihren bisherigen Analysen hat sich Marta Formula mit der Frage beschäftigt, durch welche Größen die Einstellung gegenüber Mixgetränken beeinflusst wird. Sie hat dabei einiges über die potenziellen Käufer des neuen Getränkes erfahren können. Sie vermutet jedoch, dass es mehrere Kundentypen gibt, die ganz unterschiedlich gegenüber Mixgetränken eingestellt sind. Möglicherweise gibt es bestimmte Kundentypen wie „Markenbewusste“, „Preisorientierte“, „Genussorientierte“, „Partytrinker“, „Gesundheitsbewusste“ oder „Limonadenverschmäher“. Sie fragt sich deshalb, ob die bereits erhobenen soziodemografischen, produkt- und markenrelevanten Merkmale auch dazu dienen können, um Kundengruppen zu bilden.

Bei der Clusteranalyse sind folgende Fragen zu klären:

- Anhand welcher Variablen können Konsumenten oder Objekte (z. B. Unternehmen) zu Segmenten gruppiert werden?
- Wie viele Segmente lassen sich unterscheiden?
- Wie lassen sich die Segmente beschreiben?

Marktforscher segmentieren Kunden häufig mit der Absicht, die Zielgruppe(n) eines Unternehmens näher zu definieren. Auch für Marta Formula ist das eine Frage, der sie nachgehen möchte. Steht hierfür nur ein Merkmal zur Verfügung, bedarf es keiner aufwendigen Analyse: Marta Formula teilt einfach die Stichprobe in Männer und Frauen (Merkmal = Geschlecht), Käufer und Nicht-Käufer (Merkmal = Kauf) oder gut situierte und weniger gut situierte Kunden (Merkmal = Einkommen). Wie lassen sich aber Kundengruppen auf Basis mehrerer Merkmale bilden? Antworten auf solche und ähnliche Fragestellungen liefert die Clusteranalyse. Als multivariates Analyseverfahren deckt sie Strukturen in einem Datensatz auf,

indem sie Objekte bzw. Personen (z. B. Kunden, Marken, Probanden) in Gruppen, den sog. **Clustern** zusammenfasst. Die Cluster werden so gebildet, dass sich die in einem Cluster enthaltenen Objekte in mehreren Eigenschaften untereinander ähneln und sich von den Objekten anderer Cluster unterscheiden.

Ähnlich wie die Faktorenanalyse soll die Clusteranalyse Komplexität reduzieren. Während die Faktorenanalyse aber Variablen zu Faktoren zusammenfasst, gruppiert die Clusteranalyse im Regelfall Objekte oder Personen.

Ziel der Clusteranalyse ist es demzufolge, Objekte mit ähnlichen Merkmalsausprägungen zu Gruppen zusammenzufassen, die intern möglichst homogen sind und die sich möglichst gut von anderen Gruppen abgrenzen lassen. Anders ausgedrückt: Unterschiede innerhalb der Cluster sollten möglichst gering sein, während sich die Cluster untereinander möglichst stark unterscheiden sollten. In Abb. 34 ist dieses Prinzip schematisch anhand von geometrischen Figuren dargestellt, die auf Basis der Merkmale Größe, Farbe und Form zu Clustern vereint wurden.

Die Clusteranalyse wird in der Marketing-Forschung meist zur **Marktsegmentierung** eingesetzt und es werden meist Personen „geclustert“. So könnte man fragen, ob sich Konsumenten anhand ihres Lebensstils in homogene Segmente einteilen lassen. Sung (2004, S. 344ff.) nutzt bspw. vier Merkmalsgruppen, um amerikanische Abenteuerreisende zu klassifizieren (demografische, sozioökonomische und reisebezogene Eigenschaften sowie die wahrgenommene Wichtigkeit der Komponenten einer Abenteuerreise). Es ergeben sich sechs Kundentypen, wie z. B. der „Familienurlauber“ oder der „aktive Einzelgänger“. Mai/Hoffmann (2012) unterscheiden verschiedene Segmente von Joghurt-Käufern (insb. Taste Lovers vs. Nutrition Fact Seekers), wobei die Cluster auf Basis der Angaben bestimmt werden, wie wichtig den Probanden Produktmerkmale wie Geschmack, Preis etc. sind. Es ist aber auch möglich, Nationen, Unternehmen, Marken sowie Variablen zu gruppieren. So könnte man fragen, ob sich Segmente von Unternehmen finden lassen, die ähnliche Marketing-Strategien verfolgen. Müller/Gelbrich (2015, S. 256ff.) bilden Kulturcluster, indem sie Länder auf Basis der Kultur-Dimensionen von Hofstede zusammenfassen.

Marta Formula liegen die Daten von 471 Probanden vor, die zu ihrer Einstellung gegenüber Mixgetränken, sowie zu ihren Konsumgewohnheiten und Freizeitaktivitäten befragt worden sind. Sie möchte die Befragten anhand der Dimensionen „Qualitätsbewusstsein“, „Preisbewusstsein“, „Markenbewusstsein“ und „Markentreue“ zu Kundentypen verdichten. Jeder Dimension liegen bis zu fünf Items zugrunde. Diese hatte Marta Formula vorab mit einer Faktorenanalyse auf vier Faktoren reduziert (vgl. Kap. 3 „Faktorenanalyse“).

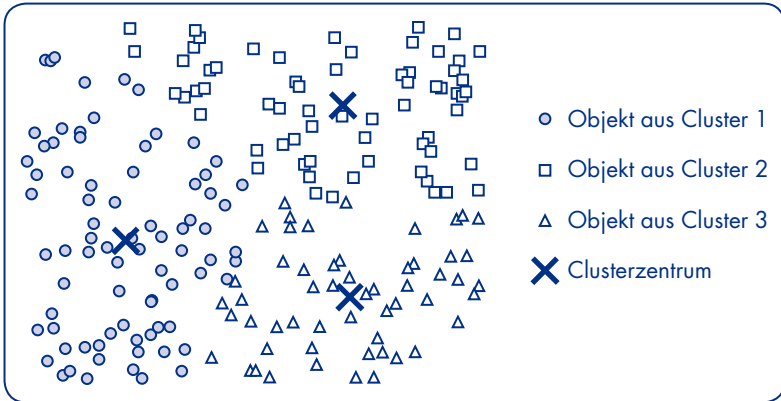


Abb. 34: Schematische Darstellung der Grundidee der Clusteranalyse

5.2 Voraussetzungen

Im Gegensatz zur Varianz- und Regressionsanalyse handelt es sich bei der Clusteranalyse nicht um eine Dependenz-, sondern um eine **Interdependenzanalyse** (Hair et al. 2014, S. 13), bei der es keine unabhängigen und abhängigen Variablen gibt.

Die Clusteranalyse stellt im Gegensatz zur Varianz- und Regressionsanalyse keine Anforderungen an das **Skalenniveau** der einbezogenen Variablen. Es lassen sich sowohl metrische als auch nominale und ordinale Daten nutzen und es können verschiedene Skalenniveaus kombiniert werden (vgl. Tab.39). Vom Skalenniveau hängt jedoch die Wahl des Proximitätsmaßes, d. h. des Ähnlichkeits- bzw. Distanzmaßes, ab (vgl. Kap. 5.3.2).

Art der Analyse	Interdependenzanalyse (Zusammenfassung mehrerer Fälle in wenige Cluster)
Anzahl Variablen	Mehrere (keine Einteilung in AV und UV)
Skalenniveau	Nominal, ordinal, metrisch

Tab. 39: Kurzcharakterisierung der Clusteranalyse

Auch für die **Anzahl der Variablen**, die in eine Clusteranalyse eingehen sollen, gibt es keine Empfehlung. Es ist aber nicht sinnvoll, so viele Variablen wie möglich in eine Clusteranalyse aufzunehmen, da irrelevante Größen das Ergebnis verzerren können (Milligan/Cooper 1987, S.350). Vielmehr sollte der Forscher nur jene Indikatoren einbeziehen, die mit seiner Forschungsfrage korrespondieren

und von denen er annimmt, dass sie die Daten in sinnvolle Gruppen trennen (Saunders 1994, S. 16). Alle Merkmale sollten allerdings dazu beitragen, die Cluster zu unterscheiden. Wenn zwei Items stark voneinander abhängen, d. h. nahezu den gleichen Sachverhalt messen, so ist eines entbehrlich. Wenn Marta Formula bspw. ihre Probanden befragt hätte, wie häufig sie Limonade trinken und wie oft sie Limonade kaufen, so wäre zu erwarten, dass diese beiden Variablen nahezu dieselben Informationen liefern. Eine der beiden hoch korrelierten Variablen wäre dann für die Clusteranalyse ausreichend. Falls die Variablen die Voraussetzungen für eine Faktorenanalyse erfüllen, empfiehlt es sich, Merkmale, die stark korrelieren, vorab durch eine Faktorenanalyse mit orthogonaler Rotation zu verdichten. Auf diesem Wege gebildete Faktoren sind unabhängig voneinander und eignen sich somit als Grundlage für eine Clusteranalyse (Breckenridge 2000, S. 266). Zudem lässt sich dann auch das Ergebnis leichter interpretieren, denn die Faktorenanalyse reduziert viele Merkmale zumeist auf wenige Faktoren (Saunders 1994, S. 16).

Die Clusteranalyse führt keinen inferenzstatistischen Test durch, sodass keine Verteilungsannahmen bestehen. Deshalb muss auch nicht durch eine ausreichend große Stichprobe die statistische Power (Teststärke) sichergestellt werden. Die **Stichprobe** sollte jedoch repräsentativ und so groß sein, dass auch kleine Cluster in der Grundgesamtheit durch die Stichprobe abgebildet werden können (Backhaus et al. 2015, S. 510).

DIE FACHBUCHHANDLUNG

5.3 Durchführung

5.3.1 Ausgangs-/Rohdatenmatrix

Der Ausgangspunkt für eine Clusteranalyse ist die Ausgangs- bzw. Rohdatenmatrix mit k Fällen (Probanden, Objekte) und x Variablen (Merkmale). Bereits daran lassen sich **per Augenschein** ähnliche Fälle identifizieren. Je mehr Variablen und je mehr Fälle vorliegen, desto aufwendiger wäre jedoch dieses unsystematische Verfahren.

In Tab. 40 ist ein Ausschnitt einer Ausgangsdatenmatrix dargestellt. In unserem Beispiel gibt sie nicht die ursprünglichen Antworten der Probanden wieder, sondern die Faktorwerte. Letztere stammen aus einer Faktorenanalyse über mehrere Items zum Marken-, Qualitäts- und Preisbewusstsein sowie der Markentreue, die Marta Formula im Vorfeld durchgeführt hat. Ein Faktorwert lässt sich nicht mehr gemäß der ursprünglichen Skala (z. B. „lehne voll und ganz ab“ bis „stimme voll und ganz zu“) interpretieren. Marta Formula erkennt aber bspw., dass Person 4 mehr Wert auf Marken legt als Person 48 (2,03 vs. -1,32).