

# planung & analyse

Zeitschrift für Marktforschung und Marketing

[www.planung-analyse.de](http://www.planung-analyse.de)

2/2013 D11700F

**Schwerpunkt**

## Qualitative Forschung

**Special**

Blick auf den Tacho:  
Aktuelles aus der  
Automobilmarktforschung

**Report**

Blick auf die  
Jüngsten:  
Kids & Youngster  
erfolgreich  
erforschen

**Forschung**

Blick ins Gehirn:  
Zusammenhang von Persönlichkeitsstruktur,  
Markenloyalität und Hirnaktivität



## Statistik KOMPAKT Identifizierung von Scheinkorrelationen

### Kontrolle von Störvariablen mit der Regressionsanalyse

Ein statistischer Zusammenhang wie beispielsweise eine positive Korrelation zwischen Produktqualität und Kundenzufriedenheit genügt allein nicht, um die Produktqualität als Treiber der Zufriedenheit auszumachen. Es muss zudem ausgeschlossen sein, dass dieser statistische Zusammenhang auf Störvariable(n) zurückzuführen ist. So könnte etwa das Markenimage beide Größen beeinflussen und damit die Korrelation zumindest zu einem Teil hervorrufen. Mithilfe der Regressionsanalyse kann der Einfluss einer Störvariable auf Ursache und Wirkung eines vermuteten kausalen Zusammenhangs kontrolliert und somit eine Scheinkorrelation aufgedeckt werden.

### Lineare Regressionsanalyse

Mit der Regressionsanalyse wird die Abhängigkeit einer metrischen Variable (Regressand) von einer oder mehreren metrischen und/oder dichotomen Variablen (Regressoren) untersucht. Für jeden Regressor bestimmt sie einen Koeffizienten, der dessen Einflussstärke auf den Regressanden misst. Damit lässt sich eine Regressionsfunktion aufstellen, mittels der für jeden Fall der Stichprobe auf Basis der beobachteten Werte der Regressoren ein Wert des Regressanden berechnet (prognostiziert) werden kann. Die Differenz zwischen dem prognostizierten und dem beobachteten Wert des Regressanden ist das Residuum. Die Regressionskoeffizienten werden so berechnet, dass die Summe der quadrierten Residuen über alle Befragten minimiert wird (Ordinary Least Squares (OLS)-Schätzung).

### Beispiel zur Kontrolle einer Störvariable

Vermutet wird, dass ein kausaler Zusammenhang zwischen der Ursache  $x$  und der Wirkung  $y$  besteht. Betrachten wir dies in einem Beispiel, bei dem die Korrelation zwischen  $x$  und  $y$  0,44 beträgt. Für die Korrelationen einer möglichen Störvariable mit der Ursache und Wirkung werden im Weiteren drei Situationen unterschieden\*.

#### Situation 1: $z_1$ korreliert mit Ursache und Wirkung.

Mit Hilfe jeweils einer einfachen Regression wird der Teil der Ursache beziehungsweise der Wirkung berechnet, der allein durch die Störvariable determiniert ist. Die Residuen für  $x$  und  $y$  spiegeln dann die Werte der Ursache beziehungsweise der Wirkung wider, die nach Herausrechnen des Einflusses von  $z_1$  übrig bleiben. Ist die Korrelation dieser Residuen 0 ist der positive statistische Zusammenhang zwischen Ursache und Wirkung vollständig auf den Einfluss der Störvariable auf die beiden Variablen zurückzuführen. Die Korrelation von 0,44 zwischen  $x$  und  $y$  ist eine *Scheinkorrelation*.

Der Umweg über die Berechnung der Residuen ist jedoch nicht notwendig, um den Einfluss der Störvariable auszuschalten. In einer multiplen Regression gibt ein Regressionskoeffizient die Veränderung des Regressanden an, wenn sich der Wert des Regressors um eine Einheit verändert – bei Konstanz der übrigen Regressoren. Um Regressoren mit unterschiedlichen Skalierungen vergleichbar zu machen, wird der standardisierte Regressionskoeffizient berechnet. Eine multiple Regression mit der Ursache und der Störvariable als Regressoren sowie der Wirkung als Regressand führt zu einem standardisierten Regressions-

koeffizienten der Ursache von 0 (siehe rote Werte in der Abbildung). Das heißt bei konstanter Störvariable verändert sich  $y$  nicht, wenn sich  $x$  ändert. Damit besteht zwischen Ursache und Wirkung offensichtlich kein kausaler Zusammenhang.

#### Situation 2: $z_2$ korreliert nur mit der Wirkung.

Der standardisierte Regressionskoeffizient der Ursache in der multiplen Regression ist gleich 0,44 (siehe violette Werte in der Abbildung). Das heißt er ist genauso hoch wie die Korrelation zwischen Ursache und Wirkung. Trotz ihres Einflusses auf  $y$  ist  $z_2$  somit keine Störvariable, die den statistischen Zusammenhang zwischen Ursache und Wirkung beeinträchtigt.

**Situation 3:** Im Vergleich zu Situation 1 ist die Korrelation zwischen Störvariable ( $z_3$ ) und Ursache gleich, die Korrelation zwischen Störvariable und Wirkung aber geringer. Der standardisierte Regressionskoeffizient der Ursache ist in der multiplen Regression mit 0,28 positiv (siehe grüne Werte in der Abbildung). Insofern ist nicht der gesamte statistische Zusammenhang zwischen Ursache und Wirkung auf die Störvariable zurückzuführen. In dieser Situation ist der durch die Störvariable erklärte Teil der Varianz der Wirkung geringer als in Situation 1, so dass Varianz übrig bleibt, die durch  $x$  erklärt wird.

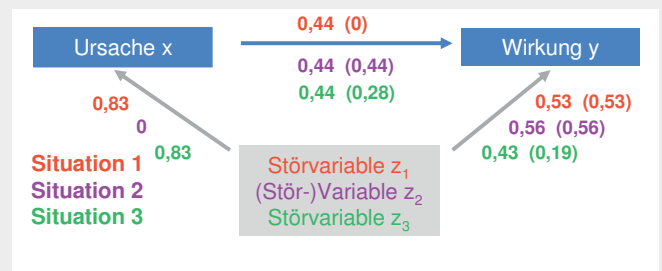


Abbildung: Korrelationen (Standardisierte Regressionskoeffizienten)

### Fazit

Eine Regression, die im Regressionsansatz gleichzeitig Ursache und Störvariable enthält, berechnet mit dem (standardisierten) Regressionskoeffizienten der Ursache den statistischen Zusammenhang zwischen Ursache und Wirkung, der um den Einfluss dieser Störvariable bereinigt ist. Dieser Koeffizient ist – neben der Korrelation zwischen Ursache und Wirkung – bestimmt durch die Korrelation zwischen Ursache und Störvariable sowie durch die Korrelation der Störvariablen mit der Wirkung, falls Ursache und Störvariable korrelieren. Bei mehreren Störvariablen kann analog vorgegangen werden, indem diese gemeinsam mit der Ursache in den Regressionsansatz aufgenommen werden. ◀

\* Der zugrunde liegende Datensatz kann bei den Autoren (hschimmelpfennig@ifad.de) angefordert werden.

**Johannes Lüken** und **Dr. Heiko Schimmelpfennig**, Experten für Multivariate Analysen bei IfaD, Institut für angewandte Datenanalyse GmbH  
In Ausgabe 3/2013: *Strukturgleichungsmodellierung*

### ► Literatur

Bortz, J.; Schuster, C.: Partielle Korrelation und multiple lineare Regression. In: Statistik für Human- und Sozialwissenschaftler, 7. Auflage, Berlin, Heidelberg, 2010, S. 339-361.

Firebaugh, G.: The Fifth Rule: Compare Like With Like. In: Seven Rules for Social Research, Princeton, Oxford, 2008, S. 120-164.