

planung & analyse

Zeitschrift für Marktforschung und Marketing www.planung-analyse.de

6/2012 D11700F

Marketing- controlling



Einblick

Special zu den
Lebensphasen Kids,
Families und Best Ager

Einsicht

Cause related Marketing
ist kein Selbstgänger

Eindruck

So blicken Lenker der
Branche auf das Jahr 2013

Statistik **KOMPAKT** *Signifikanz und Stichprobenumfang*

Nicht selten kommt es bei der Tabellierung von Marktforschungsdaten vor, dass ein Unterschied zwischen zwei Teilgruppen als signifikant und ein Unterschied zwischen zwei anderen Gruppen nicht als signifikant ausgewiesen wird, obwohl dieser größer ist. Eine mögliche Erklärung dafür ist, dass die ersten beiden Gruppen stärker besetzt sind als die anderen beiden.

Signifikanztest

Ein in einer Stichprobe beobachteter Effekt ist signifikant, wenn dieser wahrscheinlich nicht zufällig aufgetreten ist. Man kann dann davon ausgehen, dass ein Effekt auch in der entsprechenden Grundgesamtheit vorhanden ist. Die Prüfung auf Signifikanz erfolgt mit einem statistischen Test. Diesem liegt die Hypothese zugrunde, dass kein Effekt vorliegt. Es wird die Wahrscheinlichkeit p bestimmt, mit dem Verwerfen der Hypothese einen Fehler zu begehen. Ist p kleiner als ein vorgegebenes Signifikanzniveau α (zumeist ist $\alpha = 0,05$), so wird die Hypothese abgelehnt, das heißt der Effekt ist signifikant.

Abhängigkeit der Signifikanz vom Stichprobenumfang

Soll beispielsweise überprüft werden, ob sich das Ausprobieren eines Produkts positiv auf die Kaufbereitschaft auswirkt, wird die Kaufbereitschaft vor und nach dem Ausprobieren bei denselben zufällig ausgewählten Befragten auf einer Ratingskala erhoben. Im Mittel ergibt sich in der Stichprobe eine um 0,2 Skalenpunkte höhere Kaufbereitschaft nach dem Ausprobieren. Ist diese Erhöhung signifikant? Hierfür wird die einseitige Hypothese getestet, dass sich die Kaufbereitschaft nach dem Ausprobieren nicht erhöht. In Abbildung 1 sind die für mögliche Stichprobenumfänge und verschiedenen große Grundgesamtheiten resultierenden p -Werte dargestellt. Obwohl allen derselbe Mittelwert der Differenzen zwischen der Kaufbereitschaft vor und nach dem Ausprobieren sowie dieselbe Standardabweichung der Differenzen zugrunde liegt, unterscheiden sich die p -Werte. Im Fall einer sehr großen Grundgesamtheit ist die Verbesserung der durchschnittlichen Kaufbereitschaft für $n = 100$ und $n = 150$ nicht signifikant, da p größer als 0,05 ist. Für $n = 200$ ist sie dagegen signifikant.

Stichprobenumfang	Grundgesamtheit sehr groß	Grundgesamtheit N = 1000	Grundgesamtheit N = 500
n = 100	0,11	0,10	0,08
n = 150	0,06	0,05	0,03
n = 200	0,04	0,02	0,01

Abbildung 1: p -Werte für verschiedene Umfänge

Das heißt bei einem größeren ist es wahrscheinlicher als bei einem kleineren Stichprobenumfang, dass die Hypothese abgelehnt wird und ein Effekt signifikant ist. Insofern spricht das Verwerfen einer Hypothese bei einer kleinen Stichprobe für einen stärkeren Effekt in der Grundgesamtheit als bei einer großen Stichprobe.

Statistischer Hintergrund

Ob ein Effekt signifikant ist, hängt vom Ergebnis in der Stichprobe ebenso ab wie von der Stichprobenverteilung. Die Stichprobenverteilung

beschreibt die Verteilung beispielsweise des Mittelwerts für alle denkbaren Stichproben eines bestimmten Umfangs. Die Abbildung 2 zeigt die Stichprobenverteilungen für zwei unterschiedliche Stichprobenumfänge. Bei einem großen Stichprobenumfang ist die Verteilung viel schmaler, das heißt die Standardabweichung des Mittelwerts – der Standardfehler – kleiner. Die hellblaue Fläche unter der Kurve entspricht dem p -Wert beim großen Stichprobenumfang. Beim kleinen Stichprobenumfang ist der p -Wert trotz des gleichen Ergebnisses in der Stichprobe aufgrund des höheren Standardfehlers um die graue Fläche größer.

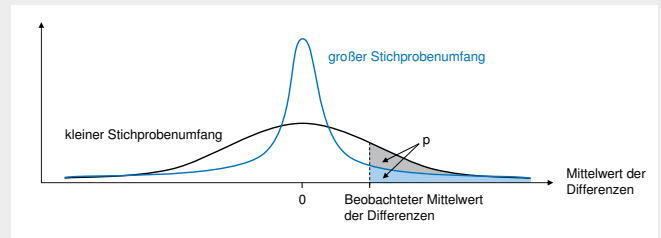


Abbildung 2: Stichprobenverteilung

Konsequenz für gewichtete Daten

Eine Fallgewichtung verändert nicht nur das Ergebnis in der Stichprobe wie beispielsweise den Mittelwert. Der gewichtete Stichprobenumfang besitzt ebenfalls Einfluss auf den Ausgang des Signifikanztests. Um die Signifikanz eines Effekts nicht ungerechtfertigt zu begünstigen, sollte unbedingt vermieden werden, dass die Summe der Gewichte größer als der ursprüngliche Stichprobenumfang ist. Um die aufgrund der Gewichtung erhöhte Unsicherheit bei der Schätzung des Mittelwerts zu berücksichtigen, wird vielmehr empfohlen, einen reduzierten Stichprobenumfang – die so genannte effective base – anzusetzen, die umso kleiner ist, je höher die Varianz der Fallgewichtungen ist.

Kleine Grundgesamtheiten und Vollerhebungen

Je kleiner die Grundgesamtheit im Verhältnis zum Stichprobenumfang ist, desto geringer ist der Standardfehler, wodurch die Stichprobenverteilung schmalgipfliger und der p -Wert damit kleiner wird. Der Vergleich der p -Werte in Abbildung 1 veranschaulicht diesen Zusammenhang.

Auf Basis einer Vollerhebung lassen sich die wahren Werte bestimmen. Das Formulieren von Hypothesen über die Werte in der Grundgesamtheit und Signifikanztests erübrigen sich dann. In dem Beispiel würde jede noch so kleine positive Abweichung des Mittelwerts von null einen positiven Effekt des Ausprobierens auf die Kaufbereitschaft bedeuten. Sinnvoll sind Signifikanztests dann nur, um zu überprüfen, ob ein beobachteter Effekt aufgrund von Messfehlern zustande gekommen ist. ◀

Johannes Lünen und **Dr. Heiko Schimmelpfennig**, Experten für Multivariate Analysen bei IfaD, Institut für angewandte Datenanalyse GmbH.

In Ausgabe 1/2013: *Einführung in Kausalanalysen*

► Literatur

Behnke, J.: Lassen sich Signifikanztests auf Vollerhebungen anwenden? Einige essayistische Anmerkungen. In: Politische Vierteljahresschrift, Jg. 46/2005, Heft 1, S. 0-1-0-15.

Kish, L.: Survey Sampling, New York et al., 1965, S. 427.

Lipovetzky, S.: Post-Stratification with Optimized Effective Base. In: Proceedings of the Survey Research Methods Section, American Statistical Association, 2007, S. 2313-2320.