

Eine Verallgemeinerung des Cochran-Armitage-Tests auf mehrere Gruppen mit PROC LOGISTIC und PROC GENMOD

Oliver Kuss

Institut für Medizinische Epidemiologie, Biometrie und Informatik
Universität Halle-Wittenberg
06097 Halle/Saale
Oliver.Kuss@medizin.uni-halle.de

Heinrich Dickel

Abteilung Klinische Sozialmedizin
Universitätsklinikum Heidelberg
Bergheimer Str. 58
69115 Heidelberg
Heinrich_Dickel@med.uni-heidelberg.de

Zusammenfassung

Das Ausgangsproblem dieser Untersuchung war die Frage nach einem Unterschied im Verlauf von Inzidenzraten in drei verschiedenen Altersgruppen in einer populationsbasierten Registerstudie. Ein Abfall der Inzidenz in allen drei Gruppen über die Zeit war offensichtlich, von Interesse war aber die Frage, ob ein Unterschied in der Stärke des Abfalls besteht. Der Cochran-Armitage-Test, der seit Version 6.12 mit der TREND-Option im TABLES-Statement von PROC FREQ aufgerufen werden kann, testet auf einen Trend von Raten in einer einzigen Gruppe, ein Vergleich von Trends in mehreren Gruppen ist aber nicht möglich. Auf der SAS-Homepage fand sich der knappe Hinweis, dass der Cochran-Armitage-Test auch als Score-Test in einem logistischen Regressionsmodell interpretiert werden kann, insofern war es möglich, nicht nur den Cochran-Armitage-Test mit PROC LOGISTIC nachzuvollziehen, sondern auch im Kontext der logistischen Modelle die gewünschte Verallgemeinerung auf mehrere Gruppen mit Hilfe von PROC LOGISTIC und PROC GENMOD.

Keywords: Cochran-Armitage-Test, PROC LOGISTIC, PROC GENMOD, Trends in Inzidenzraten, BKH-N.

1 Das inhaltliche Problem

Im BKH-N (Berufskrankheitenregister Haut-Nordbayern) wurden in den Jahren 1990 bis 1999 alle Verdachtsmeldungen auf das Vorliegen einer berufsbedingten Hautkrankheit in Nordbayern standardisiert erfasst und registriert ([1], [2]). Da der durchschnittliche Beschäftigungsstand in den jeweiligen Berufen bekannt war, konnten Aussagen zur Inzidenz von berufsbedingten Hautkrankheiten in verschiedenen Berufsgruppen und im zeitlichen Verlauf getroffen werden.

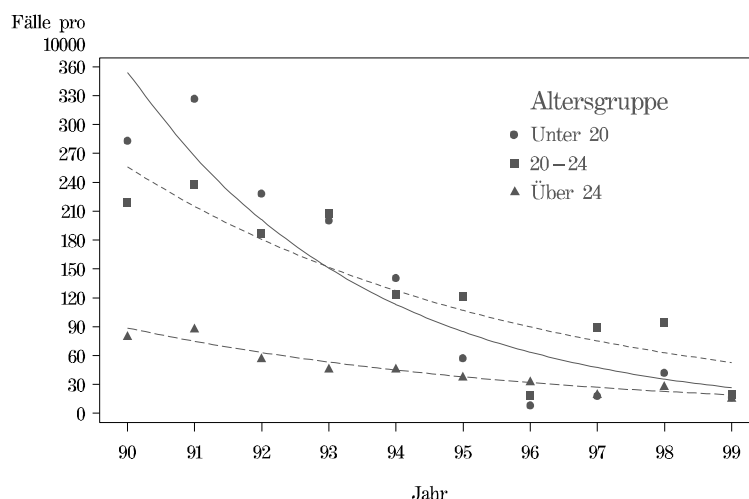


Abbildung 1: BKH-N, jährliche beobachtete und geschätzte Inzidenzraten von Friseuren, nach Altersgruppen getrennt

Abbildung 1 zeigt den Verlauf der Inzidenzraten in der besonders betroffenen Berufsgruppe der Friseure, zur Beurteilung von potentiellen 'Healthy Worker'- oder 'Hardening'-Effekten ([3], [4]) wurden die Inzidenzraten in drei Altersgruppen aufgeteilt. Ein Abfallen der Inzidenzraten über die Zeit in allen drei Altersgruppen ist offensichtlich und sehr wahrscheinlich auf verstärkte Anstrengungen der Berufsgenossenschaften zur Vermeidung von hautgefährdenden Stoffen im Friseurhandwerk Anfang der neunziger Jahre zurückzuführen. Gleichzeitig scheint aber die Stärke des Abfalls in den drei Gruppen unterschiedlich zu sein, wobei der Rückgang der Inzidenz in der niedrigsten Altersgruppe am ausgeprägtesten ist. Von inhaltlich großem Interesse war die statistische Überprüfung dieses Sachverhalts.

2 Trendtest für Raten: Der Cochran-Armitage-Test

Das statistische Standardverfahren zur Beurteilung von Trends in Raten ist der Cochran-Armitage-(CA-)Test ([5], [6]), eine aktuelle Übersicht zu diesem Thema findet man bei Williams ([7]). Der Cochran-Armitage-Test hat eine Fülle von günstigen mathematischen Eigenschaften, so wurde gezeigt, dass dieser asymptotisch lokal optimal für jede hinreichend glatte Alternative zur Nullhypothese ([8]) und für alle logistischen Alternativen der gleichmäßig beste unverfälschte Test ist ([9]). Das bedeutet also insbesondere, dass der CA-Test nicht als ein Test auf einen ausschließlich linearen Trend betrachtet werden kann, es ist vielmehr in der Lage, vielfältige monotone Abweichungen (z. B. quadratische oder exponentielle) der untersuchten Raten von einem konstanten Verlauf zu entdecken ([10]).

Seit SAS Version 6.12 kann dieser Test in PROC FREQ mit Hilfe der TREND-Option im TABLES-Statement aufgerufen werden. Der folgende Programm-Code zeigt anhand des Friseur-Beispiels, wie dies konkret zu realisieren ist, wobei der Cochran-Armitage-Test für die einzelnen Altersgruppen getrennt berechnet wird.

```
data friseure;
    input year ageclass response count;
    cards;
90 1 1 130
90 1 0 4458
90 2 1 71
90 2 0 3166
90 3 1 35
90 3 0 4281
    ...
;run;

proc sort data=friseure; by ageclass; run;

proc freq data=friseure;
    tables year*response / trend;
    weight count;
    by ageclass;
run;
```

Der Wert der CA-Statistik in den drei Altersgruppen ist 11.30 in der Gruppe der unter 20-Jährigen, 6.79 in der Gruppe der 20- bis 24-Jährigen und 5.50 in der Gruppe der über 24-Jährigen, wobei dieser Trend in allen drei Gruppen hochsignifikant ($p < 0.0001$) ist.

Dies ist jedoch bei weitem nicht die einzige Möglichkeit, den CA-Test innerhalb von SAS zu berechnen. Mit Hilfe des EXACT-Statements in PROC FREQ erhält man eine exakte Version des CA-Tests, PROC MULTTEST erlaubt sogar eine Resampling-Variante des Tests ([11]). Auf eine weitere, eher skurrile Möglichkeit weist Williams ([7]) hin, wonach man zeigen kann, dass der CA-Test gleich dem linearen Korrelationskoeffizienten (auch wenn es sich um zwei dichotome Größen handelt!) zwischen Zielgröße (hier: `response`) und Klassenvariablen (hier: `year`) ist, so dass man sogar PROC CORR bemühen kann, um den CA-Test zu berechnen.

3 Eine Verallgemeinerung des Cochran-Armitage-Tests auf mehrere Gruppen

Jedoch helfen diese Erkenntnisse in vorliegendem Fall nicht weiter, da man nicht an der Überprüfung eines einzelnen Trends interessiert sind, sondern am Vergleich von Trends in mehreren Gruppen. Der entscheidende Hinweis zur Lösung des Problems fand sich schließlich auf der SAS-Homepage ([12]), wo ohne Literaturangabe behauptet wird, dass der CA-Test auch als Score-Test in einem logistischen Regressionsmodell interpretiert werden kann. Diese Erkenntnis scheint in der Statistik allgemein bekannt zu sein und findet sich an mehreren Stellen in der Literatur (s. z. B. ([13]), S. 121), eine konkrete Herleitung dieser Äquivalenz findet man bei Chen ([14]).

Mit PROC LOGISTIC ist diese Behauptung leicht zu überprüfen, mit folgendem Code findet man im Output unter der Rubrik

```
Testing Global Null Hypothesis: BETA=0
```

die entsprechende Test-Statistik des Score-Tests, die gleich dem Quadrat des CA-Tests aus PROC FREQ ist.

```
proc logistic data=friseure;  
  freq count;  
  model response=year;  
  by ageclass;  
run;
```

Konkret wird durch diesen Score-Test geprüft, ob alle Kovariablen (mit Ausnahme des Intercepts) in diesem logistischen Regressionsmodell von Null verschieden sind. Da die Klassenvariable (hier: `year`) die einzige Kovariable in diesem Modell ist, wird also geprüft, ob diese einen signifikanten Einfluss auf die Zielgröße hat. Es besteht also eine Äquivalenz zwischen dem Vorliegen eines monotonen nicht-konstanten Ratentrends über die verschiedenen Ausprägungen der Klassenvariablen in einer Mehrfeldertafel und dem signifikanten Einfluss dieser Variablen in einem logistischen Regressionsmodell.

Es liegt nun nahe, diese Äquivalenz auszunutzen, und die spezifische Fragestellung nach dem Unterschied von zwei Ratentrends im Kontext eines logistischen Modells zu formulieren. Dazu wird die neue Variable (im folgenden „Gruppenvariable“ genannt), die die unterschiedlichen Ratentrends beschreibt (hier: `ageclass`), sowie deren *Interaktionsterm mit der Klassenvariablen* in das logistische Regressionsmodell mit aufgenommen. Dieser Interaktionsterm wird genau dann einen signifikanten Einfluss haben, wenn der Effekt der Klassenvariablen auf den verschiedenen Stufen der Gruppenvariablen unterschiedlich ist, konkret, wenn die Ratentrends in den verschiedenen Altersgruppen unterschiedlich sind. Dies ist, über den Umweg in die logistischen Regressionsmodelle, eine simple Verallgemeinerung des Cochran-Armitage-Tests für mehrere Gruppen. Da die Parameter in diesen Modellen iterativ geschätzt werden müssen, kann keine geschlossene Formel für die Test-Statistik angegeben werden, jedoch ist es überhaupt kein Problem, diese Berechnung mit PROC LOGISTIC und PROC GENMOD zu realisieren:

```
proc logistic data=friseure descending;
  freq count;
  class ageclass / param=glm;
  model response=ageclass|year;
run;
```

beziehungsweise

```
proc genmod data=friseure descending;
  freq count;
  class ageclass;
  model response=ageclass|year / d=bin link=logit type3;
run;
```

Eine kleine Unschärfe ergibt sich daraus, dass in diesem Fall nicht der Score-Test für die Interaktion zwischen Jahr und Altersklasse berechnet wird, sondern der Wald-Test (in PROC LOGISTIC) bzw. der Likelihood Ratio-Test (PROC GENMOD). Dies ist jedoch nicht wirklich eine Einschränkung, führen diese drei Testprinzipien doch asymptotisch zu denselben Ergebnissen, wobei in der Realität bei endlichen Fallzahlen in der Regel der Likelihood Ratio-Test die besten Eigenschaften hat ([15]).

Als Ergebnis findet man mit PROC LOGISTIC in der Rubrik `Type III Analysis of Effects` einen Wert der Wald-Teststatistik für `year*ageclass` von 11.60, was einem p-Wert von 0.003 entspricht, und mit PROC GENMOD in der Rubrik `LR Statistics For Type 3 Analysis` einen Wert der LR-Teststatistik für `year*ageclass` von 12.01 ($p=0.0025$). Erwartungsgemäß liegen die Werte der Teststatistiken im selben Bereich und die Hypothese von identischen Ratentrends in den drei Altersgruppen wird in beiden Fällen verworfen. Beim paarweisen Vergleich der drei Altersgruppen, der inhaltlich eigentlich noch spannenderen Fragestellung, finden wir einen hochsignifikanten Unterschied zwischen

der Altersgruppe der unter 20-Jährigen und den beiden anderen Gruppen, während letztere sich nicht signifikant voneinander unterscheiden ($p=0.875$). Vom Rückgang der Inzidenz von berufsbedingten Hautkrankheiten bei Friseuren scheinen also vor allem die Auszubildenden zu profitieren.

In Abbildung 1 sind auch die aus dem logistischen Modell geschätzten Inzidenzverläufe für die drei Altersgruppen aufgetragen. Da im logistischen Regressionsmodell die Effekte linear auf die Logits, nicht aber auf die Wahrscheinlichkeiten wirken, sind die geschätzten Inzidenzverläufe nur in Spezialfällen linear. Da jedoch auch der herkömmliche CA-Test nicht notwendig lineare Trends voraussetzt, ist zu erwarten, dass auch die Verallgemeinerung des CA-Tests beim Vergleich der Trends zu verlässlichen Ergebnissen führt.

Die Ergebnisse der paarweisen Vergleiche der Trends sind optisch gut nachzuvollziehen. Die Verläufe in den beiden älteren Altersgruppen sind nahezu parallel, während der Abfall in der jüngsten Gruppe viel steiler als in den beiden anderen Gruppen ist.

4 Diskussion/Ausblick

Es wurde gezeigt, wie durch die Interpretation des Cochran-Armitage-Tests als Score-Test in einem logistischen Regressionsmodell dieser ganz leicht auf den Vergleich von Trends in mehreren Gruppen zu verallgemeinern ist. Da leider bisher keine Erkenntnisse über die mathematisch-statistischen Eigenschaften dieser Verallgemeinerung vorliegen, ist eine Simulationstudie in Planung, die einen Eindruck von der Einsetzbarkeit und der Zuverlässigkeit dieses Vorschlages vermitteln wird.

Literatur

- [1] Dickel, H., Kuss, O., Blesius, C.R., Schmidt, A., Diepgen, T.L. (2001). Occupational skin diseases in Northern Bavaria between 1990 and 1999: a population-based study. *Br. J. Dermatol.*, **145**, 453-462.
- [2] Dickel, H., Kuss, O., Schmidt, A., Diepgen, T.L. (2001). Inzidenz berufsbedingter Hautkrankheiten in hautgefährdenden Berufsordnungsgruppen. *Der Hautarzt*, **52**, 615-623.
- [3] Arrighi, H.M., Hertz-Picciotto, I. (1994). The evolving concept of the healthy worker survivor effect. *Epidemiology*, **5**, 189-196.

- [4] McOsker, D.E., Beck, L.W. (1967). Characteristics of accommodated (hardened) skin. *Journal of Investigative Dermatology*, **48**, 372-383.
- [5] Cochran, W.G. (1954). Some methods of strengthening the common χ^2 tests. *Biometrics*, **10**, 417-451.
- [6] Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, **11**, 375-386.
- [7] Williams, P.L. (1997). Trend tests for counts and proportions. In: Armitage, P., and Colton, T., [eds.] *Encyclopedia of Biostatistics*, Section 10: Categorical Data Analysis.
- [8] Tarone, R.E., Gart, J.J. (1980). On the robustness of combined tests for trends in proportions. *Journal of the American Statistical Association*, **75**, 110-116.
- [9] Cox, D.R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society, Series B*, **205**, 215-242.
- [10] Margolin, B.H. (1988). Test for trend in proportions. In: Kotz, S., Johnson, N.L., Read, C.B., [eds.] *Encyclopedia of Statistical Sciences*, **9**, 334-336.
- [11] SAS Institute Inc. (2001). Documentation Example 1 for PROC MULTTEST. http://ftp.sas.com/techsup/download/sample/samp_lib/statsampDocumentation.Example.1_00000034.html
- [12] SAS Institute Inc. (2001). How do I compute a trend test for proportions? http://www.sas.com/service/techsup/faq/stat_proc/multproc981.html
- [13] Agresti, A. (1990). *Categorical data analysis*. John Wiley & Sons, New York.
- [14] Chen, J.J. (1993). Trend test for overdispersed proportions. *Biometrical Journal*, **35**, 949-958.
- [15] Lloyd, C.J. (1999). *Statistical Analysis of Categorical Data*. John Wiley & Sons, New York.

