Biostatistics and Biometrics
Open Access Journal
ISSN: 2573-2633

Mini Review
Volume 8 Issue 4 - December 2018
DOI: 10.19080/BBOAJ.2018.08.555745

Biostat Biometrics Open Acc J

An Application of a New and Exact Generalized Chi-Square Test to Real Data

Sascha Worz[1]*

Chair for Agricultural Systems Engineering, Technical University of Munich, Germany

Submission: June 12, 2018; **Published:** December 07, 2018

***Corresponding author:** S Worz, Chair for Agricultural Systems Engineering, Technical University of Munich, Am Staudengarten 2, 85354 Freising, Germany

Abstract

In this short communication, a new and exact generalized chi-square test which generalizes the famous chi-squared test is applied to real data.

Keywords: Applied hypothesis testing; Application of a new generalized chi-square test to real data

Mathematics Subject Classification 2010 62G10

Introduction and Short Literature Review

The famous chi-squared test homogeneity test can only be applied to absolute frequencies and underlies some restrictions on the sample size, the absolute frequencies and the expected frequencies. The former should be greater than 30, the absolute frequencies should attain a minimum of 10 and the expected frequencies should be bounded from below by 1, where at least 80 percent of them should be greater than 5, see [1,2]. In this short communication, a new and exact homogeneity test which generalizes the well-known chi-square test is applied to real data for the first time [3]. It generalizes the famous chi-squared test by assuming that the real data are realizations $x_{ij},\ i=1,\dots m,\ j=1,\dots n, m,n\in\mathbb{N}$ of normally distributed random variables $X_{ij}:\Omega\to\mathbb{R}_{\geq 0}$ with unknown means $\mu_{ij}>0$, unknown variances $\sigma_{ij}^2>0$ and $\sum_{i=1}^m\sum_{j=1}^n X_{ij}>0$ whose squared sum obeys after standardization a chi-squared distribution with m,n degrees of freedom. Especially, the test does not underlie any restrictions as reported in [1, 2].

Methodology

For the whole short communication, assume $m,n\in\mathbb{N}$. Suppose that the normally distributed and random variables $X_{ij}:\Omega\to\mathbb{R}_{\geq 0}$ with unknown means $\mu_{ij}>0$, and unknown variances $\sigma_{ij}^2>0$ with $i=1,\dots,m, j=1,\dots,n$ are given, where the index i identifies the i-th sample and the index j identifies the j-th characteristic and in addition $\sum_{i=1}^m\sum_{j=1}^n X_{ij}>0$ should hold. Then, the means μ_{ij} and variances $\sigma_{ij}^2,\ i=1,\dots,m, j=1,\dots,n,$ have to be estimated from the available realizations $x_{ij},\ i=1,\dots,m, j=1,\dots,n.$ This can be done in the following way: Define the well-defined mean $\mu_{ij}>0$ for $i=1,\dots,m$ and $j=1,\dots,n$ by

$$\mu_{ij}=\sum_{k=1}^n x_{ik}\frac{\sum_{i=1}^n x_{ij}}{\sum_{i=1}^m\sum_{j=1}^n x_{ij}}$$

and the well-defined variances $\sigma_{ij}^2>0$ for $i=1,\dots,m, j=1,\dots,n$ by

$$\sigma_{ij}^2=\sum_{k=1}^n(x_{ik}-\mu_{ik})^2\frac{\sum_{i=1}^m x_{ij}}{\sum_{i=1}^m\sum_{j=1}^n x_{ij}}.$$

Next, define the test statistic

$$\sum_{i=1}^m\sum_{j=1}^n\left(\frac{X_{ij}-\mu_{ij}}{\sigma_{ij}}\right)^2\sim\chi_{mn}^2$$

which is on the one hand exactly chi-squared distributed with mn degrees of freedom and on the other hand represents the squared sum of absolute standardized mean percentage errors. Finally, the following hypothesis test is considered: Let the null hypothesis H_0 comprise the hypothesis that the characteristic random variables vectors $X_{i,j=1,\dots,n}, i=1,\dots,m$ are identically distributed and the alternative hypothesis H_1 the hypothesis that at least both of the characteristic random variables vectors $X_{i,j=1,\dots,n}, i=1,\dots,m$ are not identically distributed. Then, the null hypothesis H_0 is rejected or the alternative hypothesis H_1 is accepted if and only if

$$\sum_{i=1}^m\sum_{j=1}^n\left(\frac{x_{ij}-\mu_{ij}}{\sigma_{ij}}\right)^2\sim\chi_{mn,\alpha}^2, \qquad (1)$$

where $0<\alpha<1$ denotes the significance level. For details see [3].

Statistical example

Consider the following deviations from skew diameters in [m] originating from 5 different samples and 6 different production machineries in biomedical engineering (Table 1). Then, for $m=5$ and $n=6$ and a significance level $\alpha:=0.05$ it holds $\chi_{30,0.05}^2=43.77$ and calculating (1) yields that the alternative hypothesis H_1 cannot be accepted or the null hypothesis H_0 cannot be rejected since

$$\chi^2_{30,0.05} = 43.77 > \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\frac{x_{ij} - \mu_{ij}}{\sigma_{ij}} \right)^2 = 24.33952.$$

Table 1: Deviations of skew diameters in [m].

Samples/Production Machineries					
0.00009	0.00930	0.01974	0.00028	0.00515	0.00211
0.00016	0.00886	0.01476	0.00015	0.00266	0.00109
0.00007	0.00940	0.00892	0.00016	0.00283	0.00116
0.00010	0.01117	0.01770	0.00018	0.00319	0.00131
0.00009	0.00932	0.01553	0.00015	0.00280	0.00115

Conclusion

In this short communication, a generalized chi-squared test which does not underlie any statistical parameter restrictions was applied for the first time to real data by considering a computational example which cannot be treated with the common chi-squared test.

Acknowledgement

This research work is/was funded by Bundesministerium fur Landwirtschaft und Erna¨hrung, Deichmanns Aue 29, 53168 Bonn.

References

1. Rinne H (2003) Taschenbuch der Statistik. 3. Auage. Verlag Harri Deutsch Pp. 562.
2. Vo W (2000) Taschenbuch der Statistik. 1. Auage. Fachbuchverlag Leipzig, Pp. 447.
3. Worz S, Hijazi O (2002) Bernhardt: A new and exact homogeneity test for real data. Journal of the American Statistical Association.

DOI: 10.19080/BBOAJ.2018.08.555746

How to cite this article: Sascha W. An Application of a New and Exact Generalized Chi-Square Test to Real Data. Biostat Biometrics Open Acc J. 2018; 8(5): 555746. DOI: 10.19080/BBOAJ.2018.08.555746.