# Bortkewitsch's horse-kicks and the generalised linear model

D. A. PREECE*, G. J. S. ROSS & S. P. J. KIRBY†

*Institute of Arable Crops Research, Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ, U.K.*

**Abstract.** Many a student of statistics has had his introduction to the Poisson distribution enlivened by reference to 'Bortkewitsch's data' on numbers of Prussian *Militärpersonen* killed by horse-kicks. However, these data, which in their fullest form are for each of 14 corps in each of 20 successive years, show differences between corps and between years, as is indicated both by classical Pearson $X^2$ tests and by an analysis based on a generalised linear model with Poisson 'errors' and a logarithmic link function. This heterogeneity calls in question the fit of a single Poisson distribution to the 280 counts of numbers of deaths. Yet the fit of a single Poisson distribution to the frequency distribution of the data, as judged by the Pearson $X^2$ test, is good. This note shows that the Negative binomial distribution may be derived as a model for the counts; the similarity of this distribution to the Poisson for modest heterogeneity is well known.

Ladislaus von Bortkewitsch (1868–1931) was born in St Petersburg (now Leningrad) of Polish ancestry (see Haight, 1967, pp. 114–117; Gumbel, 1968). After studies in Russia, he went to Germany. He received a Ph.D. from the University of Göttingen, where he was a student of Wilhelm Lexis, to whom he later dedicated his monograph *Das Gesetz der kleinen Zahlen* [The Law of Small Numbers] (Bortkewitsch, 1898). In his publications, his name was variously spelt in the German form *Bortkewitsch* and the Polish form *Bortkiewicz*.

The 1898 monograph gives the famous horse-kick data (Bortkewitsch, 1898, pp. 23–25) whose analysis has featured in so many text-book accounts of the Poisson distribution. (Johnson & Kotz, 1969, p. 88, described the horses as 'mules'; we do not know why). The full data-table shows how many Prussian *Militärpersonen* died from horse-kicks in each of the 14 corps in each of the 20 successive years 1875 to 1894; the 14 corps were 13 *Armeecorps* (denoted by the Roman numerals, I, II, . . ., XI, XIV, XV) plus the *Gardecorps* (denoted by 'G'). The table—with dashes replaced by the zeroes that they stand for, and with marginal totals inserted—is reproduced as Table 1 of this paper. (See also p. 157 of Winsor, 1947; Table 9.5-2 of Bishop, Fienberg & Holland, 1975; and Table 4.1 of Andrews & Herzberg, 1985).

Table 2 gives the frequency distribution for the 280 counts from the body of Table 1. This distribution was introduced to English readers by Keynes (1921, p. 402) and by Jeffreys (1939, p. 58); the former also reproduced the original two-way table of numbers of deaths, which total to 196.

Present address: *Institute of Horticultural Research, East Malling, Maidstone, Kent ME19 6BJ, and †42/8 Craighouse Gardens, Edinburgh EH10 5LW.

**Table 1.** Bortkewitsch's data table, giving numbers of deaths from horse-kicks

| Corps | Year | | | | | | | | | | | | | | | | | | | | Totals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1875 | 1876 | 1877 | 1878 | 1879 | 1880 | 1881 | 1882 | 1883 | 1884 | 1885 | 1886 | 1887 | 1888 | 1889 | 1890 | 1891 | 1892 | 1893 | 1894 | |
| G | 0 | 2 | 2 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 0 | 2 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 16 |
| I | 0 | 0 | 0 | 2 | 0 | 3 | 0 | 2 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 0 | 3 | 1 | 0 | 16 |
| II | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 2 | 0 | 0 | 12 |
| III | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 12 |
| IV | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 8 |
| V | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 11 |
| VI | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 2 | 0 | 1 | 1 | 3 | 1 | 1 | 1 | 0 | 3 | 0 | 0 | 17 |
| VII | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 2 | 0 | 0 | 2 | 1 | 0 | 2 | 0 | 12 |
| VIII | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 7 |
| IX | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | 2 | 1 | 1 | 0 | 1 | 2 | 0 | 1 | 0 | 0 | 13 |
| X | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 1 | 3 | 0 | 1 | 1 | 15 |
| XI | 0 | 0 | 0 | 0 | 2 | 4 | 0 | 1 | 3 | 0 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 1 | 3 | 1 | 25 |
| XIV | 1 | 1 | 2 | 1 | 1 | 3 | 0 | 4 | 0 | 1 | 0 | 3 | 2 | 1 | 0 | 2 | 1 | 1 | 0 | 0 | 24 |
| XV | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 8 |
| Total | 3 | 5 | 7 | 9 | 10 | 18 | 6 | 14 | 11 | 9 | 5 | 11 | 15 | 6 | 11 | 17 | 12 | 15 | 8 | 4 | 196 |

**Table 2.** Frequency distributions for Bortkewitsch's full table

| Number of deaths | Observed frequency | Expected frequency, as given by Bortkewitsch and by Keynes | Expected Poisson frequency, as given by Jeffreys |
|---|---|---|---|
| 0 | 144 | 143·1 | 139·0 |
| 1 | 91 | 92·1 | 97·3 |
| 2 | 32 | 33·3 | 34·1 |
| 3 | 11 | 8·9 | 8·0 |
| 4 | 2 | 2·0 | 1·4 |
| 5+ | 0 | 0·6 | 0·2 |
| Total | 280 | 280·0 | 280·0 |

If each of the 280 counts of numbers of deaths could reasonably be thought to be independent of all the others, and the number of cavalry officers and their susceptibility to death from horse-kicks could reasonably be thought to be the same for each of the 280 units of observation, then a simple Poisson model for the observed frequencies would be reasonable. The expected frequencies for a Poisson distribution with mean $196/280=0{\cdot}700$ were given by Jeffreys and are reproduced here in Table 2; these show good agreement with the observed frequencies. Table 2 also reproduces the expected frequencies given by Bortkewitsch and quoted by Keynes; these were obtained by fitting a Poisson model to the data for each corps and then summing the expected values across the corps (e.g. Winsor, 1947, p. 158).

Bortkewitsch (1898, p. 24) noted that the four corps denoted G, I, VI and XI had numerical compositions that were particularly far from the average. He therefore excluded these four corps, to give the observed frequencies in our Table 3, for which the total number of deaths is 122.

Table 3 also contains the expected Poisson frequencies as obtained by Bortkewitsch himself and by Fisher (1925, Section 15, Table 4) for a Poisson distribution with mean $122/200=0{\cdot}610$. The agreement between observed and expected is very good indeed for the smaller data-set.

**Table 3.** Frequency distribution excluding corps G, I, VI and XI

| Number of deaths | Observed frequency | Expected Poisson frequency as obtained by Bortkewitsch and Fisher |
|---|---|---|
| 0 | 109 | 108·7 |
| 1 | 65 | 66·3 |
| 2 | 22 | 20·2 |
| 3 | 3 | 4·1 |
| 4 | 1 | 0·6 |
| 5+ | 0 | 0·1 |
| Total | 200 | 200·0 |

However, a generalised linear model with logarithmic link function and Poisson errors for the observations and with terms for corps and years may be fitted to the corps-by-years table of counts. Goodness-of-fit for these two terms may be summarised by an analysis of deviance (McCullagh & Nelder, 1983, p. 17).

| Term | Deviance | d.f. | Mean deviance | Mean deviance ratio |
|---|---|---|---|---|
| Corps | 26·1 | 13 | 2·01 | 1·91 |
| Years | 38·5 | 19 | 2·03 | 1·93 |
| Residual | 258·6 | 247 | 1·05 | |

This analysis is obtainable from the computer program Genstat's facilities for generalised linear models by specifying Poisson 'errors' and a logarithmic link function (Alvey *et al.*, 1983).

The order of fitting the terms corps and years makes no difference to the deviances accounted for here, the two terms being orthogonal for the present model (a property which does not generally hold for analyses of higher-dimensional tables with interaction terms). On the assumption that the mean deviance ratio is approximately distributed as $F$, there is evidence of differences between both corps and years.

The years do not show an obvious trend except for the first 6 years, when there was a consistent increase in deaths. Corps G, I, VI and XI, which were noted as having a numerical composition particularly far from the average, have four of the five highest counts of deaths, the other corps with a high count being XIV.

The deviances, 26·1 and 38·5 for corps and years respectively, are close to the corresponding Pearson $X^2$ test statistics, 27·3 and 37·5 respectively, of Uniform distributions for the corps and years totals (these Pearson $X^2$ test statistics were calculated by Bishop, Fienberg & Holland, 1975, p. 325). This is, of course, to be expected because the deviance and the Pearson $X^2$ statistic differ only in terms of third and higher powers of the residuals, which are here relatively small.

The chi-squared distribution used to test the significance of the residual deviance is an approximation, and the derived probability levels may be too large or too small. In the present case the difference is not expected to be serious and the residual deviance thus indicates a reasonable fit.

Fitting a log-linear model to the data-set without the four anomalous corps G, I, VI and XI gives the following analysis of deviance:

| Term | Deviance | d.f. | Mean deviance | Mean deviance ratio |
|---|---|---|---|---|
| Corps | 15·6 | 9 | 1·73 | 1·73 |
| Years | 25·3 | 19 | 1·33 | 1·33 |
| Residual | 171·6 | 171 | 1·00 | |

This indicates that there is still evidence of differences between corps. Again, the residual deviance indicates a satisfactory model.

A general point worth discussing, therefore, is why a single Poisson distribution fits so well when heterogeneity seems very likely.

An answer to this question may be found by noting that the 280 fitted values from the generalised linear model for the full data-set with terms for corps and years may themselves be fitted by a Gamma distribution (see Quine & Seneta, 1987, for another answer based on the large number of cells with very few deaths). This fitting can be done by MLP (the Maximum Likelihood Program; Ross, 1980) and gives a residual deviance chi-squared value of 8·26 with 9 degrees of freedom when the 280 fitted means are grouped into the following intervals:

$<$0·20 0·20–0·39 0·40–0·59 0·60–0·79 0·80–0·99 1·00–1·19
1·20–1·39 1·40–1·59 1·60–1·79 1·80–1·99 2·00–2·19 $\geqslant$2·20

Now it is well known that a Gamma compound Poisson distribution gives a Negative binomial distribution (e.g. Plackett, 1981). We may therefore regard the Negative binomial distribution rather than the Poisson as an appropriate model for the distribution of the data. The similarity of the Negative binomial to the Poisson for modest heterogeneity is also well known (McCullagh & Nelder, 1983, p. 132).

For the full data-set, the maximum-likelihood estimate of $k$ from fitting the Negative binomial distribution,

$$pr(X=x;\ a,\ k)=(x+k-1)!a^x/x!(k-1)!\ (1+a)^{x+k}$$

is 7·6. For the reduced data-set, $k$ becomes negative because the estimated variance is less than the estimated mean, see e.g. Ross & Preece (1985).

The power of the Pearson $X^2$ test when fitting a Poisson distribution to frequency distributions of data with a Negative binomial distribution with mean=0·7 and $k=7{\cdot}6$ was assessed by simulation. Only 67 out of 500 data-sets, each with 280 observations, were correctly rejected at the 5% level as being inconsistent with the Poisson model.

What sample size, therefore, would have been adequate to detect heterogeneity for this test using the frequency distribution of the data? An empirical approach to this problem is to consider a large sample of size $N$ assumed to have a Negative binomial distribution with mean 0·7 and $k=7{\cdot}6$. The probablilities for the Poisson and Negative binomial distributions are given in Table 4.

**Table 4.** Poisson probabilities for a distribution with a mean of 0·7 and Negative binomial probabilities for a distribution with a mean of 0·7 and $k=7{\cdot}6$

| Number | Poisson probability | Negative binomial probability |
|---|---|---|
| 0 | 0·49659 | 0·51191 |
| 1 | 0·34761 | 0·32811 |
| 2 | 0·12166 | 0·11899 |
| 3 | 0·02839 | 0·03211 |
| 4 | 0·00497 | 0·00718 |
| 5+ | 0·00079 | 0·00170 |

Because the Pearson $X^2$ statistic for the fit of a Poisson distribution to the frequency distribution of data is given by

$$\Sigma[(y-\hat{\mu})^2/\hat{\mu}]$$

where $y$ is the observed frequency and $\hat{\mu}$ is the expected frequency for the Poisson model, an exact fit of the Negative binomial distribution would give

$$\Sigma N[(p_{nb}-p_p)^2/p_p]$$

where $p_{nb}$ and $p_p$ are the probabilities for the Negative binomial and Poisson distributions.

This expression is equal to

$$\begin{aligned}&N[(0{\cdot}51191-0{\cdot}49659)^2/0{\cdot}49659]+\\&N[(0{\cdot}32811-0{\cdot}34761)^2/0{\cdot}34761]+\\&N[(0{\cdot}11899-0{\cdot}12166)^2/0{\cdot}12166]+\\&N[(0{\cdot}03211-0{\cdot}02839)^2/0{\cdot}02839]+\\&N[(0{\cdot}00718-0{\cdot}00497)^2/0{\cdot}00497]+\\&N[(0{\cdot}00170-0{\cdot}00079)^2/0{\cdot}00079]\\&=0{\cdot}004144N\end{aligned}$$

0.004144 * 280 = 1.16
0.004144 * 927 = 3.84 critical value at df = 1

If this value is compared with 3·84, the 5% value of $\chi^2$ on 1 degree of freedom (from the extra parameter in the Negative binomial distribution compared with the Poisson distribution), we find that a sample size of 927 is required. This number is, of course, only a measure of location, or a typical sample size, but it does reinforce the point that Bortkewitsch's data were too few to detect heterogeneity using the Pearson $X^2$ test for the fit of a single Poisson distribution to the frequency distribution of the data.

## References

ALVEY, N.G. *et al.* (1983) *Genstat: A General Statistical Program, a manual*, Chapter 7 (Harpenden, Rothamsted Experimental Station).

ANDREWS, D.F. & HERZBERG, A.M. (1985) The number of deaths by horsekicks in the Prussian Army, in: *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, pp. 17–18 (New York, Springer-Verlag).

BISHOP, Y.M.M., FIENBERG, S.E. & HOLLAND, P.W. (1975) *Discrete Multivariate Analysis: theory and practice* (Cambridge, MA, MIT Press).

BORTKEWITSCH, L. VON (1898) *Das Gesetz der kleinen Zahlen* (Leipzig, B. G. Teubner).

FISHER, R.A. (1925) *Statistical Methods for Research Workers* (Edinburgh, Oliver & Boyd).

GUMBEL, E.J. (1968) Bortkiewicz, Ladislaus von, in: SILLS, D.L. (Ed.) *International Encyclopedia of the Social Sciences*, Vol. 2, pp. 128–131 (New York, Macmillan).

HAIGHT, F.A. (1967) Handbook of the Poisson distribution, *Operations Research Society of America, Publications in Operations Research No. 11*, pp. 14–117 (New York, Wiley).

JEFFREYS, H. (1939) *Theory of Probability* (Oxford, Clarendon Press).

JOHNSON, N.L. & KOTZ, S. (1969) *Distributions in Statistics: discrete distributions* (New York, Houghton Mifflin).

KEYNES, J.M. (1921) *A Treatise on Probability* (London, Macmillan).

MCCULLAGH, P. & NELDER, J.A. (1983) *Generalized Linear Models* (London, Chapman & Hall).

PLACKETT, R.L. (1981) *The Analysis of Categorical Data* (London, Griffin).

QUINE, M.P. & SENETA, E. (1987) Bortkiewitcz's data and the law of small numbers, *International Statistical Review*, 55, pp. 173–181.

ROSS, G.J.S. (1987) MLP: *Maximum Likelihood Program*, User Manual (Oxford, Numerical Algorithms Group).

ROSS, G.J.S. & PREECE, D.A. (1985) The negative binomial distribution, *The Statistician*, 34, pp. 323–336.

WINSOR, C.P. (1947) Quotations: Das Gesetz der kleinen Zahlen, *Human Biology*, 19, pp. 154–161.