1. **Goodness of fit**. In 1898, Ladiaslus Bortkiewicz published *Das Gesetz der Kleinen Zahlen* [The Law of Small Numbers] in which he first showed how the Poisson distribution occurs in real life. This table is his most famous example; it counts the number of deaths due to horse kicks in 14 Prussian army corps over a period of 20 years.

Jahres-ergebnis	Zahl der Fälle, in denen das nebenstehende Jahresergebnis	
	eingetreten ist	zu erwarten war
0	144	143,1
1	91	92,1
2	32	33,3
3	11	8,9
4	2	2,0
5 u. mehr	—	0,6

Yearly result	Number of years in which the result was	
	observed	expected
0	144	143.1
1	91	92.1
2	32	33.3
3	11	8.9
4	2	2.0
5+	-	0.6

For instance, there were 32 times where exactly two deaths occured in a year in one corps. Bortkiewicz compared this distribution to a Poisson distribution with a parameter that he estimated from the data.

(a) How many degrees of freedom are there?

(b) Set up a test statistic for goodness of fit. You don't need to compute it, an estimate is enough.

(c) How good is the fit? Are there any problems with the test? How could these be resolved?

(d) Give a rough guess as to the p-value.

Here's the relevant section of a chi-square table:

DOF	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81

Solutions

(a)

We have six categories, so typically that would mean $6 - 1 = 5$ degrees of freedom. But we lose a degree of freedom by estimating the parameter. So there are really four degrees of freedom.

(b)

The test statistic is

$$\sum_{\text{categories}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$
$$= \frac{(144-143.1)^2}{143.1} + \frac{(91-92.1)^2}{92.1} + \frac{(32-33.3)^2}{33.3} + \frac{(8.9-11)^2}{11} + \frac{(2-2)^2}{2} + \frac{(0-0.6)^2}{0.6}$$
$$\approx 1.07$$

(c/d)

By the table, this corresponds to a p-value of about 0.9. So we don't reject the null hypothesis, which was that the data are Poisson-distributed. That means the fit is very good.

Note that we do have a category (five or more deaths) where the expected number is very small, and that piece does contribute most of the test statistic. This might be a problem, but we don't really mind in this case as it doesn't lead to a rejection.