Solutions to Methods in Spring 2007, Jan 4 2007, 8:00-12:00

1. Here we assume $n_i \sim Poisson(\theta x_i)$. Then, the MLE of θ is

$$\hat{\theta} = \frac{\sum_{i=1}^{5} n_i}{\sum_{i=1}^{5} x_i} = 0.0001414$$

Thus, the predicted values of n_i ($\hat{n}_i$) are 34.96, 134.13, 204.10, 109.81, 10.00. The Pearson χ^2 statistic is

$$X^2 = \sum_{i=1}^{5} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i} = 220.6$$

and the deviance goodness of fit is

$$G^2 = 2\sum_{i=1}^{5} n_i \log \frac{n_i}{\hat{n}_i} = 142.2.$$

2. (a) Using the concept of odds ratios. The risks of nodal involvement increases $e^{1.801} - 1 = 4.0557$ times if the X-ray reading is serious.

   (b) This is the deviance goodness of fit which is 47.611 in the output.

   (c) For all the three predictor variables, they assign 1 to more serious cancer and 0 to less serious cancer. If one of them shows serious, then the others also show serious.

   (d) The null model does not fit the data well at significance level 0.05.

   (e) First, we calculate $e^{-3.079-0.292} = 0.034$ and then we calculate the probability as $0.034/(1 + 0.034) = 0.033$.

3. Let y be HOURS1 and x be ORDERS.

   (a) The regression line is

$$y = 514.37 + 2.8932x.$$

   The company needs 2.8932 hours more to handle one orders increases per month. When the ORDER is 0, it needs 514.37 HOURS. This is the starting point of HOURS1 for the ORDER. It may indicate the technique part for the handling.

   (b) The model only fits when ORDERs is not 0. When ORDERs is 0, it is out of the bound of the data. Thus, the model can not predict the case when ORDERs is small.

   (c) Note that the t-value is

$$\frac{2.89315 - 2.4}{0.33628} = 1.466$$

   the p-value is between 0.05. Thus, we can say that the statement of the CEO is not significant violated from the data.

(d) Model assumption: linearly dependent, normally distributed with constant variance of the error term. Diagnostic method: qq-plot, residual plot, partial residual plot and etc. You need to explain what they are.

(e) This method does not work since they are always not correlated

(f) The predicted value is

$$\hat{y} = 514.37 + 2.8932(2800) = 8615.33.$$

The variance of the mean of the predicted value is

$$V(\hat{y}) = (1, 2800) \begin{pmatrix} 807759 & -276.34 \\ -276.34 & 0.1131 \end{pmatrix} \begin{pmatrix} 1 \\ 2800 \end{pmatrix} = 146959.$$

Thus, the 95% confidence interval of the expected number of work hours is

$$8615.33 \pm t_{0.025,14}\sqrt{146959} = [7793, 9437].$$

In this problem, it asks the confidence interval for the real work hours which is

$$8615.33 \pm t_{0.025,15}\sqrt{146959 + 1455^2} = [5408, 11822].$$

4. In this study, we have 6 patients, each with one lobule, and the counts of grade according 1, 2, 3 and 4 are recorded.

   (a) The table is 6×4, where 6 rows indicate 6 different patients, 3 columns indicate three polarity makers and 4 heights indicate the index of grades of acinus. The counts of different patient and different of grades are given in the table.

   (b) i. Yes. Suppose we already known the information about the population of normal tissue. We want to compare the proportion of the counts with the proportion of counts of normal tissue. We can carry out a χ^2 test for the proportion of the samples of each patients with the normal tissues. The χ^2 test could be either a Pearson or a deviance goodness of fit statistic.

      ii. Yes. Collapse into a 3×4 table and see which combination is dominant. We assume they are equally likely and then calculate the p-value for the equally likely assumption.

      iii. No. Since in this example, one patient only contributes one lobule. If we want to answer this, we need to get several lobules for one patient.

      iv. Yes. We can look at whether the polarity maker effect is significant.

5. (a) The residual degree of freedom is 16.

(b) It is clear that 346.37 is too large. Thus they are not independent.

(c) It is not reasonable since the Brand is a nominal variable.

(d) Look at the data. It shows that people like to choose the same brand.

(e) The fit is much better. At level 0.01, we accept the symmetry model.

(f) Since the table is not symmetric, the nonzero appears at the non-symmetric part.

(g) The pair is (High Point, Sanka).