Statistical Models STAT2011

Semester 1, 2012

Week 9 Computer Exercise

Poisson fitting and checking

1. A diluted suspension of yeast cells was spread on a microscopic slide divided into 400 squares of equal areas. The number of cells per square gave a sample of 400 counts (Student (1907)).

2	2	4	4	4	5	2	4	7	7	4	7	5	2	8	6	7	4	3	4
3	3	2	4	2	5	4	2	8	6	3	6	6	10	8	3	5	6	4	4
7	9	5	2	7	4	4	2	4	4	4	3	5	6	5	4	1	4	2	6
4	1	4	7	3	2	3	5	8	2	9	5	3	9	5	5	2	4	3	4
4	1	5	9	3	4	4	6	6	5	4	6	5	5	4	3	5	9	6	4
4	4	5	10	4	4	3	8	3	2	1	4	1	5	6	4	2	3	3	3
3	7	4	5	1	8	5	7	9	5	8	9	5	6	6	4	3	7	4	4
7	5	6	3	6	7	4	5	8	6	3	3	4	3	7	4	4	4	5	3
8	10	6	3	3	6	5	2	5	3	11	3	7	4	7	3	5	5	3	4
1	3	7	2	5	5	5	3	3	4	6	5	6	1	6	4	4	4	6	4
4	2	5	4	8	6	3	4	6	5	2	6	6	1	2	2	2	5	2	2
5	9	3	5	6	4	6	5	7	1	3	6	5	4	2	8	9	5	4	3
2	2	11	4	6	6	4	6	2	5	3	5	7	2	6	5	5	1	2	7
5	12	5	8	2	4	2	1	6	4	5	1	2	9	1	3	4	7	3	6
5	6	5	4	4	5	2	7	6	2	7	3	5	4	4	5	4	7	5	4
8	4	6	6	5	3	3	5	7	4	5	5	5	6	10	2	3	8	3	5
6	6	4	2	6	6	7	5	4	5	8	6	7	6	4	2	6	1	1	4
7	2	5	7	4	6	4	5	1	5	10	8	7	5	4	6	4	4	7	5
4	3	1	6	2	5	3	3	3	7	4	3	7	8	4	7	3	1	4	4
7	6	7	2	4	5	1	3	12	4	2	2	8	7	6	7	6	3	5	4

The data appears on the course webpage and can be read into R as a vector using the command

```
yeast=scan(url("http://maths.usyd.edu.au/stat2011/r/yeast.txt"))
```

(Questions appear over the page...)

(a) Obtain the mean and call it `mn`. This serves as an estimate of the Poisson parameter.

(b) Obtain a vector of Poisson probabilities with parameter `mn` for 0:11. Append to this the corresponding Poisson probability for 12 or more (a command like `c(prob,1-sum(prob))` will do this).

(c) Obtain from this vector of probabilities a vector of expected (Poisson) frequencies `pois.f`.

(d) Obtain a corresponding vector of *observed* frequencies `f`; you can use `table()` but you will need to append a zero to the beginning.

(e) Hence obtain a vector `sr` of standardised residuals (consult the handouts accompanying lectures in weeks 8 and 9 for some examples), and display all 3 vectors using `cbind(f,pois.f,sr)`. Comment on how the standardised residuals indicate a good or poor fit.

2. The data below relate to testing milk for bacterial contamination. 0.01 mL of milk is spread over a 1cm^2 slide which is mounted on a microscope slide. A 20-by-20 grid is overlaid on the slide and the count of bacterial clumps in each grid-square is obtained. The counts for such a test are summarised in the frequency table below.

Count	0	1	2	3	4	5	6	7	8	9	10	19
Freq.	56	104	80	62	42	27	9	9	5	3	2	1

We are modelling the original 400 counts $y_1, y_2, \ldots, y_{400}$ as being the observed values of 400 i.i.d. Poisson(λ) random variables $Y_1, Y_2, \ldots, Y_{400}$.

(a) Compute the mean of the *original* counts. Hence obtain an estimate of the Poisson parameter for this data.

(b) Obtain fitted Poisson probabilities and hence expected frequencies for each cell in the above table implied by our fitted Poisson model (regard the column labelled '19' as representing the range of values $\{11, 12, \ldots\}$).

(c) Compute standardised residuals for each category, present in a matrix using `cbind` as for the yeast data, and comment on the goodness (or poorness) of fit.

3. Compute the sample variance in both cases (i.e. questions 1 and 2) and comment on how comparing it to the sample mean complements the analysis in each case.