Semester 2	Applied Statistics	2008

Exercise 6 Solution

Tutorial Exercise.

1. (a) $16 * 125.3875^2 + ... + 14 * 157.3071^2 - (16 * 125.3875^2 + ... + 14 * 157.3071)^2/67$
   (b) 3
   (c) $P(\chi^2_{15} \geq 63 * 10.142) = 0$ so there is strong evidence of a difference.
   (d) The first canonical variable separates all groups and the second separates group 3 from the others.
   (e) $P(\chi^2_5 \geq 63 * 2.81) = 0$ so there is strong evidence that all three dimensions are required.

2. (a) We have

$$\bar{y} = \frac{\sum_i y_i}{n} = \frac{177}{10} = 17.7,\ \bar{x} = \frac{\sum_i x_i}{n} = \frac{179}{10} = 17.9 \text{ and } \overline{X} = \frac{X}{N} = \frac{16,300}{1,000} = 16.3.$$

The ordinary estimate and ratio estimate of the average number of man-hours lost Y because of sickness this year are respectively:

$$\widehat{\overline{Y}} = \bar{y} = 17.7 \quad \text{and} \quad \widehat{\overline{Y}}_r = \overline{X}\frac{\bar{y}}{\bar{x}} = 16.3 \times \frac{17.7}{17.9} = 16.12$$

The discrepancy lies in the ratio of the population mean to sample mean,

$$\frac{\overline{X}}{\bar{x}} = \frac{16.3}{17.9} = 0.911$$

which is quite close to 1 showing that the discrepancy is not large.

   (b) Ratio estimator for the total number of man-hours lost Y because of sickness this year:

$$\begin{aligned}
\widehat{Y}_r &= Xr = X \times \frac{\sum_i y_i}{\sum_i x_i} = 16300 \times \frac{177}{179} = 16300(0.9888) = 16117.877 \\
s_r^2 &= \frac{1}{n-1}(\sum_i y_i^2 - 2r\sum_i x_i y_i + r^2 \sum_i x_i^2) \\
&= \frac{1}{9}\left[3,843 - 2 \times \frac{177}{179} \times 3,927 + \left(\frac{177}{179}\right)^2 \times 4,067\right] = 5.9310 \\
se(\widehat{Y}_r) &= N\sqrt{\left(1 - \frac{n}{N}\right)\frac{s_r^2}{n}} = 1,000\sqrt{\left(1 - \frac{10}{1,000}\right)\frac{5.9310}{10}} = \sqrt{587,171.562} \\
&= 766.2712
\end{aligned}$$

(c) Regression estimator for the total number of man-hours lost Y because of sickness this year:

$$\begin{aligned}
\sum_i x_i y_i - n\overline{x}\overline{y} &= 3,927 - 10 \times 17.9 \times 17.7 = 758.7 \\
\sum_i x_i^2 - n\overline{x}^2 &= 4,067 - 10 \times 17.9^2 = 862.9, \\
\sum_i y_i^2 - n\overline{y}^2 &= 3,843 - 10 \times 17.7^2 = 710.1.
\end{aligned}$$

We have

$$\begin{aligned}
s_y^2 &= \frac{1}{n-1}\left(\sum_i y_i^2 - n\overline{y}^2\right) = \frac{710.1}{9} = 78.9 \\
\hat{\rho} &= \frac{\sum_i x_i y_i - n\overline{x}\overline{y}}{\sqrt{(\sum_i x_i^2 - n\overline{x}^2)(\sum_i y_i^2 - n\overline{y}^2)}} = \frac{758.7}{\sqrt{862.9 \times 710.1}} = 0.9692 \\
s_{reg}^2 &= s_y^2(1-\hat{\rho}^2) = 78.9(1-0.9692^2) = 4.7854, \\
b &= \frac{\sum_i x_i y_i - n\overline{x}\overline{y}}{\sum_i x_i^2 - n\overline{x}^2} = \frac{758.7}{862.9} = 0.8792 \\
\widehat{Y}_{reg} &= N[\overline{y} + b(\overline{X} - \overline{x})] \\
&= 1,000\left[17.7 + 0.8792\left(\frac{16,300}{1,000} - 17.9\right)\right] = 16,293.209 \\
se(\widehat{Y}_{reg}) &= 1,000\sqrt{\left(1-\frac{n}{N}\right)\frac{s_{reg}^2}{n}} \\
&= 1,000\sqrt{\left(1-\frac{10}{1,000}\right)\frac{4.7854}{10}} = \sqrt{473189.9408} = 687.888 < 766.2712.
\end{aligned}$$

The estimator $\widehat{Y}_{reg}$ is more efficient than $\widehat{Y}_r$ because the assumption that the linear relationship between x and y passes through the origin can be dropped. Since $\hat{\rho} = 0.9692$ is positive and close to 1 which shows a strong and positive relationship between x and y, the use of x as an auxiliary variable is suitable.

(d) The rate of change of man-hours lost is

$$\text{Rate of change} = \frac{\text{Lost in current year} - \text{lost in last year}}{\text{Lost in last year}} = \frac{\text{Lost in current year}}{\text{Lost in last year}} - 1.$$

Hence the ratio estimator is $R' = R - 1$ and hence s_r^2 in part (a) is used for S^2. The error bound for R' is $\delta\mu_{r'} = 0.05$. Hence the error bound for R is $\delta\mu_r = 0.05$ and the error bound for the mean estimate is

$$\delta\mu_{\bar{X}r} = \overline{X}(\delta\mu_r) = \frac{16,300}{1,000} \times 0.05 = 0.815,$$

i.e. $\delta\mu = 0.815$.

$$n = \frac{NS_r^2}{N(\delta\mu)^2/z_{\alpha/2}^2 + S_r^2} = \frac{1,000 \times 5.931}{1,000 \times \frac{0.815^2}{1.96^2} + 5.931} = 33.16.$$

Take $n = 34$.