統計應用方法

Homework 1

1. (55 points)

Bird colonies. One of nature's patterns connects the percent of adult birds in a colony that return from the previous year and the number of new adults that join the colony. Here are data for 13 colonies of sparrowhawks:[2]

Percent returning	New adults	Percent returning	New adults	Percent returning	New adults
74	5	62	15	46	18
66	6	52	16	60	19
81	8	45	17	46	20
52	11	62	18	38	20
73	12				

a. (10 points) Please plot the data with X indicating "Percent returning" and Y indicating "New adults". Comment on the main features of the plot. Any possible outliers? (可用軟體畫)

b. (10 points) Apply the formula to compute $\bar{X}$, $\bar{Y}$, $S_x^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$, $S_y^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)$ and Pearson's correlation:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^{n}X_iY_i - n\bar{X}\bar{Y}}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

可以用電腦協助操作公式 例如:

```
> x <- c(1,2,3)
> sum(x)
[1] 6
> x-2 -> y
> y
[1] -1  0  1
>
> sum(y^2)
[1] 2
```

c. (10 points) Use R (軟體) to compute Pearson's correlation, Kendall's tau and Spearman's rho. (相關係數指令在講義)

d. (15 points) Please fit the regression model:

$$Y = \alpha + \beta X + \varepsilon .$$

Find $\hat{\alpha}$, $\hat{\beta}$ (用上課給的公式) and

$$\hat{\sigma}^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 / (n-2) = \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 / (n-2)$$

e. (5 points) What is the value of

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2} ?$$

Is it true that $R^2 = r^2$

f. (5 points) Please plot $e_i = Y_i - \hat{Y}_i$ (Y axis) versus $\hat{Y}_i$ (X axis). Comment on the residual plot. What does this plot indicate?

2. (45 points) 全部用軟體的指令

Milk or soda? The presence of soda vending machines in schools, under contracts with soft drink companies, is the subject of hot debate. Many see a link to childhood obesity as well as tooth decay and caffeine dependence. Has the soft drink industry changed our drinking habits? The Census Bureau reports U.S. per capita consumption of milk and carbonated soft drinks (in gallons per year) between 1980 and 2000:

Year	1980	1985	1990	1995	1998	1999	2000
Milk	27.6	26.7	25.7	23.9	23.0	22.9	23.3
Soda	35.1	35.7	46.2	47.4	47.9	49.7	49.3

a. (10 points) Please plot the data with X indicating "Soda" and Y indicating "Milk". Comment on the main features of the plot. Any possible outliers? 這裡我建議把 Milk 當成 Y, 是因為牛奶攝取是營養學家或家長比較在意的.

b. (10 points) Use R (軟體) to compute Pearson's correlation, Kendall's tau and Spearman's rho.

c. (10 points) Use R to fit the regression model:

$$Y = \alpha + \beta X + \varepsilon .$$

Show your output and explain the meaning of β.

d. (15 points) Show the residual plot and normal plot. Make your comments.
Hint: You need to know the purposes of these two plots and judge whether the plots support the assumptions that

$$\varepsilon_i \sim^{iid} N(0, \sigma^2) .$$