CEG2002

Statistical and Numerical Methods for Civil Engineers

Statistics Project, Semester 2, 2011-2012

Deadline and submission procedure

Deadline is **4.30pm, Thursday 8th March 2012**. Attach a completed cover sheet to your work, and submit it to the Cassie common room before the deadline. Work not handed in by this deadline might not be marked.

Instruction

There are two sections to this project, **Section A** and **Section B**. All questions in Section A must be hand written, and any graphs must be produced by hand. Questions in Section B are to be done in `Minitab`, and must be word processed and written up in a short 'report style'. Only relevant output from `Minitab` should be included.

Questions in Section B should be answered using your own personal datasets. To find your personal datasets, go to

`https://www.mas.ncl.ac.uk/modules/`

You may be asked to enter your university user name and password to access the above website. In the website, click 'CEG2002' and then click on the 'ice' or 'rope' link, the data set can be saved to disk, or opened immediately in Minitab.

Section A.1 Questions 1 to 4 in Project A.1 in Exercise 1.

Section A.2 (Some similar questions will be discussed in Tutorial 2)

1. The following data represent the running times (minutes) of films produced by a motion-picture company (A).

   `81 165 97 92 87 114`

   (a) Test the hypothesis that the average running time of films produced by company A is 100 minutes (note that we use the same data set as Question 1 in Project A.1).

   (b) The data from another company (B) are given as follows.

   `102 86 98 109 92`

   i. Test the hypothesis that the average running time of films produced by company A exceeds the average running time of films produced by company B by 10 minutes.

   ii. What assumptions are implicit in the test you carried out in part (ii)?

2. One of the problems when estimating the size of animal populations from aerial surveys is that animals may bunch together, making it difficult to distinguish and count them accurately. For example, a horse standing alone is easy to spot; if seven horses are huddled close together

some may be missed, resulting in an undercount. The relative frequency of undercounts is typically reported as a percent. For example, if there are 10 horses in a group, a person in the plane may typically count fewer than 10 horses 7% of the time. In a recent study, the percent of sightings that resulted in an undercount was related to the size of the 'group' of horses and donkeys; the following data were gathered:

x	2	3	4	5	6	7	8	9	10	11	12	14	16	18
p	5	5	6	10	5	7	5	6	7	5	5	14	13	23
y	1.61	1.61	1.79	2.30	1.61	1.95	1.61	1.79	1.95	1.61	1.61	2.64	2.56	3.14

Here, p represents % undercount, $y = \log(p)$ and x represents group size.

(a) Produce a scatter plot of y against x on graph paper, and comment.

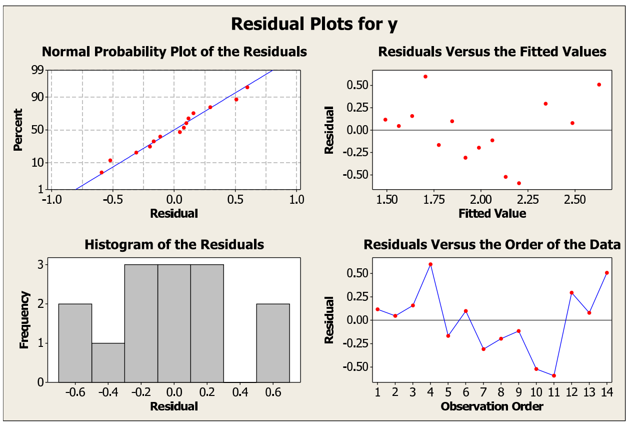
(b) Perform a linear regression analysis by using a simple linear regression $y = \alpha + \beta x + \epsilon$, and obtain the linear regression equation. Superimpose this line on your diagram in part (a). You may use the following summaries.

$$\sum x = 125, \quad \sum y = 27.77$$

$$\sum x^2 = 1425, \quad \sum y^2 = 58.21 \quad \sum xy = 269.91.$$

(c) Using the output from `Minitab` belew, would you say the assumption underlying your regression analysis are valid? Explain and justify.

(d) Calculate the predicted log (%undercount) (i.e. y) for a group size of 10, and then calculate the predicted value of %undercount.



3. A company is planning to develop a new project. There are two alternatives. Plan A will gain benefit 2m and have one fatality with probability 2/100. Plan B will gain benefit 1m and have on fatality with probability 1/100. We know that one fatality will cost 0.9m.

(a) Draw a decision tree;

(b) Which plan you would use based on minimax criterion?

(c) Which plan you would use based on expected-value decisions?

Section B

The questions in this Section require you to use your own personal dataset.

Following the instruction given at the begining of this course work sheet, go to

`https://www.mas.ncl.ac.uk/modules/`

and open or save the data sets of `rope` and `ice`, which are for questions 1 and 2 respectively.

1. Open the file `rope`. An engineer in a company which manufactures cranes wishes to compare wire ropes, of the same diameter, from two manufacturers A and B. She took a random sample of ropes from both manufacturers, and the load (kN) which caused each rope to break was recorded. The results are shown in the first two columns of the `Minitab` worksheet.

   (a) Use `Minitab` to obtain some numerical summaries of both samples, and copy and complete this table:

	Mean St. dev.	Median IQR
Manufacturer A		
Manufacturer B		

   (b) Produce histograms for the data in both samples, and compare and contrast the two sample distributions.

   (c) Use `Minitab` to perform an appropriate hypothesis test to compare the population mean breakage loads of rope from both manufacturers. Clearly state your hypotheses and summarise your conclusion in the context of the question. Be sure to include any relevant `Minitab` output.

   (d) What assumptions are implicit in the hypothesis test you have carried out? Are these assumptions verified?

2. Open the file `ice`. The data are the results obtained from 30 ice breaker trials, where the ice thrust, y (in thousand Newtons), was recorded for various ship speeds, x (in metres per second). We are interested in making predictions of ice thrust based on ship speed.

   (a) Use a scatterplot to determine whether there is any association between ship speed and ice thrust. Include the plot in your solutions, and make appropriate comments.

   (b) Calculate the correlation coefficient between ship speed and ice thrust, and test it significance.

   (c) Perform a regression analysis on these data, and include the regression table in your solutions. State the estimated regression equation.

   (d) Use `Minitab` to check the residual assumptions implicit in your regression, and include any relevant plots in your solutions.

   (e) Use your estimated regression equation to predict the ice thrust of a ship traveling at a speed of 6.8 metres per second.