

PROJECT 1 - DATA

Statistics 401: Fall 2006

Due Friday, September 8

Read Chapter 4 in Course Notes: Statistics for Researchers STAT401 FALL 2006 and write up your report for this assignment accordingly. Your write-up must be typed. The write-up should not exceed two pages, but you may attach R output on additional, separate pages. Your grade will be determined by how well you answer the questions and by the professionalism and clarity of your write-up.

1. Read the AP article at the STAT401 web site entitled “Illegal immigrants expanding footprint.”
 - (a) The graphic in the article shows three different explanatory variables. Classify each variable as either categorical or numerical. For each categorical variable, give the number of categories. For each numerical variable, classify it as either discrete or continuous.
 - (b) What are the response variables?
2. Table 1 displays the dimensions (in mm) of jellyfish from two samples taken from the Hawkesbury River in New South Wales, Australia. One of the samples is from Dangar Island and the other is from Salamander Bay. The investigator wants to know to what degree one can determine the location of the jellyfish given its dimensions.

Table 1: Dimension of Jellyfish

Dangar Island		Salamander Bay	
Breadth	Length	Breadth	Length
6.5	8.0	12.0	14.0
6.0	9.0	15.0	16.0
6.5	9.0	14.0	16.5
7.0	9.0	13.0	17.0
8.0	9.5	15.0	17.0
7.0	10.0	15.0	18.0
8.0	10.0	15.0	18.0
8.0	10.0	16.0	18.0
7.0	11.0	14.0	19.0
8.0	11.0	15.0	19.0
9.0	11.0	16.0	19.0
10.0	13.0	16.5	19.0
11.0	13.0	18.0	19.0
12.0	13.0	18.0	19.0
11.0	14.0	16.0	20.0
11.0	14.0	16.0	20.0
13.0	14.0	17.0	20.0
14.0	16.0	18.0	20.0
15.0	16.0	19.0	20.0
15.0	16.0	15.0	21.0
15.0	19.0	16.0	21.0
16.0	16.0	21.0	21.0
		19.0	22.0
		20.0	22.0

- (a) List the variables in this study. Classify each variable as either categorical or numerical. Give the number of categories and list the categories of each categorical variable. Classify each numerical variable as either discrete or continuous. If a variable is continuous but rounded, then classify it as continuous.
 - (b) Give the number of rows and the number of columns of the electronic data file (suitable for reading into R) that would be constructed from this data set.
 - (c) A data file which contains the data from Table 1 is at the STAT401 web site. Read the data file into a data frame in R. In your report, include commands that (a) make it possible to refer to the variables by their names alone, rather than as components of the data frame, and (b) display the contents of the data frame. Remember, you'll need to click on tab **File** → (**Change dir ...**) if the data file does not reside in the working directory.
 - (d) Attach a print-out of the output from the previous question to your report.
 - (e) Consider the main question that was asked of the data. Of the variables listed in (2a), which are response variables and which are explanatory variables?
3. Use R to find the mean and standard deviation of the Length of the jellyfish from Dangar Island and from Salamander Bay. Do these values suggest that there may be a true difference in the mean length of jellyfish from these two locations? In your report, include the R code you used to answer this question.
 4. Find a data set in your research area. Write a self-contained description of the data. Assume that you are writing for an educated reader, but one who is not familiar with your research area. Make sure that I will be able to understand the meaning of your variables.
 5. Answer the questions in #2 for your chosen data set.