

Assignment 1, Math 3346, 2007

Lecturer: John Maindonald

July 26, 2007

This exercise will work with three groups of data – experimental control and treatment groups in the data set `nswdemo`, and a non-experimental control group in `psid3`. The datasets `nswdemo` and `psid3` are both included in the *DAAGextras* package. Be sure to examine their help pages.

- (a) Look up the help page for the function `rbind()`. Create small data frames `df1`, `df2`, `df11`, and `df12`, thus:

```
> df1 <- data.frame(x1 = letters[1:4], x2 = 11:14)
> df2 <- data.frame(x1 = letters[5:9], x2 = 25:29)
> df11 <- data.frame(x1 = letters[1:4], x2 = 11:14, stringsAsFactors = FALSE)
> df12 <- data.frame(x1 = letters[5:9], x2 = 25:29, stringsAsFactors = FALSE)
```

Comment on the different results from

```
> rbind(df1, df2)$x1
> rbind(df11, df12)$x1
```

[1 mark]

- (b) Combine `nswdemo` and `psid3` into a single data frame, taking care that the three groups of data are distinguished. Ensure that the group identifier is a factor, rather than an integer or other numeric value.

[1 mark]

- (c) Write brief notes on each of the columns in the data frame, noting whether columns should be treated as numeric or categorical, ranges of values in numeric columns, and numbers of NAs. For the variables that record incomes, note numbers of zeros.

[3 marks]

- (d) Provide graphs that conveniently summarise differences between the three groups, with respect to `age`, `re74` and `re75`.

NB: Marks will be given for layout, with a preference for a layout that lays information out in a compact and readily comprehended form.

[4 marks]

- (e) Use tables to summarise differences in categorical variables between the three groups, or to summarise differences between continuous variables that are not well captured in the graphs.

[2 marks]

- (f) What are the major differences between the three groups, as evident from examining columns one at a time?

[4 marks]

Due Date: August 9, 2007, 5pm

In addition to any R code that may be included in the main document, please provide the R code separately from the output. Please provide assignments in a pdf file, either as hard copy or emailed to john.maindonald@anu.edu.au