Biostat II: Lab 11, Entering data into R and Adjusted Variables plots
Date: 9 May 2008

The data set below consists of measurements of weight in pounds, height in inches and age in years for 12 nutritionally deficient children. (Data are from *Applied Regression Analysis and Other Multivariable Methods* by Kleinbaum, Kupper, and Muller 1988 p.103).

The purpose of your analysis is to investigate the dependence of weight on height and age in this particular population of children.

weight	height	age
64	57	8
71	59	10
53	49	6
67	62	11
55	51	8
58	50	7
77	55	10
57	48	9
56	42	10
51	42	6
76	61	12
68	57	9

1. In order to read in the data, we would like to save it in a text (.txt) file format. Here is one way you can do this:

   (a) Highlight and copy the data and the column headings from this .pdf document.

   (b) Open the 'Notepad' or 'Wordpad' text editor program in Windows, or if you're using a Mac, use whatever is your favorite text editor. To find and open Notepad or Wordpad, you usually go to 'Start' ⇒ 'All programs' ⇒ 'Accessories' and then choose either 'Notepad' or 'Wordpad'.

   (c) Paste your data and headings directly into the text editor.

   (d) Save the data you have entered as a text file. To do so, go to the "File" menu and choose "Save". Make sure the file format is set to be text (TXT). Save the file as "weight.txt" to the **Desktop**.

2. Now we are ready to load the data into `R`. Open the `R` software. Change your working directory in `R` to select the folder where you have stored the data. To do so, click on the "File" menu and choose "Change dir...". Navigate through the menus using your mouse, and select Desktop folder as your working directory.

3. Read the data file you've created into R using the command:

```
child <- read.table("weight.txt", header=T)
```

Recall: The purpose of your analysis is to investigate the dependence of weight on height and age in this particular population of children.

4. Create scatter plots of weight versus height, weight versus age, and height versus age using the `plot` command as follows

   - `par(mfrow=c(1,3))`
   - `plot(child$height, child$weight, xlab="Height", ylab="Weight")`
   - `plot(child$age, child$weight, xlab="Age", ylab="Weight")`
   - `plot(child$age, child$height, xlab="Age",ylab="Height")`

   Resize the graphics window by dragging one of the corners to get a good view of your data.

5. Our goal is to make an adjusted variables plot (AVP) of the relationship between weight and height, after adjusting for age. We will do this by plotting residuals from the regression of weight on age versus residuals from the regression of height on age.

   (a) To see what the relationship between weight and height looks like before controlling for age, obtain the simple linear regression of weight on height:

   ```
   out.wh = lm(weight ~ height, data=child)
   ```

   (b) As the first step of making the AV plot, obtain the simple linear regression of weight on age:

   ```
   out.wa = lm(weight ~ age, data=child)
   ```

   (c) As the second step of making the AV plot, obtain the simple linear regression of height on age.

   ```
   out.ha = lm(height ~ age, data=child)
   ```

   (d) Make the AV plot:

   ```
   plot(out.ha$resid, out.wa$resid)
   ```

6. Perform linear regression of the age adjusted weight values on the age adjusted height values. To do so, regress residuals from `out.wa` on the residuals from `out.ha`. Also, add this regression line to the plot you've just made:

   ```
   out.ageadjusted = lm(out.wa$resid ~ out.ha$resid)
   abline(out.ageadjusted)
   ```

7. Look at a summary of the age adjusted regression that we performed using the residual values, and compare it to a summary of the full regression model. Comment.

   ```
   summary(out.ageadjusted)
   summary(lm(weight ~ height + age, data=child))
   ```