Assignment 2

Anh Le

October 3, 2014

```
## knitr configuration: http://yihui.name/knitr/options#chunk_options
opts_chunk$set(error= TRUE, warning = FALSE, message = FALSE,
               tidy = FALSE, cache = F, echo = T,
               fig.width = 4, fig.height = 4, fig.align="center")
library(DAAG)

## Loading required package:  lattice
```

1 Question 1: Problem 1.2 in book

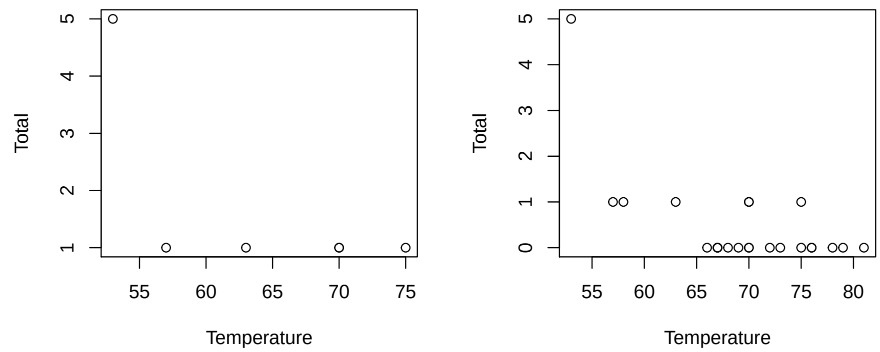
The `orings` data frame gives data on the damage that had occurred in US space shuttle launches prior to the disastrous Challenger launch of January 28, 1986. Only the observations in rows 1, 2, 4, 11, 13, and 18 were included in the pre-launch charts used in deciding whether to proceed with the launch.

Create a new data frame by extracting these rows from `orings`, and plot `total` incidents against `temperature` for this new data frame. Obtain a similar plot for the full data set.

Solution

Use the following to extract rows that hold the data that were presented in the pre-launch charts:

```
orings86 <- orings[c(1,2,4,11,13,18), ]
par(mfrow=c(1, 2))
with(orings86, plot(Temperature, Total))
with(orings, plot(Temperature, Total))
```




```r
par(mfrow=c(1, 1))
```

2 Question 2: Problem 1.11 in book

Explain the output from the final `table(gender)`.

```r
gender <- factor(c(rep("female", 91), rep("male", 92)))
table(gender)

## gender
## female   male
##     91     92

gender <- factor(gender, levels=c("male", "female"))
```

```r
table(gender)

## gender
##   male female
##     92     91

gender <- factor(gender, levels=c("Male", "female"))
table(gender)

## gender
##   Male female
##      0     91

rm(gender)                    # Remove gender
```

Solution

Notice that before the final `table(gender)`, we make a typo, specifying `levels=c("Male", "female")` instead of `levels=c("male", "female")`. Since there is no `"Male"` in `gender`, R correctly says that there is 0 `"Male"`.

The moral of this problem is that you should avoid specifying the `levels` by yourself. In most cases, R is smart enough to rely on.

3 Question 3: Endogeneity

1. When do we have a problem with endogeneity?

   When our independent variable is correlated with the error term. In that case, OLS regression estimate is biased.

2. Show why reverse causality leads to endogeneity.

   We have:

$$y = \beta_1 x + u \tag{1}$$
$$x = \beta_2 y + v \tag{2}$$

   where u, v are error terms, and $\beta_1, \beta_2 \neq 0$ (indeed, if $\beta_1, \beta_2 = 0$ then x and y don't cause one another, and we don't have reverse causality).

   Our goal is to show that x is correlated with u. The overall strategy is:

   - Based on Eq 2, we can express y in terms of x

$$y = \beta_2^{-1} x - \beta_2^{-1} v \tag{3}$$

   - Substitute Eq 3 into y in Eq 1, so that Eq 1 consists of only x and u.

$$y = \beta_1 x + u \tag{4}$$
$$\beta_2^{-1} x - \beta_2^{-1} v = \beta_1 x + u \tag{5}$$
$$(\beta_2^{-1} - \beta_1) x = \beta_2^{-1} v + u \tag{6}$$
$$x = \frac{\beta_2^{-1} v}{\beta_2^{-1} - \beta_1} + \frac{u}{\beta_2^{-1} - \beta_1} \tag{7}$$

     Notice that since $\beta_1, \beta_2 \neq 0$, $\frac{1}{\beta_2^{-1} - \beta_1} \neq 0$. Therefore, Eq 7 means that x and u are correlated. $\square$

3. Discuss one empirical paper – what is the dependent variable, the independent variable. Is there a potential endogeneity problem? Of what kind (ommited variable bias, selection bias, reverse causality)?

4. Fun fact: Endogeneity can also be caused by measurement error and simultaneity bias.