STATS 500 - Homework 1

Due in class Wed, September 20

1. Chapter 1, problem 2 (page 12)
The dataset `uswages` is drawn as a sample from the Current Population Survey in 1988. Make a numerical and graphical summary of the data, commenting on any features that you find interesting. Limit the output you present to a quantity that a busy reader would find sufficient to get a basic understanding of the data.

Hints:

- After you install R (if it is not already installed on the system you are using) and the `faraway` package, you can type `library(faraway)` to load the package, `data(uswages)` to load the data, and `attach(uswages)` to be able to use variable names directly.
- Type `help(uswages)` to read a description of the variables.
- Useful R functions for this homework: `data(), summary(), hist(), plot()`.
- Do NOT include the output of `plot(uswages)`: it has too many meaningless plots. Rather, select a few plots that you can comment on.
- To plot a scatter plot of two variables, use `plot(educ, wage)`. To get side-by-side boxplots, use `plot(factor(educ), wage)`.
- You can always type `help(subject)` to get detailed help on the `subject`, e.g. `help(plot)`. Or you can type `help.start()` to get interactive help with a search engine.

Solutions for this problem should be no more than 3 pages.

2. Determine which of the following models for the dependency of y on $x = (x_1, x_2, \ldots, x_p)$ are linear models according to the definitions in class. If they are not linear models, determine if there are transformations that could be applied to make them linear models – if this is the case, be explicit about the transformations and the (unknown) parameters in this transformed model.

(a) $x = x_1$ and $f(x) = \beta_o + \beta_1 \sin(x_1)$.

(b) $x = x_1$ and $f(x) = \beta_1 e^{\beta_2 x_1}$.

(c) $x = x_1$ and $f(x) = \beta_o + \beta_1 e^{\beta_2 x_1}$.

(d) $x = (x_1, x_2)$ and $f(x) = \beta_1 e^{\beta_2 x_1} + \beta_3 x_2$.

(e) $x = (x_1, x_2)$ and $f(x) = \beta_1 e^{\beta_2 x_1} x_2^{\beta_3}$.