Statistics 135 – Assignment 2
Due: Friday, October 28, 2005

For this assignment, please submit your answers in LaTeX.

1. Download the data furnace.dat from the course web site (either from the Datasets page or the Assignments page) and read the file into **R** using the `read.table`. This dataset describes a study by Wisconsin Power and Light of the effectiveness of two devices for improving the efficiency of gas home-heating systems. The electric vent damper (EVD) reduces heat loss through the chimney when the furnace is in its off cycle by closing off the vent. It is controlled electrically. The thermally activated vent damper (TVD) is the same as the EVD, except that it is controlled by the thermal properties of a set of bimetal fins set in the vent. Ninety test houses were used, 40 with TVDs and 50 with EVDs. For each house, energy consumption was measured for a period of several weeks with the vent damper active and for a period with the damper not active. This should help show how effective the vent damper is in each house.

   Both overall weather conditions and the size of a house can greatly affect energy consumption. A simple formula was used to try and adjust for this. Average energy consumed by the house during on period was recorded as (consumption)/[(weather)(house area)], where consumption is total energy consumption for the period, measured in BTUs, weather is measured in number of degree days, and house area is measured in square feet. In addition, various characteristics of the house , chimney, and furnace were recorded for each house. A few observations were missing and recorded as NA.

   A further description of the dataset is in the file `furnace.desc`, also available on the course web site.

   (a) Consider the subset of the data where House is either ranch or two story. For this subset of the data, generate box plots of BTUIn against Damper for each combination of Type and House in a single figure. Is there any difference in the relationship between BTUIn and Damper across the difference combinations of Type and House?

   (b) For the whole dataset, plot the relationship between BTUIn and BTUOut for the different combinations of CHShape and Damper, again in a single figure. Is there any evidence from the plot that relationship between BTUIn and BTUOut is different for some combinations of CHShape and Damper?

   (c) Generate the scatterplot matrix for all of the continuous variables in the dataset (Omit the variables that should be treated as factors). Are there any interesting relationships suggested by the plot?

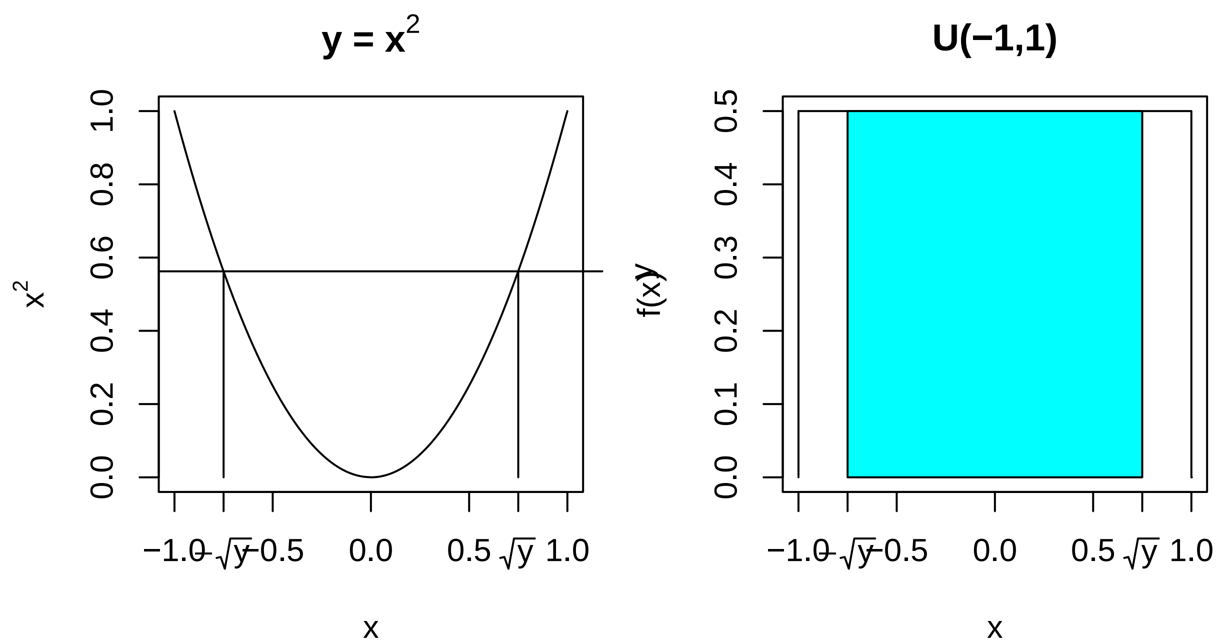
   (d) Fit the linear model predicting BTUIn by BTUOut, CHShape, and Damper where all two-way interactions are included in the model and give a summary of the fitting results. Does this output agree with the conclusions you made in part (b)?

   (e) Fit the additive model predicting BTUIn by BTUOut and Damper. Is there any evidence that the EVD type of damper reduces energy consumption after accounting for BTUOut? What is the estimated difference between the different dampers?

2. Create in LaTeX as closely as possible, with the necessary figures created in **R**, the following examples. To help with the figure in a), see the help pages for `axis`, `polygon`, `plotmath`.

   (a) Now often this can be written in terms of $F_X(x)$. For example, let $X \sim U(-1, 1)$ and $Y = X^2$, $A = [-\sqrt{y}, \sqrt{y}]$. Thus

$$\begin{aligned} F_Y(y) &= P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= \sqrt{y} \end{aligned}$$



   (b) Then $P[Y \in (y_1, y_2)] = P[X \in (x_1, x_2)]$, so

$$f_Y(y)(y_2 - y_1) \approx f_X(x)(x_2 - x_1)$$

   or

$$f_Y(y) \approx f_X(x) \frac{\Delta x}{\Delta y}$$

   The transformation g changes the scale of measurement. To keep keep the probabilities the same, which is needed since both descriptions give the same event, we must account for this change of scale. As we know from calculus, this change of scale is given by

$$\frac{d}{dy} g^{-1}(y) = \lim_{\Delta y \to 0} \frac{\Delta x}{\Delta y}$$

(c) Lets consider the case where X and Y are both $Bern(p)$ marginally.

- If X and Y are independent

$$\text{Var}(X+Y) = \text{Var}(Bin(2,p)) = 2p(1-p)$$

- If $X = Y$ (positive dependence), then

$$\text{Var}(X+Y) = \text{Var}(2X) = \text{Var}(2Bin(1,p)) = 4p(1-p)$$

- If $X = -Y$ (negative dependence), then

$$\text{Var}(X+Y) = \text{Var}(0) = 0$$

(d) If X and Y are independent (and X and Y are continuous RVs)

$$\begin{aligned}
E[XY] &= \int_{\mathcal{X}}\int_{\mathcal{Y}} xyf_{X,Y}(x,y)dydx \\
&= \int_{\mathcal{X}}\int_{\mathcal{Y}} xyf_X(x)f_Y(y)dydx \\
&= \left(\int_{\mathcal{X}} xf(x)dx\right)\left(\int_{\mathcal{Y}} yf(y)dy\right) = E[X]E[Y]
\end{aligned}$$