**R homework 1**

Learning R

This homework will require use of the Ischemic Heart Disease dataset which can be found on the course website. The codebook is also in the course website.

1. Read in the dataset IschemicHeartDisease.csv and save it as an R object.
2. what is the dimension of the dataset? (that is, how many rows and columns does it have)?
3. what are the names of the variables in the dataset?
4. find the mean, median, 1st and 3rd quartile for the variable totalcost.
5. Create a boxplot and a histogram of totalcost.
6. Generate a new variable that is the log of totalcost+1. (we add 1 because the minimum value of totalcost is 0. taking the log of 0 would give us values of –Inf)
7. Summarize log of totalcost using the statistics in question 4 above
8. Create a new variable that creates categories of the number of interventions carried out. Use the quartiles of the data to come up with 4 categories. Hint: use the cut command and use the option include.lowest=T. See help(cut) for more info!). Refer to this new variable as intervention category.
9. Make a table of the intervention category by gender.
10. Test that there is an association between intervention category and gender using a fisher’s exact test. What is the p-value?