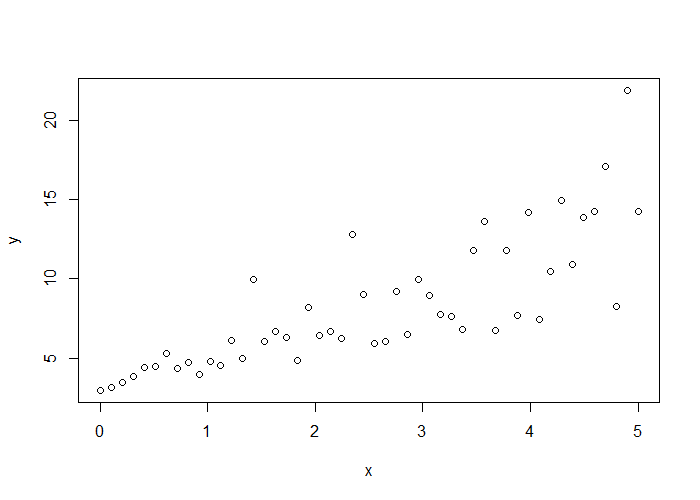
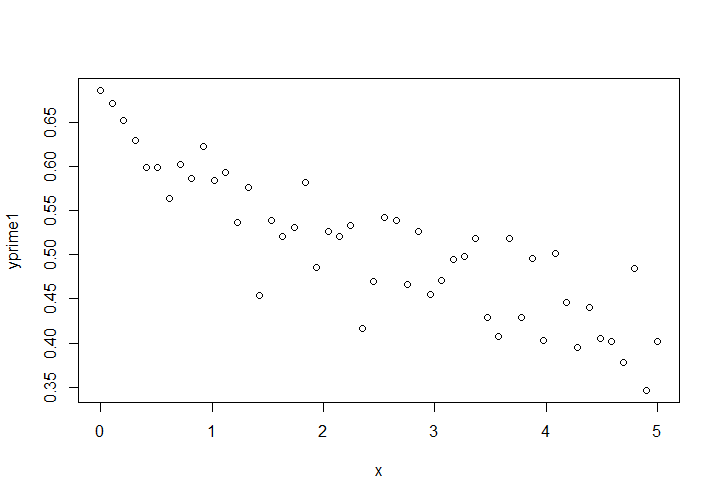
STA 5207 Assignment 10

Due Friday November 19

1. (60 points total) The data set A10Q1.txt contains a predictor variable *x* and then a response variable *y*. There are 50 observations.

If you are using SAS, make a note of this, as your solution to b. and c. will be slightly different than if you use R.

1. (20 points) Perform simple linear regression with the original variables. Do there appear to be problems with nonconstant variance, nonnormality, and nonlinearity? Explain your answer and include the scatterplot.
   1. There do appear to problems with the value. We can see the relationship is not linear as y increases with x, but at a much faster rate than linear. We can also see that the variance is nonconstant as it increases over time.
2. (10 points) Estimate the value of λ for a Box-Cox transformation. Give the value of .
   1. Because we see that y is increasing with increasing rate of increase, we can assume we’ll use something of the form y’ = which would be. From the boxcox function, we get.
3. (30 points) Use the transformation on *y* suggested in part b. Does this fix the problems you saw in the original data? Explain your answer and include the scatterplot.
   1. The transformation fixes the nonlinearity in the original data. There is some improvement in the nonconstant variance as well. However, the normal probability and histogram of the residuals indicate there is still a problem in the errors as evidenced with the long tails and skewness of the QQ plot.
4. (40 points total) The data set teengamb in the faraway package has 47 observations and 5 variables. The columns do not have headings. A survey was conducted to study teenage gambling in Britain.

The variables, in order, are:

sex: 0=male, 1=female

status: Socioeconomic status score based on parents' occupation

income: in pounds per week

verbal: verbal score in words out of 12 correctly defined

gamble: expenditure on gambling in pounds per year

The data is stored in SAS Studio as gamble.txt.

gamble is the response variable, and the others are the predictors. Some of the values of gamble are 0, so create a new variable gamble2=gamble+1.

In Assignment 4, we saw that there were problems with nonconstant variance, nonnormality, and possibly nonlinearity.

If you are using SAS, please make a note of this as your answers will be different than if you use R.

1. (10 points) Estimate the value of λ for a Box-Cox transformation. Remember to use gamble2 as the response variable. Give the value of.
   1. From the plot it looks like λ = 0 should be the estimated value
   2. The value is estimated
2. (30 points) Use the transformation on gamble2 suggested in part a. Does this fix the original problems of nonconstant variance, nonnormality, and nonlinearity? Explain your answer and include the plot of residuals by fitted values.
   1. The value for R2 is a little lower for the transformed model, but there is much less of a pattern in the residual plot, suggesting the transformation has made the error variance more constant. Additionally, none of the partial regression plots from the original model suggested there was a nonlinear relationship between any predictors and the response, so we have no need to transform any of the predictors.

