Homework 5

Psychology 313

Instructions. Show your R code, your input, and your output. Feel free to email me for hints if you get stumped. DUE APRIL 10.

1. (25 points.) In Thursday's class, in the notes on WLS regression, we worked a problem involving "strong interaction" data in the `physics` data set. We concluded our analysis in class by showing that adding a quadratic term significantly improved the fit over a model with just a linear predictor. Go back to the output for that example in the notes, and look at it closely.

   (a) In the second model, with both x and x^2 as regressors, is the regression coefficient for x signficantly different from 0?

   (b) The ANOVA showed that x was a significant regressor (over and above the intercept), and that x^2 improves the model significantly over and above a model with just x as a regressor. So why is x no longer significant?

   (c) Try dropping x out of the model, and fitting a third model with just x^2. Look at the R^2 value. How does it compare to a model with both x and x^2 in the model?

   (d) Does adding x to a model with just x^2 in it improve the model significantly?

   (e) Add the model with just x^2 as a regressor to the plot shown in the lecture notes with linear and quadratic models. What do you see?

2. (25 points). Use the ALR4 data set `stopping`. This data set contains stopping `Distance` in feet and `Speed` in miles per hour for $n = 62$ automobiles.

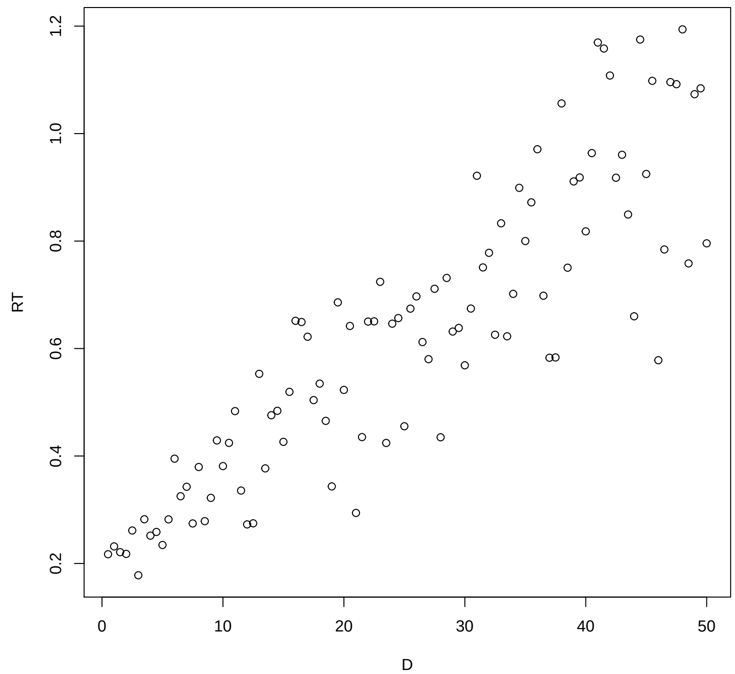
   (a) Using `Speed` as the only regressor, find an appropriate transformation for `Distance` that can linearize this regression. I suggest using the visual (slider-based) transformation software that we used in class.

   (b) Using `Distance` as the response, transform the predictor Speed using a power transformation with $\lambda \in \{-1, 0, 1\}$, and show that none of these transformations is adequate.

(c) Show that using $\lambda = 2$ does match the data well. (Simply copy the plot from the RStudio graphics pane.) This suggests using a quadratic polynomial for regressors, including both `Speed` and `Speed`^{2}.

(d) Hald (1960) suggested on the basis of a theoretical argument using a quadratic mean function for `Distance` given `Speed`, with $\text{Var}(Distance|Speed) = \sigma^2 Speed^2$. Draw the plot of `Distance` versus `Speed`, and add a line on the plot of the fitted curve from Hald's model. Then obtain the fitted values from the fit of the transformed `Distance` on `Speed`, using the transformation you found in part(a) above. Transform these fitted values to the `Distance` scale (for example, if you fit the regression $\sqrt{Distance} \sim Speed$, then the fitted values would be in square-root scale and you would square them to get the original `Distance` scale). Add to your plot the line corresponding to these transformed fitted values. Visually compare the fit of the two models and describe what you see.

3. (25 points.) Data for this question are in a file called `RTData.csv`. A group of $n = 40$ participants performed a "mouse pursuit" task. Their job ws to point and click on a round dot on the screen as quickly as possible. A well known formula produces an index of difficulty (`D`) for each point, based on the size of the point and its distance from the cursor starting position. The scatterplot below shows mean reaction time (`RT`) of the 40 participants as a function of difficulty `D` for 100 points of various sizes and starting positions.



(a) Perform a Breusch-Pagan test for the assumption of nonconstant variance.

(b) Suppose the variance is a simple function of D itself. Break the data into 10 groups of 10 observations each based on the values of D. For example, the first 10 observations would have values of D ranging from 1 to 5.5, the next from 6 to 10.5, etc. Fit a straight line to the group data from these 10 groups, with the mean D on the horizontal axis and the variance of RT on the vertical axis. This will give you a very rough estimate of the variance function. Is the variance a function of the mean itself?

(c) Suppose that, in the population, the conditional variance is directly proportional to the value of D. You decide to do a WLS regression. What simple weights could you use?

(d) Use these weights, standard errors for the regression weights, and compare them to robust standard errors obtained using a sandwich estimator.

4. (25 points). I found the table shown below online at

`http://www.healthdiscovery.net/links/calculators/ideal_bw_men.htm`

Ideal Body Weight Charts

For Men 25-59 years of age.

Height in Feet&Inches	Small Frame	Medium Frame	Large Frame
5'2"	128-134	131-141	138-150
5'3"	130-136	133-143	140-153
5'4"	132-138	135-145	142-156
5'5"	134-140	137-148	144-160
5'6"	136-142	139-151	146-164
5'7"	138-145	142-154	149-168
5'8"	140-148	145-157	152-172
5'9"	142-151	151-163	155-176
5'10"	144-154	151-163	158-180
5'11"	146-157	154-166	161-184
6'0"	149-160	157-170	164-188
6'1"	152-164	160-174	168-192
6'2"	155-168	165-178	172-197
6'3"	158-172	167-182	176-202
6'4"	162-176	171-187	181-207

(a) Plot the regression of maximum desirable weight for men with a medium frame versus height in inches.

(b) Look at the plot. Look at the chart! See an obvious error?

(c) Fix up the data with your best estimate for the correct value of the erroneous observation.

(d) Fit the repaired data with a simple linear regression model. Superimpose the line of best fit on the data.

(e) Add a quadratic term. Is there a significant improvement? (Hint: use the `anova` function.)

(f) Plot the predictions from the quadratic.