UNIT 10: STAT 6371: Chapter 7

1. Black Wheatears, are small birds of Spain and Morocco. Males of the species demonstrate an exaggerated sexual display by carrying many heavy stones to nesting cavities. This 35-gram bird transports, on average, 3.1 kg of stones per nesting season! Different males carry somewhat different sized stones, prompting a study of whether larger stones may be a signal of higher health status. M. Soler et al. calculated the average stone mass (g) carried by each of 21 male black wheaters, along with T-cell response measurements reflecting their immune system’ strengths. Analyze the data and write a statistical report (by answering the questions below); treat the T-cell as the response and the stone mass as the explanatory variable.
   1. You can find the data for this problem on Canvas. (Male Display Data Set.csv).
   2. Analyze the data providing at least the following:
      1. A Scatterplot with confidence intervals of the regression line and prediction intervals of the regression line. Please do in SAS and R!
      2. A table showing the t-statistics and pvalues for the significance of the regression parameters: . Please do in SAS and R!
      3. Using the data in ii show all 6 steps *of each* hypothesis test.
      4. The regression equation.
      5. Interpretation of the slope and intercept in the model (regression equation.)
      6. Find and interpret the 95% confidence interval for the mean t-cell response conditional on a stone mass of 4.5 grams. Please do in SAS and R!
      7. Find and interpret the 95% prediction interval for the predicted t-cell response given a stone mass of 4.5 grams. Please do in SAS and R!
      8. Using the graphical method, find and interpret the calibration intervals for the t-cell response of 0.3. (Both for mean t-cell response and for a single t-cell response. Please do in R as SAS does not provide calibration intervals.! (R: package investr)
      9. Find the same calibration intervals analytically using the SE equations given in class and in the book (Version 3 page 194).
      10. A scatterplot of residuals. Please do in SAS and R!
      11. A histogram of residuals with normal distribution superimposed. (from SAS).
2. Using the data in problem 1: Calculate by hand (using excel) the following:

(An example of this was in the PowerPoints and in the videos below.)

* 1. The t-statics and p-value for each.
  2. 99% Confidence Interval Calculations for X = {3,4,5,6,7,8,9} grams.
  3. 99% Prediction Interval Calculations for X = {3,4,5,6,7,8,9} grams.
  4. Calibration Interval Calculations for and for Y = 0.3 mm.
     1. Graphical Method
     2. Analytical Method
  5. Provide Plots for the Confidence Intervals and Prediction Intervals (From Excel … fully labeled!) … an example is shown for the Movie data on Canvas as well as in the two videos below:

http://screencast.com/t/ztSxTImiOk6s

http://screencast.com/t/V9gnhSwb

1. Problem 29 Chapter 7

The data can be found on Canvas.

Exit pollsters predict election results before final counts are tallied by sampling voters leaving voting locations. The pollsters have no way of selecting a random sample, so they instruct their interviewers to select every third exiting voter, of fourth, or tenth, or some other specified number. Some voters refuse to participate of avoid the interviewer. If the refusers and avoiders have the same voting patterns as the rest of the population, then it shouldn’t matter; the sample, although not random, wouldn’t be biased. If, however, one candidate’s voters are more likely to refuse or avoid interview, the sample would be biased and could lead to misleading conclusion.

On November 4, 2004, exit pollsters incorrectly predicted that John Kerry would win the U.S. presidential election over George W. Bush. The exit polls overstated the Kerry advantage in 42 of 50 states. No one expects exit polls to be exact, but chance alone cannot reasonable explain this discrepancy. Although fraud is a possibility, the data are also consistent, with Bush supports being more likely than Kerry supporters to refuse or avoid participation in the exit poll.

In a postelection evaluation, the exit polling agency investigated voter avoidance of interviewers. Display 7.21 shows the average Kerry exit poll overestimate (determined after the actual counts were available) for a large number of voting precincts, grouped according to the distance of the interviewer from the door. If Bush voters were more likely to avoid interviews in general, one might also expect a greater avoidance with increasing distance to the interviewer (since there is more opportunity for escape). A positive relationship between distance of the interviewer from the door and amount of Kerry overestimate, therefore, would lend credibility to the theory that Bush voters were more likely to avoid exit poll interviews. How strong is the evidence that the mean Kerry overestimate increases with increasing distance of interviewer from the door? (Data from Evaluation of Edison/Mitofsky Election System 2004 prepared by Edison Media Research and Mitofsky International for the National Election Pool (NEP), January 15, 2005.

https://abcnews.go.com/images/Politics/EvaluationofEdisonMitofskyElectionSystem.pdf (accessed May 9, 2008).)

NOTE: Provide both the SAS and R Code as was done above.

* 1. Provide the analysis described in Question 1.b for this data.
     1. For the Confidence and Prediction intervals use X = {37 ft}
     2. For the Calibration intervals use and for Y = 6.0 points.
  2. Answer the questions in the text.

BONUS 1pt. I couldn’t figure out from the text what unit of measure this was … ft, meters, yards etc. Look up the reference in the text and find out what unit of distance was used.

1. For questions 1 and 3 go back and provide a measure of the amount of variation in the response that is accounted for by the explanatory variable. Interpret this measure clearly indicating the units of the response and the explanatory variables.

BONUS: Problem 23 Chapter 7 (4 pts)