STA-371G
Homework 9

This assignment contains 3 questions; there are 166 total points.

Instructions

Submit your assignment at the **beginning** of class on May 8, 2019. Please make sure that the pages of your homework assignment are in the correct order and securely stapled together. You may (but are not required to) work in groups of up to four students. Each group should hand in one copy of the homework.

Grading

A small number of the problems on each assignment will be graded for correctness, and the remainder graded for completeness. A complete response answers the question posed and also shows your work. This means showing the steps of a mathematical calculation, or including the R code you used to arrive at your answer. For questions that are not just calculations (e.g., more than computing an expected value from a table) you should answer in complete, concise sentences. Detailed solutions will be available – you should always check your work against these solutions.

---

1. (70 points)

   Read the "Oakland A's (A)" case in the course packet. The data are available on the course website; you can use the hw9.R script to read in the data.

   (a) (15 points) Run the regression

   $$Ticket_t = \beta_0 + \beta_1 Nobel_t + \epsilon_t$$

   where $Nobel$ is a dummy variable that takes the value 1 when Nobel starts on day t.

   What are the estimates of β_0 and β_1? What are the estimated expected ticket sales for games where Nobel is pitching? When he is not pitching?

   (b) (10 points) Plot Ticket against Time (i.e. create a time series plot of Ticket). Do you see any patterns in the data?

   (c) (10 points) Now plot a time series of the residuals from part a along with their autocorrelation function. Do they appear to be independent? Why or why not? If they are not independent, what factors might explain the pattern?

(d) (15 points) Run the regression

$$Ticket_t = \beta_0 + \beta_1 Pos_t + \beta_2 GB_t + \beta_3 Temp_t + \beta_4 Prec_t + \beta_5 TOG_t + \beta_6 TV_t + \\ + \beta_7 Promo_t + \beta_8 Nobel_t + \beta_9 Yanks_t + \beta_{10} Weekend_t + \beta_{11} OD_t + \beta_{12} DH_t + \epsilon_t$$

Do the residuals from this regression appear to be independent? Why might these residuals be independent even if the residuals from the first model were dependent?

(e) (15 points) Now suppose the residuals are in fact independent, so we have a valid model. What evidence is there about Nobel pitching in a game being related to the attendance at the game? Do you have more confidence in drawing a conclusion from the first model (with Nobel alone) or the second model?

(f) (5 points) Do you think Nobel's agent has a legitimate case that Nobel should be paid more because he brings fans to the games? Justify your answer using the regression output.

2. (75 points)

Read the "Oakland A's (B)" case in the course packet. The data file is available in the class website.

(a) (10 points) Run a regression of Attendance against Wins. What is the practical problem associated with using this model to forecast Attendance for the next season (i.e. to forecast attendance in the 1981 season)?

(b) (15 points) Now run a regression of Attendance against Roddey's forecast of the number of wins for that season. Why is the R2 value obtained from this regression so much lower than the R2 obtained from the regression in part (a)?

(c) (10 points) Why is it more appropriate to use the model in part (b) for forecasting Attendance than the model in part (a)?

(d) (8 points) Before the 1981 season starts Roddey forecasts 95 wins for the season. Using the model from part (b), what is the prediction for attendance in the 1981 season? What is its standard deviation, s?

(e) (32 points) Using the prediction and standard deviation for the prediction from the model in part (b), what is the probability associated with a bonus to Nobel of \$0, \$50,000, \$100,000 and \$150,000? (Use plug-ins to estimate predictive distributions.) What is the mean of this distribution? What is the expected bonus if the lump-sum incentive plan is used?

3. (21 points)

Read the case "Northern Napa Valley Winery, Inc." in the course packet. The data file is available in the course website. The file contains the monthly wine sales for the Northern Napa Valley Winery for the period January, 1988 through August, 1996.

The goal of this case is to provide monthly forecasts for wine sales in the next twelve months, i.e. September, 1996 through August, 1997, and to combine the monthly forecasts to provide a forecast of annual sales for these twelve months.

(a) (3 points) Using the appropriate time series plots decide whether sales or the percentage change in sales, denoted PctChange, should be analyzed. Fit a model to the appropriate variable (Sales or PctChange) that accounts for the trend and seasonality in the data.

(b) (5 points) Are the residuals from this model independent? If not, modify the model so the residuals are independent.

(c) (3 points) Using the final model obtained above, provide numerical forecasts of wine sales in September, 1996 and October, 1996.

(d) (10 points) Provide forecasts and 95% prediction intervals for the forecast of sales in September through August, 1997.