ADM2304-Winter 2017-Sections M-N-P-Q-R-S (45 points)

Assignment 4 Due Date: April 7, 2017

Instructions

1. *You may use Minitab or other software for any calculations. However, **you must show your manual calculations if requested**. You may paste your output onto your assignment to show your use of Minitab; however, this output does not replace any of the steps outlined below. This means that answers that are exclusively Minitab output may receive only one mark.*
2. *If you are performing a hypothesis test, make sure you state the hypotheses, the level of significance, the rejection region, the test statistic (and/or p-value, if requested), your decision (whether to reject or not to reject the null hypothesis), and a conclusion in managerial terms that answers the question posed. These steps must be completed in addition to any Minitab output.*
3. *Please remember to complete and include your signed statement of Academic Integrity in your final report. Submit a PDF of your type-written solutions via blackboard learn by the due-date indicated above.*

Context for the Assignment

All questions relate to the analysis of a dataset (see file on BBL: *ASSIGN4-Data.MTW*) originating from the Labour Force Survey conducted by Statistics Canada and which include annual salaries of 530 managers by Age, Gender, Education, Province, Sector and Industry. A respondent ID variable is also included.

Variable definitions

Quantitative variables (2):

- ANNUAL SALARY: dollars (January 2015 $Can)
- AGE: age of respondent (in years, rounded based on date of birth)

Categorical variables (5):

- GENDER: Female=0; Male=1;
- EDUCATION: No university graduation=0; University graduation=1;
- PROVINCE: Quebec=0; Ontario=1;
- SECTOR: Private=0; Public=1
- INDUSTRY[1]:
  - 08 Retail Trades
  - 10 Finance, insurance and property management (Finance)
  - 12 Business management and administration (Administration)
  - 13 Education and Training
  - 14 Health and Social
  - 16 Hotels and Restaurants
  - 18 Government

[1] Please note that this variable has been re-coded in the file using dummy variables with Industry 16 (Hotel and Restauration) as the reference category.

QUESTIONS

1. (10 points) Using Minitab or other software, run a simple linear regression model to predict annual salaries of managers based on their age and answer the following questions:
   a. Using an appropriate graph, plot annual salary versus age and comment on the relationship between these two variables;
   b. Write down your estimation of the regression equation for predicting annual salary from age. Draw the regression line on the plot in part (a);
   c. Test whether age has statistically significant predictive power in estimating the average salary. Assume $\alpha = 0.01$. State the hypotheses, provide a test statistic and p-value, and state your conclusion. (show your calculations);
   d. Interpret the meaning of the regression coefficients (slope and intercept).
2. (7 points) Perform a multiple regression analysis of annual salaries (response variable) against all the remaining variables as predictors (including the industry dummy variables but excluding the ID variable).
   a. Write down the regression equation and provide at least two summary measures of the fit of the model;
   b. Examine and comment on the usual residual plots and how they fit the usual model assumptions;
   c. The variable Annual Salary_New corresponds to the Annual Salary variable from which some values have been recoded as missing values. Identify those missing values and explain what they are and why they were recoded as missing.
3. (16 points) Perform the same multiple regression as in 2 above this time using the Annual Salary_New variable as the response variable.
   a. Briefly compare the two models in terms of the resulting regression equation and fit;
   b. Examine the residuals and report whether or not the model complies with the usual assumptions of a multiple regression and why;
   c. Do you see a justification to drop any variable from the model? Explain *(hint: multicollinearity; significance of predictors);*
   d. Run a final model using the Annual Salary_New variable as the response variable and include only the significant predictors *(hint: those with a p-value ≤ 5%*);

e. Test the overall significance of the final model (99% confidence level) and comment (show your calculations);

f. Perform a hypothesis test for the coefficient of the Age variable at the 99% confidence level (show your calculations);

4. (7 points) Based on your final model in 3 above:

   a. Test the marginal contribution of the education variable (University degree or not), assuming that other variables in the model remain constant. Use a 1% significance level and show the ratio for calculating the t-statistics.

   b. Assuming all other predictors would remain the same, what would be the estimated difference in mean salary between managers with or without a university degree? Calculate a confidence interval for the difference in mean salary. Could you say that if more managers had degrees it would increase the overall manager salary level?

5. (5 points) Again, based on your final model in 3 above:

   a. Calculate a 99% prediction interval for the **actual** annual salary of a new manager with the same characteristics as # 20853 in the data file (show how the interval is calculated manually using the Minitab output results);

   b. Check if it includes the actual salary value of # 20853; explain why it does or does not.