ECON 339: Quiz 2

Note: *Create a Word file for the answers that includes the relevant Excel or R output. Convert the file into pdf, using your last name followed by your first name in the filename, and upload to the designated subfolder on Canvas in the Assignments folder. Late submissions via email will be penalized. Read the questions carefully!*

Quantitative analysis can help managers and organizations make well-informed decisions about their workforce. Instead of making personnel decisions based on perception and instinct, managers can use data to help identify, attract, and retain employees. The **DataQ2** file contains information on 800 sales representatives from the software product groups of a high-tech company. For each employee, the data includes socio-demographic, feedback, salary, and personality variables. A deep understanding of the role of personality types is important for professionals wishing to pursue a managerial position in a firm.

1. Estimate a multiple linear regression model using salary as the dependent variable. The explanatory variables consist of Age, Female, Feedback, and three dummy variables created from Personality; the fourth personality type is used as reference. Report the regression output.

Note – reference personality type: Sentinel

```
Call:
lm(formula = Salary ~ Age + Female + Feedback + Diplomat + Explorer +
    Analyst, data = q2_data)

Residuals:
   Min     1Q Median     3Q    Max
-46314 -11719  -2670  10070  89303

Coefficients:
            Estimate Std. Error t value
(Intercept) 14812.18    3851.68   3.846
Age           621.25      66.33   9.366
Female      -2685.94    1381.84  -1.944
Feedback     8064.10     848.94   9.499
Diplomat    13396.40    1976.79   6.777
Explorer    14834.31    2003.04   7.406
Analyst      -488.80    2480.10  -0.197
                      Pr(>|t|)
(Intercept)              0.00013 ***
Age         < 0.0000000000000002 ***
Female                   0.05228 .
Feedback    < 0.0000000000000002 ***
Diplomat       0.000000000023962 ***
Explorer       0.000000000000333 ***
Analyst                  0.84381
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19150 on 793 degrees of freedom
Multiple R-squared:  0.2545,      Adjusted R-squared:  0.2488
F-statistic: 45.11 on 6 and 793 DF,  p-value: < 0.00000000000000022
```

2. Use an appropriate goodness of fit measure to show that the above model is superior to the one that does not include the personality dummy variables.

When I remove the dummy variables and I run the exact same linear regression model, I obtained an adjusted r squared value of 0.1619. The adjusted r-squared for the model with the dummy variables is 0.2488. Since the model with the dummy variables has an adjusted r-squared that is closer to 1, this means that this model is superior to the one without the personality dummy variables. Also, since we have a different number of predictor variables between the two models, I am using adjusted r-squared to compare the models.

Here is the regression model without the dummy variables:

```
Call:
lm(formula = Salary ~ Age + Female + Feedback, data = q2_data)

Residuals:
   Min     1Q Median     3Q    Max
-48259 -13384  -2196  11325  83120

Coefficients:
            Estimate Std. Error t value
(Intercept) 24774.58    3820.81   6.484
Age           596.04      69.82   8.537
Female      -2238.56    1453.60  -1.540
Feedback     8273.49     895.71   9.237
                      Pr(>|t|)
(Intercept)       0.000000000157 ***
Age          < 0.0000000000000002 ***
Female                    0.124
Feedback     < 0.0000000000000002 ***
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
  0.1 ‘ ’ 1

Residual standard error: 20230 on 796 degrees of freedom
Multiple R-squared:  0.1651,      Adjusted R-squared:  0.1619
F-statistic: 52.45 on 3 and 796 DF,  p-value: < 0.00000000000000022
```

3. Use the estimated model to predict the salary of a 50-year-old male sales rep with a feedback score of 3 and the personality type of an analyst. Make similar predictions for a diplomat, an explorer, and a sentinel.

Personality Type	Prediction
Analyst	69577.94
Diplomat	83463.14
Explorer	84901.05
Sentinel	70066.74

4. Report 90% confidence and prediction intervals for the salary of a 50-year-old male sales representative with a feedback score of 3 and the personality type of an analyst.

Functions I used:

predict(q2_lm, data.frame(Age = 50, Female = 0,Feedback = 3, Diplomat=0,Analyst=1 ,Explorer=0), level=0.9, interval="confidence")

predict(q2_lm, data.frame(Age = 50, Female = 0,Feedback = 3, Diplomat=0,Analyst=1 ,Explorer=0), level=0.9, interval="prediction")

Confidence Interval:

```
fit          lwr         upr
1 69577.94   65555.84    73600.04
```

Prediction Interval:

```
fit          lwr         upr
1 69577.94   37855.57    101300.3
```

5. Do females get paid less than males with other things being equal? Report the appropriate hypotheses in terms of the regression coefficient, test statistics, *p*-value, and the inference. Use a 5% level of significance for the test.

Hypothesis:

H_0: $\beta_{female} >= 0$

- the null is that beta for the female variable is greater than 0, meaning that females do not get paid less than males

H_1: $\beta_{female} < 0$

- the alternative is that beta for the female variable is less than 0, meaning that females do get paid less than males

Test-Statistic: -1.9437428

P-value: 0.026139

Inference: We reject the null hypothesis that females are not paid less than males since the p-value for the hypothesis test is less than 5% level of significance at 0.026139. This means we conclude that females are paid less than males with all other things being equal, as this is our alternative hypothesis.