Test 1

Ayato Tanemura

2022-04-13

Q1

```
>              Df Sum Sq Mean Sq F value   Pr(>F)    
> female        1     96    95.6   1.186    0.277    
> prog          2   4626  2312.9  28.701 1.24e-11 ***
> female:prog   2    224   111.8   1.387    0.252    
> Residuals   192  15472    80.6                     
> ---
> Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

1

In the results, you can conclude the following, based on the p-value and a significance level of 0.05.

- The p-value for female is 0.277. There is non-significant differences between female in relationship with read, which indicates that the types of gender are not associated with different reading score.
- The p-vale for prog is 0.000. There is significant differences between prog and read, which indicates that the types of programs are associated with different reading score.
- The p-value for the interaction between female*prog is 0.252. There is non-significant interaction between female and prog, which indicates that the relationship between gender types and reading score does not depend on the types of programs.

2

```
>   Tukey multiple comparisons of means
>     95% family-wise confidence level
> 
> Fit: aov(formula = read ~ female * prog, data = df.Noout)
> 
> $female
>          diff       lwr      upr     p adj
> 1-0 -1.394269 -3.919137 1.130598 0.2774368
> 
> $prog
>           diff        lwr         upr     p adj
> 2-1   7.066584   3.411871 10.72129770 0.0000262
> 3-1  -4.320406  -8.584456 -0.05635647 0.0462636
> 3-2 -11.386991 -15.116321 -7.65766042 0.0000000
```

```
>
> $‘female:prog‘
>                diff        lwr        upr     p adj
> 1:1-0:1  -5.4384615 -12.677026  1.8001026 0.2601712
> 0:2-0:1   4.5069767  -1.992438 11.0063913 0.3482993
> 1:2-0:1   3.9928571  -2.223021 10.2087355 0.4368772
> 0:3-0:1  -6.7478261 -14.214217  0.7185646 0.1018037
> 1:3-0:1  -7.4400000 -14.749186 -0.1308141 0.0433927
> 0:2-1:1   9.9454383   3.525547 16.3653293 0.0002005
> 1:2-1:1   9.4313187   3.298639 15.5639980 0.0002295
> 0:3-1:1  -1.3093645  -8.706634  6.0879053 0.9957642
> 1:3-1:1  -2.0015385  -9.240103  5.2370256 0.9679255
> 1:2-0:2  -0.5141196  -5.753903  4.7256642 0.9997547
> 0:3-0:2 -11.2548028 -17.930519 -4.5790871 0.0000367
> 1:3-0:2 -11.9469767 -18.446391 -5.4475622 0.0000049
> 0:3-1:2 -10.7406832 -17.140678 -4.3406883 0.0000405
> 1:3-1:2 -11.4328571 -17.648735 -5.2169788 0.0000048
> 1:3-0:3  -0.6921739  -8.158565  6.7742168 0.9998144
```

According to the result in (1), interaction effect is not sgnificant.

Q2

1

```
>
> Call:
> lm(formula = Alumni.Giving.Rate ~ Graduation.Rate, data = al_df)
>
> Residuals:
>      Min       1Q   Median       3Q      Max
> -17.6801  -6.5096   0.1953   6.6465  23.6122
>
> Coefficients:
>                 Estimate Std. Error t value Pr(>|t|)
> (Intercept)     -68.7612    12.5827  -5.465 1.82e-06 ***
> Graduation.Rate   1.1805     0.1507   7.832 5.24e-10 ***
> ---
> Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
>
> Residual standard error: 8.894 on 46 degrees of freedom
> Multiple R-squared:  0.5715,  Adjusted R-squared:  0.5621
> F-statistic: 61.34 on 1 and 46 DF,  p-value: 5.238e-10
```

The coefficient table shows that Graduation.Rate is significant at 0.05 level of significance.

So the model of predicting would be as following:

Alumni.Giving.Rate = -68.76 + 1.181(Graduation.Rate)

Overall Result:

F(61.34, 46) and p-value = 0.000 which is less than 0.05 significance level. Thus, null hypothesis is rejected, we can conclude that this model is statistically significant.

AD-R2 = 0.5621. We can conclude that approximately 56% variation in Alumni.Giving.Rate can be explained by this model.

2

```
>
> Call:
> lm(formula = Alumni.Giving.Rate ~ Graduation.Rate + Percentage.of.Classes.U20 +
>     Student.Faculty.Ratio, data = al_df)
>
> Residuals:
>      Min       1Q   Median       3Q      Max
> -11.9800  -5.9024  -0.6273   3.7644  20.6281
>
> Coefficients:
>                            Estimate Std. Error t value Pr(>|t|)
> (Intercept)               -20.72013   17.52137  -1.183  0.24333
> Graduation.Rate             0.74818    0.16596   4.508  4.8e-05 ***
> Percentage.of.Classes.U20   0.02904    0.13932   0.208  0.83584
> Student.Faculty.Ratio      -1.19201    0.38672  -3.082  0.00354 **
> ---
> Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
>
> Residual standard error: 7.61 on 44 degrees of freedom
> Multiple R-squared:  0.6999,  Adjusted R-squared:  0.6795
> F-statistic: 34.21 on 3 and 44 DF,  p-value: 1.432e-11
```

Estimate model:

Alumni.Giving.Rate = -20.72 + 0.7482(Graduation.Rate) + 0.02904(Percentage.of.Classes.U20) - 1.192(Student.Faculty.Ratio)

The above coefficient table shows that only Graduation.Rate and Student.Faculty.Ratio (p-value = 0.000) are significant at 0.05 level of significance. So those two variables play an important role in predicting Alumni.Giving.Rate. So the model of predicting will only include Graduation.Rate and Student.Faculty.Ratio in the model.

```
>
> Call:
> lm(formula = Alumni.Giving.Rate ~ Graduation.Rate + Student.Faculty.Ratio,
>     data = al_df)
>
> Residuals:
>      Min       1Q   Median       3Q      Max
> -11.9304  -6.1594  -0.5521   3.5910  20.5412
>
> Coefficients:
>                       Estimate Std. Error t value Pr(>|t|)
> (Intercept)           -19.1063    15.5501  -1.229    0.226
> Graduation.Rate         0.7557     0.1602   4.717 2.35e-05 ***
> Student.Faculty.Ratio  -1.2460     0.2843  -4.382 6.95e-05 ***
> ---
> Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
>
```

```
> Residual standard error: 7.528 on 45 degrees of freedom
> Multiple R-squared:  0.6996,  Adjusted R-squared:  0.6863
> F-statistic: 52.41 on 2 and 45 DF,  p-value: 1.765e-12
```

Estimate equation:

Alumni.Giving.Rate = -19.11 + 0.7557(Graduation.Rate) - 1.246(Student.Faculty.Ratio)

Overall result:

F(52.41, 45), p-value = 0.000 which is less than 0.05 significance level. Thus, null hypothesis is rejected. We conclude that this model is statistically significant.

AD-R2 = 0.6863. We can conclude that approximately 68% variation in Alumni.Giving.Rate can be explained by this model.

To compare between part one and two, the variation that can be explained is increased after adding other independent variables.

```
>
> Call:
> imcdiag(mod = reg3.fit)
>
>
> All Individual Multicollinearity Diagnostics Result
>
>                          VIF    TOL      Wi  Fi Leamer    CVIF Klein   IND1
> Graduation.Rate       1.5772 0.6341 26.5495 Inf 0.7963 -3.8637     0 0.0138
> Student.Faculty.Ratio 1.5772 0.6341 26.5495 Inf 0.7963 -3.8637     0 0.0138
>                       IND2
> Graduation.Rate          1
> Student.Faculty.Ratio    1
>
> 1 --> COLLINEARITY is detected by the test
> 0 --> COLLINEARITY is not detected by the test
>
> * all coefficients have significant t-ratios
>
> R-square of y on all x: 0.6996
>
> * use method argument to check which regressors may be the reason of collinearity
> ===================================
```

In this case, we can observe smaller tolerance and larger VIF values for both Graduation.Rate and Student.Faculty.Ratio. These results confirms the multicollinearity issue that we detected before when assessing the significance of the age coefficient and the correlation matrix.