|  |  |
| --- | --- |
| **Module:** | ST4400 |
| **Name:** | James Henry Hehir |
| **Student Number:** | 120224791 |
| **Chapter:** | Chapter 2: Multiple Regression |

**Maximum 2 pages! Do not delete the page number in the footer.**

**Exam Question (Winter 2019-20, Question 2)**

**(a)**

Graphical user interface, text, application

Description automatically generated

Strong positive linear correlation between Salary and Experience. (0.7088156)

Slightly positive linear correlation between Salary and Education. (0.2550112)

This would suggest multiple regression would be suitable or worthwhile in this model.

There is a strong positive linear correlation between Profits and Sales. (0.90271274)

This would suggest that there could be a potential problem with collinearity.

**(b)**

Text

Description automatically generated

2805.5 is β1. This suggests that with each year of Experience you should gain €2805.5 in Salary. This is assuming all other variables stay constant. In comparison to Question1 β1 was 2892.0. This would suggest the other variables have an influence on the Experience predictor in a multiple regression model.

**(c)**

In this model the multiple R-squared value is 0.5501 or 55%. In the model in Question 1 the multiple R-Squared value is 0.5024 or 50%. As we should expect when you add more variables to a model it made the R-squared bigger. If you add the values of the all the predictor variables’ estimated coefficients to the model it should improve the quality of the fit. Whenever you add a variable to a model, the R-squared cannot get smaller.

**(d)**

**Table

Description automatically generated**

The hypothesis I suggest is to compared the full model to a 3 variable model excluding Experience. This will show the significance of the variable Experience that we left out. So my hypothesis is **H0: β2 = β3 = β4 = 0.**

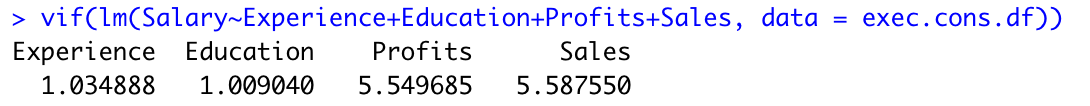
F-value=**5.547** p-value=**0.001194.** Because the p-value is less than 5% significance we reject **H0.** This suggests that Experience is needed in the model as it makes the p-value smaller. If we compare **H0** to the full model partial F-test or even a model with only the Experience predictor (as in Question 1) the p-value is smaller than the **H0** ‘s p-value. This suggests that Experience could be the only predictor variable in the model.

**(e)**

Collinearity can be defined as 2 or more predictors that can be expressed as a linear combination.

Difficulties caused by collinearity include : 1. Variances of predictors concerned are greatly inflated. 2. Regression coefficients are not defined for perfect collinearity.

**(f)**

****

The largest VIF (5.587550) which is Sales does not exceed 10 which would suggests that collinearity is not a serious problem in the linear model.