**Case Study Deadline:** Wednesday 30th September 2020 by 23:55 PM

**Case Study Deliverables:**

*One professionally written case study report per group, inclusive of the following*:

1. One-page *executive summary*, highlighting main features of your analysis, modelling, and comments on results obtained.
2. Main case study report, not exceeding *15* one-sided pages, including a succinct understanding of the problem, regression models, graphical representations (residual plots, P-P plots, etc.), results, and conclusions. Specific recommendations for the different questions listed below must also be included in the report.
3. Show all the calculations in the report to receive full credit.

|  |  |
| --- | --- |
| **Q1** | (a) Why do you disagree with Jack’s comments about the uselessness of the regression due to the low R-squared?   * 1. Can you think of a situation in which a useless regression has a high R-squared?   2. There are techniques to determine the validity of a regression model—in particular, whether the relationship is linear and the error terms display equal variance (homoskedasticity). Does the regression in Table 1 violate either of these two assumptions? Justify your answer. |
| **Solution** | **a)**  R-square / Adjusted R-squared is one of the checks to assess the usefulness of the model . But R square is not the right metric to measure the validity of the model.  To test the validity of the model following tests must be performed   1. F test to check the validity of the model. Looking at the F statisics value, model is valid at **(α=0.05)**      1. T-test to check the existence of statistically significant relationship between response variable and individual explanatory variables. In this case at least two variables are statistically significant at **(α=0.05)**      1. Residual analysis to check that variance of residual is constant for all values of Xi. (Homoscedasticity). The residual plot shows that the values are mostly centered around 0. We conclude that there is Homoscedasticity      1. Test to check if residuals follow a normal distribution. This can be done using a pp plot by plotting observed cum probability against expected cum probabilities. The plot shows that though there is some deviation at the edge, the errors are mostly normal. (The deviation could be because of some outliers)      1. Multicolinearity test- to check if there is high correlation between the independent variables. Multicolinearity can be tested using Variance Inflation Factor(VIF). In this model the VIF < 4 for all variables.     **b)**  R square/adjusted r square can be spurious i.e. High R-square does not necessarily mean that the Regression model is valid. Seemingly unrelated can show a high r square if the correlation between the variables is high.  **c)**  **Homoscedasticity:**  The residual plot shows that the values are mostly centered around 0. We conclude that there is Homoscedasticity. From the plot below it is clear that there is no visible pattern and we can conclude that relationship is linear    **Normality of Residuals:**  Normality of residuals can be tested using a **pp plot** by plotting observed cum probability against expected cum probabilities. The plot shows that though there is some deviation at the edge, the errors are mostly normal. (The deviation could be because of some outliers) |

|  |  |
| --- | --- |
| **Q 2** | (a) Estimate the excess return (RET) of the funds that Bob and Putney currently manage. Assume that Princeton’s average composite SAT score is 1355, while Ohio State’s is 1042. Between Bob and Putney, who is expected to obtain higher returns *at their current funds* and by how much?  (b)Between Bob and Putney, who is expected to obtain higher returns *if hired* by AMBTPM and by how much? |
| **Solution** | a)  Using Mulitple Linear Regression ( and selecting only statistically significant variables at **α=0.05**), we get the following equation  **RET** = **-2.64216** - **GRI** \***2.11046** + **SAT** \* **0.005735**  It follows from the above equation, that Excess return is dependent on SAT and GRI. For the 2 Fund managers with same GRI, Excess Return(RET) varies with SAT of the college. Summarizing the RET calculation for Bob and Putney using the above equation we get   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | **College** | **SAT** | **GRI** | **Equation** | **Return** | | **Bob** | Ohio | 1042 | 1 | -2.64216 - 1 \*2.11046 + (1042) \* 0.005735 | 1.2232499999999993 | | **Putney** | Princeton | 1355 | 1 | -2.64216 - 1 \*2.11046 + (1355) \* 0.005735 | **3.018304999999999** |   From the table it is clear that Putney will generate a higher return and the return would be  **1.7950549999999996**% higher on an average. at **(α = 0.05)**  b)  if hired by AMBTPM, there is no change in the above variables. On joining the **AMBTPM,** There will be change in TEN variable( TEN = 0 ). But at **α = 0.05,** TEN is not statistically significant . Therefore, we can conclude that **Putney** would outperform Bob after joining AMBTPM and will on an average generate a higher return of **1.75%**  at **α = 0.05** |

|  |  |
| --- | --- |
| **Q 3** | 1. Can you prove at the 5 percent significance level that if Bob had attended Princeton instead of Ohio State, then the return of his current fund would be greater? 2. Can you prove at the 10 percent level of significance that if Bob were managing a growth fund instead of a growth and income fund, then he would achieve at least 1 percent higher average returns? |
| **Solution** | **a)**  Using Mulitple Linear Regression ( and selecting only statistically significant variables at **α=0.05**), we get the following equation  **RET** = **-2.64216** - **GRI** \***2.11046** + **SAT** \* **0.005735**  It follows from the above equation, that Excess return is dependent on SAT and GRI. If **GRI** is held constant, the Excess Return is varies with the SAT of the college. Princeton has a Higher SAT average than Ohio State. Substituting the SAT values, we get the following Returns   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **College** | **SAT** | **GRI** | **Equation** | **Return** | | Ohio | 1042 | 1 | -2.64216 - 1 \*2.11046 + (1042) \* 0.005735 | 1.2232499999999993 | | Princeton | 1355 | 1 | -2.64216 - 1 \*2.11046 + (1355) \* 0.005735 | **3.018304999999999** |   As per the table(and from the **MLR equation**) Bob would generate a higher return if he had attended Princeton  **Important Observations:**   * It is important to note that Regression only establishes **association relationship** and does not establish a **causal relationship** between the dependent and independent variables. Therefore, it is not accurate to conclude that Bob would generate higher returns had he attended Princeton. * There is no variable in the given data that can be used to show an association between the University and Excess returns. An Association does exist between SAT scores and Excess Return. However students score SAT without help of universities. Therefore, the above equation does not conclusively prove that Bob would have generated higher return had he attended Princeton   b)  Using Multiple Linear Regression ( and selecting only statistically significant variables at **α=0.1**), we get the following equation  **RET** = **-2.64216** - **GRI** \***2.11046** + **SAT** \* **0.005735**  Keeping SAT contant, we get the following results for Growth and Growth and Income funds   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Fund type** | **SAT** | **GRI** | **Equation** | **Return** | | Growth | **Const** | 0 | -2.64216 - 0 \*2.11046 + (const) \* 0.005735 | -2.64216 + const | | Growth & Income | **Const** | 1 | -2.64216 - 1 \*2.11046 + (const) \* 0.005735 | **-4.75262 + const** |   Therefore, if Bob were managing only 'Growth' he would have generated a return **2.11046%** higher on average. Thus he would have generated at least 1% higher average returns |