BUSN. 3421 – TR-2

Assignment #3

Group #4

Okunola Uthman

Christian Perez

Maria Ramirez – Teles

Versai Rodriguez

**QUESTIONS**

1. Develop an estimated simple linear regression model (*Analyze/Regression/Linear*) that can be used to predict the *alumni giving rate*, given the *graduation rate* (10 points):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .756a | .571 | .562 | 8.894 |
| a. Predictors: (Constant), Graduation Rate | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Anova** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 4852.462 | 1 | 4852.462 | 61.339 | <.001b |
| Residual | 3639.017 | 46 | 79.109 |  |  |
| Total | 8491.479 | 47 |  |  |  |
| a. Dependent Variable: Alumni Giving Rate | | | | | | |
| b. Predictors: (Constant), Graduation Rate | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficients** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -68.761 | 12.583 |  | -5.465 | <.001 |
| Graduation Rate | 1.181 | .151 | .756 | 7.832 | <.001 |
| a. Dependent Variable: Alumni Giving Rate | | | | | | |

1. What is the estimated regression equation? Interpret the y-intercept and slope in the current context.

* **y-intercept b0**= -68.761 (The **y-intercept** represents the predicted alumni giving rate when the **graduation rate is 0%**)
* **slope b1**= 1.181(The **Slope** means for every **1% increase in graduation rate**, the predicted **alumni giving rate increases by 1.181%**)
* Estimated Regression Equation is: ŷ = b0 + b1x  
  **ŷ = -68.761 + 1.181x**(where x = *Graduation Rate*, **ŷ** = Est *Alumni Giving Rate*)

1. What is the coefficient of determination? Interpret it in the current context.

* The **R Squared value** is the coefficient of determination. This tells you how much variation in the **Alumni Giving Rate** is explained by the **Graduation Rate**. It is .571 which means that only 57.1% of the variation in alumni giving rate is explained by graduation rate.

1. Using the output of the estimated simple linear regression model in question-1, perform hypothesis testing to see if there is a significant linear relationship between the variables (10 points):
2. State the null and the alternate hypothesis in complete statements for the current context.

* **Null Hypothesis**: There is **no significant linear** relationship between Graduation Rate and Alumni Giving Rate. (Graduation Rate has no effect on Alumni Giving Rate.)
* **Alternative Hypothesis:** There **is a significant** linear relationship between Graduation Rate and Alumni Giving Rate. (Graduation Rate affects Alumni Giving Rate.)

1. Is there a significant linear relationship between the variables? Explain your reasoning.

* We look at the **p-value** for the Graduation Rate coefficient. Since the **p-value (.001)** is **less than** the alpha level **(α = 0.05)**, we **reject** the null hypothesis and accept the Alternate hypothesis

1. Develop an estimated multiple linear regression model (*Analyze/Regression/Linear)* that could be used to predict the *alumni giving rate* using *Graduation Rate*, *% of Classes under 20*, and *Student/ Faculty Ratio* as independent variables (20 points):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .837a | .700 | .679 | 7.610 |
| a. Predictors: (Constant), Student-Faculty Ratio, Graduation Rate, % of Classes Under 20 | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 5943.531 | 3 | 1981.177 | 34.213 | <.001b |
| Residual | 2547.948 | 44 | 57.908 |  |  |
| Total | 8491.479 | 47 |  |  |  |
| a. Dependent Variable: Alumni Giving Rate | | | | | | |
| b. Predictors: (Constant), Student-Faculty Ratio, Graduation Rate, % of Classes Under 20 | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficients** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -20.720 | 17.521 |  | -1.183 | .243 |
| Graduation Rate | .748 | .166 | .479 | 4.508 | <.001 |
| % of Classes Under 20 | .029 | .139 | .029 | .208 | .836 |
| Student-Faculty Ratio | -1.192 | .387 | -.430 | -3.082 | .004 |
| a. Dependent Variable: Alumni Giving Rate | | | | | | |

What is the estimated regression equation? Interpret the slopes and the y-intercept?

* **Y-intercept (b₀ = -20.720):** This is the predicted Alumni Giving Rate when Graduation Rate, % of Classes Under 20, and Student-Faculty Ratioare all 0.
* **Graduation Rate (b₁ = 0.748):** For every 1% increase in Graduation Rate, the predicted Alumni Giving Rate increases by **0.748 percentage,** holding all other variables constant.
* **% of Classes Under 20 (b₂ = 0.029):** For every 1% increase in the percentage of classes with fewer than 20 students, the predicted Alumni Giving Rate increases by **0.029 percentage**, assuming all other factors remain constant. (Note: This effect is small and not statistically significant, p = 0.836)
* **Student-Faculty Ratio (b₃ = -1.192):** For every 1 unit increase in the student-Faculty Ratio, the predicted Alumni Giving Rate decreases by **1.192 percentage**, assuming all other factors remain constant.
* **Estimated Regression Equation**: ŷ = -b0 + b1x1 + b2x2  - b3x3  
  ŷ = −20.720 + 0.748 x1  ​ + 0.029 x2  ​−1.192 x3 ​ (ŷ = Estimated Alumni Giving Rate, x1 = Graduation Rate, x2 = % of Classes Under 20, x3 = Student-Faculty Ratio)

1. What is the coefficient of determination? Interpret it in the current context.

* The **R Squared value** isthe coefficient of determination. This tells you how much variation in the **Alumni Giving Rate** is explained by Graduation rates, % of class under 20 & Student faculty ratio. It is .700 which means that 70% of the variation in alumni giving rate is explained by Graduation rates, % of class under 20 and Student faculty ratio.

1. For the multiple regression model output from question-3, perform hypothesis testing to see if there is a significant relationship between the dependent variable and each independent variable (30 points): *For each independent variable* you need to do the following:

a) show the null and alternate hypotheses

b) is there a significant relationship between the dependent variable and the independent variable, keeping the others constant. Explain the reasoning.

**% of classes under 20**

* **Null Hypothesis**: There is **no significant** relationship between % of classes under 20 and alumni giving rate
* **Alternate Hypothesis:** There is **a significant** relationship between % of classes under 20 and alumni giving rate
* **Comparing P value:** The P value for **% of** **classes under 20 is** .836 and since its greater than significance level **(α = 0.05)**, we accept the null hypothesis thereby concluding that there is **no significant statistical relationship** between **alumni giving rate** and **% of class under 20**. It should not be used to predict alumni giving rates

**Student Faculty Ratio**

* **Null Hypothesis:** There is **no significant** relationship between student faculty ratio and predicted alumni giving rate
* **Alternative Hypothesis:** There is **a significant** relationship between student faculty ratio and predicted alumni giving rate
* **Comparing P value**: The P value for student faculty ratio is .004 and since its less than the Alpha level **(α = 0.05)**, we reject the null hypothesis thereby concluding that there is **a significant statistical relationship** between **alumni giving rate** and **student faculty ratio** and it can be used to predict alumni giving rate.

**Graduation Rates**

* **Null Hypothesis:** There is **no significant** relationship between graduation rates and predicted alumni giving rate
* **Alternative Hypothesis:** There is **a significant** relationship between student Graduation rates and predicted alumni giving rates.
* **Comparing P value**: The P value for graduation rates is <.001 and since its less than the Alpha level **(α = 0.05)**, we reject the null hypothesis thereby concluding that there is a **significant statistical relationship** between **alumni giving rate** and **graduation rates** and it can be used to predict alumni giving rate.

5). From the multiple regression model in question-3, test for multicollinearity between the independent variables. Show the output. Discuss if there is multicollinearity and explain your reasoning (10 points).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | Graduation Rate | % of Classes Under 20 | Student-Faculty Ratio |
| 1 | 1 | 3.802 | 1.000 | .00 | .00 | .00 | .00 |
| 2 | .186 | 4.527 | .00 | .00 | .03 | .19 |
| 3 | .010 | 19.568 | .03 | .24 | .90 | .34 |
| 4 | .003 | 37.973 | .97 | .76 | .07 | .47 |
| a. Dependent Variable: Alumni Giving Rate | | | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Coefficients** | | | |
| Model | | Collinearity Statistics | |
| Tolerance | VIF |
| 1 | Graduation Rate | .604 | 1.656 |
| % of Classes Under 20 | .365 | 2.742 |
| Student-Faculty Ratio | .350 | 2.856 |
| a. Dependent Variable: Alumni Giving Rate | | | |

**Tolerance** values below **0.10** indicate **serious multicollinearity**. **VIF (Variance Inflation Factor)** values above **10** typically indicate **high multicollinearity**

* All the observed **Tolerance** values are **above** **0.10** and **VIF** values are **well below 10,** indicating **no severe multicollinearity**.Based on **VIF** and **Tolerance** there is **no severe multicollinearity**.

6). Develop an estimated multiple linear regression model (*Analyze/Regression/Linear)* to predict the alumni giving rate, *excluding* the “% of classes under 20” (20 points):

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summary** | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .836a | .700 | .686 | 7.528 | .700 | 52.411 | 2 | 45 | <.001 |
| a. Predictors: (Constant), Graduation Rate, Student-Faculty Ratio | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 5941.015 | 2 | 2970.508 | 52.411 | <.001b |
| Residual | 2550.464 | 45 | 56.677 |  |  |
| Total | 8491.479 | 47 |  |  |  |
| a. Dependent Variable: Alumni Giving Rate | | | | | | |
| b. Predictors: (Constant), Graduation Rate, Student-Faculty Ratio | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficients** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -19.106 | 15.550 |  | -1.229 | .226 |
| Student-Faculty Ratio | -1.246 | .284 | -.450 | -4.382 | <.001 |
| Graduation Rate | .756 | .160 | .484 | 4.717 | <.001 |
| a. Dependent Variable: Alumni Giving Rate | | | | | | |

a). What is the estimated regression equation? Interpret the slopes.

* ŷ = -19.106 -1.246x1 + 0.756x2

**Interpretation of the Slopes:**

* **Graduation Rate (b₁ = 0.756):**  
  For every 1 percentage point increase in the **Graduation Rate**, the predicted **Alumni Giving Rate** increases by **0.756**, holding **Student-Faculty Ratio constant**.
* **Student-Faculty Ratio (b₂ = -1.246):**  
  For every 1 unit increase in the **student-Faculty Ratio**, the predicted **Alumni Giving Rate** decreases by **1.246**, holding **Graduation Rate constant**.

b). How does this model compare to the earlier one in question-3 in terms of offering a good fit? Explain your reasoning.

* The removed variable (**% of Classes Under 20**) had an **insignificant** effect in the earlier model (p = .836), so removing it improves the clarity of the model and provides a better fit. The remaining predictors (**Graduation Rate and Student-Faculty Ratio**) are both **highly significant** and have even **stronger t-values**. This suggests the new model is more (simpler with no loss of predictive power) and has significantly **better fit** and **interpretability**.