**Case Problem Alumni Giving**

Alumni donations are an important source of revenue for colleges and universities. If administrators could determine the factors that influence increases in the percentage of alumni who make a donation, they might be able to implement policies that could lead to increased revenues. Research shows that students who are more satisfied with their contact with teachers are more likely to graduate. As a result, one might suspect that smaller class sizes and lower student-faculty ratios might lead to a higher percentage of satisfied graduates, which in turn might lead to increases in the percentage of alumni who make a donation. Table 14.13 shows data for 48 national universities (America’s Best Colleges, Year 2000 ed.).The column labelled % of Classes under 20 shows the percentage of classes offered with fewer enrolled divided by the total number of faculty. Finally, the column labelled Alumni Giving Rate is the percentage of alumni that made a donation to the university.   
  
**Managerial Report**

**Introduction:**

The data set under consideration are % of class under 20, Student/Faculty Ratio and Alumni Giving Rate 48 national universities. First of all we have to summaries the data set using summary statistics and graphs. Then we develop an estimated regression equation that could be used to predict the alumni giving rate giving the percentage of classes with fewer than 20 students. Also an estimated regression equation that could be used to predict the alumni giving rate given the student-faculty ratio is also provided. The residuals are analysed to assess the validity of the assumptions of the regression analysis.

1. Develop numerical and graphical summaries of the data.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **% of Classes Under 20** | **Student/Faculty Ratio** | **Alumni Giving Rate** |
| Mean | 55.72916667 | 11.54166667 | 29.27083333 |
| Standard Error | 1.904348231 | 0.700150904 | 1.940091114 |
| Median | 59.5 | 10.5 | 29 |
| Mode | 65 | 13 | 13 |
| Standard Deviation | 13.19371156 | 4.850787755 | 13.44134552 |
| Sample Variance | 174.0740248 | 23.53014184 | 180.6697695 |
| Kurtosis | -0.956615467 | -0.440337538 | -0.071942009 |
| Skewness | -0.500612795 | 0.581838493 | 0.370106739 |
| Range | 48 | 20 | 60 |
| Minimum | 29 | 3 | 7 |
| Maximum | 77 | 23 | 67 |
| Sum | 2675 | 554 | 1405 |
| Count | 48 | 48 | 48 |

Mean is defined as the arithmetic average of data, that is, the sum of all the numbers divided by the number of observations contributing to that sum. The mean represents the balance point, or centre of gravity of the distribution and it is the most common measure of central tendency.

Median is the middle most observation of the data, determined after all items are arranged in ascending or descending order of magnitude. In other words, median is that observation above which and below which lie 50% of all the data.

The Mode is simply the most frequently occurring item in a distribution.

Range measures the scatter of the number of coupons used among them and not about an average.

The value of the variance is interpreted as the average squared deviation of the observations used and can be used to compare different-sized distributions. Variance gives an idea of how much the observations in the data differ from the mean of the data. The variance and the standard deviation are both measures of the spread of the distribution about the mean.

The standard deviation is the square root of variance, a measure of the degree of dispersion of the data from the mean value. A large standard deviation indicates that the data points are far from the mean and a small standard deviation indicates that they are clustered closely around the mean.

The histogram and box plot suggests that the distribution of % of classes under 20 is slightly negatively skewed, since most of the scores tend to occur toward the upper end of the scale while increasingly fewer scores occur toward the lower end.

The histogram and box plot suggests that the Student/Faculty Ratio is approximately normally distributed.

The histogram and box plot suggests that the Alumni Giving Rate is approximately normally distributed.

2. Use regression analysis to develop an estimated regression equation that could be used to predict the alumni giving rate giving the percentage of classes with fewer than 20 students.

The general form of simple linear regression is Y= a + bX

Where Y is the dependent variable and X is the independent variable, a and be are known as the regression coefficients .They are estimated by the method of least squares. The estimates of a and b are given by

 

The parameter b measures the impact of unit change in X on the dependent variable Y. It is the slope of the regression line. The parameter a is the value of Y when X=0. It is known as the Intercept term.

The regression equation can be used to predict the value of Y for a given X. The predicted value of Y is given by 

N = 48, = 2675,= 1405, = 157257,= 49617,= 83681 (Please see the excel spreadsheet)

The estimated value of slope b = 0.6578

Estimated intercept a = -7.3861

The regression equation, Y = 0.6578\*X – 7.3861

That is, Alumni giving rate = 0.6578 \* % of class under 20 – 7.3861

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | -7.386067617 | 6.565472287 | -1.124986489 | 0.266430663 |
| % of Classes Under 20 | 0.65776869 | 0.114704802 | 5.734447737 | 7.22812E-07 |

3. Use regression analysis to develop an estimated regression equation that could be used to predict the alumni giving rate given the student-faculty ratio.

The general form of simple linear regression is Y= a + bX

Where Y is the dependent variable and X is the independent variable, a and be are known as the regression coefficients .They are estimated by the method of least squares. The estimates of a and b are given by

 

The parameter b measures the impact of unit change in X on the dependent variable Y. It is the slope of the regression line. The parameter a is the value of Y when X=0. It is known as the Intercept term.

The regression equation can be used to predict the value of Y for a given X. The predicted value of Y is given by 

N = 48, = 2675,= 1405, = 157257,= 49617,= 83681 (Please see the excel spreadsheet)

The estimated value of slope b = -2.0572

Estimated intercept a = 53.0138

The regression equation, Y = -2.0572\*X + 53.0138

That is, Alumni giving rate = -2.0572 \* Student/Faculty Ratio + 53.0138

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | 53.01382714 | 3.421450432 | 15.49454776 | 7.05881E-20 |
| Student/Faculty Ratio | -2.057154698 | 0.273716035 | -7.515652861 | 1.54423E-09 |

4. Which of the two estimated regression equations provide the best fit? For this estimated regression equation, perform an analysis of the residuals and discuss your findings and conclusions.

**Regression Analysis: Alumni giving rate Vs % of class under 20**

The model adequacy is measured using the R2 value. Here R2 = 0.4169. Thus 41.69% variability in the Alumni giving rate can be explained by the regression model.

Standard error of estimate = 10.37522473

As the standard error of estimate increases, the accuracy of prediction decreases.

|  |  |
| --- | --- |
| **Regression Statistics** | |
| Multiple R | 0.645650419 |
| R Square | 0.416864464 |
| Adjusted R Square | 0.404187604 |
| Standard Error | 10.37522473 |
| Observations | 48 |

**Regression Analysis: Alumni giving rate Vs Student/Faculty Ratio**

The model adequacy is measured using the R2 value. Here R2 = 0.5512. Thus 55.12% variability in the Alumni giving rate can be explained by the regression model.

Standard error of estimate = 9.102515793

|  |  |
| --- | --- |
| **Regression Statistics** | |
| Multiple R | 0.742397463 |
| R Square | 0.551153993 |
| Adjusted R Square | 0.541396472 |
| Standard Error | 9.102515793 |
| Observations | 48 |

Here the R2 value is higher and standard error of estimate is lower for the second model. Hence the regression analysis between Alumni giving rate and Student/Faculty Ratio provides the best fit.

The normal probability plot and the histogram of residuals suggest that the residuals are approximately normally distributed.

Since there is no specific pattern for the points, the points on the plot of residuals against the fitted value are at random. Hence we can conclude that the errors are independent and have a constant variance.

5. What conclusions and recommendations can you derive from your analysis?

The estimated regression equation is given by

Alumni giving rate = -2.0572 \* Student/Faculty Ratio + 53.0138

The regression coefficient can be interpreted as

For a unit increase in the Student/Faculty Ratio, the Alumni giving rate decreases by 2.0572 units.

The significance of the regression coefficient is tested using t test.

Here the regression coefficient “Student/Faculty Ratio” is highly significant with p-value less than 0.05.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | 53.01382714 | 3.421450432 | 15.49454776 | 7.05881E-20 |
| Student/Faculty Ratio | -2.057154698 | 0.273716035 | -7.515652861 | 1.54423E-09 |

The overall significance of the model is tested using F test.

Here the F statistic is highly significant with p-value less than 0.05. Hence the regression model is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ANOVA** | | | | | |
|  | **df** | **SS** | **MS** | **F** | **Significance F** |
| Regression | 1 | 4680.112653 | 4680.112653 | 56.48503792 | 1.54423E-09 |
| Residual | 46 | 3811.366513 | 82.85579377 |  |  |
| Total | 47 | 8491.479167 |  |  |  |