V506 Homework Exercise 6

# Authors: Jivitesh Poojary and Qiwen Zhu

Part I: Bivariate Regression without SAS

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | Quit (Y) | Unemployment (X) | Quit x Unemployment |  | Quit2 | Unemployment2 |
| 1960 | 1.1 | 6.2 | 6.82 |  | 1.21 | 38.44 |
| 1961 | 1.2 | 7.8 | 9.36 |  | 1.44 | 60.84 |
| 1962 | 1.5 | 6 | 9 |  | 2.25 | 36 |
| 1963 | 1.5 | 5.7 | 8.55 |  | 2.25 | 32.49 |
| 1964 | 1.6 | 5 | 8 |  | 2.56 | 25 |
| 1965 | 1.8 | 4.6 | 8.28 |  | 3.24 | 21.16 |
| 1966 | 2.6 | 3.2 | 8.32 |  | 6.76 | 10.24 |
| 1967 | 2.4 | 3.6 | 8.64 |  | 5.76 | 12.96 |
| 1968 | 2.5 | 3.3 | 8.25 |  | 6.25 | 10.89 |
| 1969 | 3 | 3.3 | 9.9 |  | 9 | 10.89 |
| 1970 | 2 | 5.6 | 11.2 |  | 4 | 31.36 |
| 1971 | 1.8 | 6.8 | 12.24 |  | 3.24 | 46.24 |
| 1972 | 2.5 | 5.6 | 14 |  | 6.25 | 31.36 |
|  |  |  |  |  |  |  |
| Sum | 25.5 | 66.7 | 122.56 |  | 54.21 | 367.87 |
| Count | 13 | 13 |  |  |  |  |
| Average | 1.961538 | 5.130769231 |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Quit\_diff | Unemployment\_diff | | Quit\_diff2 | Unemployment\_diff2 |
| 1960 | -0.861538462 | | 1.069230769 | 0.742248521 | -0.921183432 |
| 1961 | -0.761538462 | | 2.669230769 | 0.579940828 | -2.032721893 |
| 1962 | -0.461538462 | | 0.869230769 | 0.213017751 | -0.401183432 |
| 1963 | -0.461538462 | | 0.569230769 | 0.213017751 | -0.262721893 |
| 1964 | -0.361538462 | | -0.130769231 | 0.130710059 | 0.047278107 |
| 1965 | -0.161538462 | | -0.530769231 | 0.026094675 | 0.085739645 |
| 1966 | 0.638461538 | | -1.930769231 | 0.407633136 | -1.232721893 |
| 1967 | 0.438461538 | | -1.530769231 | 0.192248521 | -0.671183432 |
| 1968 | 0.538461538 | | -1.830769231 | 0.289940828 | -0.985798817 |
| 1969 | 1.038461538 | | -1.830769231 | 1.078402367 | -1.901183432 |
| 1970 | 0.038461538 | | 0.469230769 | 0.00147929 | 0.018047337 |
| 1971 | -0.161538462 | | 1.669230769 | 0.026094675 | -0.26964497 |
| 1972 | 0.538461538 | | 0.469230769 | 0.289940828 | 0.252662722 |
|  |  | |  |  |  |
| Sum |  | |  | 4.190769231 | -8.274615385 |

**= 66.7  = 25.5 = 122.56 = 367.87**

**= 54.21 = 5.13076 = 1.96154**

**The Regression Equation** (equations 3.23, 3.22, and 3.19):





Indicates that an increase of one percent increase in the unemployment rate decreases quit rate/100 employee by 1.97448. It also indicates that if the unemployment rate is 0 we will have a quite rate at the level of 12.09217. However, 0 unemployment rate was not included in our sample so extrapolations of this nature are unreliable.

**Quite Rate /100 employee = 12.09217 – 1.97448 (Unemployment rate) + u(error term)**

**Coefficient of Determination** (equation 3.34):





This *r* value indicates the existence of a strong positive relationship between *X* and *Y*, and implies that 63.70% of the variation in Quite Rate is associated with (or “explained by”) the unemployment rate, and that the model fits the data quite well.

***F* Test** using the formula on page 87 of V506 notes:

*H*0: *ρ* = 0 Critical Value: *F*.05(1, 11) = 4.84

*H*1: *ρ≠* 0



Since 19.30 > 4.84, reject *H*0 at the .05 level of significance. The *F* test examines the significance or fit of the entire regression.

**Standard Error of the Slope Coefficient** (equation 3.38):

The standard error of  measures the average amount of variation in the estimate of *β*2.



***t* test** (equation 3.36):

*H*0: *β*2 = 0 Critical Value: *t*.025(11)  = -2.201

*H*1: *β*2 ≠ 0



-4.3937 < -2.201, Therefore reject *H*0. The *t* test examines the significance of a single parameter. In this case, we conclude that  is statistically significant at the .05 level.

**95% Confidence Interval for *β*2** (equation 3.39):



This interval will contain the actual value of *β*2 95% of the time.

**Standard Error of the Regression** on page 88 of V506 notes:



This is the average amount by which the predicted values of Quit Rate derived from the regression equation differ from the observed or actual levels. Thus, it is the “average error” in predicting the dependent variable.

Question 4.

Yi = 3.1237 – 0.1714 Xi

Se(B2) = 0.0210

R2 = 0.8575

**Equation:**

Yes, this equation is different from the one obtained by us in step 1.

The obtained equation is: Yi = 12.09217 – 1.97448 Xi + u(error term)

The Given equation is: Yi = 3.1237 – 0.1714 Xi

The obtained equation has a steeper slope implying the quite rate will be more for a unit change in the employment rate. Similarly, the intercept in the obtained equation is also more implying when the unemployment rate is 0 the Quite rate is also higher than the given equation.

**Standard Error of Slope:**

SE(B2) obtained : 0.4494

SE(B2) Given: 0.0210

This implies that the standard error in the slope obtained is higher implying a higher variability in the slope.

R2 Value;

R2 obtained: 0.6370

R2 given: 0.8575

Is the value of the R2 obtained after calculation is lesser than the given value we can say that there is lesser correlation between the two variables in our obtained equation than in the given equation.

Part II: Bivariate Regression with SAS

1A)

**SAS CODE:**

TITLE "V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.MFtemp;

SET V506.Mutualfund;

**RUN**;

**PROC** **REG** ;

MODEL EXP\_RTRN / ALPHA=**0.01**;

**RUN**;

**OUTPUT:**

|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: EXP\_RTRN**

|  |  |
| --- | --- |
| **Number of Observations Read** | 34 |
| **Number of Observations Used** | 34 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 136.88841 | 136.88841 | 74.23 | <.0001 |
| **Error** | 32 | 59.01394 | 1.84419 |  |  |
| **Corrected Total** | 33 | 195.90235 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 1.35801 | **R-Square** | 0.6988 |
| **Dependent Mean** | 13.64118 | **Adj R-Sq** | 0.6893 |
| **Coeff Var** | 9.95521 |  |  |

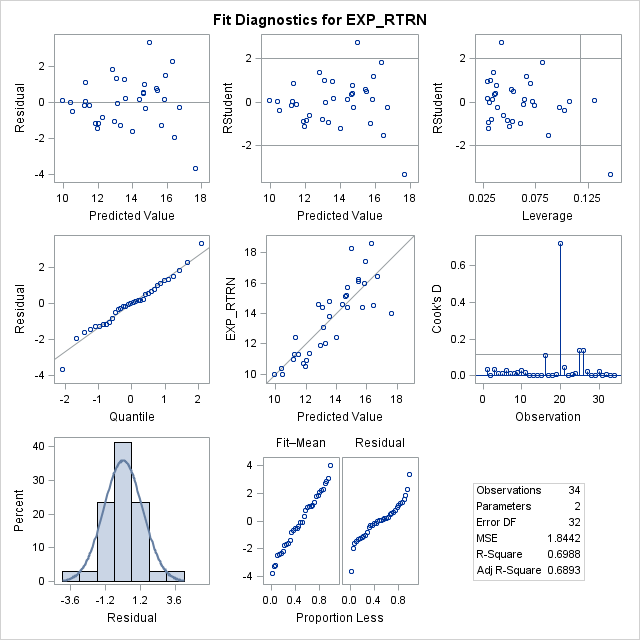
| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 5.54094 | 0.96861 | 5.72 | <.0001 |
| **SD\_RTRN** | 1 | 0.47451 | 0.05508 | 8.62 | <.0001 |

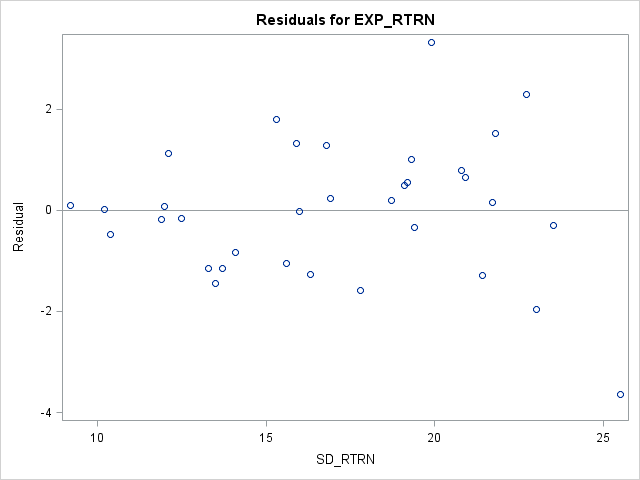
|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

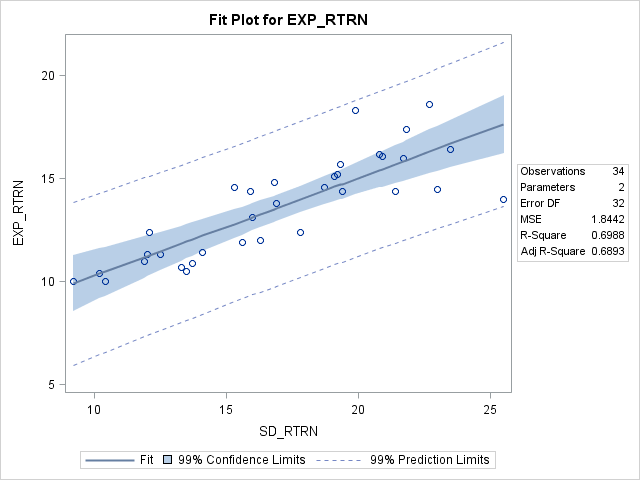
**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: EXP\_RTRN**







**INTERPRETATION:**

The **F statistic** = 74.23 is significant (p-value) at the .0001 level, indicating that we can reject the null hypothesis that H0: ρ = 0. We can conclude that *the data are consistent with some type of relationship between the dependent and independent variables*.

The **R2 value** of 0.6988 indicates that *69.88 percent of the variation in the expected return on portfolio is explained by or associated with standard deviation of return*. This is a strong relationship.

The **regression equation** would be **EXP\_RTRN = 5.54094 + 0.47451 SD\_RTRN + u-hat**

This implies that *the expected return on portfolio would increase by 0.47451 percent given a one-unit increase in standard deviation of return*. The intercept (constant) value of 5.54094 implies that *if standard deviation of return was equal to zero, the expected return on portfolio would be* 5.54094 *percent.* This is unrealistic because the standard deviation of a real world data cannot be 0.

The **t-statistic** of 8.62 and **p-value** of .0001 **for SD\_RTRN** indicates that *this independent variable has a statistically significant relationship with expected return on portfolio*; i.e., we can reject the null hypothesis H0: β2 = 0 at the .0001 level of significance. Thus, we can conclude that *there appears to be a significant relationship between expected return on portfolio and standard deviation of return*.

The **root mse** of 1.35801 represents the **standard error of the regression**, and is *the average error made in predicting the dependent variable using the regression equation* (i.e., the average difference between y-hat and y). The smaller this is as a percentage of the **dependent variable mean**, the better the prediction. That percentage is provided as the **coefficient of variation**, and indicates that *the average error in predicting expected return on portfolio is 1.358 percent*. This implies good predictive accuracy.

We have plotted the residuals against the predicted values of the dependent variable.

1B)

**SAS CODE:**

TITLE "V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.MFtemp;

SET V506.Mutualfund;

**RUN**;

**DATA** V506.MFtemp2;

SET V506.MFtemp;

LNSD\_RTRN = LOG(SD\_RTRN);

**RUN**;

**PROC** **REG** ;

MODEL LNSD\_RTRN = EXP\_RTRN / ALPHA=**0.01**;

**RUN**;

**OUTPUT:**

|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: EXP\_RTRN**

|  |  |
| --- | --- |
| **Number of Observations Read** | 34 |
| **Number of Observations Used** | 34 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 139.66505 | 139.66505 | 79.47 | <.0001 |
| **Error** | 32 | 56.23731 | 1.75742 |  |  |
| **Corrected Total** | 33 | 195.90235 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 1.32568 | **R-Square** | 0.7129 |
| **Dependent Mean** | 13.64118 | **Adj R-Sq** | 0.7040 |
| **Coeff Var** | 9.71819 |  |  |

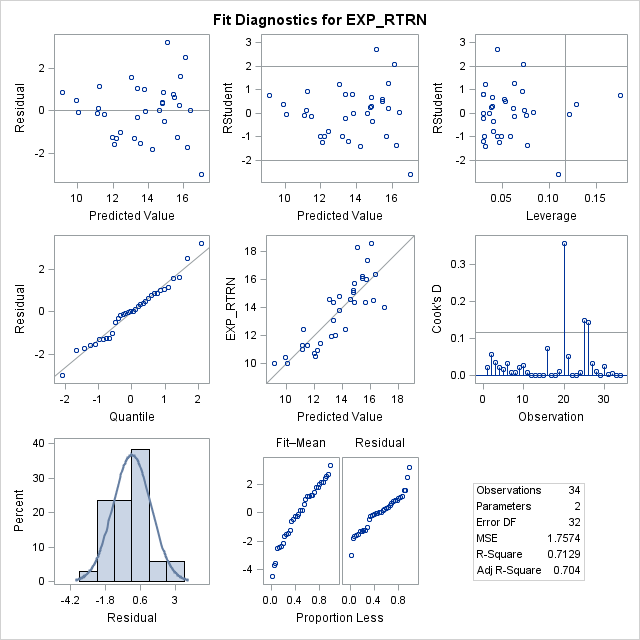
| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | -8.05553 | 2.44441 | -3.30 | 0.0024 |
| **LNSD\_RTRN** | 1 | 7.73656 | 0.86784 | 8.91 | <.0001 |

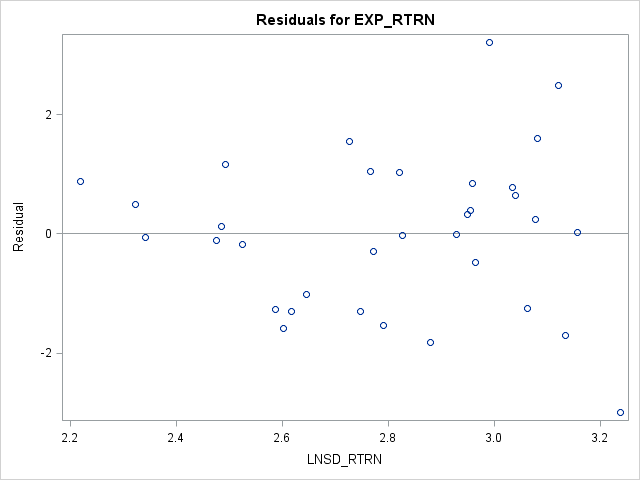
|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

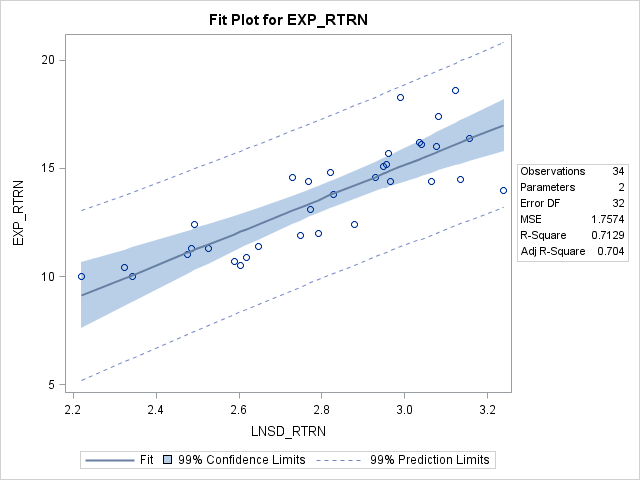
**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: EXP\_RTRN**







**INTERPRETATION:**

The **F statistic** = 79.47 is significant (p-value) at the .0001 level, indicating that we can reject the null hypothesis that H0: ρ = 0. We can conclude that *the data are consistent with some type of relationship between the dependent and independent variables*.

The **R2 value** of 0.7129 indicates that *71.29 percent of the variation in the expected return on portfolio is explained by or associated with natural log of standard deviation of return*. This is a strong relationship.

The **regression equation** would be **EXP\_RTRN = -8.05553 + 7.73656** **LNSD\_RTRN + u-hat**

This implies that *the expected return on portfolio would increase by 7.73656 percent given a one-unit increase in natural log standard deviation of return*. The intercept (constant) value of **-8.05553** implies that *if natural log standard deviation of return was equal to zero, the expected return on portfolio would be* **-8.05553** *percent.* This means that we will have a negative value of *expected return on portfolio*. When the *standard deviation of return is 1.*

The **t-statistic** of 8.91 and **p-value** of .0001 **for SD\_RTRN** indicates that *this independent variable has a statistically significant relationship with expected return on portfolio*; i.e., we can reject the null hypothesis H0: β2 = 0 at the .0001 level of significance. Thus, we can conclude that *there appears to be a significant relationship between expected return on portfolio and natural log standard deviation of return*.

The **root mse** of 1.32568 represents the **standard error of the regression**, and is *the average error made in predicting the dependent variable using the regression equation* (i.e., the average difference between y-hat and y). The smaller this is as a percentage of the **dependent variable mean**, the better the prediction. That percentage is provided as the **coefficient of variation**, and indicates that *the average error in predicting expected return on portfolio is 1.358 percent*. This implies good predictive accuracy.

We have plotted the residuals against the predicted values of the dependent variable.

2 A)

**SAS CODE:**

TITLE "V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.REtemp;

SET V506.RealEstate;

**RUN**;

**PROC** **REG** ;

MODEL PRICE = SIZE;

**RUN**;

**OUTPUT:**

|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Price Price**

|  |  |
| --- | --- |
| **Number of Observations Read** | 105 |
| **Number of Observations Used** | 105 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 33340 | 33340 | 17.39 | <.0001 |
| **Error** | 103 | 197472 | 1917.20824 |  |  |
| **Corrected Total** | 104 | 230812 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 43.78594 | **R-Square** | 0.1444 |
| **Dependent Mean** | 221.10056 | **Adj R-Sq** | 0.1361 |
| **Coeff Var** | 19.80363 |  |  |

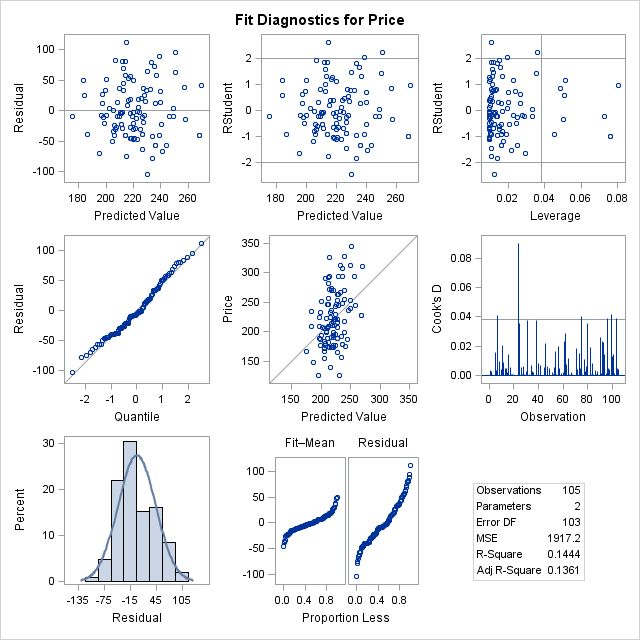
| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | 1 | 60.85146 | 38.66491 | 1.57 | 0.1186 |
| **Size** | Size | 1 | 0.07182 | 0.01722 | 4.17 | <.0001 |

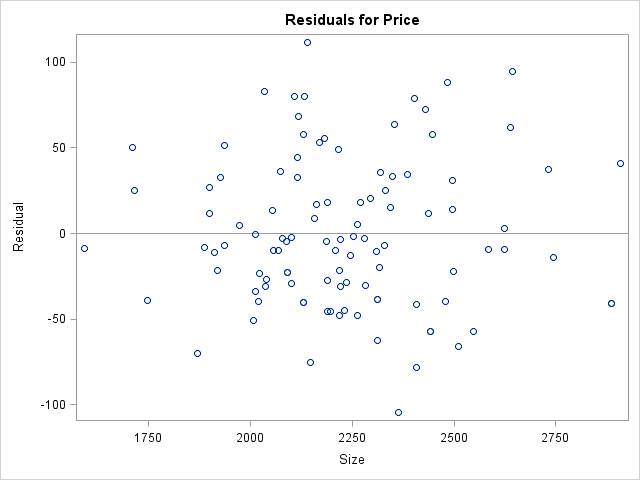
|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

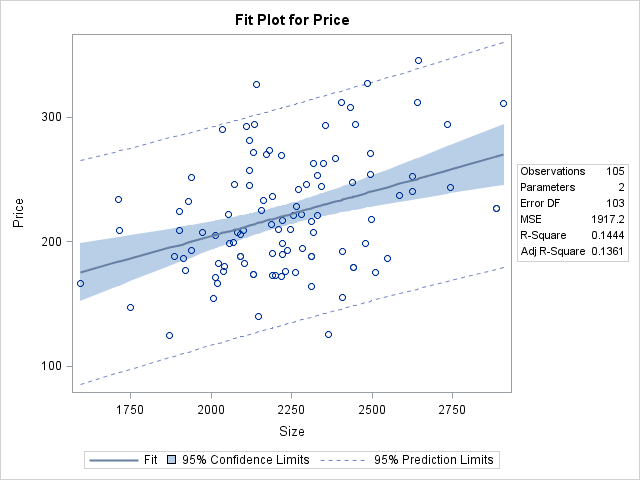
**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Price Price**







**INTERPRETATION:**

The **F statistic** = 17.39 is significant (p-value) at the .0001 level, indicating that we can reject the null hypothesis that H0: ρ = 0. We can conclude that *the data are consistent with some type of relationship between the dependent and independent variables*.

The **R2 value** of 0.1444 indicates that *14.44 percent of the variation in the selling price of the house is explained by or associated with size of the house*. This is a weak relationship.

The **regression equation** would be **Price = 60.85146 + 0.07182 Size + u-hat**

This implies that *the selling price of the house would increase by 0.07182 units given a one-unit increase in size of the house*.Note that since this variable is measured in thousands, a one-unit increase is actually equivalent to a $1,000 increase *total team salary*. It is critical to interpret the results in terms of the units of measurement.The intercept (constant) value of 60.85146 implies that *if size of the house was equal to zero, the selling price of the house would be 60.85146 percent.* This is unrealistic because if the size of the house is 0, the house would not exist.

The **t-statistic** of 4.17 and **p-value** of .0001 **for Size** indicates that *this independent variable has a statistically significant relationship with Price*; i.e., we can reject the null hypothesis H0: β2 = 0 at the .0001 level of significance. Thus, we can conclude that *there appears to be a significant relationship between size of the house and selling price of the house*.

The **root mse** of 43.78594 represents the **standard error of the regression**, and is *the average error made in predicting the dependent variable using the regression equation* (i.e., the average difference between y-hat and y). The smaller this is as a percentage of the **dependent variable mean**, the better the prediction. That percentage is provided as the **coefficient of variation**, and indicates that *the average error in predicting selling price of the house is 43.78594 percent*. This implies poor predictive accuracy.

If the Size of the house id 2200 ft we can estimate the selling price of the house using the above equation.

**Price = 60.85146 + 0.07182 (2200)**

**Price = 218.85546 (figure in $ 1000)**

2 B)

**SAS CODE:**

TITLE "V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.REtemp;

SET V506.RealEstate;

**RUN**;

**PROC** **REG** ;

MODEL PRICE = DISTANCE;

**RUN**;

**OUTPUT:**

|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Price Price**

|  |  |
| --- | --- |
| **Number of Observations Read** | 105 |
| **Number of Observations Used** | 105 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 27788 | 27788 | 14.10 | 0.0003 |
| **Error** | 103 | 203025 | 1971.11378 |  |  |
| **Corrected Total** | 104 | 230812 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 44.39723 | **R-Square** | 0.1204 |
| **Dependent Mean** | 221.10056 | **Adj R-Sq** | 0.1119 |
| **Coeff Var** | 20.08011 |  |  |

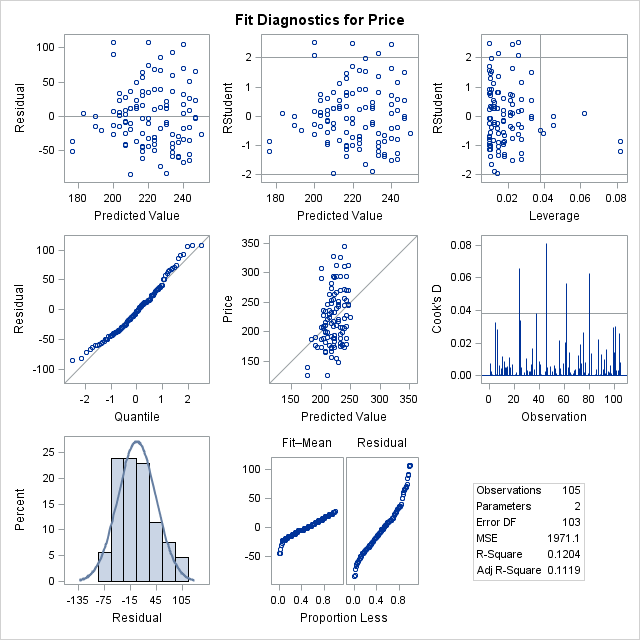
| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | 1 | 270.16123 | 13.76625 | 19.62 | <.0001 |
| **Distance** | Distance | 1 | -3.35376 | 0.89323 | -3.75 | 0.0003 |

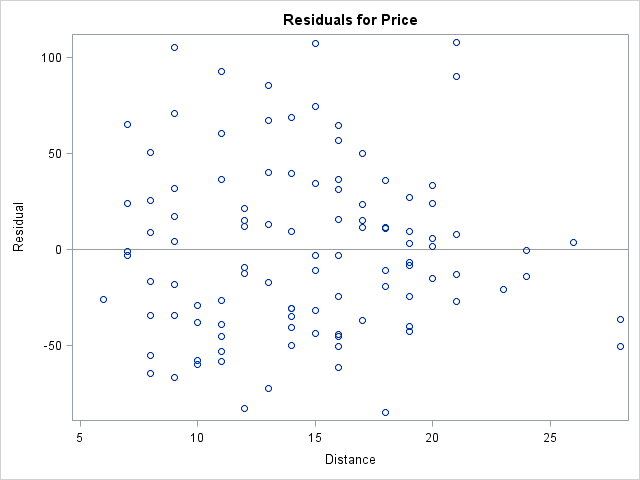
|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

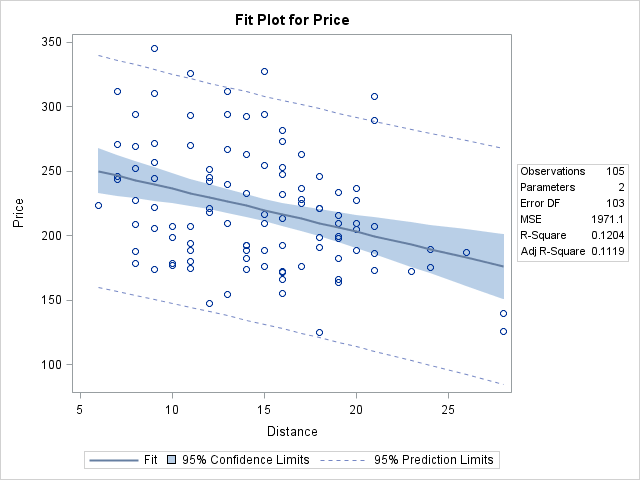
**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Price Price**







**INTERPRETATION:**

The **F statistic** = 14.10 is significant (p-value) at the .0001 level, indicating that we can reject the null hypothesis that H0: ρ = 0. We can conclude that *the data are consistent with some type of relationship between the dependent and independent variables*.

The **R2 value** of 0.1204 indicates that *12.04 percent of the variation in the selling price of the house is explained by or associated with distance from the center of the city*. This is a weak relationship.

The **regression equation** would be **Price = 270.16123 + (-3.35376) Distance + u-hat**

This implies that *the selling price of the house would decrease by 3.35376 units given a one-unit increase in distance from the center of the city*. Note that since this variable is measured in thousands, a one-unit increase is actually equivalent to a $1,000 increase *total team salary*. It is critical to interpret the results in terms of the units of measurement.The intercept (constant) value of 270.16123 implies that *if distance from the center of the city to zero, the selling price of the house would be 270.16123 percent.* This is true for house located exactly at the center of the city, which is practically not feasible and is an estimated location.

The **t-statistic** of -3.75 and **p-value** of .0003 **for Size** indicates that *this independent variable has a statistically significant relationship with Price*; i.e., we can reject the null hypothesis H0: β2 = 0 at the .0001 level of significance. Thus, we can conclude that *there appears to be a significant relationship between distance from the center of the city and selling price of the house*.

The **root mse** of 44.39723 represents the **standard error of the regression**, and is *the average error made in predicting the dependent variable using the regression equation* (i.e., the average difference between y-hat and y). The smaller this is as a percentage of the **dependent variable mean**, the better the prediction. That percentage is provided as the **coefficient of variation**, and indicates that *the average error in predicting selling price of the house is 44.39723 percent*. This implies poor predictive accuracy.

If the distance from the center of the city is 20 miles, we can estimate the selling price of the house using the above equation.

**Price = 270.16123 - 3.35376 (20)**

**Price = 203.08603 (figure in $1000)**

2 C)

Yes, we can conclude that the independent variables “*distance from the center of the city*” and “selling price” are negatively correlated; and “*size of the house*” and “selling price” are positively correlated.

Equation:

**Price = 270.16123 - 3.35376 Distance + u-hat**

**Price = 60.85146 + 0.07182 Size + u-hat**

3 )

**SAS CODE:**

TITLE "V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU";

**DATA** V506.BBtemp;

SET V506.Baseball;

**RUN**;

**PROC** **REG** ;

MODEL ATTENDANCE = SALARY;

**RUN**;

**PROC** **REG** ;

MODEL ATTENDANCE = BA;

**RUN**;

**PROC** **REG** ;

MODEL ATTENDANCE = ERA;

**RUN**;

**OUTPUT:**

|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Attendance Attendance**

|  |  |
| --- | --- |
| **Number of Observations Read** | 30 |
| **Number of Observations Used** | 30 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 7.70274 | 7.70274 | 33.50 | <.0001 |
| **Error** | 28 | 6.43860 | 0.22995 |  |  |
| **Corrected Total** | 29 | 14.14134 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.47953 | **R-Square** | 0.5447 |
| **Dependent Mean** | 2.44767 | **Adj R-Sq** | 0.5284 |
| **Coeff Var** | 19.59135 |  |  |

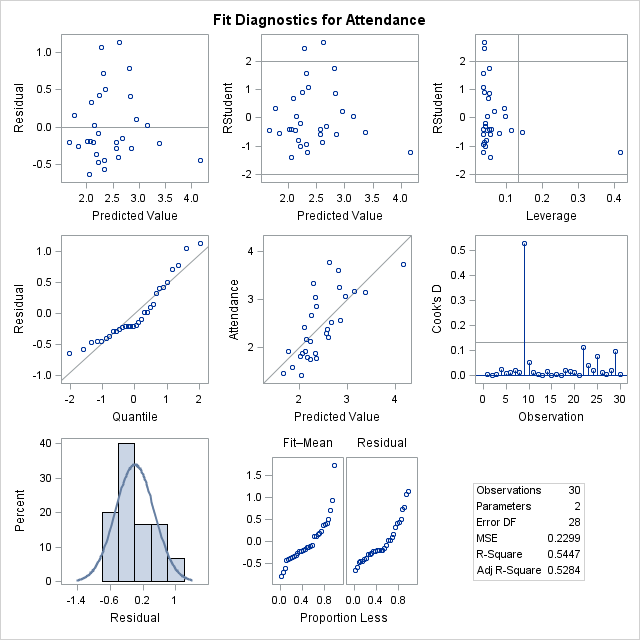
| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | 1 | 1.10206 | 0.24843 | 4.44 | 0.0001 |
| **Salary** | Salary | 1 | 0.01520 | 0.00263 | 5.79 | <.0001 |

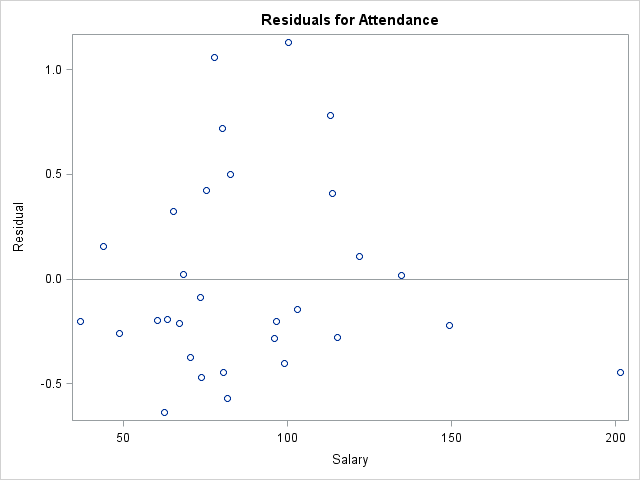
|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

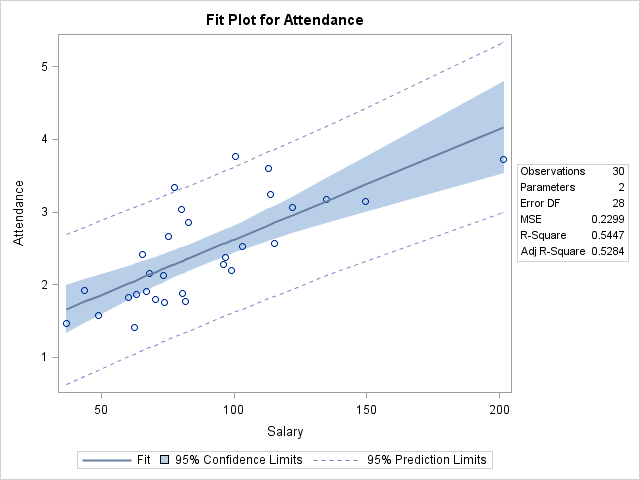
**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Attendance Attendance**







|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Attendance Attendance**

|  |  |
| --- | --- |
| **Number of Observations Read** | 30 |
| **Number of Observations Used** | 30 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 2.74613 | 2.74613 | 6.75 | 0.0148 |
| **Error** | 28 | 11.39521 | 0.40697 |  |  |
| **Corrected Total** | 29 | 14.14134 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.63794 | **R-Square** | 0.1942 |
| **Dependent Mean** | 2.44767 | **Adj R-Sq** | 0.1654 |
| **Coeff Var** | 26.06333 |  |  |

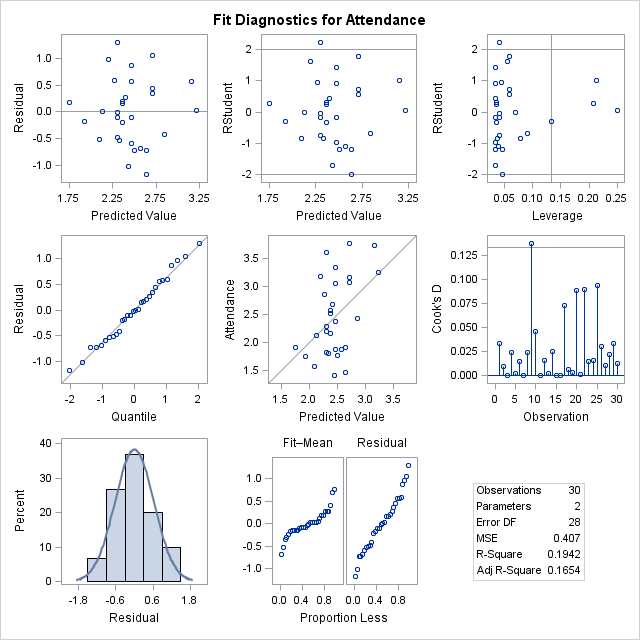
| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | 1 | -6.49215 | 3.44349 | -1.89 | 0.0698 |
| **BA** | BA | 1 | 34.07809 | 13.11889 | 2.60 | 0.0148 |

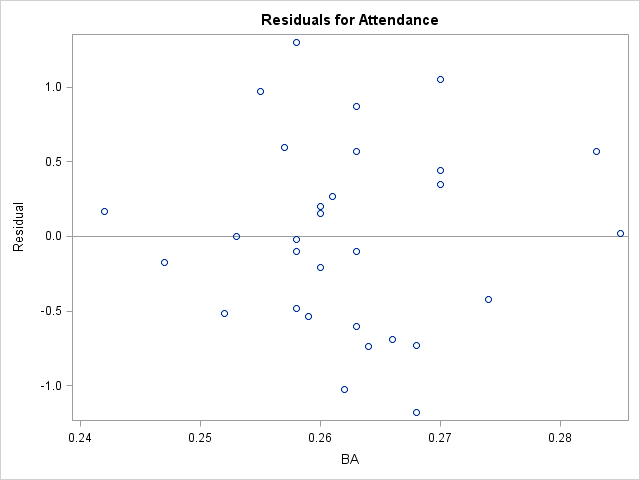
|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

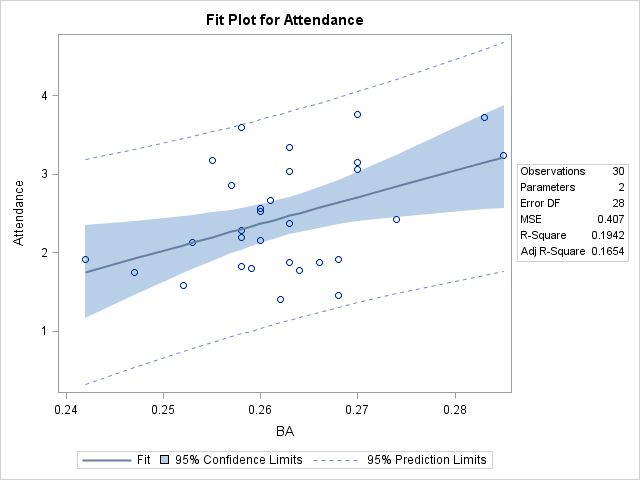
**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Attendance Attendance**







|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Attendance Attendance**

|  |  |
| --- | --- |
| **Number of Observations Read** | 30 |
| **Number of Observations Used** | 30 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 2.85523 | 2.85523 | 7.08 | 0.0127 |
| **Error** | 28 | 11.28611 | 0.40308 |  |  |
| **Corrected Total** | 29 | 14.14134 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.63488 | **R-Square** | 0.2019 |
| **Dependent Mean** | 2.44767 | **Adj R-Sq** | 0.1734 |
| **Coeff Var** | 25.93826 |  |  |

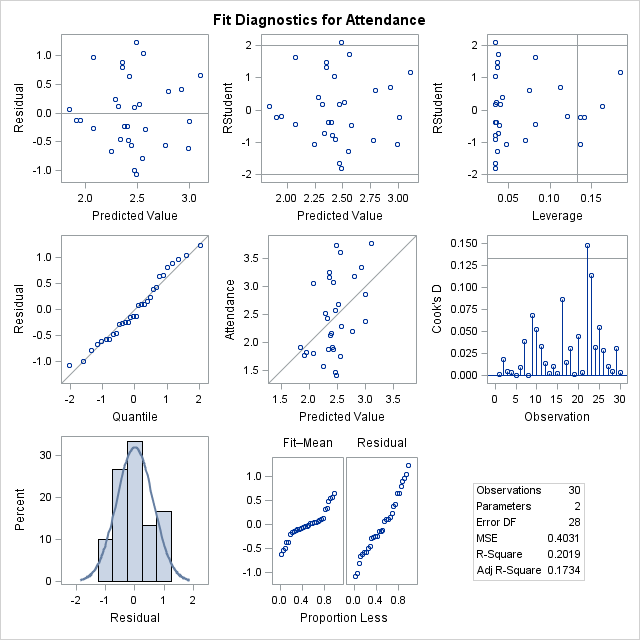
| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | 1 | 5.58641 | 1.18500 | 4.71 | <.0001 |
| **ERA** | ERA | 1 | -0.72757 | 0.27337 | -2.66 | 0.0127 |

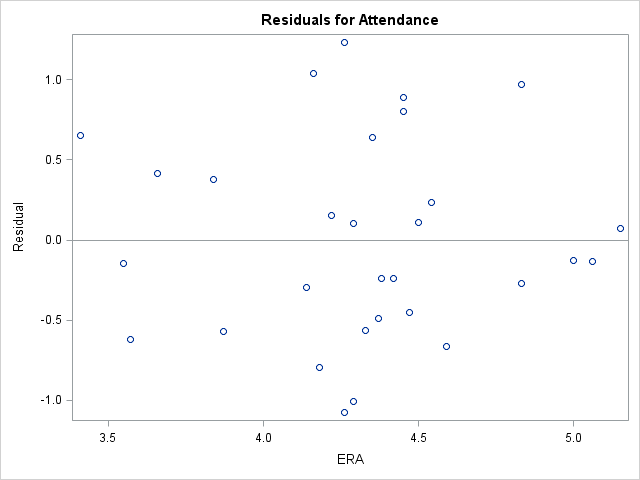
|  |
| --- |
| **V506 HOMEWORK06 PART 2 - JIVITESH POOJARY AND QIWEN ZHU** |

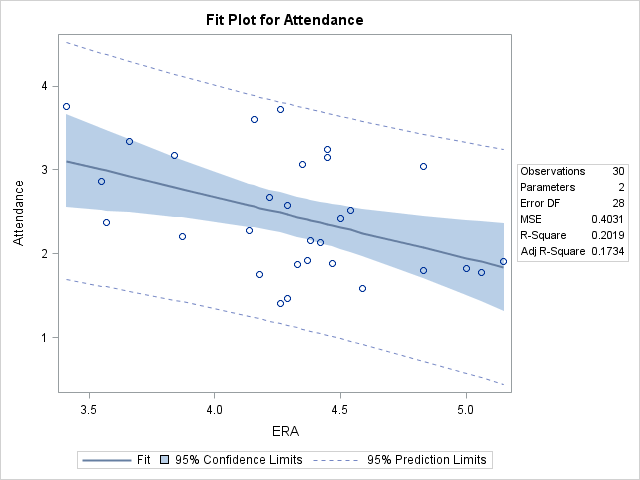
**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: Attendance Attendance**







**INTERPRETATION:**

The **F statistic** = 33.50 is significant (p-value) at the .0001 level, indicating that we can reject the null hypothesis that H0: ρ = 0. We can conclude that *the data are consistent with some type of relationship between the dependent and independent variables*.

The **R2 value** of 0.5447 indicates that *54.47 percent of the variation in the attendance is explained by or associated with total team salary*. This is a moderate relationship.

The **regression equation** would be **Attendance = 1.10206 + 0.01520 Salary + u-hat**

This implies that *the attendance would increase by* ***0.01520*** *(total for team in millions) given a one-unit increase in total team salary*. Note that since this variable is measured in millions, a one-unit increase is actually equivalent to a $1,000,000 increase *total team salary*. It is critical to interpret the results in terms of the units of measurement.The intercept (constant) value of 1.10206 implies that *if total team salary was equal to zero, the attendance would be 1.10206 percent.*

The **t-statistic** of 5.79 and **p-value** of .0001 **for Salary** indicates that *this independent variable has a statistically significant relationship with deaths per month*; i.e., we can reject the null hypothesis H0: β2 = 0 at the .0001 level of significance. Thus, we can conclude that *there appears to be a significant relationship between total team salary and attendance*.

The **root mse** of 0.47953 represents the **standard error of the regression**, and is *the average error made in predicting the dependent variable using the regression equation* (i.e., the average difference between y-hat and y). The smaller this is as a percentage of the **dependent variable mean**, the better the prediction. That percentage is provided as the **coefficient of variation**, and indicates that *the average error in predicting attendance is 0.47953 percent*. This implies good predictive accuracy.

Expected attendance with a team salary of $80 million:

**Attendance = 1.10206 + 0.01520 Salary**

**Attendance = 1.10206 + 0.01520 (80)**

**Attendance = 2.31806 (total for team in millions)**

Expected attendance with a team salary of additional $30 million so the total becomes $110 million:

**Attendance = 1.10206 + 0.01520 Salary**

**Attendance = 1.10206 + 0.01520 (1100)**

**Attendance = 2.77406 (total for team in millions)**

Yes, from the above plot and hypothesis test we can show that the slope of the regression line is positive.

From the value of **Coeff Var** we can see that salary accounts for **19.59135 %** of the total variation in attendance.

**INTERPRETATION:**

Correlation between attendance and team batting average: 0.1942

Correlation between attendance and team ERA: 0.2019

From the above value, we can conclude that the correlation between attendance and team ERA is stronger than attendance and team batting average