

# **[Measuring Students Performance by using Preparation Test]**

---

**By : Ghadeer Alharbi**

# Abstract

Several empirical studies are conducted to explore the factors that influence student performance. This research focuses on the marks obtained by students from the United States in various subjects in high school, to understand the influence of parents’ background, test preparation etc. on students’ performance.

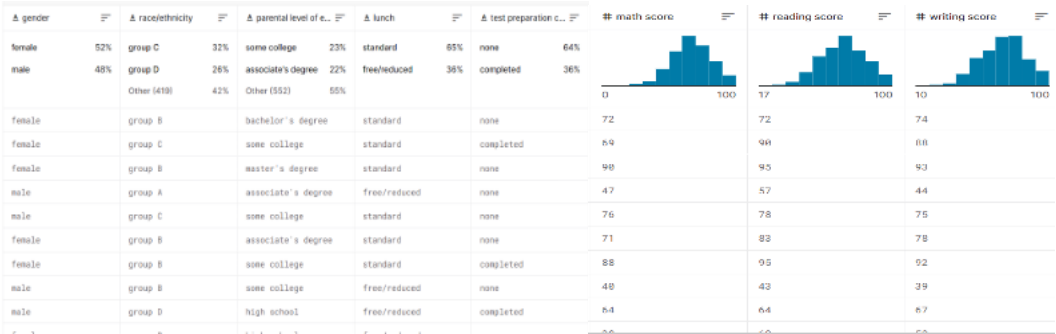
## 1. Introduction

This study was conducted to identify the changes in student’s marks in math, reading, and writing due to a certain course. The target population: All students university, the data set includes scores from three exams and a variety of personal, social, and economic factors that have interaction effects upon them. The main objective is to understand the influence of the parents’ level of education, test preparation etc. on students’ performance.

## 2. Data Source:

The data set consists of the marks secured by the students in various subjects. A sample of 1000 students was selected randomly and obtained their gender (Male/ Female), parent’s level of education, whether the students received the course or not?, math exam scores, reading exam scores, and writing exam scores. Data set is available on the [Kaggle](#) website

Figure 1: sample of students' performance dataset on Kaggle



### **3. Data preparation**

includes standardizing data formats, enriching source data, and/or removing outliers. we choose clean, easy to understand and high-quality data, that help us to processed and analyzed more quickly and efficiently leads to more timely, efficient, better-quality business decisions. The data is clean as percentage of missing values is 0%.

### **4. Goals of the study:**

The main goals of the study are to answer these questions:

- 1- Identify whether completing the course affect the students' score in math, reading and writing exam or not? (i.e.: Is there a significant difference in math, reading, and writing scores among those who completed the course and those who did not take the course or not?)
- 2- Does the score of the student in reading exam significantly affect his math score or not?
- 3- Predicting the students' math score at different reading scores.

### **5. Methodologies:**

In order to answer each of these questions each has its suitable methodology:

- 1- Independent samples t-test is used to test whether there a significant difference in math scores among those who completed the course and those who did not take the course or not?
- 2- Regression analysis is used to study whether students' score in reading exam significantly affect his math score or not?
- 3- The estimated regression line is used to predict the students' math score at different reading scores.

## 6. Statistical Analysis:

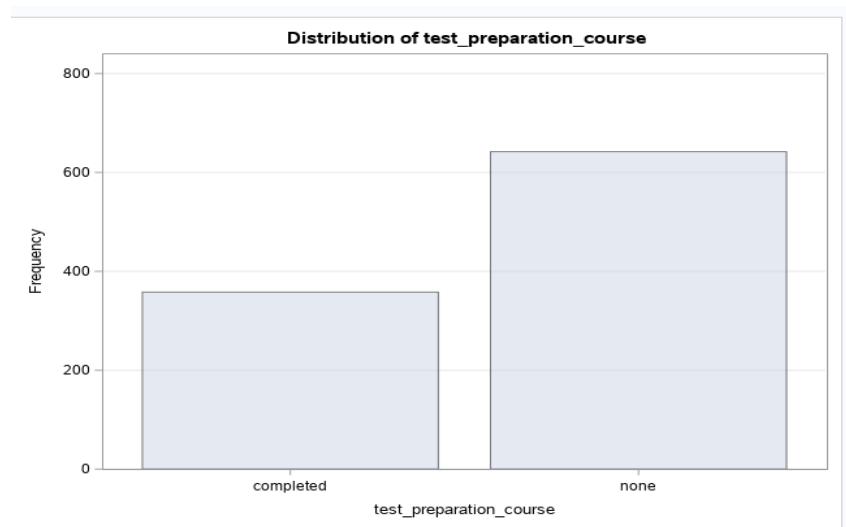
### ➤ 6.1 Sample Characteristics:

A sample of 1000 patients was drawn and the following variables were measured from the sample observations: Gender (Male/Female), the parents' level of education, test preparation etc. students' math, reading, and writing scores. For the Gender 48% of the sample individuals are Females while 52% are Males which means that percentage of males and females in the sample are close to each other. For the course preparation Table.1 and Figure.1 show that around 35.8% completed the course, while 64.2% did not take the course.

Table.1: The percentage of students completed the course

Categorical Variable Frequency Analysis				
The FREQ Procedure				
test_preparation_course	Frequency	Percent	Cumulative Frequency	Cumulative Percent
completed	358	35.80	358	35.80
none	642	64.20	1000	100.00

Figure.1: The percentage of students completed the course



Regarding to as shown from Figure.2, there exists correlation between students' scores and the normal distribution line. This means that our data is **very close to Gaussian!** . (See Table.2)

**Table.3: Descriptive for student scores in math exam**

<b>Pearson Correlation Coefficients, N = 1000</b> <b>Prob &gt;  r  under H0: Rho=0</b>			
	<b>math_score</b>	<b>reading_score</b>	<b>writing_score</b>
<b>math_score</b>	1.00000	0.81758 <.0001	0.80264 <.0001
<b>reading_score</b>	0.81758 <.0001	1.00000	0.95460 <.0001
<b>writing_score</b>	0.80264 <.0001	0.95460 <.0001	1.00000

Regarding the students' math score is normally distributed as shown from Figure.3, and the mean score is 66.09, with standard deviation 15.16 which means that there are small variations among students' scores in math exam. Classifying observations into two groups who completed the preparation course and who did not complete the course, the mean score for math exam in completed course group is 69.70 while mean math score for students did not complete the course is 64.08, then from the sample the mean score in math exam for those completed the preparation course is higher than that for those who did not take the course. (See Table.3)

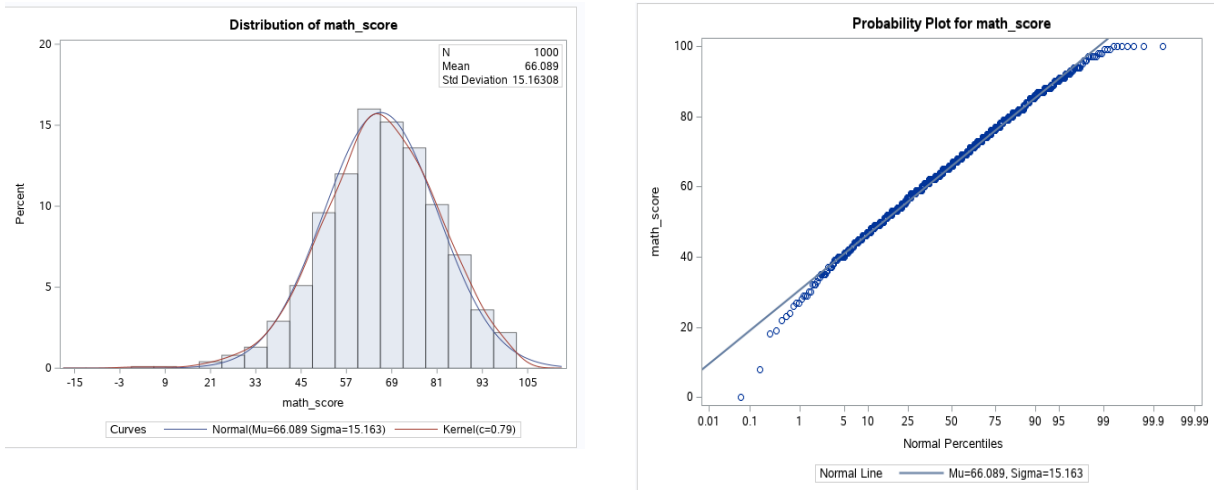
**Table.3: Descriptive for student scores in math exam**

Selected Descriptive Statistics math score							
The MEANS Procedure							
Analysis Variable : math_score							
N Obs	N	Mean	Std Dev	Lower Quartile	Upper Quartile	Quartile Range	
1000	1000	66.09	15.16	57.00	77.00	20.00	

Analysis Variable : math_score							
test_preparation_course	N Obs	N	Mean	Std Dev	Lower Quartile	Upper Quartile	Quartile Rang
completed	358	358	69.70	14.44	60.00	79.00	19.0
none	642	642	64.08	15.19	54.00	75.00	21.0

**Figure.3: Distribution of math-score**



Regarding the students' reading score is normally distributed as shown from Figure.4, the mean score is 69.17, with standard deviation 14.60 which means that there are small variations among students' scores in reading exam (See Table.3), the mean score for reading exam inn completed course group is 73.89 while mean reading score for students did not complete the course is 66.53, then from the sample the mean score in reading exam for those completed the preparation course is higher than that for those who did not take the course.

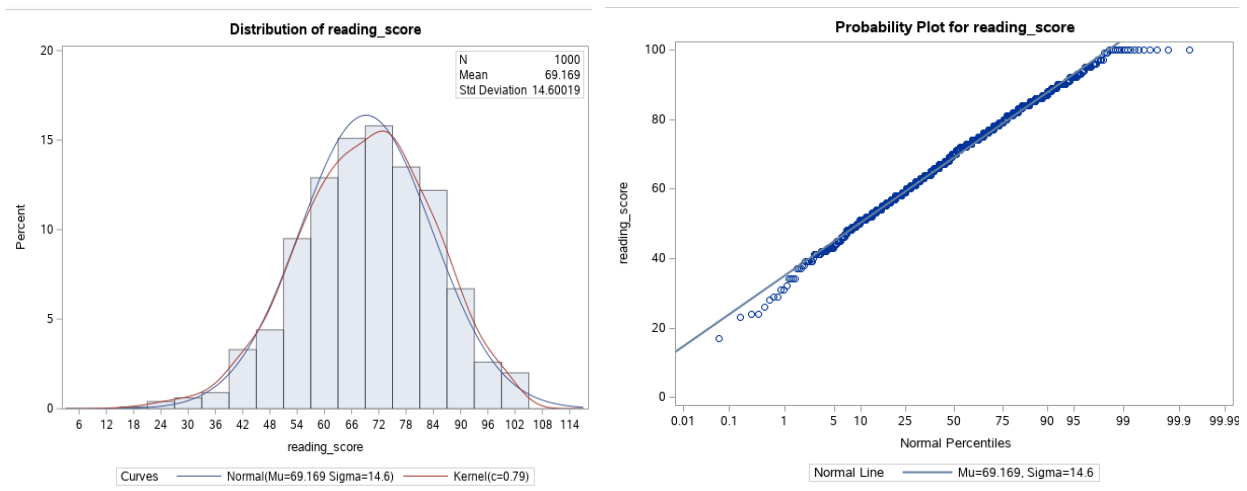
**Table.4: Descriptive for student scores in reading exam**

Selected Descriptive Statistics reading score							
The MEANS Procedure							
Analysis Variable : reading_score							
N Obs	N	Mean	Std Dev	Lower Quartile	Upper Quartile	Quartile Range	
1000	1000	69.17	14.60	59.00	79.00	20.00	

Analysis Variable : reading_score							
test_preparation_course	N Obs	N	Mean	Std Dev	Lower Quartile	Upper Quartile	Quartile Range
completed	358	358	73.89	13.64	65.00	84.00	19.00
none	642	642	66.53	14.46	57.00	78.00	19.00

**Figure.4: Distribution of reading-score**



Finally, students' writing score is normally distributed as shown from Figure.5, the mean score is 68.05, with standard deviation 15.20 which means that there are small variations among students' scores in writing exam (See Table.5), the mean score for writing exam inn completed course group is 74.42 while mean writing score for students did not complete the course is 64.50, then from the sample the mean score in writing exam for those completed the preparation course is higher than that for those who did not take the course.

Table.5: Descriptive for student scores in writing exam

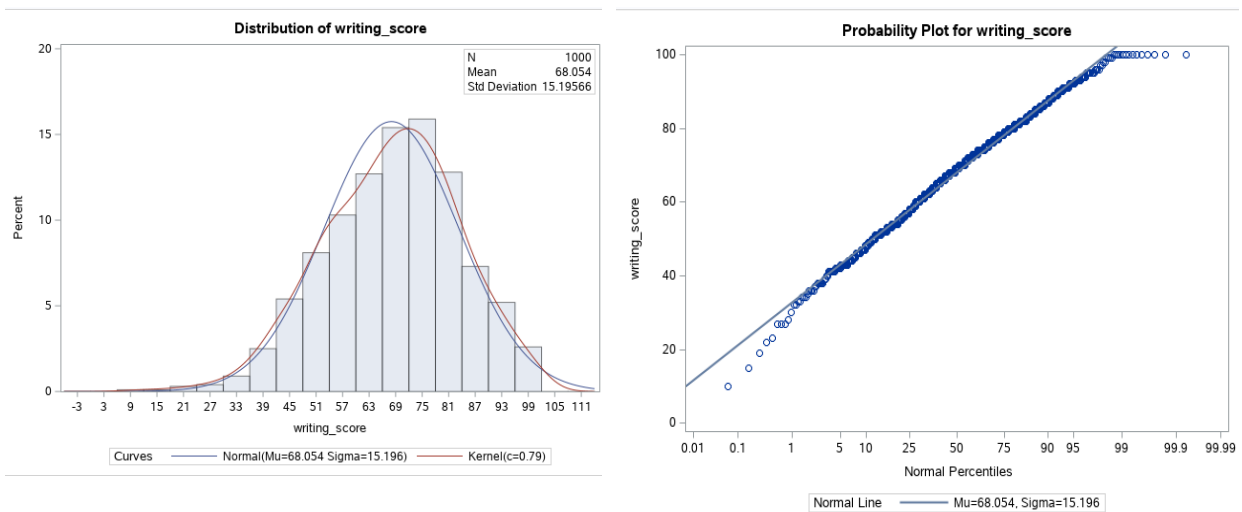
Selected Descriptive Statistics writing score

Analysis Variable : writing_score						
N Obs	N	Mean	Std Dev	Lower Quartile	Upper Quartile	Quartile Range
1000	1000	68.05	15.20	57.50	79.00	21.50

Analysis Variable : writing_score							
test_preparation_course	N Obs	N	Mean	Std Dev	Lower Quartile	Upper Quartile	Quartile Range
completed	358	358	74.42	13.38	66.00	83.00	17.00
none	642	642	64.50	15.00	54.00	74.00	20.00

Figure.5: Distribution of writing-score



## 7. Results &Analysis

### ➤ 7.1 Independent Samples t-test:

Here, the independent samples t-test is used to test whether there a significant difference in math scores among those who completed the course and those who did not take the course or not?



### The Hypothesis: Math score

**H0:** Mean math-score completed preparation course = Mean math-score did not take preparation course

**H1:** Mean math-score completed preparation course  $\neq$  Mean math-score did not take preparation course

### Test statistics and p-Value:

First the p-value for the test of equality of variances of the two groups is equal to

$0.2864 > 0.05$ , then we do not reject that the two groups (completed and none) have equal variances, So we depend on the results of the test based on the equal variances for the two groups.

For the equal variances test, test statistics  $t = 5.7$  with  $p\text{-value} < 0.001$ , so we reject the null hypothesis that the mean score for math exam is equal for the two groups.

### Decision:

Then it could be decided that there is a significant difference in the mean score of math exam between the two groups (completed and none).

Results are shown in Table.6

**Table.6: Test of hypothesis for the two groups in math exam**

T-Test(two sides ) for math score , test preparation course							
Variable: math_score							
test_preparation_course	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
completed		358	89.6955	14.4447	0.7834	23.0000	100.0
none		642	84.0779	15.1924	0.5998	0	100.0
Diff (1-2)	Pooled		5.6176	14.9292	0.9848		
Diff (1-2)	Satterthwaite		5.6176		0.9707		

test_preparation_course	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
completed		89.6955	88.1942 71.1989	14.4447	13.4585 15.5881
none		84.0779	82.9005 65.2553	15.1924	14.4044 16.0723
Diff (1-2)	Pooled	5.6176	3.6852 7.5501	14.9292	14.3021 15.6143
Diff (1-2)	Satterthwaite	5.6176	3.7120 7.5233		

Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	998	5.70	<.0001
Satterthwaite	Unequal	770.08	5.79	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	641	357	1.11	0.2864

### The Hypothesis: Reading score

**H<sub>0</sub>:** Mean reading-score completed preparation course = Mean reading-score did not take preparation course

**H<sub>1</sub>:** Mean reading-score completed preparation course  $\neq$  Mean reading-score did not take preparation course

### Test statistics and p-Value:

First the p-value for the test of equality of variances of the two groups is equal to 0.2144 > 0.05, then we do not reject that the two groups (completed and none) have equal variances, So we depend on the results of the test based on the equal variances for the two groups.

For the equal variances test, test statistics  $t = 7.67$  with p-value < 0.001 , so we reject the null hypothesis that the mean score for reading exam is equal for the two groups.

### Decision:

Then it could be decided that there is a significant difference in the mean score of reading exam between the two groups (completed and none). Results are shown in Table.8

**Table.8 Test of hypothesis for the two groups in reading exam**

T-Test(two sides ) for reading score , test preparation course							
Variable: reading_score							
test_preparation_course	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
completed		358	73.8939	13.6384	0.7208	37.0000	100.0
none		642	66.5343	14.4639	0.5708	17.0000	100.0
Diff (1-2)	Pooled		7.3596	14.1741	0.9349		
Diff (1-2)	Satterthwaite		7.3596		0.9195		

test_preparation_course	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
completed		73.8939	72.4763 75.3114	13.6384	12.7072 14.7180
none		66.5343	65.4133 67.6552	14.4639	13.7137 15.3018
Diff (1-2)	Pooled	7.3596	5.5249 9.1943	14.1741	13.5787 14.8248
Diff (1-2)	Satterthwaite	7.3596	5.5546 9.1645		

Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	998	7.87	<.0001
Satterthwaite	Unequal	775.37	8.00	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	641	357	1.12	0.2144

### The Hypothesis: Writing score

**H0:** Mean writing-score completed preparation course= Mean writing-score did not take preparation course

**H1:** Mean writing-score completed preparation course  $\neq$  Mean writing-score did not take preparation course

### Test statistics and p-Value:

First the p-value for the test of equality of variances of the two groups is equal to  $0.0157 < 0.05$ , then we reject that the two groups (completed and none) have equal variances, So we depend on the results of the test based on the unequal variances for the two groups.

For the equal variances test, test statistics  $t = 10.75$  with  $p\text{-value} < 0.001$ , so we reject the null hypothesis that the mean score for writing exam is equal for the two groups.

### Decision:

Then it could be decided that there is a significant difference in the mean score of writing exam between the two groups (completed and none).

**Table.7: Test of hypothesis for the two groups in reading exam**

T-Test(two sides ) for writing score , test preparation course							
Variable: writing_score							
test_preparation_course	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
completed		358	74.4190	13.3753	0.7069	36.0000	100.0
none		642	64.5047	14.9997	0.5920	10.0000	100.0
Diff (1-2)	Pooled		9.9143	14.4396	0.9525		
Diff (1-2)	Satterthwaite		9.9143		0.9220		

test_preparation_course	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
completed		74.4190	73.0288 75.8092	13.3753	12.4621 14.4341
none		64.5047	63.3422 65.6671	14.9997	14.2217 15.8684
Diff (1-2)	Pooled	9.9143	8.0453 11.7834	14.4396	13.8330 15.1022
Diff (1-2)	Satterthwaite	9.9143	8.1044 11.7242		

Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	998	10.41	<.0001
Satterthwaite	Unequal	811.13	10.75	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	641	357	1.26	0.0157

### ➤ 7.2 Regression Analysis:

Regression analysis is used to study the effect of students' score in reading exam on his math score.

Before running the regression analysis, we have a look at the correlations matrix among the three variables, math, reading, and writing score.

**Table.8: Correlation matrix**

Pearson Correlation Coefficients, N = 1000 Prob >  r  under H0: Rho=0			
	math_score	reading_score	writing_score
math_score	1.00000	0.81758 <.0001	0.80264 <.0001
reading_score	0.81758 <.0001	1.00000	0.95480 <.0001
writing_score	0.80264 <.0001	0.95480 <.0001	1.00000

P-value for the correlation between the math score and reading score is less than 0.0001, which means that we reject H0: correlation-coefficient=0, then there will be a significant correlation between reading and math score.

**The proposed regression line:**

$$\text{Mathscore} = B0 + B1 (\text{readingscore})$$

We consider the reading score as our independent variable (x) and math score as the dependent variable(y).

**Results of the regression:**

Table.9 shows the ANOVA results for studying the significance of the effect of reading score on math score, the test statistics F=29.86 with p-value <.0001, then we reject H0: there is no significant effect for the reading score on math score, then it could be concluded that there is a relation between reading and math score. Also R-square for the model is 0.6958 which is a large value indicating goodness of fit of the regression line.

**Table.9: ANOVA results for the regression model**

Least Squares Model (No Selection)					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	71	159764	2250.20283	29.86	<.0001
Error	928	69925	75.34987		
Corrected Total	999	229689			

Root MSE	8.68043
Dependent Mean	66.08900
R-Square	0.6956
Adj R-Sq	0.6723
AIC	5393.41864
AICC	5405.08602
SBC	4744.77702

**The estimated regression line:**

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	7.357588119	1.338180781	5.498201906	4.87332E-08
X Variable	0.849100202	0.01892984	44.85511738	1.7878E-241

$$\hat{y} = 7.35 + 0.849x$$

Where y is the students' math score, and x is the students' reading score.

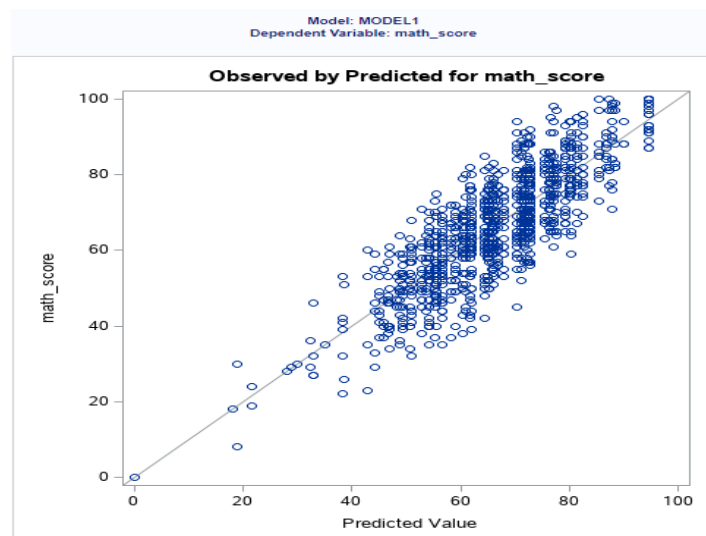
**Adequacy of the regression mode:**

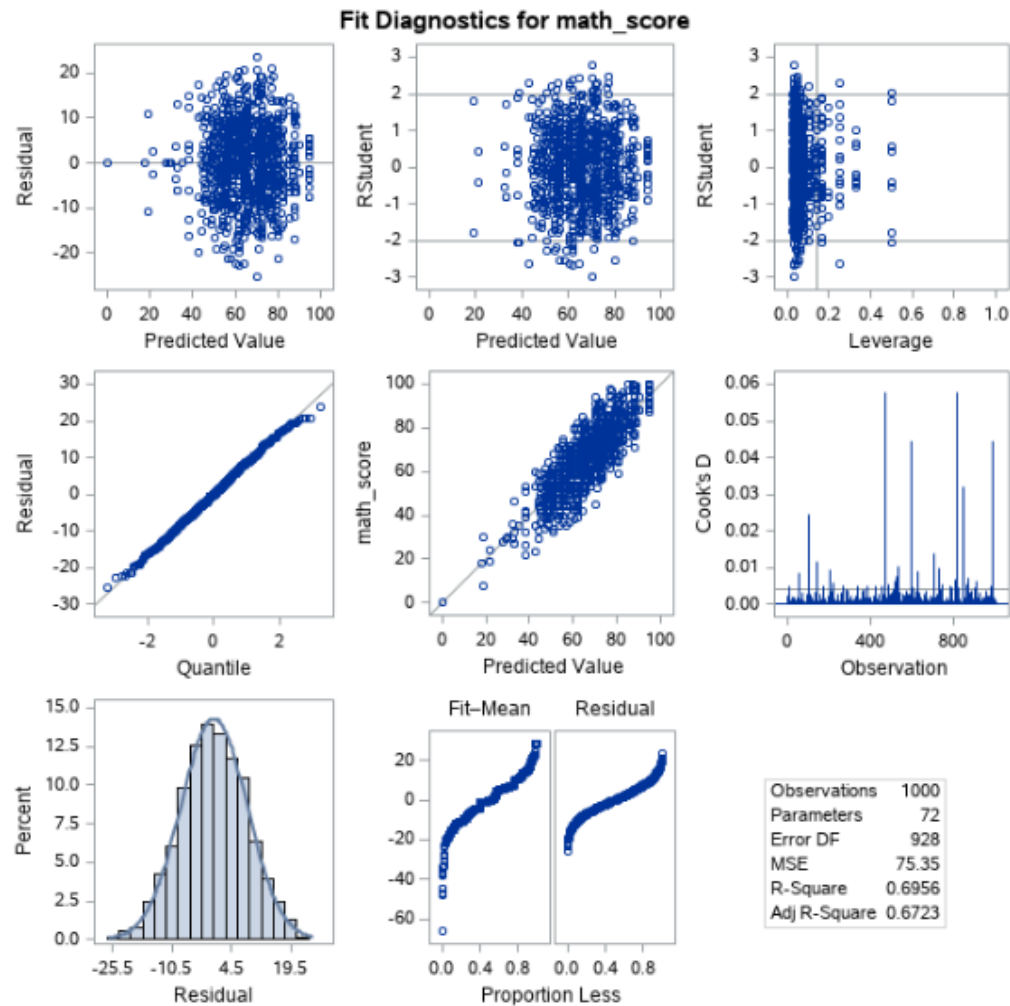
From Figure.6 we could detect the adequacy of the model, the observed values for the math score (y) are very close to the fitted regression line. Also, residual analysis shows that the model is adequate for describing the relation between reading and math score.

### **Prediction:**

Depending on the estimated regression line, we could predict students' math score based on his reading score, For example, for a student whose reading score is 50, and then his math score will be approximately 49.81.

**Figure.6: Regression model adequacy**





## Conclusions:

- There are a significant difference in math, reading, and writing scores among those who completed the course and those who did not take the course. This means that the preparation course is effective on the students' scores in the exams.
- It could be concluded that there is a relation between reading and math score, and it is approximately an increasing relation which means that when the reading score of the student increases, then his math score increases.



**References:**

Dancey, C., Reidy, J., and Rowe, R. (2012, April 4). Statistics for the Health Sciences: A Non-Mathematical Introduction (1st ed.). SAGE Publications Ltd.

Ebrahim, G. J. (2009). Medical Statistics from Scratch, David Bowers.