Summary

Science Score had a mean of 51.85 (*SD* = 9.90). Math Score had a mean of 52.65 (*SD* = 9.37). Read Score had a mean of 52.23 (*SD* = 10.25). Further, shape of histograms together box-plots suggested an approximate normal distribution for all these variables.

Among 200, 84% of schools were of Public type and 16% of schools were of Private type. In Types of Programs, 22.5% had general program, 52.50% had academic programs and 25% had vocational programs. There was a significant association prevailing between Types of programs and Types of schools.

Math and reading scores had highest coefficient of correlation in the tune of 0.662.

The prediction equation thrown by regression analyis was , *Science score = 16.76 + 0.667\*Math score*. Overall the model was found fit as the *F*-statistics was significant with a *p*-value as 0.000. Math score was able to explain approximately 40% of total variance in science score.

Social Studies scores had mean 52.405 and standard deviation as 10.736. There were 109 females and 91 males out of total of 200 cases. 147 cases had shown writing score less than 60, whereas 53 cases had scored equal to more than 60. Low, middle and high categories of ses were 47, 95 and 58 respectively.

The prediction equation for predicting Science scores thrown by multiple linear regression was: *Science score = 12.325 + 0.389\*math score + 0.05\*social studies score + 0.335\*reading score -2.010\*female*. Predictors were able to explain approximately 49% of total variance in science score. There was no linear relationship between female and science score found through t-test. Confidence Interval for regression coefficient of female were found between -4.0202 and 0.0002 at 95% level.

Logistic regression Model was found fit as Hosmer Lameshow Chi-square statistics was found non-significant. The Logistic regression equation was found as: *Honcomp = -10.201 + 0.098\*read + 0.066\*science + 0.110\*ses.* Overall classification accuracy was found as 78.51% with 91.80% for group-0 and 41.50% for group-1.

Other than attsc4, first 5 items measuring attitude towards schools were found in second factor. Factor analysis had thrown twelve factors. However, based on scree plot, it is recommended that we should consider only three factors. Factor loading tells us the relationship between individual items with their corresponding factors. When only two factors were extracted, all ten items measuring attitude towards school belonged to second factor. Internal consistency of variables measuring attitude towards school was found ok.

Conclusions:

1. Science, Math, Read, Write, Social studies were found approximately normal.
2. Significant association between Types of programs and Types of schools was found through Chi-square test.
3. The coefficients of correlation between science, math and reading score were found between 0.630 and 0.661.
4. Scatter plot showed a positive relationship between science and math scores.
5. The simple linear regression for predicting Science score with math score as independent variable had R-square as 39.80%. The model was found fit.
6. The multiple linear regression for predicting Science score with math score, social studies score, reading score and female as independent variables had R-square as 48.90%. the model was found fit. Linear relationship between female and science score was not confirmed by T-test.
7. Overall classification accuracy was found as 78.51% with 91.80% for group-0 and 41.50% for group-1 through Logistic regression model for predicting honcomp.
8. Based on scree plot, we should consider only three factors. The internal consistency of variables measuring attitude towards school was ok.

Recommendations

1. Science score was better predicted by Multiple regression model hence should be used for prediction.
2. Classification accuracies were comparatively very low in group-0 against group-1, hence, alternate models should be explored.
3. Analyst/researcher should use the factor analysis results along with his past experience and judgment for data reduction purpose.

**3.1.1 Summary of science, math and read scores.**

Descriptive statistics has been shown in Table 3.1.1-1: *Descriptives* and Histograms and Box plots are shown in Figure 3.1.1-1: *Histograms and Box Plots.*

Science Score had a mean of 51.85 (*SD* = 9.90). The median value was found as 53.23 which was almost same as that of mean. Hence, mean can be considered as a reasonable estimator of central tendency. Minimum and maximum values were found as 26 and 74. Skewness and kurtosis values were found within +,- 2 and shape of histogram together suggested an approximate normal distribution. Further, box-plots had not shown any outliers.

Math Score had a mean of 52.65 (*SD* = 9.37). The median value was found as 52.00 which was almost same as that of mean. Hence, mean can be considered as a reasonable estimator of central tendency. Minimum and maximum values were found as 33 and 75. Skewness and kurtosis values were found within +,- 2 and shape of histogram together suggested an approximate normal distribution. Further, box-plots had not shown any outliers.

Read Score had a mean of 52.23 (*SD* = 10.25). The median value was found as 50.00 which was almost same as that of mean. Hence, mean can be considered as a reasonable estimator of central tendency. Minimum and maximum values were found as 28 and 76. Skewness and kurtosis values were found within +,- 2 and shape of histogram together suggested an approximate normal distribution. Further, box-plots had not shown any outliers.

Table 3.1.1-1

*Descriptives*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Descriptive Statistics | | science | math | read |
| Mean | | 51.85 | 52.65 | 52.23 |
| 95% Confidence Interval for Mean | Lower Bound | 50.47 | 51.34 | 50.80 |
| Upper Bound | 53.23 | 53.95 | 53.66 |
| 5% Trimmed Mean | | 51.96 | 52.39 | 52.14 |
| Median | | 53.00 | 52.00 | 50.00 |
| Variance | | 98.03 | 87.77 | 105.12 |
| Std. Deviation | | 9.90 | 9.37 | 10.25 |
| Minimum | | 26.00 | 33.00 | 28.00 |
| Maximum | | 74.00 | 75.00 | 76.00 |
| Range | | 48.00 | 42.00 | 48.00 |
| Interquartile Range | | 14.00 | 14.00 | 16.00 |
| Skewness | | -0.19 | 0.29 | 0.20 |
| Kurtosis | | -0.56 | -0.65 | -0.62 |

Figure 3.1.1-1

*Histograms and Box Plots*

|  |  |  |
| --- | --- | --- |
| Variable↓ | Histogram | Box plot |
| Science |  |  |
| Math |  |  |
| Read |  |  |

**3.1.2 (a) Table of *type of school* and type of *program*.**

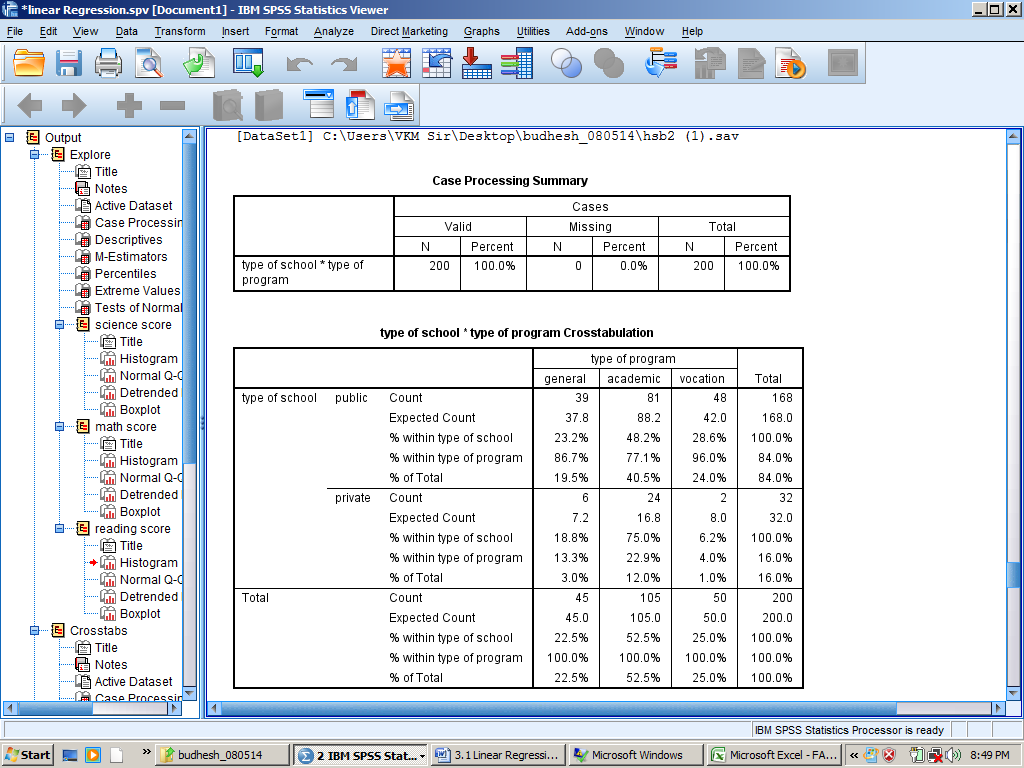
Interpretation of Row and Column percentages:

Screen shot 3.1.2-1: *Cross tabulation* shows percentages within each cells. Overall Public schools were 84% and Private schools were 16%. Under types of program, general. Academic and vocation were 22.50%, 52,50% and 25% respectively.

Within Public type of schools, general were 19.50%, academic were 40.50% and vocation were 24%. Within Private type of schools, general were 3%, academic were 12% and vocation were only 1%.

Screen shot 3.1.2-1

*Cross tabulation*



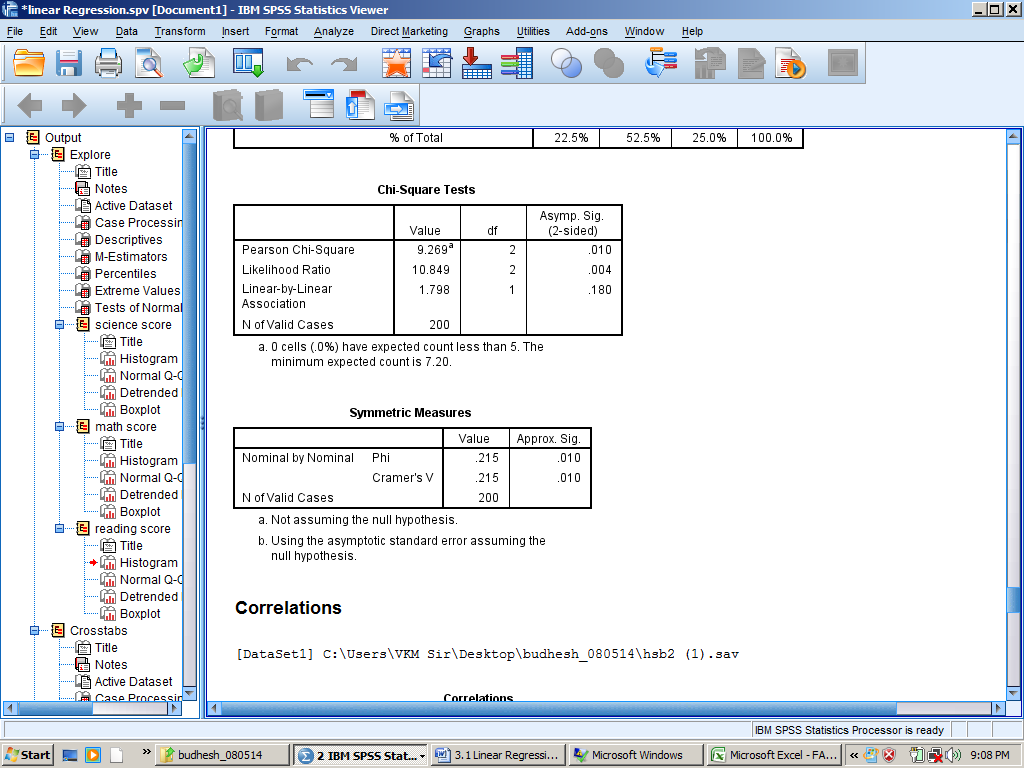
**3.1.2 (b) Association between *type of school* and type of *program*.**

Hypotheses were designed as follows:

SPSS Output is shown in Screen shot 3.1.2-1 for Chi-square test. The chi-square statistics was found as 9.269 with 2 degrees of freedom and significance value was less than the level of significance, 5%. This indicated the rejection of null hypothesis and we can conclude that a significant association between both the types was existing. Further, the strength of association, as indicated by Phi was found as 0.215.

Screen shot 3.1.2-1:

*Chi-square tests*

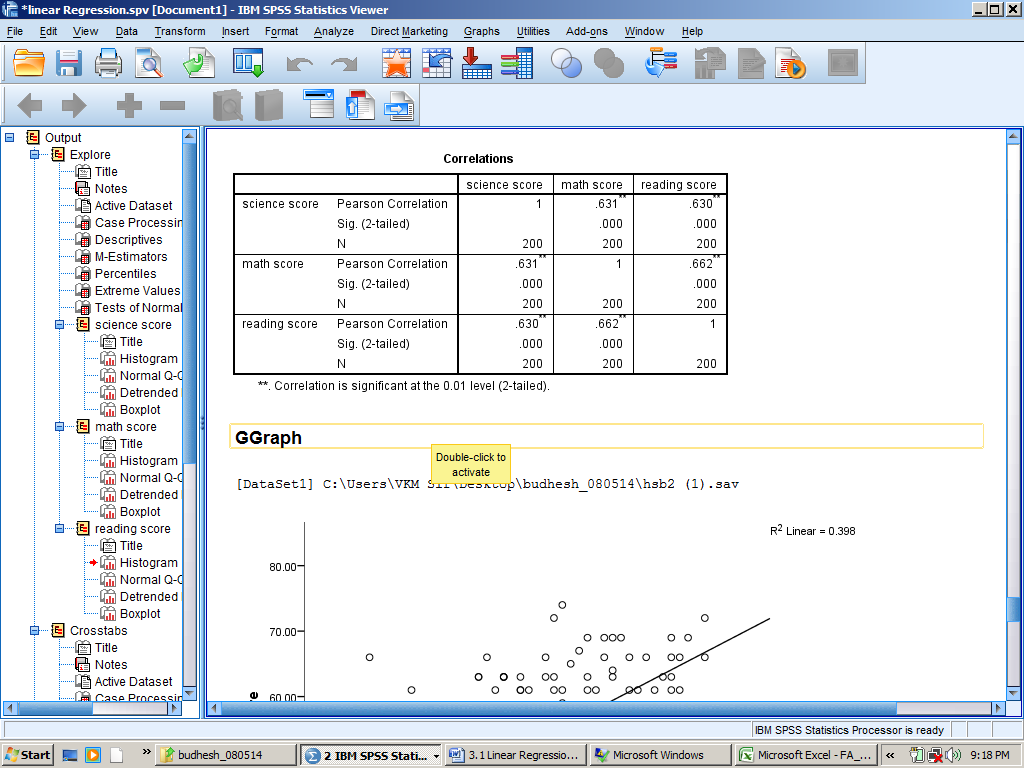


**3.1.3 Correlation between *science, math* and *read*.**

Screen shot 3.1.3: *Correlations* shows the SPSS output for bi-variate correlation analysis. The maximum correlation was found between Math and Read in the tune of 0.662 followed by Science and Math as 0.631. Almost the same correlation was found between Science and Read as 0.630.

Screen shot 3.1.3:

*Correlations*

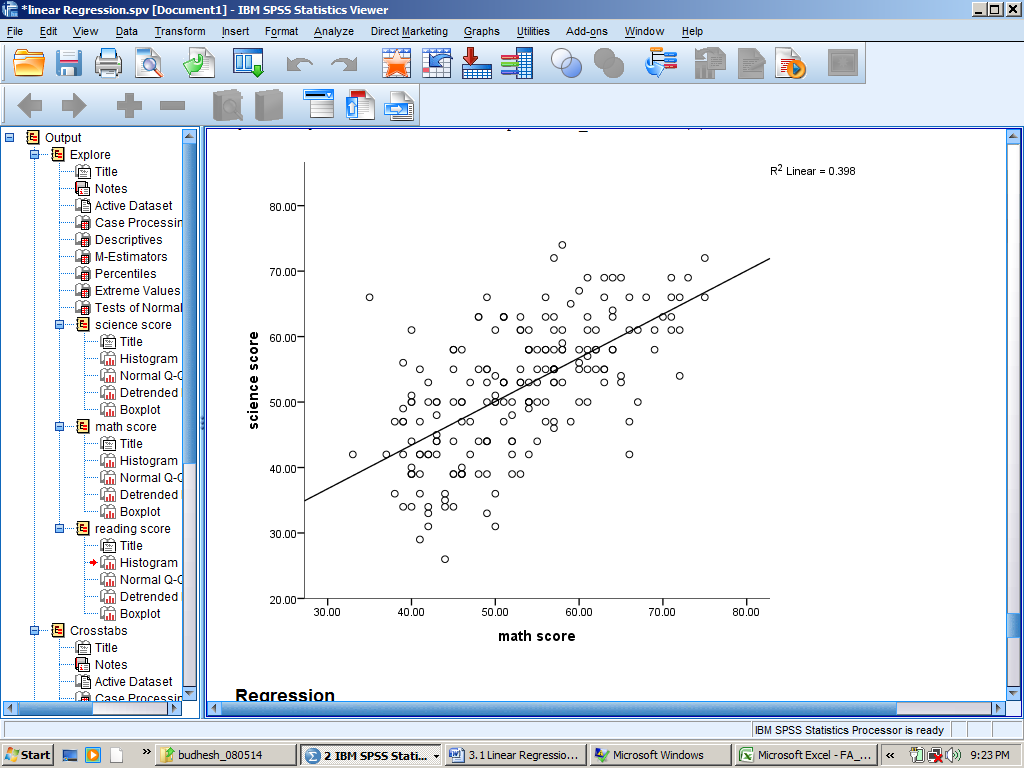


**3.1.4 Scatter plot between *science* and *math*.**

Screen shot 3.1.4: *Scatter Plot* shows the SPSS output for scatter plot between *Science* and *Math* scores. The trend line shows that there is a Positive relationship existing between two variables. Further, the coefficient of determination is 0.398 which shows that 39.80 percent of total variance in Science score was explained by Math scores.

Screen shot 3.1.4:

*Scatter Plot*



**3.1.5 Simple Linear Regression between *science* and *math*.**

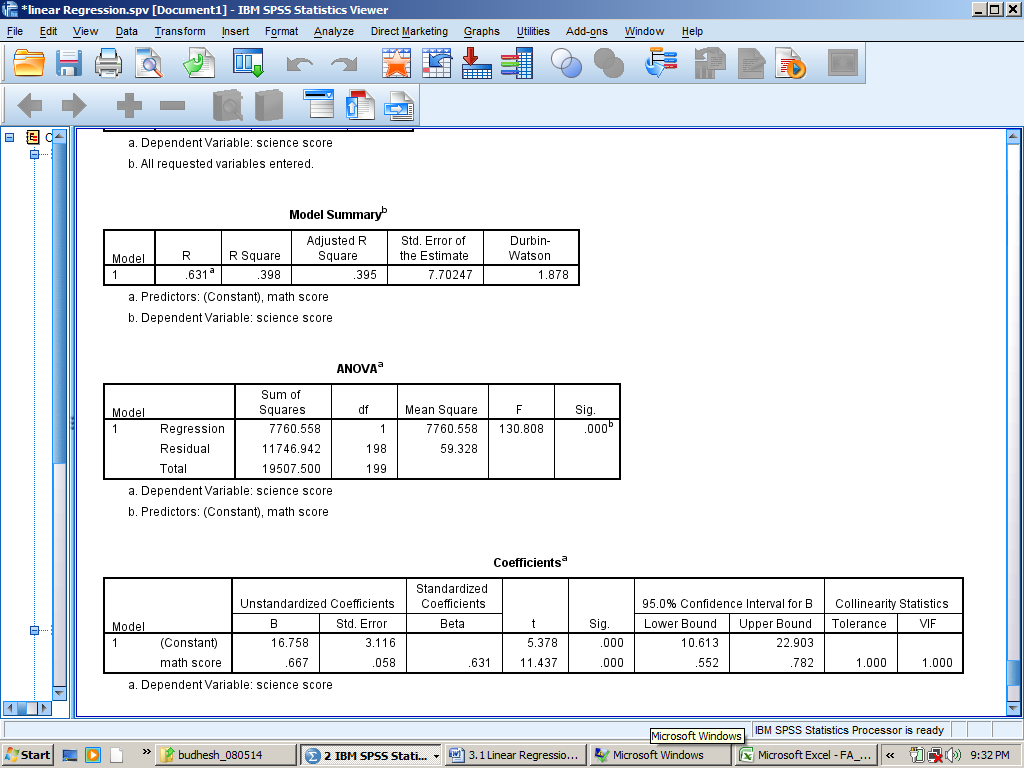
Screen shot 3.1.5: *Regression Output* shows the SPSS output for Simple Linear Regression between *Science* (Response Variable) and *Math* scores (Explanatory Variable).

The relationship between Science and Math scores were found positive as indicated by the positive sign of regression coefficient 0.667. The linearity of relationship was further supported by the significance value of regression coefficient. It was found less than 0.05. The coefficient of correlation was indicated a 0.631 which was same as found in previous analysis under section 3.1.3. R-square value was found as 0.395 which was reflected in scatter plot also.

The Durbin Watson Statistics was 1.878 which showed that there was no autocorrelation. Overall model was found good as F-statistics was significant (p-value was <.= 0.01). Standard Error of estimate was found as 7.702.

Screen shot 3.1.5:

*Regression Output*

****

3.2

**Descriptive Analysis**

As I have already presented the Descriptive Analysis for science, math and read under 3.1.1, I am presenting for Descriptive Analysis for *socst*, *female*, *honcomp* and *ses* under the following section.

**Variable: Social Studies Score (socst)**

Social Studies Score had a mean of 52.405 (*SD* = 10.736). The median value was found as 52.41 which was almost same as that of mean. Hence, mean can be considered as a reasonable estimator of central tendency. Minimum and maximum values were found as 45 and 71. Descriptive statistics is shown in Screen shot 3.2-1.

Skewness and kurtosis values were found within +,- 2 and shape of histogram together suggested an approximate normal distribution. Further, box-plots had not shown any outliers as shown in Figure 3.2-1.

Screen shot 3.2-1

*Descriptive Statistics of Social Studies Score (socst)*

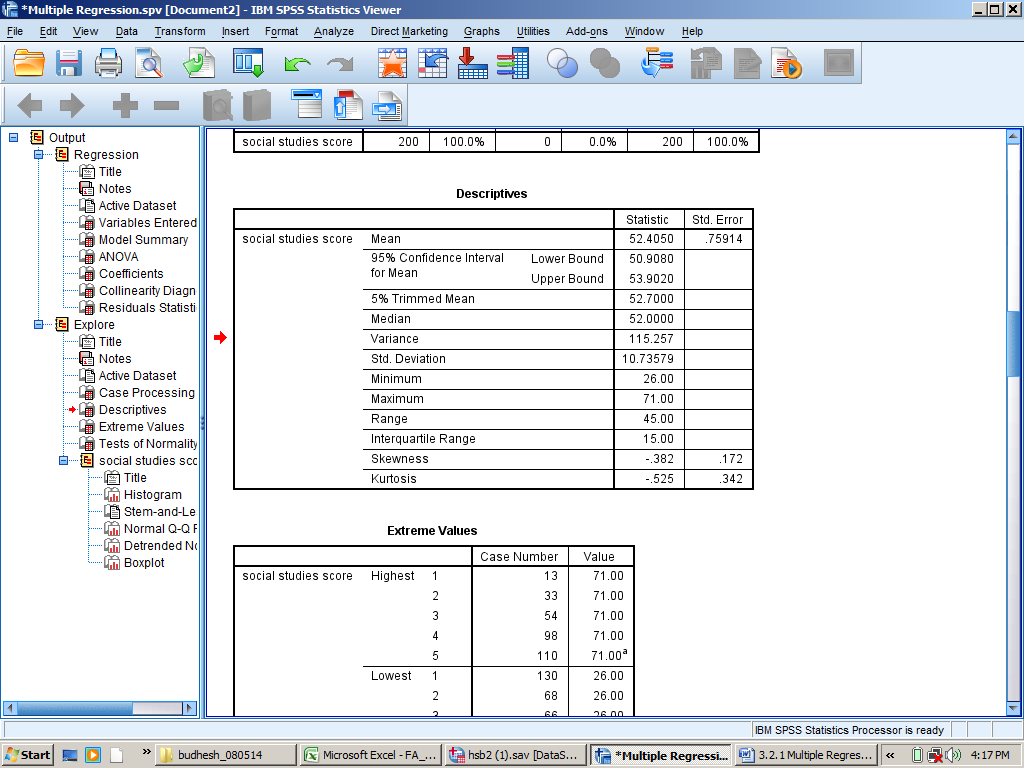


Figure 3.2-1

*Histogram & Box Plot of Social Studies Score*

|  |  |
| --- | --- |
| *Histogram* | Box Plot |
|  |  |

**Variable: Female, honcomp and ses**

The frequencies of female, honcomp and ses were shown in Bar Diagrams in Figure 3.2-2. There were 91 males and 109 female respondents. 53 respondents have scored more than or equal to 60 in writing score. Rest 147 scored less than 60 in writing score. There were 47, 95 and 58 respondents under low, middle and high categories of ses variable.

Figure 3.2-2

*Bar Diagrams of female, honcomp and ses*

|  |  |
| --- | --- |
| Bar Diagram of *female* | Bar Diagram of *honcomp* |
|  |  |
| Bar Diagram of *ses* | |
|  | |

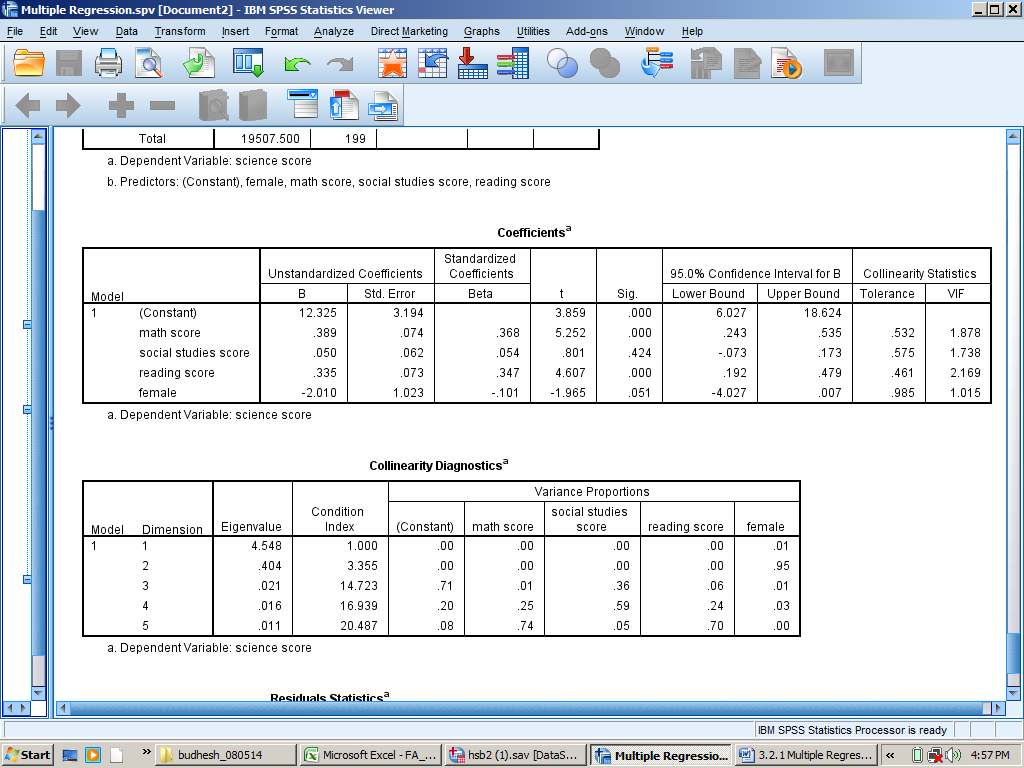
**3.2.1 (a) Regression Equation and Output**

The regression equation found is as follows:

Regression output is shown in Screen shots 3.2-2 and 3.2-3.

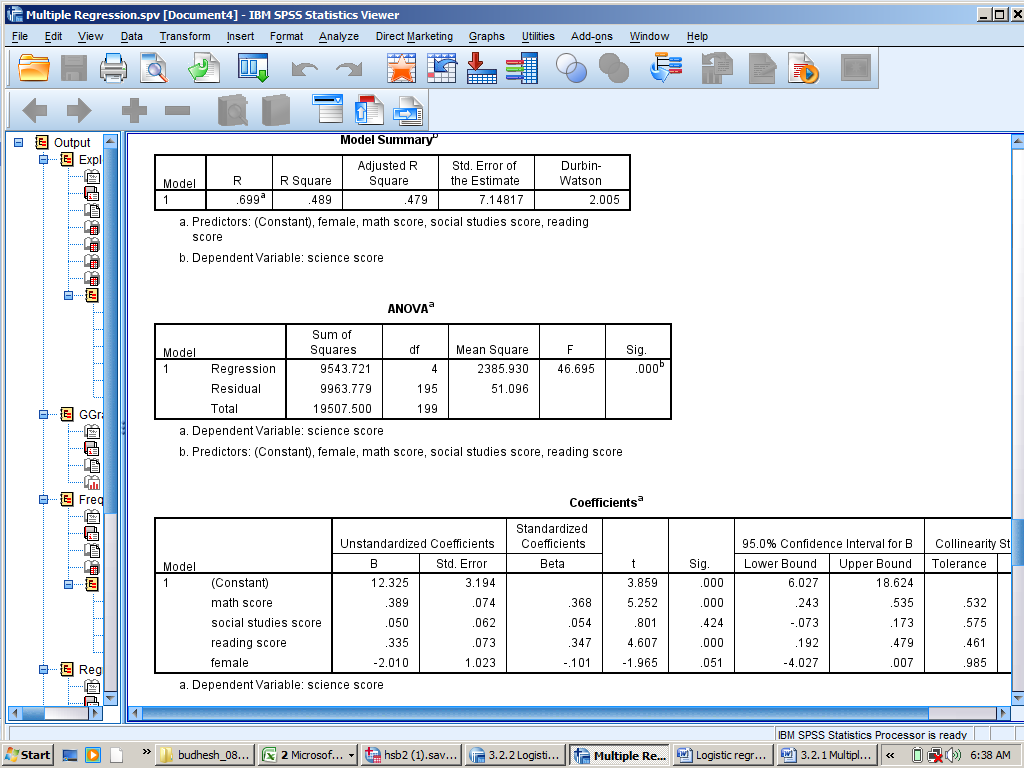
Screen shot 3.2-2

*Regression Output-Coefficients*



Screen shot 3.2-3

*Regression Output-Model Summary and ANOVA results*



**3.2.1 (b) Interpreting Regression Coefficients**

There were 4 explanatory variables in the model. Their regression coefficients can be interpreted as follows:

1. Explanatory Variable- *Math Score*: For unit increase in *math score*, keeping other explanatory variables as constant, there will be 0.389 times increase in the *science score*.
2. Explanatory Variable- *Social Studies Score*: For unit increase in *Social Studies Score*, keeping other explanatory variables as constant, there will be 0.05 times increase in the *science score*.
3. Explanatory Variable- *Reading Score*: For unit increase in *Reading Score*, keeping other explanatory variables as constant, there will be 0.335 times increase in the *science score*.
4. Explanatory Variable- *female*: For unit increase in *female*, keeping other explanatory variables as constant, there will be -2.010 times decrease in the *science score*.

**3.2.1 (c) Predicting Science score**

By using the Regression Equation:

**3.2.1 (d) Hypothesis testing for the slope of *female***

The hypotheses are as follows:

The *t*-statistics is calculated as:

,

Where is slope of the variable j with dependent variable, holding all other independent variables constant, is the standard error of the regression coefficient and is the hypothesized population slope.

The *t*-statistics follow a *t*-distribution with *n-k-1* degrees of freedom.

Plugging, as -2.010, as 1.023 and , we found *t*-statistics as -1.965.

We could not found critical value of *t* at 5% level of significance with 200-4-1=195 degrees of freedom from table; however, significance value as 0.051 shows the acceptance of null hypothesis.

Thus it has been established that female was not having a linear relationship with science score.

**3.2.1 (e) 95% confidence interval of *female***

The 95% Confidence Interval can be built as:

Hence, the upper bound will be = 0.0002 and the lower bound will be = -4.0202

Interpretation: With 95% of confidence, we can say that the regression coefficient of female will lie between -4.0202 and 0.0002.

**3.2.2 (a) Generate a new dichotomous variable ‘honcomp’ (1 for higher or equal to 60 in *write* score else 0)**

This was done first in excel with IF function as =IF(cell>=60,1,0) and then copied the result into SPSS file with a variable name *honcomp*.

**3.2.2 (b) Use logistic regression to model the effect of *read, science* and *ses* on the probability of being in honours composition (*honcomp*). Report the prediction equation and interpret the model fit.**

The SPSS Output is shown in SPSS Output 3.2.2-1 for Prediction Equation.

The Prediction Equation found was as follows:

SPSS Output 3.2.2-1

*Variables in Equation*

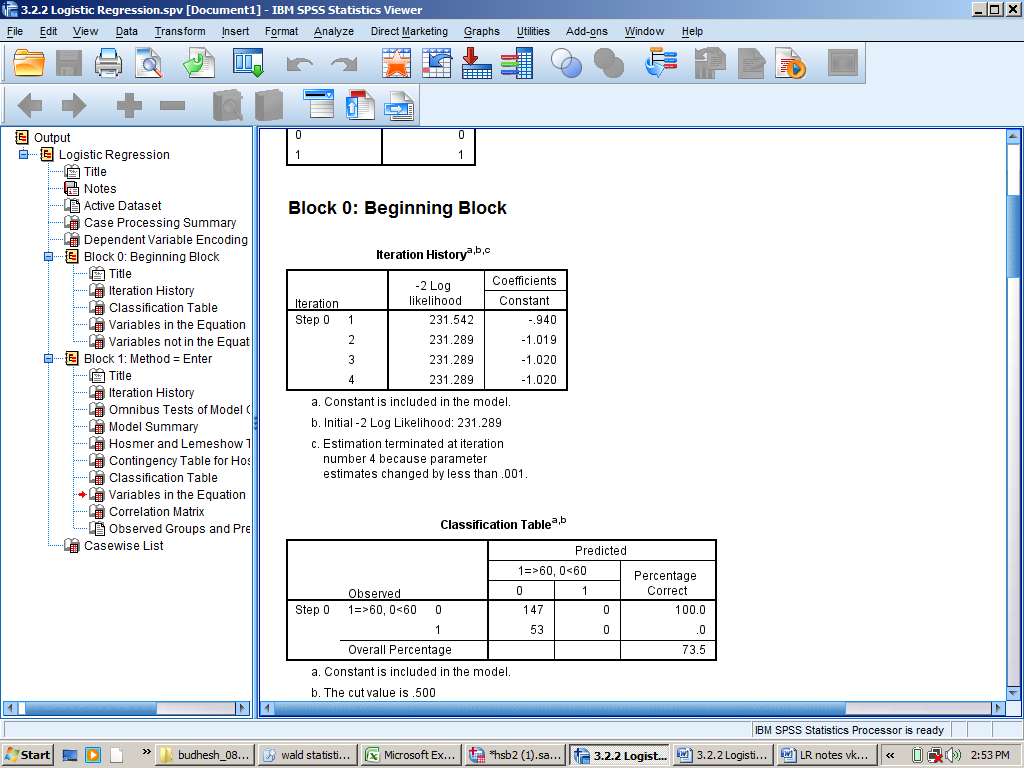
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| Lower | Upper |
| Step 1a | read | .098 | .025 | 15.567 | 1 | .000 | 1.103 | 1.050 | 1.158 |
| science | .066 | .027 | 5.805 | 1 | .016 | 1.068 | 1.012 | 1.126 |
| ses | .110 | .275 | .160 | 1 | .689 | 1.116 | .652 | 1.912 |
| Constant | -10.201 | 1.584 | 41.468 | 1 | .000 | .000 |  |  |
| a. Variable(s) entered on step 1: read, science, ses. | | | | | | | | | |

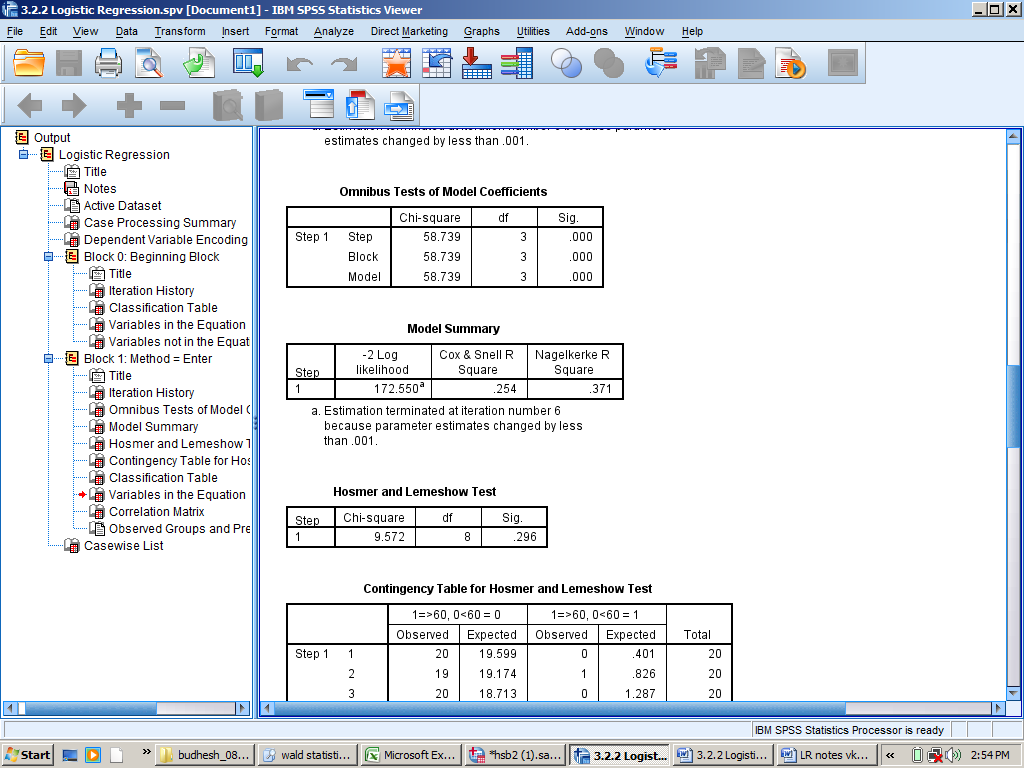
**Model Fit**

The overall model was found fit as the significance value for Hosmer Lameshow Chi-square test was found more than 0.05 at 5 percent level of significance. Further, the Nagelkerke R Square was found as 0.371and -2Log Likelihood came down to 172.550 in the final model from 231.289 of base or null model. This indicates that final model was a much improvised model for prediction.

SPSS Output: 3.2.2-2

*-2LL and Model Summary*





**3.2.2 (c) Interpret the effects of the explanatory variables on the odds of being in honours composition.**

There were 3 explanatory variables in the model for prediction of odds of being 1 or getting honors composition. Referring the **SPSS Output 3.2.2-1: *Variables in Equation****, t*hese can be interpreted as follows:

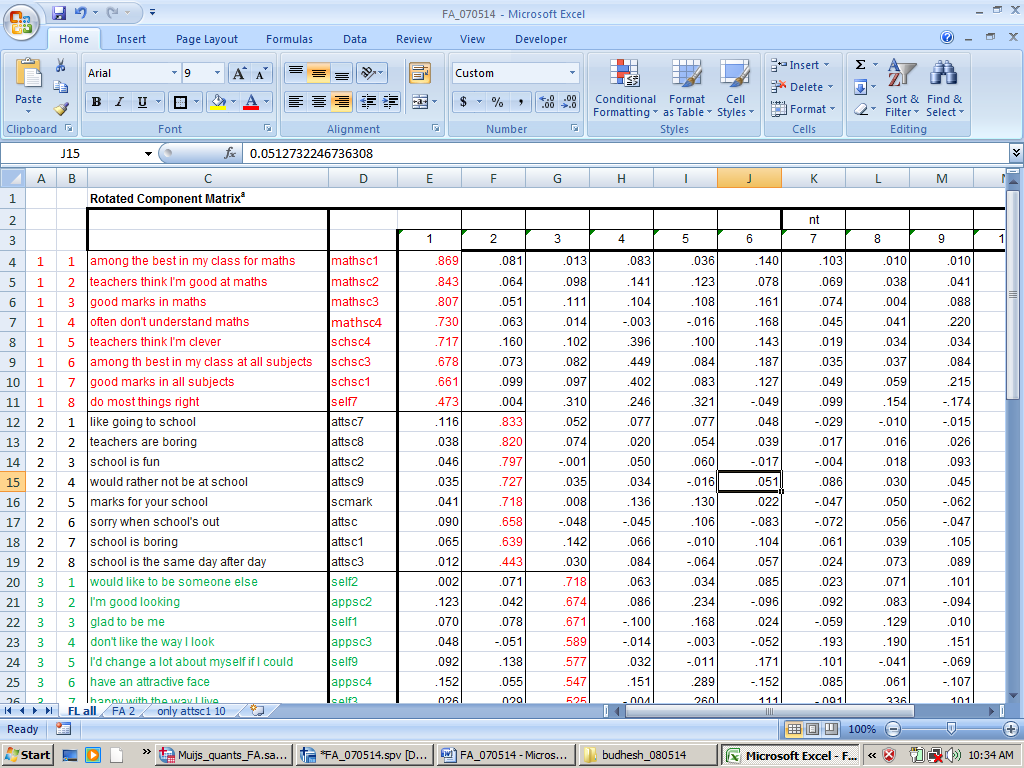
1. Explanatory Variable- *read*: The odds of getting honors composition increases by 1.103 times if *read* score is increased by 1 unit, keeping other Explanatory Variables as constant.
2. Explanatory Variable- *science*: The odds of getting honors composition increases by 1.068 times if *science* score is increased by 1 unit, keeping other Explanatory Variables as constant.
3. Explanatory Variable- *ses*: The odds of getting honors composition increases by 1.116 times if *ses* score is increased by 1 unit, keeping other Explanatory Variables as constant.

**3.3.2 (a) How many from attsc1 to attsc5 form a factor?**

When all the 54 scale and ordinal variables were applied for factor analysis, second Factor had extracted eight items/variables and other that attsc4, ‘teacher’s don’t try hard enough’, attsc1 to attsc5 were present. This is Screen shot 3.3.2-1. However, when only variables belonging to ‘attitude towards school’, attsc1 through attsc10 were applied for factor analysis, only attsc1, attsc2 and attsc3 got appeared as second factor. This is shown in Screen shot 3.3.2-2

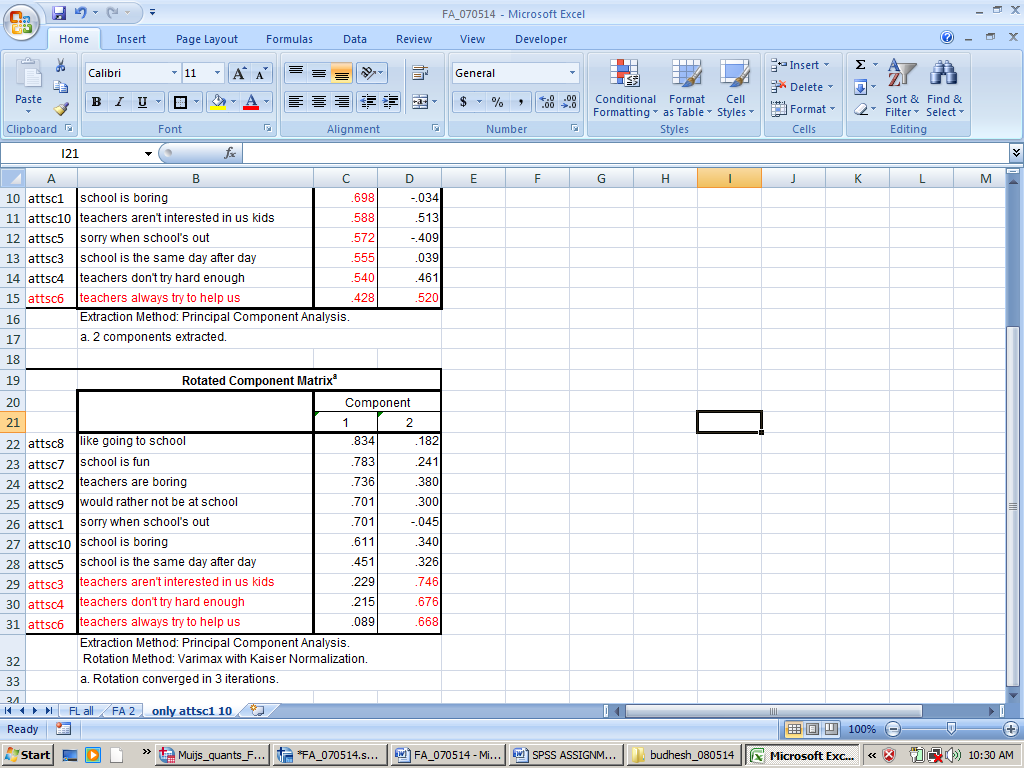
Screen shot: 3.3.2-1

*All 54 variables used in Factor Analysis*



Screen shot 3.3.2-2:

*Variables attsc1 through attsc10 applied for factor analysis*

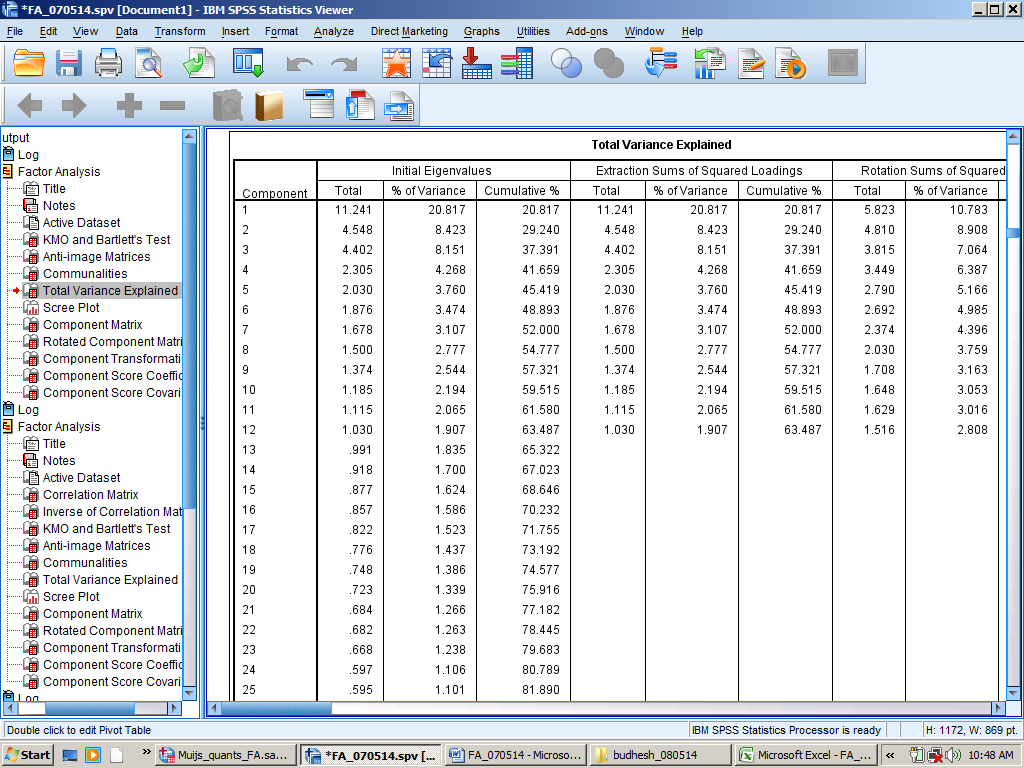


**3.3.2 (b) How many factors will you extract? Why?**

Though 12 factors were extracted through factor analysis after applying Principal Component analysis and Varimax rotation, Scree Plot suggested that **only three factors** should be considered. Cumulative sum of squared loading was found as 26.755 percent after rotation.

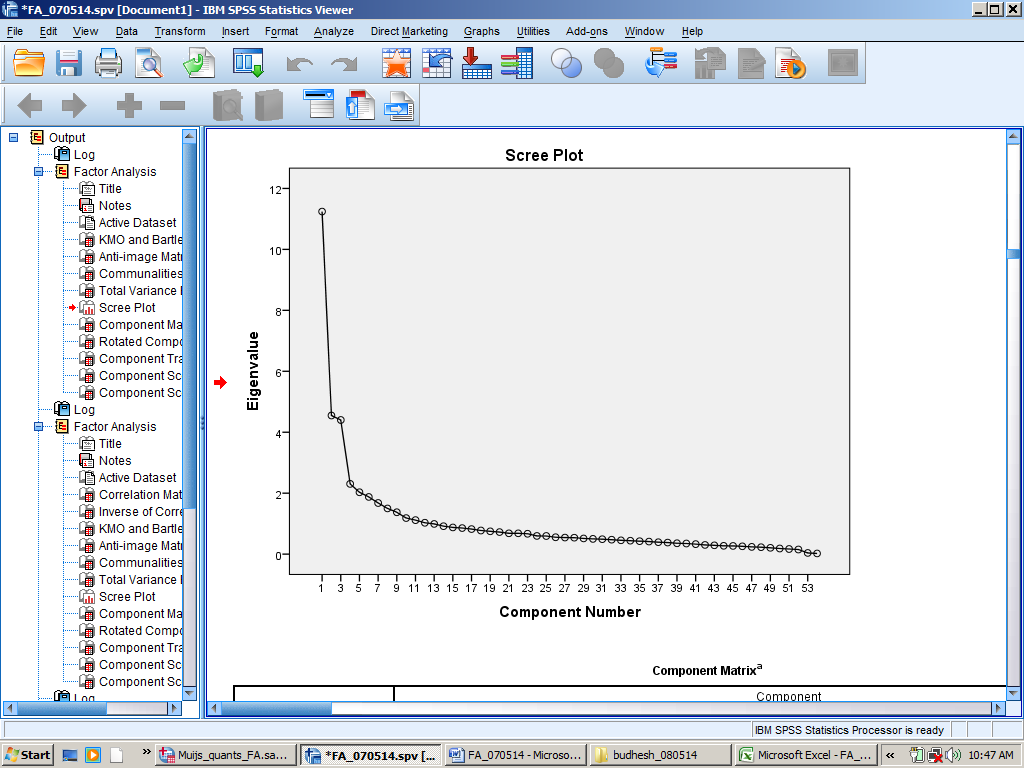
Screen shot 3.3.2- 3:

*Total Components extracted*



Screen shot 3.3.2- 4:

*Scree Plot*



Factor 1

Factor 2

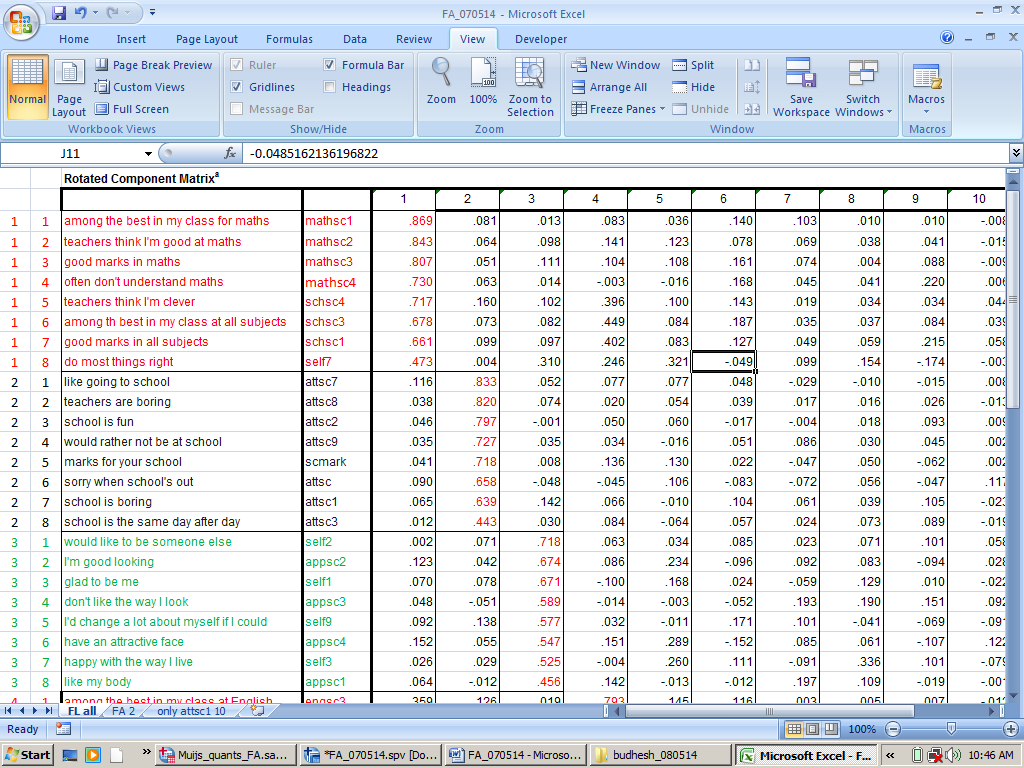
Factor 3

**3.3.2 (c) What is the factor loading telling you?**

Factor loading is the correlation between items/variables and the factor to which they belong. Factor loadings have been shown in Screen shot 3.3.2-5: *Rotated Component Matrix* for the first 3 extracted factors or components. Factor loadings tell the strength of relationship between item and factor. For example, in Factor 1, mathsc1 (among the best in my class for maths) had the highest factor loading as 0.869 and schsc1 (good marks in all subjects) had the lowest factor loading as 0.661. This means that mathsc1 was the most strongly related to Factor 1 among all 8 extracted variables and schsc1 was the most least related to Factor 1 among all 8 extracted variables.

Screen shot 3.3.2- 5:

*Rotated Component Matrix*



**3.3.2 (d) Two extracted factors.**

Factor analysis was applied to all 54 scale and ordinal variables and requested SPSS to extract only two factors. In this attempt 27 items were belonging to first Factor and 25 items were belonging to Factor 2, further, all items representing ‘attitude towards school’ from attsc1 through attsc10 fall under Factor 1.

**3.3.2 (e) Would the five variables measuring attitudes to school form an internally consistent scale? Why?**

The first five variables measuring ‘attitude towards school’ were tested for their reliability and Cronbach’s Alpha was found more than 0.709 which is slightly higher than the generally accepted threshold of 0.700. The SPSS Output 3.3.2-2: Reliability Test shows the Cronbach’s Alpha. Hence, this can be safely assumed that the scale of measurement had internal consistency.

SPSS Output 3.3.2- 2:

*Reliability Test for first five items*

|  |  |
| --- | --- |
| **Reliability Statistics** | |
| Cronbach's Alpha | N of Items |
| .709 | 5 |

**3.3.2 (f) Would internal consistency improve if we left any of the five variables out of the scale?**

We have tested for all 10 items measuring attitude towards school and for reliability and found Cronbach’s Alpha as 0.857 which showed very good internal consistency. Later we had taken out the first five items and run the reliability test. This time we found Cronbach’s Alpha as 0.785 which showed that even after deleting first five items, the internal consistency remained intact.

SPSS Output 3.3.2- 3:

*Reliability Tests*

|  |  |
| --- | --- |
| All 10 items used for Reliability Test | Items attsc6 through attsc10 used for Reliability Test |
| |  |  | | --- | --- | | **Reliability Statistics** | | | Cronbach's Alpha | N of Items | | .857 | 10 | | |  |  | | --- | --- | | **Reliability Statistics** | | | Cronbach's Alpha | N of Items | | .785 | 5 | |