## *TABLE OF CONTENT*

1.Introduction Regression Analysis

1.1 General Information

1.2 Analysis problem and aim

1.3 Define Data Set

2.Regression Analysis

2.1 Model Summary

2.2 Anova Table

2.3 Correlations

2.4 Coefficients

3. FORECASTING

3.1 FORECASTING’s CONFIDENCE INTERVAL

4.CHECK THE ASSUMPTION

# 1.Introduction Regression Analysis

## 1.1 General Information

Regression analysis, two or more variables with cause-effect relationship between them to determine the relationship between and estimation using that relationship or to make predictions.Numerical correlation between a criterion variable and one or more forecast variables It is the statistical analysis used in converting into halo. Regression analysis mainly variables It aims to determine the nature of the relationship between. A variable as a prediction variable simple regression, if used, two or more variables as prediction variables If used, it is possible to talk about multiple regression analysis. The goal is every guess variable's contribution to the total change in the criterion variable and hence the criterion based on the value of the linear combination of prediction variables value is estimated. Regression with one dependent variable and more than one independent variableRegression with one dependent variable and more than one independent variablemodels are called multivariate regression analysis.The change of independent variables simultaneously (simultaneously) in the dependent variable.I try to explain the univariate regression analysis in terms of computation and interpretation.Interpretation of multivariate regression analysis to univariate regression analysisHowever, there are some differences.For example, in univariate regression analysis multiple regression coefficient is expressed as R (multiple R).The coefficient R is multiple (simultaneously) handled simultaneously with the change in a dependent variable.shows the degree of the relationship between the change in the independent variable.In other words, the change in a group of independent variables that are considered together with the dependent variableIt is an indication of the relationship (correlation).Multivariate regression analysis in many branches of social sciences Behavioral movements in branches of science such as marketing, sociology and psychology.determining the factors affecting the economic variables of the time series type in the economy.It finds use in its detection and projections (forecasting) for the future.

### 1.1.1 Multiple Regression Analysis Models

- Enter method: Entering arguments as a block in one step method in which it is evaluated.
- Forward selection: Highest with dependent variable the positive or negative correlated argument is selected first.The hypothesis that the coefficient of the entered variable is 0 is examined with the F test.The F value is compared to the predicted F values of SPSS. SPSS has two F criteria. The F value is compared with a minimum F value that you determine. (F-toenter, FIN). The normal setting is 3.84."Use F Value" in the window passed through the "options" key.It is part.Setting the probability associated with the F statistic (Porbability of F-toenter, PIN). The normal setting for this is 0.05.If the F value obtained is less than or equal to these criteria, that argument the regression is taken into consideration and the selection continues forward, otherwise the process is there.It is stopped.

- Backward Elimination Method: Forward selection on the contrary, here all arguments are selected first, and then in order to certain criteria.Screening is done according to SPSS; it sets two criteria for screening.The least squares F value (F-toremove, FOUT) to keep the variable in the formula. Normal setting is 2.71. The greatest probability of F (Probabability of F-to-remove, POUT).Setting is 0.10.First, the variable with the smallest partial correlation coefficient is examined.A variable with a large value is eliminated from the values.
- Stepwise Selection Method: First argument if this meets the FIN or PIN requirements in the right choice,the second variable is selected, or the process ends there.The variable with partial correlation is taken. Selections from high correlation to low.If the independent variables meet the criteria, regression analysis is started.Selecting Step by Step after the variable is selected differs from choosing forward.The variable matches the FOUT or POUT criteria as in the backward elimination so it is checked both step by step and forward selection.Reverse sieving operations are done.

## 1.2 Analysis problem and aim

In this research, I decided to do it because of the incident we have experienced in recent months. In recent months, I think that there has been a decrease in the level of people happiness due to earthquakes and viruses. So I decided to investigate the happiness level of the first hundred countries of 2019,and one of my reasons for this research was to make people understand that it is necessary to work so that they do not become unhappy every year. In this study, we take the happiness score as a dependent variable and look at the factors affecting the happiness score. I think other factors that I think affect the happiness score are GDP per capita, Social support, Freedom to make life choices, and Perceptions of corruption,so independent variables are GDP per capita,Social support, Freedom to make life choices, and Perceptions of corruption.

## 1.3 Define Data Set

This is my data set definetion.In this all data are numeric but country is string,and country didnot effect regression analysis.Also score is dependent variable,and other variables are independent.Also in this research used to enter method.

| Name | Type | Width | Decimals | Label | Values |
|---|---|---|---|---|---|
| Countryorr... | String | 22 | 0 | Country or reg... | None |
| Score | Numeric | 18 | 15 | | None |
| GDPperca... | Numeric | 19 | 15 | GDP per capita | None |
| Socialsupport | Numeric | 19 | 15 | Social support | None |
| Freedomto... | Numeric | 21 | 15 | Freedom to m... | None |
| Perception... | Numeric | 21 | 15 | Perceptions of... | None |

**Descriptive Statistics**

| | Mean | Std. Deviation | N |
|---|---|---|---|
| Score | 6,072569999999998 | ,737496012726767 | 100 |
| GDP per capita | 1,111710000000000 | ,274388145597517 | 100 |
| Social support | 1,362510000000000 | ,172099307424991 | 100 |
| Freedom to make life choices | ,435210000000000 | ,122162443186327 | 100 |
| Perceptions of corruption | ,118720000000000 | ,107420732344244 | 100 |

This table figure out my date set variables mean, standard deviation,and values of variables.

## 2.Regression Analysis

## 2.1 Model Summary

**Model Summary[b]**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,817[a] | ,668 | ,654 | ,433950554393937 |

a. Predictors: (Constant), Perceptions of corruption, Social support, Freedom to make life choices, GDP per capita

b. Dependent Variable: Score

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.The definition of R-squared is fairly straight-forward; it is the percentage of the response variable variation that is explained by a linear model. Or:

R-squared = Explained variation / Total variation

R-squared is always between 0 and 100%:

- 0% indicates that the model explains none of the variability of the response data around its mean.
- 100% indicates that the model explains all the variability of the response data around its mean.

In general, the higher the R-squared, the better the model fits your data.In myresearch R squared is %66 but we use adjusted R square,because we use multiple regression analysis,so this R value was %65.Also this value sad that Taken as a set the predictors GDP per capita, Social support, Freedom to make life choices, and Perceptions of corruption for %65 of the variance in score.

2.2 Anova Table

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 35,956 | 4 | 8,989 | 47,735 | ,000[b] |
| | Residual | 17,890 | 95 | ,188 | | |
| | Total | 53,846 | 99 | | | |

a. Dependent Variable: Score

b. Predictors: (Constant), Perceptions of corruption, Social support, Freedom to make life choices, GDP per capita

ANOVA is used to compare differences of means among more than 2 groups. It does this by looking at variation in the data and where that variation is found (hence its name). Specifically, ANOVA compares the amount of variation between groups with the amount of variation within groups. It can be used for both observational and experimental studies.Here is my Anova table. Anova table is a test of whether that R square is greater than 0. My P value is less than 0.05. So that means that the test is significant, regression is significant. Also that means my predictors are able to account for a signigicant amount of variance in score. My regression model is significant. So i can predict score significantly.Also The difference between the observed value of the dependent variable ($y$) and the predicted value ($\hat{y}$) is called the **residual** ($e$).Residual = Observed value - Predicted value $\boldsymbol{e = y - \hat{y}}$ *(in general).*Also in regression part df 4 is related to independent variables. **Residual df** is the total number of observations (rows) of the dataset subtracted by the number of variables being estimated. **Regression SS** is the total variation in the dependent variable that is explained by the regression model. It is the sum of the square of the difference between the predicted value and mean of the value of all the data points. **Residual SS** — is the total variation in the dependent variable that is left unexplained by the regression model. It is also called the **Error Sum of Squares** and is the sum of the square of the difference between the actual and predicted values of all the data points. Mean Squared Errors (MS) — are the mean of the sum of squares or the sum of squares divided by the degrees of freedom for both, regression and residuals.Regression MS = $\sum (\hat{y} - \bar{y})^2$/Reg. Df Residual MS = $\sum (y - \hat{y})^2$/Res. Df.**F** — is used to test the hypothesis that the slope of the independent variable is zero. Mathematically, it can also be calculated as F = Regression MS / Residual MS.This is otherwise calculated by comparing the F-statistic to an F distribution with regression df in numerator degrees and residual df in denominator degrees.**Significance F** — is nothing but the p-value for the null hypothesis that the coefficient of the independent variable is zero and as with any p-value, a low p-value indicates that a significant relationship exists between dependent and independent variables.

**Test Using Alpha = .05**

**This overall regression model is significant.**

**F(4,95) = 48,p < 0.01, R Square = .668**

## 2.3 Correlations

**Correlations**

| | | Score | GDP per capita | Social support | Freedom to make life choices | Perceptions of corruption |
|---|---|---|---|---|---|---|
| Pearson Correlation | Score | 1,000 | ,646 | ,620 | ,530 | ,638 |
| | GDP per capita | ,646 | 1,000 | ,579 | ,201 | ,496 |
| | Social support | ,620 | ,579 | 1,000 | ,324 | ,364 |
| | Freedom to make life choices | ,530 | ,201 | ,324 | 1,000 | ,484 |
| | Perceptions of corruption | ,638 | ,496 | ,364 | ,484 | 1,000 |
| Sig. (1-tailed) | Score | . | ,000 | ,000 | ,000 | ,000 |
| | GDP per capita | ,000 | . | ,000 | ,023 | ,000 |
| | Social support | ,000 | ,000 | . | ,001 | ,000 |
| | Freedom to make life choices | ,000 | ,023 | ,001 | . | ,000 |
| | Perceptions of corruption | ,000 | ,000 | ,000 | ,000 | . |
| N | Score | 100 | 100 | 100 | 100 | 100 |
| | GDP per capita | 100 | 100 | 100 | 100 | 100 |
| | Social support | 100 | 100 | 100 | 100 | 100 |
| | Freedom to make life choices | 100 | 100 | 100 | 100 | 100 |
| | Perceptions of corruption | 100 | 100 | 100 | 100 | 100 |

This table show that correlation between each variable.The important correlation is score between other variables,because score is dependent variable in this research.Also statistics expert sad that your correlation values must not be over .800.I look the my values.This are not over than .800.

**Coefficients[a]**

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95,0% Confidence Interval for B | |
|---|---|---|---|---|---|---|---|
| | B | Std. Error | Beta | | | Lower Bound | Upper Bound |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | (Constant) | 2,732 | ,369 | | 7,401 | ,000 | 1,999 | 3,464 |
| | GDP per capita | ,843 | ,213 | ,314 | 3,954 | ,000 | ,420 | 1,266 |
| | Social support | 1,113 | ,322 | ,260 | 3,459 | ,001 | ,474 | 1,752 |
| | Freedom to make life choices | 1,544 | ,421 | ,256 | 3,671 | ,000 | ,709 | 2,379 |
| | Perceptions of corruption | 1,812 | ,524 | ,264 | 3,456 | ,001 | ,771 | 2,852 |

a. Dependent Variable: Score

## 2.4 Coefficients

Regression coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant. This statistical control that regression provides is important because it isolates the role of one variable from all of the others in the model. The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). A low p-value (< 0.05) indicates that you can reject the null hypothesis. In other words, a predictor that has a low p-value is likely to be a meaningful addition to your model because changes in the predictor's value are related to changes in the response variable,so I evaluated each of these test for alpha of .05.Also we look to significant level.If their significant level is less than 0.05,we sad that it is significant. Standard Error — provides the estimated standard deviation of the distribution of coefficients. It is the amount by which the coefficient varies across different cases. A coefficient much greater than its standard error implies a probability that the coefficient is not 0.

**t-Stat** — is the t-statistic or t-value of the test and its value is equal to the coefficient divided by the standard error.

t-Stat = Coefficients/Standard Error

Again, the larger the coefficient with respect to the standard error, the larger the t-Stat is and higher the probability that the coefficient is away from 0.

**p-value** — The t-statistic is compared with the t distribution to determine the p-value. We usually only consider the p-value of the independent variable which provides the likelihood of obtaining a sample as close to the one used to derive the regression equation and verify if the slope of the regression line is actually zero or the coefficient is close to the coefficient obtained.

A p-value below 0.05 indicates 95% confidence that the slope of the regression line is not zero and hence there is a significant linear relationship between the dependent and independent variables.A p-value greater than 0.05 indicates that the slope of the regression line may be zero and that there is not sufficient evidence at the 95% confidence level that a significant linear relationship exists between the dependent and independent variables.

**Lower and Upper 95%** — Since we mostly use a sample of data to estimate the regression line and its coefficients, they are mostly an approximation of the true coefficients and in turn

the true regression line. The lower and upper 95% boundaries give the 95th confidence interval of lower and upper bounds for each coefficient.

**Score Significant p = .000**

**GDP per capita Significant p=.000**

**Social Support Significant p=.001**

**Freedom to make life Choices Significant p=.000**

**Perceptions of Corruption Significant p=.001**

## 3. CALCULATION OF FORECAST

Recall that predictions of yy can be obtained using^yt=^β0+^β1x1,t+^β2x2,t+⋯+^βkxk,t,yt^=β^0+β^1x1,t+β^2x2,t+⋯+β^kxk,t,which comprises the estimated coefficients and ignores the error in the regression equation. Plugging in the values of the predictor variables x1,t,…,xk,tx1,t,…,xk,t for t=1,…,Tt=1,…,T returned the fitted (training-sample) values of yy. What we are interested in here, however, is forecasting *future* values of yy.

First of all let us calculate estimation of row of my dataset;

According to formula and my coefficients table

**Coefficients[a]**

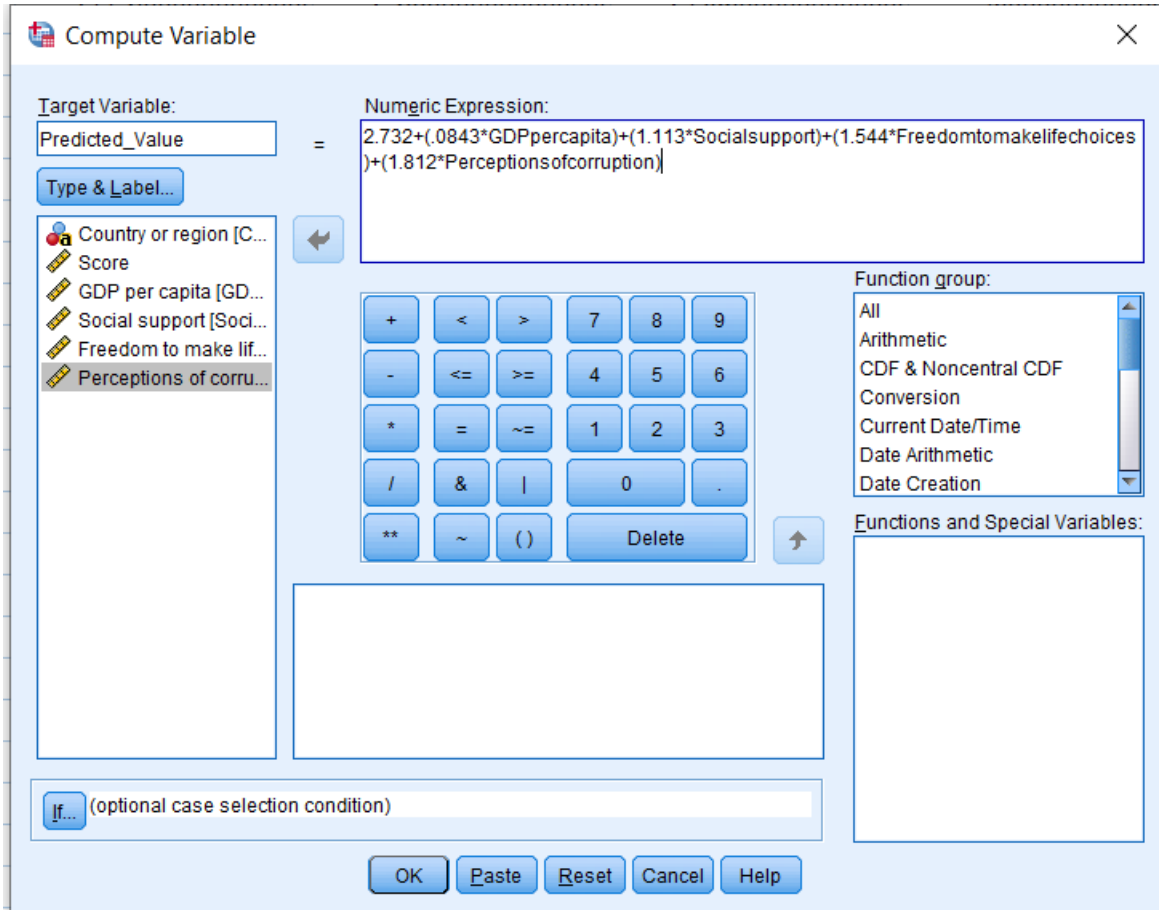
| Model | | Unstandardized Coefficients | | Standardized Coefficients | | | 95,0% Confidence Interval for B | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. | Lower Bound | Upper Bound |
| 1 | (Constant) | 2,732 | ,369 | | 7,401 | ,000 | 1,999 | 3,464 |
| | GDP per capita | ,843 | ,213 | ,314 | 3,954 | ,000 | ,420 | 1,266 |
| | Social support | 1,113 | ,322 | ,260 | 3,459 | ,001 | ,474 | 1,752 |
| | Freedom to make life choices | 1,544 | ,421 | ,256 | 3,671 | ,000 | ,709 | 2,379 |
| | Perceptions of corruption | 1,812 | ,524 | ,264 | 3,456 | ,001 | ,771 | 2,852 |

a. Dependent Variable: Score

B1=2.732 B3=1.113 B5=1.812

B2=.0843 B4=1.544

Then i calculated my predictions via SPSS. In order to make that calculation i clicked "Transform" from top menu then i clicked "Compute Variable" option.




After that all my predictions are calculated. As you can see down there my predictions are under the new variable which is “Predicted_Value”.

| | Countryorregion | Score | GDPpercapita | Socialsupport | Freedomtomakelifechoices | Perceptionsofcorruption | Predicted_Value |
|---|---|---|---|---|---|---|---|
| 1 | Finland | 7,769000000000000 | 1,340000000000000 | 1,587000000000000 | ,596000000000000 | ,393000000000000 | 6,24 |
| 2 | Denmark | 7,600000000000000 | 1,383000000000000 | 1,573000000000000 | ,592000000000000 | ,410000000000000 | 6,26 |
| 3 | Norway | 7,554000000000000 | 1,488000000000000 | 1,582000000000000 | ,603000000000000 | ,341000000000000 | 6,17 |
| 4 | Iceland | 7,494000000000000 | 1,380000000000000 | 1,624000000000000 | ,591000000000000 | ,118000000000000 | 5,78 |
| 5 | Netherlands | 7,488000000000000 | 1,396000000000000 | 1,522000000000000 | ,557000000000000 | ,298000000000000 | 5,94 |
| 6 | Switzerland | 7,480000000000000 | 1,452000000000000 | 1,526000000000000 | ,572000000000000 | ,343000000000000 | 6,06 |
| 7 | Sweden | 7,343000000000000 | 1,387000000000000 | 1,487000000000000 | ,574000000000000 | ,373000000000000 | 6,07 |
| 8 | New Zealand | 7,307000000000000 | 1,303000000000000 | 1,557000000000000 | ,585000000000000 | ,380000000000000 | 6,17 |
| 9 | Canada | 7,278000000000001 | 1,365000000000000 | 1,505000000000000 | ,584000000000000 | ,308000000000000 | 5,98 |
| 10 | Austria | 7,246000000000000 | 1,376000000000000 | 1,475000000000000 | ,532000000000000 | ,226000000000000 | 5,72 |
| 11 | Australia | 7,228000000000000 | 1,372000000000000 | 1,548000000000000 | ,557000000000000 | ,290000000000000 | 5,96 |
| 12 | Costa Rica | 7,167000000000000 | 1,034000000000000 | 1,441000000000000 | ,558000000000000 | ,093000000000000 | 5,45 |
| 13 | Israel | 7,139000000000000 | 1,276000000000000 | 1,455000000000000 | ,371000000000000 | ,082000000000000 | 5,18 |
| 14 | Luxembourg | 7,090000000000000 | 1,609000000000000 | 1,479000000000000 | ,526000000000000 | ,316000000000000 | 5,90 |
| 15 | United Kingdom | 7,054000000000000 | 1,333000000000000 | 1,538000000000000 | ,450000000000000 | ,278000000000000 | 5,75 |
| 16 | Ireland | 7,021000000000000 | 1,499000000000000 | 1,553000000000000 | ,516000000000000 | ,310000000000000 | 5,95 |
| 17 | Germany | 6,985000000000000 | 1,373000000000000 | 1,454000000000000 | ,495000000000000 | ,265000000000000 | 5,71 |
| 18 | Belgium | 6,923000000000000 | 1,356000000000000 | 1,504000000000000 | ,473000000000000 | ,210000000000000 | 5,63 |
| 19 | United States | 6,892000000000000 | 1,433000000000000 | 1,457000000000000 | ,454000000000000 | ,128000000000000 | 5,41 |
| 20 | Czech Republic | 6,852000000000000 | 1,269000000000000 | 1,487000000000000 | ,457000000000000 | ,036000000000000 | 5,26 |
| 21 | United Arab Emirates | 6,825000000000000 | 1,503000000000000 | 1,310000000000000 | ,598000000000000 | ,182000000000000 | 5,57 |
| 22 | Malta | 6,726000000000000 | 1,300000000000000 | 1,520000000000000 | ,564000000000000 | ,151000000000000 | 5,68 |
| 23 | Mexico | 6,595000000000000 | 1,070000000000000 | 1,323000000000000 | ,433000000000000 | ,073000000000000 | 5,10 |
| 24 | France | 6,592000000000000 | 1,324000000000000 | 1,472000000000000 | ,436000000000000 | ,183000000000000 | 5,49 |
| 25 | Taiwan | 6,446000000000000 | 1,368000000000000 | 1,430000000000000 | ,351000000000000 | ,097000000000000 | 5,16 |
| 26 | Chile | 6,444000000000000 | 1,159000000000000 | 1,369000000000000 | ,357000000000000 | ,056000000000000 | 5,01 |
| 27 | Guatemala | 6,436000000000000 | ,800000000000000 | 1,269000000000000 | ,535000000000000 | ,078000000000000 | 5,18 |
| 28 | Saudi Arabia | 6,375000000000000 | 1,403000000000000 | 1,357000000000000 | ,439000000000000 | ,132000000000000 | 5,28 |
| 29 | Qatar | 6,374000000000000 | 1,684000000000000 | 1,313000000000000 | ,555000000000000 | ,167000000000000 | 5,49 |
| 30 | Spain | 6,354000000000000 | 1,286000000000000 | 1,484000000000000 | ,362000000000000 | ,079000000000000 | 5,19 |

Also,we see that the predicted values are roughly close to the scores.This shows that our prediction is almost correct.

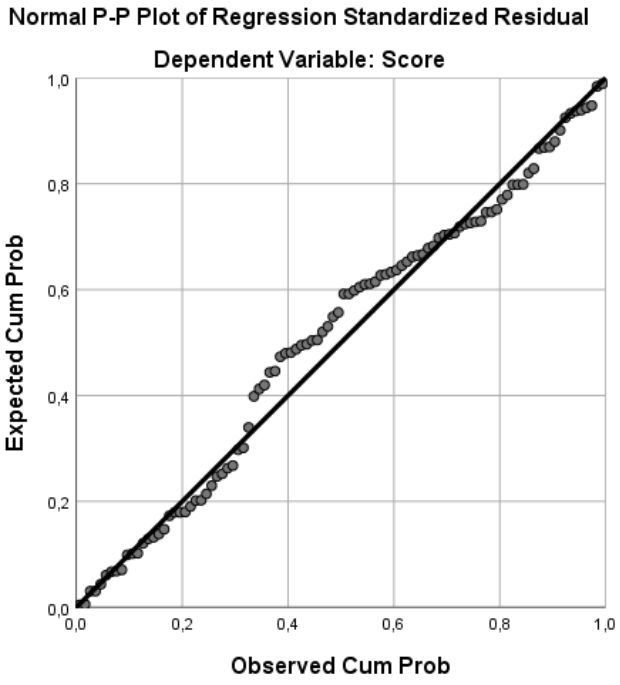
## 3.1 FORECASTING's CONFIDENCE INTERVAL

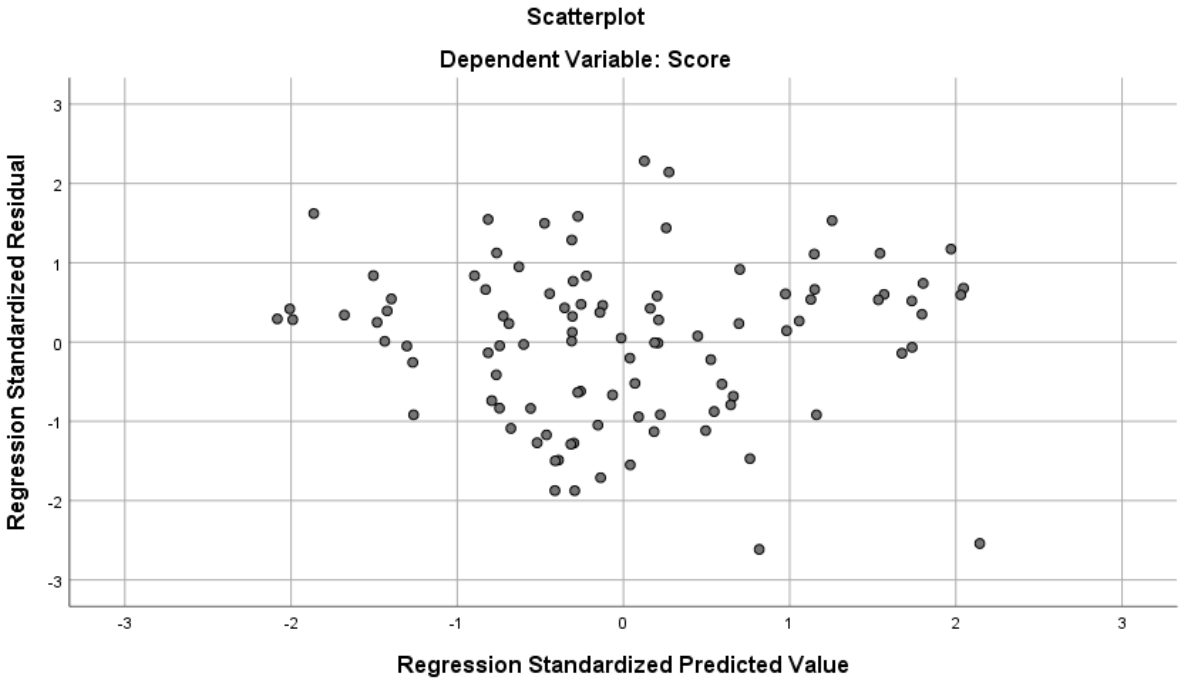
**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| Predicted_Value | Mean | | 5,2293 | ,04562 |
| | 95% Confidence Interval for Mean | Lower Bound | 5,1388 | |
| | | Upper Bound | 5,3198 | |
| | 5% Trimmed Mean | | 5,2218 | |
| | Median | | 5,1929 | |
| | Variance | | ,208 | |
| | Std. Deviation | | ,45620 | |
| | Minimum | | 4,28 | |
| | Maximum | | 6,26 | |
| | Range | | 1,97 | |
| | Interquartile Range | | ,52 | |
| | Skewness | | ,300 | ,241 |
| | Kurtosis | | -,153 | ,478 |

Then i got the results. Here %95 Confidence Interval for my prediction values. Lower Bound is 5.1388. Upper bounds is 5.3198. %95 of all values from my predicted values are falling between 5.1388 and 5.3198.

## 4.CHECK THE ASSUMPTION






This show that heteroscedasticity,because Heteroscedasticity means unequal scatter. In regression analysis, we talk about heteroscedasticity in the context of the residuals or error term. Specifically, heteroscedasticity is a systematic change in the spread of the residuals over the range of measured values. Heteroscedasticity is a problem because ordinary least squares (OLS) regression assumes that all residuals are drawn from a population that has a constant variance (homoscedasticity).