Data Analysis
10676333
Section 1 (9:30-11:00)

MIS

Main Assignment

Descriptive & Inferential Analysis

Prepared by:

Ali Abu Nimah (12112503)

Presented for:

Dr.Mohammed Dwikat

Due Date is 04/06/2024

Analyzing High School Data

- ## Introduction

  We receive a file that contains 190 cases. We want to summarize it using Descriptive Statistics and Inferential, in addition to using visualization chart and G-Power.
  We investigate for outliers and missing values, no outliers exist, no missing values exist.

- ## Executive Summary

  This study aims to analyze high school student data to uncover patterns and relationships within the dataset, focusing on variables such as gender, school type, native language, and subject scores (Math, Reading, Writing, Physics, Art). The research addresses the reliability and quality of the data, employing various statistical methods including normality tests, t-tests, ANOVA, regression analysis, correlation analysis, and chi-square tests to investigate the research problems. The analysis reveals significant gender differences in Reading and Physics scores, with males generally outperforming females in Physics. Socio-economic status (SES) also significantly affects Physics scores, indicating better performance among higher SES students. Additionally, a significant positive correlation between Math and Physics scores suggests that proficiency in one subject is associated with higher performance in the other. The study finds no significant association between native language and reading performance, and the gender distribution does not match the hypothesized proportions. These findings highlight the need for targeted educational interventions to address gender disparities and support lower SES students, providing a basis for further research and policy-making to enhance educational outcomes and equity.

- ## Reliability Analysis
  - We check for reliability and the Cronbach alpha = 76.3% which is acceptable.

Table 1 Reliability Check Table

Reliability Statistics

Cronbach's Alpha	N of Items
.763	10

- ## Missing Values
  - We check if there are any missing values in the data we received.

Table 2 Missing Values

Statistics

		Gender	Native Language	SES	School	Field	Reading	Writing	Math	Physics	Art
N	Valid	190	190	190	190	190	190	190	190	190	190
	Missing	0	0	0	0	0	0	0	0	0	0

As we notice there are no missing values in any of the data we have.

- **Descriptive Analysis**

- **Categorical Data**

Table 3 Gender Frequency Table

Gender

	Frequency	Percent	Cumulative Percent
Male	100	52.6	52.6
Female	90	47.4	100.0
Total	190	100.0	

We notice that there's no intangible difference between the number of males and females as their percentage most likely the same as follows in the bar chart too:

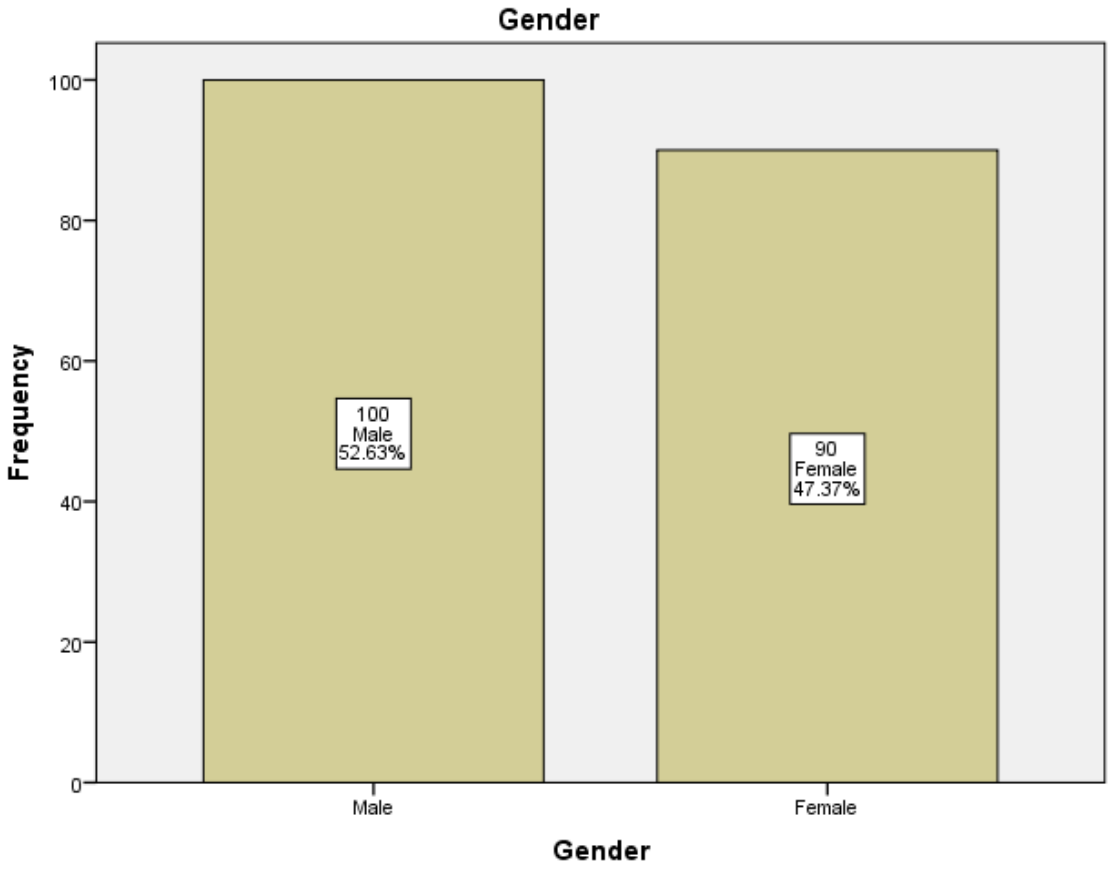

Figure 1 Bar Chart Gender Frequency

Table 4 School Frequency Table

School

	Frequency	Percent	Cumulative Percent
Government	162	85.3	85.3
Private	28	14.7	100.0
Total	190	100.0	

There are huge differences of the students attempted to Government schools rather than going to Private ones so the Government schools are so much higher as follows in the pie chart below:

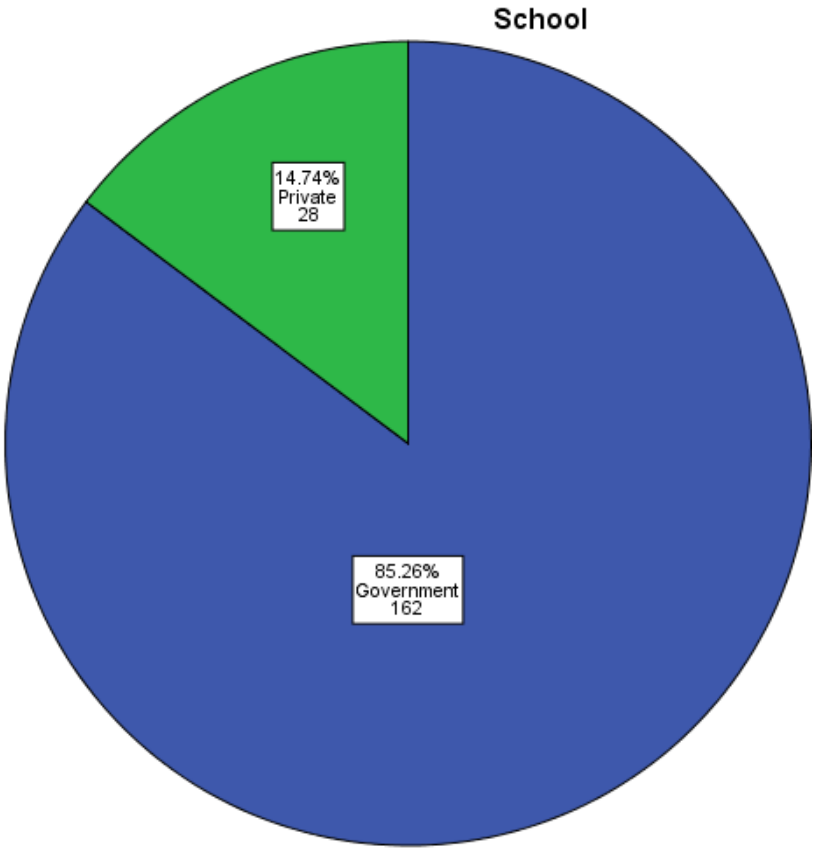

Figure 2 Pie Chart School Frequency

Table 5 Crosstab Table of Gender & Native Language

Gender * NativeLanguage Crosstabulation

Count

		NativeLanguage				Total
		French	Chinese	Other	English	
Gender	Male	9	4	4	83	100
	Female	8	5	8	69	90
Total		17	9	12	152	190

We notice that the student's main language is English according to the chart as it's the highest percentage for both males and females.

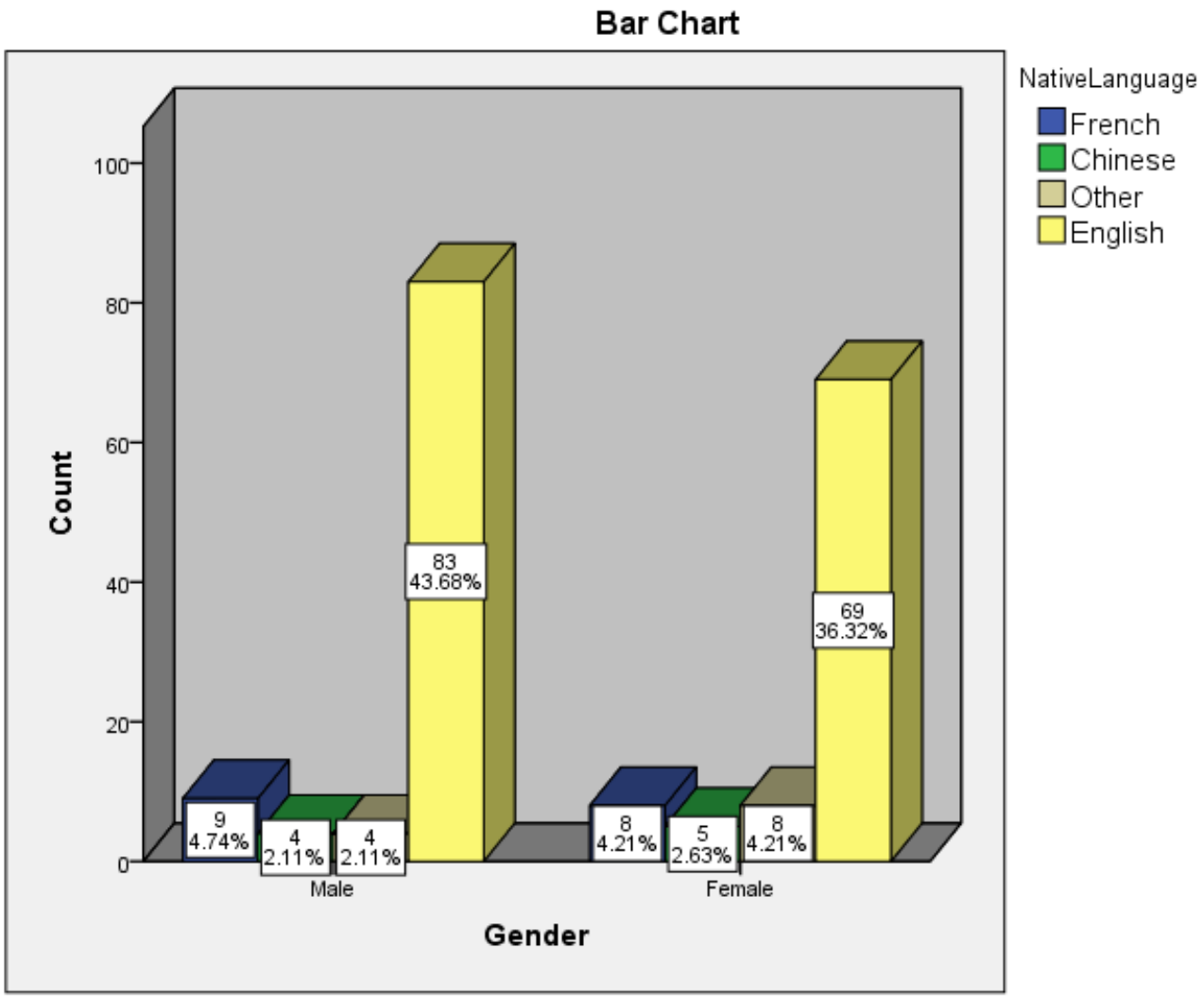

*Figure 3 Clustered Bar Chart Gender * Native Language Chart*

- **Scale Data**

Table 6 Average Math Score of Total Students

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Math	190	44	97	69.79	12.441

The average score of all students is 69.79 for Math subject.

- **"Using Split file"**

Table 7 Comparing Students According to Fields with their Average of Both Art and Reading

Descriptive Statistics

Field		N	Minimum	Maximum	Mean	Std. Deviation
Business	Art	50	52	92	71.70	12.334
	Reading	50	37	90	65.50	12.554
Literature	Art	53	45	98	81.06	13.099
	Reading	53	45	101	77.53	13.479
Vocation	Art	51	38	92	67.18	15.246
	Reading	51	41	97	63.78	12.971
Scientific	Art	36	58	98	78.86	9.960
	Reading	36	52	90	69.67	10.337

We notice that the Literature filed students with the highest average scores for both Art and Reading subjects, while the lowest is the Vocation field and both the Scientific and Business fields come in between them.

- **“Using Explore”**

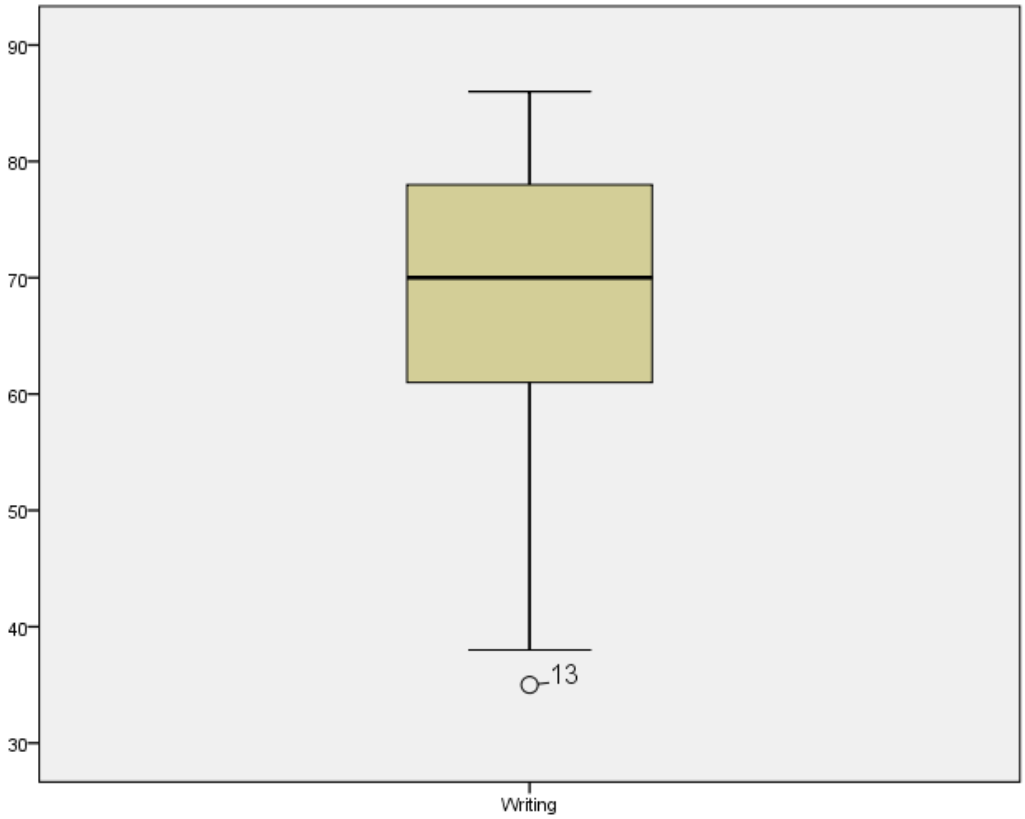

Figure 4 Writing Box Plot

We notice that in the Writing subject there is a down outlier that exist which is in the 13th cell and its value is “35”, but every other subject which are Reading, Math, Physics, no outliers exist, no missing value exist when we checked.

- **Inferential Statistics**

- **One Sample T-Test/Wilcoxon**

  - Checking for normality distribution for Writing Subject

    Confidence level = 95%, Alpha = 5%

    H_0: Writing Grades are normally distributed.

    H_a: Writing Grades are NOT normally distributed.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Writing is normal with mean 68.384 and standard deviation 11.70.	One-Sample Kolmogorov-Smirnov Test	.008	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 5 Normality Distribution Check for Writing Subject

As we notice from the Sig. the null hypothesis is rejected so the grades ARE NOT normally distributed and we will attend to use Wilcoxon signed rank test which uses the median.

  - We want to check if the median of Writing grades equal 70.

    H_0: Is the Median of Writing Grades = 70

    H_a: Is the Median of Writing Grades ≠ 70

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of Writing equals 70.000.	One-Sample Wilcoxon Signed Rank Test	.303	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 6 Wilcoxon for Writing

We notice from the Sig. that the median of Writing grades = 70 so we do nothing as the exams are as we expected and hoped for.

- Checking for normality distribution for Math Subject

  Confidence level = 95%, Alpha = 5%

  H_0: Math Grades are normally distributed.

  H_a: Math Grades are NOT normally distributed.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Math is normal with mean 69.789 and standard deviation 12.44.	One-Sample Kolmogorov-Smirnov Test	.300	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 7 Normality Distribution Check for Math Subject

As we notice from the Sig. the null hypothesis is retained so the grades ARE normally distributed and we will attend to use One Sample T-Test which uses the mean.

- We want to check if the mean of Math grades equal 70.

  H_0: Is the Average of Math Grades = 70

  H_a: Is the Average of Math Grades ≠ 70

Table 8 Statistics of Math Grades

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Math	190 190	53	44	97	13260	69.79	.903	12.441	154.781	.163	.176	-.751	.351

Table 9 One Sample Test for Math Grades

One-Sample Test

	Test Value = 70					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Math	-.233	189	.816	-.211	-1.99	1.57

As we notice from the Sig. the average of Math grades = 70 as the null hypothesis is retained and we are satisfied with what we predicted as the result is positive thought.

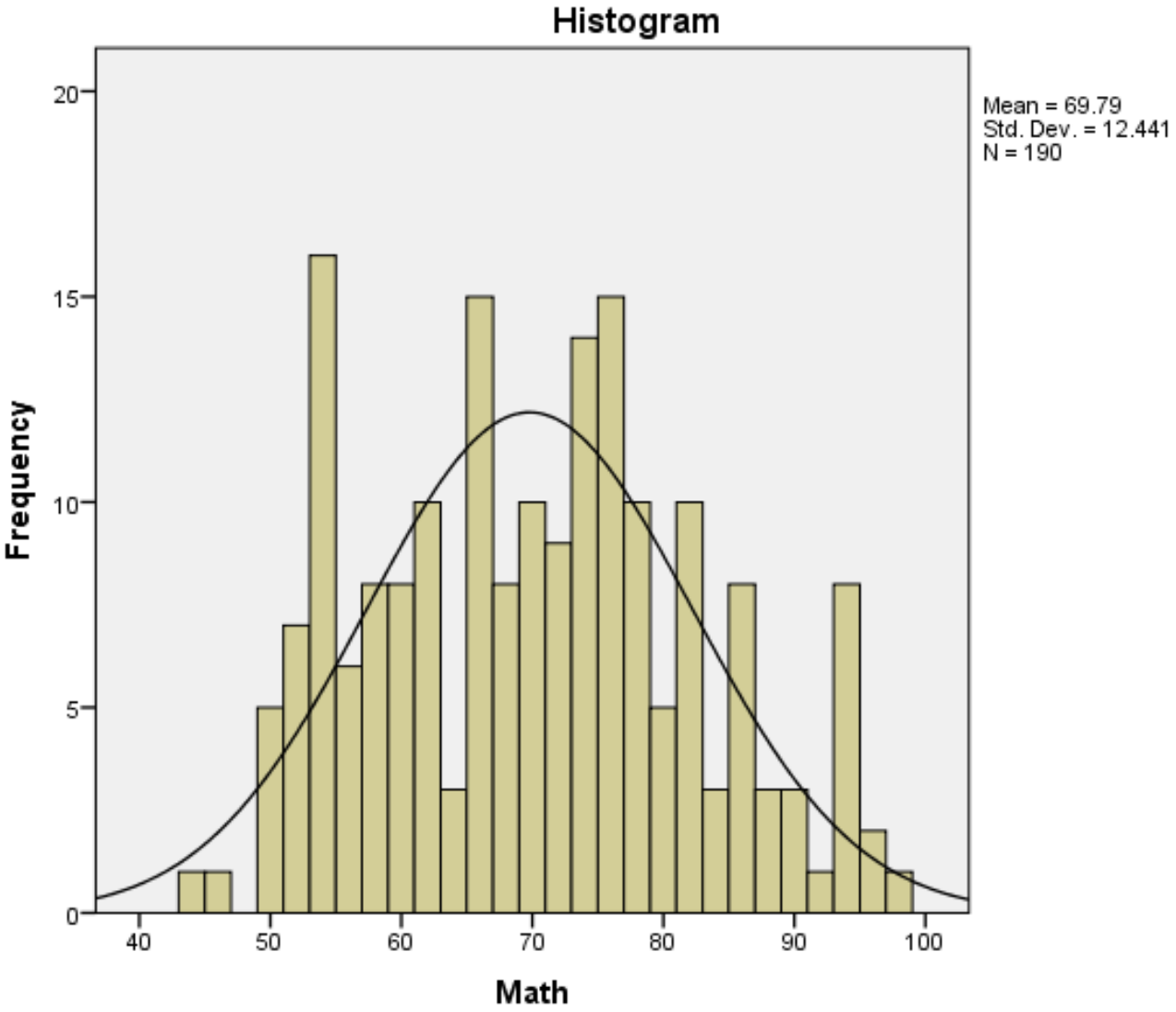

Figure 8 Histogram of Math

The Math grades normally distributed. The histogram shows a peak around the 60-80 range, which corresponds to the mean score of approximately 69.79. Additionally, the standard deviation of 12.41 indicates moderate variability around the mean. While there may be slight deviations from perfect normality (as indicated by skewness and kurtosis).

- **Paired Samples T-Test**
  - We want to check if Math grades change according to Physics grades

    H_0: Is Average score of Physics = Average score of Math

    H_a: Is Average score of Physics ≠ Average score of Math

Table 10 Physics & Math Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Math	69.79	190	12.441	.903
	Physics	65.93	190	12.497	.907

Table 11 Math & Physics Correlations

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Math & Physics	190	.585	.000

Table 12 Paired Samples Test

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Math - Physics	3.858	11.363	.824	2.232	5.484	4.680	189	.000

As we notice from the Sig. the null hypothesis is rejected as there is impact from Physics on Math and its increasing.

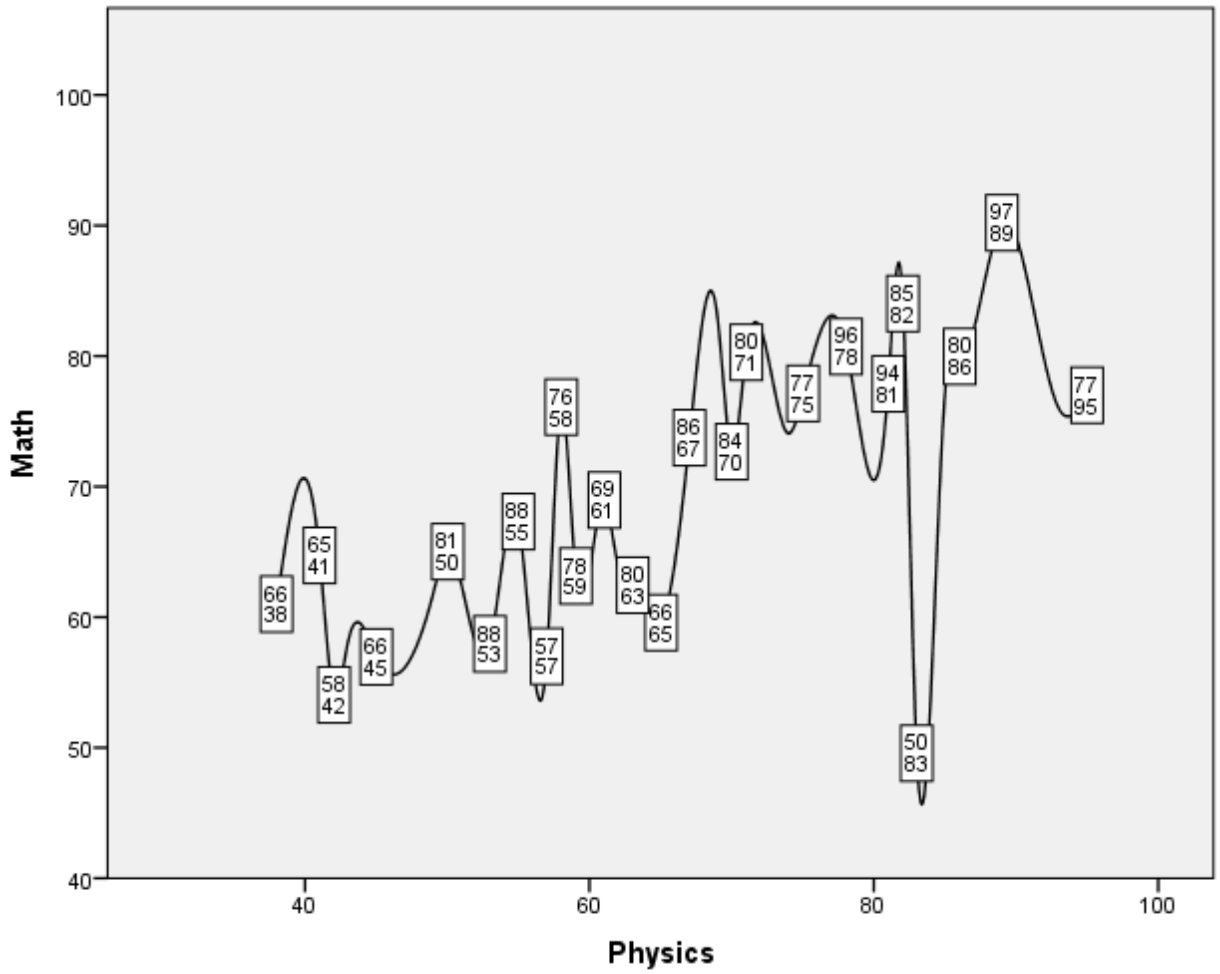

Figure 9 Math Impact on Physics

According to this graph, we could notice that if you get better at Physics subject, the better you will be at Math subject.

- **Independent-Samples T-Test/Mann-Whitney U Test**
- Reading between Genders

o Checking for normality distribution for Reading Subject

Confidence level = 95%, Alpha = 5%

H_0: Reading Grades are normally distributed.

H_a: Reading Grades are NOT normally distributed.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Reading is normal with mean 69.184 and standard deviation 13.64.	One-Sample Kolmogorov-Smirnov Test	.004	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 10 Normality Distribution Check for Writing Subject

As we notice from the Sig. the null hypothesis is rejected so the grades ARE NOT normally distributed and we will attend to use Mann-Whitney U Test which uses the median.

o We want to check if the median of grades of Reading are equal between both genders

H_0: Is Median of Reading grades for Males = Median of Reading grades for Females

H_a: Is Median of Reading grades for Males ≠ Median of Reading grades for Females

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Reading is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	.011	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 11 Mann-Whitney U Test

As we notice from the Sig. the null hypothesis is rejected so there is a statistically significant difference in reading performance across different genders, based on this result, we recommend acknowledging the observed difference in reading distribution between males and females.

- Physics between Genders

  - Checking for normality distribution for Physics Subject

    Confidence level = 95%, Alpha = 5%

    H_0: Physics Grades are normally distributed.

    H_a: Physics Grades are NOT normally distributed.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Physics is normal with mean 65.932 and standard deviation 12.50.	One-Sample Kolmogorov-Smirnov Test	.360	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 12 Normality Distribution Check for Physics Subject

As we notice from the Sig. the null hypothesis is retained so the grades ARE normally distributed and we will attend to use Independent-Samples T-Test which uses the mean and homoscedasticity from Levene's Test for Equality of Variances.

  - We want to check if the average grades of Physics are equal between both genders

    H_0: Is Average of Physics grades for Males = Average of Physics grades for Females

    H_a: Is Average of Physics grades for Males ≠ Average of Physics grades for Females

    H_0: The variance in Physics scores for males is equal to the variance in Physics scores for females

    H_a: The variance in Physics scores for males IS NOT equal to the variance in Physics scores for females

Table 13 Physics between Genders Statistics

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Physics	Male	100	68.22	13.296	1.330
	Female	90	63.39	11.072	1.167

Table 14 Independent Samples Test for Physics between Genders

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Physics	Equal variances assumed	4.680	.032	2.705	188	.007	4.831	1.786	1.308	8.355
	Equal variances not assumed			2.731	186.906	.007	4.831	1.769	1.341	8.321

Based on the Sig. that is highlighted with a Yellow Color, the null hypothesis of homoscedasticity is rejected and therefore the variances aren't equal, and therefore we check that the Sig. of the null hypothesis average Physics grades between Genders which is the highlighted Green Color and the null hypothesis is rejected and the Physics grades between Genders are not equal either, so as a recommendation we should consider targeted interventions, mentorship, or additional resources to help bridge the gap in average grades.

- ## One Way ANOVA

  - We want to check if the average grades of Physics are equal to the 3 SES and the homoscedasticity among variances

    H_0: The variances in Physics scores are EQUAL among the 3 SES High, Middle and Low

    H_a: The variances in Physics scores are NOT EQUAL among the 3 SES High, Middle and Low

    H_0: Is Average of Physics grades are EQUAL among the 3 SES High, Middle and Low

    H_a: Is Average of Physics grades are NOT EQUAL among the 3 SES High, Middle and Low

Table 15 Homoscedasticity Test

Test of Homogeneity of Variances

Physics

Levene Statistic	df1	df2	Sig.
.095	2	187	.910

As we notice from the Sig. the null hypothesis is retained which means the variances are equal

Table 15 One Way ANOVA Physics among SES

ANOVA

Physics

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2271.653	2	1135.826	7.795	.001
Within Groups	27246.458	187	145.703		
Total	29518.111	189			

As we notice from the Sig. the null hypothesis is rejected which means the grades of Physics among the 3 SES are NOT EQUAL, the table down below will show the differences:

Table 16 Physics Statistics among SES

Descriptive Statistics

SES		N	Minimum	Maximum	Mean	Std. Deviation
Low	Physics	46	38	85	61.54	12.346
Middle	Physics	82	42	95	64.93	12.055
High	Physics	62	38	89	70.52	11.884

As we notice there is a tangible differences among the 3 SES, as the best is High Status and the lowest is the Low Status.

- ## Simple Linear Regression

  - We want to check if there is a linear regression equation that make us predict the Physics score according to the Math score. => Physics = a * Math + b??

    H_0: Is a = 0 (there is no linear equation)

    H_a: Is a ≠ 0 (there is linear equation)

    H_0: Is b = 0 (b constant doesn't have impact in the equation)

    H_a: Is b ≠ 0 (b constant has impact in the equation)

Table 17 Simple Linear Regression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	24.936	4.213		5.919	.000
	Math	.587	.059	.585	9.885	.000

a. Dependent Variable: Physics

As we notice from the Sig. that is highlighted as Green Color that the null hypothesis is rejected which means there is a linear equation and from the Yellow highlighted Sig. which is rejected then "b" constant has impact in the equation as we could predict the Physics score according to this equation => Physics = 0.587 * Math + 24.936

So if the Math score is 74 and we want to predict the Physics score, we do this:

Physics = 0.587 * 74 + 24.936 = 68.374

- ## Pearson Correlation

  - We want to check if there is a significant relationship between Writing and Art Scores

    H_0: $\rho = 0$ there is NO significant correlation between Writing and Art Scores

    H_a: $\rho \neq 0$ there is significant correlation between Writing and Art Scores

Table 18 Correlation between Writing and Art

Correlations

		Writing	Art
Writing	Pearson Correlation	1	.457**
	Sig. (2-tailed)		.000
	N	190	190
Art	Pearson Correlation	.457**	1
	Sig. (2-tailed)	.000	
	N	190	190

**. Correlation is significant at the 0.01 level (2-tailed).

We notice from the Sig. that the null hypothesis is rejected which means there is significant correlation between Writing and Art.

As in the Green highlighted color $\rho = 0.457$

Then it's a weak relationship and directly proportional which means both Writing and Art scores tend to move in the same direction.

- ## Chi Square Test of Independence

  - We want to investigate if there is a significant association between Native Language and Reading

    H_0: Native Language is independent of Reading

    H_a: Native Language is NOT independent of Reading

Table 19 NativeLanguage and Reading_Ord

NativeLanguage * Reading_Ord Crosstabulation

Count

		Reading_Ord					Total
		F	D	C	B	A	
NativeLanguage	French	6	7	1	3	0	17
	Chinese	2	3	2	2	0	9
	Other	3	9	0	0	0	12
	English	40	39	24	32	15	150
Total		51	58	27	37	15	188

Table 20 Chi-Square Tests

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	19.538[a]	12	.076
N of Valid Cases	188		

a. 14 cells (70.0%) have expected count less than 5. The minimum expected count is .72.

According to the Sig. the null hypothesis is retained which means there is no association between choose Native Language and Reading, so Native Language is independent of Reading.

- ## One Sample Binomial Test

  - We want to investigate if the proportion of Males = 0.6 and Females = 0.4 or not

    H_0: Is the distribution of gender occurs as 0.6 for Males and 0.4 for Females

    H_a: Is the distribution of gender NOT occurs as 0.6 for Males and 0.4 for Females

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories defined by Gender = Male and Female occur with probabilities 0.6 and 0.4.	One-Sample Binomial Test	.023	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 13 Proportion of Males Binomial Test

As we notice from the Sig. the null hypothesis is rejected which means the proportion of Males doesn't equal 0.6 and for Females 0.4, and we could confirm that by looking at Table 19 down below which is the Percentage of Males = 52.6% and Females = 47.4%.

Table 21 Gender Statistics of Percentage

Gender

	Frequency	Percent	Cumulative Percent
Male	100	52.6	52.6
Female	90	47.4	100.0
Total	190	100.0	

- ## Total Grades Average

  - What is the average of all students in all exams

Table 22 Total Grades Statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
TotalGradesAVG	190	48.00	91.40	69.5484	10.21971
Valid N (listwise)	190				

The average of exams for all students is 69.5%, while our best student achieved average of a 91.4% and the poorest one is 48%.

- ## Chi Square of Goodness Test

  - We want to investigate if the proportion of 3 SES Levels are equal or not

    H_0: Is proportion of 3 SES Levels are EQUAL

    H_a: Is proportion of 3 SES Levels NOT EQUAL

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories of ses occur with equal probabilities.	One-Sample Chi-Square Test	.006	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 14 Proportion of 3 SES Levels

As we notice from the Sig. the null hypothesis is rejected, which means the proportion of 3 SES Levels are not equal which we could confirm it from Table 20 down below which is the Low Level SES is 24.2%, Middle one is 43.2% and High one is 32.6%.

Table 23 SES Statistics of Percentage

SES

	Frequency	Percent	Cumulative Percent
Low	46	24.2	24.2
Middle	82	43.2	67.4
High	62	32.6	100.0
Total	190	100.0	

- ## Risk-(Odds Ratio)

  - What is the probability of a male/female student to be in Private school more than to be in public school?

Table 24 Crosstab Gender and School

Gender * School Crosstabulation

Count

		School		Total
		Government	Private	
Gender	Male	83	17	100
	Female	79	11	90
Total		162	28	190

Table 25 Risk Estimation

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Gender (Male / Female)	.680	.300	1.541
For cohort School = Government	.946	.841	1.063
For cohort School = Private	1.391	.689	2.809
N of Valid Cases	190		

The percentage is 0.68% which is actually low so there are any preferences for the student to go to Private or Public School.

- **G-Power**

- **One Sample T-Test**

Case Value 1: At High School based on expert researches it claims that the average score of Reading is 70% for students, so the school expects from the teacher to lead his students to achieve average Reading score 73% with a standard deviation 4.

- How many students' Reading scores does the school need to test to determine whether their expectation of achieving a 73% average score is reasonable?

  H_0: 70%

  H_a: 73%

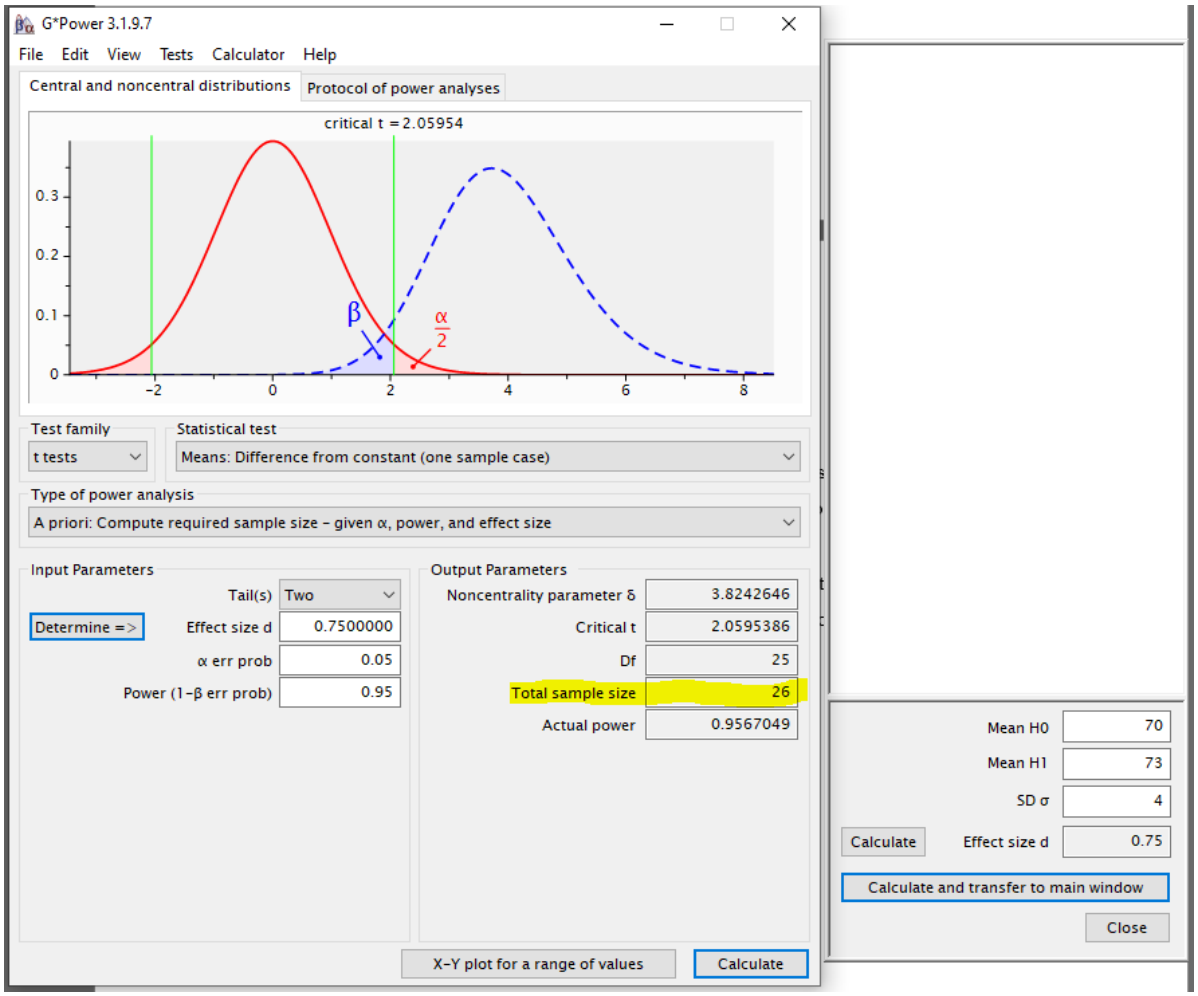

Figure 15 Priori G-Power One Sample T-Test

This means that testing 26 students will allow the school to confidently determine if their teaching efforts are indeed leading to a higher average score than the one claimed by expert researches.

- What if only 14 students' Reading scores are available for testing?

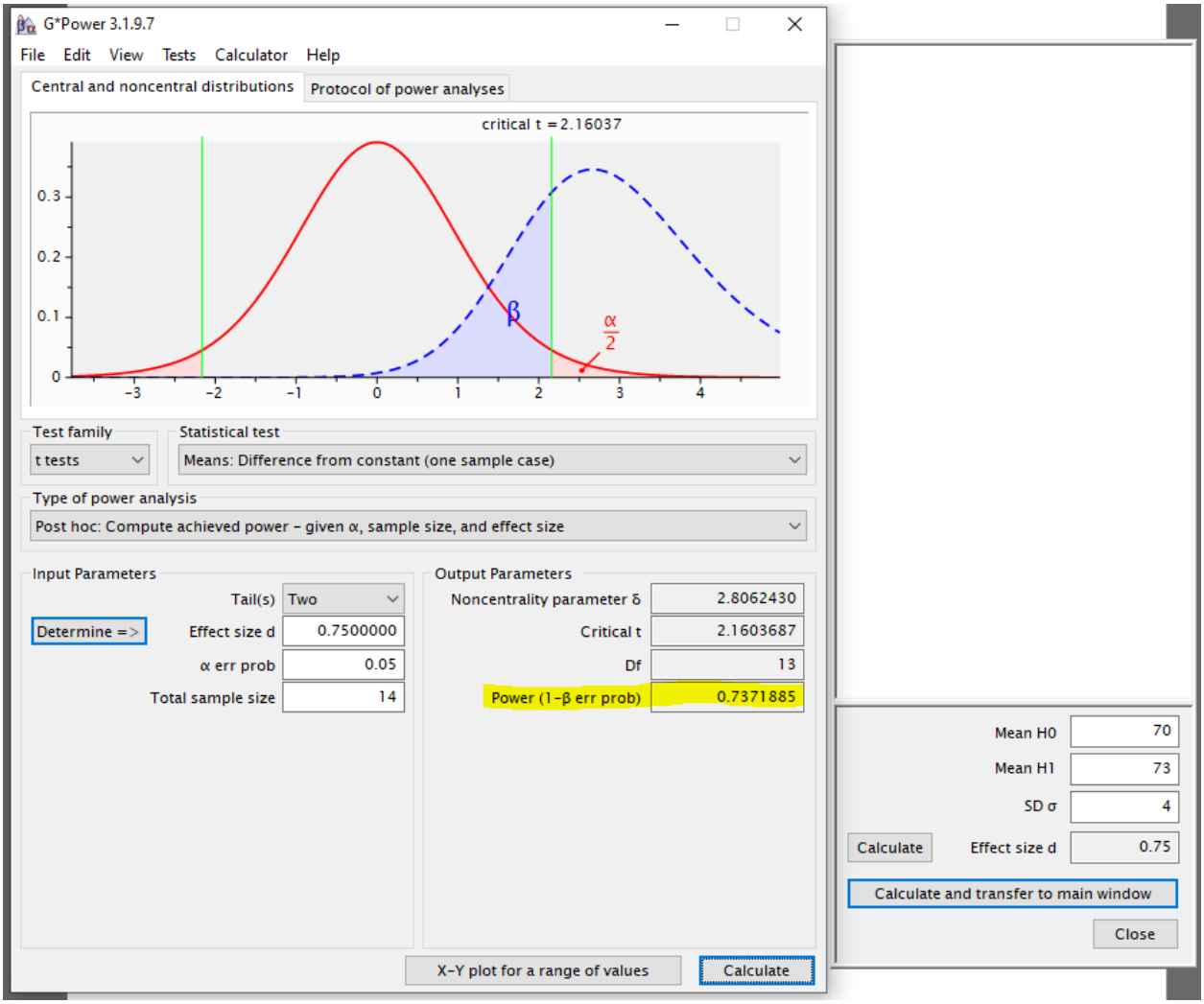

Figure 16 Post hoc G-Power One Sample T-Test

Testing 14 students provides an achieved power of 73.7%, which is acceptable but not ideal. Additional data collection, if feasible, would improve the reliability of the results.

- ## Paired Samples T-Test

  Case Value 2: A high school has implemented a new teaching method aimed at improving students' Physics. Based on expert research, it is claimed that the average Physics score for high school students before the new teaching method is 70% with a standard deviation of 13. The school expects that, after applying the new teaching method, the average Physics score will improve to 79% with a standard deviation of 10.

  - How many students' Physics scores before and after implementing the new teaching method do we need to test to determine if there is a significant improvement in Physics subject?

    H_0: 70%

    H_a: 79%

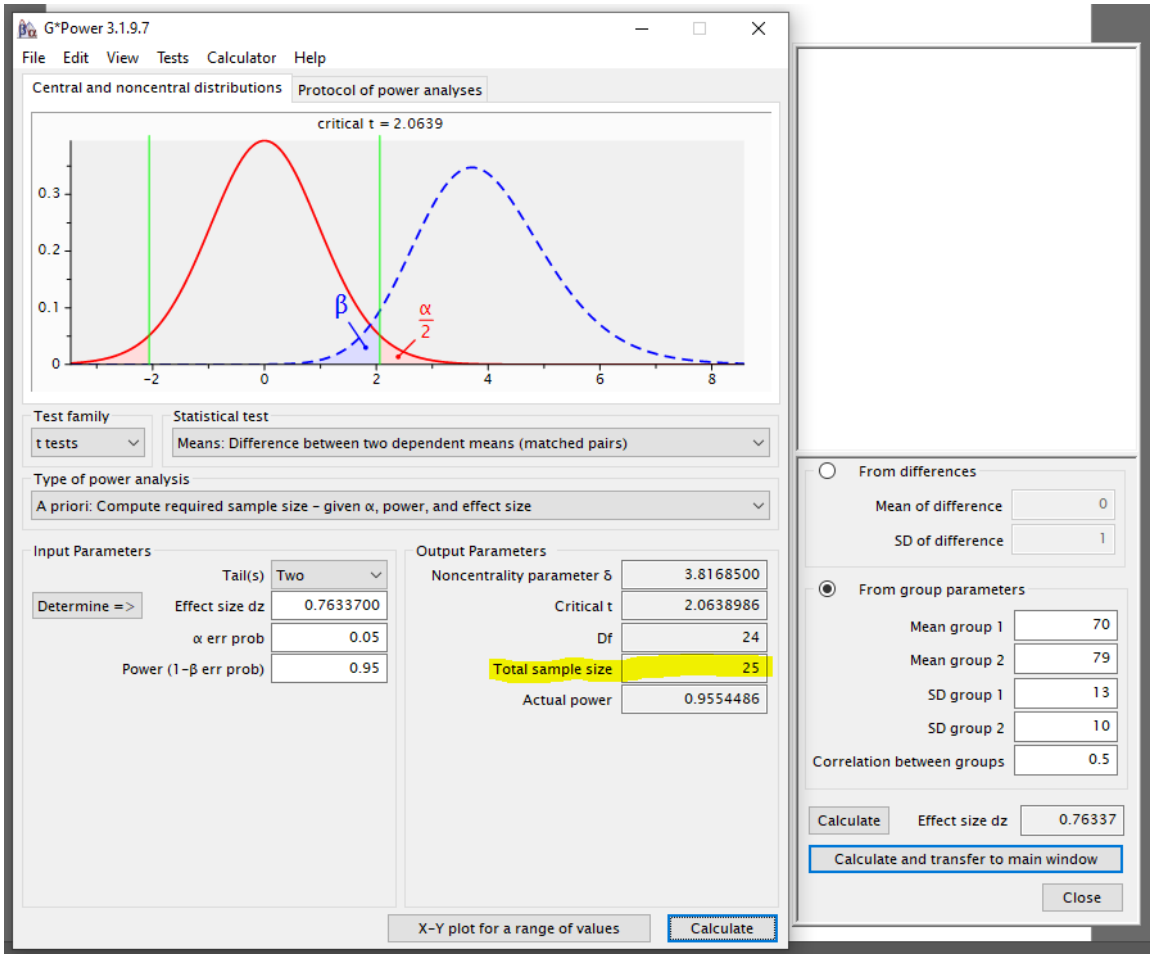

Figure 17 Priori G-Power Paired Samples T-Test

This means you need to test the Physics scores of 25 students both before and after implementing the new teaching method to determine if there is a statistically significant improvement.

- What if only 11 students' Physics scores are available for testing?

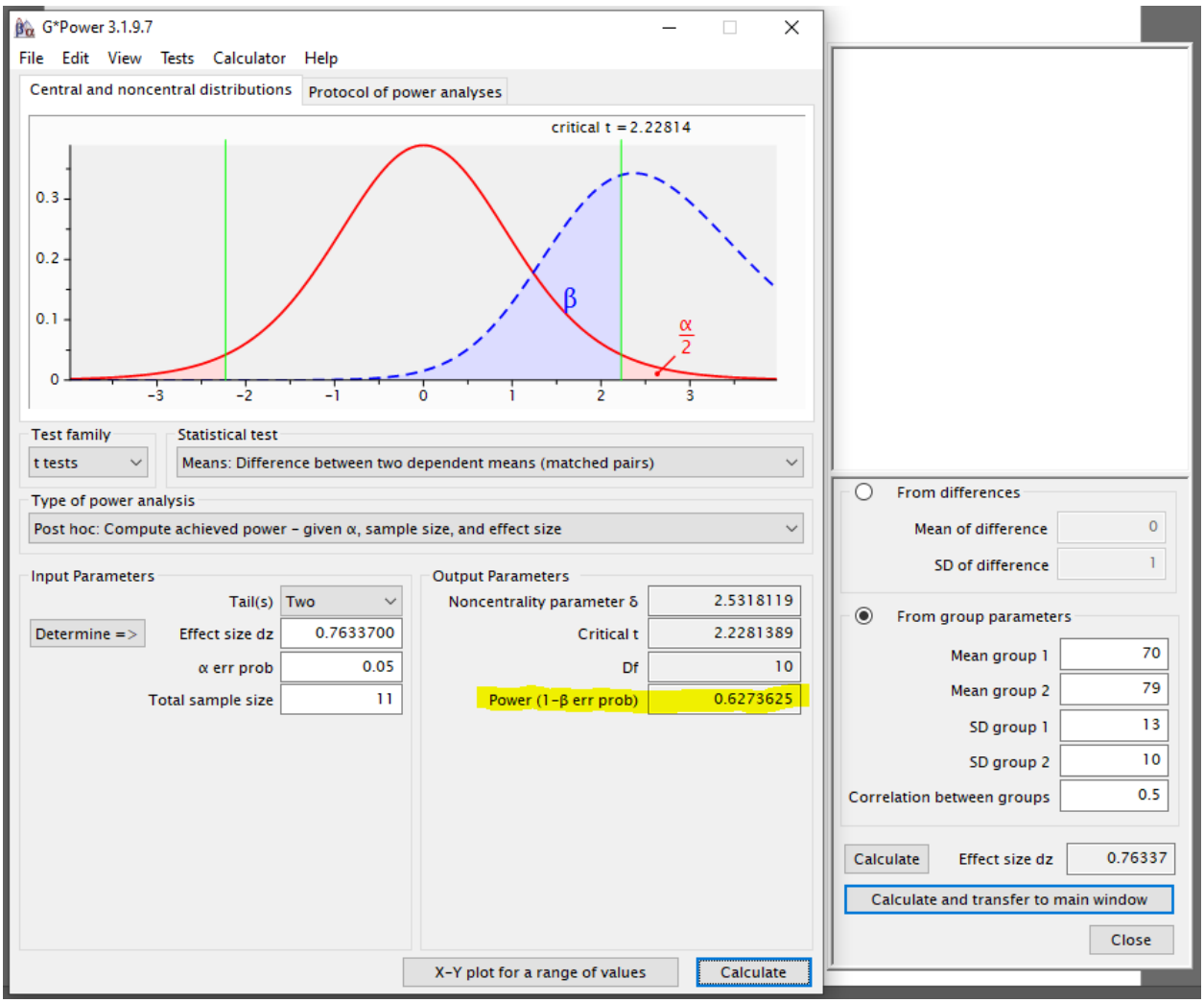

Figure 18 Post hoc G-Power Paired Samples T-Test

The power of 62.7% is lower than the conventional standard of 80%, suggesting that with only 11 students, the study might not have sufficient power to reliably detect a significant improvement in Physics scores.