# A COMPREHENSIVE ANALYSIS OF FACTORS AFFECTING JAMB SCORES


## Investigating Key Variables that Influence Student Performance in JAMB Examinations

A Statistical Analysis of the Impact of Study Habits, Socioeconomic Status, Teacher Quality, Access to Extra Tutorials, Parent Involvement and Education Level, and IT knowledge on JAMB Scores

**By**

**GROUP B, SkillUp Imo, Data Analytics, Cohort 3**

**November, 2024.**

## Executive Summary

### Background

The Joint Admissions and Matriculation Board (JAMB) examination is a critical milestone in Nigerian education, determining tertiary education opportunities. Annually, over 1.5 million candidates sit for the examination, underscoring its significance. However, persistent disparities in academic performance warrant investigations into factors influencing JAMB success.

Nigeria's education sector faces numerous challenges, including inadequate infrastructure, teacher shortages, and socioeconomic disparities. These factors contribute to varying levels of academic preparedness, ultimately affecting JAMB performance.

This study aims to contribute to the ongoing conversation on improving educational outcomes in Nigeria, focusing on JAMB performance.

### Research Questions/Objective

This study addresses the following research questions:

1. Is there a significant positive correlation between study hours per week and JAMB scores?

2. What is the relationship between socioeconomic status (SES) and JAMB scores among students in different school types and locations?

3. How does the quality of teachers impact students' attendance and JAMB scores?

4. Is there a significant difference in JAMB score between students who have access to extra tutorials and those who do not?

5. How do parent involvement and education level interact to influence assignment completion and JAMB scores?

6. What impact does IT knowledge have on study habits and JAMB performance among students without access to learning materials?

The primary objective is to identify significant predictors of JAMB performance, providing actionable insights for stakeholders.

**Scope and Significance**

This study focuses on Nigerian secondary school students, exploring factors influencing JAMB performance. The findings will contribute to the existing body of knowledge, informing policy decisions, educational interventions, and resource allocation.

The study's outcomes will benefit various stakeholders, including:

- Policymakers: Informing evidence-based decisions
- Educators: Enhancing teaching practices
- Students: Improving academic outcomes
- Parents: Supporting informed decision-making

By investigating these research questions, this study aims to shed light on the complex interplay of factors influencing JAMB performance.

**Methodology**

**Data Collection**

This study utilized a secondary dataset obtained from Kaggle, a publicly available data repository. The dataset contained 5024 observations and 17 variables related to JAMB performance.

**Data Description**

The dataset consists of the following variables:

[Student ID, JAMB scores, Study Hours Per Week, Attendance Rate, Teacher Quality, Distance to School, School Type, School Location, Extra Tutorials, Access to Learning Materials, Parent Involvement, IT Knowledge, Age, Gender, Socioeconomic Status (SES), Parent Education Level, Assignments Completed]

**Data Cleaning and Preprocessing**

To ensure data quality, the dataset underwent thorough cleaning and preprocessing. The detailed cleaning process is documented in the excel file accessible via this link:

https://docs.google.com/spreadsheets/d/1vLBYVs0VIlnZRxZ2YyjTdVeqVrPnGusf/edit?usp=drive_link&ouid=107230697908290031929&rtpof=true&sd=true

The cleaned dataset was then used for statistical analysis.

**Statistical Analysis**

To address the research questions, the following statistical tests were employed:

- Correlation analysis (Pearson's r) for Research Question 1
- One-way Analysis of Variance (ANOVA) Question 2
- Two-way ANOVA for Research Question 3
- Two-Tailed T-Test for Research Question 4
- Two-way ANOVA for Research Question 5
- Two-way Analysis of Variance (ANOVA) for Research Question 6

**Software Used**

Data cleaning, preprocessing, and statistical analysis were performed using:

***Excel*** *and* ***Python***

## Analysis Theme and Data Cleaning Process

The dataset was cleaned with Microsoft Excel software. The summary is in tables 1 and 2.

| S/N | Analyses Theme | Columns to be Cleaned |
|---|---|---|
| 1 | Is there a significant positive correlation between study hours per week and JAMB scores? | Study_Hours_Per_Week, JAMB_Score |
| 2 | What is the relationship between socioeconomic status (SES) and JAMB scores among students in different school types and locations? | Socioeconomic_Status, JAMB_Score, School_Type, School_Location |
| 3 | How does the quality of teachers impact students' attendance and JAMB scores? | Teacher_Quality, Attendance_Rate, JAMB_Score |
| 4 | Is there a significant difference in JAMB score between students who have access to extra tutorials and those who do not? | JAMB_Score, Extra_Tutorials |
| 5 | How do parent involvement and education level interact to influence assignment completion and JAMB scores? | Parent_Involvement, Parent_Education_Level, Assignments_Completed, JAMB_Score |
| 6 | What impact does IT knowledge have on study habits and JAMB performance among students without access to educational resources? | IT_Knowledge, Study_Hours_Per_Week, JAMB_Score, Access_To_Learning_Materials |

**Table 1: Analysis Theme**

| Cleaning Effort | |
|---|---|
| The dataset contained; 5024 rows, 17 columns | |
| 24 duplicates were found and removed: 5000 unique rows remained<br>13 relevant columns were selected for cleaning based on the problem statements | |
| **Columns To Be Cleaned** | **Remarks** |
| JAMB Score, Study Hours Per Week, Attendance Rate, Teacher Quality, School Type | There are no empty cells |
| | There are no inconsistent values |
| | There are no inaccurate values |
| School Location | There are no empty cells |
| | Inconsistent value "Rurale" was dictated in cell F13 and corrected to "Rural" |
| | Inaccurate value "village" was dictated in cells F4992, F4993, F4994, F5000 and changed to "Rural"<br>Inaccurate value "City" was dictated in cells F5, F6, F7, F8, F9, F4965, F4970, F4972, F4973, F4975, F4976, F4977, F4978, and changed to "Urban" |
| Extra Tutorials, Access To Learning Materials, Parent Involvement, IT Knowledge | There are no empty cells |
| | There are no inconsistent values |
| | There are no inaccurate values |
| Socioeconomic Status | There are no empty cells |
| | There are no inconsistent values |
| | There are no inaccurate values |
| Parent Education Level | There are no empty cells |
| | There are no inconsistent values |
| | 891 inaccurate values "None" was changed to '"Uneducated" |
| Assignments Completed | There are no empty cells |
| | There are no inconsistent values |
| | There are no inaccurate values |

**Table 2: Column Observations and Cleaning Efforts.**

## Results from Analysis with Pandas

### Q1: Is there a significant positive correlation between study hours per week and JAMB scores?

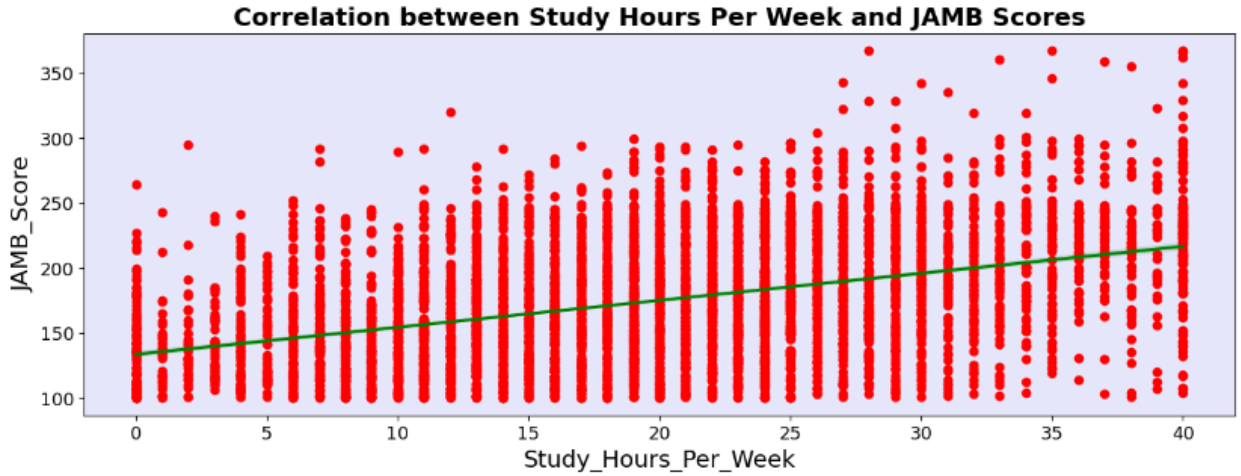
This question aims to investigate the relationship between study time and academic performance, specifically focusing on JAMB scores.

### Choice of Statistical Testing Tool

To address this research question, Pearson's correlation coefficient (r) was employed. This statistical test was chosen for the following reasons:

**1. Research Question Objective:** The research question seeks to identify a relationship between two continuous variables (study hours and JAMB scores), making Pearson's correlation coefficient an appropriate choice.

**2. Data Type:** Both variables are quantitative and continuous, meeting the requirements for Pearson's correlation.



**Pearson Correlation Results Interpretation**

1. Correlation Coefficient (r): 0.4201 (or 0.42)

Positive correlation: As study hours increase, JAMB scores tend to increase.

Moderate correlation: The strength of the relationship is moderate ($0.4 \leq |r| < 0.7$ is considered moderate).

42.01% of the variance in JAMB scores can be explained by study hours

2. p-value: 5.447e-213 (extremely small)

The correlation is statistically significant (p-value << 0.05), indicating a reliable relationship.

Reject the null hypothesis: The correlation is not due to chance as Confidence Level > 99.99%

**Key Findings**

1. Increasing study hours is associated with an increase in JAMB scores.

2. The positive moderate correlation suggests that study hours are an important factor, but not the only factor influencing JAMB scores.

**Conclusion**

There is a statistically significant positive moderate correlation between study hours per week and JAMB scores.

**Q2: What is the relationship between socioeconomic status (SES) and JAMB scores among students in different school types and locations?**

To address this research question, a statistical testing tool was required to compare mean JAMB scores across different SES groups, school types, and locations.

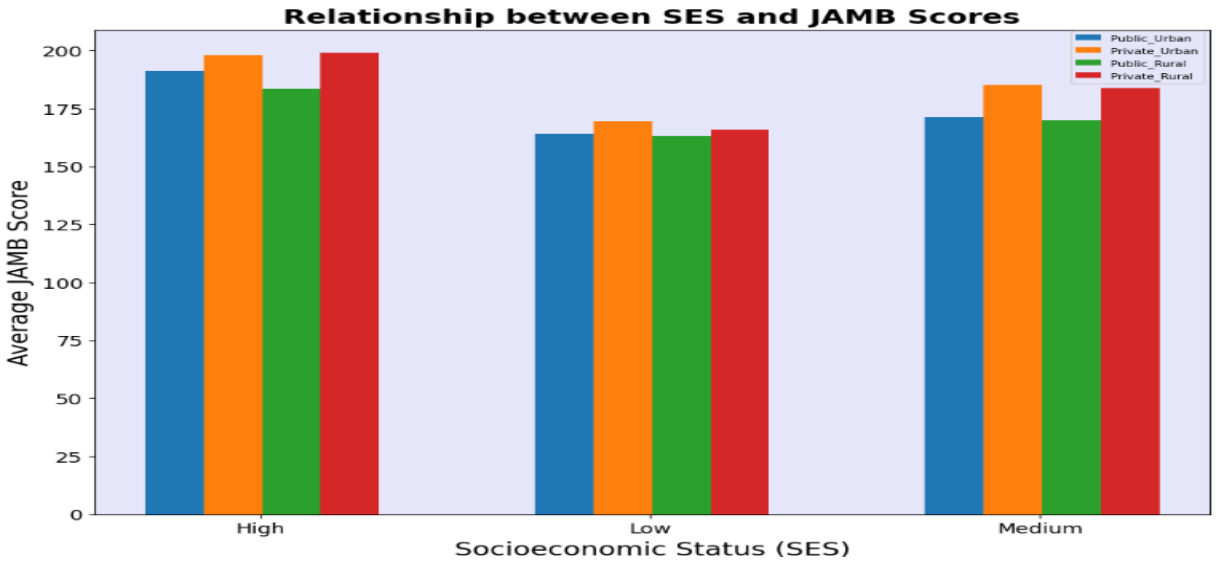
**Choice of Statistical Testing Tool**

For this analysis, a one-way Analysis of Variance (ANOVA) was employed.

**Reasons**

1. Research Question Objective: The question aims to compare mean JAMB scores across different SES groups, school types, and locations.

2. Data Type: The data consists of continuous (JAMB scores) and categorical (SES, school type, location) variables.




**ANOVA Results Interpretation**

**Public Urban Schools:** F-value: 49.55, p-value: 9.35e-22

Very strong, statistically significant effect of SES on JAMB scores for Public schools in Urban areas.

**Public Rural Schools:** F-value: 22.90, p-value: 1.56e-10

A strong, statistically significant effect of SES on JAMB scores for Public schools in Rural areas.

**Private Urban Schools:** F-value: 19.96, p-value: 3.67e-9

A strong, statistically significant effect of SES on JAMB scores for Private schools in Urban areas

**Private Rural Schools:** F-value: 20.07, p-value: 3.94e-9

A strong, statistically significant effect of SES on JAMB scores for Private schools in Rural areas

## Key Findings

1. Socioeconomic Status (SES) Impact: The chart reveals differences in average JAMB scores across High, Low, and Medium SES groups.

2. School Type and Location Effects: Public and Private schools in Urban and Rural areas show varying performance levels.

3. Urban-Rural Disparities: Rural schools tend to have lower average JAMB scores compared to Urban schools.

4. Public-Private School Differences: Private schools generally outperform Public schools across SES groups.

## Conclusion

Socioeconomic status (SES) has a statistically significant impact on JAMB scores in all school types and locations.

**Q3: How does the quality of teachers impact students' attendance and JAMB scores?**

This question investigates the relationship between teacher quality and student outcomes, specifically attendance and JAMB scores.

**Choice of Statistical Testing Tool**

For this analysis, a one-way Analysis of Variance (ANOVA) was employed.

**Reasons:**

1. Research Question Objective: Compare mean attendance and JAMB scores across different teacher quality groups.

2. Data Type: Categorical (teacher quality) and continuous (attendance, JAMB scores) variables.

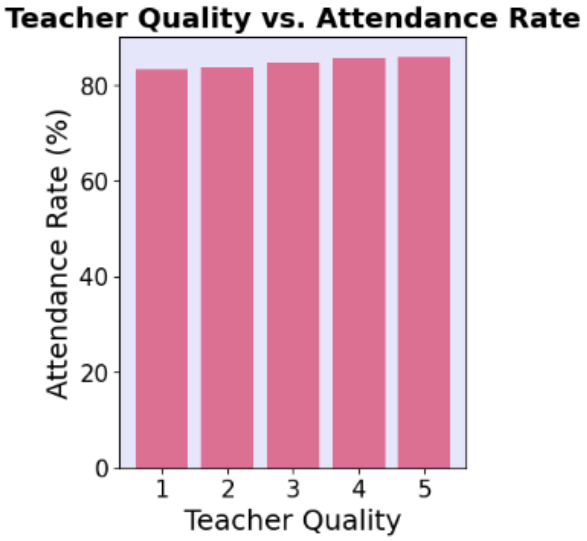
3. Interpretation: ANOVA provides a clear and interpretable measure of the differences between teacher quality groups.

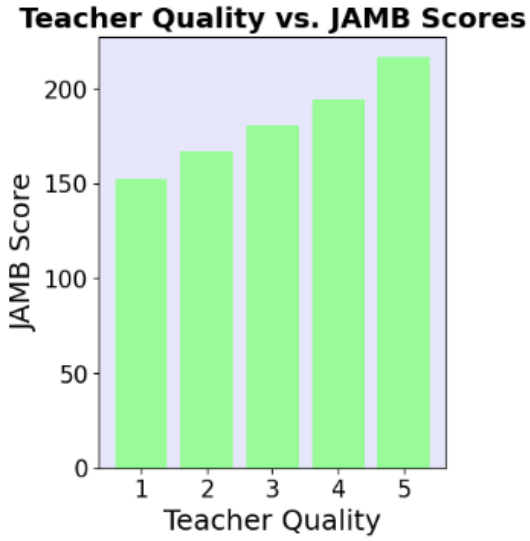
**Why One-way ANOVA?**

- Suitable for comparing mean differences across 3+ groups.

- Controls for variability within groups.

- Provides F-statistic and p-value for significance testing.




**ANOVA Results Interpretation: Teacher Quality vs. Attendance Rate**

**Teacher Quality Effect**

- Statistically Significant: p-value = 7.520418e-07 (less than 0.05)
- F-Statistic: 8.522404 (indicating strong evidence)
- Variation Explained: 3048.972846 units (by Teacher Quality)

**Residual Variation**

- Unexplained Variation: 446752.431954 units
- Residual Mean Square: 89.439926

**Key Findings**

1. Teacher Quality explains a moderate portion of the variation.

2. Other factors contribute to the remaining (residual) variation.

**Conclusion**

Teacher quality has a statistically significant effect on the attendance rate.

## ANOVA Results Interpretation: Jamb Score vs. Teacher Quality

### Teacher Quality Effect

- df (Degrees of Freedom): 4
- sum_sq (Sum of Squares): 1,023,407
- mean_sq (Mean Square): 255,851.86
- F-Statistic: 123.94
- p-value (PR(>F)): 5.157675e-101 (extremely small)

### Residual Variation

- df: 4995
- sum_sq: 10,310,970
- mean_sq: 2064.26

### Key Findings

1. Strong Teacher Quality Effect: F-Statistic = 123.94, p-value ≈ 5.16e-101

2. Significant Variation Explained: Teacher Quality accounts for 1,023,407 units

3. Residual Variation: 10,310,970 units remain unexplained

### Conclusion

Teacher quality has an extreme statistical significant effect on JAMB scores, explaining a substantial portion of the variation.

**Q4: Is there a significant difference in JAMB score between students who have access to extra tutorials and those who do not?**

This question investigates the effect of extra tutorials on JAMB performance.

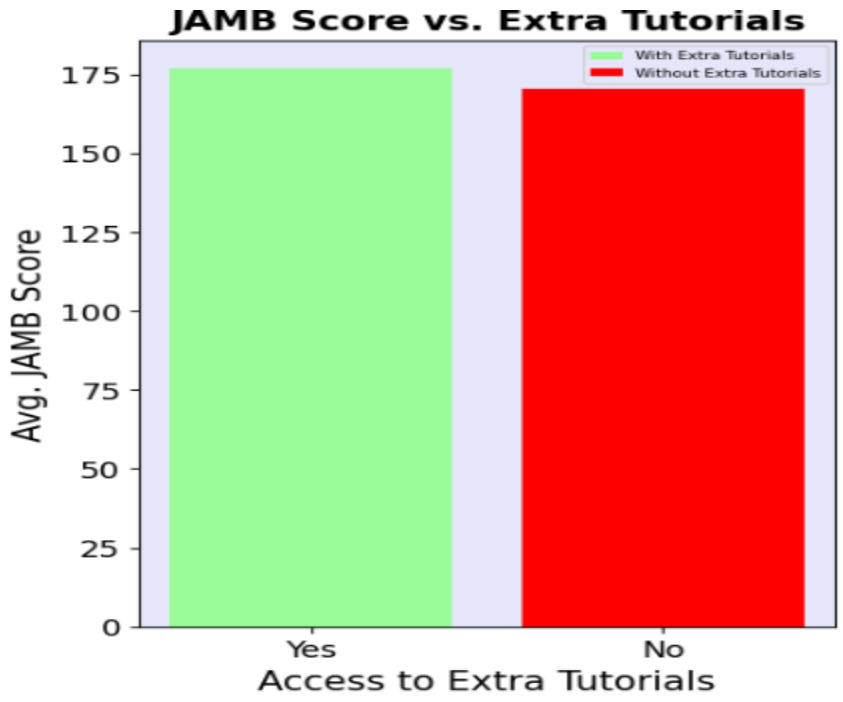
**Choice of Statistical Testing Tool**

For this analysis, a Two-Tailed T-Test was employed because of the following reasons:

1. Research Question Objective: Compare mean JAMB scores between two groups (extra tutorials vs. no extra tutorials).

2. Data Type: Continuous (JAMB scores) and categorical (extra tutorials yes/no) variables.

3. Interpretation: T-Test provides a clear and interpretable measure of the difference between group means.



**T-Test Results Interpretation**

T-statistic of 4.78 indicates strong evidence against the null hypothesis (there is no significant difference)

P-value of 1.82e-06 is extremely small, indicating strong statistical significance.

**Key Findings**

1. Students with access to extra tutorials have a higher mean JAMB score.
2. The difference in JAMB scores is statistically significant (p-value = 1.82e-06).
3. The t-statistic (4.78) indicates a moderate to large effect size.

**Conclusion**

Access to extra tutorials has a statistically significant positive impact on JAMB scores.

**Q5: How do parent involvement and education level interact to influence assignment completion and JAMB scores?**

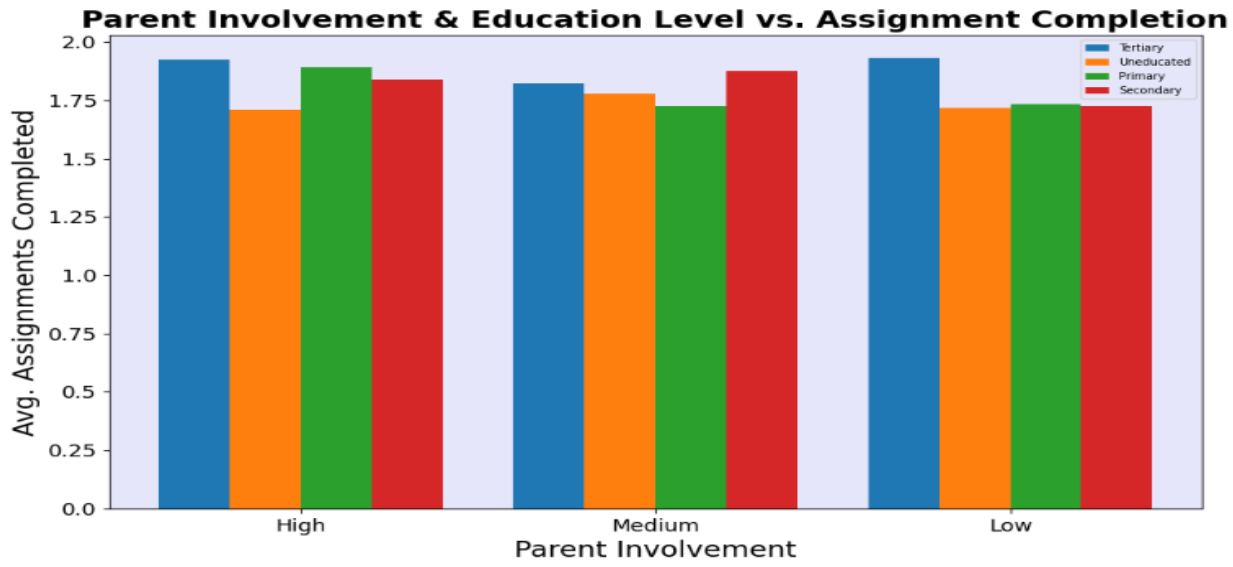
This question explores the combined effects of parent involvement and education level on academic outcomes.

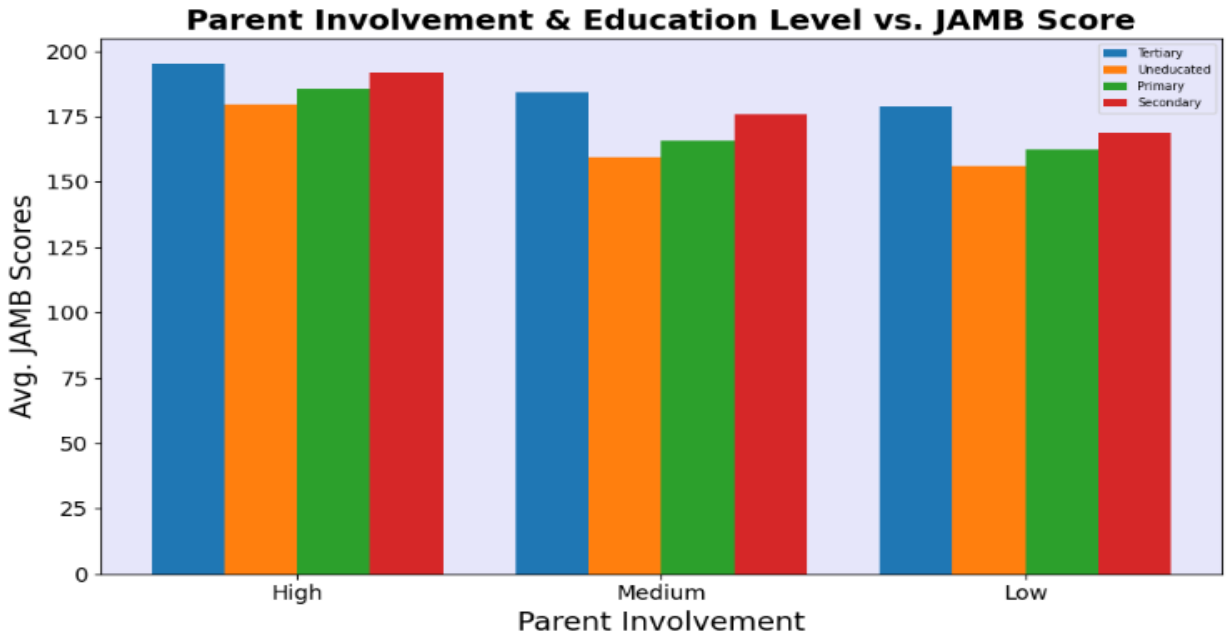
**Choice of Statistical Testing Tool**

For this analysis, a Two-way Analysis of Variance (ANOVA) was employed for the following reasons:

1. Research Question Objective: Examine the interaction between two independent variables (parent involvement and education level) on two dependent variables (assignment completion and JAMB scores).

2. Data Type: Categorical (parent involvement and education level) and continuous (assignment completion and JAMB scores) variables.




## Anova Results Interpretation

## Main Effects:

1. Parent Involvement (df=2):

- sum_sq: 395,266.3 (or 335,498.3 in the second table)
- mean_sq: 197,633.1

- F: 2.016 (or 78.277 in the second table)
- p-value: 0.133 (or 3.37e-34 in the second table)

2. Parent Education Level (df=3):

- sum_sq: 139,628.04 (or 297,189.5 in the second table)
- mean_sq: 46,542.7
- F: 4.749
- p-value: 0.002611 (or 1.85e-29 in the second table)

**Interaction Effect:**

1. Parent Involvement:Parent Education Level (df=6):

- sum_sq: 13,493.3 (or 12,282.51 in the second table)
- mean_sq: 2,248.9
- F: 2.295
- p-value: 0.032509 (or 0.454059 in the second table)

**Residual Variation**

- df: 4988
- sum_sq: 4888.685429 (or 1.068941e+07 in the second table)
- mean_sq: 980.089

**Key Findings**

**Assignment Completion:**

1. Parent Education Level has a significant main effect (p = 0.002611) on assignment completion.

2. Parent Involvement has no significant main effect ($p = 0.133232$) on assignment completion.

3. There is a significant interaction between Parent Involvement and Parent Education Level ($p = 0.032509$) on assignment completion.

**JAMB Scores:**

1. Both Parent Involvement ($p < 0.000001$) and Parent Education Level ($p < 0.000001$) have significant main effects on JAMB scores.

2. Parent Involvement has a stronger effect ($F = 78.276826$) on JAMB scores compared to Parent Education Level ($F = 46.225862$).

3. There is no significant interaction between Parent Involvement and Parent Education Level ($p = 0.454059$) on JAMB scores.

**Conclusion**

- Parent education level has a statistically significant effect on assignment completion, while parent involvement has no direct effect.
- However, the interaction between parent involvement and education level significantly influences assignment completion.
- Parent involvement and education level are both strong predictors of JAMB scores, with parent involvement having a slightly stronger effect.

**Q6: What impact does IT knowledge have on study habits and JAMB performance among students without access to learning materials?**

This question investigates the relationship between IT knowledge and academic outcomes for disadvantaged students.

**Choice of Statistical Testing Tool**

For this analysis, a Two-way Analysis of Variance (ANOVA) was employed for the following reasons:

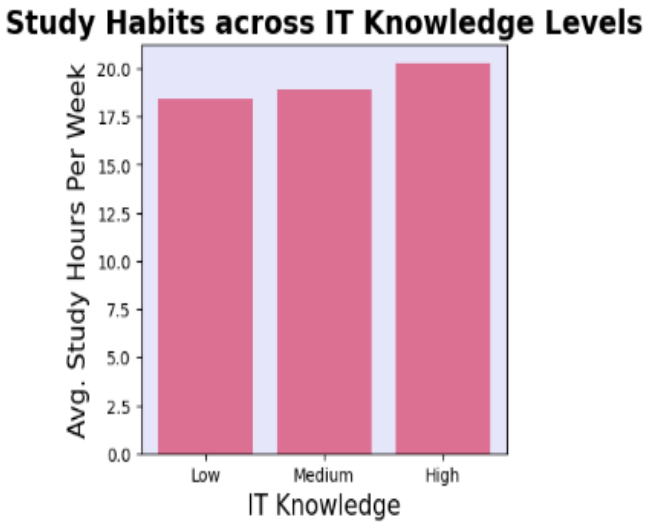
1. Research Question Objective: Examine the interaction between IT knowledge and study habits on JAMB performance.

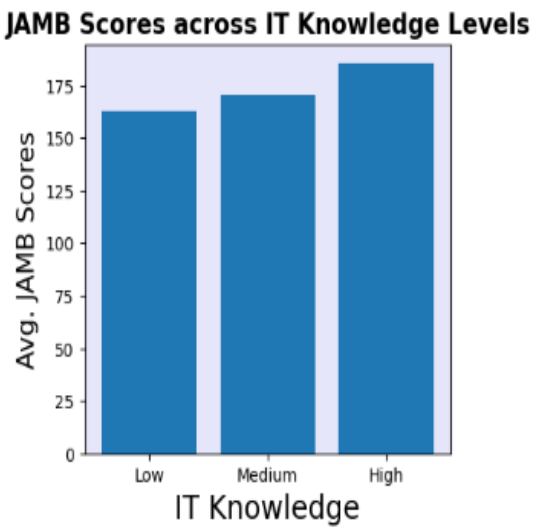
2. Data Type: Categorical (IT knowledge, study habits) and continuous (JAMB performance) variables.

3. Effect: Examines main effects of IT knowledge and study habits.

4. Variability Check: Accounts for variability within groups.

5. Nature and Significance: Provides F-statistic and p-value for significance testing.




**ANOVA Results: JAMB Scores across IT Knowledge Level**

**IT Knowledge Effect**

- df (Degrees of Freedom): 2
- sum_sq (Sum of Squares): 119,640.7

- mean_sq (Mean Square): 59,820.35
- F-Statistic: 27.79
- p-value (PR(>F)): 1.346364e-12

**Residual Variation**

- df: 1651, sum_sq: 3,553,768, mean_sq: 2,152.49

**Key Findings**

1. There is a highly significant main effect of IT Knowledge on JAMB Scores (p = 1.346364e-12).
2. The F-statistic (27.791179) indicates that IT Knowledge is a strong predictor of JAMB Scores.
3. IT Knowledge explains 119,640.7 units of variation

**Conclusion**

- IT knowledge has a strong positive impact on JAMB scores

## Summary of findings and Recommendations

This study aimed to identify and analyze key factors influencing students' performance in the Joint Admissions and Matriculation Board (JAMB) examination, a pivotal determinant for tertiary education in Nigeria. Through statistical analysis of variables such as study habits, socioeconomic status (SES), teacher quality, access to extra tutorials, parent involvement and education level, and IT knowledge, several critical insights were uncovered.

The findings confirmed that there are multiple significant factors impacting JAMB scores. Study hours exhibited a moderate positive correlation with JAMB performance, suggesting that students who dedicate more time to studying are likely to achieve better outcomes. Socioeconomic status (SES) was also found to significantly affect JAMB scores across different school types and locations, with urban students generally outperforming their rural counterparts. Additionally, private schools demonstrated higher average scores compared to public schools, emphasizing the role of resources and school type in academic success. Teacher quality emerged as another influential factor, with highly qualified teachers positively impacting both student attendance and JAMB performance. Moreover, access to extra tutorials was found to have a statistically significant positive effect on student scores. Parent involvement and education level also played a crucial role in shaping both assignment completion rates and JAMB outcomes, although the impact was stronger for parental involvement. Finally, IT knowledge was identified as a key predictor of JAMB success, particularly for students who lacked access to traditional learning materials.

**Recommendations**

Based on the findings of this study, several recommendations can be made to improve student performance in the JAMB examination and overall educational outcomes:

1. **Enhanced Access to Learning Resources**: Efforts should be made to ensure that students, especially those from disadvantaged backgrounds, have access to additional learning materials, including textbooks, online resources, and extra tutorials. This could be achieved through partnerships between government agencies, private institutions, and non-governmental organizations (NGOs) to provide affordable or free resources.

2. **Increase Teacher Training and Quality**: Teacher quality was found to have a significant impact on both attendance and academic performance. Therefore, regular professional development programs for teachers should be implemented, with a focus on teaching methodologies that promote engagement, critical thinking, and exam preparation.

3. **Focus on Socioeconomic Disparities**: Policymakers should address the socioeconomic disparities that affect student performance. This can include providing financial support or scholarships for students from lower-income households, improving school infrastructure, and ensuring equitable distribution of educational resources across rural and urban areas.

4. **Promote Parent Involvement**: Parents, particularly those with higher educational levels, have been shown to positively influence students' academic performance. Schools should encourage greater parental involvement in students' academic lives by organizing workshops and seminars that help parents support their

children's education. Furthermore, parent-teacher collaborations could be strengthened to ensure consistent support for students both at home and school.

5. **Integration of IT in Education**: Given the positive impact of IT knowledge on JAMB scores, schools should integrate more technology into the curriculum, providing students with the tools and skills necessary to excel in an increasingly digital world. Special emphasis should be placed on providing IT training to students in underfunded schools.

6. **Improvement in Study Habits**: Since study hours were found to be positively correlated with JAMB scores, initiatives should be developed to foster better study habits among students. Educational institutions can organize study skill workshops to teach students time management, effective study techniques, and stress management, which can contribute to improved academic performance.

**Limitations**

While this study provides valuable insights into the factors affecting JAMB scores, there are some limitations that should be acknowledged:

1. **Data Limitations**: The study relied on secondary data obtained from Kaggle, which may not fully represent the diversity of all JAMB candidates across Nigeria. The dataset was limited in scope, and certain potentially influential factors, such as mental health, access to physical infrastructure (e.g., libraries, internet), and exam-related anxiety, were not included in the analysis.

2. **Causal Inferences**: This study employed statistical analyses such as correlation and ANOVA, which can reveal associations between variables but do not establish direct causal relationships. For example, while a correlation between study hours

and JAMB scores was found, it is impossible to definitively conclude that increased study hours directly cause higher JAMB scores without further experimental or longitudinal research.

3. **Generalizability**: The study focused on secondary school students in Nigeria, and the results may not be directly applicable to students in other countries or regions with different educational systems, socioeconomic contexts, or cultural practices.

4. **Unaccounted Variables**: While this study considered several key variables, there may be other unmeasured factors influencing JAMB scores. For instance, individual student motivation, peer influences, and specific teaching practices at individual schools were not captured in the analysis.

5. **Cross-sectional Nature**: The data used in this study is cross-sectional, meaning it represents a snapshot of student performance at a single point in time. Longitudinal studies tracking changes over time would provide a more comprehensive understanding of how these factors evolve and influence student outcomes over the course of their education.

Despite these limitations, the study provides a comprehensive analysis of the key factors influencing JAMB scores and offers actionable recommendations for improving student performance in Nigeria. Future research could address these limitations by incorporating additional variables, using longitudinal data, and exploring causal relationships more deeply.

# Appendix

In [19]:
```
# Importing the necessary modules for data analysis and visualization
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings as w
from scipy.stats import pearsonr
w.filterwarnings('ignore')
```

In [20]:
```
# Loading JAMB Exam Results data from excel file
Jamb_df = pd.read_excel("Jamb_Exam_Results.xlsx",
                        sheet_name = 'Cleaned_Data')
```

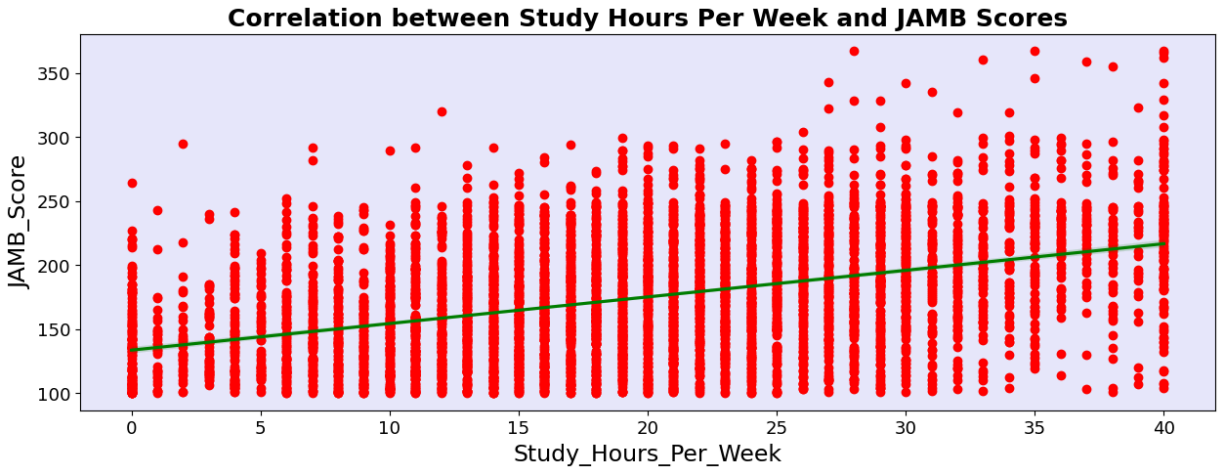
In [21]:
```
# Displaying the JAMB Exam Results DataFrame
Jamb_df.head(3)
```

Out[21]:

| | JAMB_Score | Study_Hours_Per_Week | Attendance_Rate | Teacher_Quality | School_Type | School_Location | Extra_Tutor |
|---|---|---|---|---|---|---|---|
| **0** | 192 | 22 | 78 | 4 | Public | Urban | |
| **1** | 207 | 14 | 88 | 4 | Public | Rural | |
| **2** | 182 | 29 | 87 | 2 | Public | Rural | |

## Q1: Is there a significant positive correlation between study hours per week and JAMB scores?

In [22]:
```
# Importing pearsonr for calculating correlation between two variables

# Creating scatter plot
plt.figure(figsize=(15,5))
plt.scatter(Jamb_df['Study_Hours_Per_Week'], Jamb_df['JAMB_Score'],
            color = 'red')
plt.xlabel('Study Hours Per Week', fontsize=17)
plt.xticks(fontsize=13)
plt.ylabel('JAMB Scores', fontsize=17)
plt.yticks(fontsize=13)
plt.title('Correlation between Study Hours Per Week and JAMB Scores',
          fontsize=18, fontweight = "bold")

# plotting regression line between Study Hours Per Week and JAMB Scores
sns.regplot(x = 'Study_Hours_Per_Week',
            y = 'JAMB_Score', data = Jamb_df, color = 'green',
            scatter = False)

# Setting background color
plt.rcParams['axes.facecolor'] = 'lavender'


# Displaying the plot
plt.show()
```




## Observation

Increasing study hours is associated with an increase in JAMB scores.

In [23]:
```
# Calculating Pearson correlation coefficient and p-value between study hours per week and JAMB sc
correlation, p_value = pearsonr(Jamb_df['Study_Hours_Per_Week'],
                                Jamb_df['JAMB_Score'])
print(f"Correlation coefficient: {correlation}")
print(f"p-value: {p_value}")
```

```
Correlation coefficient: 0.42006122096711646
p-value: 5.4470122437216606e-213
```

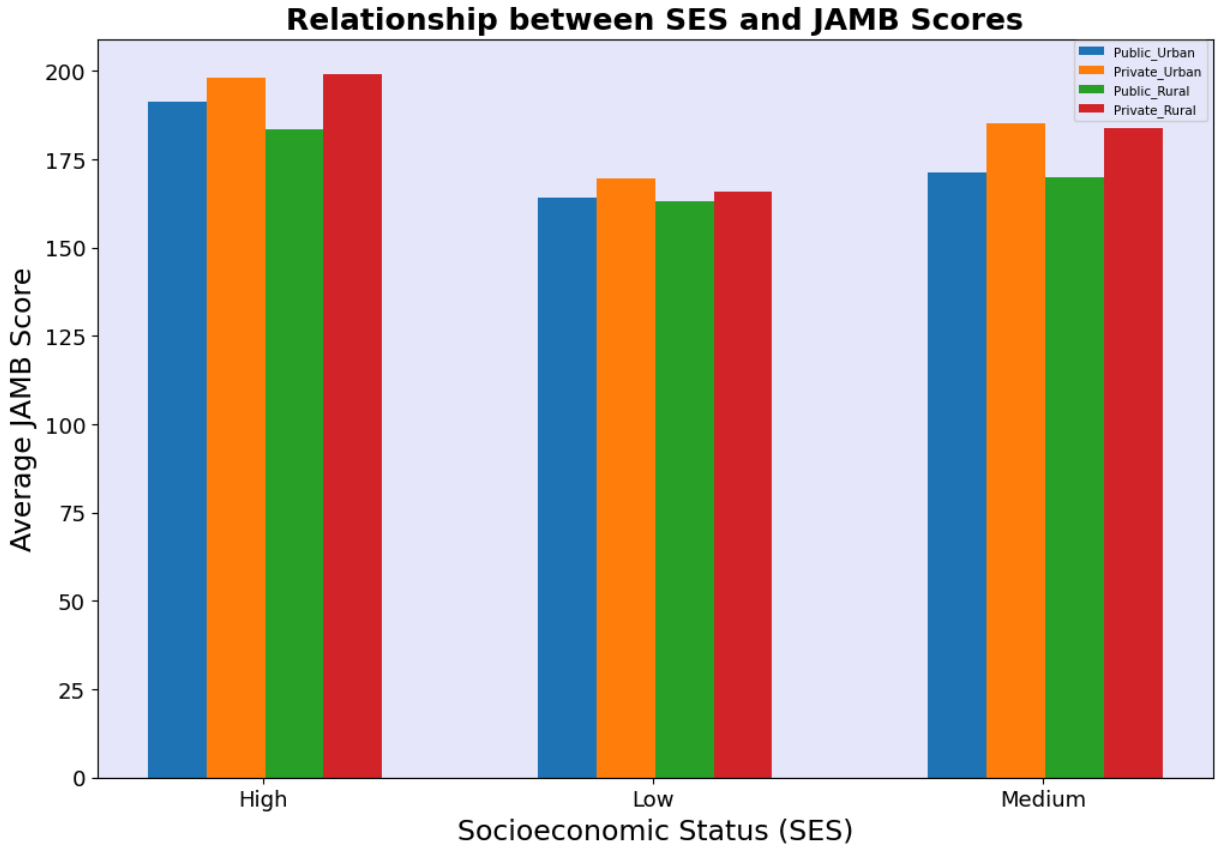
### Observation

Correlation coefficient: 0.4201, p-value: 5.447e-213 indicates that there is a statistically significant positive moderate correlation between study hours per week and JAMB scores.

## Q2: What is the relationship between socioeconomic status (SES) and JAMB scores among students in different school types and locations?

In [24]:
```
# Importing required modules for numerical computations and statistical analysis
import numpy as np
from scipy.stats import f_oneway

# Calculating mean JAMB scores by socioeconomic status for different school types and locations
df_public_urban = Jamb_df[(Jamb_df['School_Type'] == 'Public') \
                          & (Jamb_df['School_Location'] == 'Urban')]\
                    .groupby('Socioeconomic_Status')['JAMB_Score'].mean()
df_private_urban = Jamb_df[(Jamb_df['School_Type'] == 'Private')\
                           & (Jamb_df['School_Location'] == 'Urban')]\
                    .groupby('Socioeconomic_Status')['JAMB_Score'].mean()
df_public_rural = Jamb_df[(Jamb_df['School_Type'] == 'Public')\
                          & (Jamb_df['School_Location'] == 'Rural')]\
                    .groupby('Socioeconomic_Status')['JAMB_Score'].mean()
df_private_rural = Jamb_df[(Jamb_df['School_Type'] == 'Private') \
                           & (Jamb_df['School_Location'] == 'Rural')]\
                    .groupby('Socioeconomic_Status')['JAMB_Score'].mean()

# Configuring bar chart parameters
x = np.arange(3)  # x-coordinates for bars
width = 0.15  # bar width
width_2 = 0.3  # offset for fourth bar group

# Creating figure and axis
fig, ax = plt.subplots(figsize=(12, 8))  # 12x8 inch figure

# Ploting bar groups for each School Type
rects1 = ax.bar(x - width, df_public_urban.values,
                width, label='Public_Urban')  # Public urban bars
rects2 = ax.bar(x, df_private_urban.values, width,
                label='Private_Urban')  # Private urban bars
```

```
rects3 = ax.bar(x + width, df_public_rural.values,
                width, label='Public_Rural')  # Public rural bars
rects4 = ax.bar(x + width_2, df_private_rural.values,
                width, label='Private_Rural')  # Private rural bars

# Customizing chart appearance
ax.set_ylabel('Average JAMB Score', fontsize = 17)  # y-axis label
ax.set_xlabel('Socioeconomic Status (SES)', fontsize = 17)  # x-axis label
ax.set_title('Relationship between SES and JAMB Scores',
             fontsize = 18, fontweight='bold')  # title

# Configuring x-axis and y-axis ticks and labels
ax.set_xticks(x+0.07)  # tick locations
ax.set_xticklabels(['High', 'Low', 'Medium'], fontsize = 13)  # tick labels
ax.set_yticklabels([0,25,50,75,100,125,150,175,200], fontsize = 13)

# Adding legend
ax.legend(fontsize=7.5, bbox_to_anchor=(0.87, 0.88))  # legend properties

# Setting background color
plt.rcParams['axes.facecolor'] = 'lavender'

# Displaying the plot
plt.show()
```




## Observation

The relationship between socioeconomic status (SES) and JAMB scores among students in different school types and locations varies

In [25]:
```
# Looping through each School Type and Location combination
for school in Jamb_df['School_Type'].unique():
    for location in Jamb_df['School_Location'].unique():
        public_urban = Jamb_df[(Jamb_df['School_Type'] == school)\
                               & (Jamb_df['School_Location'] == location)]\
                                ['JAMB_Score']
        low_ses = public_urban[Jamb_df['Socioeconomic_Status'] == 'Low']
        medium_ses = public_urban[Jamb_df['Socioeconomic_Status']\
                                  == 'Medium']
        high_ses = public_urban[Jamb_df['Socioeconomic_Status'] == 'High']
```

```
        # Performing one-way ANOVA to compare JAMB scores across socioeconomic statuses
        f_value, p_value = f_oneway(low_ses, medium_ses, high_ses)

        # Display ANOVA results
        print(f"School Type: {school}, Location: {location},\
                  F-value: {f_value}, p-value: {p_value}")
```

```
School Type: Public, Location: Urban,                F-value: 49.55206114709313, p-value: 9.354
983056174925e-22
School Type: Public, Location: Rural,                F-value: 22.895171775087764, p-value: 1.56
44177385173075e-10
School Type: Private, Location: Urban,                F-value: 19.955593953374287, p-value: 3.6
65636033139722e-09
School Type: Private, Location: Rural,                F-value: 20.066824546194194, p-value: 3.9
43983281889071e-09
```

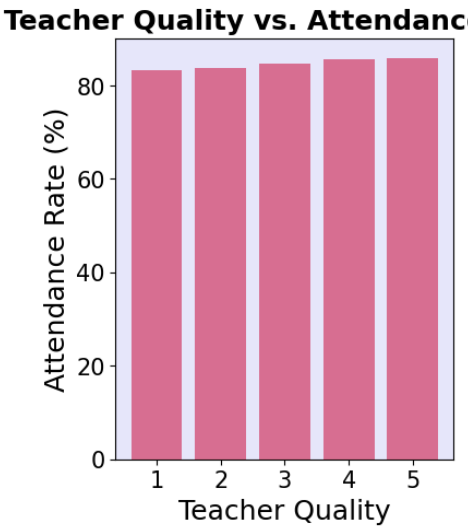
### Observation

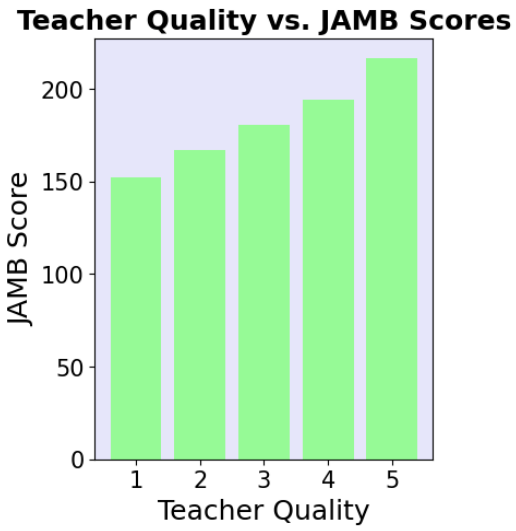
All F-values correspond to extremely small p-values, indicating statistical significance.

## Q3: How does the quality of teachers impact students' attendance and JAMB scores?

In [26]:

```
# Importing necessary libraries for statistical analysis and visualization
import statsmodels.api as sm
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm

# Creating figure with two subplots
plt.figure(figsize=(12,5))

# Adjust spacing between subplots horizontally
plt.subplots_adjust(wspace = 1)

# Subplot 1: Teacher Quality vs. Attendance Rates
plt.subplot(1,2,1)
Jamb_df1 = Jamb_df.groupby('Teacher_Quality')\
                ['Attendance_Rate'].mean()
                # Calculate mean attendance rate by teacher quality
plt.bar(Jamb_df1.index, Jamb_df1.values, color = "palevioletred")
                # Bar chart
plt.title('Teacher Quality vs. Attendance Rate',
          fontsize=18, fontweight='bold')
plt.xlabel('Teacher Quality', fontsize = 18)
plt.ylabel('Attendance Rate (%)', fontsize = 18)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)


# Subplot 2: Teacher Quality vs. JAMB Scores
plt.subplot(1,2,2)
Jamb_df2 = Jamb_df.groupby('Teacher_Quality')['JAMB_Score'].mean()
                    # Calculate mean JAMB score by teacher quality
plt.bar(Jamb_df2.index, Jamb_df2.values, color = "palegreen")
                    # Bar chart
plt.title('Teacher Quality vs. JAMB Scores',
          fontsize=18, fontweight='bold')
plt.xlabel('Teacher Quality', fontsize = 18)
plt.ylabel('JAMB Score', fontsize = 18)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)

# Exporting Teacher Quality vs. JAMB Scores to image file
plt.savefig("Teacher Quality vs. JAMB Scores.png")

# Setting background color
plt.rcParams['axes.facecolor'] = 'lavender'

# Displaying the plot
plt.show()
```




### Observation

Teacher quality has more effect on students' JAMB Scores than their attendance

In [27]:
```
# Using OLS model to run Regression analysis: Attendance Rate vs. Teacher Quality
model = ols('Attendance_Rate ~ C(Teacher_Quality)', data=Jamb_df).fit()

# Running analysis of variance (ANOVA) on the OLS model
anova_table1 = anova_lm(model)

# Printing ANOVA table
anova_table1
```

Out[27]:

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| **C(Teacher_Quality)** | 4.0 | 3048.972846 | 762.243211 | 8.522404 | 7.520418e-07 |
| **Residual** | 4995.0 | 446752.431954 | 89.439926 | NaN | NaN |

### Observation

F-Statistic: 8.522404, p-value = 7.520418e-07 (less than 0.05) indicates that teacher quality has a moderate effect on students' attendance)

In [28]:
```
# Using OLS model to run Regression analysis: Jamb Score vs. Teacher Quality
model = ols('JAMB_Score ~ C(Teacher_Quality)', data=Jamb_df).fit()

# Running analysis of variance (ANOVA) on the OLS model
anova_table2 = anova_lm(model)

# Printing ANOVA table
anova_table2
```

Out[28]:

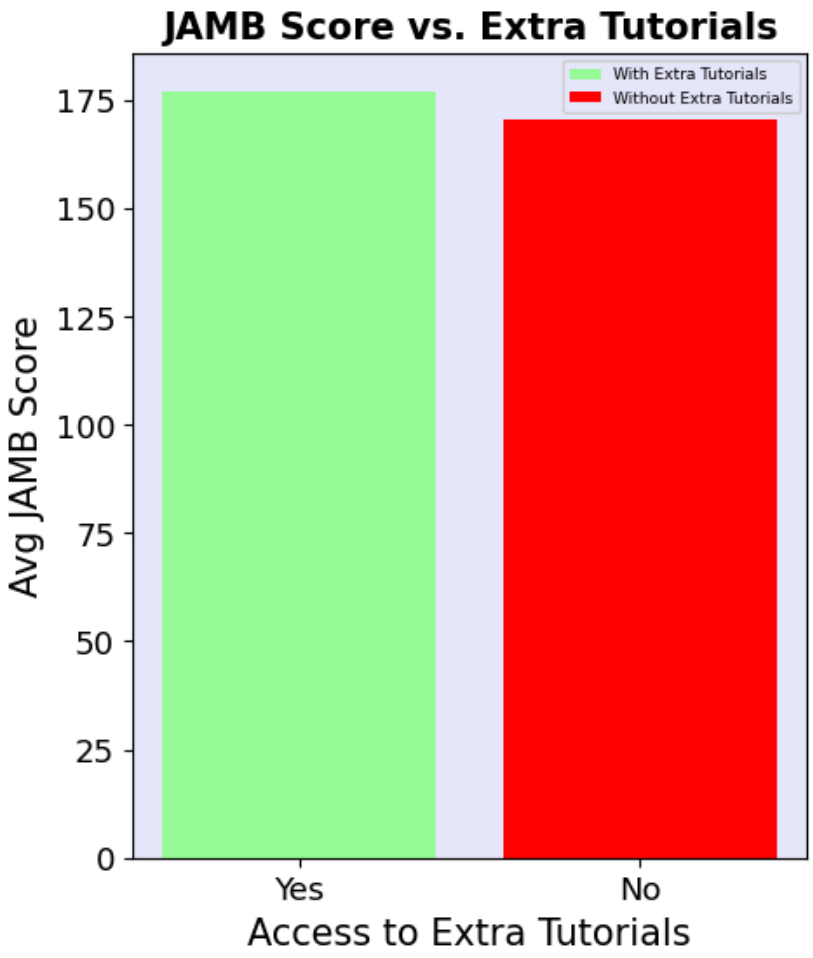
| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| **C(Teacher_Quality)** | 4.0 | 1.023407e+06 | 255851.855612 | 123.943727 | 5.157675e-101 |
| **Residual** | 4995.0 | 1.031097e+07 | 2064.258209 | NaN | NaN |

### Observation

The F-statistic (123.94) and highly significant p-value (5.16e-101) indicate a strong, statistically significant relationship between teacher quality and JAMB scores

## Q4: Is there a significant difference in JAMB score between students who have access to extra tutorials and those who do not?

In [29]:
```
# Importing t-test function for comparing means of two independent samples
from scipy.stats import ttest_ind
from matplotlib.patches import Patch

# Creating figure
plt.figure(figsize=(5,6))

# Visualizing mean JAMB scores by access to extra tutorials
plt.bar(['Yes', 'No'],
[Jamb_df.loc[Jamb_df['Extra_Tutorials'] == 'Yes', 'JAMB_Score'].mean(),
Jamb_df.loc[Jamb_df['Extra_Tutorials'] == 'No', 'JAMB_Score'].mean()],
        color = ['palegreen','red'])

# Setting plot title and labels
plt.title('JAMB Score vs. Extra Tutorials', fontsize=15, fontweight='bold')
plt.xlabel('Access to Extra Tutorials', fontsize=15)
plt.ylabel('Avg JAMB Score', fontsize=15)
plt.xticks(fontsize=13)
plt.yticks(fontsize=13)

# Defining patches for legend
patch1 = Patch(facecolor='palegreen', label='With Extra Tutorials')
patch2 = Patch(facecolor='red', label='Without Extra Tutorials')

# Adding legend
plt.legend(handles=[patch1, patch2], loc='upper right', fontsize=6.5)

# Exporting JAMB Score vs. Extra Tutorials to image file
plt.savefig("JAMB Score vs. Extra Tutorials.png")

# Setting background color
plt.rcParams['axes.facecolor'] = 'lavender'


# Displaying the plot
plt.show()
```



### Observation

Students with access to extra tutorials have a higher mean JAMB score.

In [30]:
```
t_stat, p_val = ttest_ind(Jamb_df.loc[Jamb_df['Extra_Tutorials']\
                                      == 'Yes', 'JAMB_Score'],
                          Jamb_df.loc[Jamb_df['Extra_Tutorials']\
                                      == 'No', 'JAMB_Score'])

print(f'T-statistic: {t_stat}, p-value: {p_val}')
```

```
T-statistic: 4.778116229149867, p-value: 1.820122571245901e-06
```

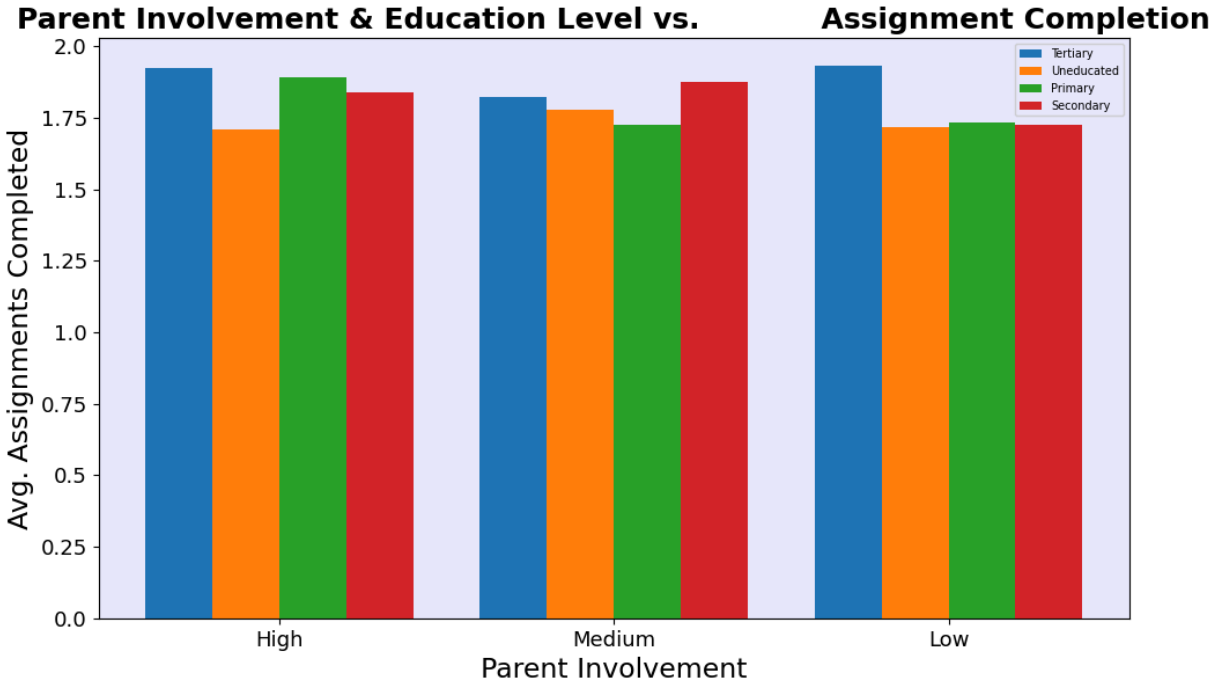
### Observation

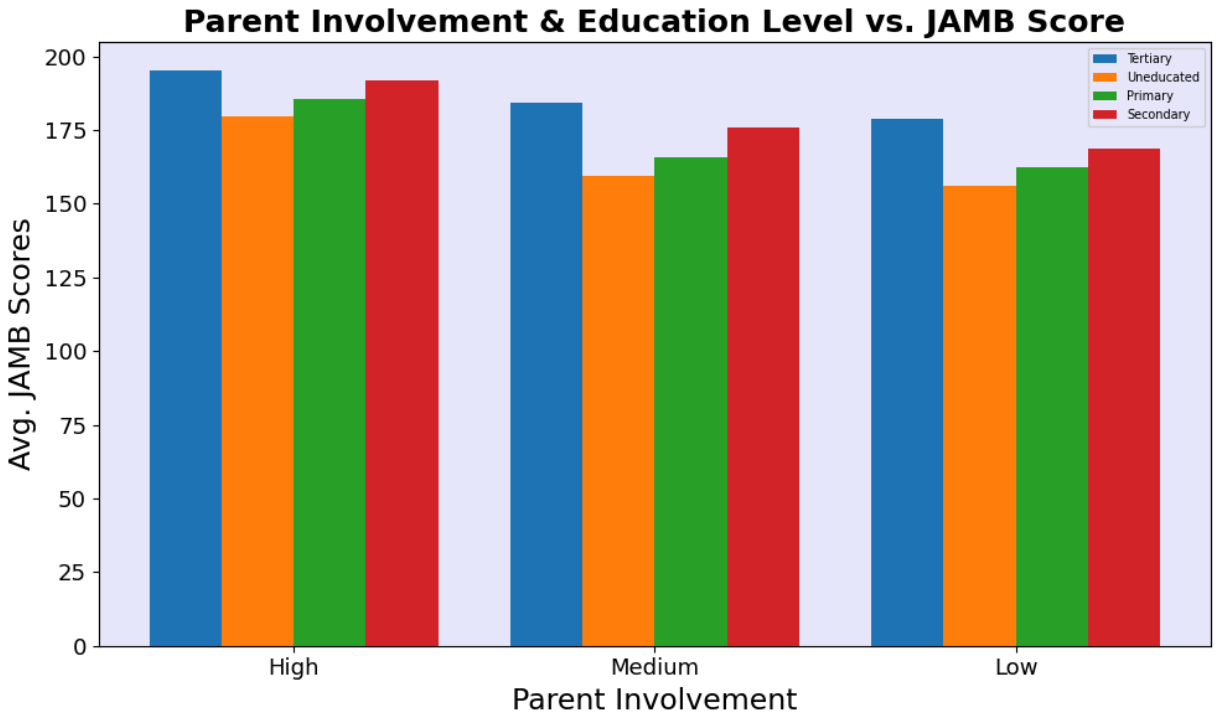
T-statistic of 4.78, p-value of 1.82e-06 indicates that access to extra tutorials has a statistically significant positive impact on JAMB scores.

## Q5: How do parent involvement and education level interact to influence assignment completion and JAMB scores?

In [38]:
```
# Creating clustered bar charts to visualize the relationship
#between Parent Involvement,
# Parent Education Level, and academic performance
#(Assignments Completed and JAMB Score)

# Assignment Completion Chart
# Setting x-axis values for unique Parent Involvement levels
x = np.arange(len(Jamb_df['Parent_Involvement'].unique()))

# Setting bar width
width = 0.2
# Setting the right margin for aligning bar
right = 0.3

# Creating figure and axis
fig, ax = plt.subplots(figsize=(11.5, 6.5))

# Iterating over unique Parent Education Levels
for i, level in enumerate(Jamb_df['Parent_Education_Level'].unique()):
    # Plotting bar chart for each education level
    ax.bar([j + width * i for j in x],
           [Jamb_df.loc[(Jamb_df['Parent_Education_Level'] == level) \
                        & (Jamb_df['Parent_Involvement'] == inv)
                        , 'Assignments_Completed'].mean()
            for inv in Jamb_df['Parent_Involvement'].unique()],
           width, label=level)

# Setting chart labels and title
ax.set_xlabel('Parent Involvement', fontsize=17)
ax.set_ylabel('Avg. Assignments Completed', fontsize=17)
ax.set_title('Parent Involvement & Education Level vs.\
            Assignment Completion', fontsize=18, fontweight='bold')

# Setting x-axis and y-axix ticks and labels
ax.set_xticks(x+right)
ax.set_xticklabels(Jamb_df['Parent_Involvement'].unique(), fontsize=13)
ax.set_yticklabels([0.00,0.25,0.50,0.75,1.00,1.25,1.50,1.75,2.00],
                   fontsize=13)

# Adding Legend
ax.legend(fontsize=7)

# JAMB Score Chart
# Repeating same process for JAMB Score
x = np.arange(len(Jamb_df['Parent_Involvement'].unique()))
width = 0.2
fig, ax = plt.subplots(figsize=(10, 6))

for i, level in enumerate(Jamb_df['Parent_Education_Level'].unique()):
```

```
    ax.bar([j + width * i for j in x],
           [Jamb_df.loc[(Jamb_df['Parent_Education_Level'] == level) \
                        & (Jamb_df['Parent_Involvement'] == inv),
                        'JAMB_Score'].mean()
            for inv in Jamb_df['Parent_Involvement'].unique()],
           width, label=level)

# Setting chart labels and title
ax.set_xlabel('Parent Involvement', fontsize=17)
ax.set_ylabel('Avg. JAMB Scores', fontsize=17)
ax.set_title('Parent Involvement & Education Level vs. JAMB Score',
             fontsize=18, fontweight='bold')

# Setting x-axis and y-axis ticks and labels
ax.set_xticks(x+right)
ax.set_xticklabels(Jamb_df['Parent_Involvement'].unique(), fontsize=13)
ax.set_yticklabels([0,25,50,75,100,125,150,175,200], fontsize=13)

# Adding legend
ax.legend(fontsize=7)

# Exporting Parent Involvement & Education Level vs. JAMB Score & Assignment Completion to image f
plt.savefig("Parent Involvement & Education Level vs.\
            JAMB Score & Assignment Completion.png")

# 1. Adjusting plot layout
plt.tight_layout()

# Setting background color
plt.rcParams['axes.facecolor'] = 'lavender'

# Displaying the plot
plt.show()
```






## Observation

1. The interaction between parent involvement and education level influences assignment completion.
2. Parent involvement and education level are both strong predictors of JAMB scores, with parent involvement having a slightly stronger effect.

In [39]:
```
# Two-way ANOVA for Assignment Completion
model = ols('Assignments_Completed ~ C(Parent_Involvement) \
            * C(Parent_Education_Level)', data=Jamb_df).fit()
anova_table1 = anova_lm(model)
anova_table1

# Two-way ANOVA for JAMB Scores
model = ols('JAMB_Score ~ C(Parent_Involvement) *\
        C(Parent_Education_Level)', data=Jamb_df).fit()
anova_table2 = anova_lm(model)
```

In [40]:
```
anova_table1
```

Out[40]:

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| **C(Parent_Involvement)** | 2.0 | 3.952663 | 1.976331 | 2.016481 | 0.133232 |
| **C(Parent_Education_Level)** | 3.0 | 13.962804 | 4.654268 | 4.748820 | 0.002611 |
| **C(Parent_Involvement):C(Parent_Education_Level)** | 6.0 | 13.493304 | 2.248884 | 2.294571 | 0.032509 |
| **Residual** | 4988.0 | 4888.685429 | 0.980089 | NaN | NaN |

In [41]:
```
anova_table2
```

Out[41]:

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| **C(Parent_Involvement)** | 2.0 | 3.354983e+05 | 167749.166356 | 78.276826 | 3.368120e-34 |
| **C(Parent_Education_Level)** | 3.0 | 2.971895e+05 | 99063.159771 | 46.225862 | 1.852279e-29 |
| **C(Parent_Involvement):C(Parent_Education_Level)** | 6.0 | 1.228251e+04 | 2047.084838 | 0.955232 | 4.540587e-01 |
| **Residual** | 4988.0 | 1.068941e+07 | 2143.024630 | NaN | NaN |

## Observation

### Assignment Completion

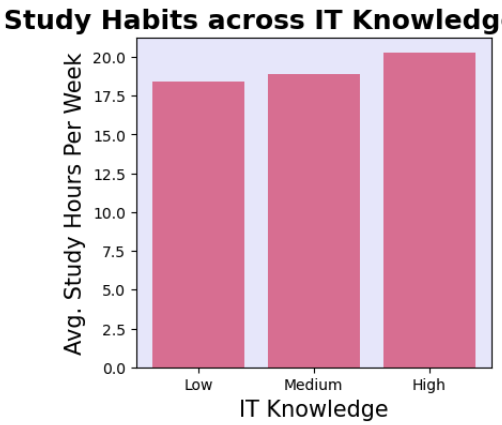
1. "Parent Involvement and Parent Education Level significantly interact to influence assignment completion (p = 0.033)."

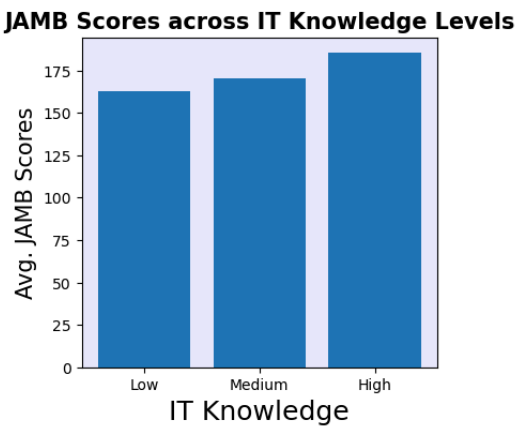
**JAMB Scores**

1. "No significant interaction exists between Parent Involvement and Parent Education Level on JAMB scores (p = 0.454)."

## Q6: What impact does IT knowledge have on study habits and JAMB performance among students without access to learning materials?

In [42]:
```
#Creating a dataframe of students without access to learning materials
Jamb_df1 = Jamb_df[Jamb_df['Access_To_Learning_Materials'] == 'No']


# Creating a figure with specified size
plt.figure(figsize=(12,4))

# Adjust spacing between subplots horizontally
plt.subplots_adjust(wspace = 0.8)

# Dividing the figure into 2 subplots in 1 row and 2 columns
plt.subplot(1,2,1)

# Plotting bar chart of mean Study Hours Per Week by IT Knowledge level
plt.bar(Jamb_df1['IT_Knowledge'].unique(),
        [Jamb_df1.loc[Jamb_df1['IT_Knowledge'] == level,
                      'Study_Hours_Per_Week'].mean()
         for level in Jamb_df1['IT_Knowledge'].unique()],
                        color = 'palevioletred')
plt.xlabel('IT Knowledge', fontsize=15)  # Set x-axis label
plt.ylabel('Avg. Study Hours Per Week', fontsize=15)  # Set y-axis label
plt.title('Study Habits across IT Knowledge Levels',
          fontsize=18, fontweight='bold')  # Set plot title

# Dividing the figure into 2 subplots in 1 row and 2 columns
#(second subplot)
plt.subplot(1,2,2)

# Plotting bar chart of mean JAMB Scores by IT Knowledge level
plt.bar(Jamb_df1['IT_Knowledge'].unique(),
        [Jamb_df1.loc[Jamb_df1['IT_Knowledge'] == level,
                      'JAMB_Score'].mean()
         for level in Jamb_df1['IT_Knowledge'].unique()])
plt.xlabel('IT Knowledge', fontsize=18)  # Set x-axis label
plt.ylabel('Avg. JAMB Scores', fontsize=15)  # Set y-axis label
plt.title('JAMB Scores across IT Knowledge Levels',
          fontsize=15, fontweight='bold')  # Set plot title

# Setting background color
plt.rcParams['axes.facecolor'] = 'lavender'

# Displaying the plots
plt.show()
```




## Observation

Study Hours per week and JAMB scores tend to increase with a higher IT knowledge

In [36]:
```
# Two-way ANOVA for Study Hours Per Week
model = ols('Study_Hours_Per_Week ~ C(IT_Knowledge)', data=Jamb_df1).fit()
anova_table = anova_lm(model)
anova_table
```

Out[36]:

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| **C(IT_Knowledge)** | 2.0 | 811.564608 | 405.782304 | 4.312185 | 0.013556 |
| **Residual** | 1651.0 | 155361.264896 | 94.101311 | NaN | NaN |

## Observation

F-Statistic: 4.31, p-value = 0.013556 indicates IT knowledge has a moderately significant effect on Study hours per week

In [37]:
```
# Two-way ANOVA for JAMB Scores
model = ols('JAMB_Score ~ C(IT_Knowledge)', data=Jamb_df1).fit()
anova_table = anova_lm(model)
anova_table
```

Out[37]:

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| **C(IT_Knowledge)** | 2.0 | 1.196407e+05 | 59820.347062 | 27.791179 | 1.346364e-12 |
| **Residual** | 1651.0 | 3.553768e+06 | 2152.494033 | NaN | NaN |

## Observation

The F-statistic =27.791179, p-value = 1.346364e-12 indicates that IT Knowledge has a significantly strong effect on JAMB Scores.