Analysis of Mean Test Score

Anoushka Khayar, Gayatri Pant, and Subha Raut

Youngstown State University

STAT 6949/4849: Design of Experiment

Dr. Lucy Kerns

ABSTRACT

The main purpose of this paper is to determine the interaction between race and test preparation and their interaction on the mean score of the students. If there exists significant interaction, we simply use simple effects. Simple effects are comparisons of the cell means across levels of one factor for some or all levels of the other factor (Montgomery, 2021). If there doesn't exist interaction then, we use main effects. The main effects are comparisons of marginal means for one of the factors (Montgomery, 2021). Test preparation and Race are considered as the two fixed factors in our project. Test preparation consists of two levels (completed and none) whereas race consists of five levels (A, B, C, D, and E). There were three scores in the dataset: Reading scores, writing scores, and Math scores. We created a new variable and calculated the mean of three scores called "Mean" which represents the mean score of the students. The mean scores of the students were considered as the response variable.

A 2-factorial ANOVA test was performed to determine the interaction between race and test preparation and its effect on the mean score of students. We further conducted a multiple comparison test to meet the final conclusions. We analyzed the data using R statistical software throughout the procedure. The study found that there was no significant interaction between race and test on the mean score of the students (p-value =0.775>0.05). However, we can say that the race (p-value =1.06e-7<0.05) and the test(p-value=2e-16<0.05) individually have a significant effect on the mean score of the students. Thus, from Tukey's post hoc test, we concluded that there is a significant difference in mean groups between race and mean groups between tests.

Keyword: Race, Test preparation, interaction, main effect

Contents

1. Introduction

The results of a student's standardized high school tests have long been a deciding element in their college entrance. Higher test scores boost a student's chances of being accepted to school and of being eligible for merit-based scholarships. GPAs in high school are frequently viewed as a significant component, whereas test scores are viewed as equivalent. Because all students are graded on the same task, standardized test scores are thought to be more reliable. Various test preparation courses have been provided to high school students over the years in order to assist them to improve their test scores. Despite the growth of test preparation classes, the question of whether they have a positive impact on a student's average score continues to be debated. Test preparation has led to a small boost in student's score. Research has demonstrated a favorable increase for students who took the ACT test a second time with test preparation, particularly for the ACT test. Various studies have also revealed mixed results depending on the demographic being tested for the variation in test average score.

The purpose of this study is to see how race and test preparation affect the mean score obtained by high school students. We investigated whether the completion of test preparation courses is likely to affect the differences in average scores among different race groups to do this, we examined students' average scores in Reading, Writing, and Math scores they obtained in their Standardized tests. We also examined whether certain race groups are more likely than others to obtain higher mean scores or whether these results are consistent across all race groups. For this study, we assumed that all the students completed the test preparation course.

1.1 Hypothesis

1.1.1 We wish to answer the following questions.

a. Do the factors, "Test Prep" and "Race" interact?
b. What effects do "Test Prep" and "Race" have on mean test scores?
c. Does completing a test prep result in a higher mean score?

2.Methodology

2.1 Data Description

The dataset was obtained from the Kaggle website which consists of data related to students from high school and their scores in Reading, Writing and Mathematics. There are total of 8 columns and 1000 rows in the original dataset. The columns are Gender, Race/Ethnicity, Parental education, Lunch, Test preparation, Math score, reading score, Writing score. The data didn't consist of any null values. We added a column "Mean" with the average scores from the 3 subject score column. The variables of interest in this study are race/ethnicity and test preparation. We will be performing two factor ANOVA method to determine the impact of race/ethnicity and test preparation on student's mean scores

3.Descriptive analysis

3.1 Descriptive analysis for test

Table 3.1: Descriptive analysis for Test Preparation

Test Preparation	Mean	Median
Completed	72.67	73.50
None	65.04	65.33

From the Table (3.1) we observe the mean test score of students who completed the test preparation is 72.67. The median is 73.50. The students who didn't complete the test preparation had a mean of 65.04 and median of 65.33. Group of Students who completed the Test Preparation scored higher than the group of students who didn't complete any test preparation. The boxplot (3.1) also indicates that the student who completed the test preparation has a higher mean test score. From the boxplot, we can observe that we have outliers in the dataset.

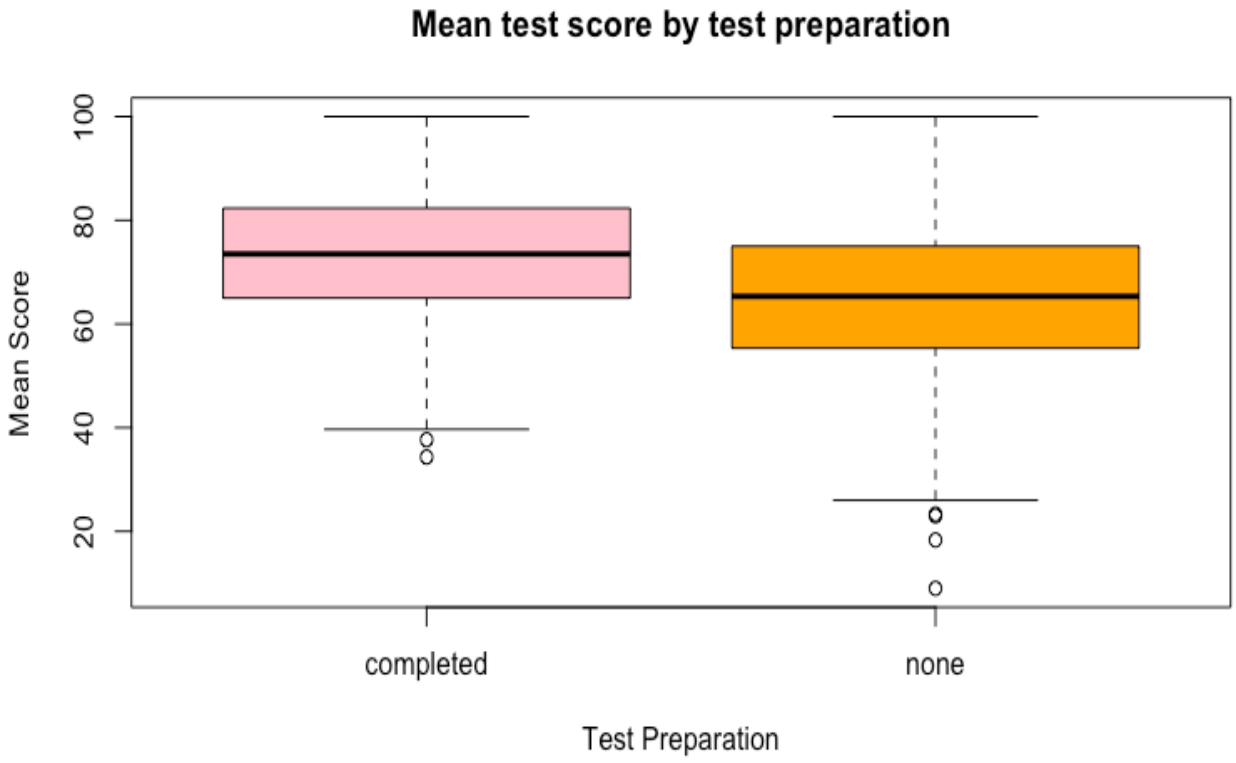

Figure 3.1: Boxplot for Mean score vs Test preparation

3.2 Descriptive analysis for race

Table 3.2: Descriptive analysis for Race

Race	Mean	Median
Group A	62.99	61.33
Group B	65.47	65.00
Group C	67.13	68.33
Group D	69.18	70.00
Group E	72.75	73.50

From the Table (3.2) we observe the mean test score for Group A is 62.99 which is the lowest mean score. Meanwhile the mean test score for Group E is 72.75 which is the highest mean test score. The median for Group E is 73.50 which is the highest median. Group D has the mean of 69.18 which is the second highest mean test score. The boxplot (3.2) also indicates that the Group E has a highest mean test score. From the boxplot, we can observe that we have outliers in the dataset.

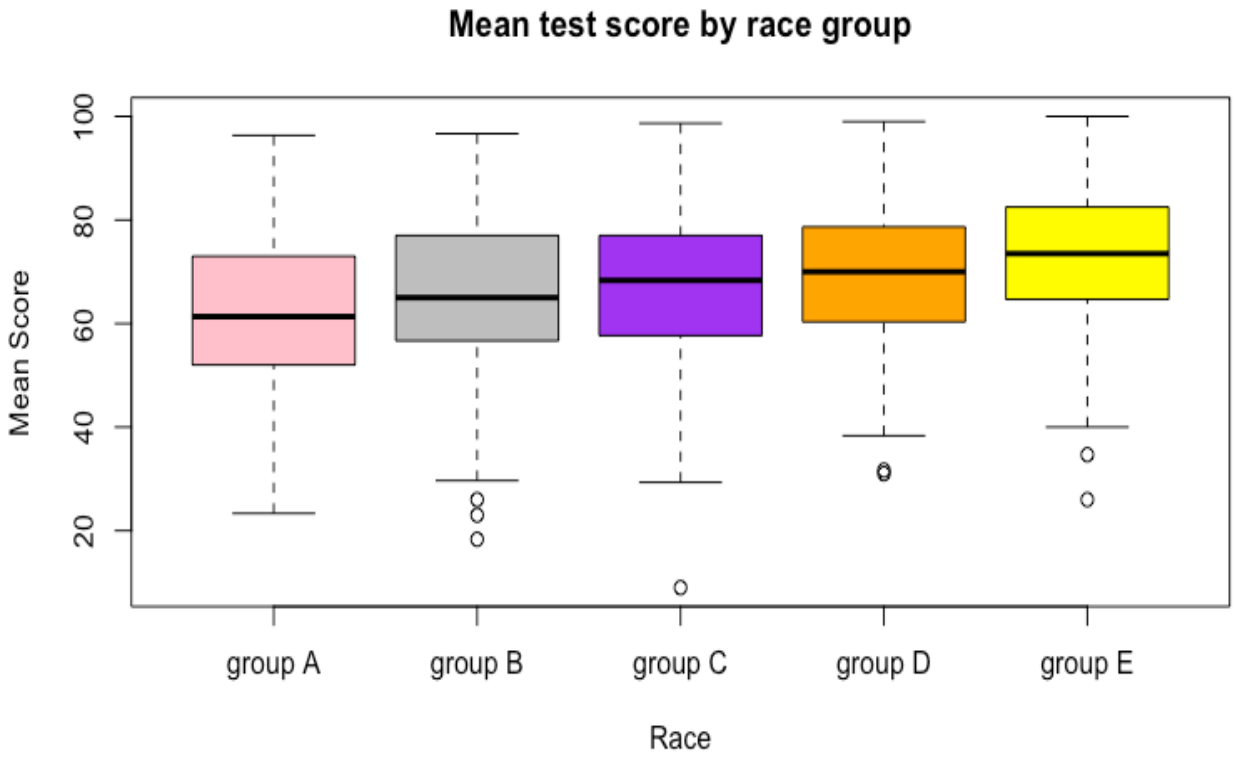

Figure 3.2: Boxplot for Mean score vs Test preparation

3.3 Descriptive analysis for Race and Test Preparation

Table 3.3.1 Mean for Race and Test Preparation

Race /Test preparation	Completed	None
Group A	70.06	59.21
Group B	70.64	62.58
Group C	71.87	64.39
Group D	73.53	67.20
Group E	76.69	69.80

Table 3.3.2 Standard Deviation for Race and Test Preparation

Race /Test preparation	Completed	None
Group A	15.30	12.54
Group B	12.88	14.96
Group C	13.21	13.53
Group D	11.82	13.42
Group E	12.61	15.29

From the table 3.3, we observe that the highest mean test score is for Group E (76.69) who have completed the test preparation. Meanwhile, Group A (59.21) who did not complete any test preparation has the lowest mean test score. In contrast to the mean, Group A (15.30) who completed the test score has the highest standard deviation. Meanwhile, Group D (11.82) who completed the test preparation has the lowest standard deviation.

3.4 Interaction Plot

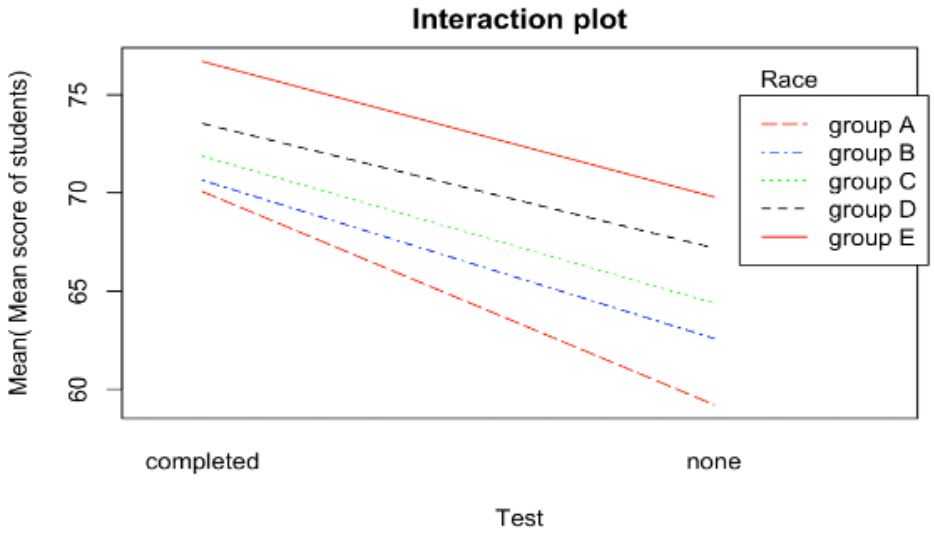

Figure 3.4: Interaction plot of the Variables

As illustrated in Figure (3.4), the lines are parallel, so we do not anticipate finding any interaction in our model. Group E showed the highest mean score for the completed and non-completed test. Group A showed the lowest mean score for the completed and non-completed test. The graphic shows that the test of the five groups appear to be different. There appears to be a small difference in the mean score of Races of who completed the test. Because there is no interaction in the plot, we can now investigate how the main effect influences the Mean.

3.5 Main effects of test and race

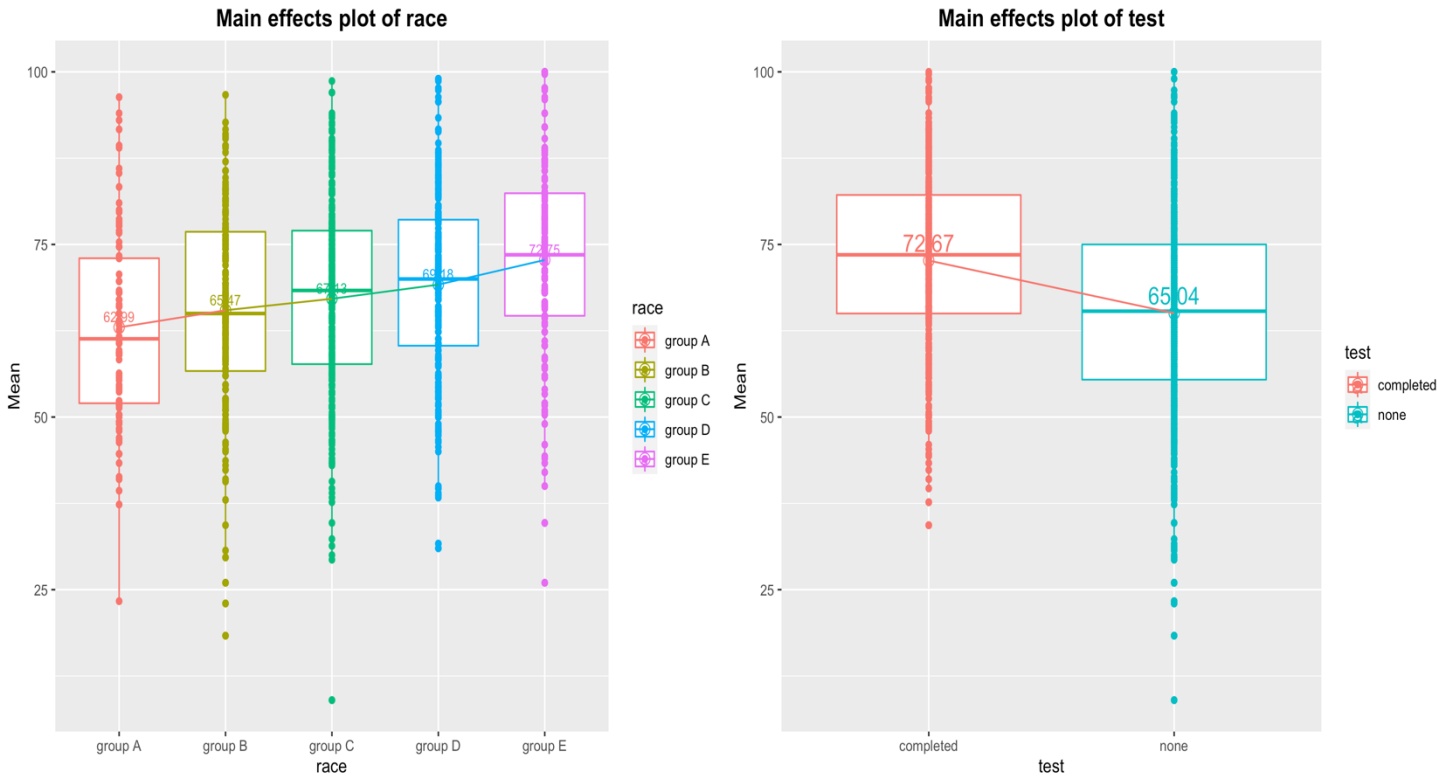

Figure 3.5: Main effects plot of race and test

From the plots shown in Figure (3.5), we can see that Group E has a higher mean score and Group A has a lower mean score. Also, the mean score of the students who completed the test preparation is higher than the mean score of the students who did not complete the test preparation.

4.Analysis

4.1 A Two-Factor Fixed Effects Model

In this experiment, five levels of race were used: A, B, C, D, E. In addition, two levels of the test were used: completed, and non-complected. Race and test both were considered as our fixed variable. So, we fit a two-factor fixed effects model. The model is:

$$Y_{ijk} = \mu_i + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \text{ where, } i = 1,...,a, j = 1,...,b, k = 1,...,n$$

- μ is the overall mean.
- τ_i is the effect of the ith level of factor A.
- β_j is the effect of the jth level of factor B.
- $(\tau\beta)_{ij}$ is the effect of interaction between race and test. We know that if interaction exists, we cannot rely on the p-values for the race or for the test prep.
- $\varepsilon_{ijk} \sim$ iid $N(0,\sigma^2)$.

4.2 The F-test

- Testing whether factors A(test) and B(race) interact:

  H_0: $(\tau\beta)_{ij} = 0$ for all i, j

  H_1: At least one $(\tau\beta)_{ij} \neq 0$

If there is no significant interaction, then we test the equality of effects of test and race.

- Testing the equality of the effects of factor A(test):

  H_0: $\tau_1 = \tau_2 = \cdots = \tau_a = 0$

  H_1: At least one $\tau_i \neq 0$

- Testing the equality of the effects of factor B(race):

  H_0: $\beta_1 = \beta_2 = \cdots = \beta_b = 0$

  H_1: At least one $\beta_b \neq 0$

4.3 ANOVA Results

The results from the two-factor ANOVA are listed below

Table 4.3: Two-factor ANOVA results

	DF	**Sum Sq**	**Mean Sq**	**F-value**	**Pr(>F)**
Race	4	7164	1791	9.707	1.06e-07 ***
Test Prep	1	12925	12925	70.054	<2e-16 ***
Race: Test Prep	4	329	82	0.446	0.775
Residuals	990	182651	184		
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1

Table (4.3) shows the ANOVA findings in conjunction with the interaction and main effect plots. We looked to determine if there was a substantial relationship between the test and race. The null hypothesis cannot be rejected as the p-value (0.775) is significantly high. This means we'll argue that there isn't any interaction between race and test preparation. We can check the equivalence of the effects of test and race because there is no interaction. Race has a p-value of 1.06e-07 which is less than 0.05. We have sufficient evidence to reject the null hypothesis and adopt the alternative hypothesis, which states that the there is an impact of race in test score. The test preparation has a p-value less than 2e-16, so we have sufficient evidence to reject the null hypothesis and support the alternative hypothesis, that the test preparation has impact on score. In our model, using the results from table above, we can conclude that both the race and test preparation have an impact on the mean test score. In the next section, we will run the Tukey's test. The purpose of Tukey's test is to figure out which groups in our sample differ from each other. It is, however, a number that represents the difference between groups, to compare different mean test score.

4.4 Tukey test

We used a Tukey test to compare the Means from the test and race.

4.4.1 Tukey test for test

Table 4.4.1: Tukey test for test

Variable Name	diff	lwr	upr	p adj
Group B-Group A	2.475912	-2.2920516	7.243875	0.6154456
Group C-Group A	4.139152	-0.3106668	8.588971	0.0823759
Group D-Group A	6.186880	1.6326966	10.741063	0.0020177
Group E-Group A	9.759872	4.7276319	14.792111	0.0000014
Group C-Group B	1.663240	-1.7384075	5.064888	0.6685986
Group D-Group B	3.710968	0.1738924	7.248044	0.0342884
Group E-Group B	7.283960	3.1495003	11.418420	0.0000168
Group D-Group C	2.047728	-1.0471583	5.142614	0.3693360
Group E-Group C	5.620720	1.8575864	9.383853	0.0004631
Group E-Group D	2.690947 -	1.1950367	6.576931	0.3220332

From the table 4.1.1, we can conclude that there is significant difference between Group D- Group A, Group D- Group B, Group E- Group A, Group E-Group B, and Group E-Group C The confidence interval for Group D-Group A is [1.6326966, 10.741063]. This interval does not include zero and the associated p-value is 0.0020177 which is less than 0.05 alpha level. The confidence interval for the mean of Group E- Group A is [4.7276319, 14.792111]. This interval does not include zero and the associated p-value is 0.0000014. The confidence interval for group D-group B is [0.1738924, 7.248044]. This interval does not include zero and the associated p-

value is 0.0342884. The confidence interval for group E-group B is [3.1495003, 11.418420]. This interval does not include zero and the associated p-value is 0.0000168 which is less than 0.05 alpha level. The confidence interval for group E-group C is [1.8575864, 9.383853]. This interval does not include zero and the associated p-value is 0.0004631. Besides these groups, all other p-value is higher than 0.05, indicating that there is no difference in mean test score between those race groups.

4.4.2 Tukey test for race

Table 4.4.2: Tukey test for race

Variable Name	diff	lwr	upr	p adj
None-completed	-7.47832	-9.236508	-5.720148	0

From table 4.4.2, we can see that the confidence interval for the mean of none−completed is [-9.236508 -5.720148]. Because this interval does not include zero and the associated p-value is 0, we can say that there is significant difference in the means score between the groups of students who completed the test preparation and those who didn't complete the test preparation.

4.5 Model Adequacy

4.5.1 Normality

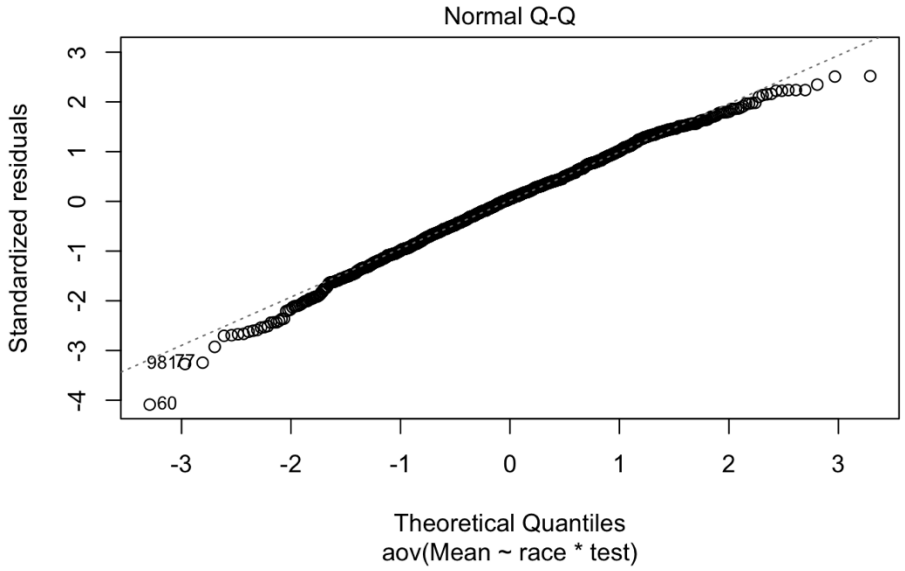

Figure 4.5.1: Normal Quantile-Quantile plot

With reference to Figure (4.5.1) above, we can observe that a good number of points are along the line, which indicates verification of normality assumption. However, we can also observe few outliers on our dataset. To confirm the normality assumption, we conduct Shapiro-Wilk normality test.

Table 4.5.1: *Shapiro-Wilk normality test*

Shapiro-Wilk normality test
data: results$res
W = 0.99369, p-value = 0.0003106

From the Shapiro-Wilk test we got the p-value = *0.0003106*, since the p-value is less than 0.05 it gives statistical evidence that the normality assumption is not satisfied.

4.5.2 Constant Variance

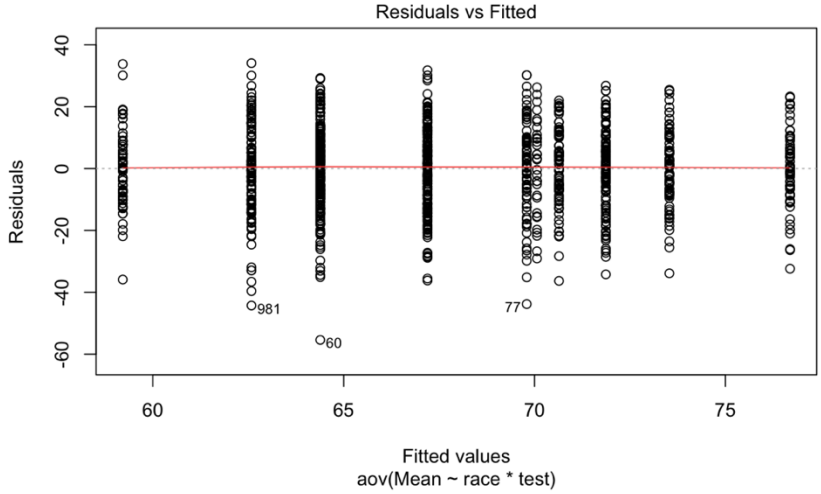

Figure 4.5.2: Constant variance plot

In the Residuals versus Fitted line, we can see that there is no pattern to the spread of the residuals for all combinations, which indicates that the variance is constant across all combinations of factor level. In order to confirm the constant variance assumption, Levene's test was performed. For this dataset, the normality assumption was not met which is why Levene's test was performed instead of Bartlett's test.

Table 4.5.2: Levene's Test for Homogeneity of Variance (center = median)

Levene's Test for Homogeneity of Variance (center = median)			
	Df	**F value**	**Pr(>F)**
group	**9**	**1.0694**	**0.3828**
	990		

From the table 4.5.1, we got the p-value = 0.3828. Thus, we do not have sufficient evidence to reject the null hypothesis which states constant variance assumption. Therefore, constant variance assumption is satisfied

5. Conclusions

The major goal of this study was to see how race and test preparation interacted with the students' mean score and to examine the interaction between race and test preparation and its effect on students' mean score. This research concludes that the test preparation has a positive impact on a student's average score. The study showed that there is no interaction between the two factor "Race" and "Test Preparation". We assumed that all the students were enrolled in same test preparation course. It also concludes that certain race groups are likely to obtain higher average test scores than others. We found out that the student who completed the test preparation has a higher mean test score. The Group E in "Race" has a highest and Group A has lowest mean test score.

Future Research Possibilities:

More accurate analysis can be done by collecting more data from more diverse student group. To calculate accurate significance, we could collect data from students who took different kinds of test preparation classes. In addition to the different kinds of test preparation, economical background of the family, and educational background of the students can be analyzed as well.

References

Jakki Seshapanpu. (2018, December 24). *Student performance in exams*. Kaggle. Retrieved April 22, 2022, from https://www.kaggle.com/code/spscientist/student-performance-in-exams/data?fbclid=IwAR1JfUC8FUBs95dlSbwv9e7RAJOOgl0UfwJOasrlkxJl3V7ssBnw1YBT3pU

Montgomery, Douglas C. (2021). *Design and analysis of Experiments*. John Wiley & Sons.

O'Sullivan, B., Dunn , K., & Berry, V. (n.d.). Test preparation: An international comparison of Test Takers' preferences. Taylor & Francis. Retrieved April 30, 2022, from https://www.tandfonline.com/doi/full/10.1080/0969594X.2019.1637820