STAT2183

Min Yang

Assignment 4

Assignment 4 Report

**Introduction**

The researcher wants to study whether an individual’s interest in politics was influenced by their education level and gender. He recruited a random sample of participants, then he asked them to score between 0 to 100 for their level of interest in politics, with a higher score represent a higher interest. The researcher divided the participants by gender as Female and Male, and education level (High School or lower/Some College/Undergraduate Degree/Graduate Degree).

**Method**

First, we will perform a two-way factorial design, which contains two factors—education level and gender. For education, it contains four levels— High School or lower/Some College/Undergraduate Degree/Graduate Degree. For gender, it contains two levels—female and male. To justify the experiment, there are three assumptions we need to satisfy. First, for each treatment group, the samples are random and independent. Second, all treatment samples are from normal distribution. Third, the variance of all treatments are equal.

Then, we test the treatment means are equal and see if there are difference among treatments. If they have equal mean, we can conclude neither factor has an effect on individual’s interest in politics. If there are difference, we further test the interaction effect to see if they have influence on individual’s interest in politics. If the interaction exists, we perform a relevant comparison of means procedures. If the interaction does not exit, we take a step further and test individual main effect for education and gender, and perform the comparison of means. For all test, we use 0.05 significance level.

**Test for Assumption**

1. Samples are random and independent

As stated in the question, the researcher recruited a random sample of participants, so this assumption is satisfied. The sample is selected randomly, and individual’s interest of politics is independent from each other.

1. All treatment sample are from normal distribution

This assumption is satisfied. From the box-plot, we can see for each group, the sample is normally distributed. Also from Shapiro-Wilk test, all treatment groups failed to reject the null hypothesis., so the distributions of these groups are normally distributed.

Hypothesis of Shapiro-Wilk test:

H0 : The distribution of data is normally distributed;

Ha : The distribution of data is not normally distributed.

|  |  |
| --- | --- |
| Female Graduate |  |
| Female Undergraduate |  |
| Female Some College |  |
| Female High School |  |
| Male Graduate |  |
| Male Undergraduate |  |
| Male Some College |  |
| Male High School |  |

Since for every group, their p-value is greater than 0.05, we fail to reject null hypothesis, so the distribution of data is normally distributed.

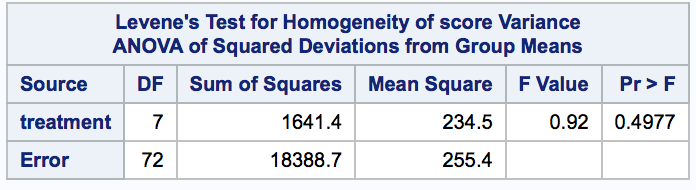
1. Variance of all treatments are equal

This assumption is satisfied. From the box-plot, the variance of each group is basically equal.

Also, we can perform testing for equal variance.

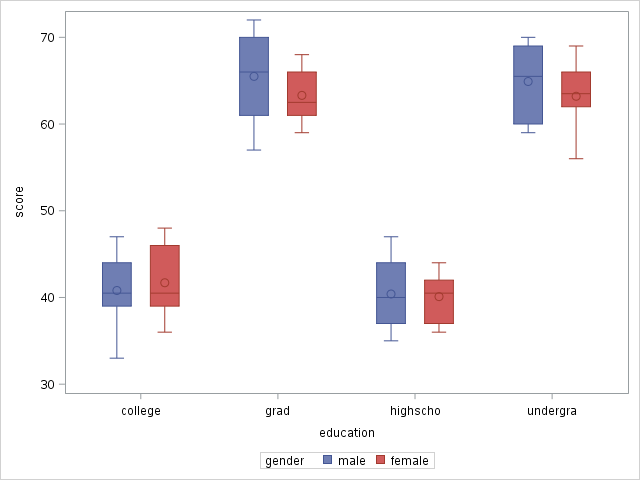
H0: =

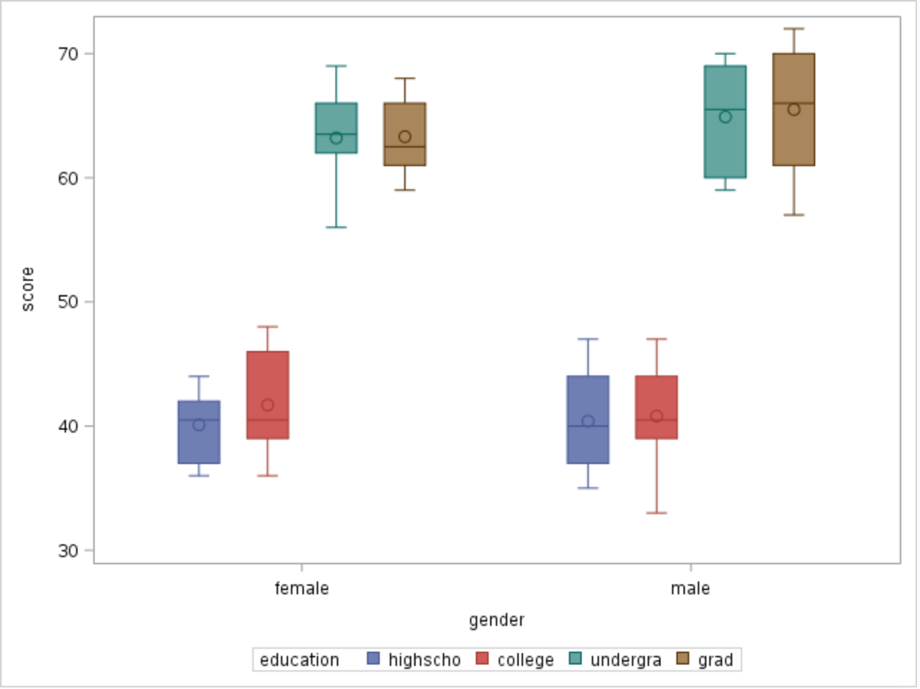
Ha: At least one is different



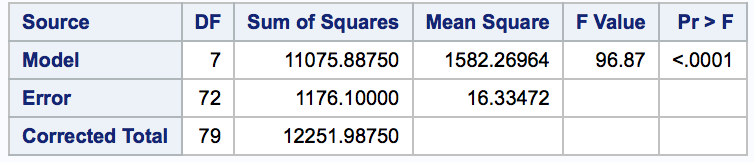
With F value=0.92, and p-value=0.4977 > 0.05, we failed to reject null hypothesis. So the variance of all treatments are equal, the assumption is satisfied.

Box-Plot





**Test for treatment means**

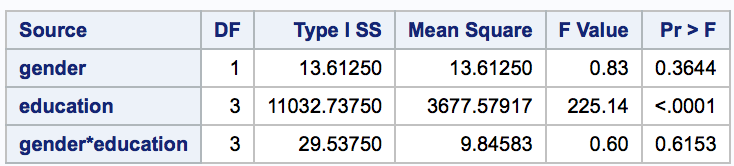
****

H0: all treatment means are equal

Ha: at least one treatment mean differs

With a F value of 96.87, and p-value <0.0001, we reject the null hypothesis and conclude that at least one treatment mean is different and has effect. Thus, we continue to test interaction.

**Test for interaction**

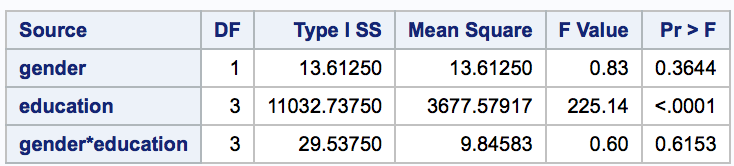


H0: Factor gender and education do not interact to affect the politic interest

Ha: Factor gender and education do interact to affect the politic interest

Interaction here means the effect of gender differs depending on the level of education, or vice versa. We want to test if there are interaction of gender and education. With F value of 0.6, and p-value =0.6153 < 0.05, we do not reject the null hypothesis, so gender and education do not have intersection effect on individual’s political interest. Thus, we need to test the main effect.

**Test for main effects – factor gender**

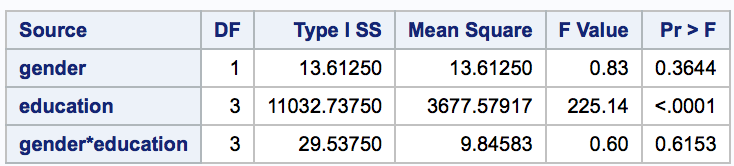


H0: male=female

Ha: malefemale

With a F value of 0.83, and p-value of 0.3644>0.05, we do not reject null hypothesis, so gender does not have an effect on individual’s interest of politics.

**Test for main effects – factor education**

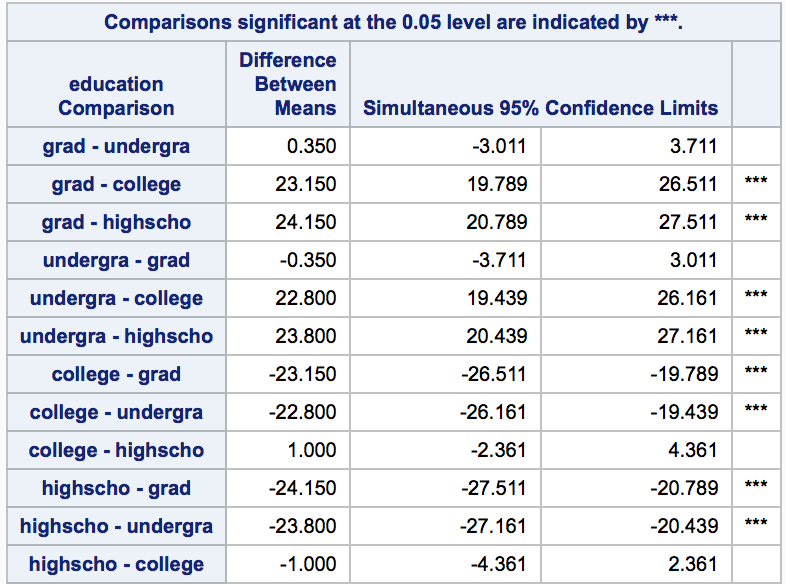


H0: high school=college=undergradute=graduate

Ha: at least one differs

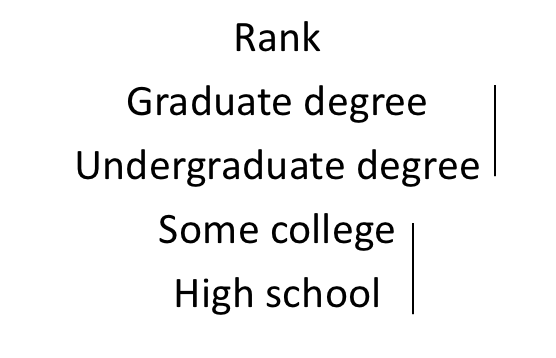
With F-value of 225.14, and p-value <0.0001, we reject the null hypothesis and conclude that the education has an effect on individual’s interest of politics. Since we reject the null hypothesis, we need to perform comparison of means for education factor.

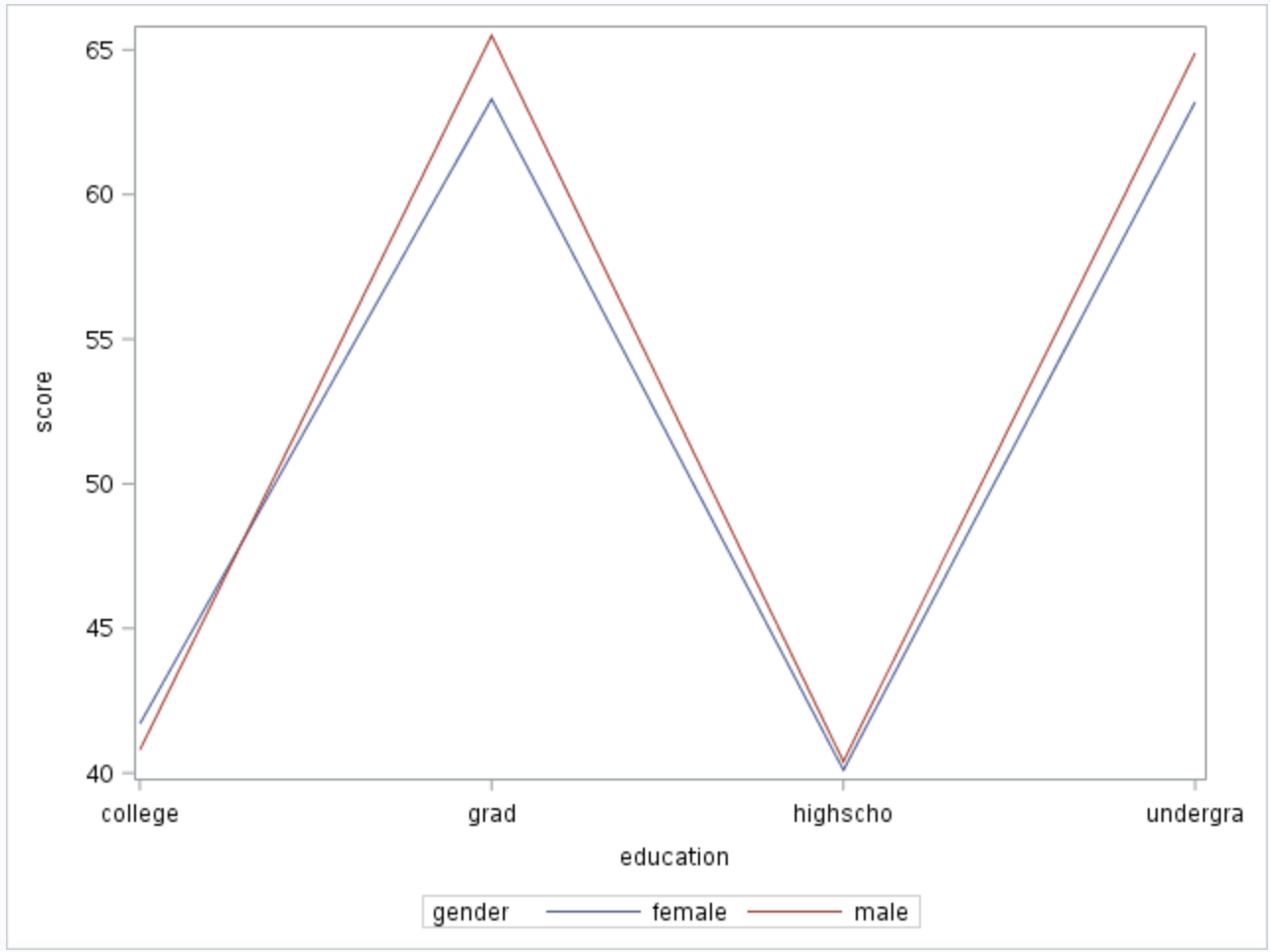
**Comparison of means**



By comparing the means of different education levels, we can rank the education effect that individuals with graduate and undergraduate degrees have higher interest of politics than individual with some college and high school. While graduate and undergraduate do not show significant different, neither does college and high school have significant difference.

Visualization of the comparison:





The Profile plot of the mean interest score matches with the conclusion I make above. The gender does not have an impact on individual’s interest of politics as two lines almost overlap on each other, while education does. People with graduate and undergraduate degree have a much higher interest than people with some college and high school education.

**Conclusion**

From the study, all assumptions of the test are satisfied. Not all treatment means are equal. There is no interaction effect as we fail to reject the null hypothesis. We failed to reject the test of main effect for gender, so gender has no effect on interest in politics. We found that education level has effect on individual’s interest of politics. We also found that individual with graduate or undergraduate degree have a much higher interest than people with some college or high school education. There is no significant difference between graduate degree group and undergraduate group, and there is no significant difference between some college group and high school group.

Appendix

DATA politics;

INPUT gender $ education $ score @@;

CARDS;

male highschool 44

male highschool 42

male highschool 35

male highschool 47

male highschool 38

male highschool 40

male highschool 45

male highschool 40

male highschool 36

male highschool 37

male college 40

male college 41

male college 43

male college 39

male college 47

male college 40

male college 44

male college 46

male college 35

male college 33

male undergrad 63

male undergrad 70

male undergrad 60

male undergrad 68

male undergrad 59

male undergrad 66

male undergrad 59

male undergrad 69

male undergrad 70

male undergrad 65

male grad 66

male grad 70

male grad 65

male grad 71

male grad 61

male grad 60

male grad 57

male grad 66

male grad 67

male grad 72

female highschool 41

female highschool 37

female highschool 43

female highschool 37

female highschool 42

female highschool 44

female highschool 42

female highschool 36

female highschool 39

female highschool 40

female college 47

female college 40

female college 41

female college 38

female college 48

female college 42

female college 40

female college 36

female college 46

female college 39

female undergrad 56

female undergrad 64

female undergrad 69

female undergrad 66

female undergrad 57

female undergrad 63

female undergrad 66

female undergrad 62

female undergrad 66

female undergrad 63

female grad 66

female grad 62

female grad 63

female grad 61

female grad 68

female grad 62

female grad 64

female grad 60

female grad 68

female grad 59

;

RUN;

PROC GLM DATA=politics;

CLASS gender education;

MODEL score=gender education gender\*education;

means education / CLDIFF TUKEY;

RUN;

PROC MEANS DATA=politics;

CLASS gender education;

OUTPUT OUT=poli;

RUN;

DATA poli2;

SET poli;

IF \_FREQ\_ = 10;

IF \_STAT\_ = "MEAN";

KEEP gender education score;

RUN;

PROC SGPLOT DATA=poli2;

SERIES X=education Y=score / GROUP=gender;

RUN;

PROC SGPLOT DATA=politics;

vbox score / group=gender category=education;

RUN;

proc univariate data = politics normal;

class gender education;

var score;

run;

DATA politics2;

SET politics;

if gender="male" and education="highschool" then treatment="mHS";

if gender="male" and education="college" then treatment="mSC";

if gender="male" and education="undergrad" then treatment="mUn";

if gender="male" and education="grad" then treatment="mGr";

if gender="female" and education="highschool" then treatment="fHS";

if gender="female" and education="college" then treatment="fSC";

if gender="female" and education="undergrad" then treatment="fUn";

if gender="female" and education="grad" then treatment="fGr";

RUN;

PROC GLM DATA=politics2;

CLASS treatment;

MODEL score = treatment;

MEANS treatment / HOVTEST;

RUN;