**Openness to Experience by Age and Gender Factorial ANOVA Analysis**

**Zachary Denton**

**Abstract** Classification of people’s personality has been an interest of many philosophers and phycologists over the years. The big five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) created by Paul Costa and Robert R. McCrae have become an industry standard in classifying personality. This research report aims to evaluate the effects of gender and age on openness to experience. The question at hand is: Is there a significant difference in openness to experience based on gender and age? There are three hypotheses in question: Main effect A: Ignoring age are there differences in openness to experience across gender? Main effect B: Ignoring gender are there differences in openness to experience across age levels? Interaction: does the effect of gender on openness to experience depend on age? To determine this a 2X5 two-way factorial ANOVA was conducted to evaluate mean differences between subjects. The dependent variable used was openness to experience (Oscore) converted to a NEO-FFI-R measurement used to measure the five personality traits. While, factor one gender (M/F) and factor two age (18-24, 25-34, 35-44, 45-54, 55-64) were used as independent variables. The study encompasses 1866 participants questioned via online survey. Males ages 18-24 had the highest averages in openness and females ages 45-54 had the lowest averages in openness. The three assumptions of ANOVA testing: normality, homogeneity of variance, and independence, were tested and met. The factorial ANOVA was then conducted an in conclusion, gender and age do not depend on each other in one’s openness to experience. However, when gender and age are each isolated there is significant differences in openness to experience levels.

Z. Denton (MABA)

University of Arkansas, Walton College of Business, Department of Information Systems

Email: zpdenton@uark.edu

**1 Introduction**

Personality types have been characterized by many throughout the course of history. Starting in the 1970’s Paul Costa and Robert R. McCrae’s developed the big five personality traits, (openness, conscientiousness, extraversion, agreeableness, and neuroticism) to classify the most human characteristics of one’s personality. This study servers the purpose to analysis if there is a significance difference between gender and age related to openness to experiences trait from the big five personality traits. People that are “open-minded” are usually curious, creative, imaginative people who enjoy musical, and artistic branches of culture. These people are open to try new things and experience as much as possible. As we get older, our personalities change. We know a thirty-year old will have a different personality than a twelve-year old but how to the differences look across only adult ages. The objective of the study is to convey results of a two-way factorial ANOVA test to explain the relationship that age and gender have on openness to experience. In the report the data is explained, research questions are defined, hypotheses are established, assumptions of the factorial ANOVA are defined, test results are interpreted, and conclusions against the hypothesis are made.

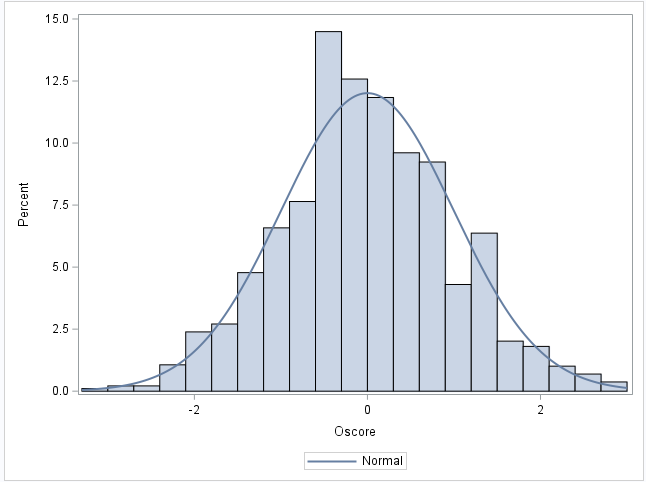
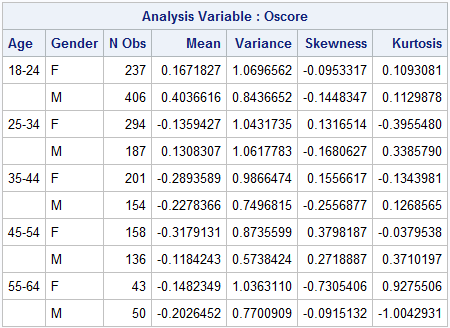
**Data** The data used in this study is a subset of variables from a web scraped dataset from Kaggle (Khadija, 2021). Data collection procedure of questioned participants were conducted using an online survey. There are 1866 observations (N) in the dataset. The three variables used in the study are age, gender, and openness to experience. The age variable is categorical and six levels that bin age as follows: 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. Distribution of records for the age variable are 34% for 18-24, 26% for 25-34, and 40% other. The 65+ age bin was removed from analysis due to low numbers of observations affecting variance and the chances of committing a type one two error (k=5). The gender variable is categorical with two levels being “M” male or “F” female (k=2). The distribution of records for gender is 50% male and 50% female. The variable for openness to experience has been converted to a quantitative input and represented as “Oscore”. Oscore has been converted from the Likert scale ranging 0 (strongly disagree) to 5 (strongly agree) to a NEO-FFI-R measurement. The NEO-FFI-R is a reliable test used to measure P.Costa and R.McCrae’s big five personality traits.

**Research Question and Hypothesis** The research question tested in the study is “Do age and gender have a significant impact on openness to experience?”. In this test, gender and age serve as the independent variables (IV) and Oscore as the dependent variable (DV). Gender being factor one and age being factor two. The hypothesis for the two-way factorial ANOVA conducted in the study for factors one, two, and interaction. The hypothesis for factor one is: Ho: No main effect of gender: µ(male)=µ(female). Ha: Main effect of gender: µ(male)≠µ(female). The hypothesis for factor two is: Ho: No main effect of age: µ(18-24)=µ(25-34)=µ(35-44)=µ(45-54)=µ(55-64). Ha: Main effect of age: µ(18-24)≠µ(25-34) ≠µ(35-44)≠µ(45-54)≠µ(55-64)≠µ(65+). The hypothesis for interaction is: Ho: No interaction between gender and age: µ(male, 18-24)=µ(male, 25-34)=µ(male, 35-44)=µ(male, 45-54)=µ(male, 55-64)=µ(female, 18-24)=µ(female, 25-34)=µ(female, 35-44)=µ(female, 45-54)=µ(female, 55-64). Ha: There is interaction between gender and age: µ(male, 18-24)≠µ(male, 25-34)≠µ(male, 35-44)≠µ(male, 45-54)≠µ(male, 55-64)≠µ(male, 65+)≠µ(female, 18-24)≠µ(female, 25-34)≠µ(female, 35-44)≠µ(female, 45-54)≠µ(female, 55-64)≠µ(female, 65+). An alpha (α) level of 0.05 was used for all testing purposes.

**Test Assumptions** The two-way factorial ANOVA test has several associated assumptions about the model. They are:

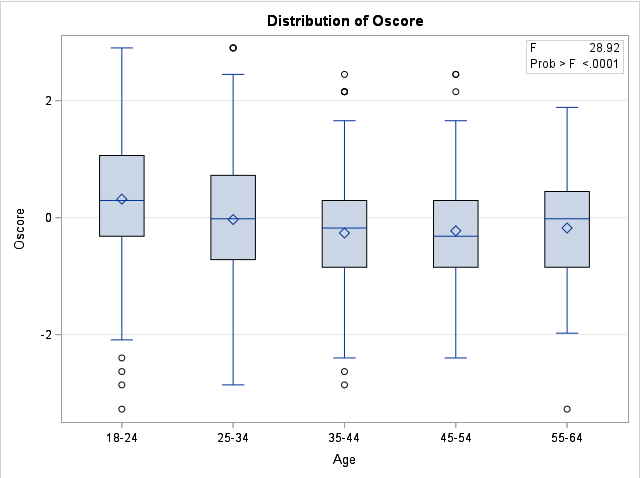
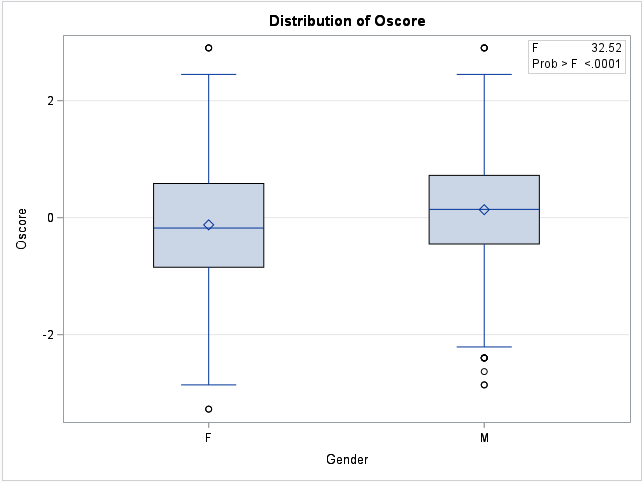
1. Dependent variable is normally distributed.
2. Homogeneity of variance – Equal variance of the dependent variable.
3. Independent observations.

Figure 1 below displays a histogram for Oscore distributions. Assumption one of normality is met meaning the dependent variable Oscore is normally distributed. Figure 2 below shows the summary statistics for the dependent variable Oscore grouped by factors gender and age. With all skewness and kurtosis values being between -1 and positive +1 for each cell, assumption one of normality has been met for cell factors.



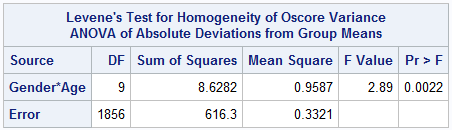
**Fig 2.** Summary statistics for Oscore with gender and age factors.

**Fig 1.** Histogram showing distribution of Oscore.



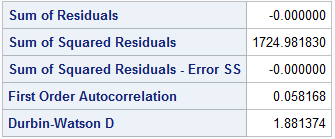
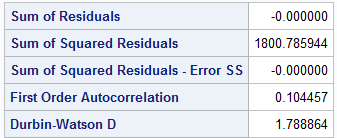
**Fig 4.** Box Plot of main effect age.

**Fig 3.** Box Plot of main effect gender.

Looking at assumption two, we can conclude that homogeneity of variance for Oscore between groups not been met. Figure 5 below shows the Levenes test for homogeneity of variance resulting in a p-value of 0.0022, indicating that we would reject the null hypothesis that there is equal homogeneity of variance between groups (α=0.05). However, with the f-max being less than three (1.07/0.58=1.84 <3) the statistical test is robust and will still produce accurate results even with the assumption failing to be met. With robustness being true I would conclude that assumption two has been met.

**Fig 5**. Levenes test for homogeneity of variance

When looking at figures 6 and 7 assumption three, independence. The test for independence resulted in a DW statistic of 1.880 for age and 1.789 for gender. With a DW statistic between 1.25 and 2.75 independence is met. For both age and gender assumption three has been met and fail to reject the null hypothesis that there is independence.



**Fig. 7** Independence test for gender

**Fig. 6** Independence test for age

All three assumptions of the factorial ANOVA have been met.

**2 Methods**

**Quality of Measurement** The data used in the study has been collected via online survey. There is a relatively large number of observations(N) which adds integrity to the study. The 65+ age bin was removed from analysis due to low numbers of observations affecting variance and the chances of committing a type one two error (k=5). Participants are 50% male and 50% female giving an equal distribution. Distribution of records for the age variable are 34% for 18-24, 26% for 25-34, and 40% other. The quantitative dependent variable has been converted to the NEO-FFI-R scale supporting equal variance in openness to experience.

**Analytical Strategy** There are several sections in the analysis of this research question. The first part is testing for the assumptions of a factorial ANOVA. Two is evaluating the sampling and power. The third section will be the two-way factorial ANOVA testing, results and conclusion.

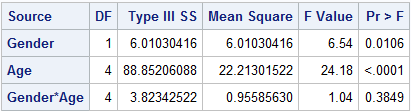
**Experiment Design** The design of the two-way factorial ANOVA is a 2x5 design. There are two factors (gender and age), gender having two levels, and age having five. Number of observations for each cell are indicated on figure 8 below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Observations | Age | | | | | |
| Gender |  | 18-24 | 25-34 | 35-44 | 45-54 | 55-64 |
| Male | 406 | 187 | 154 | 136 | 50 |
| Female | 237 | 294 | 201 | 158 | 43 |

**Fig. 8** Factorial ANOVA Design

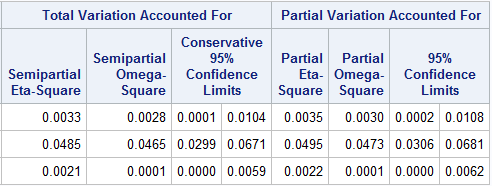
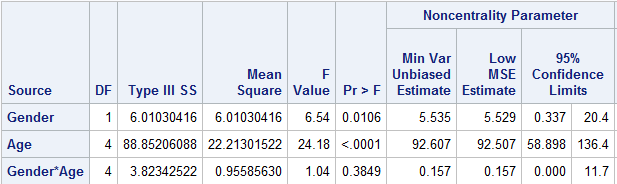
**3 Results**

Now that we have covered all assumptions for the two-way factorial ANOVA we can move on to the formal hypothesis testing for the omnibus test using SAS software PROC GLM. The factorial ANOVA will test all three hypotheses for all the effects and interaction. As previously stated the hypothesis for factor one is: Ho: No main effect of gender: µ(male)=µ(female). The hypothesis for factor two is: Ho: No main effect of age: µ(18-24)=µ(25-34)=µ(35-44)=µ(45-54)=µ(55-64).The hypothesis for interaction for both factors is: Ho: No interaction between gender and age: µ(male, 18-24)=µ(male, 25-34)=µ(male, 35-44)=µ(male, 45-54)=µ(male, 55-64)=µ(female, 18-24)=µ(female, 25-34)=µ(female, 35-44)=µ(female, 45-54)=µ(female, 55-64). The omnibus test results are shown in figure 9 below. Looking at the interaction first, the test produced a p= 0.3849 [0,11.7] which is not significant. Therefore, we would fail to reject null hypothesis (Ho). Factor one gender has a p-value of 0.011, [0.34,20.4] which is significant. Therefore, we would reject the null hypothesis for factor one, gender. Factor two age has a p <0.0001, [58.9,136.4] which is significant. Therefore, we would also reject the null hypothesis for factor two age.



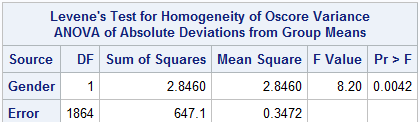
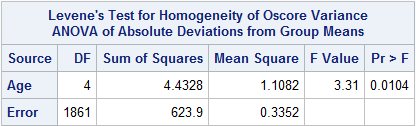
**Fig 9**. Two-way factorial ANOVA omnibus test

**Sampling Effect Size** There are 1866 observations used in the model. Effect size measures the strength in relationship between age and gender. The test is at α = 0.05 and produced a partial Eta statistic of 0.0022, [0,0.0062] seen below by figure 10. Age and gender explain less than 1% of Oscore level. This is a small effect.



**Fig 10.** Effect size for gender and age

**Follow-up Tests** Since we concluded that the interaction was not significant but factor one and two were significant we need to check homogeneity for each main effect. Figure 11 shows the output of the HOV test for the main effect of gender and Figure 12 shows the output for the main effect of age. Both main effects had a significant p-value, meaning we would reject the null hypothesis for age and gender. In conclusion, we do not have homogeneity of variance for each main effect.



**Fig 12.** HOV test for main effect age

**Fig 11.** HOV test for main effect gender

**4 Discussion**

The analysis done in this research report had the goal to answer the question of: Do age and gender have a significant impact on openness to experience? To answer this, the experiment was designed around a two-way factorial ANOVA (2x5). Oscore being the dependent variable and gender and age being the two independent variables. The highest overall means for openness score were males, ages 18-24, while the lowest mean openness score was held by females, ages 45-54. We have determined and met the three assumptions of the ANOVA test for the relationship between independent factors. From there the omnibus ANOVA test was conducted. A two-way factorial ANOVA was conducted to compare the effect of and gender and age on openness to experience. The statistical test concluded that there was not a significant interaction between gender and age on openness to experience (oscore). However, it did find that there is a significant effect of gender on openness to experience at the α=0.05 level for the conditions (F(1,1864) 6.45, p=0.0106). And there is a significant main effect of gender on openness to experience at the α=0.05 level for the conditions (F(4, 1861) = 24.18, p<.0001). After running post tests for each main effect, gender and age have not met homogeneity of variance. This raises questions about the validity of the test because when homogeneity is violated there is risk of committing a type one error (falsely reject the null hypothesis).

The study presents important understandings of contributing factors to Paul Costa and Robert R. McCrae’s big fiver personality trait, openness to experience. I found that gender and age do not have a significant effect on each other when determining a person’s openness to experience. However, age alone, and gender alone do have a significant effect on openness to experience levels.

**References and Appendix**

1. Khadija (2021, October) *Drug Consumption Data*.https://www.kaggle.com/code/
2. obeykhadija/drug-consumptions-eda/data
3. Fehrman, Elaine & Muhammad, Awaz & Mirkes, Evgeny & Egan, Vincent & Gorban, Alexander. (2017). The Five Factor Model of Personality and Evaluation of Drug Consumption Risk. 10.1007/978-3-319-55723-6\_18.
4. StudyCorgi. (2020, December 22). Two-Way Factorial Anova Analysis. Retrieved from <https://studycorgi.com/two-way-factorial-anova-analysis/>

The code for this paper was generated using SAS software, Version 9.4 of the SAS System for Windows. Copyright © 2013 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.”

**proc** **import**

datafile = "\\uem.walton.uark.edu\UEMProfiles\_Lab$\zpdenton\RedirectedFolders\Desktop\Drug\_Consumption.csv"

out = Drug;

**run**;

**proc** **sgplot** data=drug;

histogram oscore;

density oscore;

**run**;

/\* Descriptive Statistics \*/

**PROC** **MEANS** DATA=drug MEAN VAR SKEW KURT;

CLASS age gender;

VAR Oscore;

**RUN**;

/\* Homogeneity of Variance\*/

**PROC** **GLM** DATA=drug alpha=**0.05**;

CLASS gender age;

MODEL oscore=gender\*age;

MEANS gender\*age/ HOVTEST=LEVENE(TYPE=ABS);

**RUN**;

/\* follow up test for significant interaction slice gender\*/

**pROC** **GLM** DATA=drug;

CLASS gender age;

MODEL oscore=gender\*age;

LSMEANS gender\*age/adj=tukey slice=gender;

**RUN**;

/\* follow up test for significant interaction slice age\*/

**pROC** **GLM** DATA=drug;

CLASS gender age;

MODEL oscore=gender\*age;

LSMEANS gender\*age/adj=tukey slice=age;

**RUN**;

/\*Independence\*/

**PROC** **GLM** data=drug;

class age;

model oscore=age/p;

**run**;

**PROC** **GLM** data=drug;

class gender;

model oscore=gender/p;

**run**;

/\* omnibus test (factorial ANOVA)\*/

**PROC** **GLM** data=drug;

class gender age;

model oscore=gender age Gender\*Age;

**run**;

/\* HOV for gender and age\*/

**PROC** **GLM** DATA=drug;

CLASS GENDER age;

MODEL oscore=GENDER;

MEANS GENDER/HOVTEST=LEVENE(TYPE=ABS);

**RUN**;

**PROC** **GLM** DATA=drug;

CLASS GENDER age;

MODEL oscore=age;

MEANS age/HOVTEST=LEVENE(TYPE=ABS);

**RUN**;