**EXPLORING THE ECONOMIC OUTCOMES OF DEMOGRAPHIC GROUPS USING THE 2021 GENERAL SOCIAL SURVEY**

By Ryan Bresnahan

May 2022

STAT613: Multivariate Statistics

Final Report

**TABLE OF CONTENTS**

[INTRODUCTION 3](#_Toc104496598)

[METHODS 5](#_Toc104496599)

[RESULTS 10](#_Toc104496600)

[DISCUSSION 21](#_Toc104496601)

[REFERENCES 27](#_Toc104496602)

[APPENDIX A 28](#_Toc104496603)

[APPENDIX B 34](#_Toc104496604)

[APPENDIX C 36](#_Toc104496605)

[APPENDIX D 37](#_Toc104496606)

[APPENDIX E 39](#_Toc104496607)

**CHAPTER 1**

# INTRODUCTION

In recent years, American society has become increasingly focused on the observed and perceived disparities of socio-economic outcomes between different groups of people. This topic, while historically reserved for analysis by social scientists, has become a subject of intense investigation by people across all levels of the educational and vocational stratum. While there are infinitely many ways to group people together – down to the most minute biological, social, or psychological characteristics – the analysis of these differences has been largely focused on broader human characteristics that are not necessarily given to us by choice, but rather by chance.

The primary unchosen human characteristics that have come to the forefront of public debate have been ethnicity and sex. A brief review of recent literature makes salient the disparities in economic outcomes between racial groups as well as the sexes. According to one recent study published in 2020, there are marked differences in intergenerational economic mobility between the various ethnic groups living in the U.S. (Chetty et al., 2020). Additionally, a 2020 Pew Research Center study found that, in the aggregate, American women earned 84% of what American men earned in wages (Barroso & Brown, 2021).

There are also factors that affect economic outcomes which are derived from some level of personal choice, such as immigration/citizenship status and educational attainment. One analysis by the St. Louis Federal Reserve found that foreign-born immigrants are simultaneously more likely to have lower levels of high school education and higher levels of graduate education than the native-born population (Subhayu Bandyopadhyay & Rodrigo Guerrero, 2017). In the same St. Louis Federal Reserve study, it was noted that aggregate income of the immigrant population is lower than the native-born population. This likely reflects the fact that immigrants are more likely to work in lower-skilled occupations than the native-born population (Bennett, 2020). Additionally, it is widely known amongst both researchers and the general population that those with higher levels of education also have higher incomes and lower rates of unemployment (U.S. Bureau of Labor Statistics, 2021). Disparities even exist between those who own land and those who rent their living space, as landowners are 89% more wealthy than their renter counterparts (U.S. Census Bureau, n.d.).

Across all mediums of the public forum, there is an intense debate about what mechanisms cause these economic disparities in our society. Attempting to uncover these mechanisms would be out of the scope of this paper, and the purpose of this paper is not to find the reasons for why these disparities exist. Rather, this paper intends to simply explore the magnitude of these economic differences exists between the various groups discussed earlier, or if said differences exist in our data at all. For this paper, I will be using data from the 2021 General Social Survey to analyze the differences and similarities of family incomes across ethnic, sexual, immigration, educational, and economic backgrounds. This dataset was chosen due to its public availability, reliability, and usefulness for this paper’s specific purposes.

The research question this project aims to answer is the following; how do economic outcomes vary amongst different groups of ethnicities, men and women, native- and foreign-born citizens, different degree holders, and people of differing economic backgrounds?

**CHAPTER 2**

# METHODS

**Dataset and Respondents**

This paper will analyze the differences of family incomes between groups using data from the 2021 General Social Survey (2021 GSS). This dataset contains a total of 568 variables and 4,032 observations (or respondents). Note that this survey is generally considered a sociological survey, but our analysis is restricted to a small subset of economic- and demographic-related variables. Questions asked in the 2021 GSS primarily pertain to demographic information (race, income, sex, etc.), opinions on current issues (abortion, voting, religion, social attitudes, etc.), and personal information (status of physical/mental health, drug use, number of partners, etc.). All respondents of the GSS are 18 years old or older and are randomly sampled from households across the U.S. using an area probability design. The survey itself is conducted face-to-face or virtually by the National Opinion Research Center (NORC) at the University of Chicago. The dataset used in this analysis was made available by NORC in Stata format, and all analysis was completed using Stata.

**Statistical Methods**

The first analysis will utilize MANOVA to analyze the variations in real income and the number of rooms in the sample’s houses between different racial groupings (White, Black, Asian, and Hispanic), degree attainment levels (less than high school, high school diploma, associate/junior college, bachelor’s degree, graduate degree), and sex (male or female). The number of rooms and real income were chosen as the dependent variables since they are moderately positively correlated (correlation coefficient=0.33), and conceptually speaking it makes sense to group them together since higher incomes hypothetically implies a larger dwelling, which implies more rooms. More succinctly, both relate to economic status.

To ascertain the differences in income between said groups, a separate ANOVA model will be estimated with the dependent variable being real income, and the independent variables being race, degree level, sex, and a nested variable of the three independent variables in the order of race, degree, and sex (race|degree|sex). From this, I will contrast the values for each group in the nested variables race|degree and race|degree|sex (a factorial of all levels of each of the nested variables) against the mean income of the sample (using an array of Wald tests). This will test if the mean income of each specific segmentation is significantly different from the average income of all respondents of the same overall group. For example, this tests the average income of Black males with graduate degrees vs. all males with graduate degrees.

The second analysis will analyze economic status by using a multivariate regression. The primary intent of this analysis is to analyze the upward (or downward) economic mobility of different groups. The dependent variables will be real income and number of rooms in the respondent’s house (same as the MANOVA and ANOVA models) and the first independent variable a two-way interaction between “born” (yes/no if a respondent was born in the U.S.) and ethnic group. The next independent variable will be an interaction between ethnicity and the income of the respondent’s family at age 16 (five levels, low income to high income). Note that the regression will be coded to include the coefficients for the first variable in the interaction (main effect), then the interaction afterwards as a separate coefficient (interaction effect). Sex will be included as the last independent variable. From this, I will calculate separate linear predictions of real income and number of rooms for each ethnicity by their respective family’s incomes at age 16. These linear predictions will be visually shown as well. This will allow the reader to visualize how current adult incomes are sloped for each ethnicity by the respective income levels of their family (or parental income) at age 16.

**Hypotheses**

Given that our analysis is exploratory in nature and that we have no *a priori* hypotheses, all hypotheses will be stated as null hypotheses. Shown below in Table 1 are the null hypotheses that we will be testing.

**Table 1: Null Hypotheses**

|  |  |
| --- | --- |
| MANOVA Model | |
| H10 | The means of real income and number of rooms are equal for all ethnic groups. |
| H20 | The means of real income and number of rooms are equal for all degree level groups. |
| H30 | The means of real income and number of rooms are equal for men and women. |
| ANOVA Model | |
| H40 | The mean of real income for each ethnic group is not significantly different from the grand mean of each degree level. |
| H50 | The mean of real income for each ethnic group is not significantly different from the grand mean of each degree level and sex. |
| Multivariate Regression Model | |
| H60 | The coefficients for all independent variables (race, native status, sex, and income at age 16) are equal to zero. |

**Data Transformations and Pre-Modeling Diagnostics**

Prior to computation, the assumption of normal distribution of the dependent variables *REALINC* (family income in real dollars) and *NUMROOMS* (number of rooms in abode) was checked. This is necessary for the MANOVA, ANOVA, and multiple regression models. Histograms, quantile-normal plots, and Shapiro-Wilks W tests of *REALINC* and *NUMROOMS* prior to transformation are available in Appendix A as Figures 3-6 and Table 8. According to the visuals provided by the histograms and quantile-normal plots and the Shapiro-Wilks W tests (p<.0001), *REALINC* and *NUMROOMS* are not normally distributed in their natural state. Intentionally or not, it appears that the methodology used by NORC led to many observations of *REALINC* to take on the same values, primarily on either tail of the distribution. Furthermore, being that *NUMROOMS* is a count variable, its distribution is right-skewed and trends towards zero. Hence, *REALINC* was log transformed and *NUMROOMS* was log+1 transformed (to account for values equal to zero). Normality issues persist after transformation, but to a much smaller magnitude. Histograms, quantile-normal plots, and Shapiro-Wilks W tests of *REALINC* and *NUMROOMS* after transformation are available in Appendix A as Figures 7-10 and Table 9. Values shown in the results section will reflect the transformed values of the variables rather than the actual dollar amount of family incomes or the number of rooms in the respondents’ abodes.

For the MANOVA and ANOVA models, it was necessary to test the assumption of equality of variances. To check this assumption, I elected to use robust Levene’s tests of equality of variances, the results of which are shown in Appendix B as Tables 10-15. Robust Levene’s tests were chosen due to the violation of normality amongst the dependent variables (described in the previous paragraph). The array of the robust Levene’s tests produced mixed results. For real income by race (p<0.0001), degree attainment (p<0.0001), and sex (p<0.001), we reject the null hypothesis of equal variances. This indicates that amongst the subgroups contained within race, degree attainment, and sex, the variances are not equal. To account for this violation, the Welch’s F-test statistic will be prioritized in our analysis of the MANOVA model, rather than the other statistics provided in the output (though they will be shown in the output).

To check for collinearity, correlations between the variables included in the MANOVA, ANOVA, and multivariate regression models was computed. The results of this correlation are shown in Appendix C in Table 16. The correlation matrix that was returned does not appear to have any extreme correlations between the variables, although several variables were moderately correlated. *REALINC* and *NUMROOMS* have a correlation coefficient equal to 0.338, *DEGREE* and *REALINC*’s correlation coefficient is equal to 0.428, and *BORN* and *RACE*’s correlation coefficient is equal to 0.344. Overall, these correlation coefficients are not high enough to imply multicollinearity.

**CHAPTER 3**

# RESULTS

**MANOVA and ANOVA Models**

To find how similar or dissimilar the family incomes and number of rooms in abode are across all of our selected independent variables (ethnic, educational, and sexual backgrounds), the MANOVA model shown in Table 2 was estimated. The relevant null hypotheses are displayed in Table 1.

The results indicate that we may reject H10, H20, and H30. The Welch’s F-statistics along with the other test statistics shown in the output (indicated by W, L, P, and R) indicate that there are significant differences among *RACE*, *DEGREE*, and *SEX* (all p<0.0001). These results suggest that the included ethnic groups (White, Black, Hispanic, and Asian), degree attainment levels (less than high school to graduate degree), and sexual groups (male and female) significantly vary in real income and the number of rooms in their abodes.

**Table 2: MANOVA Model**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Type | Statistic | | df | F(df1, df2) | = | | F | Prob>F |
| Model | W | 0.767 | | 3 | 6 | 3166.000 | | 74.940 | 0.000 | e |
|  | P | 0.240 | |  | 6 | 3168.000 | | 72.070 | 0.000 | a |
|  | L | 0.295 | |  | 6 | 3164.000 | | 77.820 | 0.000 | a |
|  | R | 0.260 | |  | 3 | 1584.000 | | 137.450 | 0.000 | u |
| Residual | | | 1584 | | | |
| race | W | 0.960 | | 1 | 2 | 1583.000 | | 33.290 | 0.000 | e |
|  | P | 0.040 | |  | 2 | 1583.000 | | 33.290 | 0.000 | e |
|  | L | 0.042 | |  | 2 | 1583.000 | | 33.290 | 0.000 | e |
|  | R | 0.042 | |  | 2 | 1583.000 | | 33.290 | 0.000 | e |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| degree | W | 0.825 | 1 | 2 | 1583.000 | 168.080 | 0.000 | e |
|  | P | 0.175 |  | 2 | 1583.000 | 168.080 | 0.000 | e |
|  | L | 0.212 |  | 2 | 1583.000 | 168.080 | 0.000 | e |
|  | R | 0.212 |  | 2 | 1583.000 | 168.080 | 0.000 | e |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sex | W | 0.977 | | 1 | 2 | 1583.000 | | 18.560 | 0.000 | e |
|  | P | 0.023 | |  | 2 | 1583.000 | | 18.560 | 0.000 | e |
|  | L | 0.023 | |  | 2 | 1583.000 | | 18.560 | 0.000 | e |
|  | R | 0.023 | |  | 2 | 1583.000 | | 18.560 | 0.000 | e |
| Residual | | | 1584 | | | |
| Total | | | 1587 | | | |
| Number of obs = 1,588 W = Wilks' lambda  L = Lawley-Hotelling trace  P = Pillai's trace  R = Roy's largest root  e = exact  a = approximate  u = upper bound on F | | |

Shown in Table 3 are the results from the ANOVA model, wherein the only dependent variable is real income. The ANOVA model includes the same independent variables as the MANOVA model shown in Table 2, with the addition of a nested variable of race|degree|sex (a factorial of those three variables). Null hypotheses for the overall ANOVA model are not included since the MANOVA model also measures the overall differences in real income (ANOVA hypotheses relate to the contrasts that will be shown on the following pages).

The independent variables contained in the model explains 21.85% (R2=0.2185) of the variance amongst the dependent variable *REALINC*. All independent variables are highly significant (all p<0.0001) apart from the nested variable race|degree|sex. These results confirm previous results in the MANOVA model that the included ethnic groups (White, Black, Hispanic, and Asian), degree attainment levels (less than high school to graduate degree), and sexual groups (male and female) significantly vary in real income.

**Table 3: ANOVA Model**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Number of obs | 3,428 | R-squared = | 0.2185 |  |  |  |
| Root MSE | 1.137 | Adj R-squared = | 0.2095 |  |  |  |
| Source | Partial SS | d.f. | MS | F | Prob>F | Sig. |
| Model | 1224.58 | 39 | 31.399487 | 24.29 | 0.0000 | \*\*\* |
| race | 81.000212 | 3 | 27.000071 | 20.89 | 0.0000 | \*\*\* |
| degree | 219.89991 | 4 | 54.974978 | 42.53 | 0.0000 | \*\*\* |
| sex | 24.139922 | 1 | 24.139922 | 18.67 | 0.0000 | \*\*\* |
| race|degree|sex | 49.88495 | 31 | 1.609192 | 1.24 | 0.1656 |  |
| Residual | 4379.8899 | 3,388 | 1.2927656 |  |  |  |
| Total | 5604.47 | 3,427 | 1.6353866 |  |  |  |
| Note: Significance levels of p-values are indicated as such: p<0.10 (\*), p<0.05 (\*\*), p<0.01 (\*\*\*) | | | | | | |

Shown in Table 4 are the contrasts (or comparisons) of each ethnicity to the grand mean of each degree level. These contrasts are computed, for example, by comparing the mean of White people who have less than a High School degree to the grand mean of all respondents that have less than a High School degree (this would be shown as (white vs. mean) <High School in the output). The relevant hypothesis is shown in Table 1.

The results suggest that many of the ethnic subgroups specified in Table 4 are significantly different from their relevant degree attainment groupings. Amongst the White category, we reject H40 for those with less than high school degrees (Contrast=0.594098, p<.01), high school degree (Contrast=0.1522, p<.05), associate degrees (Contrast=0.240092, p<.10), bachelor’s degrees (Contrast=0.231939, p<.05), and graduate degrees (Contrast=0.243623, p<.05). Amongst the Black category, we reject H40 for those with less than a high school degree (Contrast=-0.50174, p<.10) and high school degrees (Contrast=-0.29959, p<.01). Amongst the Asian category, we reject H40 for those with bachelor’s degrees (Contrast=0.316489, p<.05) and graduate degrees (Contrast=0.377705, p<.05). Amongst the Hispanic category, we reject H40 for those with bachelor’s degrees (Contrast=-0.40976, p<.05) and graduate degrees (Contrast=-0.43112, p<.10).

**Table 4: Contrasts of Ethnicity by Respective Degree Levels**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| race|degree | Contrast | Std. Err. | t | P>t | Sig. |
| (white vs mean) <High School | 0.594098 | 0.191071 | 3.11 | 0.002 | \*\*\* |
| (white vs mean) High School | 0.1522 | 0.072378 | 2.10 | 0.04 | \*\* |
| (white vs mean) Associate | 0.240092 | 0.169848 | 1.41 | 0.16 | \* |
| (white vs mean) Bachelor | 0.231939 | 0.082893 | 2.80 | 0.01 | \*\* |
| (white vs mean) Graduate | 0.243623 | 0.104805 | 2.32 | 0.02 | \*\* |
| (black vs mean) <High School | -0.50174 | 0.265876 | -1.89 | 0.06 | \* |
| (black vs mean) High School | -0.29959 | 0.094104 | -3.18 | 0.00 | \*\*\* |
| (black vs mean) Associate | -0.31372 | 0.199793 | -1.57 | 0.12 |  |
| (black vs mean) Bachelor | -0.13867 | 0.116389 | -1.19 | 0.23 |  |
| (black vs mean) Graduate | -0.19021 | 0.146687 | -1.30 | 0.20 |  |
| (asian vs mean) <High School | -0.02674 | 0.438049 | -0.06 | 0.95 |  |
| (asian vs mean) High School | 0.252968 | 0.161355 | 1.57 | 0.12 |  |
| (asian vs mean) Associate | 0.486812 | 0.384012 | 1.27 | 0.21 |  |
| (asian vs mean) Bachelor | 0.316489 | 0.130266 | 2.43 | 0.02 | \*\* |
| (asian vs mean) Graduate | 0.377705 | 0.141223 | 2.67 | 0.01 | \*\* |
| (hispanic vs mean) <High School | -0.06563 | 0.253782 | -0.26 | 0.80 |  |
| (hispanic vs mean) High School | -0.10558 | 0.124099 | -0.85 | 0.40 |  |
| (hispanic vs mean) Associate | -0.41319 | 0.314621 | -1.31 | 0.19 |  |
| (hispanic vs mean) Bachelor | -0.40976 | 0.186383 | -2.20 | 0.03 | \*\* |
| (hispanic vs mean) Graduate | -0.43112 | 0.25536 | -1.69 | 0.09 | \* |
| Note: Significance levels of p-values are indicated as such: p<0.10 (\*), p<0.05 (\*\*), p<0.01 (\*\*\*) | | | | | |

Shown in Table 5 are the contrasts for each ethnicity by the grand mean for each level of degree attainment and sex. In the case of Table 5, these contrasts are computed, for example, by comparing the mean of White males who have less than a High School degree to the grand mean of all male respondents that have less than a High School degree (this would be shown as (white vs. mean) <High School#Male in the output). The relevant hypothesis is shown in Table 1.

Amongst the White category, we may reject H50 for females with less than a high school degree (Contrast=0.7522887, p<.01), females with a high school degree (Contrast=0.2068488, p<.05), females with bachelor’s degrees (Contrast=0.3398581, p<.01), and females with graduate degrees (Contrast=0.3714185, p<.01). Amongst the Black category, we reject H50 for males with less than a high school degree (Contrast=-1.286043, p<.01), females with a high school degree (Contrast=-0.3779682, p<.01), males with associate degrees (Contrast=-0.6058961, p<.01), and females with graduate degrees (Contrast=-0.3296885, p<.10). Amongst the Asian category, we reject H50 for males with less than a high school degree (Contrast=1.313957, p<.05), females with less than a high school degree (Contrast=-1.36743, p<.05), males with a high school degree (Contrast=0.4396078, p<.05), men with bachelor’s degrees (Contrast=0.5056068, p<.01), and males with graduate degrees (Contrast=0.4674516, p<.05). Amongst the Hispanic category, we may reject for males with high school degrees (Contrast=-0.3159478, p<.10) and males with bachelor’s degrees (Contrast=-0.5055043, p<.10).

**Table 5: Contrasts of Ethnicity by Respective Degree Levels and Sex**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| race|degree|sex | Contrast | Std. Err. | t | P>t | Sig. |
| (white vs mean) <High School#Male | 0.4359079 | 0.2885691 | 1.51 | 0.131 |  |
| (white vs mean) <High School#Female | 0.7522887 | 0.2505206 | 3 | 0.003 | \*\*\* |
| (white vs mean) High School#Male | 0.0975509 | 0.1005086 | 0.97 | 0.332 |  |
| (white vs mean) High School#Female | 0.2068488 | 0.1041745 | 1.99 | 0.047 | \*\* |
| (white vs mean) Associate#Male | 0.1412177 | 0.2686025 | 0.53 | 0.599 |  |
| (white vs mean) Associate#Female | 0.3389667 | 0.207958 | 1.63 | 0.103 |  |
| (white vs mean) Bachelor#Male | 0.1240204 | 0.1261528 | 0.98 | 0.326 |  |
| (white vs mean) Bachelor#Female | 0.3398581 | 0.1075655 | 3.16 | 0.002 | \*\*\* |
| (white vs mean) Graduate#Male | 0.115827 | 0.1585896 | 0.73 | 0.465 |  |
| (white vs mean) Graduate#Female | 0.3714185 | 0.1370621 | 2.71 | 0.007 | \*\*\* |
| (black vs mean) <High School#Male | -1.286043 | 0.4436026 | -2.9 | 0.004 | \*\*\* |
| (black vs mean) <High School#Female | 0.2825704 | 0.2932162 | 0.96 | 0.335 |  |
| (black vs mean) High School#Male | -0.2212109 | 0.1417975 | -1.56 | 0.119 |  |
| (black vs mean) High School#Female | -0.3779682 | 0.1237573 | -3.05 | 0.002 | \*\*\* |
| (black vs mean) Associate#Male | -0.6058961 | 0.32269 | -1.88 | 0.061 | \* |
| (black vs mean) Associate#Female | -0.0215359 | 0.2356688 | -0.09 | 0.927 |  |
| (black vs mean) Bachelor#Male | -0.1241229 | 0.1776323 | -0.7 | 0.485 |  |
| (black vs mean) Bachelor#Female | -0.1532229 | 0.1504413 | -1.02 | 0.309 |  |
| (black vs mean) Graduate#Male | -0.0507302 | 0.2284692 | -0.22 | 0.824 |  |
| (black vs mean) Graduate#Female | -0.3296885 | 0.1840398 | -1.79 | 0.073 | \* |
| (asian vs mean) <High School#Male | 1.313957 | 0.6250585 | 2.1 | 0.036 | \*\* |
| (asian vs mean) <High School#Female | -1.36743 | 0.6138795 | -2.23 | 0.026 | \*\* |
| (asian vs mean) High School#Male | 0.4396078 | 0.2113685 | 2.08 | 0.038 | \*\* |
| (asian vs mean) High School#Female | 0.0663289 | 0.2438552 | 0.27 | 0.786 |  |
| (asian vs mean) Associate#Male | 0.9823392 | 0.6239373 | 1.57 | 0.115 |  |
| (asian vs mean) Associate#Female | -0.0087154 | 0.447841 | -0.02 | 0.984 |  |
| (asian vs mean) Bachelor#Male | 0.5056068 | 0.1887578 | 2.68 | 0.007 | \*\*\* |
| (asian vs mean) Bachelor#Female | 0.1273721 | 0.1795758 | 0.71 | 0.478 |  |
| (asian vs mean) Graduate#Male | 0.4674516 | 0.2106258 | 2.22 | 0.027 | \*\* |
| (asian vs mean) Graduate#Female | 0.2879585 | 0.188181 | 1.53 | 0.126 |  |
| (hispanic vs mean) <High School#Male | -0.4638211 | 0.3732654 | -1.24 | 0.214 |  |
| (hispanic vs mean) <High School#Female | 0.3325704 | 0.3439402 | 0.97 | 0.334 |  |
| (hispanic vs mean) High School#Male | -0.3159478 | 0.1741051 | -1.81 | 0.07 | \* |
| (hispanic vs mean) High School#Female | 0.1047905 | 0.1768889 | 0.59 | 0.554 |  |
| (hispanic vs mean) Associate#Male | -0.5176608 | 0.4771813 | -1.08 | 0.278 |  |
| (hispanic vs mean) Associate#Female | -0.3087154 | 0.4101736 | -0.75 | 0.452 |  |
| (hispanic vs mean) Bachelor#Male | -0.5055043 | 0.2930934 | -1.72 | 0.085 | \* |
| (hispanic vs mean) Bachelor#Female | -0.3140072 | 0.2303272 | -1.36 | 0.173 |  |
| (hispanic vs mean) Graduate#Male | -0.5325484 | 0.3899919 | -1.37 | 0.172 |  |
| (hispanic vs mean) Graduate#Female | -0.3296885 | 0.3297594 | -1 | 0.317 |  |
| Note: Significance levels of p-values are indicated as such: p<0.10 (\*), p<0.05 (\*\*), p<0.01 (\*\*\*) | | | | | |

Table 6 shows the pairwise comparisons of each ethnicity by their respective average real incomes. The results indicate that there are significant differences in real income between several ethnic groups. Note that the contrasts are not absolute values and represent the order in which each group is presented in Table 6 (Hispanic vs. Asian would be x̄Hispanic - x̄Asian). There are significant differences for Black vs. White (Contrast=-0.581, p<.01), Hispanic vs. White (Contrast=-0.577, p<.01), Asian vs. Black (Contrast=0.570, p<.01), and Hispanic vs. Asian (Contrast=-0.567, p<.01).

**Table 6: Pairwise Comparisons of Ethnicities by Real Income**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Contrast | Std.Err. | | t | | P>t | [95%Conf. | Interval] | Sig. |
| race | | |  | |
| black vs white | -0.581 | 0.084 | | -6.890 | | 0.000 | -0.747 | -0.416 | \*\*\* |
| asian vs white | -0.011 | 0.164 | | -0.070 | | 0.947 | -0.333 | 0.311 |  |
| hispanic vs white | -0.577 | 0.131 | | -4.420 | | 0.000 | -0.833 | -0.321 | \*\*\* |
| asian vs black | 0.570 | 0.180 | | 3.180 | | 0.002 | 0.218 | 0.922 | \*\*\* |
| hispanic vs black | 0.004 | 0.149 | | 0.020 | | 0.980 | -0.289 | 0.297 |  |
| hispanic vs asian | -0.567 | 0.205 | | -2.760 | | 0.006 | -0.969 | -0.164 | \*\*\* |
| Note: Significance levels of p-values are indicated as such: p<0.10 (\*), p<0.05 (\*\*), p<0.01 (\*\*\*) | | | | | | | | |  |

**MULTIVARIATE REGRESSION MODEL**

To ascertain the linear effects of native-born status (*BORN*), ethnicity (*RACE*), childhood family income (*INCOM16*), and sex (*SEX*) on economic outcomes (defined by real income and number of rooms), the multivariate regression model shown in Table 7 was estimated. The relevant null hypotheses are displayed in Table 1.

Analyzing the entire model, we first observe that the independent variable *BORN*, *RACE, INCOM16,* and *SEX* explain only 7.96% (R2=0.0796) and 6.03 (R2=0.0603) of the variance of the dependent variables *REALINC* and *NUMROOMS* respectively. The p-values are highly significant for both dependent variables (p<.01), however, the R2 values indicate that economic outcomes are determined by a much larger number of factors that are unspecified in the model.

For the real income dependent variable, most coefficients are significant, and we may reject H60. Note that *RACE* is specified as a factor variable (whereas it would naturally be ordinal), so the model treats each level as an indicator variable. Each level of *RACE* represents the discrete change (or difference) from the base level (reference group) of real income or number of rooms from the White category (whose coefficient is represented by the constant, or \_cons). The results suggest that the Black category (β=-0.5596254, p<.01) and Hispanic category (β=-0.5339266, p<.01) earn significantly less in real income compared to the White category, while the Asian category (β=0.3969462, p<.05) earn significantly more in real income compared to the White category. An increased level of *INCOM16* (family income at age 16) corresponds with a significant increase in a respondent’s real income (β=0.1344612, p<.01). Additionally, the female category (β=-0.3605994, p<.01) earn significantly less in real income than the male category.

We see a slightly similar pattern for the results regarding the number of rooms in an abode. First, we see that foreign-born respondents (β =-0.2311444, p<.01) have significantly fewer rooms in their abode than native-born respondents. Additionally, the results indicate that the Black (β=-0.1689302, p<.01), Asian (β =-0.1281495, p<.05), and Hispanic (β=-0.2607037, p<.01) categories all have significantly fewer rooms in their abode than the White category.

**Table 7: Multivariate Regression Model**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Equation | Obs | Parms | RMSE | R-sq | F | P | Sig |
| realinc | 1,566 | 7 | 1.135932 | 0.0796 | 22.48079 | 0.00 | \*\*\* |
| numrooms | 1,566 | 7 | 0.4486402 | 0.0603 | 16.67962 | 0.00 | \*\*\* |
|  | Coef. | Std. Err. | t | P>t | [95% Conf. | Interval] |  |
| realinc |  |  |  |  |  |  |  |
| born | 0.0205005 | 0.1068229 | 0.19 | 0.85 | -0.1890313 | 0.2300322 |  |
| race |  |  |  |  |  |  |  |
| black | -0.5596254 | 0.0902277 | -6.20 | 0.00 | -0.7366057 | -0.3826450 | \*\*\* |
| asian | 0.3969462 | 0.1442966 | 2.75 | 0.01 | 0.1139102 | 0.6799821 | \*\* |
| hispanic | -0.5339266 | 0.1539767 | -3.47 | 0.00 | -0.8359499 | -0.2319034 | \*\*\* |
| incom16 | 0.1344612 | 0.0300243 | 4.48 | 0.00 | 0.0755690 | 0.1933534 | \*\*\* |
| sex | -0.3605994 | 0.0579700 | -6.22 | 0.00 | -0.4743068 | -0.2468920 | \*\*\* |
| \_cons | 10.3568300 | 0.1785143 | 58.02 | 0.00 | 10.0066800 | 10.7069900 | \*\*\* |
|  | Coef. | Std. Err. | t | P>t | [95% Conf. | Interval] |  |
| numrooms |  |  |  |  |  |  |  |
| born | -0.2311444 | 0.0421901 | -5.48 | 0.00 | -0.3138996 | -0.1483891 | \*\*\* |
| race |  |  |  |  |  |  |  |
| black | -0.1689302 | 0.0356357 | -4.74 | 0.00 | -0.2388292 | -0.0990312 | \*\*\* |
| asian | -0.1281495 | 0.0569904 | -2.25 | 0.03 | -0.2399355 | -0.0163635 | \*\* |
| hispanic | -0.2607037 | 0.0608136 | -4.29 | 0.00 | -0.3799888 | -0.1414186 | \*\*\* |
| incom16 | 0.0033382 | 0.0118582 | 0.28 | 0.78 | -0.0199214 | 0.0265979 |  |
| sex | -0.0113682 | 0.0228955 | -0.50 | 0.62 | -0.0562773 | 0.0335410 |  |
| \_cons | 2.0578590 | 0.0705048 | 29.19 | 0.00 | 1.9195640 | 2.1961530 | \*\*\* |

Prior to discussion of Figures 1 and 2, we must take note a couple of important pieces of information. One, these linear predictions were modeled separately in one-way linear regression models, wherein the output is shown in Appendix D. Hence, these estimates are slightly different than the ones shown in the multivariate regression model discussed earlier. Second, the independent variable is a categorical variable defined by five increasing levels of income (far below average to far above average).

Shown in Figure 1 is a visualization of predicted income by increasing levels of income at age 16. The coefficients and significance level are indicated in each section of the figure, which is sorted by ethnicity. The results shown in Figure 1 and its corresponding output in Appendix D suggest that a one-unit increase in income at age 16 is associated with a significant linear increase in real income for the White category (β=0.159, p<.01) and Black category (β=0.177, p<.01). Observing the results without accounting for significance, all coefficients indicate that increase incomes at age 16 correspond with linear increases in real income.

**Figure 1: Predicted Income by Age 16 Family Income (Sorted by Ethnicity)**



Shown in Figure 2 is a visualization of predicted number of rooms in abode by income at age 16. This visualization takes on quite a different composition from the previous linear predictions shown in Figure 1. The results shown in the figure below and in Appendix D indicate that a one-unit increase in family income at age 16 is associated with a significant linear increase in the number of rooms in abode for the White category (β=0.021, p<.10), but a significant linear decrease for the Black category (β=0.030, p<.10). Interestingly, increases in family income at age 16 also lead to linear increases for the Hispanic group and linear decreases for the Asian group. Note that the fifth age 16 income category was omitted due to low sample size for Hispanics.

**Figure 2: Predicted Number of Rooms by Age 16 Family Income (Sorted by Ethnicity)**



Post-estimation diagnostics for the linear regression models are shown in Appendix E. VIF estimates did not indicate collinearity, however, the Breusch-Pagan test indicates that heteroskedasticity exists in the multivariate regression model. This is a logical outcome since sample sizes varied across the subgroups, as the methodology used in surveying did not require all respondents to answer each question. Furthermore, data was collected to be representative of each ethnic population’s true proportion of the entire American population. Hence, estimations of the confidence intervals (along with standard deviations) should be viewed with caution.

**CHAPTER 4**

# DISCUSSION

Overall, the results of our statistical analysis display varying levels of economic inequality in the 2021 GSS amongst various ethnicities, men and women, native- and foreign-born citizens, and people of different levels of degree attainment and economic background. Our discussion of these results, however, will be limited to interpreting these differences as meaningfully as possible within the parameters of our empirical analysis, rather than interpreting them as a systemic issue (which is not in the scope of this paper nor my area of expertise).

The results of the MANOVA model indicated that there are significant differences across ethnic groups, degree attainment levels, and between males and females. While we delve further into our analysis further on, the results from the MANOVA suggest several takeaways. One, real incomes and number of rooms in abode are significantly different between White, Black, Asian, and Hispanic respondents. Two, real income and number of rooms in abode are significantly different between those with less than a high school degree, high school degree holders, associate degree holders, bachelor’s degree holders, and graduate degree holders. Three, real income and the number of rooms in abode are significantly different between men and women. The results from the overall ANOVA model with real income as the dependent variable confirm these the results of the MANOVA model. Overall, this suggests that there are marked differences between respondents depending on their innate characteristics (ethnicity, sex, etc.) and characteristics of choice and/or circumstance (degree attainment level).

The results from the contrasts against the grand means (first defined by degree attainment level) reveal these disparities in greater focus. Without accounting for statistical significance, we see that the White and Asian categories are likely to have above average incomes when compared to their respective degree attainment levels, while the Black and Hispanic categories are likely to have below average incomes when compared to their respective degree attainment levels. In the results, we see that White respondents with less than a high school degree, high school degree, associate degree, bachelor’s degree, and graduate degree have significantly higher incomes than the average income of respondents in each of those degree categories. This also applies for Asian respondents with bachelor’s degrees and graduate degrees. Furthermore, we see that Black respondents with less than a high school degree and high school degrees have significantly lower incomes than the average income for those respective degree categories. Lastly, Hispanic respondents with a bachelor’s or graduate degree earn significantly lower incomes than the average incomes of bachelor’s and graduate degree holders.

Next, we contrasted the mean of each group (defined here by ethnicity, degree level, and sex) to the grand mean of all ethnicities (defined only by degree level and sex). These results revealed all possible differences between groups within the realm of statistical possibility (any further nesting of the variable would likely lead to n=0 for some groups). In the White category, White females with less than a high school degree, high school degree, bachelor’s degree, and graduate degree earned significantly higher incomes than their female/similarly educated counterparts. No significant differences were observed amongst White men. In the Black category, Black men with less than a high school degree and associate degrees have significantly lower incomes than their male/similarly educated peers. Additionally, Black women with high school and graduate degrees have significantly lower incomes than their counterparts. In the Asian category, Asian men with less than a high school degree, high school degree, bachelor’s degree, and graduate degree have significantly higher incomes than their male/similarly educated counterparts. On the other hand, Asian women with less than a high school degree have significantly lower incomes compared to their peers. Finally, we observed only two significant differences in the Hispanic category. Hispanic men with high school and bachelor’s degrees have significantly lower incomes in comparison to men with similar degree attainment levels. The segmentation done with this series of contrasts was important to our analysis, as it shows precisely where the largest differences in income lie. Hypothetically, further nesting could be possible (such as native-born vs. foreign-born), but that would require a much larger sample size.

The final analysis that was undertaken post-ANOVA was the pairwise comparison of each ethnicity’s average real income. These simple tests serve to buttress the results of the MANOVA and ANOVA models and show where the largest disparities exist. The results are relatively straightforward – the incomes of the Black and Hispanic category are significantly less than the White and Asian categories. There were no significant differences between Black and Hispanic respondents, as well as no significant differences between Asian and White respondents.

The results from the multivariate regression model were similar to those discussed throughout this section, albeit with slight variations. In terms of real income, Black and Hispanic respondents made significantly less income than White respondents, while Asian respondents made significantly more income than White respondents. Interestingly, this result confounds the pairwise comparisons that were discussed in the previous paragraph. The explanation for this discrepancy is due to the pairwise comparisons needing to be weighted according to sample size, which was necessary due to the large difference in White and Asian sample sizes. Moving forward, an increase in age 16 family income is positively linearly associated with higher incomes (adult incomes to be precise – minimum age to participate in the survey is 18 years old). In other words, the wealthier a respondent is when they are a minor, the wealthier they are likely to be as an adult. Finally, females earn significantly less in real income than the males in the aggregate. This is an expected result, as it is known in existing research that American females earn less than their male counterparts as a whole.

The results of the multivariate regression when the number of rooms in abode is the dependent variable reveal some unexpected trends (at least to the author). Respondents who were not born in the United States have significantly less rooms in their home than native-born respondents. Contextualizing this with what was discussed in the introduction section, this is a logical result as immigrants tend to have less wealth than natives, and those who have less wealth tend to have smaller houses or rent apartments (hence having fewer rooms). Further into the results, we see that Black, Asian, and Hispanic respondents have significantly less rooms in their abodes than White respondents. To fully make sense of this, we will first discuss the linear predictions of the number of rooms in abode by income at age 16 (sorted by ethnicity).

Observing the results of the linear regressions that were visualized by Figures 1 and 2, we see that as age 16 family income increases for White respondents, the number of rooms in their home significantly increases simultaneously. Hispanic respondents of increasing family wealth at age 16 follows the same positive trend (although this is not significant). Conversely, as age 16 family income at age 16 increases for Asian and Black respondents, the number of rooms in their home decreases (significantly decreases in the case of Black respondents). These trends confound earlier results that indicated that the number of rooms in abode increases as familial income increases, as well as the intuitive notion that as people earn more, they are likely to have more rooms in their house/apartment/etc. One possible explanation are trends in urbanization. It is possible that the Black and Asian respondents with greater familial income tend to live in more urban settings (and therefore apartments), as opposed to suburban or rural settings wherein most dwellings tend to be houses (with more rooms than urban apartments). This is merely a theory, though, and it is possible that underpowered sample sizes lead to incorrect predictions (though that is unlikely given the number of observations).

Finally, the results from the linear predictions of real income by increasing family incomes at age 16 reveal several notable takeaways. For both White and Black respondents, a one-unit increase in familial income leads to a significant increase in current income. This suggests that for White and Black respondents, a potential key predictor of adult income is parental income (family income at age 16 could be viewed as a proxy of parental income). The slope (or coefficient) of predicted income by family income at age 16 is also positive for Asian and Hispanic respondents, however, their coefficients are not significant. Note that this does not necessarily mean that familial income is an important predictor, though, as these regression models were one-way (only included one variable, so more susceptible to omitted variable bias) and had relatively low explained variances. Additionally, one must also note the fairly subjective nature of the *INCOM16* variable, as they were self-reported and could be entirely relative to the economic conditions a respondent is accustomed to.

Overall, the results of this analysis highlight many of the economic inequalities that are discussed or pondered upon daily by many. While this is in no way a substitute for analyses done with, for example, census data, this paper should at the very least demonstrate the usefulness of the General Social Survey. This analysis did, of course, come with some limitations. For one, the analysis was restricted to the 2021 GSS dataset, which included upwards of 4,000 observations. This was not adequate, as some of the sample sizes (particularly for the groups specified by the nested variables) were relatively low, which could skew results. This restriction was due to the Stata license version that is provided by the University of Delaware, which would not have allowed for multi-year datasets (too many observations). The assumption of homoskedasticity for the linear regression model could also not be fully met, which likely stems from the fact that the *REALINC* and *NUMROOMS* having the presence of many outliers. Non-parametric methods would usually be employed in this case, however, the goal of the analysis was to display the linear relationship of income at age 16 and the dependent variables, so it was necessary to proceed with the linear model (with a warning of caution so as not to view its results as absolute). In the future, these analyses should use a dataset from multiple years (preferably recent years) to bolster sample sizes, and a thorough power analysis should be conducted beforehand.

# REFERENCES

Barroso, A., & Brown, A. (2021, May 25). Gender pay gap in U.S. held steady in 2020. *Pew Research Center*. <https://www.pewresearch.org/fact-tank/2021/05/25/gender-pay-gap-facts/>

Bennett, J. (2020, February 24). The share of immigrant workers in high-skill jobs is rising in the U.S. *Pew Research Center*. <https://www.pewresearch.org/fact-tank/2020/02/24/the-share-of-immigrant-workers-in-high-skill-jobs-is-rising-in-the-u-s/>

Bureau, U. C. (n.d.). *2017 Data Show Homeowners Nearly 89 Times Wealthier Than Renters*. Census.Gov. Retrieved May 26, 2022, from <https://www.census.gov/library/stories/2020/11/gaps-in-wealth-of-americans-by-household-type-in-2017.html>

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and Economic Opportunity in the United States: An Intergenerational Perspective\*. *The Quarterly Journal of Economics*, *135*(2), 711–783. <https://doi.org/10.1093/qje/qjz042>

Subhayu Bandyopadhyay & Rodrigo Guerrero. (2017, July 25). *Immigrants vs. Natives on Jobs, Income, Education | St. Louis Fed*. <https://www.stlouisfed.org/publications/regional-economist/second-quarter-2017/comparing-income-education-and-job-data-for-immigrants-vs-those-born-in-us>

U.S. Bureau of Labor Statistics. (2021, April 21). *Education pays: U.S. Bureau of Labor Statistics*. <https://www.bls.gov/emp/chart-unemployment-earnings-education.htm>

# APPENDIX A

**VARIABLE DISTRIBUTIONS AND TRANSFORMATIONS**

**Figure 3: Distribution of Real Income (Untransformed)**



**Figure 4: Distribution of the Number of Rooms in Abode (Untransformed)**

****

**Figure 5: Quantile-Normal Plot of Real Income (Untransformed)**

****

**Figure 6: Quantile-Normal Plot of Number of Rooms in Abode (Untransformed)**

****

**Table 8: Shapiro-Wilk W Test for Normal data (Untransformed)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Obs | W | V | z | Prob>z |
| realinc | 3,509 | 0.746 | 502.020 | 16.132 | 0.000 |
| numrooms | 1,812 | 0.970 | 32.460 | 8.820 | 0.000 |

**Figure 7: Distribution of Real Income (Transformed)**

****

**Figure 8: Distribution of the Number of Rooms in Abode (Transformed)**

****

**Figure 9: Quantile-Normal Plot of Real Income (Transformed)**

****

**Figure 10: Quantile-Normal Plot of Number of Rooms in Abode (Transformed)**

****

**Table 9: Shapiro-Wilk W Test for Normal data (Transformed)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Obs | W | V | z | Prob>z |
| realinc | 3,509 | 0.924 | 149.237 | 12.985 | 0.000 |
| numrooms | 1,812 | 0.985 | 15.901 | 7.011 | 0.000 |

# APPENDIX B

**LEVENE’S TESTS OF EQUAL VARIANCES**

**Table 10: Number of Rooms by Race**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| race | Mean | | Std. Dev. | |  | Freq. |
| white | | 1.798 | | 0.450 | | 1,398 |
| black | | 1.589 | | 0.468 | | 218 |
| asian | | 1.542 | | 0.477 | | 92 |
| hispanic | | 1.450 | | 0.443 | | 74 |
| Total | | 1.745 | | 0.465 | | 1,782 |

W0 = 0.99793694 df(3, 1778) Pr > F = 0.39283955  
W50 = 0.81644508 df(3, 1778) Pr > F = 0.48469999  
W10 = 0.73198734 df(3, 1778) Pr > F = 0.53288971

**Table 11: Number of Rooms by Degree Attainment**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| degree | Mean | | Std. | | Dev. | | Freq. |
| 0 | | 1.560 | | 0.460 | | 91 | |
| 1 | | 1.686 | | 0.451 | | 689 | |
| 2 | | 1.709 | | 0.458 | | 170 | |
| 3 | | 1.786 | | 0.466 | | 472 | |
| 4 | | 1.882 | | 0.445 | | 379 | |
| Total | | 1.749 | | 0.463 | | 1,801 | |

W0 = 0.45559448 df(4, 1796) Pr > F = 0.76836927  
W50 = 0.72384323 df(4, 1796) Pr > F = 0.57560713  
W10 = 0.23841823 df(4, 1796) Pr > F = 0.91669634

**Table 12: Number of Rooms by Sex**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| sex | Mean | | Std. | | Dev. | | Freq. |
| Male | | 1.751 | | 0.476 | | 803 | |
| Female | | 1.743 | | 0.455 | | 1,006 | |
| Total | | 1.747 | | 0.465 | | 1,809 | |

W0 = 2.4281305 df(1, 1807) Pr > F = 0.11935006  
W50 = 2.7654012 df(1, 1807) Pr > F = 0.09649595  
W10 = 2.8062900 df(1, 1807) Pr > F = 0.09406831

**Table 13: Real Income by Race**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| race | Mean | | Std. | | Dev. | | Freq. |
| white | | 10.157 | | 1.137 | | 2,755 | |
| black | | 9.550 | | 1.370 | | 406 | |
| asian | | 10.429 | | 1.354 | | 167 | |
| hispanic | | 9.482 | | 1.461 | | 125 | |
| Total | | 10.074 | | 1.213 | | 3,453 | |

W0 = 9.7542920 df(3, 3449) Pr > F = 0.00000209  
W50 = 7.4699750 df(3, 3449) Pr > F = 0.00005541  
W10 = 7.4297729 df(3, 3449) Pr > F = 0.00005868

**Table 14: Real Income by Degree Attainment**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| degree | Mean | | Std. | | Dev. | | Freq. |
| 0 | | 8.705 | | 1.623 | | 166 | |
| 1 | | 9.692 | | 1.159 | | 1,363 | |
| 2 | | 9.933 | | 1.018 | | 335 | |
| 3 | | 10.443 | | 0.958 | | 940 | |
| 4 | | 10.711 | | 1.022 | | 693 | |
| Total | | 10.072 | | 1.210 | | 3,497 | |

W0 = 25.932735 df(4, 3492) Pr > F = 0.00000000  
W50 = 20.562491 df(4, 3492) Pr > F = 0.00000000  
W10 = 26.268935 df(4, 3492) Pr > F = 0.00000000

**Table 15: Real Income by Sex**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| sex | Mean | | Std. | | Dev. | | Freq. |
| Male | | 10.253 | | 1.148 | | 1,572 | |
| Female | | 9.927 | | 1.237 | | 1,918 | |
| Total | | 10.074 | | 1.209 | | 3,490 | |

W0 = 11.0739295 df(1, 3488) Pr > F = 0.00088458  
W50 = 8.9675184 df(1, 3488) Pr > F = 0.00276753  
W10 = 9.4767006 df(1, 3488) Pr > F = 0.00209716

# APPENDIX C

**CORRELATION MATRIX**

**Table 16: Variable Correlations**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| (1) numrooms | 1.000 |
| (2) realinc | 0.338 | 1.000 |
| (3) race | -0.201 | -0.085 | 1.000 |
| (4) degree | 0.167 | 0.428 | -0.036 | 1.000 |
| (5) sex | -0.002 | -0.163 | -0.004 | -0.091 | 1.000 |
| (6) born | -0.192 | 0.025 | 0.344 | 0.053 | -0.085 | 1.000 |
| (7) incom16 | 0.044 | 0.128 | -0.120 | 0.218 | 0.000 | -0.097 | 1.000 |
|  | | | | | | | |

# APPENDIX D

**ONE-WAY LINEAR REGRESSION OUTPUTS**

**Table 17: Linear Predictions of Real Income by Family Income at 16yo (Sorted by Ethnicity)**

|  |  |  |
| --- | --- | --- |
| -> race = white | | |
| Source | SS | df | | MS | Number of obs = 2,705 | | | |
| Model | 60.153 | 1 | | 60.153 | Prob>F = 0.000 | | | |
| Residual | 3358.513 | 2,703 | | 1.243 | R-squared = 0.018 | | |
| Total | 3418.666 | 2,704 | | 1.264 | Root MSE = 1.115 | | | | |
| realinc | Coef. | Std.Err. | | t | P>t | [95%Conf. | Interval] |
| incom16 | 0.159 | 0.023 | | 6.960 | 0.000 | 0.114 | 0.203 |
| \_cons | 9.721 | 0.068 | | 143.770 | 0.000 | 9.589 | 9.854 |

-> race = black

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | Number of obs = 392 | | | |
| Model | 12.364 | 1 | 12.364 | Prob>F = 0.009 | | | |
| Residual | 700.892 | 390 | 1.797 | R-squared = 0.017 | | |
| Total | 713.257 | 391 | 1.824 | Root MSE = 1.341 | | | | |
| realinc | Coef. | Std.Err. | t | P>t | [95%Conf. | Interval] |
| incom16 | 0.177 | 0.068 | 2.620 | 0.009 | 0.044 | 0.310 |
| \_cons | 9.149 | 0.179 | 51.210 | 0.000 | 8.797 | 9.500 |

|  |  |  |
| --- | --- | --- |
| -> race = asian | | |
| Source | SS | df | | MS | Number of obs = 158 | | | |
| Model | 1.387 | 1 | | 1.387 | Prob>F = 0.390 | | | |
| Residual | 290.889 | 156 | | 1.865 | R-squared = 0.005 | | |
| Total | 292.275 | 157 | | 1.862 | Root MSE = 1.365 | | | | |
| realinc | Coef. | Std. Err. | | t | P>t | [95%Conf. | Interval] |
| incom16 | 0.095 | 0.110 | | 0.860 | 0.390 | -0.123 | 0.313 |
| \_cons | 10.205 | 0.313 | | 32.570 | 0.000 | 9.586 | 10.824 |
|  | | |

-> race = hispanic

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | Number of obs = 120 | | | |
| Model | 3.767 | 1 | 3.767 | Prob>F = 0.155 | | | |
| Residual | 216.935 | 118 | 1.838 | R-squared = 0.017 | | |
| Total | 220.702 | 119 | 1.855 | Root MSE = 1.356 | | | | |
| realinc | Coef. | Std.Err. | t | P>t | [95%Conf. | Interval] |
| incom16 | 0.199 | 0.139 | 1.430 | 0.155 | -0.076 | 0.473 |
| \_cons | 9.075 | 0.360 | 25.230 | 0.000 | 8.362 | 9.787 |

**Table 18: Linear Predictions of # of Rooms by Family Income at 16yo (Sorted by Ethnicity)**

|  |  |  |
| --- | --- | --- |
| -> race = white | | |
| Source | SS | df | | MS | Number of obs = 1,374 | | | |
| Model | 0.547 | 1 | | 0.547 | Prob>F = 0.100 | | | |
| Residual | 276.844 | 1,372 | | 0.202 | R-squared = 0.002 | | |
| Total | 277.391 | 1,373 | | 0.202 | Root MSE = 0.449 | | | | |
| numrooms | Coef. | Std.Err. | | t | P>t | [95%Conf. | Interval] |
| incom16 | 0.021 | 0.013 | | 1.650 | 0.100 | -0.004 | 0.047 |
| \_cons | 1.742 | 0.038 | | 45.410 | 0.000 | 1.667 | 1.817 |

-> race = black

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | Number of obs = 209 | | | |
| Model | 0.205 | 1 | 0.205 | Prob>F = 0.333 | | | |
| Residual | 45.112 | 207 | 0.218 | R-squared = 0.004 | | |
| Total | 45.317 | 208 | 0.218 | Root MSE = 0.467 | | | | |
| numrooms | Coef. | Std.Err. | t | P>t | [95%Conf. | Interval] |
| incom16 | -0.030 | 0.031 | -0.970 | 0.333 | -0.091 | 0.031 |
| \_cons | 1.675 | 0.084 | 20.010 | 0.000 | 1.510 | 1.840 |

|  |  |  |
| --- | --- | --- |
| -> race = asian | | |
| Source | SS | df | | MS | Number of obs = 87 | | | |
| Model | 0.274 | 1 | | 0.274 | Prob>F = 0.280 | | | |
| Residual | 19.703 | 85 | | 0.232 | R-squared = 0.014 | | |
| Total | 19.977 | 86 | | 0.232 | Root MSE = 0.481 | | | | |
| numrooms | Coef. | Std. Err. | | t | P>t | [95% Conf. Interval] | | |
| incom16 | -0.057 | 0.052 | | -1.090 | 0.280 | -0.160 | 0.047 |
| \_cons | 1.674 | 0.144 | | 11.640 | 0.000 | 1.388 | 1.960 |
|  | | |

-> race = hispanic

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | Number of obs = 71 | | | |
| Model | 0.243 | 1 | 0.243 | Prob>F = 0.275 | | | |
| Residual | 13.842 | 69 | 0.201 | R-squared = 0.017 | | |
| Total | 14.086 | 70 | 0.201 | Root MSE = 0.448 | | | | |
| numrooms | Coef. | Std.Err. | t | P>t | [95%Conf. | Interval] |
| incom16 | 0.069 | 0.063 | 1.100 | 0.274 | -0.056 | 0.194 |
| \_cons | 1.295 | 0.159 | 8.130 | 0.000 | 0.977 | 1.613 |

# APPENDIX E

**POST ESTIMATION DIAGNOSTICS FOR MULTIVARIATE REGRESSION**

**Table 19: VIF Output**

|  |  |  |
| --- | --- | --- |
| Real Income | VIF | 1/VIF |
| Born | 1.230 | 0.810 |
| Black | 1.210 | 0.829 |
| Asian | 1.050 | 0.955 |
| Hispanic | 1.030 | 0.974 |
| Sex | 1.010 | 0.992 |
| Mean VIF | 1.090 |
| # of Rooms | VIF | 1/VIF |
| Born | 1.230 | 0.810 |
| Black | 1.210 | 0.829 |
| Asian | 1.050 | 0.955 |
| Hispanic | 1.030 | 0.974 |
| Sex | 1.010 | 0.992 |
| Mean VIF | 1.090 |

**Table 20: Correlations of Residuals and Breusch-Pagan Test**

|  |
| --- |
|  |
| Correlation matrix of residuals: |
|  |
| realinc numrooms |
| realinc 1.0000 |
| numrooms 0.3350 1.0000 |
|  |
| Breusch-Pagan test of independence: chi2(1) = 175.767, Pr = 0.0000 |
|  |