Analyzing the Effect of Sex and Education on Income using a Factorial ANOVA Test

To answer the research questions 'Do sex and education each have an effect on income?' and 'Do highly educated men earn more than highly educated women?', a 2014 dataset on higher education, gender, and income compiled by the Pew Research Center will be used to conduct a full-factorial two-way ANOVA. With the ANOVA results, an omnibus significance test and multiple corrected pairwise t-tests would be conducted and evaluated with respect to the effect size of the effect data and limitations of the tests performed to come to a conclusion to answer the initial research question. Hence, the dependent variable would be a ratio value of each person's approximate income while the independent variables would consist of nominal values for the sex and level of education of each person.

Full-Factorial Two-Way ANOVA

Main Effect – Education:

Null Hypothesis (H_0) : All population means (μ) of income for each level of education would be equal

$H_0 : \mu_{less\ than\ highschool} = \mu_{highschool\ grad} = \mu_{college}$

Alternative Hypothesis (H_a) : At least one population mean (μ) income for a level of education is different

Main Effect – Sex:

Null Hypothesis (H_0) : Male and female population means (μ) of income are equal

$H_0 : \mu_{male} = \mu_{female}$

Alternative Hypothesis (H_a) : Male and female population mean (μ) incomes are not equal

Interaction – Sex X Education:

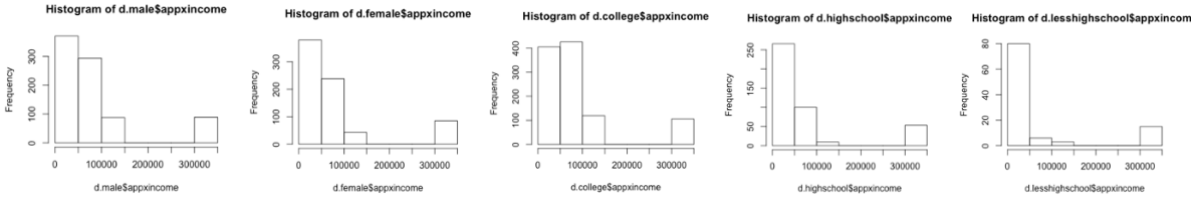
Null Hypothesis (H_0) : Difference between male and female population means (μ) are equal across conditions

Alternative Hypothesis (H_a) : At least one difference between male and female population mean (μ) incomes is not equal across conditions.

Assumptions for Full-Factorial Two-Way ANOVA:

- Normality: Tested using histograms for income for each sex and education level as shown below

Male | Female | College | Highschool Grad | Less Than Highschool



- Homogeneity of Variance: Levene's Test with p-values of 0.7579 and 0.9541 for variance among education levels and sex respectively, failing to reject the null hypothesis and hence, the results are inconclusive.
- Independence of Observations: While each income data recorded is unique to each person, the observations are technically dependent on other variables such as sex and education and hence, not independent.

Summary Table:

Categories	Df	Sum Sq	Mean Sq	Test Statistic F	p-value
Sex	1	13,650,000,000	13,650,000,000	1.376	0.24092
Education	2	103,100,000,000	51,560,000,000	5.197	0.00563
Sex:Education (Interaction)	2	81,690,000,000	40,850,000,000	4.117	0.01646
Error	1582	15,690,000,000,000	9,921,000,000	-	-
Total	1,587	15,888,440,000,000	-	-	-

Effect Size: $Overall\ R^2$ = 0.01248958362, $Sex\ \eta^2$ = 0.0008591151806, $Education\ \eta^2$ = 0.006488994514, $Sex:Education\ \eta^2$ = 0.005141473927

Bonferroni Correction: Pairwise $\alpha_{sex:education} = \frac{0.05}{15} = 0.003333$

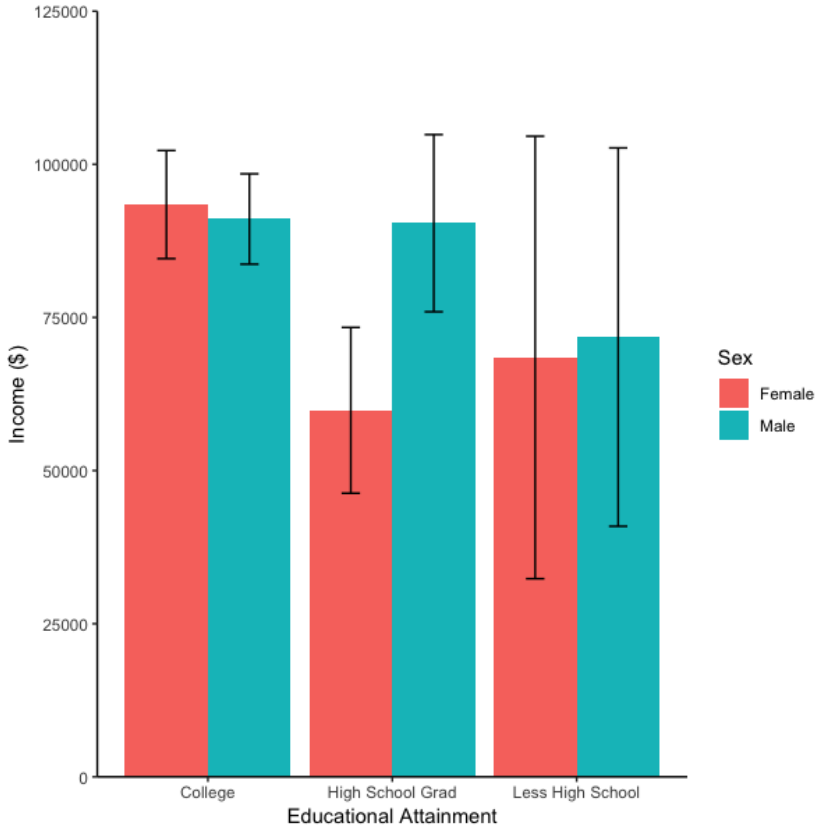
Corrected Pairwise t-tests:

Sex:Education	Female:Less Than High School	Female:High School Grad	Female: College	Male:Less Than High School	Male:High School Grad
Female:High School Grad	0.9955491	-	-	-	-
Female:College	0.6013950	0.0012111	-	-	-
Male:Less Than High School	0.9999814	0.9661573	0.6034768	-	-
Male:High School Grad	0.7605591	0.0210186	0.9988057	0.7877326	-
Male:College	0.6977587	0.0032860	0.9989145	0.7134713	0.9999992

95% Confidence Intervals:

Genre	Mean Approx. Income	SD Approx. Income	Sample Size	$t_{critical}$	Upper Interval	Lower Interval
Female:Less Than High School	68465.91	118792.6	44	±2.016692	104582.1579	32349.66212
Female:High School Grad	59837.84	93309.33	185	±1.972941	73372.68557	46302.99443
Female:College	93428.43	102111.2	517	±1.964572	102251.0192	84605.84079
Male:Less Than High School	71791.67	119560.5	60	±2.000995	102677.4164	40905.92362
Male:High School Grad	90390.95	114438.3	243	±1.969815	104851.796	75930.104
Male:College	91071.43	87223.94	539	±1.964383	98451.61902	83691.24098

Bar Graph Visualization of the Effect of Sex and Educational Attainment on Income:



As per the results above, the dataset for approximate income violates all assumptions for the two-way ANOVA since the mean approximate incomes are not approximately normally distributed, there is no homogeneity of variances, and independence of observations does not exist. Since the p-values are 0.24092 for sex, 0.00563 for education, and 0.01646 for the interaction between them, the result is statistically significant ($p < 0.05$) for education and the interaction between sex and education only. Hence, the null hypothesis is only rejected for the main effect of education and the interaction which means at least one of the population mean incomes for an education level is different and at least one difference between male and female population mean incomes is not equal across conditions. Since F is relatively big with values of 5.197 and 4.117, the differences in means amongst groups of educational attainment levels and sex-education interactions is unlikely to be due to sampling errors.

As overall R^2 is 0.01248958362, the effect size, which is the strength of the phenomenon, is very small and shows that all observations only explains 1.25% of the variance in income. Moreover, with $Education\ \eta^2$ = 0.006488994514 and $Sex{:}\,Education\ \eta^2$ = 0.005141473927, education explains 0.65% and the interaction explains 0.51% of the variance. As observed by the 95% confidence intervals table and graph, the significant difference in population means for approximate income for each interaction between education and sex can be observed. However, the error bars for the education level of less than high school are larger signifying greater uncertainty within that domain which is a limitation for our hypothesis tests. Hence, more income data resulting in greater and equal sample sizes among education levels would be required to solve this limitation.

Thus, in response to the research questions, our results provide substantial evidence to prove that education has a significant effect on income while sex does not. In addition, our results show that highly educated men earn more than highly educated women generally until and as 'High School Grads'. On the other hand, females in college actually earn more than men in college on average as per our graph's results however, since sex was found to be a non-significant factor toward determining an individual's income, this slight difference may not truly be the case. Additionally, the difference is miniscule when compared to the income disparity between the two sexes as high school graduates. Meanwhile, the probability of getting at least 1 Type I error across all tests is = 0.5367 which represents a familywise error rate of 53.7%. However, for the purposes of the Bonferroni correction which resulted in a pairwise $\alpha_{sex:education}$ of 0.003333, familywise α is assumed to be 0.05 since we aim to preserve the Type I error rate across all hypothesis tests. Based on the corrected pairwise t-tests, only the highlighted pairwise t-test results on the table, which are Female:College-Female:High School Grad and Male:College-Female:High School Grad, are statistically significant under the pairwise $\alpha_{sex:education}$ and reject the H_0.

Works Cited

Suh, Michael. "Higher Education, Gender & Work." *Pew Research Center's Social & Demographic Trends Project*, 22 Dec. 2014, https://www.pewsocialtrends.org/2014/12/22/higher-education-gender-work/.