Assignment #2: Module 4 Problem Set

Charles Titus

Grand Canyon University: RES – 866

September 24, 2019

**Module 4 Problem Set**

**7.14 Learning Activity**

1. **We want to see whether men and women differ in their mean socioeconomic index (sei) and their age when their first child was born (agekdbrn). First, use the Explore procedure to view the distributions of these two variables by gender. Are they similar or different? Do you see any problems with doing a t test?**

Below the charts and graphs are listed and they show the descriptive statistics for the socioeconomic index (sei) and age when the child who was born first (agekdbrn) for both males and females. The socioeconomic index scores summarize the differences in prestige between occupations. The sei range for males was 78.9 with a maximum of 96 and a minimum being 17.1. The male average of SEI was 49.039. This would be considered at the bottom part of the range. On the other hand, females had an sei range of 80.1 with a maximum of 97.2 and a minimum of 17.1. The women’s minimum was the same as the males while the maximum was higher for women. The female average of sei was 47.184 and this is also in the bottom half of the range just like the males.

When looking at the average age of people when they had their first child it is clear to see that female’s ages ranged from 13 to 44 with the average being 23.37. The average of males when they had their first child was 14 to 43 with the average being 25.43.

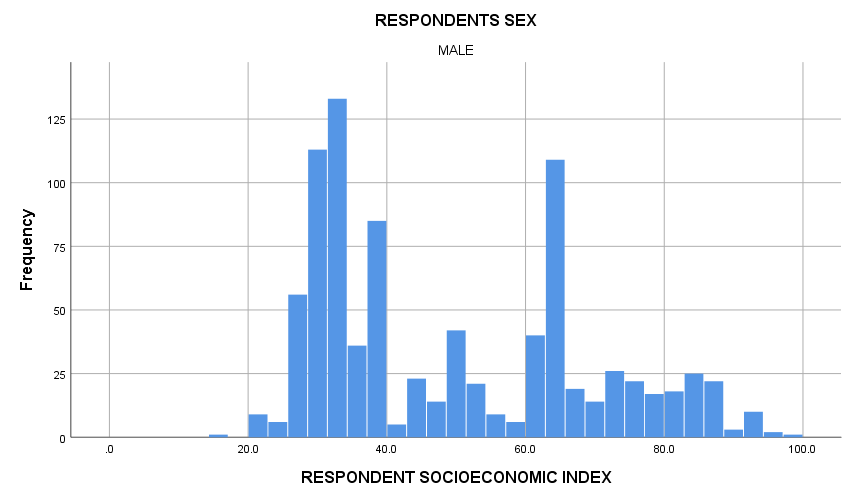
**Do you see any problems with doing a t test?**

As a result of the above information it can be determined that there are no issues with carrying out a t-test on the samples. The reason is because the samples t-test can be applied when there are two different populations that can be compared on a dependent variable (IBM, 2010).

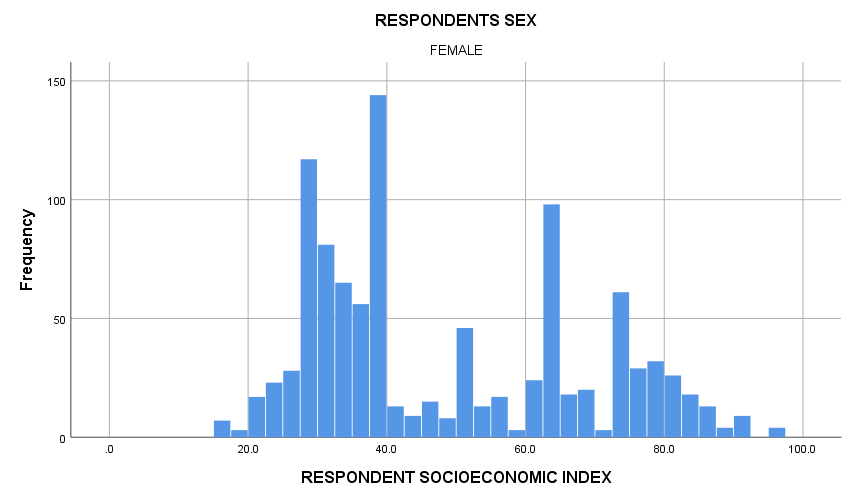
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Case Processing Summary** | | | | | | | | |
|  | RESPONDENTS SEX | Cases | | | | | | |
|  | Valid | | Missing | | Total | |
|  | N | Percent | N | Percent | N | Percent |
| RESPONDENT SOCIOECONOMIC INDEX | MALE | 598 | 64.4% | 331 | 35.6% | 929 | 100.0% |
| FEMALE | 817 | 74.7% | 277 | 25.3% | 1094 | 100.0% |
| R'S AGE WHEN 1ST CHILD BORN | MALE | 598 | 64.4% | 331 | 35.6% | 929 | 100.0% |
| FEMALE | 817 | 74.7% | 277 | 25.3% | 1094 | 100.0% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptives** | | | | | |
|  | RESPONDENTS SEX | | | Statistic | Std. Error |
| RESPONDENT SOCIOECONOMIC INDEX | MALE | Mean | | 49.039 | .7894 |
| 95% Confidence Interval for Mean | Lower Bound | 47.489 |  |
| Upper Bound | 50.589 |  |
| 5% Trimmed Mean | | 48.192 |  |
| Median | | 43.400 |  |
| Variance | | 372.604 |  |
| Std. Deviation | | 19.3030 |  |
| Minimum | | 17.1 |  |
| Maximum | | 96.0 |  |
| Range | | 78.9 |  |
| Interquartile Range | | 32.0 |  |
| Skewness | | .510 | .100 |
| Kurtosis | | -1.001 | .200 |
| FEMALE | Mean | | 47.184 | .6745 |
| 95% Confidence Interval for Mean | Lower Bound | 45.860 |  |
| Upper Bound | 48.508 |  |
| 5% Trimmed Mean | | 46.521 |  |
| Median | | 38.400 |  |
| Variance | | 371.655 |  |
| Std. Deviation | | 19.2784 |  |
| Minimum | | 17.1 |  |
| Maximum | | 97.2 |  |
| Range | | 80.1 |  |
| Interquartile Range | | 32.2 |  |
| Skewness | | .564 | .086 |
| Kurtosis | | -.976 | .171 |
| R'S AGE WHEN 1ST CHILD BORN | MALE | Mean | | 25.00 | .221 |
| 95% Confidence Interval for Mean | Lower Bound | 24.56 |  |
| Upper Bound | 25.43 |  |
| 5% Trimmed Mean | | 24.72 |  |
| Median | | 24.00 |  |
| Variance | | 29.280 |  |
| Std. Deviation | | 5.411 |  |
| Minimum | | 14 |  |
| Maximum | | 46 |  |
| Range | | 32 |  |
| Interquartile Range | | 7 |  |
| Skewness | | .819 | .100 |
| Kurtosis | | .735 | .200 |
| FEMALE | Mean | | 23.01 | .181 |
| 95% Confidence Interval for Mean | Lower Bound | 22.66 |  |
| Upper Bound | 23.37 |  |
| 5% Trimmed Mean | | 22.72 |  |
| Median | | 22.00 |  |
| Variance | | 26.644 |  |
| Std. Deviation | | 5.162 |  |
| Minimum | | 13 |  |
| Maximum | | 44 |  |
| Range | | 31 |  |
| Interquartile Range | | 7 |  |
| Skewness | | .869 | .086 |
| Kurtosis | | .439 | .171 |

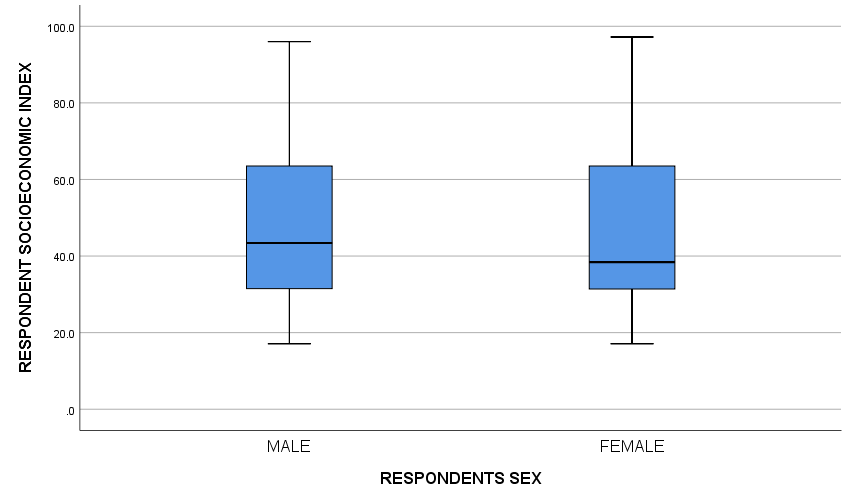
**Graphs:**

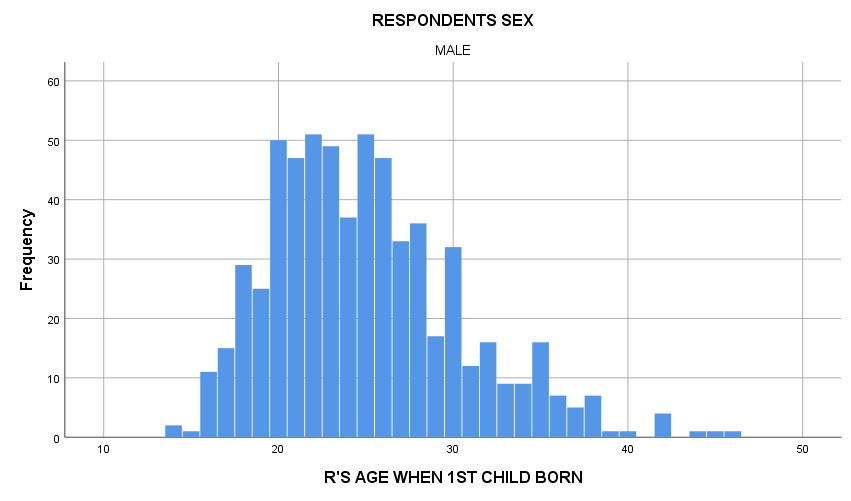


Mean: 49.04 Std Dev: 19.303 N= 598

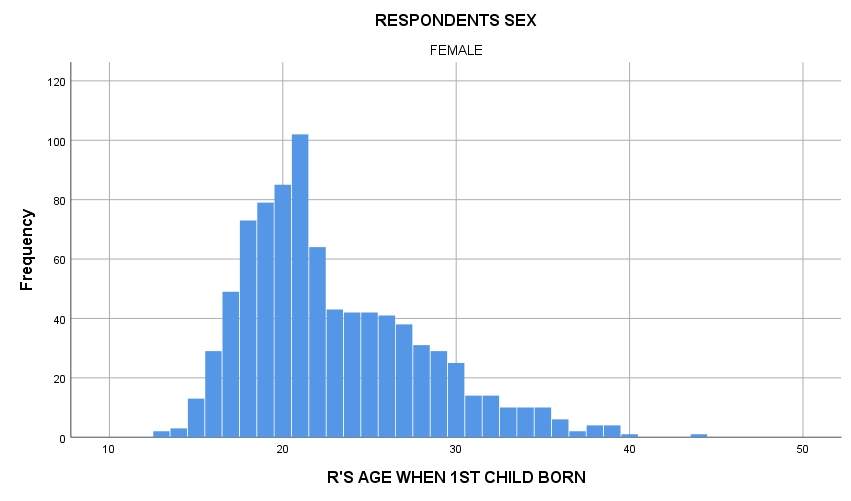


Mean: 47.18 Std Dev: 19.278 N = 817





Mean: 25 Std Dev- 5.411 N = 598



Mean: 23.01 Std Dev: 5.162 N = 817



1. **Now do a t test for each variable, by gender.** (See Below)

T-Test

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | RESPONDENTS SEX | N | Mean | Std. Deviation | Std. Error Mean |
| RESPONDENT SOCIOECONOMIC INDEX | MALE | 887 | 49.109 | 19.4399 | .6527 |
| FEMALE | 1024 | 48.458 | 19.5677 | .6115 |
| R'S AGE WHEN 1ST CHILD BORN | MALE | 623 | 25.00 | 5.444 | .218 |
| FEMALE | 866 | 22.87 | 5.128 | .174 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Independent Samples Test** | | | | | | | | | | |
|  | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| RESPONDENT SOCIOECONOMIC INDEX | Equal variances assumed | .256 | .613 | .728 | 1909 | .467 | .6515 | .8948 | -1.1034 | 2.4065 |
| Equal variances not assumed |  |  | .728 | 1873.685 | .466 | .6515 | .8944 | -1.1026 | 2.4057 |
| R'S AGE WHEN 1ST CHILD BORN | Equal variances assumed | .724 | .395 | 7.704 | 1487 | .000 | 2.130 | .276 | 1.588 | 2.672 |
| Equal variances not assumed |  |  | 7.629 | 1291.153 | .000 | 2.130 | .279 | 1.582 | 2.677 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test of Homogeneity of Variances** | | | | | |
|  | | Levene Statistic | df1 | df2 | Sig. |
| RESPONDENT SOCIOECONOMIC INDEX | Based on Mean | .256 | 1 | 1909 | .613 |
| Based on Median | .341 | 1 | 1909 | .559 |
| Based on Median and with adjusted df | .341 | 1 | 1851.970 | .559 |
| Based on trimmed mean | .241 | 1 | 1909 | .624 |
| R'S AGE WHEN 1ST CHILD BORN | Based on Mean | .724 | 1 | 1487 | .395 |
| Based on Median | 1.867 | 1 | 1487 | .172 |
| Based on Median and with adjusted df | 1.867 | 1 | 1483.393 | .172 |
| Based on trimmed mean | 1.162 | 1 | 1487 | .281 |

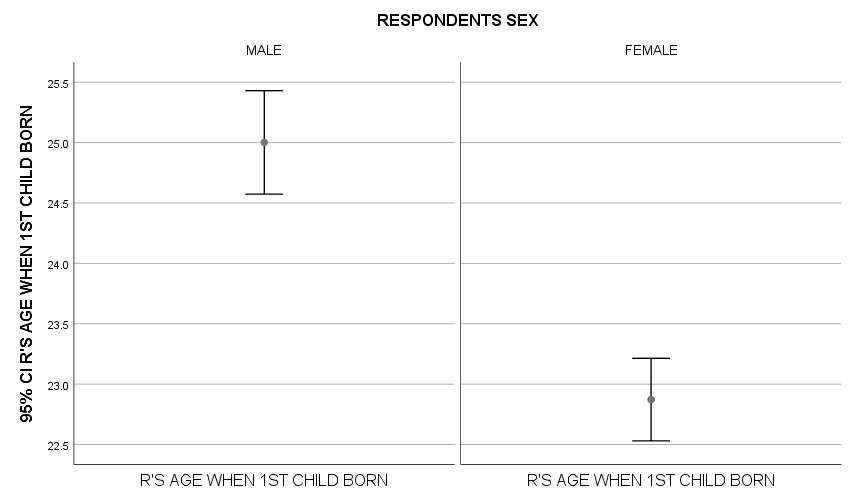
**Is the homogeneity of variance assumption met, or not?**

In order to determine the homogeneity of variance assumption an independent samples t-test was carried out. By carrying this action out it is clear to see that the level of significance of the independent samples is .000 which is very low for the age of when the first child was born. This shows that the F value is statistically significant. By looking at the test for homogeneity of variances one can see that there is a high level of significance which is .613 and .395. This data would mean that the variables are not statistically significant as a result.

**What do you conclude about mean differences by gender?**

When looking at the gender it is clear to see the following: One can make a conclusion that there is indeed a significant difference between the sei of males and females than their age when their child was first born.

1. **Create an error bar chart for each variable by gender.**





**Is the graph consistent with the result from the t test?**

By looking at the graph and the t test it is clear to see that there is a consistency between the two. This can be seen by looking at the data and seeing that there are many similarities amongst them.

**8.10 Learning Activity**

1. **One variable in the customer survey asked about agreement that SPSS products are a good value (gdvalue). A second question asked about agreement that SPSS offers high quality products (hiqualty). Use a paired-samples t test to see whether the means of these two questions differ (they are measured on a five-point scale). What do you conclude?**

**See data below:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Paired Samples Statistics** | | | | | |
|  | | Mean | N | Std. Deviation | Std. Error Mean |
| Pair 1 | SPSS prods are a gd val | 2.66 | 890 | 1.055 | .035 |
| SPSS offers hi-quality prods | 2.08 | 890 | .793 | .027 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Paired Samples Correlations** | | | | |
|  | | N | Correlation | Sig. |
| Pair 1 | SPSS prods are a gd val & SPSS offers hi-quality prods | 890 | .451 | .000 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Paired Samples Test** | | | | | | | | | |
|  | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Pair 1 | SPSS prods are a gd val - SPSS offers hi-quality prods | .579 | .993 | .033 | .513 | .644 | 17.381 | 889 | .000 |

**What do you conclude?**

It can be concluded that the researchers can accept the hypothesis that customers do indeed feel SPSS is a high-quality product and it is a good value as well. The way that this can be concluded is that the results of the paired samples t-test actually show that the average for good value is 2.66. The average response for high-quality is 2.08. The statistical significance of .000 is what allows one to accept the hypothesis that is presented.

1. **Then test whether there is a mean difference between agreement that SPSS products are easy to learn (easylrn) and SPSS products are easy to use (easyuse).**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Paired Samples Statistics** | | | | | |
|  | | Mean | N | Std. Deviation | Std. Error Mean |
| Pair 1 | SPSS prods are easy to learn | 2.62 | 900 | 1.050 | .035 |
| SPSS prods are easy to use | 2.63 | 900 | 1.050 | .035 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Paired Samples Correlations** | | | | |
|  | | N | Correlation | Sig. |
| Pair 1 | SPSS prods are easy to learn & SPSS prods are easy to use | 900 | .811 | .000 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Paired Samples Test** | | | | | | | | | |
|  | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Pair 1 | SPSS prods are easy to learn - SPSS prods are easy to use | -.006 | .646 | .022 | -.048 | .037 | -.258 | 899 | .796 |

**What do you conclude?**

One can conclude by looking at this data that the researchers can indeed accept the hypothesis that customers feel that SPSS is indeed average when using the variable of easy to learn and easy to use. This can be determined based upon the paired t-test that shows that the average for the easy to learn is 2.62 and the average for easy to use is 2.63 with the statistical significance of .000.

1. **Could we use a paired-sample t test to compare how long a customer has used SPSS products (usespss) and how frequently they use SPSS (freqspss)? Why or why not?**

One could use a paired t test to compare how long a customer has used SPSS products and how frequently they use it. The reason is because the variables have met the assumptions of justifications that are pointed out by IBM in 2010. Both variables have an ordinal scale of measurement, and both variables are from the same sample group. As a result of both of these it means that a paired-sample t test can be used in the situation at hand.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| How long have you used SPSS products? | .197 | 935 | .000 | .873 | 935 | .000 |
| How frequently do you use SPSS | .189 | 935 | .000 | .914 | 935 | .000 |
|  | | | | | | |

**9.20 Learning Activity**

1. Investigate how the number of siblings (sibs) varies by highest degree (degree). Ask for appropriate statistics.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Descriptives** | | | | | | | | |
| NUMBER OF BROTHERS AND SISTERS | | | | | | | | |
|  | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
| Lower Bound | Upper Bound |
| Less than primary/high school | 297 | 5.73 | 4.691 | .272 | 5.19 | 6.26 | 0 | 55 |
| Primary/High School | 1002 | 3.76 | 2.950 | .093 | 3.58 | 3.94 | 0 | 37 |
| Post Primary/High School | 173 | 3.38 | 2.720 | .207 | 2.97 | 3.79 | 0 | 17 |
| BACHELOR | 354 | 2.47 | 1.928 | .102 | 2.26 | 2.67 | 0 | 14 |
| GRADUATE | 194 | 2.40 | 1.755 | .126 | 2.15 | 2.65 | 0 | 8 |
| Total | 2020 | 3.66 | 3.187 | .071 | 3.52 | 3.80 | 0 | 55 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test of Homogeneity of Variances** | | | | | |
|  | | Levene Statistic | df1 | df2 | Sig. |
| NUMBER OF BROTHERS AND SISTERS | Based on Mean | 31.134 | 4 | 2015 | .000 |
| Based on Median | 27.494 | 4 | 2015 | .000 |
| Based on Median and with adjusted df | 27.494 | 4 | 1388.138 | .000 |
| Based on trimmed mean | 29.668 | 4 | 2015 | .000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ANOVA** | | | | | |
| NUMBER OF BROTHERS AND SISTERS | | | | | |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 2106.493 | 4 | 526.623 | 57.653 | .000 |
| Within Groups | 18405.814 | 2015 | 9.134 |  |  |
| Total | 20512.307 | 2019 |  |  |  |

1. **Is the assumption of homogeneity of variance met? Is the ANOVA test significant at the .01 level?**

By looking at the table above it indicates that the p-value is less compared to a 0.01 level of significance. This means that the variances are not homogeneous. By looking at the ANOVA test it shows that the p-value is significant.

1. **Do a post hoc analysis, if justified. Ask for both the Bonferroni and Scheffe tests? What do you conclude from these tests? Which education groups have different mean numbers of children? Are the Bonferroni and Scheffe tests consistent?**

Based on the results below the majority of the mean differences are significant at a 0.05 level of significance. The education groups that have different mean numbers of children are primary/high school and post primary/high school as compared to the bachelor and graduate categories. The two analyses do indeed appear to be consistent because both provide similar results to one another.

**Post Hoc Tests**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Multiple Comparisons** | | | | | | | |
| Dependent Variable: NUMBER OF BROTHERS AND SISTERS | | | | | | | |
|  | (I) RS HIGHEST DEGREE | (J) RS HIGHEST DEGREE | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|  | Lower Bound | Upper Bound |
| Scheffe | Less than primary/high school | Primary/High School | 1.969\* | .200 | .000 | 1.35 | 2.58 |
| Post Primary/High School | 2.346\* | .289 | .000 | 1.45 | 3.24 |
| BACHELOR | 3.261\* | .238 | .000 | 2.53 | 3.99 |
| GRADUATE | 3.330\* | .279 | .000 | 2.47 | 4.19 |
| Primary/High School | Less than primary/high school | -1.969\* | .200 | .000 | -2.58 | -1.35 |
| Post Primary/High School | .377 | .249 | .682 | -.39 | 1.14 |
| BACHELOR | 1.292\* | .187 | .000 | .72 | 1.87 |
| GRADUATE | 1.362\* | .237 | .000 | .63 | 2.09 |
| Post Primary/High School | Less than primary/high school | -2.346\* | .289 | .000 | -3.24 | -1.45 |
| Primary/High School | -.377 | .249 | .682 | -1.14 | .39 |
| BACHELOR | .915\* | .280 | .031 | .05 | 1.78 |
| GRADUATE | .985\* | .316 | .046 | .01 | 1.96 |
| BACHELOR | Less than primary/high school | -3.261\* | .238 | .000 | -3.99 | -2.53 |
| Primary/High School | -1.292\* | .187 | .000 | -1.87 | -.72 |
| Post Primary/High School | -.915\* | .280 | .031 | -1.78 | -.05 |
| GRADUATE | .069 | .270 | .999 | -.76 | .90 |
| GRADUATE | Less than primary/high school | -3.330\* | .279 | .000 | -4.19 | -2.47 |
| Primary/High School | -1.362\* | .237 | .000 | -2.09 | -.63 |
| Post Primary/High School | -.985\* | .316 | .046 | -1.96 | -.01 |
| BACHELOR | -.069 | .270 | .999 | -.90 | .76 |
| Bonferroni | Less than primary/high school | Primary/High School | 1.969\* | .200 | .000 | 1.41 | 2.53 |
| Post Primary/High School | 2.346\* | .289 | .000 | 1.53 | 3.16 |
| BACHELOR | 3.261\* | .238 | .000 | 2.59 | 3.93 |
| GRADUATE | 3.330\* | .279 | .000 | 2.55 | 4.11 |
| Primary/High School | Less than primary/high school | -1.969\* | .200 | .000 | -2.53 | -1.41 |
| Post Primary/High School | .377 | .249 | 1.000 | -.32 | 1.08 |
| BACHELOR | 1.292\* | .187 | .000 | .77 | 1.82 |
| GRADUATE | 1.362\* | .237 | .000 | .70 | 2.03 |
| Post Primary/High School | Less than primary/high school | -2.346\* | .289 | .000 | -3.16 | -1.53 |
| Primary/High School | -.377 | .249 | 1.000 | -1.08 | .32 |
| BACHELOR | .915\* | .280 | .011 | .13 | 1.70 |
| GRADUATE | .985\* | .316 | .019 | .10 | 1.87 |
| BACHELOR | Less than primary/high school | -3.261\* | .238 | .000 | -3.93 | -2.59 |
| Primary/High School | -1.292\* | .187 | .000 | -1.82 | -.77 |
| Post Primary/High School | -.915\* | .280 | .011 | -1.70 | -.13 |
| GRADUATE | .069 | .270 | 1.000 | -.69 | .83 |
| GRADUATE | Less than primary/high school | -3.330\* | .279 | .000 | -4.11 | -2.55 |
| Primary/High School | -1.362\* | .237 | .000 | -2.03 | -.70 |
| Post Primary/High School | -.985\* | .316 | .019 | -1.87 | -.10 |
| BACHELOR | -.069 | .270 | 1.000 | -.83 | .69 |
| \*. The mean difference is significant at the 0.05 level. | | | | | | | |

**Homogeneous Subsets**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **NUMBER OF BROTHERS AND SISTERS** | | | | | |
|  | RS HIGHEST DEGREE | N | Subset for alpha = 0.05 | | |
|  | 1 | 2 | 3 |
| Scheffea,b | GRADUATE | 194 | 2.40 |  |  |
| BACHELOR | 354 | 2.47 |  |  |
| Post Primary/High School | 173 |  | 3.38 |  |
| Primary/High School | 1002 |  | 3.76 |  |
| Less than primary/high school | 297 |  |  | 5.73 |
| Sig. |  | .999 | .709 | 1.000 |
| Means for groups in homogeneous subsets are displayed. | | | | | |
| a. Uses Harmonic Mean Sample Size = 275.864. | | | | | |
| b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed. | | | | | |

**References**

IBM, (2010). *Introduction to Statistical Analysis Using IBM SPSS Statistics* (Student Guide).