Statistics for Social Research II

David D. Reedy

Omega Graduate School

March 6, 2025

Professor

Sean Taladay, Ed.D

Assignment #3 – Essay

Answer the following questions in an essay format, with 1-2 fully developed paragraphs for

each question. Include citations/references from your Developmental Reading log.

1. What is normality? Why is it relevant to parametric (generalizable to the target

population) versus nonparametric (applicable only to the sample) statistics in social

science research?

2. What is a histogram? What is a box and whisker plot? How are they useful for

understanding the normality of a given sample?

3. Describe the following statistical procedures:

a. Quasi-Experimental: t-Tests (Independent and Dependent) - Parametric

b. Quasi-Experimental: Mann Whitney U and Wilcoxon - Nonparametric

c. Correlational: Pearson’s r - Parametric

d. Correlational: Spearman’s Rank - Parametric

4. Navigate to OGS’s Practical Statistics for Social Research (PSSR) tool. Click on

“Example Datasets” and load the “Independent t-Test: Ethical Decision-Making”

dataset.

5. Scroll to “Step Three: Run Statistical Procedures” and click “t-Test”.

6. Copy and paste the output's contents into your assignment document. Read it carefully

and expand on it based on your understanding. Answer the following questions:

a. What might be good problem and purpose statements for this dataset?

b. What might be good research questions related to the hypotheses generated

by the PSSR software?

c. What does the output tell you about the comparison of the two groups?

d. How was the p-value used to test the hypotheses?

7. Repeat steps 10-12 for the “Correlational: Life Satisfaction Index” dataset. Note that

this is a correlational design with two continuous variables. What does the scatterplot

tell you about the relationship between the two variables?

8. Repeat steps 10-12 for both datasets, but use the nonparametric equivalents of the

statistical tests (Mann Whitney U and Spearman’s Rank, respectively). How did the

results change?

9. For each dataset, click on “Assumptions” under “Step Two: Run Descriptives and

Assumptions.” Based on the output of each dataset, should you use parametric or

nonparametric procedures? Are the datasets normally distributed? Note that this is a

judgment call on behalf of the researcher and not a black-and-white decision.

10. Finally, navigate to OGS’s Practical Statistics for Social Research (PSSR) tool. Click

on “Example Datasets” and load the “Example: Perfect Correlation” dataset. Scroll to

“Step Three: Run Statistical Procedures” and click “Linear Regression.” What does the

scatterplot graph show you about the relationship between the X and Y variables?

11. Summarize what you learned from conducting these statistical tests.

# Part 1: Normality

What is normality? Why is it relevant to parametric (generalizable to the target

population) versus nonparametric (applicable only to the sample) statistics in social

science research?

Scientists have learned that many aspects of human life, including physical and behavioral aspects, approximate normal distributions. This indicates that when measurements of a human population are placed on a graph they will normally assume a bell shaped curve, with the mean, median, and mode having the save value and the values being distributed symmetrically around the mean in both directions. (Frey, 2022, p. 1088-1089; Hahs-Vaughn, 2020, p. 121) This gives rise to the central limit theorem which explains that as a sample size grows closer to the size of the population the sampling distribution will increasingly take on this shape, the “normal distribution.” (Willard, 2020, p. 94)

In statistics the area under the normal curve established by a set of observations is directly related to the probability that additional observational values will likely continue to fall within the same area making the prediction of the likelihood of an event or observation occurring as likely. (p. 92) This phenomenon allows the researcher to compare two or more normal distributions including allowing the examination of a sample to understand traits that are likely reflective of the wider population. (Hahs-Vaugh, 2020, pp. 120-121) Parametric procedures are designed around assumptions of normality to evaluate data observations between two or more variables for changes that are not likely due to chance. When such change is statistically significant in a sample indicated by parametric analysis, the normality of the data allows for conclusions drawn regarding the sample to likely apply also to the population. (Frey, pp. 1089-1093) When normality is violated, non-parametric models must be utilized, and one cannot assume the measurements of the sample will necessarily reflect a wider population. (p. 1094)

A researcher can examine the data for normality in a variety of ways including using graphs to determine if the distribution has a bell-shaped curve or utilizing a histogram or a box and whisker plot. (p. 1093)

# Part Two: Histograms and Box and Whisker Plots

A histogram is a visual graphing method using a bar to display a count or frequence where the independent variable consisting of interval or ratio data is displayed on the x-axis and the frequency is shown on the y-axis. The height of each bar represents the count. The data may be grouped into bins, for example age range, with the bar width corresponding to the bin size. Unlike a traditional bar chart which is based on nominal or ordinal data and has separated bars, histograms use continuous data with adjacent bars. (Frey, p. 666) Histograms are not only useful for the researcher to access normality but for conveying information from the research to others. Thus, using a scale on the y-axis that accommodates the biggest bin in easy to understand increments, such as divisions of 5 for example, allows for clear visualization. (p. 668) Histograms effectively display key distribution attributes such as shape, symmetry indicating balanced or skewed tails, and modality indicating if the data is unimodal, bimodal or multimodal. Visual inspection assists the researcher when determining if the data approximates normality. (p. 668)

A box and whisker plot is a visual graphical representation of a data set that represents the distribution’s range, symmetry, and central tendency. It is based on the five-number summary of the values including: the minimum, the first quartile (Q1), median, the third quartile (Q3) and the maximum. Outliers are also indicated. The graph visually illustrates both variability and value concentrations. (p. 148) The box plot is constructed by ordering the data from smallest to largest and then calculating the median, the lower quartile, the upper quartile, the minimum and maximum values. The interquartile range is calculated by subtracting Q1 from Q3 establishing the upper and lower fences that help identify outliers. Whiskers are drawn with short horizontal lines at the minimum and maximum values within the fence and are connected to the box with vertical lines. Outliers are then marked. (p. 149)

A box plot provides five key inferences: First, the median can be identified by the line in the middle of the box. Second, the variability of the data is reflected in the box's length, with longer boxes indicating greater variability. Third, the symmetry of the middle 50% of the data is assessed by the position of the median within the box; if centered, the distribution is symmetrical. Fourth, outliers are indicated by asterisks outside the whiskers. Fifth, skewness is identified by the relative length of the box halves; longer lower or upper halves suggest left or right skewness, respectively. If skewness is suspected, further investigation using a histogram is recommended. (p. 150, 1093)

# Part Three: Statistical Procedures

 A t-test describes a group of parametric tests that can be utilized with normal data to determine if the means of one or more group are statistically different helping the researcher to assess if observed differences are likely due to change or are probably capturing a true difference or change in the sample or population. (Frey, p. 604) The t-test is useful because it does not require complete understanding of a population’s variability thus is used when the standard deviation is not known. (Willard, p. 133) If a researcher is looking for change in one direction, for instance a value increases or decreases, a one-tailed test is utilized. If the researcher is looking for an effect without an indication of direction a two-tailed test is utilized. (p. 138) The statistical probabilities are examined for significance against a table of critical values that are necessary for rejecting the null hypothesis, that is concluding that a change is observed and it is likely not due to chance. (p. 134)

A one-sample t-test is used to determine if the mean of a sample is significantly different from a known population mean, such as a bench mark, for example a passing grade of 70 among a group of students. The obtained “t-value” must be greater than the critical value and in the same direction as the alternative hypothesis to reject the null hypothesis. In other words, the change indicated by the t-value must be large enough and in the right direction to not be attributed to chance. (p. 136)

A two sample t-test applied to independent samples is used to compare the means between two groups, for instance when new medications are tested against one group that receives the medication and another a placebo. (p. 147) To determine if the difference in the means is statistically significant, the observed differences are evaluated using a theoretical concept regarding the sampling distribution of the difference between the means. This begins with drawing random samples form a larger population and calculating the means for the sample. If it were possible to repeat this for every single possible sample combination, the data could be arranged on a frequency chart. This distribution helps us understand what kinds of differences are normal, and by comparing our observed difference to it, we can determine if it is large enough to be considered statistically significant, meaning it is unlikely to have occurred by chance. (p. 147) The t-test helps to simulate the theoretical scenario of having tested every possible sample combination, but this requires that the data meet strict assumptions. Participants must be independent of one another and selected randomly. The dependent variable must be measured on an interval or ratio scale. The dependent variable must be normally distributed. The population variances of the dependent variable must be approximately equal. This is important because if the variances are drastically different, it will throw off the averages of the sample variances. (p. 152)

A related samples, or dependent t-test, is similar to the above but in this case compares the means of two related groups. Scores may come from the same subjects as in a repeated measures design or from subjects that are logically related such as husbands and wives in matched pairs. (p. 160) Here we compare the means between two related set of scores against a theoretical sampling distribution. In this design, we take repeated measures from the same subjects or related samples (like husbands and wives), calculate the difference between each pair of scores, and then compute the mean of these difference scores. These mean differences are arranged into a sampling distribution, called the sampling distribution of mean differences, to assess if the observed mean difference is significant. The standard deviation for this distribution is the standard error of the mean difference (SMD), and the calculations for the t-statistic and SMD are similar to those in the one-sample t-test, with only different symbols used because here we are working with repeated observations of a single mean. (pp. 160-161)

The Mann Whitney U test is the nonparametric equivalent of an independent t-test. It is utilized when data is not normally distributed. It is used when both sample sizes are 20 or fewer and produces a “U statistic”. If either sample size is greater than 20, the U statistic approximates a normal distribution allowing for the conversion of the U value into a z-value. The test is highly effective for comparing small populations examining two independent samples based on ranked scores. Rather than comparing the means as in a t-test, the Mann Whitney U test compares the difference in the ranks between the two groups. (p. 273) When the null hypothesis is true, and the ranks of the two samples come from the same population distribution, the mean rank of scores in both samples should be the same. However, if the independent variable influences the scores, it will influence their rank order, leading to different mean ranks between the two samples. (Frey, p. 850) This test is useful when key assumptions for the t-test are not met when the data can be arranged in ranks. A large U statistic suggests that the difference in rank between the two groups is not related to chance. (p. 850-851)

The Wilcoxon matched-pairs signed-ranks test is the nonparametric alternative to the related samples t-test. This is used when the data is ordinal or when the assumptions required for the dependent t-test are not met. It can be utilized for two research designs examining matched pairs where the pairs are logically related or a repeated measures design where the same participants are tested before and after an intervention. (Willard, p. 276) The test generates a Tobt (T-observed) value which is compared to a critical value (Tcrit) to evaluate if the observed difference is statistically significant enough to not be attributable to chance. (p. 276) The test does not utilize a formula but a process. If the researcher is looking for change in one direction then the alternative hypothesis will indicate a change greater than or less than zero. If the researcher is looking for change in a non-specified direction, the alternative hypothesis will read that the median difference is not zero. Differences are calculated by computing the difference between each paired observation. These differences are then ranked from smallest to largest taking care to note positive and negative differences. The ranks are separated into positive differences (W+) and negative differences (W-) with W+ being the sum of the ranks of the positive differences and W- being the sum of the ranks for negative differences. The test statistic is computed as the smaller of the two sums. The value represents the smallest rank sum from the positive or negative difference. To determine if the value is great enough to indicate if the change or difference is unlikely due to chance, for small samples below N=25 a Wilcoxon signed rank table is utilized. For larger samples, the value can be compared against a normal distribution. (pp. 276-277) The p-value is usually 0.05. (p. 277)

This test may very well fit with my hypothetical research project of examining the potential effectiveness of an educational intervention with the parents of GenZ young adults who have failed to launch. Supposing my testing group was less than 20 parental units I could measure different variables for the group before and after the education was provided. I could measure how confident parents felt to engage their children or if after a period of time a change in young adults’ educational, employment or social activity was observed. The measurements would form two sets of paired scores, one from before and one from after the intervention. For each group, the difference between the pre-intervention and the post-intervention would be calculated. Differences would be ranked based on their magnitude possibly using a Likert type instrument to gather scores related to attitude and confidence both in parents and children or gather specific counts such as number of young adults with a job or engaging in social activities outside the home. The scores would be ranked based on their value and separated into positive and negative differences where the post-educational scores were higher and where they were lower. The ranks would then be added together to create a W+ score and a W- score. The Wilcoxon test statistic would be calculated by taking the smaller of the two sums and comparing that to the critical value in the Wilcoxon table.

The above tests are all reflective of quasi-experimental designs seeking to examine causal effects in real-world settings. This involves manipulating variables. (Frey, pp. 1327-1328) Correlational designs are used to examine relationships between two or more variables to see if a relationship exists without manipulating the variables. (Cresswell and Cresswell, 2022, p. 5, 15) Pearson’s R is the parametric test which examines correlation. Spearman’s Rank is the nonparametric test for examining correlation.

Pearson’s R is a parametric test used to examine “…linear relationships between variables on interval or ratio scales.” This means that as one variable changes another variable changes at a constant rate. (Willard, p. 236) The coefficient of determination measures how much variance the two examined values share if they are correlated. It is calculated by squaring the correlation coefficient and represents the proportion of variability in one variable that can be explained by or predicted by the other variable. (pp. 239-240) If the two values have correlation their values will change together in a predictable way. The coefficient of non-determination suggests differenced are due to other factors indicating what proportion of the variation is due to unrelated factors. The R value assists in indicating the strength of a relationship and how well change in one variable will predict change in another variable. (p. 240) It is crucial to remember that correlation does not equate to causation and so while the researcher can use the change in one value to predict the change in the other, the researcher cannot conclude that one variable causes the change in another. The change may be coming from non-observed factors. (pp. 241-242) The Pearson correlation is used for several purposes. It helps measure the validity of an assessment by checking if scores on a new test correlate with scores on established, related tests, such as testing if a new depression assessment correlates with existing depression measures. It also assesses the reliability of an instrument by examining the consistency of test scores when administered multiple times or with different versions, ensuring that a personality test gives similar results when retaken after some time. Additionally, it is used to predict one variable based on the strength of its relationship with another, with a strong correlation allowing for more accurate predictions between the two variables, such as predicting someone's performance based on past behavior. (p. 243)

Spearman rank order correlation is a nonparametric test that measures the “…direction and strength of the association between two variables. (Frey, p. 1568) It is a commonly used test given that man correlative studies do not meet the criteria for Pearson’s R. (p. 1568) The Spearman rank order correlation measures the strength of monotonic relationships, where both variables (x and y) are arranged in ascending order and compared based on their rank differences. It determines how one data set influences another by evaluating the degree of association between the ranks of the two sets. (p. 1568) An increasingly positive monotonic relationship occurs when both x and y increase in rank together, with concurring data pairs. A negative monotonic relationship happens when x increases in rank while y decreases or remains the same, with discordant data pairs. Unlike Pearson's r, the Spearman rank-order correlation does not distinguish between linear and monotonic associations due to its use of ordinal data. This makes it a proper nonparametric alternative to Pearson’s r when assumptions of linearity cannot be met or are unknown. (p. 1568) The Spearman correlation coefficient assesses the strength of the relationship between two ordinal variables, but it does not show if the relationship is statistically significant. To determine significance, a test statistic is required to gauge the confidence that the observed correlation reflects the entire population. This is crucial for small sample sizes, where there is a higher likelihood of a nonzero correlation occurring by chance. The null hypothesis posits no relationship between the variables and if rejected, the alternative hypothesis suggests a meaningful correlation between the variables in the population. (p. 1569) The test statistic used is based on the Spearman rank-order correlation with n-2 degrees of freedom and calculates a t-value which is compared against a critical value from a t-distribution table to determine if the correlation is significantly different from zero. (p. 1569-1570) Either a one tailed or two tailed hypotheses can be examined. (Willard, p. 279) The data is ranked with each data point assigned a rank based on its position against the other data points. After the ranking, for each pair of corresponding data points, the difference is calculated between the ranks. The differences are then squared to eliminate any negative values and emphasize larger sizes. Then all the squared ranked differences are added together and utilized in the formula. A resulting positive value, up to 1, indicates a positive monotonic relationship, as one variable rises so the other rises. A negative value up to -1 indicates as one variable increases another decreases. The closer to 1 or -1 the stronger the relationship. A value of 0 indicates no monotonic relationship. (pp. 279-280, Frey, pp. 1569-1570)

Related to my hypothetical study regarding GenZ young adults who fail to launch, this instrument could be useful to analyze relationships in a correlative study examining levels of parental support and young adult success with employment, further education, or social engagements. Given I may be working with samples that would not meet parametric requirements, this test would enable me to assess if higher levels of parental support are associated with better outcomes among GenZ young adults. The test may even prove useful to measure the effectiveness of the proposed educational intervention. Parents could their perception of the intervention on a Likert scale based on their observations of follow on behavior among their children or I could have their children rank their perception of the helpfulness of the program. This data would most likely be ordinal where differences in levels of effectiveness would not be linear or uniform but may still trend in a direction. The test could be helpful in measuring the significance of perceived effectiveness among the young adults and/or their parents.

# Part 4: Independent t-Test

## Results of t-Test Procedure

**Table 1**

T-Test Statistics

|  |  |
| --- | --- |
| Measure | Value |
| Group 1 (Religious Participants) Mean | 9.1625 |
| Group 2 (Non-Religious Participants) Mean | 4.3630 |
| Degrees of Freedom | 34.0000 |
| t-Statistic | 3.8475 |
| p-Value | 0.00050021 |

**Hypotheses**

H0: No statistically significant difference exists in the mean scores of the Ethical Decision-Making Scale (EDMS) between Religious Participants and Non-Religious Participants.

Ha: A statistically significant difference exists in the mean scores of the Ethical Decision-Making Scale (EDMS) between Religious Participants and Non-Religious Participants.

A two-tailed Student's t-Test was applied to independent samples assuming equal variances to test the null hypothesis that the difference in means the Ethical Decision-Making Scale (EDMS) between Religious Participants and Non-Religious Participants were not equal to zero. The means for the groups Religious Participants and Non-Religious Participants were 9.1625 and 4.3630 respectively. With 34.0000 degrees of freedom, the t-statistic was 3.8475.

The p-value of 0.00050021 suggests a statistically significant difference between the groups' means at a 0.05 alpha level. The null hypothesis was rejected.

**Post-Hoc Procedures**

**Table 2**

T-Test Post Hoc Statistics

|  |  |
| --- | --- |
| Statistic | Value |
| Bonferroni Correction Alpha | 0.050000 |
| Cohen's d (effect size) | 1.290490 |
| Power | 0.0000 |

**Probability Density Function**

A probability distribution function shows how likely different outcomes are. The total area under the curve is 1, representing total probability. When the distribution spreads over a broader range (larger standard deviation), its peak is lower, but the total area remains the same. Interpreting the curve involves noting its peak (where values are most likely) and its spread (how values are distributed around the mean).

**Figure**

*Probability Density Function Plot*



*Note:* The overlapping region under the two PDF curves shows where the distributions share similar values. A high degree of overlap means many data points from both samples fall in the same range, whereas less overlap suggests the samples differ more in their typical values. However, the exact significance of any observed difference should still be tested statistically.

Attribution**:** Statistical procedures were conducted using [PSSR (Practical Statistics for Social Research)](https://stats.ogs.edu/), statistical analysis software developed by Joshua D. Reichard for [Omega Graduate School](https://ogs.edu/) based on the [jStat](https://jstat.github.io/) library.

## Problem and Purpose Statements

Problem Statement: Ethical decision making influences both personal and professional decision making. The influence of religious belief on ethical decision making is debated. Some argue that religious beliefs strengthen and even form the basis for moral frameworks. Others argue ethical decision making operates independently of religious beliefs. Though religion and morality are often discussed and researched, there is less empirical evidence comparing ethical decision-making between religious and non-religious individuals using standardized testing methods. This results in less understanding of religion’s influence in ethical decision making.

Purpose Statement: This study will examine whether a statistically significant difference exists in the mean scores of the Ethical Decision Making Scale (EDMS) between religious and non-religious participants. Analyzing these differences will provide insight into the potential influence of religious beliefs on ethical decision making. The results of the study will contribute to the conversation on ethics and religion with potential implications for educational and professional training programs regarding ethics.

## Research Questions

1. Is there a statistically significant difference in the mean scores of the Ethical Decision Making Scale (EDMS) between religious and non-religious individuals?

2. How does religious belief affect ethical decision making based off the measurements of the EDMS?

## Evaluation of the Output

The measurements provided by the EDMS suggest that there is a significant difference between the ethical decision making scores of religious participants compared to non-religious participants. The mean score for religious participants of 9.1625 is much higher than for non-religious participants of 4.3630. This indicates the null hypothesis should be rejected providing tacit support for the alternative hypothesis that religious beliefs do indeed have a significant influence on ethical decision making. The t-statistic of 3.8475 is greater than the critical t-value of 2.034 for a two-tailed test with 95% significance level (0.05) and 34 degrees of freedom allowing for the rejection of the null hypothesis. (“t-Test, Chi-Square, ANOVA, Regression, Correlation...”)

## P-value

The p-value is much less than 0.05 indicating the difference the measurements gathered by the EDMS are related to chance is extremely low also affirming the rejection of the null hypothesis providing tacit support for the alternative hypothesis that religious belief does significantly influence ethical decision making. (Frey, pp. 1512-1513)

#

# Part 5: Correlational: Life Satisfaction Index

**Results of Statistical Procedure**

**Table 3**

|  |  |
| --- | --- |
| Measure | Value |
| Group 1 (Hours of Weekly Religious Involvement) Mean | 1.7579 |
| Group 2 (Life Satisfaction Index (LSI)) Mean | 8.9342 |
| Pearson's Correlation Coefficient (r) | 0.3651 |
| p-value | 0.12537887 |

**Hypotheses**

H0: No statistically significant relationship exists between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI).

Ha: A statistically significant relationship exists between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI).

**Findings**

The r-value of 0.3651 indicates a medium positive effect between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI).

The probability that the relationship between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI) was not statistically significant at a 95% confidence level (p = 0.12537887). The null hypothesis could not be rejected.

**Scatterplot**



**Attribution**

Statistical procedures were conducted using [PSSR (Practical Statistics for Social Research)](https://stats.ogs.edu/), statistical analysis software developed by Joshua D. Reichard for [Omega Graduate School](https://ogs.edu/) based on the [jStat](https://jstat.github.io/) library.

## The Scatterplot

The line is trending upward showing a potential positive trend, but the data points are scattered without tight grouping which could suggest that other factors are influencing life satisfaction beyond religious involvement. (Frey, pp. 1183-1184, 1466-1470)

## Problem and Purpose Statement

Problem Statement: The relationship between religious involvement and overall life satisfaction has been debated. Some studies suggest greater religious involvement may contribute to higher life satisfaction levels while others indicate no clear association. There is limited empirical evidence concerning the relationship between the number of hours spent in weekly religious activity and life satisfaction. This results in incomplete understanding of how well religious involvement impacts life satisfaction with implications for religious communities and their faith practitioners.

Purpose Statement: This study will examine the relationship between the hours of weekly religious involvement and life satisfaction measured by the Life Satisfaction Index (LSI). By determining if a statistically significant relationship exists the study will contribute to understanding how religious involvement influences individual well-being. These insights can not only aid faith based programs seeking to improve life satisfaction but also potentially guide future research.

## Research Questions:

1. Is there a statistically significant relationship between the hours of religious involvement and life satisfaction as measured utilizing the Life Satisfaction Index (LSI)?

2. To what extent do patterns of regular religious involvement predict life satisfaction scores among individuals?

## Test results:

As the results reveal, the r-value of 0.3651 “indicates a medium positive effect.” The score falls between 0.3 and 0.5, which suggests a moderate correlation. (Willard, pp. 236-240) This would suggest a moderate positive trend that could indicate that as individuals spend more time in religious involvement, their satisfaction with life also grew. However, as stated above, the scatterplot’s upward slope is not dramatic and the data points are scattered suggesting that the relationship is not very strong and could possibly be influenced by other factors that are impacting both of the variables.

## p-value:

The p-value of 0.124 is somewhat greater than the confidence threshold of 0.05 indicating the correlation between the two variables is not statistically significant, that is it cannot be ruled out that the results are produced by chance generated by the sample. (Frey, pp. 1512-1513) Therefore as the PSSR states, “The null hypothesis could not be rejected.”

# Part Six: Mann Whitney U

**Results of Mann-Whitney U Procedure**

**Table 4**

Mann Whitney U Statistics

|  |  |
| --- | --- |
| Measure | Value |
| Group 1 (Religious Participants) Mean | 9.1625 |
| Group 2 (Non-Religious Participants) Mean | 4.3630 |
| Mann Whitney U | 52.0000 |
| Z-Score | -3.4383 |
| p-value | 0.00058547 |

**Hypotheses**

H0: No statistically significant difference exists in the distributions of the Ethical Decision-Making Scale (EDMS) between Religious Participants and Non-Religious Participants.

Ha: A statistically significant difference exists in the distributions of the Ethical Decision-Making Scale (EDMS) between Religious Participants and Non-Religious Participants.

A two-tailed Mann-Whitney U procedure was applied to the samples assuming equal variances to test the null hypothesis that the difference in means the Ethical Decision-Making Scale (EDMS) between Religious Participants and Non-Religious Participants were not equal to zero. The means for the groups Religious Participants and Non-Religious Participants were 9.1625 and 4.3630 respectively.

The p-value of 0.00058547 suggests a statistically significant difference between the groups' distributions at a 0.05 alpha level. The null hypothesis was rejected.

**Post-Hoc Procedures**

**Table 5**

Mann-Whitney U Post Hoc Statistics

|  |  |
| --- | --- |
| Statistic | Value |
| Bonferroni Correction Alpha | 0.050000 |
| Cohen's d (effect size) | 1.290490 |
| Power | 0.0000 |

**Probability Density Function**

A probability distribution function shows how likely different outcomes are. The total area under the curve is 1, representing total probability. When the distribution spreads over a broader range (larger standard deviation), its peak is lower, but the total area remains the same. Interpreting the curve involves noting its peak (where values are most likely) and its spread (how values are distributed around the mean).

**Figure**



*Note:* The overlapping region under the two PDF curves shows where the distributions share similar values. A high degree of overlap means many data points from both samples fall in the same range, whereas less overlap suggests the samples differ more in their typical values. However, the exact significance of any observed difference should still be tested statistically.

**Attribution**

Statistical procedures were conducted using [PSSR (Practical Statistics for Social Research)](https://stats.ogs.edu/), statistical analysis software developed by Joshua D. Reichard for [Omega Graduate School](https://ogs.edu/) based on the [jStat](https://jstat.github.io/) library.

## Test Results and p-value

As in the t-test, the Mann Whitney U supported rejection of the null hypothesis. The religious participants’ mean of 9.1625 and the non-religious participants’ mean of 4.3630 generated a Mann Whitney U score of 52.000 and a Z-score of -3.4383. Given the p-value calculated by the PSSR is 0.00058547, which is far less than the 0.05 confidence threshold, this suggests the U value of 52.0000 is small enough to indicate a significant difference between the EDMS scores of religious and non-religious participants. The z-score being far from zero also suggests the difference is not likely due to chance. (Frey, pp. 1841-1842) Since the observed differences are less likely to be due to chance, we can reject the null hypothesis.

# Part Seven: Spearman Rank Procedure

**Results of Spearman's Rank Procedure**

**Table 6**

|  |  |
| --- | --- |
| Measure | Value |
| Group 1 (Hours of Weekly Religious Involvement) Mean | 1.7579 |
| Group 2 (Life Satisfaction Index (LSI)) Mean | 8.9342 |
| Spearman's Rank (ρ) | 0.3996 |
| p-Value | 0.0912 |

**Hypotheses**

H0: No statistically significant relationship exists between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI).

Ha: A statistically significant relationship exists between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI).

**Findings**

**Table 7**

|  |  |
| --- | --- |
| Measure | Value |
| Group 1 (Hours of Weekly Religious Involvement) Mean | 1.7579 |
| Group 2 (Life Satisfaction Index (LSI)) Mean | 8.9342 |
| Spearman's Rank (ρ) | 0.3996 |
| p-Value | 0.0912 |

The probability that the relationship between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI) was not statistically significant at a 95% confidence level (p = 0.09121576). The null hypothesis could not be rejected.

**Effect Size**

The r² statistic of 0.1596 indicates a small positive effect between Hours of Weekly Religious Involvement and Life Satisfaction Index (LSI).

**Scatterplot**

****

**Attribution:** Statistical procedures were conducted using [PSSR (Practical Statistics for Social Research)](https://stats.ogs.edu/), statistical analysis software developed by Joshua D. Reichard for [Omega Graduate School](https://ogs.edu/) based on the [jStat](https://jstat.github.io/) library.

## Test results and p-value

The Pearson’s Correlation Coefficient (r) for this data was 0.3651 compared to the Spearman’s rank Correlation of 0.3996. The stronger Spearman’s score would indicate a stronger positive relationship than the Pearson’s score. This could suggest a linear relationship between the variables is weak, that is as one rises the other might be expected to rise proportionally. The stronger Spearman’s rank Correlation could suggest that one could more confidently predict the more a subject participated in religion the higher the perceived life satisfaction but that the strength of the relationship is not necessarily proportional or changing at a constant rate. (Frey, 1568; Willard p. 236-240) However, as before the p-value of 0.0912 is greater than the 0.05 confidence threshold and does not allow for the rejection of the null hypothesis. It is possible the relationship suggested by both Pearson and Spearman’s could be due to chance. However, if I were the researcher, I would suggest that even though the p-value is not high enough to reject the null hypothesis, this does not warrant accepting the null-hypothesis and the slight and moderate relationships suggested by the analysis of the data suggests further research is needed to arrive at results that could be more conclusive as to whether the amount of religious involvement positively correlates with life satisfaction.

# Part 8: Assumptions

## The 1st Study

The first study compares the ethical decision making of individuals comparing religious and non-religious subjects utilizing the EDMS. Based on assumption testing performed by the PSSR the data collected from the religious participants appears close to normal. The skewness was 0.1708 which is slightly to the left but still close to zero. The excess kurtosis score of -3.9003 indicates “few extreme values”. The Shapiro-Wilk statistic of 0.9842 is close to 1 suggesting a normal distribution according to the PSSR. The Kolmogorov-Smirnov statistic of 0.1974 is close to 0 suggesting a normal distribution according to the PSSR. However, the PSSR also observed a larger right tale with the bulk of the data concentrated left below the mean and a “…few large values pulling the mean to the right of the median.” This could support that the distribution is not normal.

Examining the data for the non-religious group the PSSR reported skewness of 1.5084 indicating a strong right skew suggesting the data was not distributed normally. The Shapiro-Wilk statistic of 0.9087 suggested some deviation from normality. Likewise, the Kolmogorov-Smirnov statistic of 0.2148 is higher than the religious group suggesting less normality in this data set and some distance from zero.

So overall, the religious participants are close to normal but not perfectly normal and the PSSR indicates that the non-religious participant’s data is not normally distributed. If I were examining the groups independently I could use parametric tests for the religious participants and non-parametric for the non-religious, but given the design study is focused on comparing the two to see if religious people are more ethical in their behavior, I would utilize non-parametric testing for both data sets to make the research design more reliable.

## The 2nd Study

The second study is examining the strength of the relationship between hours of weekly religious involvement and life satisfaction as measured using the LSI. For the data set regarding weekly religious involvement, the PSSR reported a light right skew of 0.1660 and excess kurtosis of -4.1681 which the PSSR evaluated as not supporting normal distribution with “Most observations…below the mean, with a few large values pulling the mean to the right of the median.” The Shapiro-Wilk statistic was 0.9709 which is closer to 1 which the PSSR suggested “may follow a normal distribution.” The Kolmogorov-Smirnov statistic of 0.1827 is not very close to 0 which suggests the data may deviate from normality. The dataset for the LSI had a skewness score of 0.0447 which slows a slight skew to the left toward smaller values. The PSSR evaluates this as not visually appearing normal. However, the actual skewness value is slight. The Shapiro-Wilks Statistic is 0.9852 which is very close to one which the PSSR suggests may indicate normality. The Kolmogorov-Smirnov statistic is 0.1727 which is not extremely close to zero suggesting some deviation from normality.

Given the overall picture, I would utilize nonparametric procedures to evaluate the strength of the relationship, and more specifically Spearman’s Rank Correlation. Even though both data sets have Shapiro-Wilk values close to 1, both are noted by the PSSR to exhibit negative excess kurtosis which suggests the data does not have the spread typical of normal distribution. The PSSR assesses that neither data set visually appears to be normal. Additionally, the Kolmogorov-Smirnov statistic for both data sets indicated some deviation from normality.

# Part 9: Perfect Correlation

The scatterplot provided by the PSSR indicates a perfect linear correlation. As the values on the x axis increase, there is a corresponding linear and proportional increase in the values on the y axis. This will indicate that as the strength of one variable increases one can predict that the other variable will also increase in a consistent pattern with intervals of proportional value between the observations. The visual inspection of the graph indicates a strong positive correlation. (Frey, p. 1470, pp. 1183-1184)

# Part 10: Key Takeaways

Performing these exercises has reinforced several key aspects to remember when considering the application of statistical methods to a particular research design. First there is the reminder that the research problem, purpose, and questions will drive what type of data is collected. These exercises illustrated the vital importance of assumption testing of the collected data as the various statistics calculated by possible procedures only have predictive power if the data matches the assumptions of the test, especially the test for normality. It is noted that the different assumption tests of the various data sets did not always align perfectly in their suggestions regarding normality. This leads to another key learning: the importance of interpretation. Working with data statistically even though it is utilizing math, still has a degree of artistic interpretation which one could suspect is refined with experience. Here one could suspect the literature review and considering how other researchers approached similar data can be helpful. It was also noteworthy that the importance of interpretation not only applied to evaluating the assumption testing, but also to the results. The results of the study regarding religious involvement and life satisfaction could not support rejecting the null hypothesis. However, there was still some indication that some possible relationship might exist between the two variables or that some other external variable(s) may be related suggesting further research is necessary. While it cannot be proven with the data collected that religious activity was related to life satisfaction, the conclusion that religious activity has no influence over life satisfaction is not supported and there is a suggestion in the data that if more data could be collected resulting in conclusions falling into the 95% confidence interval, that such a correlation might be established. A final key take away, is the importance of making careful decisions regarding each step of the process and being able to articulate in one’s research writing the reasoning behind these decisions to lead to a more robust, reliable study with believable conclusions.

WORKS CITED

Frey, B. B. (2022). *The SAGE encyclopedia of research design* (2nd ed.). SAGE Publications, Inc.

Hahs-Vaughn, D. L. (2020). *Statistical concepts—A first course* (1st ed.). Routledge.

Reichard, J. (2025). *Practical statistics for social research (PSSR)*. Omega Graduate School. <https://stats.ogs.edu/>

*T-test, Chi-square, ANOVA, regression, correlation...* (n.d.). Retrieved March 8, 2025, from <https://datatab.net/tutorial/t-distribution>

Willard, C. A. (2020). *Statistical methods* (2nd ed.). Routledge.