Statistics for Social Research II

Derrick Snow

Omega Graduate School

Sept 8, 2024

Professor

Dr. Taladay

Assignment

**Assignment #2 – Developmental Readings**

1. Create Developmental Readings from seminal sources and scholarly peer-reviewed

journal articles. Review instructions for Assignment #3, the course essential elements, and course readings to identify selections of books and journals to create entries.

1. Refer to the "Student Guide to Developmental Readings" in the General Helps folder for updated information on sample comments, the grading rubric, and key definitions related to developmental readings.

**Source One:** Shepherd-Banigan, M., Ulrich, A., & Thompson, B. (2014). Macro-Activity Patterns of Farmworker and Non-Farmworker Children Living in an Agricultural Community. *Environmental Research*, *132*, 176–181. <https://doi.org/10.1016/j.envres.2014.04.001>

**Comment 1:**

**Quote/Paraphrase:** “The paired t-test is used for matched data and normality assumptions are considered by examining the distribution of the difference within children by season. For the most part, the distributions of these differences were approximately normal and met the distributional assumptions of the paired t-test. The distribution of the difference in the number of days that a child accompanied their parent to work in the fields was the only continuous difference that was not normally distributed. Therefore, the mean of this difference is presented and statistical significance is assessed using the Signed Rank Test for paired data.”

(Shepherd-Banigan et al., 2014, p. 178)

**Essential Element:** Assumptions testing, Parametric Statistics, Nonparametric Statistics

**Additive/Variant Analysis:** The use of normality testing is additive to my understanding of determining statistics that could be used given the shape of the distribution.

**Contextualization:**  In this quote the researcher examined the normality of two different variables and identified what statistics could be used in examining whether or not the distribution of the data is approximately normal to where it would fit the normalized sampling distribution. If so, then parametric statistics could be performed, and if not, then non-parametric statistics would be used. In this research, the researcher used a t-test for the approximately normal data and a Signed Rank Test for the data that did not fit the normal distribution. It is important for the researcher to dive into exploring the data they have collected, not only to understand the data, but also to test the assumptions of the data. There are several assumption tests that should be completed, a major one being normality testing.

**Source Two:** Min, R. N. J., 2022 Heat, Health and Central Valley Agricultural Populations: Examining Farmworker Vulnerability to Emerging Health Threats of Global Warming. *Vulnerability to Emerging Health Threats of Global Warming*

**Comment 2:**

**Quote/Paraphrase:** “To ensure the data is well suited for ANOVA testing, I also conducted a normality assumption test using the residuals and a probability plot. I used the probability plot (Chambers et al., 1983) as a graphical technique for assessing whether or not my data set follows a normal distribution. The data are plotted against a theoretical distribution in such a way that the points should form approximately a straight line. Departures from this straight line indicate departures from the specified distribution.” (Min, 2022, p. 15)

**Essential Element:** Assumptions testing

**Additive/Variant Analysis:** The concept of using assumption testing for examining the assumptions about the data and, in this case, the normality of the data is additive to my understanding of assumption testing.

**Contextualization:** The researcher intended to perform an ANOVA (Analysis of Variance) to test the hypothesis using residuals and a probability plot to test the normality assumption of the data. A residual probability plot, also known as a normal probability plot of residuals, is a visual display of error terms (residuals) from the predicted equation line. If the residual plots fall approximately linear to the equation line, then it supports that the error terms are normally distributed. The ANOVA is used to examine the variation between groups. Thus, using residual plots is a good way to test the normality of the variation.

**Source Three:** López-Gálvez, N., Wagoner, R., Canales, R. A., Ernst, K., Burgess, J. L., De Zapien, J., Rosales, C., & Beamer, P. (2021). Longitudinal Assessment of Kidney Function in Migrant Farm Workers. *Environmental Research*, *202*, 111686. <https://doi.org/10.1016/j.envres.2021.111686>

**Comment 3:**

**Quote/Paraphrase:** “Variable distribution was examined using histogram and summary statistics. Variables with a lognormal distribution were log-transformed prior to analyses.” ( López-Gálvez et al., 2021, p.4)

**Essential Element:** Assumptions testing

**Additive/Variant Analysis:** The use of histograms in determining the normal distribution assumptions is additive to my understanding of examining the normal distribution of data.

**Contextualization:** Using histograms to examine the distributions of the data is a common way researchers identify if the data follows an approximately normal distribution. Histograms are a visual tool that data within bins along the x-axis and the quantity within the bins on the y-axis, as well as a curve, maybe overlayed to display a bell curve. Many statistics are based on the normal distribution. Thus, it is important to see if the data meets the normal distribution assumption so one can know if they can perform parametric statistical tests and operations. Histograms are a standard graphical tool to see if the data is normally distributed.

**Comment 4:**

**Quote/Paraphrase:** “Differences between MSFWs and office workers at pre-harvest were assessed using a t-test for continuous variables and a chi-square test for categorical variables.” (López-Gálvez et al., 2021, p.4)

**Essential Element:** T-test and Pearson’s r

**Additive/Variant Analysis:** This is additive to my understanding of the use of t-test to examine if there is a statistically significant difference in the means of two data sets.

**Contextualization:**  The t-tests are a parametric statistic used for continuous variables when the data is normally distributed to examine if there is a statistically significant difference between two the means of two groups. The researchers used to test the diffidence between migrant/seasonal farm workers and office staff workers. For a researcher to examine the differences between two groups within the same sample, the t-test is commonly used to test the hypothesis of differences. A t-test could be used to test if there is a difference between the seasonal and migrant farmworkers' rate of heat-related illnesses and the rate of heat-related illnesses of non-seasonal and migrant farmworkers.

**Source Four:** Curl, C. L., Meierotto, L., Castellano, R. L. S., Spivak, M. R., & Kannan, K. (2021). Measurement of Urinary Pesticide Biomarkers Among Latina Farmworkers in Southwestern Idaho. *Journal of Exposure Science & Environmental Epidemiology*, *31*(3), 538–548. <https://doi.org/10.1038/s41370-020-00285-2>

**Comment 5:**

**Quote/Paraphrase:** “The data were not normally distributed, and we therefore employed the nonparametric Mann–Whitney U test to compare pesticide exposure between samples collected during the nonspray season and the spray season. (Curl et al., 2021, p. 541)

**Essential Element:** Nonparametric Statistics

**Additive/Variant Analysis:** The use of Mann-Whitney U test for nonparametric data in order to see if there is a significant difference between means of groups is additive to my understanding.

**Contextualization:** The above quote identifies from normality tests that the data was not normally distributed. Thus, the researchers used the nonparametric alternative test to the independent sample t-test to test the mean difference between two groups, the non-spray season and the spray season; they used the Mann-Whitney U test which is similar to the t-test of parametric data, it measures the difference between the means of two groups within the same sample.

**Source Five:** Mizelle, E., Larson, K. L., Bolin, L. P., & Kearney, G. D. (2022). Fluid Intake and Hydration Status Among North Carolina Farmworkers: A Mixed Methods Study. *Workplace Health & Safety*, *70*(12), 532–541. <https://doi.org/10.1177/21650799221117273>

**Comment 6:**

**Quote/Paraphrase:** “Parametric analyses were done using SPSS v.27 (IBM Corp., 2020). Descriptive statistics were calculated. Pearson’s product correlation coefficient was used, using Cohen’s guidelines, to examine the associations between cultural and place factors, beverage intake, WBGT with the dependent variable of USG.” (Mizelle et al., 2022, p. 535)

**Essential Element:** T-test and Pearson’s r

**Additive/Variant Analysis:** The use of the parametric correlation statistic of Pearson’s r to examine the strength of an association between two variables is additive to my understanding.

**Contextualization:** In doing correlational studies on parametric data, one uses Pearson’s r, which measures the strength of association between two variables. In this research they wanted to find if there was an association between Wet bulb globe temperature (WBGT) and urine specific gravity (USG) and what was the strength of the association between this measure intake and hydration. Correlation studies are informative in social science and healthcare research where the researchers could find associations between different variables, this is not causation research but nonetheless is important to find out what factors are associated with given outcomes.

**Source Six:** Lee, S., McLaughlin, R., Harnly, M., Gunier, R., & Kreutzer, R. (2002). Community Exposures to Airborne Agricultural Pesticides in California: Ranking of Inhalation Risks. *Environmental Health Perspectives*, *110*(12), 1175–1184. <https://doi.org/10.1289/ehp.021101175>

**Comment 7:**

**Quote/Paraphrase:** “We evaluated several predictors of the chronic inhalation risks estimated in this report, using Spearman rank correlation coefficients. The California ranking for potential pesticide toxic air contaminants (Table 1) was not significantly correlated with the child chronic risk ranking (r = 0.22, p = 0.43)..” (Lee et al., 2002, p. 1180)

**Essential Element:** Nonparametric Statistics

**Additive/Variant Analysis:** The concept of utilizing Spearman Rank for measuring the strength of association for nonparametric data is additive to my understanding.

**Contextualization:**  In this research, the researchers use the Spearman Rank for the nonparametric data to examine the strength of the association between chronic inhalation risk and potential pesticide toxic air contaminants. Spearman Rank is similar to Pearson’s r where it measures the direction and strength of an association between two variables but is normally used for ordinal variables. As like the Pearson’s r then Spearman Rank with a close to a -1 is a strong negative correlation, close to a +1 is a strong positive association.

**Source Seven:** Terrell, S. R. (2021). *Statistics Translated: A Step-by-Step Guide to Analyzing and Interpreting Data*. Guilford Publications.

**Comment 8:**

**Quote/Paraphrase:** “The first thing the empirical rule tells us is, if the distribution is approximately normally distributed, at least 68% of the values in the dataset fall within ±1 standard deviation from the mean.” (Terrell, 2021, p. 107)

**Essential Element:** Assumptions testing, Parametric Statistics

**Additive/Variant Analysis:** These facts about the normal distribution are additive to my understanding.

**Contextualization:** The normality of data is very powerful because of the shape and probability function of the normal curve. The empirical rule of the normal curve informs us that 68% of the data will fall within one standard deviation from the mean, 95% within two standard deviations, and 99.7% within 3 standard deviations. Given this probability function, we could infer much about the population from our sample given our sample distribution. We could utilize z scores to standardize our information. The normality of data gives us the opportunity to perform many statistical operations and allows us to infer the larger population through our sample given the sampling distribution.

**Comment 9:**

**Quote/Paraphrase:** “Setting up, running, and interpreting the output of the Wilcoxon test is very similar to that of the dependent-sample t-test. Again, this is not something that happens often, and like the Mann-Whitney U test, this test is something you need to keep in the back of your mind for those rare instances.” (Terrell, 2021, p. 237)

**Essential Element:** Nonparametric Statistics

**Additive/Variant Analysis:** This information about the nonparametric tests of Wilcoxon and Mann-Whitney U test is additive to my understanding of nonparametric means testing.

**Contextualization:** From this quote, it seems like the author assumes that much of the research data will follow an approximate normal distribution and would likely have a sufficient sample size thus the use for a t-test. But in some cases the data would not meet the assumptions of normality or would not have a large enough sample to move the distribution to an approximant normality, thus the need for nonparametric statistics such as the Wilcoxon and Mann-Whitney U tests. These nonparametric statistics are good to know what to use when the data does not meet the assumption of normality.

**Source Eight:** Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman and Hall/CRC.

**Comment 10:**

**Quote/Paraphrase:** “The discussion of correlational research in this section will be limited to the simplest type of correlation – specifically, bivariate correlation, which is the use of correlation to measure the degree of association between two variables. It should be noted that correlational procedures can also be employed with more than two variables. The latter type of correlation is commonly discussed within the framework of multivariate analysis (e.g., multiple regression.” (Sheskin,2011, p. 81)

**Essential Element:** T-test and Pearson’s r

**Additive/Variant Analysis:** This quote is about correlational research is additive to my understanding of research design.

**Contextualization:** Bivariate correlation is when researchers examine the relationship between two variables. Multivariate correlation is where the researchers examine the relationship between multiple variables and in the case of multiple regression, where there are multiple independent variables that may have an association with one dependent variable. Regression analysis is a powerful tool in economics and social sciences. However, there are assumptions that must be met in order to utilize regression analysis where there is much statistical logic built around the normal distribution.

**Comment 11:**

**Quote/Paraphrase:** “The inferential statistical procedures discussed in this book have been categorized as being parametric versus nonparametric tests. Some sources distinguish between parametric and nonparametric tests on the basis that parametric tests make specific assumptions with regard to one or more of the population parameters that characterize the underlying distribution(s) for which the test is employed. These same sources describe nonparametric tests (also referred to in some sources as distribution-free or assumption-free tests) as making no such assumptions about population parameters. In truth, nonparametric tests are really not assumption free, and, in view of this, Marascuilo and McSweeney (1977) suggest that it might be more appropriate to employ the term “assumption freeer” rather than nonparametric in relation to no such tests.” (Sheskin,2011, p. 109)

**Essential Element:** Parametric Statistics, Nonparametric Statistics

**Additive/Variant Analysis:** This quote addresses the different assumptions between parametric and nonparametric tests, which is additive to my understanding of parametric and nonparametric statistics.

**Contextualization:** This is an informative section within the text, where the author discusses that nonparametric test statistics are not totally free from assumptions about the data but are rather “assumption free” than parametric statistics. This is important to keep in mind to ensure that the appropriate statistics are being used for the appropriate circumstances. We as researchers need to ensure we understand the statistics we intend to use and to know about any short comings surrounding the statistics we use.

**Comment 12:**

**Quote/Paraphrase:** “Quasi-experimental designs Since they do not rule out the possible influence of all extraneous variables, quasi-experimental designs are also subject to confounding, and thus may lack internal validity. Quasi-experimental designs, however, rule out more extraneous variables than pre-experimental designs. Because of the later, quasi-experimental designs are preferable to employ when practical and/or ethical considerations do not permit a researcher to evaluate a hypothesis through use of a true experimental design. In most instances, lack of random assignment of subjects is responsible for compromising the internal validity of a quasi-experimental designs.” (Sheskin,2011, p. 89)

**Essential Element:** Parametric Statistics, Nonparametric Statistics

**Additive/Variant Analysis:** This quote informs us of the distinctions between true experimental design and quasi-experimental design namely random assignment of participants to different groups and the use of control groups. This is additive to my understanding of true experimental design and quasi-experimental design.

**Contextualization:** I know it is a gold standard in research to have true experimental design, however in social science research, I do not find that we should want a true experimental design, because what it entails is that we manipulate the social setting to where it no longer represents social reality in order to have ‘valid’ results that are not valid because we are no longer measure the social phenomena, but a synthetic construction of a real social phenomena. As social science researchers, we should be seeking natural experiments that are taking place in the real world. Admittedly, these natural experiments are challenging to find because we don’t typically think in such ways.

**Works Cited**

Curl, C. L., Meierotto, L., Castellano, R. L. S., Spivak, M. R., & Kannan, K. (2021). Measurement of Urinary Pesticide Biomarkers Among Latina Farmworkers in Southwestern Idaho. *Journal of Exposure Science & Environmental Epidemiology*, *31*(3), 538–548. <https://doi.org/10.1038/s41370-020-00285-2>

Lee, S., McLaughlin, R., Harnly, M., Gunier, R., & Kreutzer, R. (2002). Community Exposures to Airborne Agricultural Pesticides in California: Ranking of Inhalation Risks. *Environmental Health Perspectives*, *110*(12), 1175–1184. <https://doi.org/10.1289/ehp.021101175>

López-Gálvez, N., Wagoner, R., Canales, R. A., Ernst, K., Burgess, J. L., De Zapien, J., Rosales, C., & Beamer, P. (2021). Longitudinal Assessment of Kidney Function in Migrant Farm Workers. *Environmental Research*, *202*, 111686. <https://doi.org/10.1016/j.envres.2021.111686>

Min, R. N. J., 2022 Heat, Health and Central Valley Agricultural Populations: Examining Farmworker Vulnerability to Emerging Health Threats of Global Warming. *Vulnerability to Emerging Health Threats of Global Warming*

Mizelle, E., Larson, K. L., Bolin, L. P., & Kearney, G. D. (2022). Fluid Intake and Hydration Status Among North Carolina Farmworkers: A Mixed Methods Study. *Workplace Health & Safety*, *70*(12), 532–541. <https://doi.org/10.1177/21650799221117273>

Shepherd-Banigan, M., Ulrich, A., & Thompson, B. (2014). Macro-Activity Patterns of Farmworker and Non-Farmworker Children Living in an Agricultural Community. *Environmental Research*, *132*, 176–181. <https://doi.org/10.1016/j.envres.2014.04.001>

Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman and Hall/CRC.

Terrell, S. R. (2021). *Statistics Translated: A Step-by-Step Guide to Analyzing and Interpreting Data*. Guilford Publications