Statistics 101: Manual MST Primer

Gregory S. Rash, EdD

Research data can be obtained in a variety of formats. Each of these formats has unique characteristics that impact on the types of statistical techniques that can be appropriately applied. This tutorial is an overview of common statistics used for each of these types of data. It is not intended to give everyone all the tools necessary to handle all of their statistical needs, but to give them an overview of statistical options when dealing with the different types of data produced in a number of research environments. By having a better understanding of the statistical options available in the planning stages, one is able to design a better study and have data that can be analyzed. This brief tutorial is part of a larger tutorial that can be found on line that describes several non-parametric and parametric statistical techniques, as well as identifies common indications for their applications.

The **KEY** to a good project is to work with someone that understands how to correctly collect & analyze the data from a statistical standpoint from the beginning stages of your work. This can save hours of needless data collection only to find that the data CANNOT be analyzed (to few subjects in a group, not valid or reliable testing instruments, etc.), or that you must collapse numerous categories to allow for any statistical analysis, but now you've lost any meaningful practical separation of categories/groups. There are ways to determine the number of subjects needed in a study (to advanced for this brief overview), but as a basic rule of thumb, when using inferential statistics you need a minimum of 10 in any group for any chance of showing that differences exist unless the differences are quite large.

Your goal is to try to design your study to remove or control all influence of any variable that is not your independent variable to maximize the effect on your dependent variable (you want your independent variable to be the only thing that causes the change in your dependent variable). This is done by controlling or eliminating the variability caused by other variables. That being said, there are NO perfect studies & it is virtually impossible to eliminate or control all the variables that influence your dependent variable. Additionally, if your research is not basic research (meaning that it is applied research) then you want to see what happens in the real world when very little is controlled. However, now it becomes very difficult to know what really caused your change.

Control groups. This is why applied studies employ the use of control groups. That is a group that is identical to the group where you manipulate a variable. Then you can see if your manipulated variable caused the difference, or did the change occur due to the act of just running around on this big rock we call earth.

I. Types of data

Nominal - No arithmetic relationship or order between different classifications. **Examples**: Occupation (Clerk, Police Officer, Teacher, …); Gender (Male, Female)

Ordinal - Data can be ordered into discrete categories, but categories have no arithmetic relationship. **Examples**: Survey Data (5 - Strongly Agree, 4 - Somewhat Agree, 3 - Neither Agree Nor Disagree, 2 - Somewhat Disagree, 1 - Strongly Disagree); Manual Muscle Test (5 - Full Range of Motion with Maximal Resistance, 4 - Full Range of Motion with Resistance, 3 - Full Range of Motion Against Gravity, 2 - Full Range of Motion without Gravity, 1 - Partial Range of Motion without Gravity, 0 - None or Trace Movement)

Interval - Data on a measurement scale with an arbitrary zero point in which numerically equal intervals at different locations on the scale reflect the same quantitative difference.
Examples: Temperature (Fahrenheit or Celsius)

Ratio - Data on a measurement scale with an absolute zero point in which numerically equal intervals at different locations on the scale reflect the same quantitative difference.
Examples: Height, Weight, Pressure, Temperature (Kelvins)

II. **Types of Variables**

Independent – A variable that is manipulated (the treatment variable, the cause).
Dependent – A variable that is measured (the outcome, the effect).
Categorical – A classification variable that is analyzed (e.g. gender, race).
Control – A characteristic that is restricted in the study, but not compared (e.g. only stroke pts).
Extraneous – A variable that affects the dependent variable, but is not part of the design, is not controlled. (e.g. Amount of sleep).
Confounding – When an extraneous variable is systematically related to the independent variable. (e.g. Vertical GRF & ankle compression force).
Predictor – Another name for the independent variable in regression.
Response - Another name for the dependent variable in regression. Sometimes called the Criterion.
Dummy – Variables constructed to allow analysis within a specific models framework.
Endogenous – Variables not affected by other variables in the study.
Exogenous - Variables that are affected by other variables in the study.

REVIEW OF AVAILABLE STATISTICAL TESTS

The tutorial found on Dr. Rash's website discusses many different statistical tests. To select the right test, ask yourself two questions: What type of data do you have? What is your goal? Then refer to the table that follows. Most of the tests described in the tutorial & the table can be performed by most advanced statistical packages.

Review of Nonparametric Tests

Choosing the right test to compare measurements is tricky, as you must choose between two families of tests (parametric and nonparametric). Many statistical tests are based upon the assumption that the data are sampled from a Normal distribution. These tests are referred to as parametric tests. Commonly used parametric tests are listed in the 2nd column of the table (e.g. t test & ANOVA).

Tests that do not make assumptions about the population distribution are referred to as nonparametric tests. Some of these tests are covered in the tutorial found online as well. All commonly used nonparametric tests rank the outcome variable from low to high and then analyze the ranks. These tests are listed in the 3rd column of the table (e.g. Wilcoxon, Mann-Whitney test, and Kruskal-Wallis tests). These tests are also called distribution-free tests.

Choosing Between Parametric And Nonparametric Tests: The Easy Ones

Choosing between the two types of tests is sometimes easy. Definitely choose a parametric test if you are sure that your data were sampled from a population that follows a Normal distribution (at least approximately). Definitely select a nonparametric test in three situations:

- The outcome is a rank or a score and the population is clearly not Normal. Examples include class ranking of students, the visual analogue score for pain (measured on a continuous scale where 0 is no pain and 10 is unbearable pain), or a manual muscle test (measured on a continuous scale where 0 is no movement and 5 is basically normal).

- Some values are "off the scale," that is, too high or too low to measure. Even if the population is Normal, it is impossible to analyze such data with a parametric test since you don't know all of the values. Using a nonparametric test with these data is simple. Assign values too low to measure an arbitrary very low value and assign values too high to measure an arbitrary very high value. Then perform a nonparametric test. Since the nonparametric test only knows about the relative ranks of the values, it won't matter that you didn't know all the values exactly.

- The data are measurements, and you are sure that the population is not distributed in a Normal manner. If the data are not sampled from a Normal distribution, consider whether you can transform the values to make the distribution become Normal (e.g. take the logarithm or reciprocal of all values). There are often biological or chemical reasons (as well as statistical ones) for performing a particular transform.

Choosing Between Parametric And Nonparametric Tests: The Hard Ones

It is not always easy to decide whether a sample comes from a Normal population. Consider these points:

- If you collect many data points (over a hundred or so), you can look at the distribution of data and it will be fairly obvious whether the distribution is approximately bell shaped. A formal statistical test (Kolmogorov-Smirnoff test, not explained in this tutorial) can be used to test whether the distribution of the data differs significantly from a Normal distribution. With few data points, it is difficult to tell whether the data are Normal by inspection, and the formal test has little power to discriminate between Normal and non-Normal distributions.
- You should look at previous data as well. Remember, what matters is the distribution of the overall population, not the distribution of your sample. In deciding whether a population is Normal, look at all available data, not just data in the current experiment.
- Consider the source of scatter. When the scatter comes from the sum of numerous sources (with no one source contributing most of the scatter), you expect to find a roughly Normal distribution.

When in doubt, some people choose a parametric test (because they aren't sure the Normal assumption is violated), and others choose a nonparametric test (because they aren't sure the Normal assumption is met).

Choosing Between Parametric And Nonparametric Tests: Does It Matter?

Does it matter whether you choose a parametric or nonparametric test? The answer depends on sample size. Here are four situations to give some insight:

- Large sample. What happens when you use a parametric test with data from a non-Normal population? The central limit theorem ensures that parametric tests work well with large samples even if the population is non-Normal. In other words, parametric tests are robust to deviations from Normal distributions, so long as the samples are large. The snag is that it is impossible to say how large is large enough, as it depends on the nature of the particular non-Normal distribution. Unless the population distribution is really weird, you are probably safe choosing a parametric test when there are at least two-dozen data points in each group.
- Large sample. What happens when you use a nonparametric test with data from a Normal population? Nonparametric tests work well with large samples from Normal populations. The P values tend to be a bit too large, but the discrepancy is small. In other words, nonparametric tests are only slightly less powerful than parametric tests with large samples.
- Small samples. What happens when you use a parametric test with data from non-Normal populations? You can't rely on the central limit theorem, so the P value may be inaccurate.
- Small samples. When you use a nonparametric test with data from a Normal population, the P values tend to be too high. The nonparametric tests lack statistical power with small samples.

Thus, large data sets present no problems. It is usually easy to tell if the data come from a Normal population, but it doesn't really matter because the nonparametric tests are so powerful and the parametric tests are so robust. It is the small data sets that present a dilemma. It is difficult to tell if the data come from a Normal population, but it matters a lot. The nonparametric tests are not powerful and the parametric tests are not robust.

One- Or Two-Sided P Value?

With many tests, you must choose whether you wish to calculate a one- or two-sided P value (one- or two-tailed). Let's review the difference in the context of a t test. The P value is calculated for the null hypothesis that the two population means are equal, and any discrepancy between the two sample means is due to chance. If this null hypothesis is true, the one-sided P value is the probability that two sample means would differ as much as was observed (or further) in the direction specified by the hypothesis just by chance, even though the means of the overall populations are actually equal. The two-sided P value also includes the probability that the sample means would differ that much in the opposite direction (i.e., the other group has the larger mean). The two-sided P value is twice the one-sided P value.

A one-sided P value is appropriate when you can state with certainty (and before collecting any data) that there either will be no difference between the means or that the difference will go in a direction you can specify in advance (i.e. you have specified which group will have the larger mean). If you cannot specify the direction of any difference before collecting data, then a two-sided P value is more appropriate. If in doubt, select a two-sided P value. Most recommend that you always calculate a two-sided P value.

Paired Or Unpaired Test?

When comparing two groups, you need to decide whether to use a paired test. When comparing three or more groups, the term paired is not appropriate and the term repeated measures is used instead.

Use an unpaired test to compare groups when the individual values are not paired or matched with one another. Select a paired or repeated-measures test when values represent repeated measurements on one subject (before and after an intervention) or measurements on matched subjects. The paired or repeated-measures tests are also appropriate for repeated laboratory experiments run at different times, each with its own control.

You should select a paired test when values in one group are more closely correlated with a specific value in the other group than with random values in the other group. It is only appropriate to select a paired test when the subjects were matched or paired before the data were collected. You cannot base the pairing on the data you are analyzing.

Fisher's Test Or The Chi-Square Test?

When analyzing contingency tables with two rows and two columns, you can use either Fisher's exact test or the chi-square test. The Fisher's test is the best choice as it always gives the exact P value. The chi-square test is simpler to calculate but yields only an approximate P value. If a computer is doing the calculations, you should choose Fisher's test unless you prefer the familiarity of the chi-square test. You should definitely avoid the chi-square test when the numbers in the contingency table are very small (any number less than about six). When the numbers are larger, the P values reported by the chi-square and Fisher's test will he very similar. The chi-square test calculates approximate P values, and the Yates' continuity correction is designed to make the approximation better. Without the Yates' correction, the P values are too low. However, the correction goes too far, and the resulting P value is too high. Statisticians give different recommendations regarding Yates' correction. With large sample sizes, the Yates' correction makes little difference. If you select Fisher's test, the P value is exact and Yates' correction is not needed and is not available.

Regression Or Correlation?

Linear regression and correlation are similar and easily confused. In some situations it makes sense to perform both calculations. Calculate linear correlation if you measured both X and Y in each subject and wish to quantity how well they are associated. Select the Pearson (parametric) correlation coefficient if you can assume that both X and Y are sampled from Normal populations. Otherwise choose the Spearman nonparametric correlation coefficient. Don't calculate the correlation coefficient (or its confidence interval) if you manipulated the X variable.

Calculate linear regressions only if one of the variables (X) is likely to precede or cause the other variable (Y). Definitely choose linear regression if you manipulated the X variable. It makes a big difference which variable is called X and which is called Y, as linear regression calculations are not symmetrical with respect to X and Y. If you swap the two variables, you will obtain a different regression line. In contrast, linear correlation calculations are symmetrical with respect to X and Y. If you swap the labels X and Y, you will still get the same correlation coefficient.

Selecting a Statistical Test for Common Situations

	Type of Data			
Goal	**Measurement from Normal Population**	**Rank, Score, or Measure from Non-Normal Population**	**Binomial (Two Possible Outcomes)**	**Survival Time**
Describe one group	Mean, SD	Median, interquartile range	Proportion	Kaplan Meier survival curve
Compare one group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square or Binomial test	-
Compare two unpaired groups	Unpaired t test	Mann-Whitney test, Nominal data: Fisher's Exact (small sample), Chi-square (large sample). Ordinal Data: Wilcoxon Rank Sum	Fisher's (small sample), Chi-square (large sample)	Log-rank test or Mantel-Haenszel
Compare two paired groups	Paired t test	Wilcoxon Signed Rank, Sign test (small sample), McNemar's test (large sample)	Sign test (small sample), McNemar's test (large sample)	Conditional proportional hazards regression
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test	Cox proportional hazard regression
Compare three or more groups (matched or unmatched)	Repeated-measures ANOVA or MANOVA	Friedman test	Cochrane Q	Conditional proportional hazards regression
Compare groups with known association to other variables	ANCOVA, MANOVA (Principle Components & Factor Analysis)	-	-	-
Quantify association between two variables	Pearson’s r	Nominal: Relative Risk Odds Ratio. Ordinal: Spearman Rho, Kendall’s Tau	Contingency coefficients	-
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression	Simple logistic regression	Cox proportional hazard regression
Predict value from several measured or binomial variables	Multiple linear regression or Multiple nonlinear regression	-	Multiple logistic regression	Cox proportional hazard regression

Developed from: Intuitive Biostatistics, *H.J. Motulsky, Ch. 37, Oxford University Press, 1995. &*
Hermansen, M. Biostatistics: Some Basic Concepts. Caduceus Medical Publishers. 1990.