Appendix D

Basic Measurement And Statistics

The following information was developed by Steven Rothke, PhD, Department of Psychology, Rehabilitation Institute of Chicago (RIC) and expanded by Mary F. Schmidt, PhD, Midwest Regional Head Injury Center for Rehabilitation and Prevention, RIC. Used with permission of the authors.

I. MEASUREMENT

There are two different basic types of numerical data.

In **categorical data,** numbers are used to represent categorical labels, but do not necessarily have any actual rank order or mathematical significance. Race, gender, and FIM scores are good examples.

In **continuous data,** the numbers imply a rank order along a continuum. The data is rank ordered and proportionate. Temperature and blood pressure are good examples.

Numerical variables are classified into four different kinds of **scales** based on the type of data they represent.

Scales for **categorical** data:

1) ***Nominal Scale:*** The numbers represent a verbal or categorical label such as 1 = male, 2 = female. The numbers do *not* represent a quantity. The variables are different in quality not quantity.

2) ***Ordinal Scale***: The numbers represent a verbal or categorical label and a *rank ordering* is implied, e.g., 1 = normal, 2 = mild, 3 = moderate, 4 = severe. Although a rank ordering is implied, the numbers used do not convey any meaning about the relationship between categories or rankings (i.e., in the above example, 4 is not necessarily twice as impaired as 2).

Scales for **continuous** data:

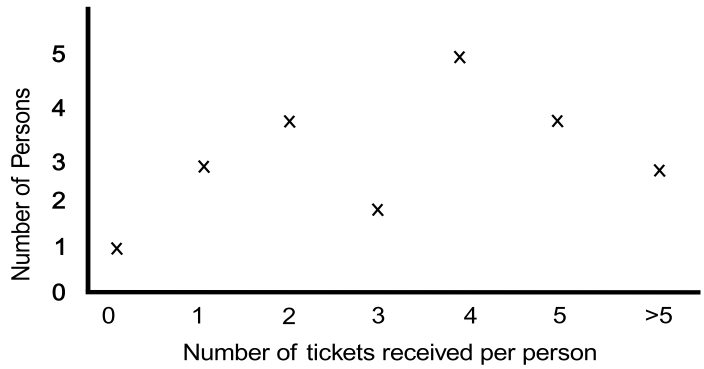
3) ***Interval Scale****:* Interval measurements have the property of equal distances or intervals between values on a scale as well as the property of ordinal ranking. For example, a thermometer measures heat in degrees which are the same “interval” or same size at any point on the scale, i.e., the distance between 11û and 12û is the same distance between 32û and 33û (equal intervals). However, an interval scale has no true or actual zero point on a scale. Therefore, one cannot say that 30û is twice as warm as 15û -- one can only speak of equal intervals and ordinal ranking. (0ûC is not an absence of heat, it is only the point at which water freezes.)

4) ***Ratio Scale***: This scale has the properties of order, equal intervals *and* an absolute zero. Thus, comparisons can be made directly from one point on the scale to another (6 feet is twice as long as 3 feet).

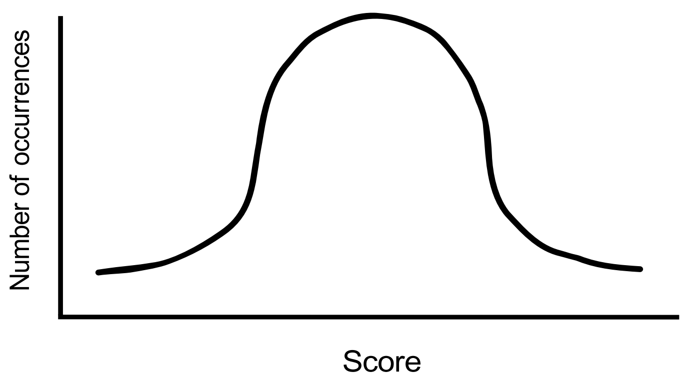
It is important to know which data types and scales you are using for your study, because this will directly affect your choice of statistical tests.

II. DISTRIBUTIONS

Many statistical assumptions and tests are based on the **distribution** of data values in a particular sample. Distributions are most easily visualized using a graph, where the X-axis shows the data values or scores and the Y-axis shows the number of times a particular score occurs. For example, imagine a group of 22 habitual traffic offenders is polled as to how many traffic tickets each has received. The resulting distribution might look like this:



While performing statistics you will frequently encounter the concept of the **normal distribution.** The normal distribution is a theoretical curve produced when members of a large population vary in the amount of a particular characteristic, such as height. Each population member's score is recorded and graphed, leading to a distinctive "bell-shaped curve."



The normal curve is the basis for **standard scores** (e.g., T-scores, Z scores). These standard scores convey information about the relative position of a score in the distribution (similar to a percentile rank) rather than information about the absolute value of a score (i.e., nothing about actual blood pressure but about where a person's blood pressure is relative to others).

Many human characteristics are normally distributed, including height. "IQ" test scores are supposedly normally distributed, with the mean IQ being 100.

Not all distributions are normal. Some are "skewed" positively (the mean is greater than the median due to extreme scores at the positive end) or negatively (mean is less than the median due to extreme low scores).

III. TYPES OF STATISTICS

Parametric and Nonparametric Statistics
Generally, scales using categorical data are analyzed using **non-parametric** statistics. Scales using continuous data are analyzed with **parametric** statistics.

An important distinction between **parametric** and **non-parametric** (or distribution-free) statistics is that parametric techniques require that certain assumptions be met. These include that the data are drawn from a normal distribution. Non-parametric statistics do not require any assumptions about the distribution of scores from which a sample was taken, which is why they are also called distribution-free statistics.

Descriptive and Inferential Statistics
Descriptive statistics *describe* the characteristics of the sample; they also you to extrapolate characteristics descriptive of the population from which the sample was drawn. Examples include the measures of central tendency and dispersion (see below).

Inferential statistics permit one to test hypotheses and make inferences about the sample. Examples include the *t*-test, chi-square test, analysis of variance, and regression analysis. Inferential statistics may be parametric or non-parametric, depending on how rigorous the researcher wants to be in meeting assumptions.

Descriptive Statistics: Types and Uses

A) **Measures of Central Tendency**

1) ***Mean:*** The numerical average of a set of scores is the sum of all scores divided by the number of scores.

   One disadvantage of using the mean is that in a small set of scores, the mean can be greatly affected by a single extreme score, or outlier.

2) ***Median:*** If all of the scores in a distribution are arranged in numerical order, the middle score is the median. (If the number of scores in the distribution is an even number, the median is the average of the two middle scores.) Unlike the mean, the median is not affected by extreme scores. However, depending on the shape of a distribution, it may not be a good descriptor of the sample.

3) ***Mode:*** The most frequently occurring score in a distribution (may or may not be near the mean or median). A distribution with two modes (two discrete peaks on the graph) is referred to as bimodal (e.g., the distribution of ages at which head injuries occur – one mode in the teens and twenties, and another in old age).

In a **normal distribution,** the mean, median and mode are **equal.**

B) **Measures of Dispersion**

1) ***Range:*** The interval between the lowest and highest scores (e.g., age range of patients in a study). A useful way of conveying the spread of scores.

2) ***Variance*** *and* ***Standard Deviation****:* Variance is a measure of the variability of scores around the mean. It is used to calculate other statistics including the standard deviation, a more useful descriptor of data distributions.

Variance:

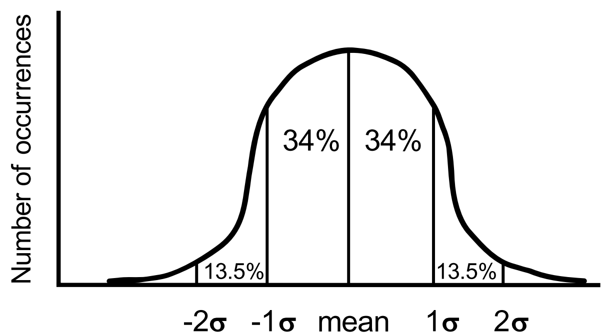
$$\sigma^2 = \frac{\Sigma(Y_1 - Y)^2}{n-1}$$

The ***standard deviation*** (s) is the average deviation (or difference) between each of the scores in the distribution and the mean.

$$\sigma = \sqrt{\frac{\Sigma(Y_1 - Y)^2}{n-1}}$$

When working with a normal distribution, the standard deviation tells you important things about your data.

For example, in a normal distribution, **68%** of the scores for your population will lie within one standard deviation of the mean, and **95%** of the scores will lie within two standard deviations of the mean.



In the standard "IQ" test, the mean IQ is considered to be 100, and the standard deviation is 15. Therefore, 68% of the population are expected to have IQs of between 85 and 115, and 95% should score between 70 and 130.

Inferential Statistics: Types and Uses

Univariate: analysis that involves only one variable at a time.

Bivariate: analysis that deals with the relationship of two variables.

Multivariate: analysis that focuses on more than two variables.

Major Univariate Statistics

Chi-Square: A nonparametric test. Used to test differences in levels of a nominal variable (e.g., political affiliation) when measured against another nominal variable (e.g., religion).

An example:

	Republican	Democrat	**Totals**
Catholic	234	590	824
Protestant	610	214	824
Totals	844	804	**1648**

Either variable can be considered the independent or dependent variable in this example. The numbers in the cells represent frequencies or percentages of total frequencies.

The chi-square test asks whether the differences in levels of the variables in question are significantly different from what we would expect to occur by **chance.** If the groups in the above example (political party and religion) were completely unrelated (chance only), the frequency table would look like this:

	Republican	Democrat	**Totals**
Catholic	422	402	824
Protestant	422	402	824
Totals	844	804	**1648**

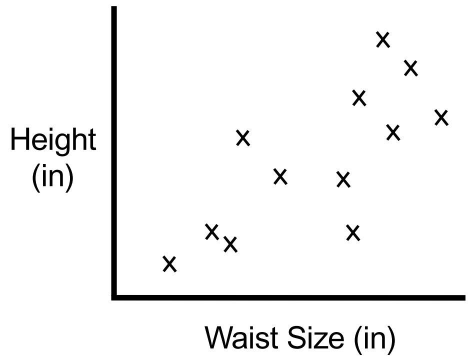
The chi-square value is calculated by comparing the actual data to the frequencies predicted by chance and arriving at a number representing the difference between the two. The *statistical significance* of this value can then be determined.

The chi-square test can also be used to test whether a frequency distribution fits a normal curve.

As with all tests of significance there are several limitations associated with the chi-square test. First, because chi-square depends on the way the categories are defined it is important to specify the ways in which your variables are classified. Chi-square is also one of the most frequently misused tests of significance, as it is often used when the expected frequencies in the cells are too small. If the sample size is not sufficiently large, the test is not valid.

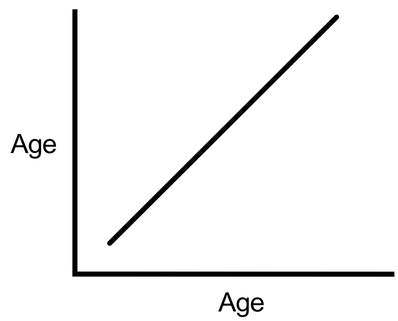
***

Correlation: For testing whether two variables **co-vary** with each other, i.e., if the level of one variable changes, there is a corresponding change in the other. Correlations are illustrated by plotting the variables against each other, one on the x-axis and one on the y-axis, like this:



A line is drawn representing the best fit between the points plotted on the graph. The *correlation coefficient* reports the slope of this line.

If two variables are perfectly correlated (for example, if you plot one variable against itself), the resulting correlation coefficient (*r*) is **1.00**, and the graph looks like this.



Correlation coefficients may range in value from -1.00 to 1.00. -1.00 is a perfect **negative** correlation; in other words, the two variables change in completely opposite directions, and the slope of the plotted line is -1.

Several types of correlation testing are available. When the scale of measurement of the variables is ordinal, use **Spearman Rank Order Correlation.** Example: order of finish in a race (first to last) versus order of weight of the racers (lowest to highest).

When the scale of measurement of the variables is interval or ratio, use **Pearson Product Moment Correlation**. Example: time to complete a race (in seconds) versus weight of the racers (in kilograms).

***

T-test: Used to test differences between the **means** of two groups. The independent variable (treatment) can be nominal (e.g., type of medication) or ordinal (high versus low dose of medication). The dependent variable must use an interval or ratio scale (e.g., blood pressure).

If the means of the two groups are judged to be significantly different from each other, the independent variable is assumed to have had an effect.

***

Analysis of variance (ANOVA): Similar to a t-test, but allows the use of more than one independent variable, or of an independent variable with more than two levels.

Example: the experimental design below assesses the effects of two multilevel independent variables (medication and type of exercise program) on mean systolic blood pressure.

		Exercise Program A	B	C	D
Drug	Diazepam	140	130	135	141
	Dantrolene	190	135	159	158
	Baclofen	125	145	150	152

***

NB: t-tests and ANOVA are used when the independent variables are categorical (nominal or ordinal) while the dependent variable is continuous (interval or ratio). When both the independent and dependent variables are continuous, use **regression** analysis (use **multiple regression** for more than one independent variable). The details of regression are beyond the scope of this discussion.

IV. TESTING AND PREDICTION: TERMS AND MEASURES

When assessing the capabilities and proposed function of a test, it is important to know these terms.

Sensitivity (true positive rate): the ability of a test or procedure to identify a specified condition when it is *actually present*.

Specificity (true negative rate): the ability of a test or procedure to identify the absence of a condition or situation when it is *actually absent.*

False Positive: When a test indicates a condition is present when it is actually absent. A large number of false positives reduces the specificity of a test.

False Negative: When a test actually indicates a condition is absent when it is actually present. A large number of false negatives reduces the sensitivity of a test.

Example:

		Disease or Condition present	absent
Test Results	positive	a: true positive	b: false positive
	negative	c: false negative	d: true negative

Sensitivity = a / (a + c)

Specificity = d / (b + d)

Hypothesis Testing and Significance

When you design a study, you are doing so with the intent to prove a particular **hypothesis**. In statistical jargon, you are also attempting to **disprove** a **null hypothesis**. The **null hypothesis** (H_0) of any study states that the independent variable has **no effect** on the dependent variable(s) other than that which would be expected to occur by chance alone. The experimenter's ultimate goal is to disprove the null hypothesis and prove his/her own hypothesis (H_1).

Hypothesis testing follows a paradigm very similar to the one used for sensitivity and specificity.

Type I Error: Rejecting the null hypothesis (i.e., accepting your alternate or desired hypothesis) when the null hypothesis is actually true. The probability of a Type I error is referred to as **alpha.**

Type II Error: Accepting the null hypothesis when it is actually false. The probability of a Type II error is referred to as beta.

		Reality	
		H_0 is True	H_0 is False
Results of Study	Reject H_0	Type I Error (alpha)	Correct Decision (power)
	Accept H_0	Correct Decision (confidence)	Type II Error (beta)

Power: The likelihood of rejecting a false null hypothesis; in other words, the likelihood of correctly identifying an experimental effect. The power of a statistical procedure is comparable to the sensitivity of a test.

$$\text{Power} = (1 - \text{beta}).$$

Confidence: The likelihood of accepting a true null hypothesis. The confidence of a statistical procedure is comparable to the specificity of a test.

More on Alpha: When statistical tests are used, the experimenter sets a criterion **alpha level** beyond which s/he will assume statistical significance of an observed effect. The alpha level most often used is **0.05.** If the **p-value** generated by a statistical test is $<= 0.05$, that means that the likelihood is less than 5% (0.05) that a Type I error will be made. In other words, there is a less than 5% chance of a false positive result (true null hypothesis, but the alternative hypothesis is accepted).

V. EPIDEMIOLOGICAL STATISTICS

Incidence: The number of new cases of a condition or disease in a specified period of time (e.g., the number of new cases of AIDS in Chicago in 1992).

Prevalence: The number of existing cases of a condition or disease at a particular point in time (e.g., the number of people living with AIDS in Chicago at the present time).