LECTURE 6: INTRODUCTION TO INFERENTIAL STATISTICS AND HYPOTHESIS TESTING

"Inferential statistics are used to help us look beyond raw data and descriptive statistics. They help us make inferences about population parameters" (Phakiti, 2010, p. 44). "Given that it is impossible to gather data from all members of the population, inferential statistics can allow researchers to generalize findings to other, similar language learners; that is, to make inferences." (Mackey & Gass, 2005, p. 269).

1. Assumptions for inferential statistical tests

There are two main categories of inferential statistical tests, namely: **1) parametric tests**, such as the t-test, One-Way ANOVA, and Pearson correlation, and **2) non-parametric tests**, such as the Mann-Whitney test, Kruskal walis test, Spearman correlation. "Although parametric tests are more preferable in quantitative research, non-parametric tests are [also] important for applied linguistics research because some data are not always strongly interval or continuous." (Phakiti, 2010, pp. 45-46).

A number of assumptions (conditions) should be met before the researcher can decide on using the appropriate inferential testing category with their data. "These assumptions (conditions) are not optional and if they are not met, there is a heightened risk of making a false inference." (Phakiti, 2010, p. 45). The conditions of each testing category are summarized in the table below:

ASSUMPTIONS	PARAMETRIC TESTS	NON-PARAMETRIC TESTS
Type of data	Numerical data	Categorical data
Sampling approach	Random sampling	Non-random sampling
Distribution of results	Normal distribution of data	Non-normal distribution of data
Homogeneity of variances	Homogeneous variances	Heterogeneous variances

Assumption 1: Type of data

The type of data is essential in determining the appropriate inferential statistical tests. Usually, numerical data are used in parametric tests; while categorical data are used in non-parametric tests.

Assumption 2: Sampling approach

Because the main aim of inferential statistics is to generalize the results from the sample to the population, it is preferable to have a sample that is randomly recruited from the population. If the sample is randomly recruited, parametric tests can be used; otherwise, non-parametric tests can serve as an alternative.

Assumption 3: Distribution of results

A distribution describes the clustering of scores in a dataset. In a normal distribution, the numbers (e.g., scores on a particular test) cluster evenly around the mean. ***Figure 10.7*** shows that the measures of central tendency coincide at the midpoint. Parametric tests require the data to be normally distributed. However, if the data are not normally distributed, non-parametric tests should be used. The two well-known tests of normality, namely, the Kolmogorov–Smirnov test and the Shapiro–Wilk test are the most widely used methods to test the normality of data.

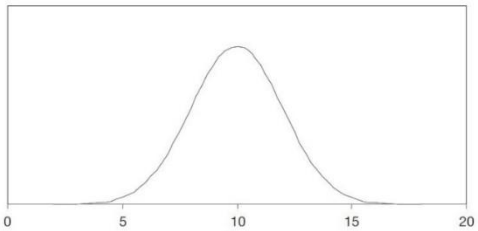


FIGURE 10.7 Normal distribution

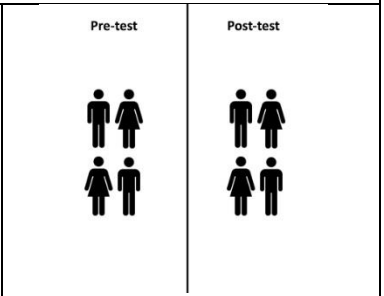
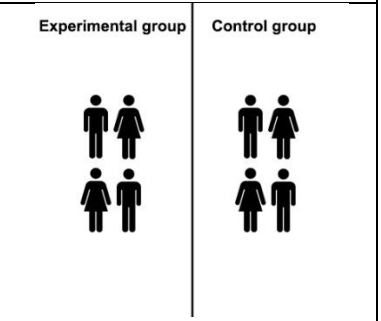
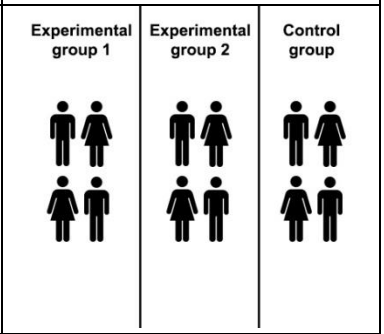
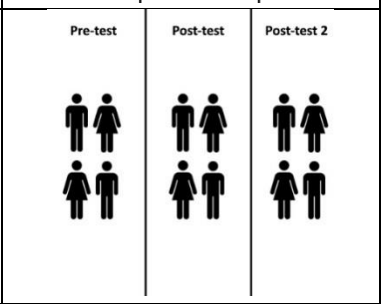
Assumption 4: Homogeneity of variance

Parametric tests require the groups being tested to have similar variances (i.e., homogenous variances). There are a number of tests that examine homogeneity of variances, but the most robust test is the Levene's test.

IMPORTANT NOTE. *Non-parametric tests should be used if any of the above assumptions are not met.*

2. Classification of inferential statistical tests

The following table classifies the most widely used inferential statistical tests and summarizes their purposes.

PARAMETRIC TESTS	EQUIVALENT NON-PARAMETRIC TESTS	PURPOSE	
Paired t-test	*Wilcoxon Rank sum test*	To test difference between the means of two related samples (e.g., testing students vocabulary scores before and after a treatment [pre-test and post-test designs])	
Independent t-test	*Mann-Whitney U test*	To test difference between the means of two independent samples (e.g., experimental group vs. control group)	
ANOVA (One way analysis of variance)	*Kruskal Wallis Test*	To test the difference between the means of more than two independent samples (e.g., control group vs. experimental group1 vs. experimental group 2)	
Repeated measures ANOVA	*Friedman test*	To test the difference between the means of more than two related samples (e.g., testing the pronunciation of EFL students in the first, second, and third semesters)	
Pearson correlation r	*Spearman correlation*	To measure correlation between two sets of data (e.g., testing the relationship between classroom attendance and Oral Expression exam results)	

N.B. *On the electronic version of this file, the test names above contain links that will direct you to YouTube tutorials demonstrating how you can conduct each test on the software package for statistics SPSS.*