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. Parametric vs. non-parametric tests

There are two main categories of inferential statistical tests, namely: **1) parametric tests** that are conducted with interval and ratio data (e.g., t-test, One-Way ANOVA, Pearson correlation), and **2) non-parametric tests** that are conducted with nominal and categorical data (e.g., 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 conditions should be met before the researcher can decide on using the appropriate inferential testing category with their data. "These [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:

	Parametric tests	**Non-parametric tests**
Type of data	Interval/ Ratio data	Nominal/ Categorical data
Sample size	Large sample sizes	Small sample sizes
Distribution of results	Normal distribution of data	Non-normal distribution of data

The subsections below provide more details on the conditions of choosing parametric or non-parametric tests.

A. Condition 1: Type of data

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

B. Condition 2: Sample size

Because the main aim of inferential statistics is to generalize the results from the sample to the population, it is preferable to have a large sample size. When choosing inferential statistical tests, studies conducted with large samples sizes often use parametric tests; meanwhile studies conducted with a small sample size resort to using non-parametric tests.

C. Condition 3: Distribution of results

Before conducting an inferential test, it is also important to check the distribution of the data set. If the data are normally distributed, the researcher can proceed to using parametric tests. However, if the data are not normally distributed, non-parametric tests should be used. A distribution describes the clustering and behavior of scores in a dataset. In a normal distribution, the numbers (e.g., scores on a particular test) cluster evenly around the mean (also known as a bell curve). As can be seen in the **Figure 10.7**, the three measures of central tendency (mean, mode, median) coincide at the midpoint. Thus, 50 percent of the scores fall above the mean and 50 percent fall below the mean.

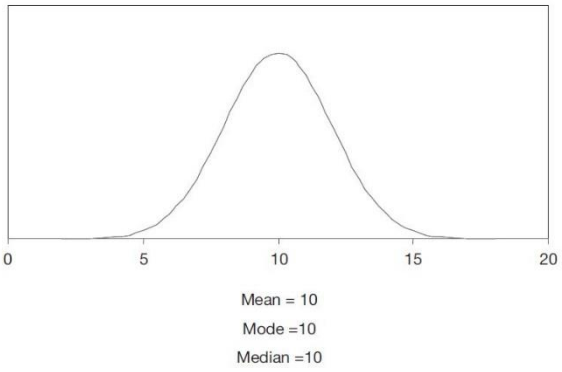
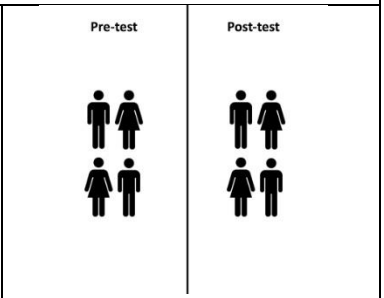
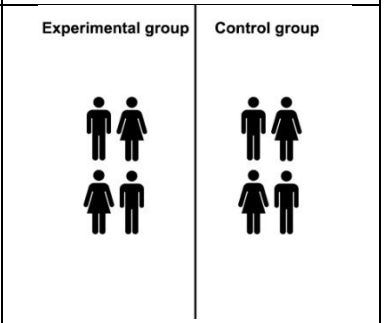
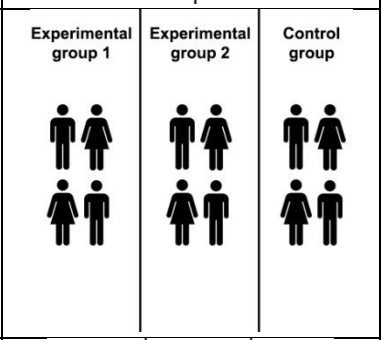
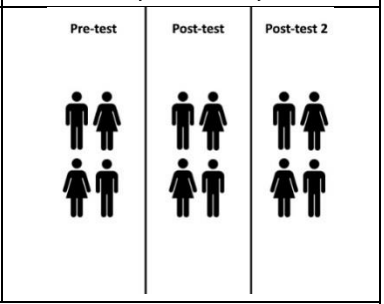


FIGURE 10.7 Normal distribution

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. Normality tests can be conducted in the statistical software "SPSS" (**analyze → descriptive statistics → explore → plots → normality plots with tests**).

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.*