1. Introduction

Welcome to the series of E-learning modules on introduction to statistical tests. In this module we are going cover the basic concepts of statistical tests, role of statistical tests, various types of tests and choice of the test.

By the end of this session, you will be able to:

- Explain statistical tests
- Describe the role and importance of statistical tests
- Describe various types of statistical tests
- Understand how to choose the test
- Understand advantages and disadvantages of Parametric and Non Parametric tests

Traditional scientific inquiry consists of four interrelated stages:

i. Problem Definition,
ii. Data Gathering,
iii. Data Analysis, and
iv. Data Interpretation

Purposes of Statistical Analysis

The general purpose of statistical analysis is to provide meaning to what otherwise would be a collection of numbers and/or values.
The "meaningfulness" of data derives from the clarity with which one specifies the problem or questions being addressed (this is Stage one of the inquiry process) and the precision with which pertinent information is gathered (this is Stage two).

Statistical procedures fall loosely into three general categories: descriptive, associative, and inferential.

Descriptive statistics portrays individuals or events in terms of some predefined characteristics. Addressing the question "**What is the average diastolic blood** pressure of middle-age women?" is a problem definition suitable for analysis by descriptive statistics.
Associative statistics seek to identify meaningful interrelationships between or among data. Addressing the question "Is there a **relationship between** salt intake and diastolic blood pressure among middle-age women?" is a problem definition suitable for analysis by associative statistics.
Inferential statistics seek to assess the characteristics of a sample in order to make more general statements about the parent population, or about the relationship between different samples or populations. Addressing the question "Does a low sodium diet **lower** the diastolic blood pressure of middle-age women?" represents a problem definition suitable for inferential statistics.

2. Levels of Measurement

Levels of Measurement
Once categorized according to the specific scheme (i.e., independent, dependent, intervening), variables must be measured. This, of course, is the basis for data gathering. Data gathering employs measurement scales or sets of rules for quantifying and assigning values to a particular variable.

Typically, four levels of measurement apply to data gathering. Data levels may be characterized as nominal, ordinal, interval and ratio.

Nominal Scale: The term **nominal** means to name. Hence, a nominal scale does not measure but rather names. A nominal variable thus consists of named categories.
Examples of nominal variables include gender, religion, and "group" assignment (such as treatment/no treatment).

Ordinal scale: The term **ordinal** means to order. In other words, an ordinal scale is a rank ordering of things, with a categorization in terms of more than or less than.
Examples of variables measured on an ordinal scale would be pain levels (on a high-medium-low scale), or the rank ordering of patients according to their diastolic blood pressure.

Interval scales not only tell the order of things; they also measure the distance between values.
For instance, assume you measure two patients' temperatures as forty one degrees centigrade and thirty seven degrees centigrade. Not only does the first patient have a higher temperature than the second, but his temperature is 4 degree centigrade higher.

Ratio measurement goes one step beyond interval scaling by providing an "absolute zero" point. With ratio measures, we can compare values not only according to the absolute interval between them, but also their relative magnitude. Thus, on a ratio scale, a variable with a value of forty represents **twice as much** of the quantity being measured as a value of twenty. Common examples of ratio level measurements include a patient's age, weight, height and pulse rate. For most statistical computations, ratio and interval data are considered equivalent.

The theory of hypothesis and the statistical tests were developed by J-Neymann and E.S Pearson.
In the theory of estimation we were mainly concerned with the estimation of certain population

parameters. However, ultimate purpose of the estimation involves some use of the estimates. Very often statistical tests are used to test certain hypothesis regarding the parameter.

For example: suppose a new method of sealing of bulbs has been developed. Now one may estimate the average life of bulbs sealed by the new method and may further make use of this estimate in comparing the new method of sealing with the old one, which is in finding whether or not the new method gives greater average life to the bulbs than the old method.

Here one may make the statement that the **new method is no better than the old method** and test whether the statement is **true or not** by using the estimate obtained from the sample values. The above problem of testing may be considered as a problem of **testing of hypothesis**.

In the example we find that estimation and testing of hypothesis are closely associated. **Statistical tests deals with the problems of testing of hypothesis.**

A **statistical hypothesis test** is a method of making decisions using data, whether from a **controlled experiment** or **an observational study** (not controlled).
In statistics, a result is called **statistically significant**, if it is **unlikely** to have occurred by chance alone, according to a pre-determined threshold probability, the significance level.

The phrase "test of significance" was coined by Ronald Fisher, who said, "Critical tests of this kind may be called **tests of significance**, and when such tests are available we may discover whether a second sample **is** or **is not** significantly different from the first."

Statistical hypothesis testing is a key technique of **frequentist statistical inference**.
The Bayesian approach to hypothesis testing is **to base rejection** of the hypothesis on **the posterior probability**.
Other approaches to reaching a decision based on data are available via **decision theory and optimal decisions.**

Once you have looked at the distribution of your data and perhaps conducted some descriptive statistics to find out the mean, median or mode, it is time to make some inferences about the data.

As previously covered in the module, inferential statistics are the set of statistical tests we use to make inferences about data. These statistical tests allow us to make inferences because they can tell us if the pattern we are observing is real or just due to chance.

3. Types of Statistical Tests

Types of statistical tests:
There are three types of Statistical tests

i. Parametric tests
ii. Non Parametric tests
iii. Sequential tests

If the distribution of the parent population is known, or if the test requires the specification of the parameters then such a test is known as the parametric tests. Some of the parametric tests are Z-test, t-test, F-test, ANOVA etc

If the test does not require the knowledge of the parent population or in other words if the test does not require any such specification of the parameters it is known as Non Parametric tests. For example: Chi-square test, Sign test, Run test, Mann Whitney U test etc.

Testing of hypothesis may be based on samples of fixed size or based on samples of varying size.
A test procedure in which a sample size is not fixed in advance which will be decided only at the end of the test procedure is known as sequential tests.
In this paper let us concentrate on two types of statistical tests, that is, Parametric tests and Nonparametric tests.

Parametric Assumptions

- The observations must be independent
- The observations must be drawn from normally distributed populations
- These populations must have the same variances
- The means of these normal and homoscedastic populations must be linear combinations of effects due to columns and/or rows

Nonparametric Assumptions

- Observations are independent
- Variable under study has underlying continuity

Parametric Methods:
These methods are very powerful statistical tools.
Common parametric statistical tests:

a. **One-sample tests** – these tests are appropriate when a sample is being compared to the population from a hypothesis. The population characteristics are known from theory or are calculated from the population.

b. **Two-sample tests**- these tests are appropriate for comparing two samples, typically experimental and control samples from a scientifically controlled experiment.

c. **Paired tests** are appropriate for comparing two samples where it is impossible to control important variables.

Rather than comparing two sets, members are paired between samples so the difference between the members becomes the sample. Typically the mean of the differences is then compared to zero.

d. **Z-tests** are appropriate for comparing means under stringent conditions regarding normality and a known standard deviation
e. **t-tests** are appropriate for comparing means under relaxed conditions (less is assumed)
f. Tests of proportions are analogous to tests of means (the 50 percent proportion)
g. Chi-square tests for variance are used to determine whether a normal population has a specified variance
h. Chi-square goodness of fit tests are used to determine the adequacy of curves fit to data
i. F-tests
   F tests (analysis of variance, ANOVA) are commonly used when deciding whether groupings of data by category are meaningful.

If the variance of test scores of the left -handed in a class is much smaller than the variance of the whole class, then it may be useful to study left-handed students as a group.
The null hypothesis is that two variances are the same - so the proposed grouping is not meaningful.

Non Parametic methods
There is at least one nonparametric test equivalent to a parametric test.
These tests fall into several categories as mentioned below:

a) Tests of differences between groups (independent samples)
b) Tests of differences between variables (dependent samples)
c) Tests of relationships between variables
d) Two samples –these compare mean value for some variable of interest

Chi square test of independenc are used for deciding whether two variables are associated or are independent.
The variables are categorical rather than numeric.
It can be used to decide whether left-handedness is correlated with libertarian politics (or not).
The null hypothesis is that the variables are independent.

4. Advantages, Critisicsm & Choosing the Right Test

Advantages of NP tests

- Probability statements obtained from most nonparametric statistics are exact probabilities, regardless of the shape of the population distribution from which the random sample was drawn
- If sample sizes as small as n equal to six are used, there is no alternative to using a nonparametric test
- Can treat samples made up of observations from several different populations
- Can treat data which are inherently in ranks as well as data whose seemingly numerical scores have the strength in ranks
- They are available to treat data which are classificatory
- Easier to learn and apply than parametric tests

Criticism

- Precision is lost or there is wastefulness of data
- Low power
- False sense of security
- Lack of software
- Tests distributions only
- Higher-ordered interactions not dealt with

Power of a test

Statistical power is the probability of rejecting the null hypothesis when it is, in fact false, and should be rejected.

- Power of parametric tests is calculated from formula, tables, and graphs based on their underlying distribution
- Power of nonparametric tests is less straightforward; calculated using Monte Carlo simulation methods

Non Parametric tests are approximate tests and in case parametric tests exist, they are more powerful than the Non parametric tests.
Nonparametric tests are designed to test the statistical hypothesis only and not for estimating the parameters.

Parametric tests are preferred because, in general, for the same number of observations, they are more likely to lead to the rejection of a false null hypothesis.

That is, they have more power.
This greater power stems from the fact that if the data have been collected at an interval or ratio level, information is lost in the conversion to ranked data (that is, merely ordering the data from the lowest to the highest value).

Nonparametric tests are also referred to as **distribution-free** tests.
These tests have the obvious advantage of **not requiring** the assumption of **normality** or the assumption of **homogeneity** of variance.
They compare **medians** rather than **means** and, as a result, if the data have one or two outliers, their influence is negated.

Generally, running nonparametric procedures is very similar to running parametric procedures, because the same design principle is being assessed in each case. So, the process of identifying variables, selecting options, and running the procedure are very similar. The final p-value determines significance or not in the same way as the parametric tests.

Some of the non parametric tests equivalent to parametric tests are
The one sample test of the parametric type has nothing quite comparable in the nonparametric type. The paired sample t test can be compared with the Wilcoxon t test.
Independent sample t test of the parametric type is cab be compared with the Mann-Whitney U test. Pearson's correlation of the parametric test can be compared to Spearman's correlation of the nonparametric analogue.

How to choose the appropriate test?
To choose the appropriate statistical test, first categorize the variables as independent and dependent (intervening or nuisance variables are usually treated as additional independent variables).
Next, determine the number of independent and dependent variables in the study.
Finally, determine the level of measurement that is nominal, ordinal or interval, applied to each relevant variable.

How do you know what kind of test to use?
There are a wide range of statistical tests.
The decision of which statistical test to use depends on the research design, the distribution of the data, and the type of variable.
In general, if the data is normally distributed, we choose from parametric tests. If the data is non-normal, we choose from the set of non-parametric tests.

5. Testing Process, Uses and Importance of Statistical Tests

The testing process

In the statistical literature, statistical hypothesis testing plays a fundamental role. The usual line of reasoning is as follows:

There is an initial research hypothesis of which the truth is unknown

1. The first step is to state the relevant null and alternative hypotheses

The second step is to consider the statistical assumptions being made about the sample in doing the test; for example, assumptions about the statistical independence or about the form of the distributions of the observations. This is equally important as invalid assumptions will mean that the results of the test are invalid

Step 3: Decide which test is appropriate, and state the relevant test statistic

Step 4: Derive the distribution of the test statistic under the null hypothesis from the assumptions. In standard cases this will be a well known result. For example the test statistic may follow a Student's t distribution or a normal distribution

Step 5: Select a significance level (alpha), a probability threshold below which the null hypothesis will be rejected. Common values are five percent and one percent.

Step 6: The distribution of the test statistic under the null hypothesis partitions the possible values of **statistic** into those for which the null-hypothesis is rejected, the so called critical region, and those for which it is not. The probability of the critical region is **alpha**.

Step 7: Compute from the observations the observed value of the test statistic**.**

Step 8: Decide to either fail to reject the null hypothesis or reject it in favour of the alternative.

An alternative process that is commonly used is as follows:

Compute from the observations the observed value **of** the test statistic**.**

From the statistic, calculate a probability of the observation under the null hypothesis (the p-value).

Reject the null hypothesis or not. The rule is to reject the null hypothesis if and only if the ‘p’ value is less than the significance level that is the selected probability threshold.

The two processes are equivalent. The former process was advantageous in the past when only tables of test statistics at common probability thresholds were available.

It allowed a decision to be made without the calculation of a probability.

It was adequate for class work and for operational use, but it was deficient for reporting results.

The latter process relied on extensive tables or on computational support, not always available.
The explicit calculation of a probability is useful for reporting.
The calculations are now trivially performed with appropriate software.

Use and Importance of statistical tests:

Statistical tests are helpful in analyzing most collections of data.
This is equally true of hypothesis testing which can justify conclusions even when no scientific theory exists.

Real world applications of hypothesis testing include:

- Testing whether more men than women suffer from nightmares
- Establishing authorship of documents
- Evaluating the effect of the full moon on behaviour
- Determining the range at which a bat can detect an insect by echo
- Deciding whether hospital carpeting results in more infections
- Selecting the best means to stop smoking
- Checking whether bumper stickers reflect car owner behaviour
- Testing the claims of handwriting analysts

Conclusion:
The choice of the correct statistical test depends upon the definition of the variables, and particularly upon their level of measurement.
It also depends upon the research design used, and the nature of the hypotheses: are they comparative or is there a relationship; is there more than one independent variable?

The answers to the following six questions will isolate the correct statistical test.

a. How many independent variables covary with the dependent variable?
b. At what level of measurement is the independent variable?
c. What is the level of measurement of the dependent variable?
d. Are the observations independent or dependent?
e. Are you comparing populations to populations, a sample to a population, or are you comparing two or more samples?
f. Is the hypothesis being tested comparative or relationship based?

Using the answers to the above six will point you in the direction of the correct statistical procedures.

Here’s a summary of our learning in this session where we have:

- Understood the basic concept of statistical tests
- Explained the different types of statistical tests
- Understood the role and importance of the tests
- Understood how to choose the right test
- Described the difference between Parametric and Non parametric tests