Course : Research statistics **Lecturer : Pr. Saliha CHELLI**

Level : Master

Lecture : 7

Inferential statistics

Parametric/ Non-parametric tests

Lecture objectives: Introducing inferential statistics / non parametric and parametric tests

Introduction

Descriptive statistics make no inferences or predictions, they simply report what has been found.Inferential statistics, by contrast, make inferences and predictions based on the data gathered. This includes, for example, hypothesis testing, correlations, different testing. Inferential statistics are often more valuable for researchers and typically they are more powerful. This lecture introduces you to this type of statistics which are based on the findings of descritptive statistics and helps you complete your research by testing the hypothese formulated in your research study.

1. Inferential statistics

Inferential statistics infer from the data whether the predicted effect of the independent variable actually occurred in the experience. We are making inferences from observable data to causal relationships between variables (Miller, 1985, p.41).

- Inferential statistics infer from the sample to the population.
- They determine probability characteristics of population based on the characteristics of the sample.
- They allow to generalize the findings to a larger group.

2. Statistical significance

The main concern of inferential statistics has traditionally been the testing of '**statistical significance'**. Statistical significance denotes whether a particular **result in a sample is true for the wholepopulation**. If the result is non-significant, this means that we cannot be certain that it did not **occur by chance.**

Significance is measured by probability coefficient (p), which can range form **0 to + 1. A p of 0.25** means that the obtained result might be due to pure chance in **25 percent of the cases**. In social sciences, significance is typically measured by **0.05**. This means that the result might be **due to pure chance in 5%of the cases.**

3. Statistical tests

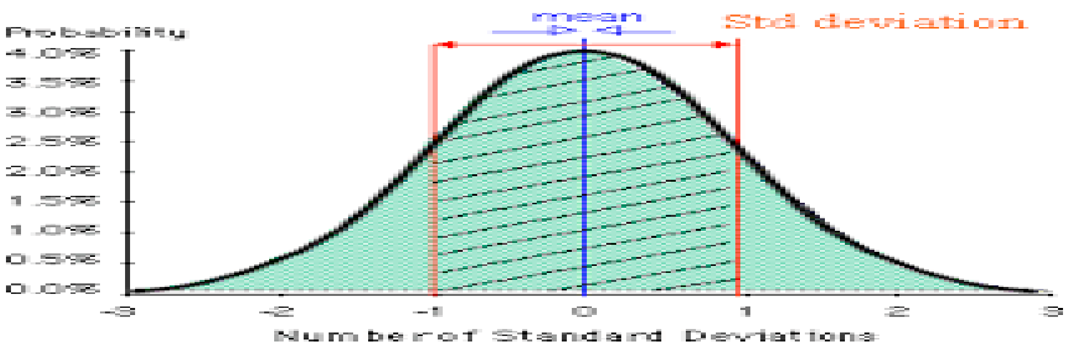
Probability theory allows to produce **test statistics** using mathematical formulas. **A test statistic** is a numerical value that is used to decide whether **to accept or reject the null hypothesis.** A statistical test is simply a device for calculating the likelihood that our **results are due to chance fluctuation** between the groups. Different tests calculate this likelihood in different ways, depending on the design of the experiment and the nature of the dependent variable (Miller, 1984, p. 42).

A statistical test is used to determine the probability that the observed results could have occurred under the null hypothesis. If this probability is **less than, or equal to 0.05**. **the null hypothesis is rejected** in favour of the alternate hypothesis and **the results are said to be significant**

4. Normal distribution

Normal distribution is the most important probability in statistics.It is an arrangement of a data set in which most values cluster around the central peak and the rest taper off symetrically toward either extreme.

A normal distribution has a bell shaped density curve by its mean and standard deviation. The density curve is symmetrical, centered about its mean, with its spread determined by its standard deviation.



The bell-shaped curve indicates that most values fall near the central value , with fewer from the centre and the rest fall symetrically.

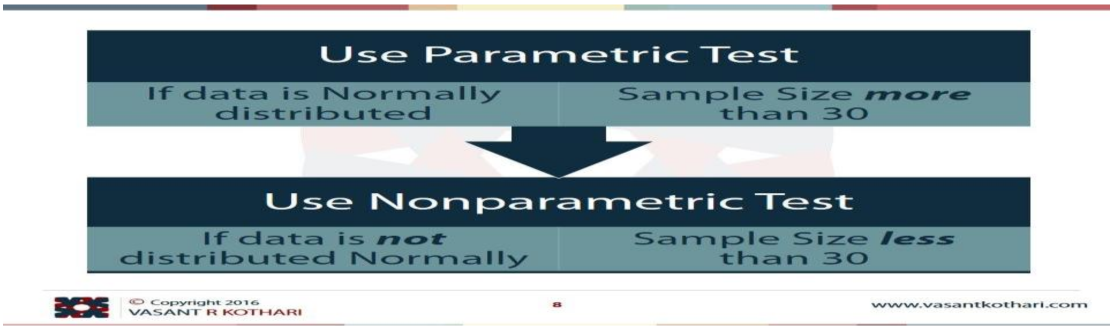
*If **the sample size is large**, it is assumed that the **distribution is normal.**

5. Testing hypotheses

To test a hypothesis, a **statistical test** is required. **A *statistical test*** is used to **determine the probability that the observed results could have occurred under the null hypothesis.** If this probability is **less than, or equal to 0·05 the null hypothesis is rejected in favour of the alternate hypothesis,** and **the results are said to be *significant***

6. Parametric versus parametric tests

There are basically two types of statistical tests—parametric and non-parametric



6.1. Difference between parametric and non- parametric tests

Parametric tests	Non-parametric tests
• The measurement of variables is done on interval and ratio level. • The measue of central tendency is the mean	• The measurement of variables is done on nominal or ordinal level. • The measue of central tendency is the median.

6.2 Examples of Parametric and non-parametric tests

	Parametric tests	Non parametric tests
Assumed distribution	Normal	Any
Typical data	Numerical	Ordinal or nominal
Usual central measure	Mean	Median , mode
Advantages	Can draw more conclusions	Simplicity ; less affected by outliers
Describe one group	Mean , SD	Median, interquartile range Proportion
Independent measures,2 groups	Unpaired t test	Chi-square test, Fisher's test Mann-Whitney U test
Independent measures ,>2 groups	ANOVA	Kruskal – Wallis test Chi-square test
Repeated measures,2 conditions	Paired t test	Wilcoxon sign rank, Mc Nemar's Chi-square test
Repeated measures,>2 conditions	ANOVA	Friedman's test Chi-square test
Regression	Simple linear regression or Non-linear regression	Non parametric regression

10

7. How to choose a statistical test?

There are two major factors that determine the correct test for any particular set of experimental results: (1) the research design and (2) the nature of the dependent variable, that is, the actual data (Miller, 1984,p.53). From the list of statistical tests seen in slide 15, the following will be presented:

- The **Man-whitneyU test** and the **chi-square test**
- The **paired and unpaired tests**

The Mann-Whitney makes no assumptions about the populations from which the samples have been drawn (**non-parametric**), and can be used with **ordinal or interval scaling** (Miller,1984,p.77).**Example,** a typical application of the test might involve the comparison of two independent groups of subjects who were taught to read by different methods and subsequently rated on a scale of reading 'fluency' from 1 to 10. (ibid.69).

The Chi-square statistic is a **non-parametric (**distribution free) tool designed to analyze group diffrences when the dependent variable is measured at a **nominal level**. The chi-square test is useful for the comparison of groups in which subjects' behaviour has been classified into **discrete, qualitatively different categories**. The observations are then *frequency* ***counts*** of the number of individuals coming within each category. The test may be used to determine whether two groups differ in terms of the proportion of individuals falling into one category rather than another ((Miller, 1984,p.77). **An example of data** suitable for chi-square is **sex of person** and **choice of favourite leisure entertainment.**

Both paired (dependent samples t-test)and unpaired (independent samples t-test) tests check if the difference between two means is significant. **The paired-samples t-test (equivalent to a one-sample-t-test)** compares scores of the same group at two different times. **The unpaired t-test, called the independent samples t-test (a two-samples t-test)** compares scores for two different groups.

The most statistical tests include:Chi-square test: can be used for nominal (categorical) data to determine whether a relationship between categorical data is likely to reflect a real association between these two variables in the population. The t-test: allows the comparison of the mean of two groups. ANOVA test: Analysis of variance : allows the comparison of three or more groups

8. Types of analyses

Univariate analysis: the analysis of one variable: mean, median, mode and standard deviation, example, How many students have the average?

Bivariate analysis: is a kind of data that explores the association between two variables

- Pearson's correlation test
- T-test
- Spearrman Rho correlation test
- Mann- Whitney test
- Linear regression test

Multivariate: the analysis of more than two variables.

Some examples:

- Multi-regression (multiple linear regression): is a statistical technique that uses several explanatory variables (independent variable) to predict the outcome of a response variable (dependent variable).

References

- Howit, D, Cramer, D. (2005). First steps in research and statistics: A practical workbook for psychology students. Taylor & Francis Group: Routledge.

- Miller, S. (1984). Experimental design and statistics (2nd ed.). London and New York: Routledge