Statistics

Lecture 7 (1)

Inferential statistics

By Dr.Chelli

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 hypotheses formulated in your research study.

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 help assess the strength of the relationship the independent (causal variable) and the dependent (effect) variable.
- They allow to generalize the findings to a larger group.

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 whole population. If the result is non-significant, this means that we cannot be certain that it did not occur by chance.

Continued

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 we typically consider a result being significant if p 0.05 ,that is, if the probability of the result not being real but only due to chance. This means that the result might be due to pure chance in 5%of the cases.

Statistical tests

Probability theory allows to produce test statistics using mathematical formulas.

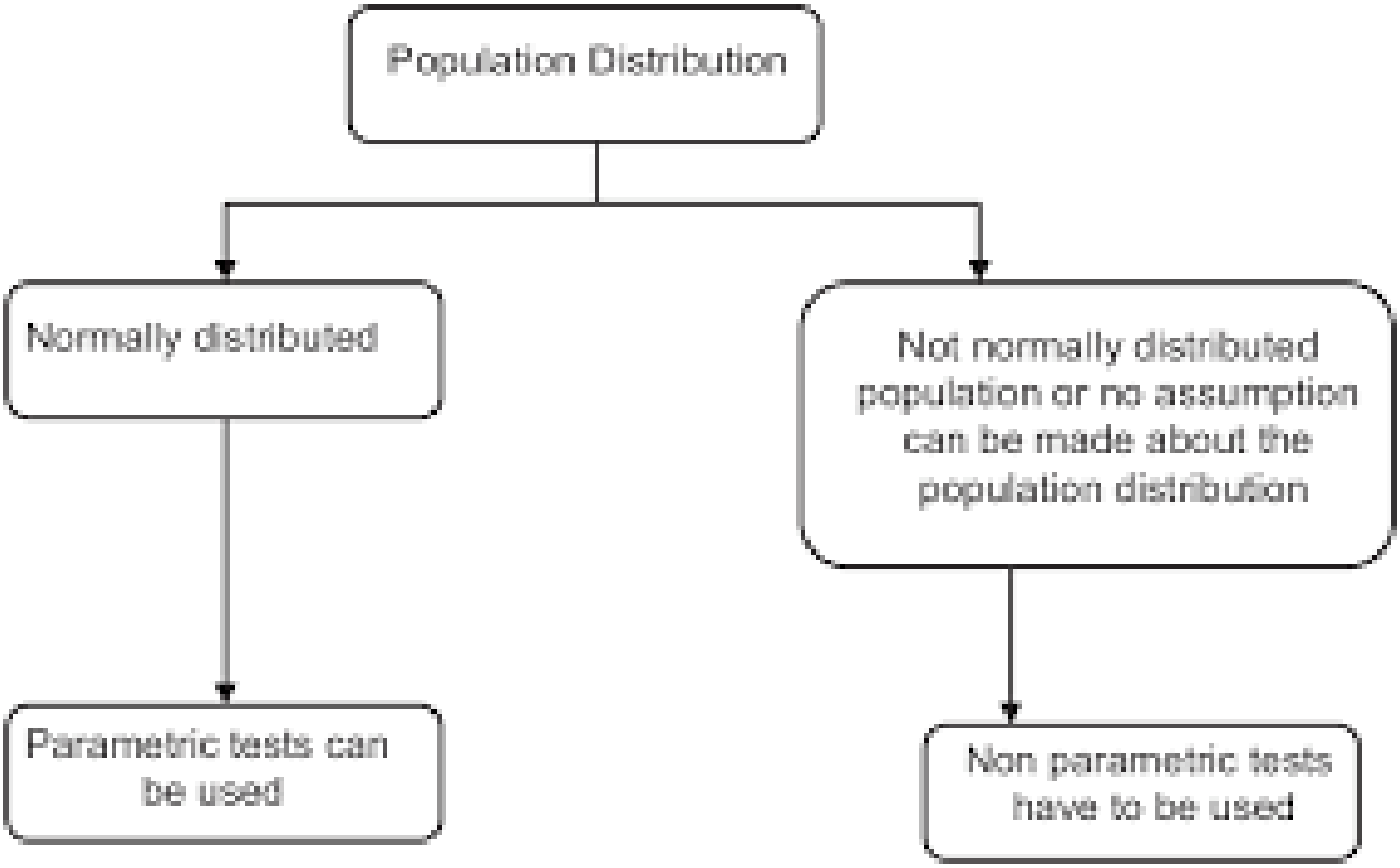
A test statistic is a numerical value that is used to decide whether to accept or reject the null hypothesis.

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).

continued

A statistical test is used to determine the probability that the observed results could have occurred under the null hypothesis. 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.

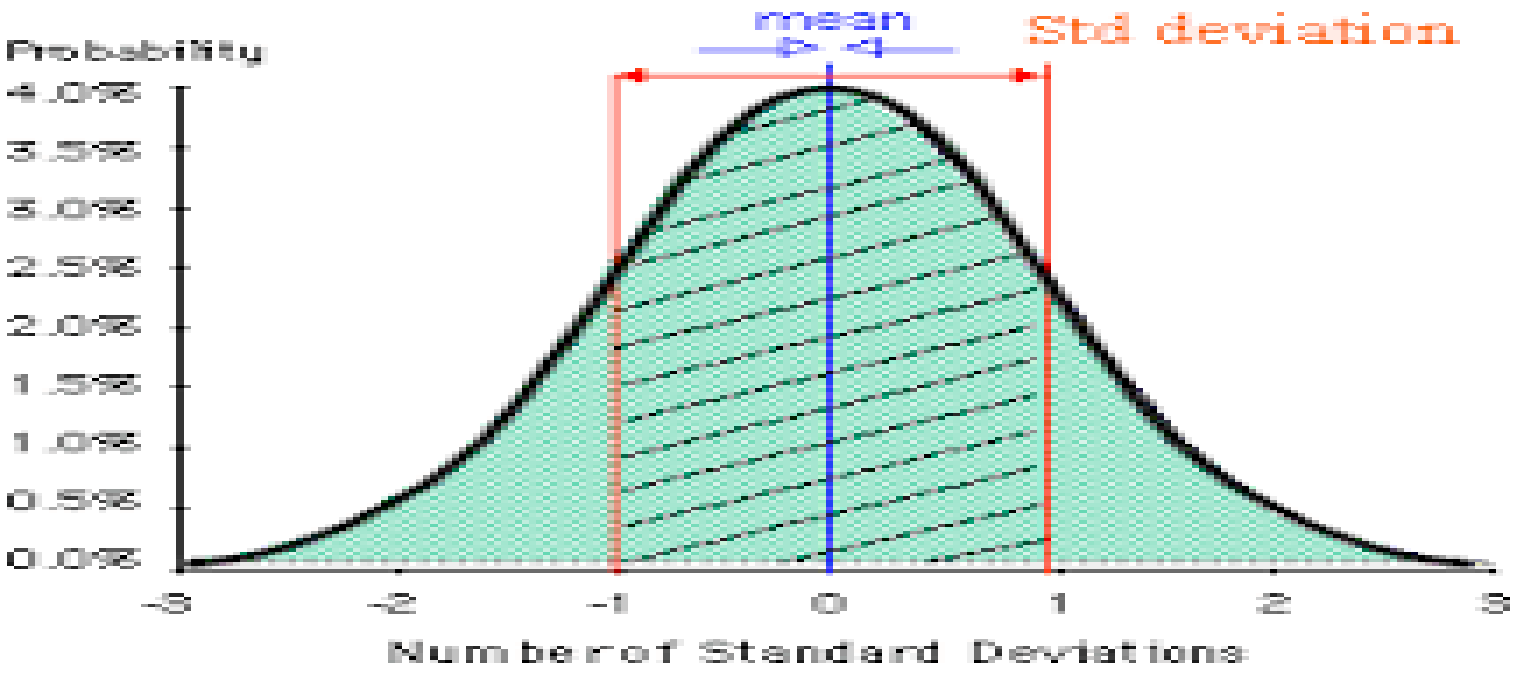
Parametric versus non-parametric tests



Continued

- 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.



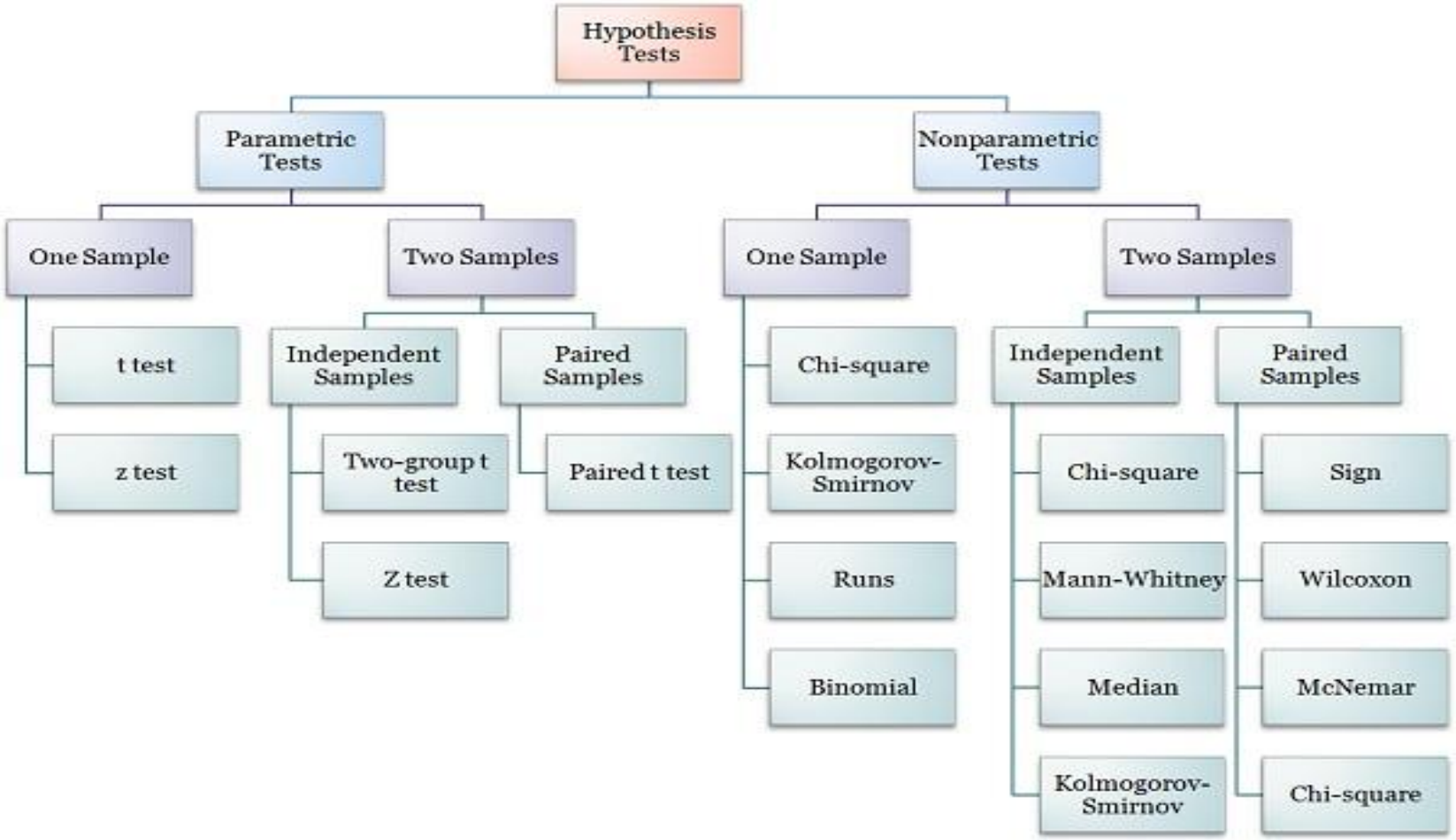
- The bell-shaped curve indicates that most values fall near the central value , with fewer from the centre and the rest fall symetrically.
- How we know that a population is normally distributed?

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

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.

Statistical tests



Difference between paired and independent samples

Both paired (dependent samples)and independent (unpaired samples) check if the difference between two means is significant.

- The paired-samples t-test compares scores on two different variables of the same group
- Independent- samples test compares scores on the same variable for two different groups.

Examples of statistical tests

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.
- T-test: T-test allows the comparison of the mean of two groups
- ANOVA test: Analysis of variance : allows the comparison of three or more groups.

Correlation: allows the examination of relationships between variables.

- Linear regression: focuses on prediction. A single independent variable is used to predict the value of a dependent variable.

  Does age predict income?

- Mann Whitney test: is frequently used as an alternative to the t-test for independent samples. It can be used with data measured on an ordinal scale.

Types of analyses

Univariate analysis: the analysis of one variable: mean, median, mode and standard deviation.

Eg. 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).

V

Task: Read the handouts you were given and sort out the main characteristics of parametric and non-parametric tests.

For what purpose is chi-square test used?

Give examples of linear regression and multi regression.