



ПРЕДГОВОР

КАКО ДА СЕ ОДБЕРЕ НАЈСООДВЕТНИОТ
СТАТИСТИЧКИ ТЕСТ ВО НАУЧНИТЕ
ТРУДОВИ

Владимир ТРАЈКОВСКИ

Дефектолошка теорија и практика
Институт за дефектологија
Филозофски факултет
Скопје, Република Македонија

Примено: 08.07.2016
Прифатено: 20.07.2016

Резиме

Статистиката е дел од математиката која се занимава со собирање, анализирање, интерпретирање и презентирање маса (голем број примероци) нумерички податоци со цел да се извлечат релевантни заклучоци од истата. Статистиката е форма на математичка анализа која користи квантифицирани модели, репрезентации и синописи за даден број експериментални податоци или истражувања кои се спроведуваат со жива материја. Студентите и младите истражувачи во биомедицинските науки како и во специјалната едукација и рехабилитација често го искажуваат своето мислење дека одбрале да се запишат на тие студии поради тоа што не поседуваат големо знаење или интерес за математика. Тоа е тажна изјава, но има и вистина во неа. Целта на овој едиторијал е да им послужи и да им помогне на младите истражувачи да ја одберат најсоодветната техника за статистичка обработка на податоците која ќе соодветствува на целите и условите на одредена анализа. Најважните статистички тестови ќе бидат прикажани во овој труд.

Адреса за кореспонденција:
Владимир ТРАЈКОВСКИ
Дефектолошка теорија и практика
Универзитет „Св. Кирил и Методиј“
Филозофски факултет
Институт за дефектологија
Бул. Гоце Делчев 9А, 1000 Скопје
Република Македонија
Е-пошта: vladotra@fzf.ukim.edu.mk

EDITORIAL

HOW TO SELECT APPROPRIATE
STATISTICAL TEST IN SCIENTIFIC
ARTICLES

Vladimir TRAJKOVSKI

Journal of Special Education and Rehabilitation
Institute of Special Education and Rehabilitation
Faculty of Philosophy
Skopje, Republic of Macedonia

Received: 08.07.2016
Accepted: 20.07.2016
Editorial

Abstract

Statistics is mathematical science dealing with the collection, analysis, interpretation, and presentation of masses of numerical data in order to draw relevant conclusions. Statistics is a form of mathematical analysis that uses quantified models, representations and synopses for a given set of experimental data or real-life studies. The students and young researchers in biomedical sciences and in special education and rehabilitation often declare that they have chosen to enroll that study program because they have lack of knowledge or interest in mathematics. This is a sad statement, but there is much truth in it. The aim of this editorial is to help young researchers to select statistics or statistical techniques and statistical software appropriate for the purposes and conditions of a particular analysis. The most important statistical tests are reviewed in the article.

Corresponding address:
Vladimir TRAJKOVSKI
Journal of Special Education and Rehabilitation
“Ss Cyril and Methodius” University
Faculty of Philosophy
Institute of Special Education and Rehabilitation
Bull. Goce Delchev 9A 1000 Skopje
Republic of Macedonia
E-mail: vladotra@fzf.ukim.edu.mk

Да знаеш како да го одбереш правилниот статистички тест е многу важна одлука во делот на обработката на добиените податоци и во пишувањето на научниот труд. Младите истражувачи и автори би требало да знаат како да ги одберат и како да ги користат овие статистички методи. Компетентниот истражувач мора да поседува одреден степен на знаење во однос на статистичките процедури. Тука може да се подразбира курс за вовед во статистиката и, се разбира, користењето на добар учебник по статистика. За оваа цел, постои потреба предметот Статистика да биде задолжителен предмет во наставната програма на Институтот за дефектологија при Филозофскиот факултет во Скопје. Младите истражувачи имаат потреба од дополнителни курсеви за да се здобијат со поголемо знаење во областа на статистиката. Тие мора да се обучат за да користат статистички компјутерски програми на соодветен начин.

Клучни зборови: *Статистичка селекција на тест, статистика, научен труд, статистички програми*

Вовед

Статистиката е дел од математиката која се занимава со собирање, анализирање, интерпретирање и презентирање маса (голем број примероци) нумерички податоци со цел да се извлечат релевантни заклучоци од истата. Статистиката е форма на математичка анализа која користи квантифицирани модели, репрезентации и синописи за даден број експериментални податоци или истражувања кои се спроведуваат со жива материја. Статистиката се користи во неколку различни дисциплини (како научни така и кај оние кои не се занимаваат со наука) за добиените податоци да се сведат на заклучоци (1). Студентите и младите истражувачи во биомедицинските науки како и во специјалната едукација и рехабилитација, често објавуваат и го искажуваат сопственото мислење дека одбрале да се запишат на тие студии поради тоа што не поседуваат големо знаење или интерес за математика. Тоа е тажна изјава, но има и вистина во неа. Тие често не знаат како да ја извршат статистичката обработка на добиените податоци од истражувањето за нивните додипломски, последипломски, па и докторски студии, па затоа најчесто бараат помош од статистичари. За оваа цел, тие мора да платат одредена сума пари. Најчесто во

Knowing how to choose right statistical test is an important asset and decision in the research data processing and in the writing of scientific papers. Young researchers and authors should know how to choose and how to use statistical methods. The competent researcher will need knowledge in statistical procedures. That might include an introductory statistics course, and it most certainly includes using a good statistics textbook. For this purpose, there is need to return of Statistics mandatory subject in the curriculum of the Institute of Special Education and Rehabilitation at Faculty of Philosophy in Skopje. Young researchers have a need of additional courses in statistics. They need to train themselves to use statistical software on appropriate way.

Keywords: *statistical test selection, statistics, scientific article, statistical software*

Introduction

Statistics is mathematical science dealing with the collection, analysis, interpretation, and presentation of masses of numerical data in order to draw relevant conclusions. Statistics is a form of mathematical analysis that uses quantified models, representations and synopses for a given set of experimental data or real-life studies. Statistics is used in several different disciplines (both scientific and non-scientific) to make decisions and draw conclusions based on data (1).

The students and young researchers in biomedical sciences and in special education and rehabilitation often declare that they have chosen to enroll that study program because they have lack of knowledge or interest in mathematics. This is a sad statement, but there is much truth in it. They often do not know to make their statistical processing of data for its undergraduate, master's and doctoral theses, and seek help from a statistician. For this purpose, they have to pay certain amount of money. There



нивните тези има погрешно одбрани статистички методи кои водат кон погрешни заклучоци. Селекцијата на правилниот статистички метод или техника може да претставува голем проблем за младите истражувачи. Во истражувањето, значајните заклучоци можат да бидат изведени од собраните податоци од валиден научен дизајн користејќи го соодветниот статистички метод или техника.

Во однос на селекцијата на статистичкиот метод кој би се користел, најважното прашање е „Која е главната хипотеза на истражувањето?“ Во некои случаи нема главна хипотеза; истражувачот само сака да „види што има таму“. На пример, во студија за преваленција нема хипотеза која би се тестираше, и големината на студијата е одредена од тоа колку прецизно истражувачот сака да ја одреди преваленцијата. Доколку нема поставено главна хипотеза, тогаш нема статистички тест. Важно е уште пред да се започне со истражувањето кои хипотези би се потврдиле како точни (тоа се однесува на некои за некои врски за кои се претпоставува дека би излегол таков резултат) и кои би биле прелиминарни (индицирани од добиените податоци од истражувањето) (2).

Во истражувачки студии користењето на погрешните статистички тестови може да се види во голем број случаи, како користењето на тестови за парен број на податоци кај податоци добиени од непарен број или користењето на параметриски статистички тест за обработка на податоците кој не ја следи нормалната дистрибуција или некомпатибилен статистички тест за добиените податоци од истражувањето (3).

Достапноста на различни типови статистички програми, го прави изведувањето на статистиката и статистичките тестови многу лесно, но изборот на соодветниот статистички тест или метод сè уште претставува проблем. Најдобар пристап е чекор по чекор систематски да се дојде до одлука на кој начин да се анализираат добиените податоци. Се препорачува да се следат овие чекори (4).

- да се одреди и специфицира во форма на прашање што сакаме да постигнеме со истражувањето;
- да се постави прашањето во форма на статистичка нулта хипотеза и да се издвојат алтернативни хипотези од главната или нулата хипотеза;
- да се одредат кои варијабли се релевантни за прашањето;
- да се одреди од кој тип е секоја варијаб-

are in the theses very often wrong selected statistical methods which then lead to erroneous conclusions. Selecting the right statistical test may represent a huge problem for younger researchers. In research, meaningful conclusions can only be drawn based on data collected from a valid scientific design using appropriate statistical tests.

Regarding to selecting a statistical test, the most important question is "what is the main study hypothesis?" In some cases there is no hypothesis; the investigator just wants to "see what is there". For example, in a prevalence study there is no hypothesis to test, and the size of the study is determined by how accurately the investigator wants to determine the prevalence. If there is no hypothesis, then there is no statistical test. It is important to decide a priori which hypotheses are confirmatory (that is, are testing some presupposed relationship), and which are exploratory (are suggested by the data) (2).

In research studies wrong statistical tests can be seen in many conditions like use of paired test for unpaired data or use of parametric statistical tests for the data which does not follow the normal distribution or incompatibility of statistical tests with the type of data (3).

The availability of different types of statistical software makes performing of the statistical tests to become easy, but selection of appropriate statistical test is still a problem. Systematic step-by-step approach is the best way to decide how to analyze data. It is recommended that you follow these steps (4):

- Specify the question you are asking.
- Put the question in the form of a statistical null hypothesis and alternate hypothesis.
- Determine which variables are relevant to the question.
- Determine what kind of variable each

- ла посебно;
- да се дизајнира студија која ги контролира или ги распределува случајните варијабли;
- да се одбере најдобриот статистички тест или метод базиран врз бројот и видот на варијаблите за да се утврди дали очекуваните резултати одговараат на претпоставките кои сме ги поставиле во параметрите и да се тестираат хипотезите;
- доколку е можно, да се направи претходна анализа за да се одреди големина на примерокот кој ќе се испитува во истражувањето;
- да се направи истражувањето;
- да се прегледаат добиените податоци и да се утврди дали соодветствуваат со претпоставките од статистичкиот тест кој е одбран. Доколку не се соодветни тогаш се бара посоодветен тест;
- да се спроведе статистичкиот тест кој ќе се покаже како најсоодветен и да се интерпретираат резултатите и
- да се презентираат добиените резултати ефективно, најчесто со графикони или табели.

Marusteri и Bacarea укажуваат и на други услови кои би требало да се земат предвид кога вршиме анализа на добиените податоци од одредено истражување:

- основно ниво на познавање на базичната статистичка терминологија и концепти;
- да се поседува знаење за неколку аспекти поврзани со податоците кои сме ги добиле за време на истражувањето / експериментот (пр. каков тип на податоци сме добиле – номинални, ординални, интервални или размерни скали (скали на односи) како се организирани добиените податоци, колку истражувачки групи се опфатени (обично експериментална и контролна група), дали групите се во пар или непар, дали примерокот или примероците припаѓаат на нормална дистрибуирана / Гаусова популација);
- добро разбирање на целта за нашата статистичка анализа;
- добра анализа на целиот статистички протокол во еден добар структуриран, разгранет, алгоритамски начин, со цел да се избегнат можни грешки (5).

Целта на овој едиторијал е да им помогне на младите истражувачи да можат да ги одберат статистичките техники или статистички

one is.

- Design a study that controls or randomizes the confounding variables.
- Based on the number of variables, the kinds of variables, the expected fit to the parametric assumptions, and the hypothesis to be tested, choose the best statistical test to use.
- If possible, do a power analysis to determine a good sample size for the study.
- Do the study.
- Examine the data to see if it meets the assumptions of the statistical test you chose. If it doesn't, choose a more appropriate test.
- Apply the statistical test you chose, and interpret the results.
- Show your results effectively, usually with a table or a figure.

Marusteri and Bacarea mentioned other things we should have in our mind when we are analyzing the data from some study:

- Decent understanding of some basic statistical terms and concepts;
- Some knowledge about few aspects related to the data we collected during the research/experiment (e.g. what types of data we have - nominal, ordinal, interval or ratio, how the data are organized, how many study groups (usually experimental and control at least) we have, are the groups paired or unpaired, and are the sample(s) extracted from a normally distributed/Gaussian population);
- Good understanding of the goal of our statistical analysis;
- We have to parse the entire statistical protocol in a well structured - decision tree /algorithmic manner, in order to avoid some mistakes (5).

The aim of this editorial is to help young researchers to select statistics or statistical



компјутерски програми кои би биле соодветни во исполнувањето на целите и условите на одредена анализа. Неколку од овие чекори ќе бидат подетално објаснети во долунаведениот текст.

Видови скали

Пред да можеме да ја спроведеме статистичката анализа, мораме да извршиме мерење на зависната варијабла. Начинот на кој се врши мерењето ќе зависи целосно од типот на варијаблата која е вклучена при самата анализа. Различни типови се мерат на различен начин. Иако процедурите за мерење се разликуваат една од друга на многу начини, можат да бидат класифицирани користејќи неколку фундаментални категории. Во секоја категорија сите процедури меѓусебно споделуваат дел од важните особини. Постојат четири типови скали.

Номинални скали

Кога при мерењата се користи номиналната скала, тогаш само се именуваат или категоризираат дадени одговори. Пол, брачен статус, омилена боја, како и религиска определба се примери на варијабли измерени со номинална скала. Есенцијалната цел на номиналните скали се состои во тоа што тие не вршат подредување на дадени одговори од субјектите кои се испитуваат. На пример, кога ги класифицираме луѓето според нивната омилена боја, нема смисла кога црвената боја е ставена пред жолтата. Одговорите само се категоризираат. Со номиналните скали се отелотворуваат најниските видови мерења во статистиката (6).

Ординални скали

Истражувач кој сака да изврши мерење на варијаблата колку се задоволни родителите од третманот на нивното дете во текот на наставата, може да им постави специфично прашање за тоа како се чувствуваат: „многу незадоволно“, „малку незадоволно“, „малку задоволно“, „многу задоволно“. Во овој случај варијаблите се подредени, рангирајќи од најмалку до најмногу задоволни. Ова е основната разлика помеѓу ординалната и номиналната скала. За разлика од номиналните скали, ординарните скали дозволуваат да се направи споредба до кој степен два субјекти кои се испитуваат ја поседуваат зависната варијабла. На пример, нашето задоволство

techniques and statistical software appropriate for the purposes and conditions of a particular analysis. In the following text it will be explained some of these steps.

Types of scales

Before we can conduct a statistical analysis, we need to measure our dependent variable. Exactly how the measurement is carried out depends on the type of variable involved in the analysis. Different types are measured differently. Although procedures for measurement differ in many ways, they can be classified using a few fundamental categories. In a given category, all of the procedures share some properties that are important to know about. There are four types of scales.

Nominal scales

When measuring using a nominal scale, one simply names or categorizes responses. Gender, marital status, handedness, favorite color, and religion are examples of variables measured on a nominal scale. The essential point about nominal scales is that they do not imply any ordering among the responses. For example, when classifying people according to their favorite color, there is no sense in which red is placed “ahead of” yellow. Responses are merely categorized. Nominal scales embody the lowest level of measurement (6).

Ordinal scales

A researcher wishing to measure satisfaction of parents with treatment of their child in regular classroom might ask them to specify their feelings as either “very dissatisfied,” “somewhat dissatisfied,” “somewhat satisfied,” or “very satisfied.” The items in this scale are ordered, ranging from least to most satisfied. This is what distinguishes ordinal from nominal scales. Unlike nominal scales, ordinal scales allow comparisons of the degree to which two subjects possess the dependent variable. For example, our satisfaction ordering makes it meaningful to assert that one person is more satisfied than another with their microwave ovens. Such an assertion reflects the first person's use of a

при купување на микробранова печка може да е поголемо од она на други купувачи. Од друга страна, со ординалните скали не можеме да ги добиеме важните информации кои се присутни кај другите видови скали. На пример, разликата помеѓу две нивоа на една ординална скала не може да се претпостави дека ќе бидат исти како и разликата помеѓу други две нивоа. Кај скалите со кои се мери задоволството, на пример, разликата помеѓу одговорите „многу незадоволен“ и „малку незадоволен“ сигурно не е еквивалентна на разликата помеѓу „малку незадоволен“ и „малку задоволен“. Ништо што е во нашата процедура за мерење не може да ни детерминира дали двете разлики ја рефлектираат истата разлика во психолошко задоволство (6).

Интервални скали

Интервалните скали се нумерички скали кои вклучуваат: возраст (години), тежина (кг) или должина на коска (цм), во која интервалите ја имаат истата интерпретација низ целата скала. Интервалните податоци се подредени по значаен редослед и го поседуваат квалитетот кој е еднаков со интервалите направени помеѓу мерењата и ја претставуваат истата промена во квантитетот на тоа што го мериме. Но кај овие типови податоци не постои природна нула. На пример, во Целзиусовата скала за температура. Во Целзиусовата скала, не постои природна нула, така што не можеме да кажеме дека 50°C е дупло од 25°C. Кај интервалните скали нулата точката може да биде поставена арбитражно. IQ-тестот исто така претставува податок за интервална скала кај која не постои природна (апсолутна) нула (7).

Размерни скали

Размерната скала за мерење на добиени податоци содржи најголем број информации. Тоа е интервална скала со дополнителна особина каде што положбата на нулата посочува на отсуство од квантитетот што се мери. За размерната скала може да се каже дека е составена од сите три претходни скали. Како и номиналната скала, ни дава име или категорија за одреден објект (броевите служат како обележја). Како кај ординалната скала, објектите се подредени (како подредување на броеви). Кај размерната скала истата разлика на две места го има истото значење, како и кај интервалната скала. Но исто така, истиот размер на две места на скалата носи исто

verbal label that comes later in the list than the label chosen by the second person. On the other hand, ordinal scales fail to capture important information that will be present in the other scales we examine. In particular, the difference between two levels of an ordinal scale cannot be assumed to be the same as the difference between two other levels. In satisfaction scale, for example, the difference between the responses “very dissatisfied” and “somewhat dissatisfied” is probably not equivalent to the difference between “somewhat dissatisfied” and “somewhat satisfied.” Nothing in our measurement procedure allows us to determine whether the two differences reflect the same difference in psychological satisfaction (6).

Interval scales

Interval scales are numerical scales including: age (years), weight (kg) or length of bone (cm), in which intervals have the same interpretation throughout. Interval data has a meaningful order and also has the quality that equal intervals between measurements represent equal changes in the quantity of whatever is being measured. But these types of data have no natural zero. Example is Celsius scale of temperature. In the Celsius scale, there is no natural zero, so we cannot say that 50°C is double than 25°C. In interval scale, zero point can be chosen arbitral. IQ test is also interval data as it has no natural zero (7).

Ratio scales

The ratio scale of measurement is the most informative scale. It is an interval scale with the additional property that its zero position indicates the absence of the quantity being measured. You can think of a ratio scale as the three earlier scales rolled up in one. Like a nominal scale, it provides a name or category for each object (the numbers serve as labels). Like an ordinal scale, the objects are ordered (in terms of the ordering of the numbers). Like an interval scale, the same difference at two places on the scale has the same meaning. And in addition, the same ratio at two places on the scale also carries the same meaning. Example of a ratio scale



значење. Пример за размерна скала е количината на пари која ја имате во овој момент (500 денари, 1000 денари итн.). Парите се мерат со размерна скала, бидејќи, покрај тоа што ги имаат особините на интервална скала, постои вистинска нулта точка: доколку имате нула денари, ова посочува на отсуство на пари. Бидејќи парите имаат вистинска нулта точка, има смисла да кажеме дека некој со 1000 денари има двапати повеќе отколку некој со 500 денари (или дека Марк Цукерберг има милион пати повеќе пари отколку што имате вие) (6).

Нормална дистрибуција или не

Ова е уште еден проблем при селекцијата на правилниот статистички тест. Доколку знаете каков е видот на податоците (номинални, ординални, интервални или размерни) и дистрибуцијата на податоците (нормална дистрибуција или ненормална дистрибуција), селекцијата на статистичкиот тест е многу лесна. Нема потреба да се проверува дистрибуцијата кај ординалните и номиналните скали на податоци добиени од истражувањето. Дистрибуцијата обично се проверува само кај интервални или размерни податоци. Ако вашите податоци ја следат нормалната дистрибуција, би бил користен параметриски (стандардизиран) статистички тест, додека доколку не се следи нормалната дистрибуција, тогаш би се користел непараметриски тест. Постојат различни методи за да се провери нормалната дистрибуција, некои од нив преку различни видови на хистограми, мерење на искривеност на кривата и куртозис, како на пример, статистичкиот тест на нормалност (Колмогоров-Смирнов тест, Шапиро-Вилк-тестот итн.). Формалните статистички тестови како Колмогоров-Смирнов-тестот и Шапиро-Вилк-тестот најчесто се користат за да се провери дистрибуцијата на добиените податоци. Сите овие тестови се базирани на нултата хипотеза дека податоците се земени од популација која ја следи нормалната дистрибуција. P вредноста се одредува за да се увиди алфа грешката. Доколку P вредноста е помала од 0,05, тогаш добиените податоци не ја следат нормалната дистрибуција и во овој случај би требало да се користи нестандардизиран тест. Доколку примерокот кој се испитува е помал, веројатноста за ненормална дистрибуција се зголемува (7).

is the amount of money you have in your pocket right now (500 denars, 1000 denars, etc.). Money is measured on a ratio scale because, in addition to having the properties of an interval scale, it has a true zero point: if you have zero money, this implies the absence of money. Since money has a true zero point, it makes sense to say that someone with 1000 denars has twice as much money as someone with 500 denars (or that Mark Zuckerberg has a million times more money than you do) (6).

Normal distribution or not

This is another issue for selection of right statistical test. If you know the type of data (nominal, ordinal, interval, and ratio) and distribution of data (normal distribution or not normal distribution), selection of statistical test will be very easy. There is no need to check distribution in the case of ordinal and nominal data. Distribution should only be checked in the case of ratio and interval data. If your data are following the normal distribution, parametric statistical test should be used and nonparametric tests should only be used when normal distribution is not followed. There are various methods for checking the normal distribution, some of them are plotting histogram, plotting box and whisker plot, plotting Q-Q plot, measuring skewness and kurtosis, using formal statistical test for normality (Kolmogorov-Smirnov test, Shapiro-Wilk test, etc). Formal statistical tests like Kolmogorov-Smirnov and Shapiro-Wilk are used frequently to check the distribution of data. All these tests are based on null hypothesis that data are taken from the population which follows the normal distribution. P value is determined to see the alpha error. If P value is less than 0.05, data is not following the normal distribution and nonparametric test should be used in that kind of data. If the sample size is less, chances of non-normal distribution are increased (7).

Параметриски и непараметриски процедури

Стандардизираниите статистички процедури се основани на претпоставки за формата на дистрибуцијата (се претпоставува нормална дистрибуција) во основната популација и за формата на параметрите кои се земени (начини и стандардни девијации) од претпоставената дистрибуција.

Нестандардизираниите статистички процедури се поткрепуваат на неколку претпоставки во однос на формата на параметрите на популацијата дистрибуција од која самиот примерок бил извлечен (8). Нестандардизираниите методи обично се послаби и помалку флексибилни за разлика од стандардизираниите. Стандардизираниите методи се користат тогаш кога претпоставките можеме да ги оправдаме. Некогаш можеме да направиме трансформација на добиените податоци за да извршиме оправдување на претпоставките, како трансформација на дневник (9). Табела 1 ни ја покажува употребата на стандардизирани и нестандардизирани статистички методи.

Табела 1. Параметриски наспроти непараметриски методи/

Table 1. Parametric vs non-parametric methods

	Parametric	Non-parametric
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Any
Typical data	Ratio or interval	Ordinal or nominal
Dataset relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers
Tests		
Choosing	Choosing parametric test	Choosing non-parametric test
Correlation test	Pearson	Spearman
Independent measures, 2 groups	Independent-measures t-test	Mann-Whitney test
Independent measures, >2 groups	One-way independent-measures ANOVA	Kruskal-Wallis test
Repeated measures, 2 conditions	Matched-pair t-test	Wilcoxon test
Repeated measures, >2 conditions	One-way repeated measures ANOVA	Friedman's test

Аритметичка средина (или просек) претпоставува мерење на локација од една група вредности добиени преку податоците; сумата на сите добиени податоци поделена со бројот на елементи во дистрибуцијата. Придружен елемент на мерење кој ја следи аритметичката средина обично е **стандардната девијација**. За разлика од **медијаната** и **модата**, не е соодветно да се користи овој тип на мерење за да се карактеризира или опише искривена (ненормална) дистрибуција.

Parametric and non-parametric procedures

Parametric statistical procedures rely on assumptions about the shape of the distribution (assume a normal distribution) in the underlying population and about the form or parameters (means and standard deviations) of the assumed distribution. Nonparametric statistical procedures rely on no or few assumptions about the shape or parameters of the population distribution from which the sample was drawn (8). Non-parametric methods are typically less powerful and less flexible than their parametric counterparts. Parametric methods are preferred if the assumptions can be justified. Sometimes a transformation can be applied to the data to satisfy the assumptions, such as log transformation (9). Table 1 shows the use of parametric and non-parametric statistical methods.

Arithmetic Mean (or average): a measure of location for a batch of data values; the sum of all data values divided by the number of elements in the distribution. Its accompanying measure of spread is usually the **standard deviation**. Unlike the **median** and the **mode**, it is not appropriate to use the mean to characterize a skewed distribution.



Медијаната е уште едно мерење на локација како и **аритметичката средина**. Вредноста која ја дели дистрибуцијата на фреквенцијата на средина кога сите податоци се подредени по редослед. Кај овој тип на мерења се гледа дека не постои сензитивност кога се мерат мали броеви во екстремно големи резултати во една дистрибуција. Затоа, таа е преферирана мерка за мерење на централната тенденција кај искривена дистрибуција (каде аритметичката средина е пристрасна) и обично оди заедно со **интеркварталниот ранг (dQ)** како придружна мерка за раст.

Интерквартален ранг (dQ) е мерка на раст и е спротивна на стандардната девијација кај искривена или ненормална дистрибуција на податоците. dQ е растојанието помеѓу горните и долните квартали ($Q_U - Q_L$).

Варијанса е нумеричка вредност која се користи за да се утврди и укаже на тоа колку индивидуите на една група се разликуваат или варираат во однос на некои особини кои ние ги мериме. Ако индивидуалната опсервација се разликува многу од средната вредност добиена за групата, тогаш разликата е голема; и обратно. Многу е важно да се прави разлика помеѓу разликата во една популација и разликата кај еден примерок. Тие се забележани на различен начин, и податоците за секој од нив се обработува посебно. Варијабилноста кај популацијата се обележува со σ^2 , а варијабилноста на еден примерок се обележува со s^2 .

Стандардна девијација (SD) претставува мерка за мерење на одреден сет податоци и нивниот раст. За разлика од варијансата која е изразена во квадратни единици, SD се изразува во истите единици како и оригиналните податоци добиени од истражувањето. Се пресметува според отстапувањата помеѓу секој податок поединечно како и од **аритметичката средина** на примерокот. Тоа е квадратниот корен од варијансата. За различни цели, n (целосниот број на вредности) или $n-1$ може да се користи при пресметувањето на варијабилноста/SD. Доколку ја имате пресметано SD делејќи ја со n но сакате да ја претворите во SD и да одговара на именителот на $n-1$, тогаш се множи резултатот со квадратниот корен од $n/(n-1)$. Доколку дистрибуцијата на SD е поголема од **аритметичката средина**, тогаш аритметичката средина не е адекватна како репрезентативна единица за мерење на централната тенденција. За податоци кои имаат нормална дистри-

Median is another measure of location just like the **mean**. The value that divides the frequency distribution in half when all data values are listed in order. It is insensitive to small numbers of extreme scores in a distribution. Therefore, it is the preferred measure of central tendency for a skewed distribution (in which the mean would be biased) and is usually paired with the **interquartile range (dQ)** as the accompanying measure of spread.

Interquartile range (dQ) is a measure of spread and is the counterpart of the standard deviation for skewed distributions. dQ is the distance between the upper and lower quartiles ($Q_U - Q_L$).

Variance is a numerical value used to indicate how widely individuals in a group vary. If individual observations vary greatly from the group mean, the variance is big; and vice versa. It is important to distinguish between the variance of a population and the variance of a sample. They have different notation, and they are computed differently. The variance of a population is denoted by σ^2 ; and the variance of a sample, by s^2 .

Standard deviation (SD): is a measure of spread (scatter) of a set of data. Unlike variance, which is expressed in squared units of measurement, the SD is expressed in the same units as the measurements of the original data. It is calculated from the deviations between each data value and the sample **mean**. It is the square root of the variance. For different purposes, n (the total number of values) or $n-1$ may be used in computing the variance/SD. If you have a SD calculated by dividing by n and want to convert it to a SD corresponding to a denominator of $n-1$, multiply the result by the square root of $n/(n-1)$. If a distribution's SD is greater than its **mean**, the mean is inadequate as a representative measure of central tendency. For normally distributed data

буција, приближно 68% од дистрибуцијата припаѓа ± 1 SD од аритметичката средина, 95% од дистрибуцијата припаѓа на ± 2 SD од аритметичката средина, и 99.7% од дистрибуцијата припаѓа на ± 3 SD од аритметичката средина (емпириско правило).

Стандардна грешка (SE) или како што и се нарекува стандардна грешка на аритметичката средина (SEM) е мерка која врши мерење за да се види до која мерка податоците добиени за примерокот кој се истражува се разликува од вистинската но непозната популациона аритметичка средина. Таа е **стандардната девијација (SD)** од случајно одбрани примероци при дистрибуцијата кај аритметичката средина (т.е. аритметичка средина на повеќе примероци од истата популација) како таква, таа ја мери прецизноста на статистиката која се употребува како проценка на одредена популација. Проценетата вредност SE/SEM зависи од големината на примерокот. Таа е поврзана со квадратниот корен од големината на примерокот: (проценето) $SE = SD / (N)^{1/2}$. Вистинската вредност на SE може да биде пресметана само доколку SD на популацијата ни е позната. Кога се користи SD на примерокот (скоро секогаш), тоа претставува проценка и би требало да се вика проценка на стандардна грешка (ESE). Кога големината на примерокот е релативно голема ($N \geq 100$), примерокот на SD ни овозможува добра поткрепена проценка на SE (10). Статистичарите користат **интервал на доверба** со цел да го изразат степенот на несигурност кој е поврзан со статистиката на примерокот. Интервалот на доверба е проценка на интервал комбинирана со изјава за веројатност. На пример, да претпоставиме дека статистичарот спроведува анкета и има пресметано проценка на интервалот базирана на податоците добиени од анкетата. Статистичарот може да го користи нивото на доверба за да ја опише несигурноста која е асоцирана со процената на интервалот. Тој / таа може да ја опишат процената на интервалот како „95% интервал на доверба“. Ова би значело дека доколку го користиме истиот метод за да селектираме други примероци и да пресметаме проценка на интервал за секој од примероците, тогаш би можеле да очекуваме вистинскиот параметар на популацијата да се наоѓа во рамките на 95% од процената на интервалот во кое било дадено време. Интервалите на доверба се преферирани за да посочат кон процени и процени на интервали,

values, approximately 68% of the distribution falls within ± 1 SD of the mean, 95% of the distribution falls within ± 2 SDs of the mean, and 99.7% of the distribution falls within ± 3 SDs of the mean (empirical rule).

Standard error (SE): or as commonly called the standard error of the mean (SEM) is a measure of the extent to which the sample mean deviates from the true but unknown population mean. It is the **standard deviation (SD)** of the random sampling distribution of means (i.e., means of multiple samples from the same population). As such, it measures the precision of the statistic as an estimate of a population. The (estimated) SE/SEM is dependent on the sample size. It is inversely related to the square root of the sample size: (estimated) $SE = SD / (N)^{1/2}$. The true value of the SE can only be calculated if the SD of the population is known. When the sample SD is used (as almost always), it is an estimate and should be called estimated standard error (ESE). When the sample size is relatively large ($N \geq 100$), the sample SD provides a reliable estimate of the SE (10).

Statisticians use a **confidence interval** to express the degree of uncertainty associated with a sample statistic. A confidence interval is an interval estimate combined with a probability statement. For example, suppose a statistician conducted a survey and computed an interval estimate, based on survey data. The statistician might use a confidence level to describe uncertainty associated with the interval estimate. He/she might describe the interval estimate as a "95% confidence interval". This means that if we used the same sampling method to select different samples and computed an interval estimate for each sample, we would expect the true population parameter to fall within the interval estimates 95% of the time. Confidence intervals are preferred to point estimates and to interval estimates, because only confidence intervals



бидејќи само интервалите на доверба ја покажуваат (а) прецизноста на процената и (б) несигурноста на процената (11).

Сите овие тестови за мерење на варијаблите се прикажани во табела 2.

indicate (a) the precision of the estimate and (b) the uncertainty of the estimate (11).

All those statistical tests for measurement variables are shown in table 2.

Табела 2. Селектирање на статистички тест за мерливите варијабли/

Table 2. Selecting statistical test for measurement variables (4)

Test	Measure-ment variables	Purpose	Notes	Example
Arithmetic mean	1	description of central tendency of data	/	/
Median	1	description of central tendency of data	more useful than mean for very skewed data	median height of trees in forest, if most trees are short seedlings and the mean would be skewed by a few very tall trees
Range	1	description of dispersion of data	used more in everyday life than in scientific statistics	/
Variance	1	description of dispersion of data	forms the basis of many statistical tests; in squared units, so not very understandable	/
Standard deviation	1	description of dispersion of data	in same units as original data, so more understandable than variance	/
Standard error of the mean	1	description of accuracy of an estimate of a mean	/	/
Confidence interval	1	description of accuracy of an estimate of a mean	/	/

Хи-квадрат (χ^2) тест е статистички тест кој најчесто се користи за да се изврши споредба на податоците добиени од набљудувањето со податоците кои би очекувале да ги добиеме според специфична хипотеза. На пример, ако, според Менделовите закони, вие очекувате 10 од 20 деца да бидат машки а вистинскиот набљудуван број е 8 машки деца, тогаш можеби би сакале да знаете за разликата помеѓу набљудуваното и очекуваното. Дали девијациите (разликите помеѓу набљудуваното и очекуваното) се резултат на одредена шанса или се резултат на други фактори. Колку девијации би можеле да се случат пред истражувачот да заклучи дека постои нешто друго освен обична шанса што го предизвикува набљудуваното да се разликува од очекуваното. χ^2 -тестот секогаш ја става на тест **нултата хипотеза**, која гласи дека не постои статистичка значајна разлика помеѓу очекуваното и набљудуваното. χ^2 е сумата од

Chi-squared test is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. For example, if, according to Mendel's laws, you expected 10 of 20 offspring from a cross to be male and the actual observed number was 8 males, then you might want to know about the "goodness to fit" between the observed and expected. Were the deviations (differences between observed and expected) the result of chance, or were they due to other factors. How much deviation can occur before you, the investigator, must conclude that something other than chance is at work, causing the observed to differ from the expected. The chi-square test is always testing what scientists call the **null hypothesis**, which states that there is no

квадратната разлика на набљудуваните (o) и очекуваните (e) податоци (или девијацијата, d), поделено со очекуваните податоци во сите можни категории.

G-статистика е апликација на **размерна веројатна статистика** за хипотезата на независноста кај $r \times c$ табела на контингенција. Исто така може да се користи за да се тестира и прикладноста на припадност. G-тестот би требало да се користи повеќе од χ^2 -тестот кога за која било клетка во табелата, $\frac{1}{2} O - E \frac{1}{2} > E$. Дистрибуцијата на χ^2 -тестот е обично послаба и посиромашна во однос на статистичкиот тест G^2 каде N/rc е помал од пет (преферабилно на χ^2 -тестот кај Hardy-Weinberg еквилибриум (HWE) тест сè додека овој услов е исполнет).

Фишеров тест на точност е тест за точната значајност за анализа на 2×2 табели за која било големина на примерок. Погрешно е мислењето дека овој тест е соодветен само за мала група на примерок. Ова потекнува од големата потреба за пресметки кај примероци со голем број испитаници, што повеќе не претставува проблем. Тоа е единствениот тест за 2×2 табела кога очекуваниот број во која било од клетките е помал од 5.

Кохран-Мантел-Ханзел тест е тест за нулата хипотеза каде нема сеопфатни врски во серии од 2×2 табели за стратификувани податоци кои доаѓаат или од кохорта или од контролна студија. Овој тест дозволува анализи на **затворени примероци** и дава **прилагоден размер** или релативен ризик. Може да се користат податоци од одредена категорија или категоризирани податоци со одредена трајност. Тестот е валиден само кога варијабилноста на набљудуваните податоци е ≥ 5 . Не е соодветен кога асоцијацијата драматично се менува низ стратумот (хетерогеноста обично се тестира со Брезлов-Деј тест). Но меѓутоа многу лесно може да се употребува кај сетови на податоци кои не се добиени преку ИТ-направи каде што асимптоматичната теорија не е поддржана за G^2 статистиката на тестот, M^2 има приближно χ^2 дистрибуција со $df = 1$ (10).

Овие се најкористените статистички тестови за номинални варијабли и се прикажани во табела 3.

significant difference between the expected and observed result. Chi-square is the sum of the squared difference between observed (o) and the expected (e) data (or the deviation, d), divided by the expected data in all possible categories.

G Statistics is an application of the **log-likelihood ratio statistics** for the hypothesis of independence in an $r \times c$ contingency table. It can also be used to test goodness-of-fit. The G-test should be preferred over Chi-squared test when for any cell in the table, $\frac{1}{2} O - E \frac{1}{2} > E$. The Chi-squared distribution is usually poor for the test statistics G^2 when N/rc is smaller than five (preferable to the Chi-squared test in Hardy-Weinberg Equilibrium (HWE) test as long as this condition is met).

Fisher's exact test is an exact significance test to analyze 2×2 tables for any sample size. It is a misconception that it is suitable only for small sample sizes. This arises from the demanding computational procedure for large samples, which is no longer an issue. It is the only test for a 2×2 table when an expected number in any cell is smaller than 5.

Cochran-Mantel-Haenszel test is test for a null hypothesis of no overall relationship in a series of 2×2 tables for stratified data derived either from a cohort or a case-control study. It allows analysis of **confounding** and gives an **adjusted odds ratio** or relative risk. It can be used on categorical or categorized continuous data. The test is only valid when the variance of observed data is ≥ 5 . It is inappropriate when the association changes dramatically across strata (heterogeneity is usually tested by Breslow-Day test). It is, however, applicable for sparse data sets for which asymptotic theory does not hold for G^2 . The test statistics, M^2 , has approximately a Chi-squared distribution with $df = 1$ (10).

These most used statistical tests for nominal variables are showed in table 3.



Табела 3. Селектирање статистички тест за номинални варијабли/
Table 3. Selecting statistical test for nominal variables (4)

Test	Nominal variables	Purpose	Notes	Example
Exact test for goodness-of-fit	1	test fit of observed frequencies to expected frequencies	use for small sample sizes (less than 1000)	count the number of red, pink and white flowers in a genetic cross, test fit to expected 1:2:1 ratio, total sample <1000
Chi-square test of goodness-of-fit	1	test fit of observed frequencies to expected frequencies	use for large sample sizes (greater than 1000)	count the number of red, pink and white flowers in a genetic cross, test fit to expected 1:2:1 ratio, total sample >1000
G-test of goodness-of-fit	1	test fit of observed frequencies to expected frequencies	used for large sample sizes (greater than 1000)	count the number of red, pink and white flowers in a genetic cross, test fit to expected 1:2:1 ratio, total sample >1000
Repeated G-tests of goodness-of-fit	2	test fit of observed frequencies to expected frequencies in multiple experiments	/	count the number of red, pink and white flowers in a genetic cross, test fit to expected 1:2:1 ratio, do multiple crosses
Fisher's exact test	2	test hypothesis that proportions are the same in different groups	use for small sample sizes (less than 1000)	count the number of live and dead patients after treatment with drug or placebo, test the hypothesis that the proportion of live and dead is the same in the two treatments, total sample <1000
Chi-square test of independence	2	test hypothesis that proportions are the same in different groups	use for large sample sizes (greater than 1000)	count the number of live and dead patients after treatment with drug or placebo, test the hypothesis that the proportion of live and dead is the same in the two treatments, total sample >1000
G-test of independence	2	test hypothesis that proportions are the same in different groups	large sample sizes (greater than 1000)	count the number of live and dead patients after treatment with drug or placebo, test the hypothesis that the proportion of live and dead is the same in the two treatments, total sample >1000
Cochran-Mantel-Haenszel test	3	test hypothesis that proportions are the same in repeated pairings of two groups	alternate hypothesis is a consistent direction of difference	count the number of live and dead patients after treatment with drug or placebo, test the hypothesis that the proportion of live and dead is the same in the two treatments, repeat this experiment at different hospitals

Student-ov t-тест е стандардизиран тест за пресметување на статистичка значајност помеѓу две вредности (**двоен примерок t-тест**) или помеѓу една пресметана аритметичка средина и хипотетизираната вредност. Една претпоставка е дека опсервациите може да бидат нормално дистрибуирани, и дека размерот на варијабилноста кај два примероци не би требало да е повеќе од три. Доколку претпоставките не се точни, тогаш нема па-

Student's t-test: a parametric test for the significance of the difference between means (**two-samples t-test**) or between a mean and a hypothesized value (**one-sample t-test**). One assumption is that the observations must be normally distributed, and the ratio of variances in two samples should not be more than three. If the assumptions are not met, there are non-

раметриски еквиваленти каде би се користел t-тестот (**Вилкохсонов тест**). Не е соодветно да се користи t-тестот за голем број споредувања како **пост хок тест**. T-тестот за независните примероци дали две вредности се или не се значајно различни помеѓу себе, но само доколку тоа се единствените два примерока кои се земени предвид.

Еднонасочна ANOVA претставува техника со која се врши споредување на средните вредности од три или повеќе примероци (користејќи ја F-дистрибуцијата). Оваа техника може да се користи само за обработка на нумерички податоци. ANOVA врши проверка на нултата хипотеза дека две или повеќе групи се извлечени од популации со исти средни вредности. За да се направи ова, две процени се прават за варијабилноста на популацијата. Овие процени се основаат на различни претпоставки. Со ANOVA тестот се добива F-статистика, размерот на варијабилноста измерена помеѓу вредностите до варијабилноста која се наоѓа во самите примероци. Ако средните вредности на групата се добиени од популации со исти средни вредности, варијансата помеѓу средните вредности на групата би требало да е пониска од варијансата на примероците, следејќи ја теоремата за централно ограничување. Повисок размер укажува дека примероците се извлечени од популации со различни средни вредности. Типично, еднонасочната ANOVA се користи за да се пронајдат разликите помеѓу барем три групи, бидејќи студија со две групи може да се тестира со t-тест. Кога има само две вредности за компарација, t-тестот и F-тестот се еквивалентни; односите помеѓу ANOVA и t се добива со $F = t^2$. Додаток на еднонасочна ANOVA е анализа на варијансата во две насоки која го истражува влијанието на две различни категориски независни варијабли врз една зависна варијабла.

Двонасочна ANOVA претставува статистички тест кој се користи за да се одреди ефектот на две номинални предвидливи варијабли на продолжена варијабла која произлегува од самиот резултат. Двонасочната ANOVA го анализира ефектот на независните варијабли врз очекуваниот исход заедно со нивните меѓусебни односи кон самиот исход. Наизменични фактори немаат статистичко влијание врз одреден сет податоци, но треба да се земе предвид дека систематските фактори имаат статистичка значајност (11).

parametric equivalents of the t-test to use (**Wilcoxon's Test**). It is inappropriate to use the t-test for multiple comparisons as a **post hoc test**. The t-test for independent samples tests whether or not two means are significantly different from each other but only if they were the only two samples taken.

One-way ANOVA: is a technique used to compare means of three or more samples (using the F distribution). This technique can be used only for numerical data. The ANOVA tests the null hypothesis that samples in two or more groups are drawn from populations with the same mean values. To do this, two estimates are made of the population variance. These estimates rely on various assumptions. The ANOVA produces an F-statistic, the ratio of the variance calculated among the means to the variance within the samples. If the group means are drawn from populations with the same mean values, the variance between the group means should be lower than the variance of the samples, following the central limit theorem. A higher ratio therefore implies that the samples were drawn from populations with different mean values. Typically, the one-way ANOVA is used to test for differences among at least three groups, since the two-group case can be covered by a t-test. When there are only two means to compare, the t-test and the F-test are equivalent; the relation between ANOVA and t is given by $F = t^2$. An extension of one-way ANOVA is two-way analysis of variance that examines the influence of two different categorical independent variables on one dependent variable.

Two-way ANOVA: a statistical test used to determine the effect of two nominal predictor variables on a continuous outcome variable. A two-way ANOVA test analyzes the effect of the independent variables on the expected outcome along with their relationship to the outcome itself. Random factors would be considered to have no statistical influence on a data set, while systematic factors would be considered to have statistical significance (11).



Париран t-тест претставува статистичка техника која се користи да се изврши споредба помеѓу две средни вредности на две популации во случај каде двата примероци се во меѓусебна корелација. Парираниот t-тест на примероци се користи во студии пред и потоа, или кога примероците се споени парови, или кога станува збор за студија на контролни случаи. На пример, доколку на одреден вработен во компанијата му се одржи обука на одредена тема и доколку сакаме да знаеме дали обуката имала или немала влијание во однос на ефикасноста во работата на тој работник, можеме да го користиме парираниот t-тест на примероци. Собираме податоци од работникот на скала од седум понудени одговори, пред обуката и по обуката. Со користењето на парираниот t-тест на примероци, ние можеме статистички да заклучиме дали обуката имала или немала влијание врз ефикасноста во работата на тој даден работник. Во медицината, со користењето на парираниот t-тест на примероци, можеме да дојдеме до заклучок дали одреден третман, лек може да ја излекува болеста.

Вилксоновиот тест за ранг се користи за да се изврши споредба на два парирани примероци кога добиените податоци се или интервални, но претпоставките за парираниот t-тест (нормалноста во внатрешните разлики) не се задоволителни или се користи ординална (рангирана) скала. Хипотезата која се става на тест е дали разликата во медијаните е нула (како спротивната разлика кај средната вредност кај парираниот t-тест) (12).

Модел на линеарна регресија: Во контекст на моделите на линеарна статистика, „линеарно“ значи линеарно подредени параметри (коефициенти), а не за варијаблите кои го даваат објаснувањето. Објаснувачките варијабли можат да бидат трансформирани (да речеме χ^2), но моделот сè уште ќе биде линеарен доколку коефициентот остане линеарен. Кога сеопфатната функција (Y) остане како сума од термини за кој секој е X-варијабла помножена со коефициентот, функцијата Y се вели дека е линеарна во коефициентите.

Пирсонов коефициент на корелација (r): Со овој коефициент се мери силната страна на линеарната врска помеѓу две квантитативни варијабли. Главната претпоставка е нормалната дистрибуција на варијаблите. Ако

Paired samples t-test: is a statistical technique that is used to compare two population means in the case of two samples that are correlated. Paired sample t-test is used in 'before-after' studies, or when the samples are the matched pairs, or when it is a case-control study. For example, if we give training to a company employee and we want to know whether or not the training had any impact on the efficiency of the employee, we could use the paired sample test. We collect data from the employee on a seven scale rating, before the training and after the training. By using the paired sample t-test, we can statistically conclude whether or not training has improved the efficiency of the employee. In medicine, by using the paired sample t-test, we can figure out whether or not a particular medicine will cure the illness.

Wilcoxon signed rank test: is used to compare two paired samples when data are either *interval* scale but assumptions for the paired *t*-test (normality of within-pair differences) are not satisfied or ordinal (ranked) scale. The hypothesis being tested is whether the median difference is zero (as opposed to mean difference in the paired *t*-test) (12).

Linear regression models: in the context of linear statistical modeling, 'linear' means linear in the parameters (coefficients), not the explanatory variables. The explanatory variables can be transformed (say, x^2), but the model will still be linear if the coefficients remain linear. When the overall function (Y) remains a sum of terms that are each an X variable multiplied by a coefficient, the function Y is said to be linear in the coefficients.

Pearson's correlation coefficient (r): a measure of the strength of the 'linear' relationship between two quantitative variables. A major assumption is the normal distribution of variables. If this assumption is invalid (for example, due to outliers), the non-parametric equivalent **Spearman's rank correlation** should be used. The **r** represents C^2 obtained from the 2x2

оваа претпоставка не е точна, тогаш би требало да се користи непараметрискиот еквивалент на **Спирмановата корелација за рангови**. Во овој случај r го претставува C^2 кое е добиено од 2x2 табелите, кое е исто така коригирано за целосната големина на примерокот. Може да се пресмета како $\pm(C^2/N)^{1/2}$. Оваа формула е еквивалентна на коваријансата поделена со производот помеѓу -1 и +1; додека 0 значи дека не постои линеарна врска (можеби сè уште нема силна линеарна врска или поврзаност). Апсолутната вредност на r ја покажува силната страна на врската. Асоцираната P вредност може да се пресмета за статистичка значајност (мала P вредност не мора да значи дека постои силна врска помеѓу вредностите). Квадратната вредност на r е r^2 (r – на квадрат или **коefficient на детерминација**) кој одговара на варијабилноста која е објаснета со другата поврзана варијабла).

Мултипна регресија: За да можеме да квантифицираме одредена врска помеѓу неколку независни (кои даваат објаснување) варијабли и зависните (исходот) варијабли. Коefициентите (a , b_1 до b_2) се проценува со методот на **најмалку квадрати**, кој е еквивалентен на **максималната шанса** за процена. Повеќекратниот модел на регресија е изграден врз три главни претпоставки:

- Варијаблата на одговори е нормално дистрибуирана.
- Резидуалната варијанса не варира кај мали или големи вредности (константна варијанса).
- Извршените набљудувања (варијаблите за објаснување) се независни.

Мултипната регресија е прототип за **генералните линеарни модели** бидејќи варијаблата на одговор би требало да биде нормално дистрибуирана и не постои **функција на поврзување**, додека едноставната линеарна регресија е посебен случај на **генерализирани линеарни модели**. Додатокот на мултипната регресија до мултиваријантна анализа на податоци се нарекува канонска корелација.

Логистичка (бинарна) регресија: Оваа статистичка анализа најчесто ги опишува врските помеѓу дихотомните (бинарни) варијабли на исход (како здрави или заболени; мртви или живи; болеста се повратила или не) и еден сет на варијабли кои даваат објаснување (возраст, ХЛА тип, крвен притисок, вид

table, corrected for the total sample size. It can then be calculated as $\pm(C^2/N)^{1/2}$. This formula is equivalent to covariance divided by the product of the standard deviations of the two variables. The **correlation coefficient**, r , can take any value between -1 and +1; 0 meaning no "linear" relationship (there may still be a strong non-linear relationship). It is the absolute value of r showing the strength of relationship. An associated P value can be computed for the statistical significance (a small P value does not necessarily mean a strong relationship). The square of the r is r^2 (r -squared or **coefficient of determination**) which corresponds to the variance explained by the correlated variable).

Multiple regression: to quantify the relationship between several independent (explanatory) variables and a dependent (outcome) variable. The coefficients (a , b_1 to b_i) are estimated by the **least squares** method, which is equivalent to **maximum likelihood** estimation. A multiple regression model is built upon three major assumptions:

- The response variable is normally distributed,
- The residual variance does not vary for small and large fitted values (constant variance),
- The observations (explanatory variables) are independent.

Multiple regression is the prototype for **general linear models** because the response variable should be normally distributed and there is no **link function**, whereas, simple linear regression is a special case for **generalized linear models**. The extension of multiple regression to multivariate data analysis is called canonical correlation.

Logistic (binary) regression: a statistical analysis most frequently models the relationship between a dichotomous (binary) outcome variable (such as diseased or healthy; dead or alive; relapsed or not relapsed), and a set of explanatory variables of any kind (such as age, HLA type, blood



на третман, фаза на болеста итн). Исто така може да се користи кога варијаблите на исход се политомични (неколку категории на прогнозата, вклучувајќи ги ординалните одговори како и ординалната логистичка регресија или моделот на пропорционални размери) и кога има неколку варијабли на исход (мултиноминална логистичка регресија – специјална класа на линеарни модели). Анализата на податоците од контролно-експериментална студија преку логистичка регресија можат да се обработуваат на ист начин како и кохортните студии (10).

Горенаведените комбинирани номинални и статистички тестови за мерења можат да се видат во табела 4.

pressure, kind of treatment, disease stage etc). It can also be used when the outcome variable is polytomous (several categories of the prognosis; including ordinal response 'ordinal logistic regression' or 'proportional odds ratio model'), and when there are several outcome variables (multinomial logistic regression - a special class of loglinear models). Analysis of data from case-control studies via logistic regression can proceed in the same way as cohort studies (10).

The above mentioned combined nominal and measurement statistical tests you can find in table 4.

Табела 4. Селектирање статистички тест за номинални и мерливи варијабли комбинирани/

Table 4. Selecting statistical test for nominal and measurement variables combined (4)

Test	Nominal variables	Measurement variables	Purpose	Notes	Example
One-sample <i>t</i> -test	/	1	test the hypothesis that the mean value of the measurement variable equals a theoretical expectation	/	blindfold people, ask them to hold arm at 45° angle, see if mean angle is equal to 45°
Two-sample <i>t</i> -test	1	1	test the hypothesis that the mean values of the measurement variable are the same in two groups	just another name for one-way ANOVA when there are only two groups	compare mean heavy metal content in mussels from Nova Scotia and New Jersey
One-way ANOVA	1	1	test the hypothesis that the mean values of the measurement variable are the same in different groups	/	compare mean heavy metal content in mussels from Nova Scotia, Maine, Massachusetts, Connecticut, New York and New Jersey
Tukey-Kramer test	1	1	after a significant one-way ANOVA, test for significant differences between all pairs of groups	/	compare mean heavy metal content in mussels from Nova Scotia vs. Maine, Nova Scotia vs. Massachusetts, Maine vs. Massachusetts, etc.
Bartlett's test	1	1	test the hypothesis that the standard deviation of a measurement variable is the same in different groups	usually used to see whether data fit one of the assumptions of an ANOVA	compare standard deviation of heavy metal content in mussels from Nova Scotia, Maine, Massachusetts, Connecticut, New York and New Jersey
Nested ANOVA	2+	1	test hypothesis that the mean values of the measurement variable are the same in different groups, when each group is divided into subgroups	subgroups must be arbitrary (model II)	compare mean heavy metal content in mussels from Nova Scotia, Maine, Massachusetts, Connecticut, New York and New Jersey; several mussels from each location, with several metal measurements from each mussel
Two-way ANOVA	2	1	test the hypothesis that different groups, classified two ways, have the same means of the measurement variable	/	compare cholesterol levels in blood of male vegetarians, female vegetarians, male carnivores, and female carnivores

Paired t -test	2	1	test the hypothesis that the means of the continuous variable are the same in paired data	just another name for two-way ANOVA when one nominal variable represents pairs of observations	compare the cholesterol level in blood of people before vs. after switching to a vegetarian diet
Wilcoxon signed-rank test	2	1	test the hypothesis that the means of the measurement variable are the same in paired data	used when the differences of pairs are severely non-normal	compare the cholesterol level in blood of people before vs. after switching to a vegetarian diet, when differences are non-normal
Linear regression	/	2	see whether variation in an independent variable causes some of the variation in a dependent variable; estimate the value of one unmeasured variable corresponding to a measured variable	/	measure chirping speed in crickets at different temperatures, test whether variation in temperature causes variation in chirping speed; or use the estimated relationship to estimate temperature from chirping speed when no thermometer is available
Correlation	/	2	see whether two variables covary	/	measure salt intake and fat intake in different people's diets, to see if people who eat a lot of fat also eat a lot of salt
Polynomial regression	/	2	test the hypothesis that an equation with X^2 , X^3 , etc. fits the Y variable significantly better than a linear regression	/	/
Analysis of covariance (ANCOVA)	1	2	test the hypothesis that different groups have the same regression lines	first test the homogeneity of slopes; if they are not significantly different, test the homogeneity of the Y-intercepts	measure chirping speed vs. temperature in four species of crickets, see if there is significant variation among the species in the slope or Y-intercept of the relationships
Multiple regression	/	3+	fit an equation relating several X variables to a single Y variable	/	measure air temperature, humidity, body mass, leg length, see how they relate to chirping speed in crickets
Simple logistic regression	1	1	fit an equation relating an independent measurement variable to the probability of a value of a dependent nominal variable	/	give different doses of a drug (the measurement variable), record who lives or dies in the next year (the nominal variable)
Multiple logistic regression	1	2+	fit an equation relating more than one independent measurement variable to the probability of a value of a dependent nominal variable	/	record height, weight, blood pressure, age of multiple people, see who lives or dies in the next year

Последната група тестови како што се: тестот на показатели, Крушкар-Валис тестот и Спирмановата рангирана корелација се прикажани во табела 5.

Тест на показатели: Ова е тест кој е базиран на веројатноста да се појават различни исходи за кој било број на плусеви и минуси, т.е. набљудуваните вредности да се под или над

The last group of tests such as: Sign test, Kruskal-Wallis test, and Spearman rank correlation are shown in table 5.

Sign test: a test based on the probabilities of different outcomes for any number of pluses and minuses, i.e., observations below or above a prespecified value. The sign test



предетерминираната вредност. Тестот на показатели може да се користи за да се испита значајноста на разликата помеѓу медијаната и нејзината специфична вредност, или помеѓу набљудуваните пол/трансмисиони односи како и 50:50 очекувана вредност. Исто така може да се користи и за парирани податоци. Сега, разликата помеѓу паровите ќе биде или негативна или позитивна, и колку е помала од двете тотални негативни или позитивни плусеви, тогаш тоталниот број на парови ќе го оформи статистичкиот тест. На пример, кога вкупниот број е 20, ако бројот за помалку фреквентниот показател е 5 или помалку $P < 0,05$. Тест на показатели кој е прикриен е **Мекнемар-тестот**, кој се користи за парирани податоци кај дихотомни одговори.

Крушкар Валис-тестот (еднонасочна ANOVA на рангови): Ова е еден од непараметриските тестови (слободна дистрибуција) и се користи кај претпоставки каде ANOVA не може да биде употребена. Овој тест врши процена на хипотезите каде различни примероци при нивна компарација кои се добиени од иста дистрибуција или од дистрибуција со иста медијана. Може да се користи за да се анализираат ординалните варијабли. Тоа е додаток на **Ман Витни (U) тестот** (се користи кај два независни примероци). Интерпретацијата на Крушкар Валис тестот е идентична како со еднонасочната ANOVA, но се базира врз ранговите, а не според просеците. Како и сите непараметриски тестови Крушкар Валис-тестот не е доволно моќен како ANOVA-тестот.

Спирманова корелација на рангови е непараметриски **корелациски коефициент (rho)** кој се пресметува со **Пирсоновиот корелациски коефициент (r)** за асоцијацијата помеѓу ранговите која е дадена преку вредностите од варијаблите кои се користат. Се користи кај ординални податоци и кај интервални и размерни податоци (податоци добиени преку овие скали). Не е соодветно да се земе квадратната вредност од Спирмановиот корелациски коефициент rho за да се добие **коефициент на детерминираност (r^2)**. **Спирмановата корелација** помеѓу две варијабли е еднаква на Пирсоновата корелација помеѓу вредности на рангови од тие две варијабли; додека Пирсоновата корелација врши процена на линеарните врски, Спирмановата корелација врши процена на монотоничните врски (без разлика дали се линеарни или не). Доколку нема вредности кај податоците кои се повторуваат, се оформува совр-

can be used to investigate the significance of the difference between a population median and a specified value for it, or between the observed sex/transmission ratio and the 50:50 expected value. It can also be used for paired data. This time, the differences between the pairs will be either negative or positive, and the smaller of the two total negatives or positives plus the total number of pairs will form the test statistics. For example, when the total number is 20, if the number for the less frequent sign is 5 or smaller, $P < 0.05$ (two-tailed). A sign test in disguise is **McNemar's test**, which is used for paired data for dichotomous response.

Kruskal-Wallis test (One-way ANOVA by ranks): it is one of the non-parametric (distribution free) tests and is used when the assumptions of ANOVA are not met. This test assesses the hypothesis that the different samples in the comparison were drawn from the same distribution or from distributions with the same median. It can be used to analyze ordinal variables. It is an extension of the **Mann-Whitney (U) test** (for two independent samples). The interpretation of the Kruskal-Wallis test is identical to that of one-way ANOVA, but is based on ranks rather than means. Like all non-parametric tests, the Kruskal-Wallis Test is not as powerful as the ANOVA.

Spearman's rank correlation: a non-parametric **correlation**

coefficient (rho) that is calculated by computing the **Pearson's correlation coefficient (r)** for the association between the ranks given to the values of the variables involved. It is used for ordinal data and interval/ratio data. It is not appropriate to take the square of Spearman's correlation coefficient rho to obtain **coefficient of determination (r^2)**. The **Spearman correlation** between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). If there are no repeated data values, a perfect

шената Спирманова корелација од +1 или -1 кога секоја од варијаблите претставува перфектна монотона функција на другата варијабла (10).

Моќност, P-вредности и проценти

Пред да се започне со истражувањето важно е да се направи пресметка за големината на примерокот за да се увиди дали ќе има доволно моќност да се детектираат значајните разлики (13). **Анализите на моќност** се директно поврзани со тестирањето на хипотезите. Додека се спроведуваат тестирања на хипотезите, истражувачот може да направи два вида грешка: тип 1 грешка и тип 2 грешка. Статистичката моќност главно се занимава со тип 2 грешки. Би требало да се забележи од страна на истражувачот дека колку е поголема големината на примерокот, толку е полесно за истражувачот да добие 0,05 ниво на статистичка значајност. Доколку примерокот е премногу мал, тогаш истражувачот би можел да направи тип 2 грешка поради немањето доволно голема моќност. Анализата на моќноста нормално се спроведува пред собирањето на податоците. Главната цел кај анализите на моќност е да му се помогне на истражувачот да го одреди најмалиот примерок кој би бил соодветен за да се детектира ефектот на кој било даден тест на посакуваното ниво на статистичка значајност. Причината за аплицирање на анализата за моќност е бидејќи, идеално гледано, истражувачот сакка помал примерок зашто големите примероци често се потешки за обработка отколку малите примероци. Помалите примероци исто така вршат оптимизација на значајноста на тестирањето (14).

Исто така, често е потребно да се комбинираат категориите за да се добијат доволен број во секоја од групите за компарација. На пример, за χ^2 -квадрат тестот да има валидни резултати, потребно е да има очекувана фреквенција во секоја клетка поголема од 5. Кога се цитираат проценти од голема важност исто така се цитира и броителот или именителот за да биде јасно како се пресметале процентите. Процентите базирани врз мали бројки, како помалку од 10, најчесто не се статистички значајни. Ниво на значајност (P-вредност) се смета за значајна доколку е помала од 0,05. Доволно е да се цитира P-вредноста до две децимали доколку е поголема од 0,01. Меѓутоа доколку P-вредноста е многу мала како на пример: $P < 0,0001$ тогаш истата би требало и да се користи (12).

Spearman correlation of +1 or -1 occurs when each of the variables is a perfect monotone function of the other (10).

Power, P values, and percentages

Before undertaking a research study it is important to make a sample size calculation so that the study will have sufficient power to detect significant differences (13). **Power analysis** is directly related to tests of hypotheses. While conducting tests of hypotheses, the researcher can commit two types of errors: Type I error and Type II error. Statistical power mainly deals with Type II errors. It should be noted by the researcher that the larger the size of the sample, the easier it is for the researcher to achieve the 0.05 level of significance. If the sample is too small, however, then the investigator might commit a Type II error due to insufficient power. Power analysis is normally conducted before the data collection. The main purpose underlying power analysis is to help the researcher to determine the smallest sample size that is suitable to detect the effect of a given test at the desired level of significance. The reason for applying power analysis is that, ideally, the investigator desires a smaller sample because larger samples are often costlier than smaller samples. Smaller samples also optimize the significance testing (14).

In addition to this it is often necessary to combine categories so that there are sufficient numbers in each group for comparison. For example, for the chi-square test to have valid results there needs to be an expected frequency in each cell of more than 5. When quoting percentages, it is essential to also quote the numerator and/or the denominator so that it is clear how the percentage has been calculated. Percentages based on small numbers such as less than 10 are not meaningful. A significance level (*P* value) is considered significant if it is less than 0.05. It is sufficient to quote *P* values to two decimal places if greater than 0.01. However, if the *P* value is very small then $P < 0.0001$ should be used (12).



Табела 5. Селектирање статистички тест за номинални и рангирани варијабли комбинирано/
Table 5. Selecting statistical test for nominal and ranked variables combined (4)

Test	Nominal variables	Measurement variables	Purpose	Notes	Example
Sign test	2	1	test randomness of direction of difference in paired data	/	compare the cholesterol level in blood of people before vs. after switching to a vegetarian diet, only record whether it is higher or lower after the switch
Kruskal–Wallis test	1	1	test the hypothesis that rankings are the same in different groups	often used as a non-parametric alternative to one-way ANOVA	40 ears of corn (8 from each of 5 varieties) are ranked for tastiness, and the mean rank is compared among varieties
Spearman rank correlation	/	1	see whether the ranks of two variables covary	often used as a non-parametric alternative to regression or correlation	40 ears of corn are ranked for tastiness and prettiness, see whether prettier corn is also tastier

Компјутерски програми за пресметување на статистички податоци

Сите погоре наведени статистички тестови можат да бидат пресметани со помош на компјутерски програми за нивно пресметување, како што се: SPSS, SAS/STAT, LISREL, BILOG-MG, STATA итн. Во продолжение ќе бидат наведени неколку карактеристики на овие програми.

SPSS (<http://www.spss.com/>)

Компјутерската програма SPSS (Статистички пакет за пресметување податоци од општествените науки) за првпат беше употребена во 1968 и е една од најшироко користените програми за статистички анализи во рамките на општествените науки. Неа ја користат истражувачи на пазарот, истражувачи на здравството, компании кои спроведуваат анкети, владите, истражувачи во образованието и наставата и други. Во прилог на статистичките анализи на оваа програма, се содржани и управувањето со податоци (селекција на случаи, преформулација на податоци, обработка на добиените податоци) и документирањето на податоците (речникот на метаподатоците се зачувува во рамките на главната база на податоци) кои се основни карактеристики на главниот софтвер. Оваа програма е доста корисна и поради тоа може да се применува и во биомедицината и во општествените науки.

SAS/STAT (<http://www.sas.com/>)

Од традиционални дескриптивни статистики, t-тестови, анализи на варијанса/разлики и модели на предвидливост до егзактни методи и визуелно статистички техники, SAS/STAT-програмата обезбедува алатки и за специјализираните и за аналитичките потреби на пошироката јавност.

Statistical Software Packages

All the statistical tests mentioned above could be measured with statistical software packages such as: SPSS, SAS/STAT, LISREL, BILOG-MG, STATA, etc (15). Bellow are listed some characteristics of these software.

SPSS (<http://www.spss.com/>)

The computer program SPSS (originally, Statistical Package for the Social Sciences) was released in its first version in 1968, and is among the most widely used programs for statistical analysis in social science. It is used by market researchers, health researchers, survey companies, government, education researchers, and others. In addition to statistical analysis, data management (case selection, file reshaping, creating derived data) and data documentation (a metadata dictionary is stored with the data) are features of the base software. It is very useful software which is widely applied in biomedical and social sciences.

SAS/STAT (<http://www.sas.com/>)

From traditional descriptive statistics, t-tests, analysis of variance, and predictive modeling to exact methods and statistical visualization techniques, SAS/STAT software provides tools for both specialized and enterprise-wide analytical needs.

LISREL

(<http://www.ssicentral.com/lisrel/index.html>)

Во последните 30 години, моделот, методите и софтверот LISREL станаа синоним за структуралното моделирање на равенките (SEM). SEM им овозможува на истражувачите од општествените науки, науките за менаџмент, бихејвиоралните науки, биолошките науки, науките за образованието и во други полиња, за емпириско потврдување на нивните теории. Овие теории се формулирани како теоретски модели на набљудуваните и латентните (без можност за набљудување) варијабли. Ако податоците се однесуваат на варијаблите кои се набљудувани во рамките на теоретскиот модел, програмата LISREL може да се користи за прилагодување на моделот со податоците.

BILOG-MG

(<http://www.ssicentral.com/irt/index.html>)

BILOG-MG се справува со повеќе групи, повеќе поттестови и повеќе форми на тестови во рамките на една анализа и содржи поедноставени форми на спецификациите. Дистрибуцијата на нормалните и емпириските латентни податоци може да биде специфицирана од страна на корисникот, така што процената на резултатите од скалите може да биде направена за конкретна група или целосно. Се пресметуваат кривите и статистичките податоци според формата и според тестот. Корисникот може да изведе слични пресметки за какви било произволно селектирани подгрупи на предметите.

STATA (<http://www.stata.com/>)

STATA е целосна, интегрална компјутерска програма за статистичка анализа која обезбедува сè што е потребно за анализирање на податоци, менаџирање на податоци и графички прикази.

Заклучок

Знаејќи како да се избере вистинскиот статистички тест е голема предност во обработката на податоците кои се добиени со истражувањето и при пишувањето на научни трудови. Младите истражувачи и автори треба да имаат познавања за тоа како да изберат и како да ги користат статистичките методи. Во денешно време, поголемите издавачи на научни списанија имаат уредник за статистичките анализи, што не е случај во поголемиот број списанија кои се издаваат во држави со слаба економија (на пример, балканските држави).

LISREL

(<http://www.ssicentral.com/lisrel/index.html>)

During the last thirty years, the LISREL model, methods, and software have become synonymous with structural equation modeling (SEM). SEM allows researchers in the social sciences, management sciences, behavioral sciences, biological sciences, educational sciences, and other fields to empirically assess their theories. These theories are usually formulated as theoretical models for observed and latent (unobservable) variables. If data are collected for the observed variables of the theoretical model, the LISREL program can be used to fit the model to the data.

BILOG-MG

(<http://www.ssicentral.com/irt/index.html>)

BILOG-MG handles multiple groups, multiple subtests, and multiple test forms in one analysis, and it includes a simplified forms specification. Both normal and empirical latent distributions can be specified by the user. Bayes estimation of scale scores can be done with a group-specific or global prior. It computes information curves and statistics by form as well as by test. The user can perform similar computations for any arbitrarily selected subset of items.

STATA (<http://www.stata.com/>)

Stata is a complete, integrated statistical package that provides everything you need for data analysis, data management, and graphics.

Conclusion

Knowing how to choose right statistical test is an important asset and decision in the research data processing and in the writing of scientific papers. Young researchers and authors should know how to choose and how to use statistical methods. Nowadays, bigger journal publishers have statistical editor in their editorial office, which is not a case in most of the journals from low income countries (for example, Balkan countries).



Потребно е добро да се знае со какви податоци располагаме, како тие податоци се организирани, со колку примероци/групи треба да се справиме и дали се во парови или не. Мора да се запрашама дали податоците се извлечени за популацијата која ѝ припаѓа или не припаѓа на Гаусовата крива и дали се исполнети потребните услови, со цел да се избере еднонасочен тест (во споредба со двонасочниот тест, кој е најчесто препорачана форма на употреба). Врз основа на таков вид податоци, ние можеме да ја следиме соодветната статистичка гранка, користејќи алгоритмички начини кои ќе не доведат до вистинскиот тест, без да се направи грешка во процесот на избирање на соодветниот тест.

Еден компетентен истражувач потребно е да има познавања од областа на статистичките постапки. Тоа може да вклучува курс за вовед во статистиката и мора да користи добри книги за статистичка обработка на податоците. За таа цел, потребно е курсевите за статистика да бидат задолжителни за сите студенти, како што и беше во претходната студиска програма на Филозофскиот факултет во Скопје. Младите истражувачи имаат потреба од дополнителни курсеви за статистика. Тие мора да ги усовршат своите вештини за да можат правилно да ги користат статистичките програми.

Издавачите на научни списанија во Македонија треба повеќе да се потрудат за да обезбедат материјални и финансиски средства за да можат да вработат уредник за статистичките податоци. На тој начин, главните уредници може да се надеваат на поголема значајност на нивните трудови и поголема цитираност што би било добар предуслов за добивање на импакт-фактор.

Конфликт на интереси

Авторот изјавува дека не постои конфликт на интереси.

There is need to know what type of data we may have, how are these data organized, how many sample/groups we have to deal with and if they are paired or unpaired; we have to ask ourselves if the data are drawn for a Gaussian on non-Gaussian population and, if the proper conditions are met, to choose an one-tailed test (versus the two-tailed one, which is, usually, the recommended choice). Based on such kind of information, we may follow a proper statistical decision-tree, using an algorithmic manner able to lead us to the right test, without any mistakes during the test selection process.

The competent researcher will need knowledge in statistical procedures. That might include an introductory statistics course, and it most certainly includes using a good statistics textbook. For this purpose, there is need course of Statistics to become mandatory (obligatory) for all students as it was in former curricula at the Faculty of Philosophy in Skopje. Young researchers have a need of additional courses in statistics. They need to train themselves to use statistical software on appropriate way.

Macedonian publishers of scientific journals shall make greater efforts to provide material-financial assets in order to employ a statistical editor. On that way, editorial offices can start counting on greater impact of their articles and greater citations which will be good prerequisite for obtaining an impact factor.

Conflict of interests

Author declares no conflict of interests.

Референци / References

1. Singurpuwalla D. A Handbook of Statistics: an overview of statistical methods. 2013. Available from: <http://bookboon.com> [Accessed on 2016 July 9].
2. British Medical Journal. Study design and choosing a statistical test. Available from: <http://www.bmj.com> [Accessed on 2016 July 10].
3. Strasak AM, Zaman Q, Karl PP, Gobel G, Ulmer H. Statistical errors in medical research-a review of common pitfalls. Swiss MED Wkly 2007;137: 44–49.
4. McDonald JH. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland. Available from: <http://www.biostathandbook.com/analysissteps.html>. [Accessed on 2016 July 10].
5. Marusteri M, Bacarea V. Comparing groups for statistical differences: how to choose the right statistical test? Biochemia Medica 2010;20(1):



- 15-32. <http://dx.doi.org/10.11613/BM.2010.004>.
6. Lane DM, Scott D, Hebl M, Guerra R, Osherson D, Heidi Zimmer H. Introduction to statistics. Rice University; 2University of Houston, Downtown Campus. Available from: <http://onlinestatbook.com> [Accessed on 2016 July 10].
7. Jaykaran. How to select appropriate statistical test? Journal of Pharmaceutical Negative Results 2010;1(2): 61–63.
8. Rosner B. Fundamentals of Biostatistics, California: Duxbury Press, 2000.
9. Siegel S. *Nonparametric statistics for the behavioral sciences*. 2nd ed. London: McGraw-Hill; 1988.
10. Dorak MT. Common Concepts in Statistics, 2015. Available from: <http://www.dorak.info/mtd/glosstat.html> [Accessed on 2016 July 11].
11. Statistics and Probability Dictionary. Available from: <http://stattrek.com/statistics/dictionary.aspx?definition=Confidence%20interval> [Accessed on 2016 July 11].
12. McCrum-Gardner E. Which is the correct statistical test to use? British Journal of Oral and Maxillofacial Surgery 2008; 46: 38–41.
13. Interactive Statistical Calculation Pages. Available from: <http://www.statpages.org/#Power> [Accessed on 2016 July 11].
14. Statistics Solutions: Advancement Through Clarity. Available from: <http://www.statisticssolutions.com/statistical-power-analysis/> [Accessed on 2016 July 12].
15. Educational Measurement and Research. Statistical Software Packages. College of Education. Available from: <http://www.coedu.usf.edu/> [Accessed on 2016 July 11].

Copyright of Journal of Special Education & Rehabilitation is the property of De Gruyter Open and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.