



Магистърска програма „Спорт за високи постижения“

MSP.1.6.2. Статистически методи в спорта

Указания за ползване на
учебното помагало

[Въведение](#)

[Тема 1. Основни понятия](#)

[Тема 2. Въведение в SPSS](#)

[Тема 3. Характеристика на разпределението на променливи величини](#)

[Тема 4. Корелационен анализ](#)

[Тема 5. Проверка на хипотези при независими извадки](#)

[Тема 6. Проверка на хипотези при зависими извадки](#)

[Тема 7. Обработка на данни от експеримент](#)





ВЪВЕДЕНИЕ В УЧЕБНАТА ДИСЦИПЛИНА

В настоящия раздел се дават общи указания за протичане на обучението.

- [Цел на обучението](#)
- [Учебна програма](#)
- [Изисквания за включване](#)
- [Учебни материали](#)
- [Протичане на заниманията](#)
- [Как да изберете своя индивидуален стил на обучение](#)
- [Изпитни изисквания](#)
- [Литература](#)
- [Контакт с преподавателя](#)

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Целта на обучението по учебната дисциплина „Статистически методи в спорта“ в ОКС „Магистър“ е да се задълбочат познанията и практическите умения на студентите по отношение на:

- приложение на статистическите методи в предметната област, в която се обучават,
- подготовка и въвеждане на данни от научно изследване,
- обработка на данни със специализирания програмен продукт **IBM SPSS Statistics**,
- представяне в таблична и графична форма на резултати от статистическа обработка,
- интерпретация на резултати от статистическа обработка.

При завършване на обучението студентите трябва **да могат**:

- да подбират подходящ за изследвания проблем и променливи величини статистически метод,
- да осъществят самостоятелно статистическата обработка с IBM SPSS,
- да представят и интерпретират резултати от статистическата обработка.

УЧЕБНА ПРОГРАМА

Въведение

Запознайте се с учебната програма и хорариума за студентите от редовната форма на обучение:

№	Наименование	Лекции	Упражнения	Самостоятелна работа	Общо
1.	Основни понятия от общата теория на статистиката (извадки, основни видове измерителни скали, общ преглед на статистическите методи в спорта).	1			1
2.	Обща характеристика на IBM SPSS 19.0, дефиниране на променливи, въвеждане на данни, трансформиране на променливи.		2		2
3.	Характеристика на разпределението на категорийни и количествени променливи (честотен и вариационен анализ).	1	1		2
4.	Изследване на зависимости (обикновена линейна и рангова корелация).	1	2		3
5.	Проверка на хипотези при независими извадки (t-критерий на Стюдънт, F-критерий на Фишер).	1	2		3
6.	Проверка на хипотези при зависими извадки (t-критерий на Стюдънт).		1		2
7.	Обработка на данни от педагогически експеримент	1	2		2
	Самостоятелна работа по изпитния проект			15	15
Общо		5	10	15	30

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Като правило е необходимо студентите да имат познания по статистика на бакалавърско ниво, но учебните материали, дават възможност да бъдат усвоени липсващите понятия. Студентите, от **присъствените форми** на обучение трябва да проведат обучение в зала 205Б – 70-ти блок, което има инсталирани необходимите за обучението програмни продукти.

За успешното провеждане на **дистанционната форма** на обучение е необходимо да имате:

- умения за работа с портала за дистанционно обучение;
- инсталирани продуктите на **MS Office** – MS Word, MS Excel и MS Power Point (версии след 2007 г.), аудио и видео плейър.
- инсталиран програмен продукт **IBM SPSS Statistics**. Програмата може да се ползва на територията на НСА (зала 205Б в 70-ти блок) или да се изтегли от интернет адрес <http://www-01.ibm.com/software/analytics/spss/downloads.html> (версията е 14 дневна).

За всяка от темите в процеса на обучение се ползват няколко вида учебни материали:



Пълно описание на изучаваното учебно съдържание в настоящото учебно помагало;



Кратко схематично представяне на материала под формата на **флаш презентации**;



Аудио лекции, в които е записан лекционният материал;



Учебни филми, в които е показан алгоритмът за изчисляване на показателите с **IBM SPSS Statistics**;

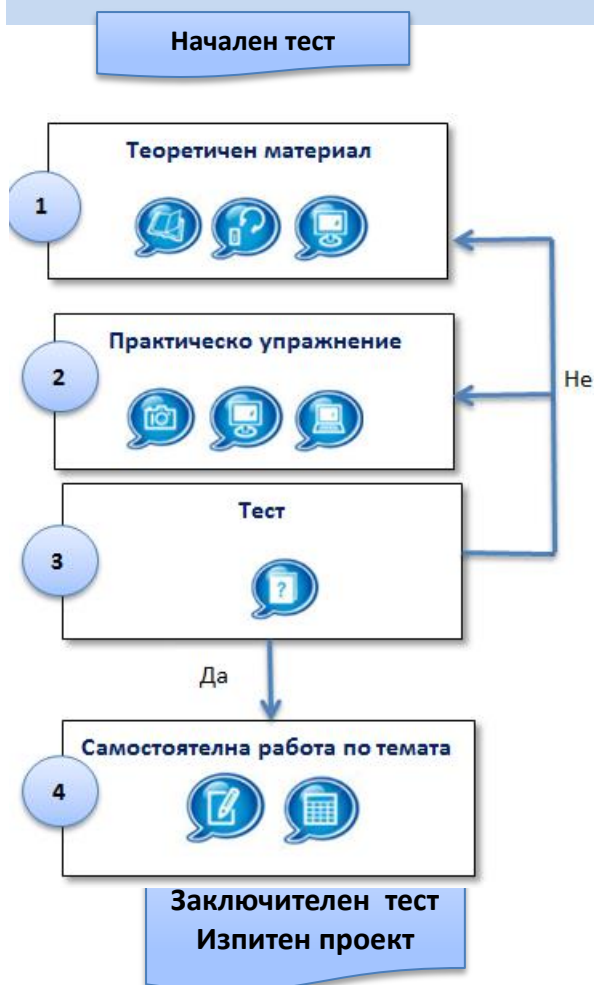


Интерактивни калкулатори ([MS Excel файл](#)), които дават възможност изучаваните показатели да бъдат изчислени от потребители, които не разполагат с **IBM SPSS Statistics**;



Предложени примерни файлове с данни от различни предметни области в сферата на спорта и физическото възпитание (в SPSS и MS Excel формат).

В случай, че се обучавате в средата за дистанционно обучение **virtual.nsa**, достъпът до учебните материали се осъществява от електронния учебник. В случай, че ползвате настоящото учебно помагало – от линковете, които са поставени на съответното място в помагалото.



Заниманията в **присъствените форми** на обучение протичат в следната последователност:

1. Студентът се запознава с **теоретичния материал** по темата;
2. Преподавателят показва как се изчисляват статистическите показатели, залегнали в темата (**задача за разучаване**), след което студентът самостоятелно решава **задачата за упражнение**.
3. Знанията и практическите умения на студентите по темата се проверяват с **тест по материала**.
4. Обучението по темата завършва, като студентът направи самостоятелна обработка и текстуален анализ на резултати от прилагане на съответния статистически метод. Финалният изпитен проект включва тези анализи.

Алгоритмът на обучение на студентите от **дистанционната форма** е аналогичен и е показан на фигурата. Теоретичният материал може да се ползва в някой от посочените на фигурата учебни материали. Студентът осъществява самостоятелно задачите за разучаване и упражнение или проследява тяхното решение от учебните филми.

Попълването на финалния изпитен проект може да стане и като се ползват Екселовите файлове. За целта трябва данните, които се обработват, да се поставят в предвидените за целта полета, а резултатите от обработката да се препишат и анализират в изпитния проект.

Студентите, които се обучават в ОКС "Магистър" се различават по ниво на своята предварителна подготовка, начин на живот, време, с което разполагат за подготовка.

Поради това Ви препоръчваме да следвате своя собствен ритъм!

Тествайте входното ниво на познания си по статистика с входящия тест,

Файлт с тест за входящо ниво може да изтеглите от [тук](#).

Ако на него имате познания по повече от въпросите - теоретичната Ви подготовка е на достатъчно високо ниво. За подготовка е достатъчно да ползвате материала в електронния учебник на virtual.nsa.

Ако установите, че нямате необходимата подготовка е добре да разгледате пълното учебно помагало, който може да изтеглите от рубриката "Учебни материали".

Тествайте предпочитания от Вас стил на обучение.

Достъп до анкетата има [тук](#)

Ако основният предпочитан от Вас стил на обучение е **визуален** - обърнете внимание на презентациите на учебното съдържание и филмите за работа със SPSS.

Ако имате предпочитание към **кинестетичния** и **тактилният** стил - изтеглете SPSS и екселовият калкулатор, въвеждайте примерни данни и проследявайте промените, които настъпват в статистическите показатели.

Ако имате предпочитание към **аудиостила** или ежедневието Ви е натоварено и нямате време за разглеждане на теоретичния материал - изтеглете аудиозаписите на учебния материал и ги слушайте в удобно за Вас време.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ИЗПИТНИ ИЗИСКВАНИЯ

Обучението приключва, когато студентът е показал достатъчно високо ниво на теоретични познания и е направил изпитен проект, който включва резултатите от обработката и анализите, които студентът е направил в стъпка 4 от обучението.

Студенти от редовна и задочна форма на обучение. В деня на изпита студентът представя оформения на хартиен носител изпитен проект. Изпитът се състои в събеседване върху някоя от задачите, решени самостоятелно от студента и теоретични въпроси, свързани с изучаваните статистически анализи. Оценката се оформя на базата на качеството на представяне на резултатите в изпитния проект (40% от оценката) и умението на студента да интерпретира получените резултати (60% от оценката).

Студенти от дистанционна форма на обучение. Оценката се поставя на базата на финалния тест и умението на студента да интерпретира статистическа информация, показано в анализите, залежали в изпитния проект.

Финалният тест съдържа въпроси, които са включени в междинните тестове за обучение към всяка от темите. Ако старателно сте попълвали междинните тестове, няма да се изненадате от въпросите, които съдържа финалният тест. Изпитният проект трябва да бъде качен на сайта за дистанционно обучение в предвидените в графика на обучението срокове.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ЛИТЕРАТУРА

1. **Брогли, Я.** Статистически методи в спорта, Миф, С. 1983
2. **Брогли, Я.** Въведение в теорията и практиката на контрола върху факторите на спортното постижение, НСА Прес, 2012
3. **Брогли, Я. Л. Петкова** Статистически методи в спорта-учебник за студентите от ВИФ „Г. Димитров“, Миф, С., 1988
4. **Гатев, К., Н. Гатева** Статистика – статистически методи в емпиричните изследвания и бизнеса, Парадигма, 2008
5. **Гигова, В., Р. Дамянова** Статистически методи в спорта – ръководство за студентите от ОКС „Бакалавър“ на НСА, НСА-прес, 2012
6. **Гоев, В.** Статистическа обработка и анализ на информация от социологически, маркетингови и политически изследвания, Стопанство, 1996
7. **Дамгалиев, Д., Ж.Телляян** Бизнесстатистика, Издателство на НБУ, 2006
8. **Калинов, К.** Статистически методи в поведенческите и социалните науки, НБУ, 2010
9. **Маринов, Кр.** Директен маркетинг – концепции и творчески решения , Университетско издателство „Стопанство“, С. 2011
10. **Мишев, Г. С. Цветков** Статистика за икономисти, Университетско издателство „Стопанство“, С. 2008
11. **Павлова, В. С. Чипева** Статистика, Нова звезда, С. 2012
12. **Петров, С. С. Стефанова** Обща теория на статистиката , Парадигма, С., 2009
13. **Съйкова, И. А. Къналиева, С. Съйкова** Статистическо изследване на зависимости, Университетско издателство „Стопанство“, С. 2002
14. **Харалампиев, К.** Работа с данни в SPSS, Университетско издателство „Св. Климент Охридски“, С., 2009
15. **Field, A.** Discovering Statistics Using SPSS for Windows, SAGE Publications, 2000
16. **Kenny, D.** Statistics for the Social and Behavioral Sciences, 1987, достъпна на <http://davidakenny.net/doc/statbook/kenny87.pdf>
17. **Vincent, W.** Statistics in kinesiology, Human Kinetics, 1995

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

АДРЕСИ ЗА КОНТАКТ

Ако имате конкретни въпроси по протичане на обучението, изискванията към студентите и приключване на модула, може да ги отправите на адреси:

доц. Валентина Гигова - v_gigova@yahoo.com

ас. Силвия Шандуркова - silvia.shandurkova@gmail.com

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ТЕМА 1. ОСНОВНИ ПОНЯТИЯ



В настоящата тема се прави преглед на основните понятия, изучавани в ОКС „Бакалавър“

Теоретичен материал

- [Генерална съвкупност и извадка](#)
- [Видове извадки](#)
- [Признак, измерване, основни видове измерителни скали;](#)
- [Систематика на статистическите методи, изучавани в ОКС „Магистър“](#)



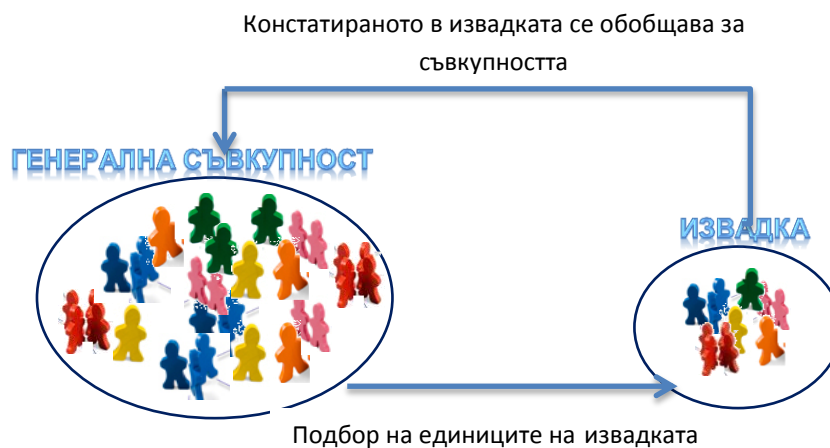
ВЪВЕДЕНИЕ В ПОНЯТИЯТА

Генерална съвкупност и извадка

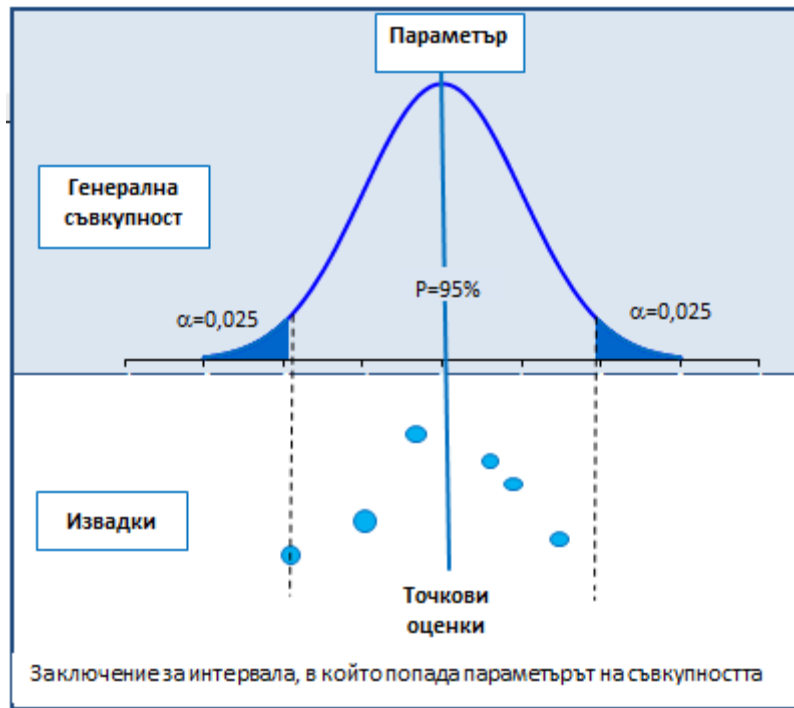
Научните изследвания в областта на спорта целят да се установи какво е състоянието и поведението на човека в условията на тренировката и състезанието. Получените изводи трябва да се основават на информация и да се отнасят за **конкретна група от състезатели от даден вид спорт, пол, възраст, квалификация**. От статистическа гледна точка, обектите (статистическите единици), които имат сходни характеристики от гледна точка на изучавания проблем представляват **генерална съвкупност**.

Генералната съвкупност (ГС) е множеството от статистически единици, в които се проявява изучаваното явление (например, висококвалифицирани състезатели по тенис). Те имат някои общи характеристики и други – по които се различават (ръст, тегло, бързина, сила, издръжливост).

Тема 1



Фиг. 1.1



Фиг. 1.2

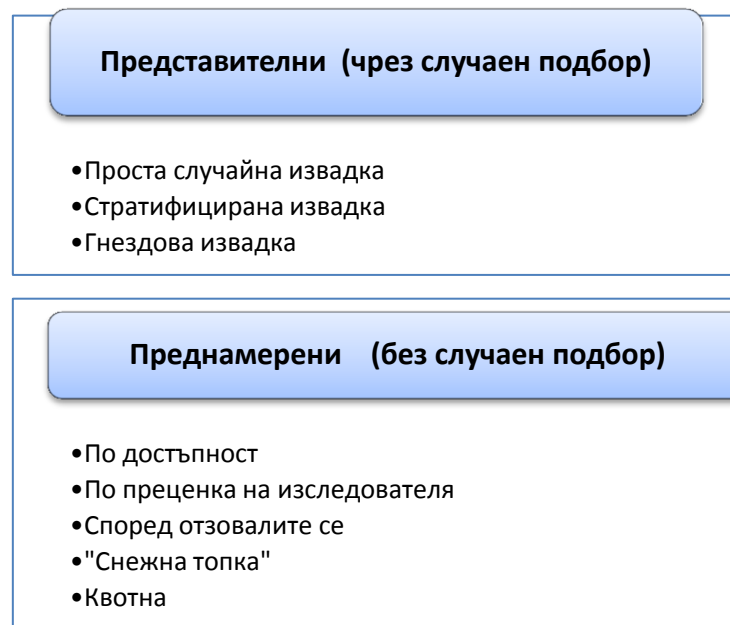
Броят на единиците, попадащи в съвкупността се нарича **обем на съвкупността** и се бележи с **N**. В някои от случаите съвкупността е съставена от конкретен брой обекти, които могат да бъдат идентифицирани (например АТР идентификационния номер на даден състезател по тенис), а в други случаи тя е абстрактно множество от обекти (например спортуващи ученици). Ако проучването обхваща цялата съвкупност, се нарича **изчерпателно**. Неговото осъществяване е трудно, а в някои случаи - невъзможно. Поради това обикновено се изследва част от съвкупността.

Извадката е онази част от съвкупността, която се изследва. Броят на обектите в извадката нарича **обем на извадката** и се бележи с **n**. Поредният номер на дадена статистическа единица се бележи с **i**, като в извадката **i** заема стойности от 1 до **n**.

Статистическите показатели, които описват състоянието на изследваното явление в извадката се наричат **точкови оценки**, а тези, които характеризират съвкупността – **параметри на съвкупността**. На базата на точковите оценки се правят вероятностни заключения за границите, в които се намира съответният параметър на генералната съвкупност. Сигурността на статистическия извод се нарича **гаранционна вероятност P(%)**, а възможността за грешка - **равнище на значимост (α)**. Обикновено се работи с гаранционна вероятност $P(\%)=95\%$, на което съответства равнище на значимост $\alpha=0,05$.

Възлово място в етапа на организация на научното изследване и решаващ фактор за коректността на получените резултати има начинът, по който са подбрани единиците в извадката. Най-общо методите за подбор на статистическите единици, а оттам и видовете извадки се делят на две големи групи:

- Такива, при които се прилага **случаен подбор**. Получените по този начин извадки се наричат **представителни (репрезентативни)**;
- Такива, при които не се прилага случаен подбор. В такива случаи извадките се наричат **преднамерени**.



Фиг. 1.3

Случаен е подборът, при който единиците на съвкупността имат равен шанс да попаднат в извадката. От тази гледна точка извадките биват:

- **Проста случайна извадка.** Тя се получава, когато единиците на ГС се номерират от **1** до **N** и **лотарийно** се изтегли номерът на единиците, който формират извадката.
- **Стратифицирана извадка.** Съвкупността се разделя на **големи подсъвкупности - страти**. Случайният подбор се провежда в рамките на всяка от стратите.
- **Гнездова извадка.** Съвкупността се разделя на голям брой **малки групи (гнезда)**. Чрез случаен подбор се селектират гнездата, които да попаднат в извадката. Изследват се всички единици в подбраните гнезда.

Преднамерен е подборът, при който единиците на съвкупността няма еднакъв шанс да попаднат в извадката. От тази гледна точка извадките се делят на:

- **Извадка по удобство (достъпност)** – изследват се тези единици на ГС, до които изследователят има достъп (например изследователят има възможност да проведе изследването сред 10 тенис клуба из страната);
- **Извадка преценка на изследователя** – изследват се тези единици, които отговарят на зададен от изследователя критерий (например, изследват се тенисистите, които са участвали във финали на турнири по тенис от Големия шлем)
- **Извадка според отзовалите се** – изследват се тези единици, които са се съгласили да бъдат изследвани (например, при доброволно анкетно проучване, извадката обхваща тези, които са попълнили анкетната карта);
- **Извадка „Снежна топка“.** Ползва се, когато не са известни единиците на съвкупността, например не са известни потребители на даден вид услуга. След като се открие първият потребител се задава въпросът известни ли са му други потребители, който са ползвали съответната услуга. При положителен отговор се изследват и тези статистически единици. По този начин извадката постепенно нараства като снежна топка, откъдето произлиза името на извадката.
- **Квотна извадка** – прилага се, когато ГС има известна на изследователя структура (например, подрастващи гимнастици, разпределени във възрастово-полови групи. Предварително се уточнява относителният дял на всяка от под съвкупностите, определя се необходимият обем на извадката и се определя по колко лица от възрастово-полова група да се изследват.

В много от случаите реално направените извадки представляват комбинация от посочените тук основни видове и формирането им се осъществява на отделни етапи (стъпки).

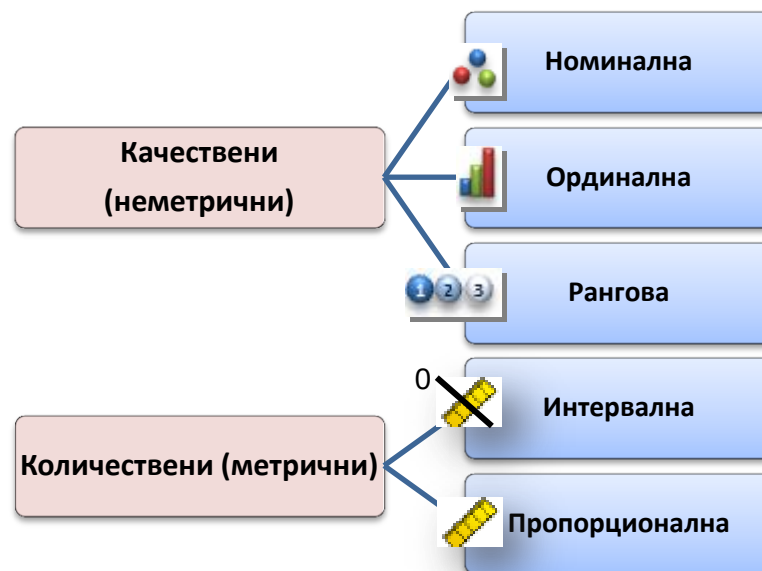
ПРИЗНАЦИ, ИЗМЕРВАНЕ, ИЗМЕРИТЕЛНИ СКАЛИ

Обектите, които се изследват, си приличат по характеристиките, по които е формирана ГС, но се различават по много други. Именно те представляват интерес при статистическото проучване.

Признаци са характеристиките (особеностите) на обектите, която се изследват. За да се получи информация за тяхното състояние е необходимо същите да бъдат измерени. От статистическа гледна точка **измерването** е процес, при който на дадено състояние на признака се присъжда число, в съответствие с някаква скала за измерване.

Тема 1

От гледна точка на това каква информация (качествена или количествена) се получава в резултат на измерването и какви свойства притежават получените в резултат на измерването стойности измерителните скали биват:



Фиг. 1.4

КАЧЕСТВЕНИ (НЕМЕТРИЧНИ), са скалите, при които в резултат на измерването се получава качествена информация за обектите:

- 
Номинална, при която обектите се причисляват към еднородни групи, между които няма разлика в интензитета на притежавания признак - вид спорт, игрови пост, националност, диагноза и др. Ако скалата има две възможни състояния, се нарича **алтернативна** (спортуващи-неспортуващи, здрави-болни, мъже-жени). Принадлежността към групата се означава с число, което е код.

- 
Ординална, при която обектите се разпределят в еднородни групи, между които има различия в интензитета на притежавания признак – оценки в българската образователна система; образователна степен (бакалавър-магистър); квалификация на състезателя (висока, средна, ниска); субективни оценъчни скали (от „категорично не“ до „категорично да“ в 5 степенна скала) и др.

- 
Рангова, при която изследваните се подреждат въз основа на определени правила (по интензитета на притежавания признак), без да има количествена информация за него – ранглисти в различните видове спорт, класиране в даден турнир, рейтингови скали и др. Резултатът от измерването е число, което показва мястото, което изследваният заема в дадена подредба и се нарича рангов номер. Характерно за тази скала е, че разликите между ранговите номера не са аналогични на разликите между величината на притежавания признак (например, разликата в класирането между 1-во и 2-ро място е аналогична на тази между 2-ро и 3-то място. Ако това са резултати в 100 м. гл. б., разликата в постиженията между 1-вия и 2-рия състезател може да е 0,05 сек., а между 2-рия и 3-тия – 0,1 сек.).

КОЛИЧЕСТВЕНИ (МЕТРИЧНИ) са скалите, при които резултатът от измерването показва величината на притежавания признак и е изразена в съответната мерна единица – м, сек, кг и др. Те биват:

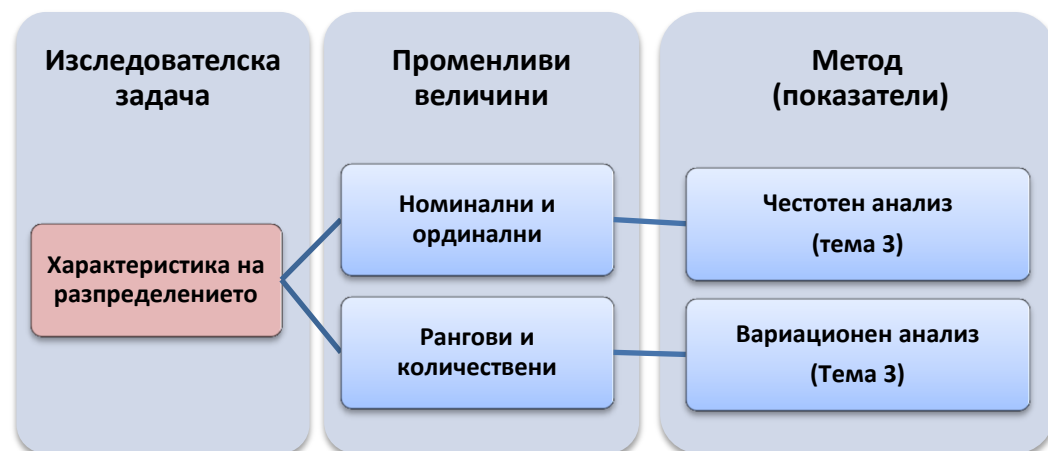
-  **Интервална** - резултатът от измерването са стойности носещи количествена информация за притежавания признак, но нулата не означава отсъствие на признака, а е условно възприета (0° по Целзий, гъвкавост на наклона на трупа напред и др.);
-  **Пропорционална** – резултатът от измерването са стойности, които отразяват величината на притежавания признак. При тях състояние 0 е абсолютно (0° по Келвин, сгъване-разгъване на ръцете от вис и др.).

Важно е да се отбележи, че скалите са подредени по способността си да долавят различия в състоянието на измерваната характеристика - качествените скали са със слаба измерителна способност, а количествените - със силна. Преминването от силна към слаба скала е възможно и се прави често при статистическа обработка. Например, ръстът може да се измери *количествено*, ако резултатът е в сантиметри, *рангово*, ако обектите се подредят по ръст и се запише само мястото на обекта в ранговата редица, *ординално*, ако обектите се разпределят в групи нисък, среден, висок ръст. Обратният преход – от слаби към силни скали не е възможен.

Ако в резултат на измерването всички изследвани имат един и същ резултат, се получава т.нар. **постоянна величина**. Обикновено това са признаците, по които е дефинирана съвкупността. Ако при измерване на статистическите единици се получават различни резултати, получената величина се нарича **променлива** (например променливата „ръст“). Променливите величини се бележат с латинските букви **X**, **Y** и т.н. и в скоби се пояснява в каква единица е измерен, например „ръст“ (X в см).

ПРЕГЛЕД НА СТАТИСТИЧЕСКИТЕ МЕТОДИ, ПРИЛАГАНИ В СПОРТА

В зависимост от изследователските задачи, които се решават и начина, по който са скалирани и разпределени променливите величини, се ползват следните статистически методи:



Фиг. 1.5

Честотният анализ се ползва за описание на състоянието на категорийни променливи (номинални и ординални) и включва описание на **броя** и **процента** от случаите, които попадат към отделните категории. Той е описан в тема 3. Пълно описание на показателите от честотния анализ се разглежда в избираемата учебна дисциплина **MPS.2.4**.

Вариационен анализ, който се ползва за характеризирание на състоянието на количествени променливи. В него се включват показателите за средно равнище, разсейване и форма на разпределението. Подробно тази проблематика е разгледана в тема 3.

Тема 1

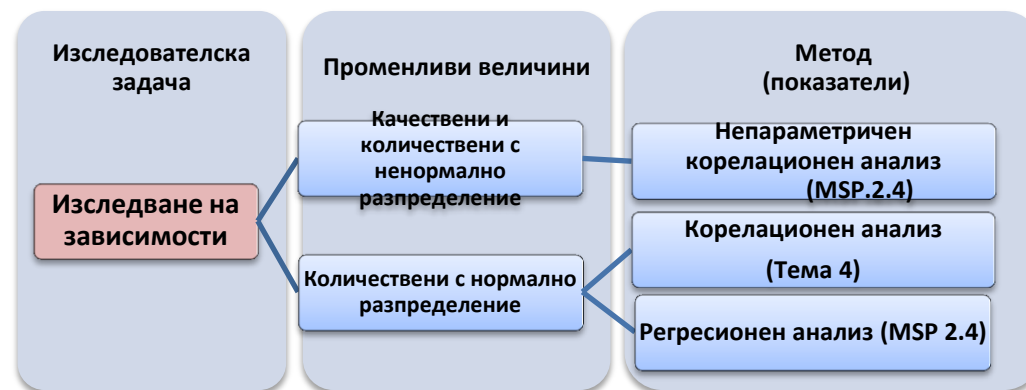
Методите за разработване на нормативи. Те се основават на състоянието на (зависимостите между) изследваните признаци и биват:

- **Персентилен метод**, който се ползва при количествени променливи с различно от нормалното разпределение;
- **Сигмален метод**, който се ползва при количествени променливи с нормално разпределение;
- Към тях трябва да добавим и **регресионния метод**, който е продължение на корелационния анализ и се основава на оценяване на състоянието на **зависимата променлива (Y)** въз основа на състоянието на **независимата променлива (X)**.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Изследване на зависимости. Условно може да се разграничат две направления:

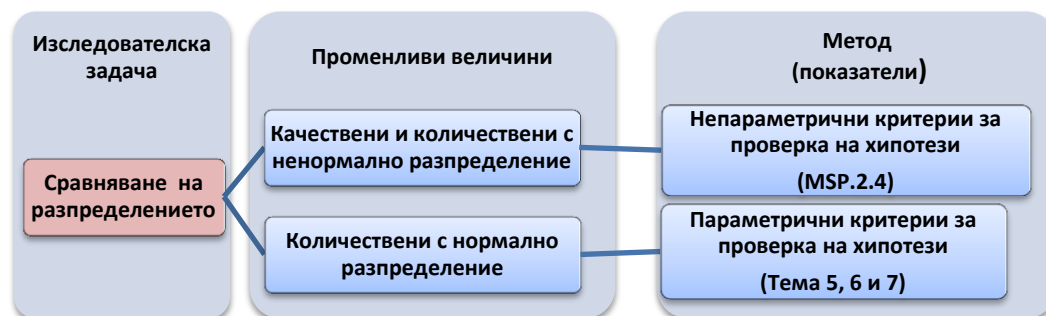
- **Корелационен анализ**, при който акцентът е поставен върху описание на *силата и посоката* на зависимостта между изследваните показатели. В зависимост от начина на скалирането им се ползват: 1) *непараметрични коефициенти на корелация* - коефициент на контингенция (C), коефициент Ф, коефициент на рангова корелация на Спирман (r_s) и 2) *параметрични коефициенти на корелация* (коефициент на Пирсън (r)). Анализът се разглежда в тема 4.



Фиг. 1.6

- За моделиране на зависимостта между количествени променливи с подходяща математическа функция се ползва **регресионният анализ**, който се изучава в избираемата учебна дисциплина **MSP. 2.4**.

Сравнителният анализ се ползва за сравняване на състоянието на променливи величини със **статистическата проверка на хипотези**. В зависимост от вида на променливите и формата на разпределение на стойностите се ползват:



Фиг. 1.7

- **непараметрични** (при качествени променливи и такива с различно от нормалното разпределение);

- **параметрични** (при количествени променливи с нормално разпределение) **критерии за проверка на хипотези**. Параметричните критерии се изучават в теми 5 и 6, а тяхното ползване за обработка на данни от експеримент с две групи и две изследвания – в тема 7. Непараметричните критерии се изучават от студентите по кинезитерапия, спортна психология и спортен мениджмънт.

ТЕМА 2. ВЪВЕДЕНИЕ В IBM SPSS STATISTICS

В настоящата тема са описани първите стъпки за опознаване на IBM SPSS Statistics



Обща характеристика

- [Първи стъпки](#)
- [Основни прозорци](#)
- [Основни менюта](#)
- [Подготовка на данните](#)
- [Дефиниране на променливи](#)
- [Въвеждане на данни](#)

Изчисляване на производни променливи

- [Изчисляване на нова променлива въз основа на съществуващи](#)
- [Трансформиране на количествена променлива в ординална](#)
- [Ранжиране на стойности](#)

Избор на случаи, за които да се отнася обработката

- [Разделяне на масива за обработка на подгрупи](#)
- [Селектиране на случаи, за които да се отнася обработката](#)

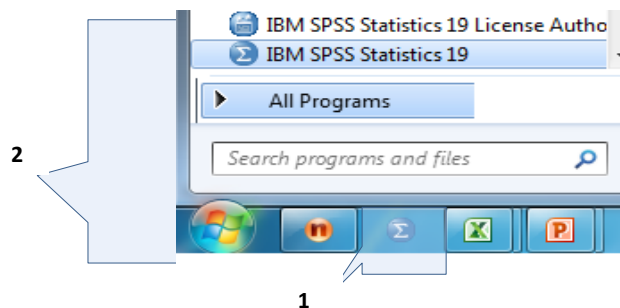


ОБЩА ХАРАКТЕРИСТИКА

IBM SPSS Statistics 19 (SPSS – статистически пакет за социални науки) е компютърна програма, която е специализирана за систематизиране, обработка и анализ на статистически данни. Както подсказва наименованието, програмата е предназначена предимно към обработка на данни от изследвания в областта на социологическите и маркетинговите проучвания, но богатото разнообразие от методи за обработка на количествени данни я прави много подходяща за приложение в областта на науката за спорта.

Стартирането на програмата става като:

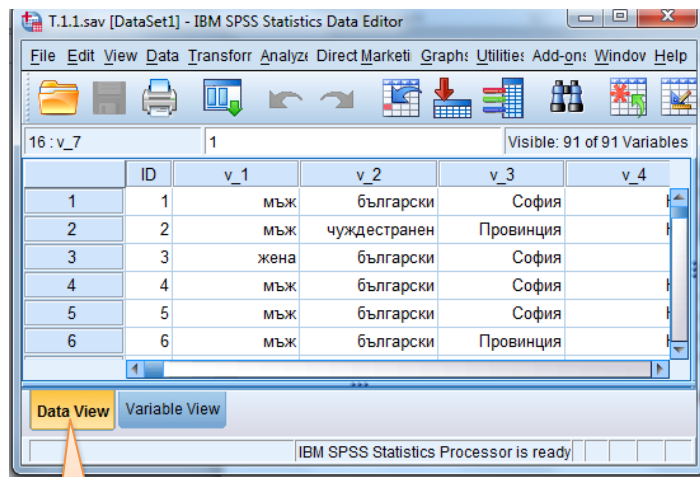
- 1) се кликне иконата за бързо избиране  или
- 2) чрез **Start>All Programs>IBM SPSS Statistics> IBM SPSS Statistics 19** (фиг. 2 1).



Фиг. 2.1

ОСНОВНИ ПРОЗОРЦИ

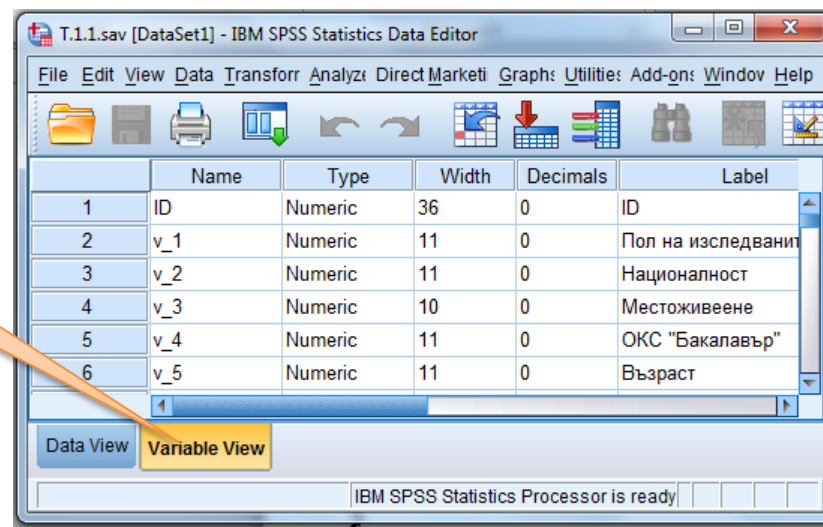
SPSS разполага с няколко основни екрани за управление на програмата. Първият от тях е за въвеждане и редактиране на данни - **IBM SPSS Statistics Data Editor**. Той има две страници – за въвеждане и редактиране на данни **Data View** (фиг. 2.2) и за дефиниране на променливи **Variable View** (фиг. 2.3). Преходът между тях се осъществява като се щракне с левия бутон на мишката върху съответния табулатор в долната лява част на екрана.



Фиг. 2.2. IBM SPSS Statistics Data Editor (Data View)

Изглед към
данните

Изглед към
променливите



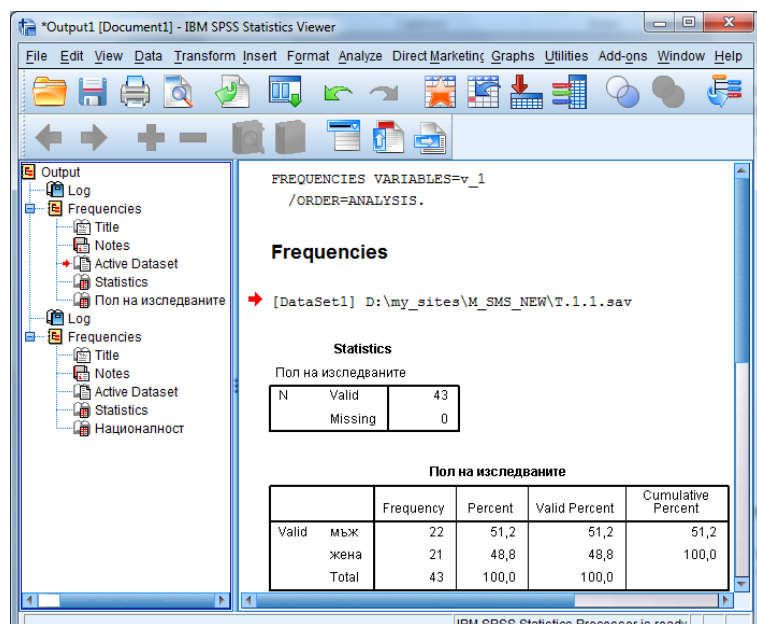
Фиг. 2.3. IBM SPSS Statistics Data Editor (Variable View)

Редакторът на данни се отваря автоматично при стартиране на програмата. Информацията, записана в него, се съхранява като самостоятелен файл с разширение **sav**.

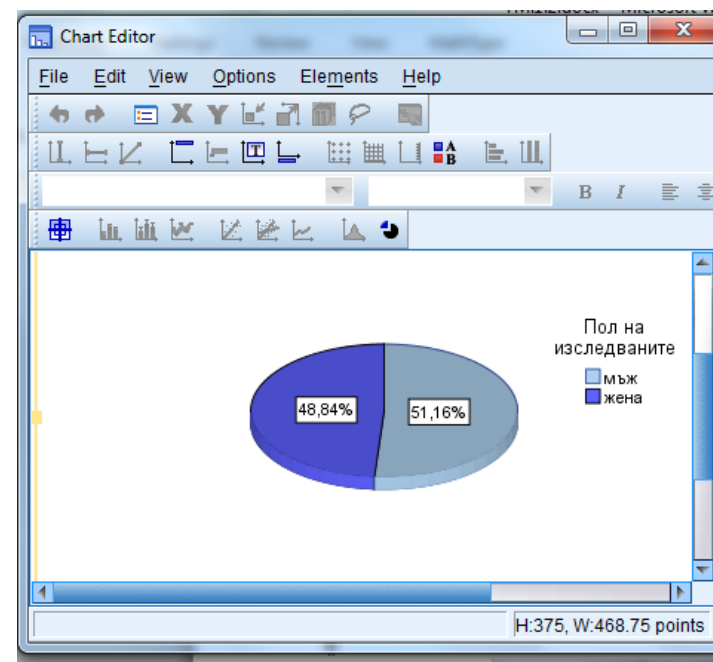
[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Вторият основен прозорец - **IBM SPSS Statistics Viewer** (фиг. 2.4) се появява след осъществяване на статистически анализ. В него се отпечатват резултатите от обработката - таблици и графики. Съдържанието му може да бъде съхранено като самостоятелен файл с разширение **.spv** или да се експортира в предпочитан от потребителя формат – **MS Word (.doc)**, **MS Excel (.xlsx)**, **Portable Document Format(.pdf)**, текстов документ (.txt) и др.

Ако дадена графика (таблица) се нуждае от редактиране, след двукратно кликване с мишката се активира редакторът на графики **Chart Editor** (фиг.2.5) или таблици.



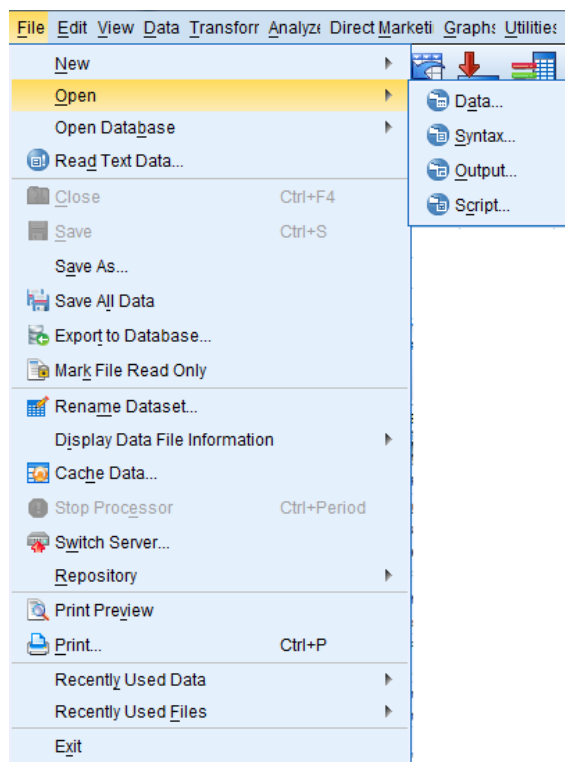
Фиг. 2.4. IBM SPSS Statistics Viewer



Фиг. 2.5. Chart Editor

ОСНОВНИ МЕНЮТА

Всеки прозорец на **SPSS** има характерна за него лента с падащи менюта. Винаги са активни **File** (фиг. 2.6), **Edit**, **Data**, **Transform**, **Analyze** (фиг.2.7) и **Graphs**. Това дава възможност за лесен достъп до заложените в тях функции. В настоящия раздел ще се спрем само на някои от основните функции за работа с файлове.



Фиг.2.6. Меню File

Меню File

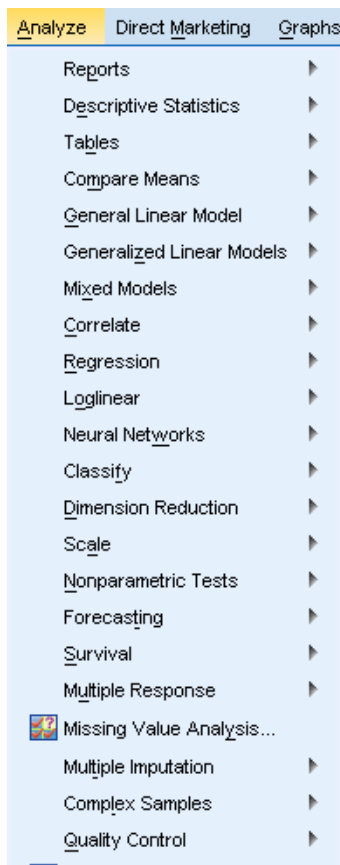
Тема 2

Менюто **File** служи за работа с файлове. Тъй като програмата поддържа файлове от различен формат (данни, резултати от обработката и др.), при създаване (**New**) и отваряне (**Open**) на файлове видът на файла трябва да бъде посочен (фиг. 2.6).

Съхраняването на файловете (**Save As**) с данни може да стане, както в собствения за SPSS формат - .sav, така и в друг предпочитан от потребителя формат.

Информация за променливите в отворения файл може да се получи, като се избере **Display Data File Information**.

Отварянето на последните файлове, с които е работено може да стане от реда **Recently Used Data**.



Фиг. 2.7. Меню Analyze

Меню Analyze

Менюто **Analyze** (фиг. 2.7) служи за задаване на статистически анализи. Най-често използваните от тях са:

- **Descriptive Statistics** (описателна статистика), в който са заложили честотен и вариационен анализ;
- **Compare Means**, в който са заложили параметричните критерии за проверка на хипотези;
- **Correlate** – служи за осъществяване на корелационен анализ с коефициентите на обикновена линейна корелация на Пирсън и коефициента на рангова корелация на Спирман;
- **Regression**, в който е заложило регресионно моделиране на зависимости;
- **Nonparametric tests**, където са заложили непараметричните критерии за проверка на хипотези;
- **Multiple Response**, който служи за обработка на въпроси с повече от един отговор.

Тема 2

Освен тези основни ленти с падащи менюта, на всеки от прозорците има и специфични за **SPSS** икони за бързо избиране. Задържането на показалеца на мишката над тях дава кратка бърза информация за функциите, които изпълнява.

ПОДГОТОВКА НА ДАННИТЕ

В резултат на провеждане на научно изследване се получава информация за състоянието на изследваните признаци (тестове, отговори на въпроси от анкета и т.н.) на статистически единици в извадката. Тези резултати, в повечето случаи, са в първични протоколи, попълнени въпросници и т.н. За да се пристъпи към въвеждане на стойностите в **IBM SPSS** е необходимо първо данните да се подготвят за въвеждане. Ще поясним как става това с резултатите от анкетно проучване, проведено сред студенти от ОКС „Магистър“ на НСА за желанието за включване в дистанционно обучение и предпочитаните от тях стилове на учене (файла [Survey_1.sav](#)).

На фиг. 2.8 е представена част от анкетната карта, която е попълнена от изследвано лице от мъжки пол, български гражданин, живеещ в София, който е завършил НСА и е на 25 години. Отговорите на първите 4 въпроса са номинално скалирани алтернативни признаци, а петият (възраст) – количествена променлива.

Тема 2

Кратко наименование	Пълно наименование	Кодове и наименование на категориите
V_1	Вашият пол е:	1 ☒ мъж 2 ☐ жена
V_2	Вашата националност:	1 ☒ български гражданин 2 ☐ чуждестранен гражданин
V_3	Постоянно местоживееене:	1 ☒ София 2 ☐ Извън София
V_4	Къде сте завършили <u>сте ОКС</u> „Бакалавър“?	1 ☒ В НСА 2 ☐ В друго висше училище
V_5	Вашата възраст:	23 години

Фиг. 2.8

В терминологията на **SPSS** условното обозначение на показателите (**V_1**, **V_2** и т.н.) се нарича **кратко наименование (Name)**, а пълното наименование (например, националност) – **етикет на променливата (Label)**. Категорийните променливи се въвеждат с **кодове (Value)**, а наименованията на категориите – **етикети на стойностите (Value Label)**.

Предварителна подготовка на данните. Процедурата преминава през следните стъпки:

	v_1	v_2	v_3	v_4
ID	Пол	Националност	Местоживееене	Бакалавърска
1	1	1	1	1
2	1	2	1	1
3	2	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	2	1
7	2	1	2	2

Фиг. 2.9. Първичните данни на хартиен носител

- Уточнява се *броят, краткото и пълното наименование* на променливите. За показатели, които имат повече от един измерител (въпрос от анкетна карта с повече от един отговор), се предвиждат толкова колони, колкото измерителя има показателят.
- Уточняват се *кодовете*, с които се обозначава принадлежността на изследваните към дадена категория за номиналните и ординални променливи (въпрос от анкетна карта (фиг. 2.8)).
- Създава се макет на таблицата на хартиен носител. В първата колона се записват поредните номера на изследваните статистически единици. Всяка следваща *колона* съответства на един изследван показател (за примера - въпрос от анкетната карта). Всеки *ред* от таблицата съответства на резултатите на едно изследвано лице, чийто пореден номер е изписан в колона ID.
- Резултатите се попълват в таблицата на хартиен носител.

Други обстоятелства, с които трябва да бъде съобразена подготовката на данните:

- Ако данните са от експеримент, в който лицата са изследвани повече от един път, резултатите от всяко следващо изследване трябва да бъде вдясно на аналогичен за изследваното лице ред.
- Ако в данните присъстват резултати на представители на различни групи (от различни градове, клубове, училища и др.), те се записват в една таблица, като с кодове се обозначава към коя група принадлежат лицата.
- Ако по даден показател изследваното лице няма резултат (по един от тестовете не е изследван, не е дал отговор на даден въпрос от анкетата), съответната клетка се оставя празна. SPSS я приема като липсваща стойност (каквато е в същност) и при всеки от анализите я третира по различен начин.

След като данните се попълнят и проверят на хартиения носител, се пристъпва към въвеждане на данните в SPSS.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)



Дефинирайте променлива с кратко наименование v_1, разширено наименование „Пол на изследваните“, в която ще се въвеждат кодове 1 и 2, като с код 1 се обозначават лицата от мъжки пол, а с 2 – от женски пол.

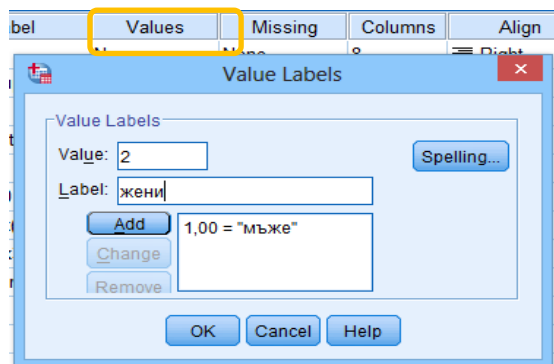
Въвеждането на данни в **SPSS** преминава през 2 етапа - първо се дефинират променливите, като по този начин се създава макетът на базата данни и след това се въвеждат стойностите.

ДЕФИНИРАНЕ НА ПРОМЕНЛИВИ В SPSS

Дефинирането на променливи се прави във **Variable View** на **Data Editor**-а (фиг. 2.3), както следва:

- В колона **Name** се записва краткото наименование на променливата (**Variable Name**). То трябва да бъде уникално, (не може две променливи да имат едно и също име), да започва с буква, като **SPSS** не прави разлика между малки и главни букви и да не съдържа празни интервали или оператори за аритметични действия (% , / , * , - , +).
- В колона **Label** се записва пълното наименование на променливата (**Variable Label**).
- В колона **Values** се задават кодовете (**Value**) и етикетите (**Value Label**) на категориите променливи. Това става, като се активира клетката в колона **Values** на реда на съответната променлива и се кликне върху появилия се син бутон .
- В отворения се диалогов прозорец се задава (фиг. 2.10):

Тема 2

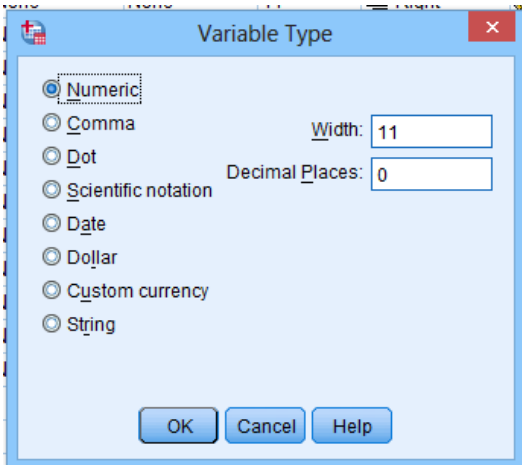


На реда **Value** – кодът, на реда **Label** – наименованието на категорията (етикет) и се натиска бутонът **Add**. Процедурата продължава докато се въведат кодовете и етикетите на всички възможни категории. След това се потвърждава с **OK**.

Тази процедура не се прилага за количествени променливи!

Фиг. 2.10. Задаване на кодове и етикети на категориите

В колона **Type** (фиг. 2.11) може да се уточни дали въвежданите резултати са числови стойности (**Number**), текст (**String**), време или дата (**Date**) и др.



Фиг. 2.11

В колона **Measure** се задава измерителната скала, в която са изразени променливите. За целта от падащото меню се избира между (фиг. 2.12):

- **Scale** - количествена,
- **Ordinal** – ординална,
- **Nominal** - номинална.

Name	Type	Width	Decimals	Label	Values	Missing	Colu...	Align	Measure
v_3	Numeric	10	0	Местоживее	{1, Co...	None	6	Right	Nominal
v_4	Numeric	11	0	ОКС "Бакалавър"	{1, HC...	None	4	Right	Ordinal
v_5	Numeric	11	0	Възраст	None	None	5	Right	Scale

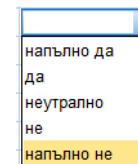
Фиг. 2.12. Задаване на измерителната скала

След като се дефинират променливите и се изгради т.нар. макет на таблицата с данни, се преминава към **Data View**, за да се въведат стойностите.

ВЪВЕЖДАНЕ НА ДАННИ В IBM SPSS

Въвеждането на данни се осъществява се в прозореца **Data View** (фиг. 2.2). Стойностите се изписват с клавиатурата. Първоначално въведеното число се появява на реда за въвеждане над прозореца за данни. След като се потвърди с **Enter**, числото се изписва в активната клетка.

Стойностите на категориите променливи се въвеждат със зададените им кодове, а потребителят може да задава дали в таблицата да се появяват кодовете или техните етикети. Това става, като се натисне иконата за етикети – **Value Labels** (фиг. 2.13). Въвеждането на категориите променливи може да стане и от падащия списък на категориите



При задържане на показалеца на мишката върху краткото име се появява пълното наименование на променливата, което е много удобно при придвижване в големи масиви от данни.

ID	v_1	v_2	v_3	v_4	v_5
1	мъж	български	София	HCA	23
2	мъж	чуждестранен	Провинция	HCA	25
3	жена	български	София	2	26
4	мъж	български	София	HCA	31
5	мъж	български	София	HCA	24
6	мъж	български	Провинция	HCA	25



Value Labels

v_8

	ID	v_1	v_2	v_3	v_4	v_5	
1	1	1	1	1	1	23	
2	2	Пол на изследваните			2	1	25
3	3	2	1	1	2	26	
4	4	1	1	1	1	31	
5	5	1	1	1	1	24	
6	6	1	1	2	1	25	

Фиг. 2.13. Състояние на редактора на данни в зависимост от иконата Value Labels

ОТКРИВАНЕ НА ГРЕШКИ ОТ ВЪВЕЖДАНЕТО

Често при измерване на резултатите или въвеждане на стойностите се допускат грешки, които, ако не бъдат открити навреме, може силно да се отразят на резултатите от осъществявания анализ. Наличието на подобни грешки може да бъде открито по някой от следните начини:

- На първо място може да се направи **логически анализ на данните**, като се прецени възможно ли е даденият признак да заеме такава стойност. Например, в таблицата с данни се открива тегло на дете 508 кг. Логично е да се предположи, че е пропусната десетичната запетая.
- Възможност за експресна и нагледна информация за проверка на количествени променливи дават **хистограмата и бокс-плот диаграмата**. Особено показателна в това отношение е втората диаграма. Познавателните и възможности се разглеждат в тема 3.
- При изследване на зависимости трябва предварително да се построи диаграма на разсейване и визуално да се открият **стойности, които влияят на заключението на зависимостта**. Подробно тези случаи се разглеждат в избираемия модул MSP.2.4.
- **Грешки при изчисляване на производни показатели** могат да се избегнат, като при изчисляването им в **SPSS**, за пълно наименование на променливата се зададе функцията по която са изчислени (виж следващия раздел). Важно е да се знае, че в случай, че се открие грешка в някоя първична променлива, изчисляването от нея производни показатели трябва да се преизчислят.

ОТСТРАНЯВАНЕ НА ГРЕШКИ ОТ ВЪВЕЖДАНЕТО

- Погрешно въведена **стойност** се коригира, като се активира съответната клетка и се изпише правилната стойност.
- Погрешно създадена **колона** се изтрива като се маркира колоната (а не само стойностите в нея) и се натисне клавиша на клавиатурата **Delete** или се натисне десният бутон на мишката и се избере **Clear**. Това може да се направи и във **Variable View**.
- Погрешно създаден **ред** се премахва, като се маркира реда и се изтрие с клавиша на клавиатурата **Delete** или се кликне с десния бутон на мишката и се избере **Clear**.
- Погрешно **изчислена производна променлива** трябва да бъде изтрита и след това да се преизчисли отново по начина, описан в следващия раздел.

Често при обработка на данни се налага преобразуване на оригиналните данни:

- При обработка на данни от педагогически експеримент е необходимо да се изчисли прирастът на резултатите на изследваните лица като разлика между второто и първото изследване (виж тема 7).
- При обработка на данни от анкетни проучвания понякога се налага трансформиране на количествена променлива (напр. възраст) в ординална (млади, средна възраст, стари).
- В случай на констатирани отклонения на стойностите от нормалното разпределение би могло съответните променливи да се преобразуват в рангови и върху тях да бъдат приложени параметрични статистически методи.

Ако данните са въведени първо в **Excel**, изчисляването на производни показатели може се извърши там. Ако данните са въведени директно в **SPSS** е необходимо, преди да се пристъпи към статистическа обработка, да се осъществят тези преобразувания. Преобразуване на данни се задава от менюто **Transform**.

Най-често използваните трансформации са:

- изчисляване на нови променливи въз основа на съществуващи,
- трансформиране на количествена променлива в ординална;
- преобразуване на количествена в рангова.

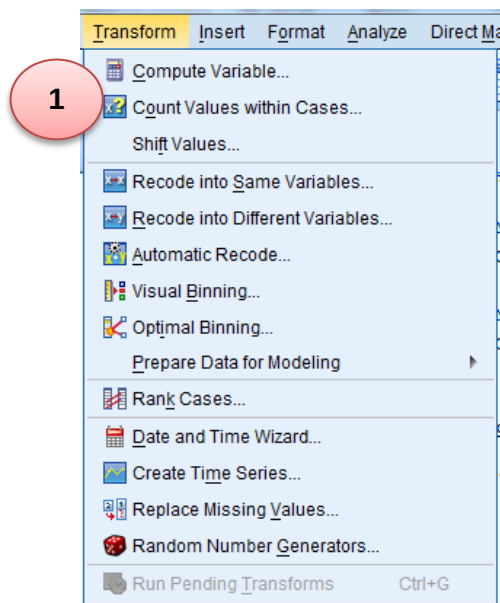
В следващите раздели са описани алгоритмите за тяхното осъществяване.

ИЗЧИСЛЯВАНЕ НА НОВИ ПРОМЕНЛИВИ, ВЪВ ОСНОВА НА СЪЩЕСТВУВАЩИ

По данните от [Survey 1.sav](#)

Да се изчислят точките за визуален стил по формулата
 $2*(v_19_6 + v_19_10 + v_19_12 + v_19_24 + v_19_29)$

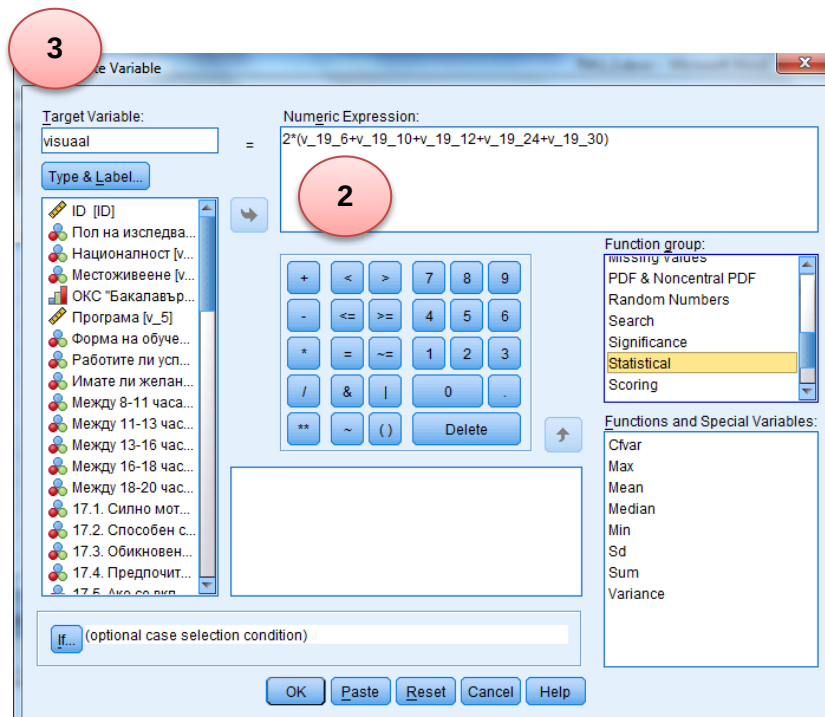
За изчисляване на производни променливи трябва:



1 Да се избере менюто **Transform > Compute** (фиг. 2.14). След неговото избиране се отваря диалоговият прозорец **Compute Variable** (фиг. 2.15).

Фиг. 2.14. Меню Transform

Тема 2

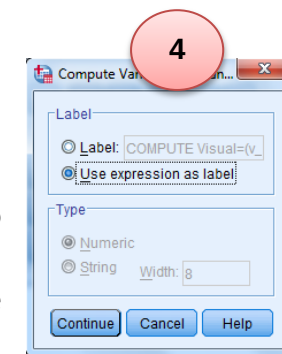


2 С мишката се избират променливите, които участват в изчислителната процедура и с бутона  се прехвърлят в полето **Numeric Expression**. С клавиатурата или с предвидените за целта бутони се задава функцията, по която се изчислява производната променлива (в случая е зададено изчисляване на резултата за предпочитание към *визуален стил* на обучение, като удвоена сума от оценките от айтеми № 6, 10, 12, 24 и 29 ($v_{19_6}, v_{19_{10}}, v_{19_{12}}, v_{19_{24}}$ и $v_{19_{29}}$).

Тема 2

Фиг. 2.15. Изчисляване на производни променливи със SPSS

- 3 Когато изразът е готов, в полето **Target Variable** се записва името на новата променлива (в случая **Visual**).
- 4 След като се кликне бутонът **Type&Label**, може да се дефинира пълното име на променливата или изразът, по който тя е получена. С оглед лесното откриване на грешки в зададената функция е уместно, при задаване на етикета на променливата, да се активира **Use expression as label**. По този начин, като пълно наименование на променливата ще се появи формулата, по която е изчислена (фиг. 2.16).



Фиг. 2. 16. Задаване на етикета на производната променлива

Когато се потвърди с **OK** в таблицата с данни се появява новата променлива.

ТРАНСФОРМИРАНЕ НА КОЛИЧЕСТВЕНА ПРОМЕНЛИВА В ОРДИНАЛНА

По данните от [Survey 1.sav](#)

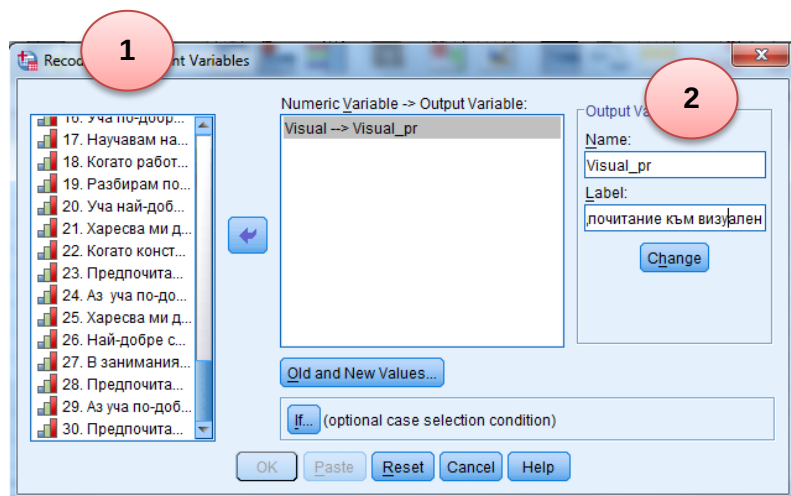
Да се оцени предпочитанието към визуален стил по следната скала:

Граници	Код	Етикет
До 24 т.	1	пренебрегнат
От 25 до 37 т.	2	второстепенен
Над 37 т.	3	основен

Тема 2

За трансформиране на количествена променлива в ординална се ползва менюто **Transform>Recode into Different Variables** (фиг. 2.17). Процедурата се осъществява в 2 стъпки:

① В диалоговия прозорец **Recode into Different Variables** се избира променливата, която ще бъде трансформирана.



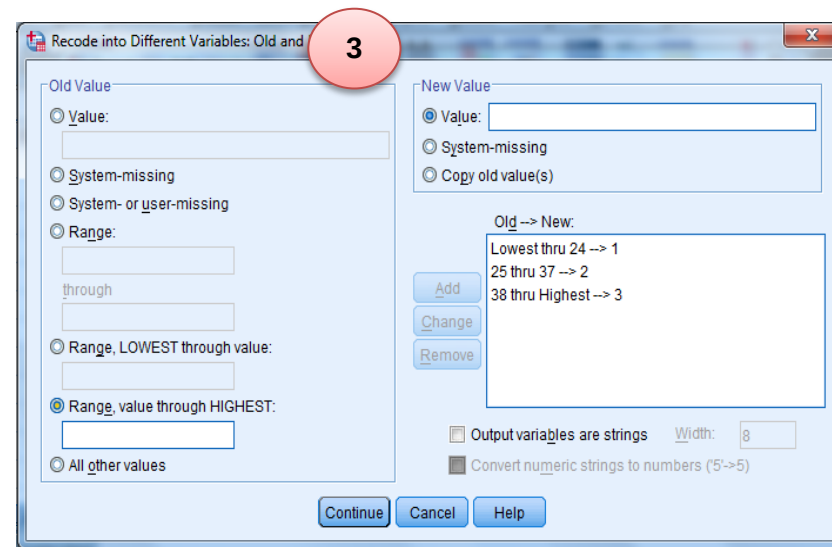
② В полето **Output variable Name** се задава краткото, а в **Label** - пълното наименование на новата променлива. Кликва се върху бутона **Change** и се избира **Old and New Values**.

Фиг. 2.17. Задаване на името и етикета на преобразуваната променлива

В отворения се диалогов прозорец **Old and New Values** (фиг. 2.18) има обособени две части – **лява**, в която се изписват *старите стойности (Old Values)* и **дясна**, в която се изписва *новата стойност (New Values)*.

3 Последователно се избира:

- каква е старата стойност:
 - конкретно число – **Value**,
 - липсваща стойност – **Missing**,
 - по-малка от – **range, LOWEST through value**,
 - между – **Range**,
 - по-голяма от – **Range, value through HIGHEST**, всички останали – **All other values**)
- с каква нова да бъде заменена (**Value**) и се потвърждава с **Add**.



Фиг. 2.18. Задаване на начина на трансформиране на променливата

Тема 2

В полето **Old>New** се появява съответствието между първичните и трансформирани стойности. Процедурата продължава докато се зададат всички граници и кодове. След това се потвърждава с **Continue>OK**.

В резултат на осъществената процедура в таблицата с данни се появява новата променлива. Ако тя е номинална или ординална, трябва да се дефинират кодовете и етикетите на категориите.

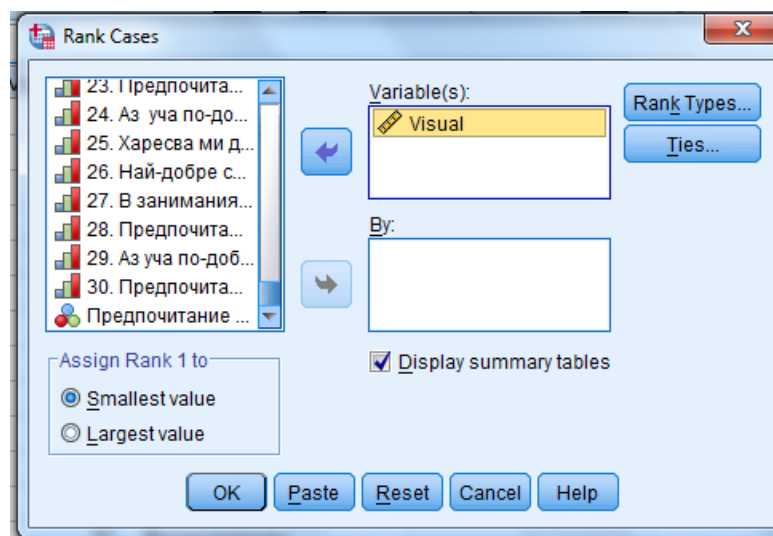
ТРАНСФОРМИРАНЕ НА КОЛИЧЕСТВЕНА ПРОМЕНЛИВА В РАНГОВА

По данните от [Survey_1.sav](#)

Да се ранжират изследваните по точките за
визуален стил.

За осъществяване на това преобразуване се ползва менюто **Transform Rank Cases** (фиг. 2.19). Променливата, която трябва да се ранжира се изпраща в полето **Variable(s)**. От полето **Assign Rank 1 to** се избира дали стойностите да се подредят във възходящ порядък (**Smallest value**) или низходящ (**Largest value**) порядък. Този избор е важен, защото влияе на посоката на зависимостите, в които участват производните променливи. След потвърждаване с **OK**, в таблицата с данни се появява нова колона с кратко име по подразбиране **Риме на изходната променлива** (в случая с ранжиране на стойностите по променливата **Visual** новата колона е **RVisual**).

Тема 2



Фиг.2.19. Задаване на променливата, която да се ранжира

В резултат на показаните преобразувания, в таблицата с данни се появяват променливите, показани на фиг. 2.20.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Важно е да се знае, че:

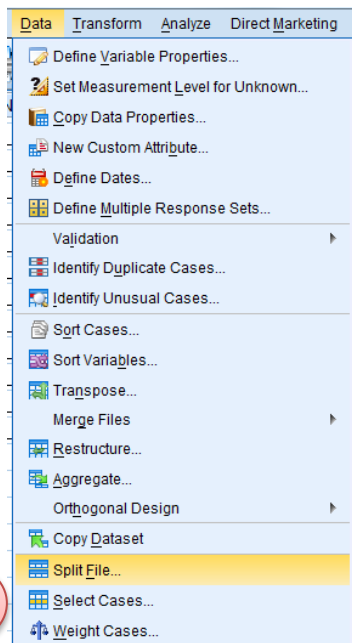
- **SPSS** прилага преобразуване само за случаите, за които в първичните променливи *няма липсващи стойности*. На първите два реда, при ползваното сортиране на стойностите, са лица, които не са дали отговор по някои от измерителите на предпочитанието към визуален стил на обучение. Поради това, в колона **Visual**, съответните им клетки са празни. Резултатите на тези лица не участват и в следващите трансформации. В колона **Visual_pr** е изписана оценката за степента на предпочитание към този стил.
- Друга важна особеност на трансформирането на променливи със **SPSS** е, че при ранжиране на стойностите, ранговите *номера за еднакви изходни резултати се осредняват*. По примера, лицата, които имат по 24 точки, в колона **RVisual** имат еднакъв рангов номер - 2,5, което е осреднена стойност на 2-то и 3-то място, което заемат.

	Visual	Visual_pr	RVisual
1	.	.	.
2	.	.	.
3	22	Маловажен	1,0
4	24	Маловажен	2,5
5	24	Маловажен	2,5
6	28	Второстепенен	6,5
7	28	Второстепенен	6,5
8	28	Второстепенен	6,5

Фиг. 2.20. Резултат от преобразуванията, описани в раздела

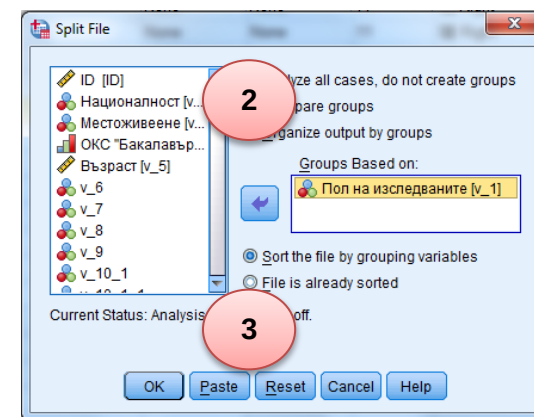
РАЗДЕЛЯНЕ НА ФАЙЛА НА ГРУПИ ЗА САМОСТОЯТЕЛНА ОБРАБОТКА

Във файловете с данни обикновено има различни групи от изследвани (мъже-жени, експериментална-контролна група и др.), чийто данни трябва да се обработват самостоятелно. Възможности за подобно селектиране има заложени в менюто **Data** (фиг. 2.21).



1 След като се избере **Data>Split File** се появява диалоговият прозорец, показан на фиг. 2.22. По подразбиране е зададено в анализа да участват всички стойности (**Analyze all cases, do not create groups**). За да се обработват поотделно зададени от потребителя групи трябва да се избере:

По данните от [Survey_1.sav](#)
Да зададе осъществяване на анализ отделно за
мъже и жени.



Тема 2

Фиг. 2.22. Разделяне на файла на подгрупи за самостоятелна обработка

2 **Compare Groups** (ако резултатите от обработката на групите трябва да се появяват в една таблица) или **Organize Output by groups** (ако резултатите от обработката трябва да се появяват в отделни таблици).

3 От списъка на променливите вляво трябва да се посочи/ат променливата/ите, по коя(и)то да се раздели обработката, да се изпрати(ят) в полет **Groups Based on** и да се потвърди с **OK**. Следващият статистическият анализ ще се осъществи поотделно за всяка от групите. Ако след това трябва да се премахне тази опция, се избира **Data>Split File>Analyze All Cases, do not create groups**.

ИЗБИРАНЕ НА СЛУЧАИ, ЗА КОИТО ДА СЕ ОТНАСЯ ОБРАБОТКАТА

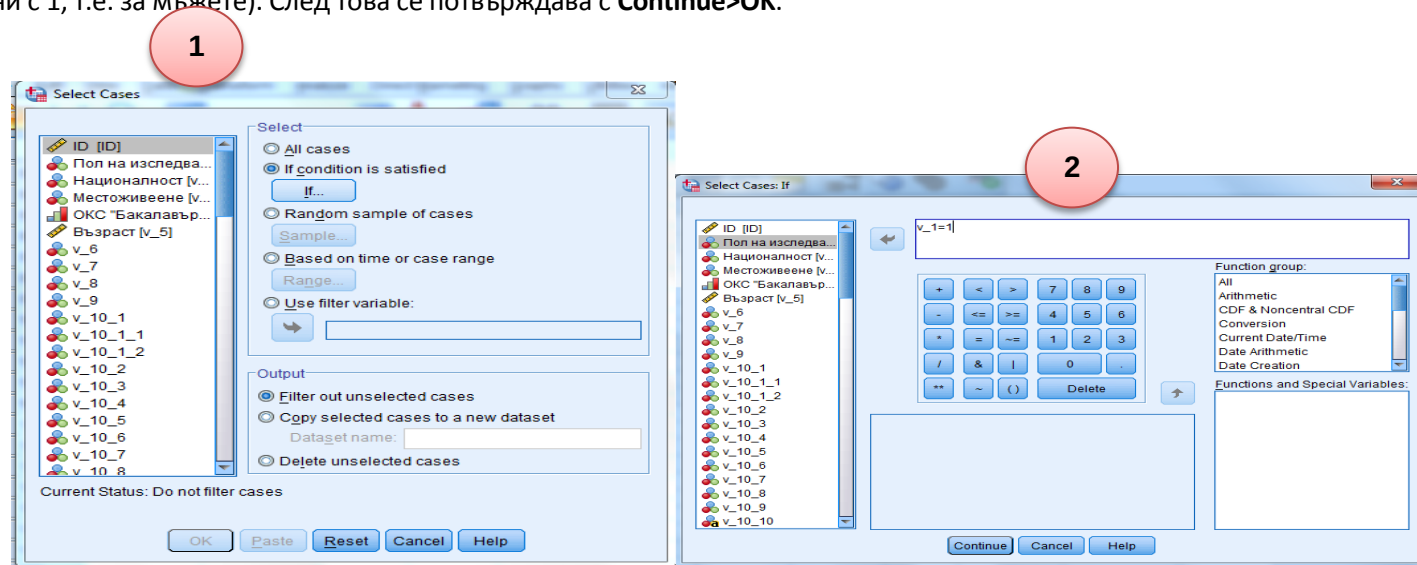
По данните от [Survey 1.sav](#)

Да се зададе осъществяване на анализ само за мъжете.

Процедурата стартира, като се избере **Data>Select Cases**. Появява диалоговият прозорец, показан на фиг. 2.23, в който по подразбиране е заложено да се обработват всички стойности.

- 1 За да се посочи за кои случаи да се отнася анализът се избира **If condition is satisfied** и се натиска бутона **If...**
- 2 В отворения се диалогов прозорец се избира променливата, в която са кодирани случаите и се изпраща се в полето горе-вдясно. След това изписва функцията, на която трябва да отговарят случаите (на фиг. 2.23 е зададено анализът да се отнася за лицата, които по променливата **пол** са кодирани с 1, т.е. за мъжете). След това се потвърждава с **Continue>OK**.

Тема 2



Фиг. 2.23. Селектиране на случаи за обработка

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

В резултат на осъществената процедура най-вдясно на променливите се появява нова колона **filter_\$** (фиг. 2.24). Както се вижда, мъжете са отбелязани като избрани, а в означенията на редовете данните за жените са зачертани.

	ID	v_1	filter_\$
1	1	мъж	Selected
2	2	мъж	Selected
3	3	жена	Not Selected
4	4	мъж	Selected
5	5	мъж	Selected
6	6	мъж	Selected
7	7	жена	Not Selected
8	8	жена	Not Selected

Фиг. 2.24. Резултат от селектиране на случаи

Освен описаните, в менюто **Data** са предвидени възможности за:

- подреждане на статистическите единици по величината на зададена от потребителя променлива (**Data>Sort Cases**)]
- подреждане на променливите величини във файла (**Data>Sort Variables**);
- обединяване на файлове (**Data>Merge Files**) чрез добавяне на случаи или променливи величини и др.

ТЕМА 3. ХАРАКТЕРИСТИКА НА РАЗПРЕДЕЛЕНИЕТО НА ПРОМЕНЛИВИ ВЕЛИЧИНИ



Изследвани са 91 млади футболисти, които са с различна националност, възраст и игрови пост.

Приложени са 3 теста:

- за бързина -30 м. гл. б.,
- за анаеробна мощност (совалково бягане 6x35 м.)
- за аеробна издръжливост (бийп тест).

Необходимо е да се установи:

- как са разпределени изследваните по националност, възрастова група и игрови пост;
- какво е състоянието на изследваните физически качества.

Честотен анализ

- [Показатели](#)
- [Едномерно разпределение на честотите със SPSS](#)
- [Двумерно разпределение на честотите със SPSS](#)
- [Задачи за упражнение](#)

Вариационен анализ

- [Предназначение](#)
 - [Показатели за средно равнище](#)
 - [Показатели за разсейване](#)
 - [Показатели за форма на разпределението](#)
- [Вариационен анализ със SPSS](#)
 - [Резултат от обработката](#)
 - [Задачи за упражнение](#)



ЧЕСТОТЕН АНАЛИЗ

В резултат на измерването на номинални и ординални признаци се получава информация за принадлежността на изследваните към дадена категория. Обобщението за извадката става като се опишат категориите, които се срещат, броят и процентът от случаите, които попадат към тях, т.е. тяхната **абсолютна (f)** или **относителна честота (W)**.

- Изследваните статистически единици може да се разпределят само по една номинална (ординална) променлива и тогава разпределението се нарича **едномерно**.
- Ако се разпределят по две или повече променливи се получава т.нар. **двумерно** (многомерно) разпределение на честотите.

В таблица 3.1 е представено едномерно разпределение на честотите на изследваните футболисти по игрови пост (данните са от файла [soccer_juniors.sav](#)).

Тема 3

Таблица 3.1

Категории	Брой	%
Нападатели	19	20,88
Среда на терена	36	39,56
Защитници	33	36,26
Вратари	3	3,30
Общо	91	100

Абсолютната честота (f) показва какъв е броят на футболистите по игрови постове. Така например 19 от изследваните са нападатели. .

Относителната честота (W) показва каква част (ако е умножена по 100 – какъв процент) от общия брой случаи са от даден игрови пост

$$W = \frac{f}{n} \quad (3.1)$$

Където:
f - абсолютна честота
n - обем на извадката

По данните от примера – нападателите са 20,88% от изследваните .

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Относителна кумулативна честота (W_{cum}) показва каква част (%) от общия брой случаи попадат в съответната и предходните категории, взети заедно (например нападателите и играчите от средата на терена са 60,4% от изследваните).

Дву (много) мерно разпределение на честотите означава, че изследваните са разпределени по две или повече категорийни променливи. Получените по този начин честотни разпределения се наричат **кростаблици**. Статистическите показатели са същите, както описаните при едномерно разпределение – абсолютна (f) и относителна честота (W).

В таблица 3.2 е представено разпределението на изследваните футболисти по игрови пост и възрастова група.

Таблица 3.2

	16-годишни		18-годишни		Общо	
	Брой	%	Брой	%	Брой	%
Нападатели	13	20,31	6	22,22	19	29,69
Среда на терена	26	40,63	10	37,04	36	56,25
Защитници	23	35,94	10	37,04	33	51,56
Вратари	2	3,13	1	3,70	3	4,69
Общо	64	100	27	100	91	100

Резултатите от обработката показват, че нападателите са 20,31% от 16 годишните и 22,22% от 18-годишните футболисти.

Честотен анализ може да се направи и за *рангови и количествени променливи*, но информацията не е обобщена по начин, който да дава възможност за ясно описание на разпределението на изучаваната променлива.

За графично представяне на честотите при едномерно разпределение на номинални променливи се ползват **кръговата и стълбовидната диаграма**.



ЧЕСТОТЕН АНАЛИЗ С IBM SPSS STATISTICS

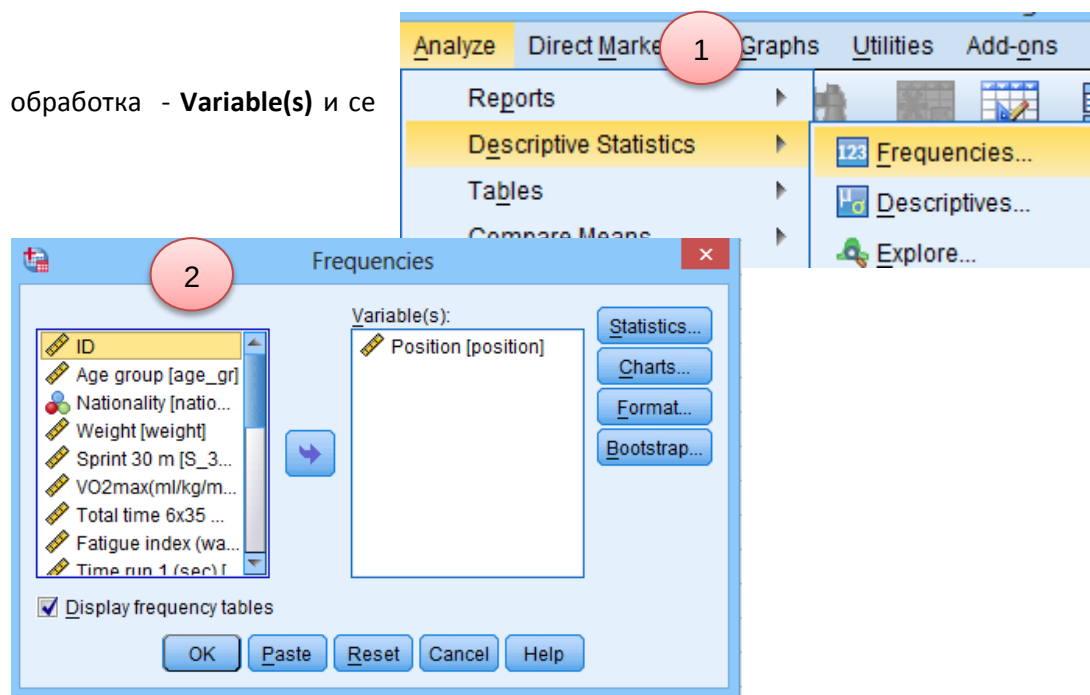
ЕДНОМЕРНО РАЗПРЕДЕЛЕНИЕ НА ЧЕСТОТИТЕ

По данните [от Soccer_juniors.sav](#)

Да се направи едномерно разпределение на изследваните по игрови пост (променливата **position**).

Осъществяването на анализа става, като се избере:

- 1 **Analyze> Descriptive statistics>Frequencies** (фиг. 3.1).
- 2 Анализираната променлива се изпраща в полето за обработка - **Variable(s)** и се потвърждава с **OK**.



Фиг. 3.1. Задаване на едномерно разпределение на честотите

Тема 3

РЕЗУЛТАТ ОТ ОБРАБОТКАТА

В резултат на осъществената процедура се появява разпечатката, представена на фиг. 3.2.

Statistics

Position

N	Valid	91
	Missing	0

Position

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	forwards	19	20,9	20,9	20,9
	midfielders	36	39,6	39,6	60,4
	defenders	33	36,3	36,3	96,7
	goalkeepers	3	3,3	3,3	100,0
	Total	91	100,0	100,0	

Фиг. 3.2. Резултат от осъществяване на честотен анализ

В таблица **Statistics** се появява информация за наличните случаи (**valid**), в случая $n=91$ и липсващите стойности (**Missing**), в няма такива.

В следващата таблица е направено едномерно разпределение на честотите, както следва:

- в първата колона са изписани възможните категории (в случая игрови постове);
- във втората колона (**Frequencies**) – броят от случаите, които попадат към дадена категория;
- в третата колона (**Percent**) е изписано процентното разпределение на стойностите, като е предвиден и процент за липсващите стойности (в случая няма такива).

Тема 3

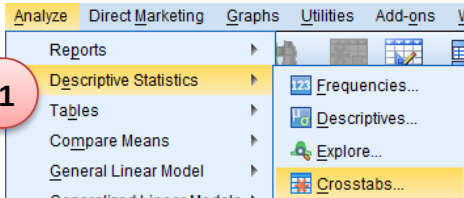
- В колона **Valid Percent** е изписано процентното разпределение на стойностите, като са изключени липсващите стойности;
- В колона **Cumulative Percent** са изписани кумулативните честоти, изчислени на базата на валидния процент от случаите.

От резултата се вижда, че 19 от изследваните футболисти са нападатели и те представляват 20,9% от изследваните случаи.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ДВУМЕРНО РАЗПРЕДЕЛЕНИЕ НА ЧЕСТОТИТЕ

Анализът се осъществява, като се избере:

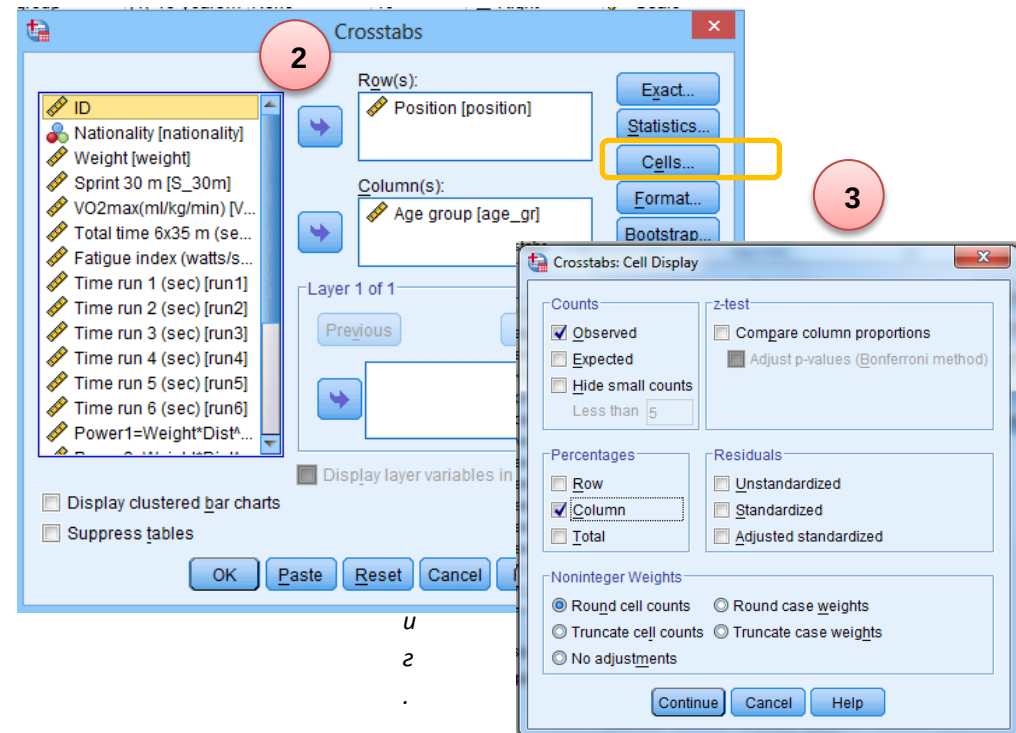


1 **Analyze>Descriptive statistics>Crosstabs**, в резултат на което се отваря диалоговият прозорец **Crosstab**, където се задава:

2 В полето **Rows** (по редове) - едната променлива (в случая игровият пост), а в полето **Columns** (по колони) – другата променлива, в случая – възрастовата група.

3 В случай, че е необходимо изчисляване на процентните разпределения се натиска бутонът **Cells**. В отворения се диалогов прозорец **Crosstabs: Cell Display** се задава изчисляването на допълнителни статистически показатели, които да включва кростаблицата. Най-често се ползват показателите от полето **Percentages**, с които се задава изчисляването на процентни разпределения. Процедурата се изпълнява след като се избере **Continue>OK**.

По данните от **Soccer_juniors.sav**
Да се направи двумерно разпределение на честотите игрови пост (**position**) и възрастова група (**age groups**).



Тема 3

3.3. Задаване на кростабулация

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

РЕЗУЛТАТИ ОТ ОБРАБОТКАТА

В резултат на осъществената процедура в **SPSS Statistics Viewer**-а се появява резултатът от обработката, представен на фиг. 3.4

			Age group		Total
			16 years	18 years	
Position	forwards	Count	13	6	19
		% within Age group	20,3%	22,2%	20,9%
	midfielders	Count	26	10	36
		% within Age group	40,6%	37,0%	39,6%
	defenders	Count	23	10	33
		% within Age group	35,9%	37,0%	36,3%
	goalkeepers	Count	2	1	3
		% within Age group	3,1%	3,7%	3,3%
Total	Count	64	27	91	
	% within Age group	100,0%	100,0%	100,0%	

Фиг. 3.4. Резултат от кростабулация

Редактираният вариант на таблицата е представен в табл. 3.2. Зададеното изчисляване на процентите **по колони** означава, че 20,3% от изследваните 16-годишни футболисти са нападатели, 40,6% - състезатели от средата на терена, 35,9% - защитници и 3,1% - вратари.



ЗАДАЧИ ЗА УПРАЖНЕНИЕ

ДАННИ ЗА УПРАЖНЕНИЕТО:

[Soccer_juniors.sav](#)

[Калкулатор](#)

Задача 3.1

По данните от файла **Soccer_juniors.sav** направете едномерно разпределение на честотите по променливата националност (**nationality**).

Националност	Брой	Процент
Български		
Чуждестранни		
Общо		

Анализирайте резултатите.

Задача 3.2

По данните от файла **Soccer_juniors.sav** направете двумерно разпределение на честотите променливите игрови пост (**position**) и националност (**nationality**).

Игрови пост	Националност	Български	Чуждестранни	Общо
Нападатели				
Среда на терена				
Защитници				
Вратари				
Общо				

Анализирайте резултатите.

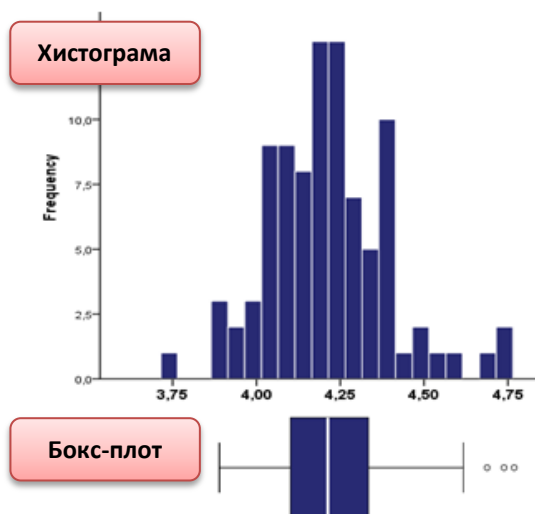
[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)



ВАРИАЦИОНЕН АНАЛИЗ

Вариационният анализ се прилага, за да се характеризира разпределението на количествени променливи. Той включва три групи от показатели - за средно равнище, разсейване на стойностите и формата на разпределението. За да поясним смисъла на тези показатели ще ползваме две графични изображения, на които са онагледени резултатите в тест 30 м. гл. б. по данните от файла [Soccer_juniors.sav](#).

Хистограмата е стълбовидна диаграма, по абсцисата на която са нанесени стойностите на X , а в полето са нанесени стълбчета, чиято височина онагледява **абсолютната (f)** или **относителна (W)** честота на стойностите. По разположението на стълбчетата на фигурата се вижда, че постиженията в тест 30 м. гл. б. варират между 3,75 сек. и 4,75 сек. С най-висока честота на резултатите около 4,2 сек.



Фиг. 3.5

Бокс-плот диаграмата онагледява 5-те обобщаващи позиционното разпределение на стойностите показатели, а именно $X_{\min}$, $X_{\max}$, P_{25} , P_{50} , и P_{75} . Графичният образ е „кутия“ с „дръжки“. „Дръжките“ онагледяват минималната и максимална стойност. Централната част, наподобяваща кутия, онагледява диапазона, в който варират централните 50% от случаите. Бялата линия в средата на „кутията“ онагледява средата на подредените стойности (P_{50}). Графичното изображение дава възможност да се открият и силно отклоняващите се стойности:

- Ако даден резултат е извън интервала $P_{25}-1,5*IRQ$ и $P_{75}+1,5*IRQ$ се обозначава с кръгче и се приема, че се отклонява от основната маса от числа, но в приемливи граници. На фигурата има 3 такива стойности;
- Ако даден резултат е извън интервала $P_{25}-3*IRQ$ и $P_{75}+3*IRQ$, тя се обозначава със звездичка. В такива случаи се приема, че отклонението на дадената стойност от останалите е неприемливо.

С понятието **вариационен ред** се обозначава фактът, че стойностите са подредени от 1 до n във възходящ или низходящ порядък.

Със символа P_i се бележат **персентилите**. Те са 99 стойности, които делят вариационния ред на 100 равни части от по 1%. P_{25} се намира на 25-тия процент от вариационния ред, P_{50} – на 50-тия, а P_{75} – на 75-тия процент от вариационния ред.

Разликата между P_{75} и P_{25} се нарича **междуквартилен размах (IQR)**.

Тема 3

ПОКАЗАТЕЛИ ЗА СРЕДНО РАВНИЩЕ

Те описват **типичното състояние** на признака – неговият център. Те са:

Мода (Mo) - стойността, която се среща най-често във вариационния ред.

Медиана (Me) - стойността, която заема средно положение във вариационния ред – под и над нея има еднакъв брой стойности. Тя е равна на P_{50} и е онагледена с бялата черта на бокс-плот диаграмата. За да се изчисли медианата е необходимо да се построи вариационен ред, да се изчисли поредният номер на медианата във вариационния ред (N_{me}) по формулата:

$$N_{me} = \frac{n + 1}{2} \quad (3.2)$$

След това трябва да се отчете стойността, която заема N_{me} място във вариационния ред. Ако обемът на извадката (n) е нечетен брой, в средата на подредените числа има само една стойност и тя е **Me**. Ако n е четно число, в средата на реда има 2 стойности. **Me** е тяхна средна стойност.

Средна аритметична величина ($\bar{X}$), която по данни от извадката се получава, като сумата на стойностите се раздели на техния брой (n).

$$\bar{X} = \frac{\sum X}{n} \quad (3.3)$$

В изчисляването на $\bar{X}$ участват всички стойности на променливата. Поради това е най-прецизният измерител на средното равнище. В същото време, при наличие на силно отклоняващи се стойности, $\bar{X}$ може да не дава вярна представа за центъра на разпределението.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ПРИМЕРИ

В резултат на изследването са получени следните резултати:

Пример 1:

X: 5, 6, 6, 7, 7, 7, 8, 9, 9

$n=9$

$Mo=7$

$N_{me}=(9+1)/2=5, Me=7$

$\bar{X} = \frac{64}{9} = 7,11$

Пример 2:

X: 5, 7, 7, 7, 8, 8, 8, 9

Две съседни стойности се повтарят еднакъв максимален брой пъти – Модата е тяхната полусума
 $Mo=(7+8)/2=7,5$

Обемът на извадката е четно число, поради което медианата попада между двете централни стойности.

$N_{me}=(8+1)/2=4,5, Me=(7+8)/2=7,5$

Пример 3:

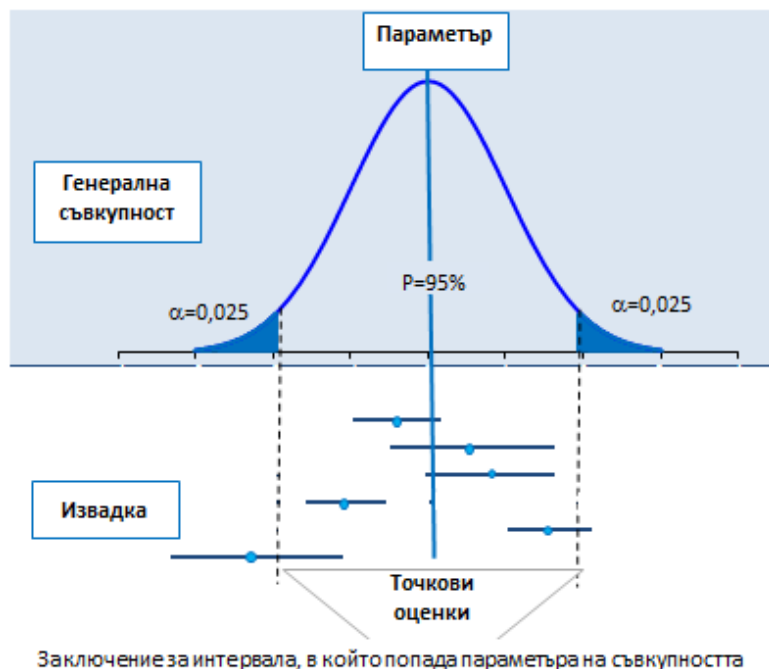
X: 5, 6, 6, 6, 7, 8, 9, 9, 9,

Две стойности, които не са съседни се повтарят еднакъв максимален брой пъти – разпределението има 2 моди – $Mo_1=6$ и $Mo_2=9$.

Тема 3

ИНТЕРВАЛНО ОЦЕЯВАНЕ НА СРЕДНОТО РАВНИЩЕ

Модата, медианата и средната стойност са **точкови оценки** за средно равнище, т.е. те описват типичното



Фиг. 3.7. Интервално оценяване на μ

състояние на признака в извадката.

Параметърът на генералната съвкупност се бележи с μ . Границите на интервала, в който попада средната на генералната съвкупност (μ) се намират по формулата:

$$\bar{X} \pm t(U)SE_{\bar{X}} \quad (3.4)$$

Където:

- $SE_{\bar{X}} = \frac{S}{\sqrt{n}}$ се нарича репрезентативна (стандартна грешка);
- Числата **t** и **U** се подбират в зависимост от обема на извадката (**n**) и гаранционната вероятност (**P**), с която се прави заключението за генералната съвкупност.

Пример

По данни от извадката с обем $n=64$ е установено, че $\bar{X} = 20$ и $S=2$. Да се намерят границите на интервала, в който попада средната на съвкупността (μ) с гаранционна вероятност **P=95%**.

Репрезентативната грешка е равна на:

$$SE_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{2}{\sqrt{64}} = 0.25$$

При **P=95%** **U=1,96** (прил. 1)

Границите на интервала, в който попада средната на съвкупността са:

$$X' = 20 - 1.96 \cdot 0.25 = 20 - 0.49 = 19.51$$

$$X'' = 20 + 1.96 \cdot 0.25 = 20 + 0.49 = 20.49$$

Тема 3

Числото U се ползва при големи извадки ($n > 20$) и се отчита от [приложение 2](#), като при **P=95%** **U=1,96**, а при **P=99%** **U=2,58**

Числото t се ползва при малки извадки ($n < 20$). Отчита се от [приложение 3](#) в зависимост от степените на свобода **df=n-1** и зададената гаранционна вероятност (**P**).

ПОКАЗАТЕЛИ ЗА РАЗСЕЙВАНЕ

Показателите за разсейване описват **различията** между стойностите.

Размахът (R) описва диапазона, в който варират стойностите (техния обхват). Визуално размахът се онагледява чрез разликата между „дръжките“ на бокс-плот диаграмата. Изчислява се по формула:

$$R = X_{max} - X_{min} \quad (3.5)$$

Стандартното отклонение (S) описва големината на отклоненията от средната аритметична величина.

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}} \quad (3.6)$$

Стандартното отклонение (S) е прецизен измерител на разсейването, защото в изчисляването му участват всички стойности на **X**. **S** се изразява в мерните единици на променливата, поради което не може да се ползва за сравняване на разсейването на променливи, изразени в различни мерни единици или относно различна средна стойност.

Коефициентът на вариация (V) описва разсейването на стойностите около средната аритметична величина, изразено като процент от нея.

$$V = \frac{S}{\bar{X}} \cdot 100 \quad (3.7)$$

Коефициентът на вариация (**V**) дава възможност за сравняване на разсейването на различни променливи. По стойността му могат да се правят заключения дали изследваната извадка е хомогенна. Ако:

- $V < 10\%$ извадката е еднородна (малко разсейване),
- V е между 10 и 30% - извадката е приблизително еднородна (средно разсейване),
- V е над 30% - извадката е силно нееднородна (разсейването е голямо).

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Примери

В резултат на изследването са получени следните резултати:

i	1	2	3	4	5	6	7	8	9
X_i	5	6	6	7	7	7	8	8	9
$(X_i - \bar{X})$	-2	-1	-1	0	0	0	1	1	2
$(X_i - \bar{X})^2$	-2 ²	-1 ²	-1 ²	0 ²	0 ²	0 ²	-1 ²	-1 ²	-2 ²
$\sum (X_i - \bar{X})^2$	4+1+1+0+0+0+1+1+4=12								

$$R = 9 - 5 = 4$$

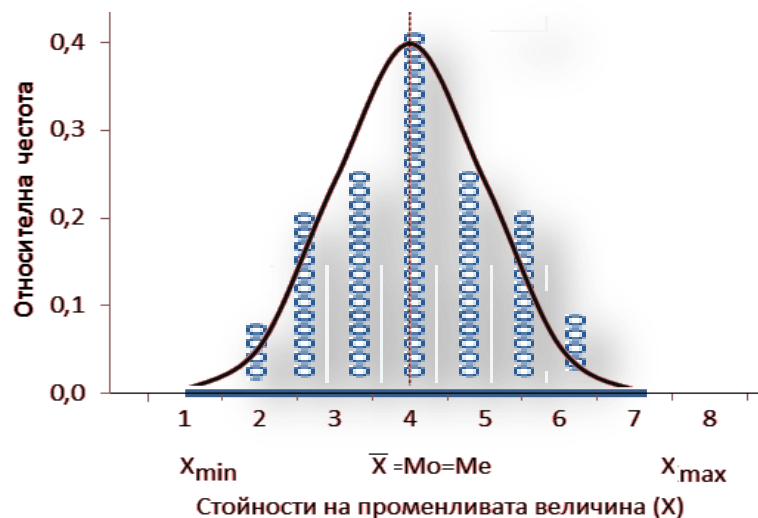
$$S = \sqrt{\frac{12}{8}} = 1,2$$

$$V = \frac{1,2}{7} \cdot 100 = 17\%$$

Тема 3

ФОРМА НА РАЗПРЕДЕЛЕНИЕТО

Третата важна познавателна задача на вариационния анализ е да се установи дали емпиричното разпределение е аналогично с теоретичното нормалното разпределение.



Фиг. 3.8

Идеалното нормално разпределение има асиметрия **As=0**. При емпиричните разпределения, стойността на **As** обикновено се различава от нула. Границите на допустимите отклонения при гаранционна вероятност **P=95%** са **$0 \pm 2 * SE_{As}$** , където **SE_{as}** е стандартната грешка на коефициента, която се получава по формулата:

$$SE_{As} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \quad (3.9)$$

Ако изчисленият по данни от извадката **As** е извън тези граници се приема, че разпределението е несиметрично.

ХАРАКТЕРИСТИКИ НА НОРМАЛНОТО РАЗПРЕДЕЛЕНИЕ:

Теоретичното нормално разпределение е едновърхово и има външен вид на напречен разрез на камбана.

Стойностите в центъра на разпределението имат най-голяма честота, като същата намалява с приближаване към крайните стойности.

Кривата е **симетрична** относно центъра на разпределението.

Индикатор за симетричността на емпиричното разпределение е равенството между **Me**, **Mo** и **$\bar{X}$** .

Измерител на симетричността на разпределението е **коефициентът на асиметрия (As)**:

$$As = \frac{\frac{\sum (x_i - \bar{x})^3}{n}}{S^3} \quad (3.8)$$

Граничната стойност на **As** при даден обем на извадката е табулирана в [приложение 1](#).

За нормалното разпределение е характерно, че стойностите, които са в центъра на разпределението имат най-висока честота, съответно на това се оформя и върхът на разпределението. Той трябва да бъде аналогичен на този на нормалното разпределение. Височината на върха се описва с **коефициента на ексцес (Ex)**, който се изчислява по формулата:

$$Ex = \frac{\frac{\sum (x_i - \bar{x})^4}{n}}{S^4} - 3 \quad (3.10)$$

Идеалното нормално разпределение има **Ex=0**. Емпирични разпределения, които имат по-висок от нормалния връх - **Ex>0**. Ако стойностите в центъра на разпределението са по-малко, отколкото е характерно за нормалното разпределение - **Ex<0**. Допустимите граници за емпиричния **Ex** се определят по формулата: $0 \pm 2 * SE_{Ex}$, където:

$$SE_{Ex} = SE_{As} \sqrt{\frac{n^2 - 1}{(n - 3)(n + 5)}} \quad (3.11)$$

Граничната стойност, получена по посочената формула е табулирана в [прил. 1](#). Ако абсолютната стойност на емпиричния **Ex** е по-голяма от посочената в приложението критична стойност се приема, че разпределението има различен от нормалния ексцес.

ПРИМЕР

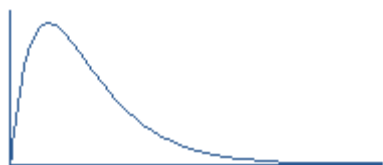
Ако е изследвана извадка с обем $n=40$ и емпиричните стойности на коефициентите са съответно:

$$As < 0,748 \text{ и } Ex < 1,465$$

се прави заключение, че изучаваната променлива има нормално разпределение.

РАЗПРЕДЕЛЕНИЯ С РАЗЛИЧНИ СТОЙНОСТИ НА КОЕФИЦИЕНТИТЕ НА АСИМЕТРИЯ И ЕКСЦЕС

Асиметрия



$$As > 0$$

$$Mo < Me < \bar{X}$$

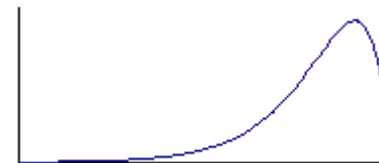
Разпределението е с: Положителна асиметрия
Дясно изтеглено рамо
Ляво изтеглен връх



$$As = 0$$

$$Mo = Me = \bar{X}$$

Разпределението е симетрично

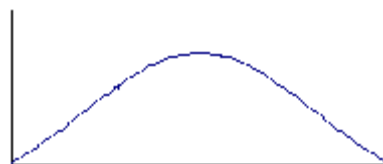


$$As < 0$$

$$\bar{X} < Me < Mo$$

Разпределението е с: Отрицателна асиметрия
Ляво изтеглено рамо
Дясно изтеглен връх

Ексцес



$$Ex < 0$$

Разпределението е с понижен връх
(platykurtic)



$$Ex = 0$$

Разпределението е с нормален връх
(mesokurtic)



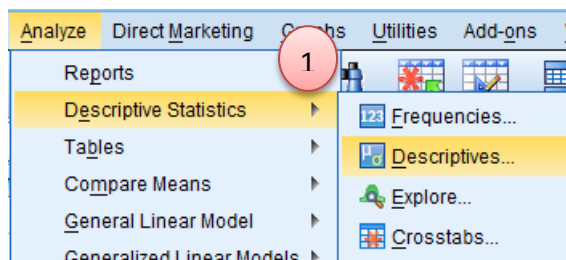
$$Ex > 0$$

Разпределението е с повишен ексцес
(leptokurtic)

Тема 3

Фиг. 3.9. Разпределения с различни стойности на коефициентите на асиметрия и ексцес

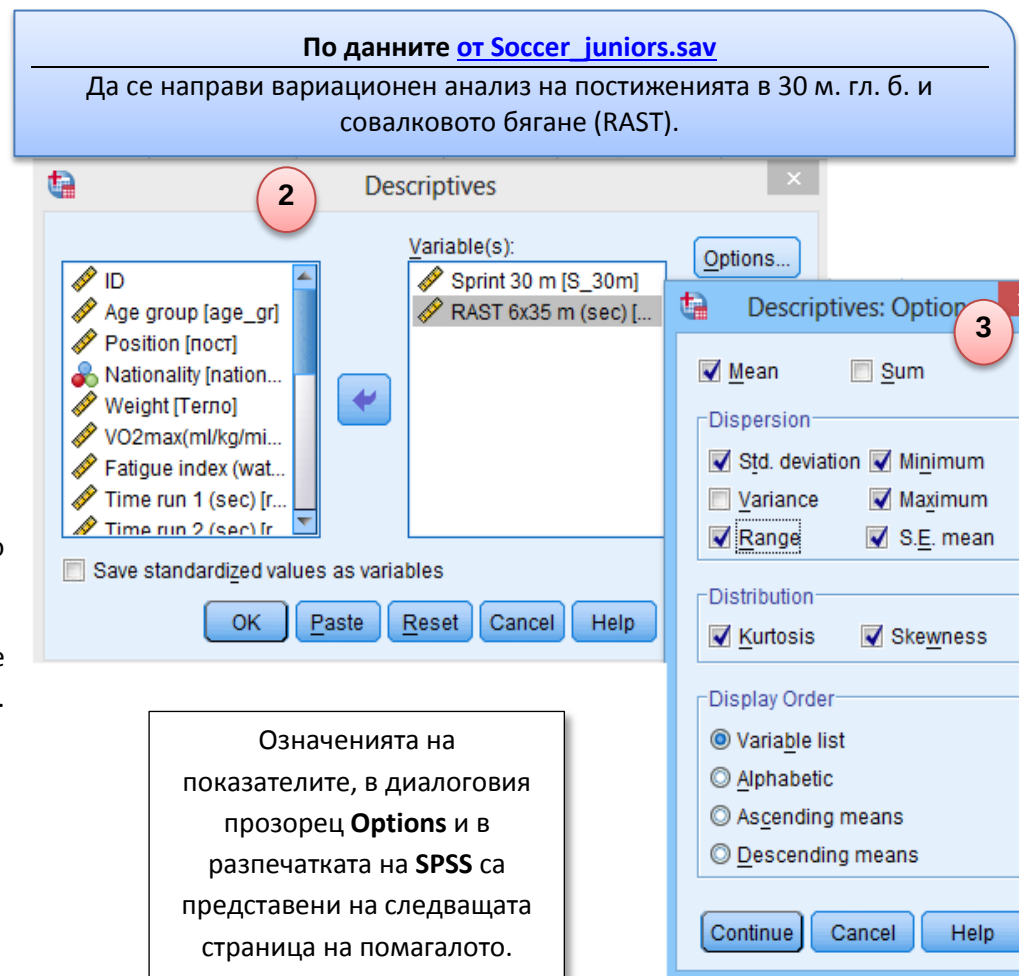
1 За осъществяване на вариационен анализ се избира менюто **Analyze>Descriptive Statistics>Descriptives**.



Отваря се диалоговият прозорец **Descriptives**.

2 Променливите, които се обработват се прехвърлят с полето **Variables** и се натиска бутона **Options**.

3 От отворения се диалогов прозорец **Descriptives: Options** се задават статистическите показатели, които да включва разпечатката. Анализът се изпълнява, като се потвърди с **Continue>OK**.



Фиг. 3.10

ОЗНАЧЕНИЕ НА ПОКАЗАТЕЛИТЕ ОТ ВАРИАЦИОННИЯ АНАЛИЗ ПРИ ЗАДАВАНЕ НА АНАЛИЗА И В РАЗПЕЧАТКИТЕ НА SPSS

Таблица 3.3

Означение при задаване	Означение в разпечатката	Наименование на показателите в разпечатката
Central Tendency		Средно равнище
Mean	Mean (Statistic)	Средна стойност
Median		Медиана
Mode		Мода
Dispersion		Дисперсия
Std. Deviation	Std. Deviation (Statistic)	Стандартно отклонение
Variance	Variance	Дисперсия
Range	Range (Statistic)	Размах
Minimum	Minimum (Statistic)	Минимална стойност
Maximum	Maximum (Statistic)	Максимална стойност
S.E. mean	Mean (Std. Error)	Стандартна грешка на средната стойност
Distribution		Разпределение
Kurtosis	Kurtosis (Statistic)	Емпирична стойност на коефициента на ексцес
-	Kurtosis (Std. Error)	Стандартна грешка на коефициента на ексцес
Skewness	Skewness (Statistic)	Емпирична стойност на коефициента на асиметрия
-	Skewness (Std. Error)	Стандартна грешка на коефициента на асиметрия

Тема 3

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

РЕЗУЛТАТИ ОТ ВАРИАЦИОНЕН АНАЛИЗ СЪС SPSS

В резултат на осъществената процедура в **SPSS Statistics Viewer**-а се появява следната разпечатка (фиг. 3.11):

Descriptive Statistics											
	N	Range	Minimum	Maximum	Mean		Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
Sprint 30 m	91	1,01	3,74	4,75	4,2156	,01909	,18211	,447	,253	,832	,500
RAST 6x35 m (sec)	91	8,09	28,08	36,17	31,6109	,16921	1,61418	,701	,253	,680	,500
Valid N (listwise)	91										

Фиг. 3.11. Разпечатка от вариационен анализ

Програмата не изчислява стойността на коефициента на вариация, поради което той трябва да бъде изчислен допълнително и резултатите от вариационния анализ да се представят по следния начин:

ПРЕДСТАВЯНЕ НА РЕЗУЛТАТИТЕ

Таблица 3.4

	n	X min	Xmax	R	$\bar{X}$	$SE_{\bar{X}}$	S	V	As	Ex
30 м. гл. б.	91	3,74	4,75	1,01	4,216	0,019	0,18	4,32	0,447	0,832
Совалково (6x35)	91	28,08	36,17	8,09	31,611	0,169	1,61	5,11	0,701*	0,68

Забележка: При $n \approx 90$ и $\alpha = 0,05$ критичните стойности на коефициентите на асиметрия и ексцес са съответно: $As_{0,05} = 0,508$ $Ex_{0,05} = 1,006$. ([прил.1](#))

Примерен анализ: Резултатите от вариационния анализ показват, че постиженията в 30 м. гл. б. варират между 3,74 сек. и 4,75 сек. Средната стойност е 4,22 сек., като коефициентът на вариация показва, че разсейването на стойностите е малко ($V = 4,32\%$). Коефициентите на асиметрия и ексцес са под критичните стойности, което означава, че стойностите имат нормално разпределение. Известна положителна асиметрия се наблюдава по отношение на постиженията в совалковото бягане, където емпиричната стойност на коефициента надвишава критичната.



ЗАДАЧИ ЗА УПРАЖНЕНИЕ

ДАННИ ЗА УПРАЖНЕНИЕТО:

[soccer_juniors.sav](#)

[Калкулатор](#)

Задача 3.3

По данните от файла **Soccer_juniors.sav** направете вариационен анализ на две променливи по избор.

	n	X min	Xmax	R	$\bar{X}$	$SE_{\bar{X}}$	S	V	As	Ex

Критичните стойности на коефициентите на асиметрия и ексцес при $n = \dots\dots\dots$ и $\alpha = 0,05$ са $As_{0,05} =$ $Ex_{0,05} =$ ([прил. 1](#))

Анализирайте резултатите.

Тема 3

Задача 3.4

По данните от файла **Soccer_juniors.sav** разделете файла на групи за самостоятелна обработка по променливата възраст и направете вариационен анализ на резултатите на една променлива по избор.

	n	X min	Xmax	R	$\bar{X}$	$SE_{\bar{X}}$	S	V	As	Ex
16-годишни										
18-годишни										

Критичните стойности на коефициентите на асиметрия и ексцес при $n = \dots\dots\dots$ и $\alpha = 0,05$ са $As_{0,05} =$ $Ex_{0,05} =$ ([прил. 1](#))

Коментирайте резултатите от обработката.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ТЕМА 4. КОРЕЛАЦИОНЕН АНАЛИЗ

В настоящата тема се разглежда начинът, по който се проучва силата и посоката на зависимостите между променливи величини с **IBM SPSS Statistics**.

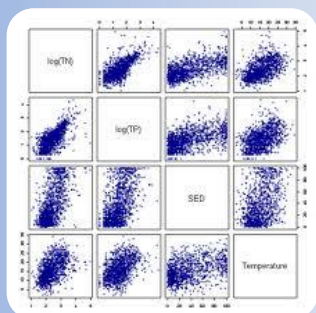
Теоретичен материал

- [Предназначение](#)
- [Основни характеристики на зависимостите](#)
- [Диаграма на разсейване](#)
- [Коефициенти на корелация](#)
- [Коефициент на детерминация](#)
- [Коефициент на обикновена линейна корелация на Пирсън \(\$r\$ \)](#)
- [Коефициент на рангова корелация на Спирман \(\$r_s\$ \)](#)

Корелационен анализ с IBM SPSS Statistics

- [Построяване на диаграма на разсейване](#)
- [Изчисляване на коефициента на Пирсън \(\$r\$ \)](#)
- [Изчисляване на коефициента на рангова корелация на Спирман \(\$r_s\$ \)](#)

[Задачи за упражнение](#)





ПРЕДНАЗНАЧЕНИЕ

Едно от най-важните направления на приложение на статистическите методи във физическото възпитание и спорта е посветено на изучаване на взаимната връзка между двигателни качества, физическо развитие, игрова дейност, изясняване на силата и посоката на влияние на различни фактори върху спортното постижение. За целта се прилагат статистическите методи за изследване на зависимости.

С понятието **зависимост** се означава връзката между променливи величини. Факторът, чието влияние се изследва се нарича **независима променлива (X)**, а изучаваната променлива (например спортният резултат) се нарича **зависима променлива** и се бележи с **Y**.



Фиг. 4.1

На фигура 4.1 е представена графика, на която са изобразени резултати от изследване на взривната сила на долните крайници (вертикален отскок) и бързината (20 м. гл. б.) на 12 младежки футболни отбора ($n=12$). С точки са означени резултатите в двата теста на изследваните отбори.

От диаграмата ясно личи, че отборите, които имат по-високи скоростно-силови възможности имат по-високо развитие на бързината, т.е. между тези променливи величини съществува зависимост.

Отборът, чиято точка е тук има среден вертикален отскок 75 см. и средно постижение в тест 20 м. гл. б. – 2,95 сек.

Тема 4

ОСНОВНИ ХАРАКТЕРИСТИКИ НА ЗАВИСИМОСТИТЕ

Основните характеристики на зависимостите са вид, форма и сила (фиг. 4.2).



Фиг. 4.2

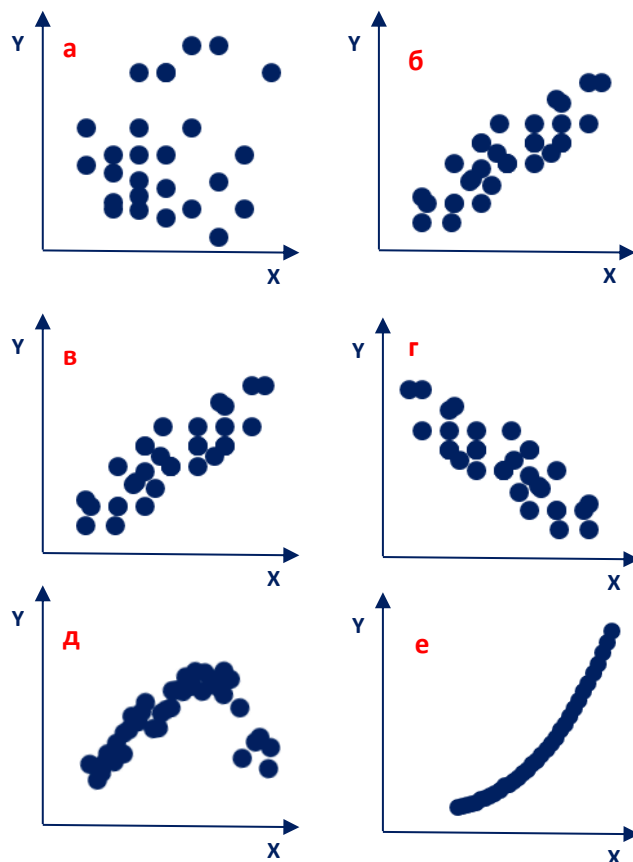
Характеристиката **ВИД** описва броя на променливите, които включва зависимостта. Ако се изследва връзката между две променливи (**X** и **Y**), зависимостта е **обикновена (simple, bivariate)** по вид. Ако се изследва връзката между повече от две променливи, едната от които е зависима, а останалите – независими променливи, зависимостта е **множествена**.

Характеристиката **ФОРМА** се отнася за математическия модел на зависимостта. Ако нарастването на **X**, във всеки един от диапазоните на нейната вариация, води до еднаква промяна на **Y** - зависимостта е **линейна**. Графиката на такива зависимости е права линия. Ако нарастването на **X** в различни диапазони от нейната вариация води до различни промени на **Y**, функцията, която описва зависимостта е криволинейна.

Тема 4

По **СИЛА** зависимостите биват **функционални** и **корелационни**. Ако върху **зависимата променлива (Y)** влияе единствено **независимата променлива X**, чието влияние се изследва, зависимостта е **пълна (функционална)**. Ако **зависимата променлива (Y)**, освен от **X**, зависи и от влияние на други фактори, които не са изследвани, зависимостта е **непълна (корелационна)**.

ДИАГРАМА НА РАЗСЕЙВАНЕ



Много ясно наличието на зависимост и нейните характеристики се установява след като се построи т. нар. **диаграма на разсейване**.

Тя представлява правоъгълна координатна система, в полето на която са нанесени точки с координати $X_i:Y_i$ (резултатите по двете променливи на изследваните статистически единици). Получилата се конфигурация от точки се нарича **облак на разсейване** и дава възможност визуално да се отговори на някои важни въпроса:

Има ли зависимост между променливите величини:

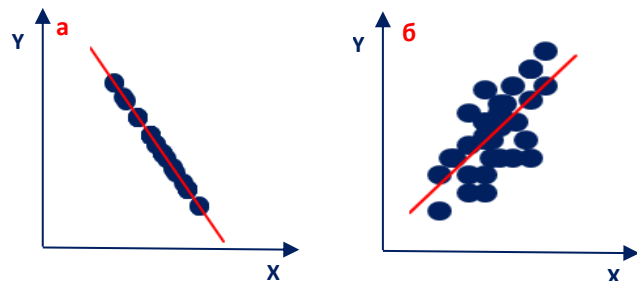
- Ако точките са безразборно разпръснати в полето на координатната система, (фиг.4.3.а) се прави заключение, че **няма зависимост** между изследваните величини.
- Ако облакът на разсейване показва, че промяната на X е свързана с промяна на Y , се прави извод, че **зависимост съществува** (фиг .4.3.б, в, г,д, е).

Каква е зависимостта по форма:

- Ако облакът на разсейване наподобява елипса (фиг. 4.3.б, в и г) се прави заключение, че зависимостта е **линейна**. Като такава, тя би могла да бъде **възходяща** (б, в) и **низходяща** (г).
- Ако конфигурацията от точки има изкривяване (д и е) – зависимостта е **нелинейна**.

Фиг. 4.3. Анализ на диаграмата на разсейване

Колко силна е зависимостта:



Фиг. 4.4. Функционална и корелационна зависимост

Колкото по-близо са точките от диаграмата на разсейване до графиката на функцията, толкова по-тясно са свързани двете променливи и толкова по-точен моделът на връзката между тях.

- Ако точките се намират на графиката на функцията (фиг. 4.4.а), зависимостта е **функционална** (пълна). При такава зависимост Y зависи напълно от X .
- Ако върху Y влияят и други фактори, те водят до отклонения на точките (т. нар. фактически стойности) от графиката на функцията (фиг. 4.4.б). Такава зависимост се нарича **корелационна** (непълна).

Тема 4

След като се построи диаграма на разсейване и се установи, че зависимост съществува, може да се пристъпи към:

- Описание на **силата** на зависимостта. За тази цел се ползва **корелационният анализ**, описан в настоящото помагало.
- Ако желанието на изследователя е да се моделира математически връзката между променливите, се прилага **регресионният анализ**. Той се изучава в избираемия модул MSP.2.4.

ВИДОВЕ КОЕФИЦИЕНТИ НА КОРЕЛАЦИЯ

Корелационният анализ служи за описание на **силата** и **посоката** на зависимост между променливи величини. Статистическите показатели, които носят тази информация се наричат **коефициенти на корелация**. В зависимост от вида, формата на зависимостта и измерителната скала, в която са изразени променливите величини се ползват различни коефициенти на корелация (табл. 4.1).

Таблица 4.1. Видове коефициенти на корелация

Скала	Променлива Y				
					
Променлива X		Коефициент Φ	Контингенция (C)	Гама (G)	Рангово бисериален (r_{rb})
			Контингенция (C)	Гама (G)	Коефициент η (ета) по ранжирани стойности
				Гама (G)	Рангова корелация на Спирман (r_s)
					Рангова корелация на Спирман (r_s)
					Обикновена линейна корелация на Пирсън (r) R на криволинейна корелация R на множествена корелация

Сила на зависимостта

Абсолютната стойност на коефициентите на корелация е в границите от 0 до 1. **Степента** (силата на зависимост) се оценява по следната скала:

При стойност 0 - липсва зависимост;
 До 0,3 – зависимостта е слаба;
 От 0,3 до 0,5 – умерена;
 от 0,5 до 0,7 – значителна;
 от 0,7 до 0,9 – голяма (силна);
 Над 0,9 – много голяма (много силна);
 При 1 – зависимостта е функционална.

Посока на зависимостта

Някои коефициенти на корелация имат **знак**, който показва **посоката** на зависимостта:

- Ако коефициентът на корелация е **положително число**, зависимостта е възходяща, еднопосочна - на по-големи стойности на едната променлива, съответстват по-големи стойности на другата променлива;
- Ако той е **отрицателно число** – зависимостта е отрицателна, низходяща, разнопосочна - на по-големи стойности на едната променлива съответстват по-ниски стойности на другата променлива.

• [Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

КОЕФИЦИЕНТИ НА ДЕТЕРМИНАЦИЯ И НЕОПРЕДЕЛЕНОСТ

Стойностите на **зависимата променлива (Y)** варират, тъй като върху тях оказват влияние много фактори. Като се изследва един (част) от тях (**независимата променлива X**), се обяснява **само част** от вариацията на Y.

Частта от дисперсията на зависимата променлива, която се *обяснява с влиянието на изучавания фактор (X)* се нарича **обяснена дисперсия** и се описва с т. нар. **коэффициент на детерминация**. Той се получава, като коэффициентът на корелация се повдигне на квадрат (r^2). Ако r^2 се умножи по 100, обяснената дисперсия се изразява в проценти.

Тъй като зависимостите обикновено са непълни (корелационни), **X** не обяснява цялата вариация на зависимата променлива. **Необяснената част от дисперсията** се нарича **коэффициент на неопределеност (k^2)** и се получава, като от като от 1 се извади коэффициентът на детерминация ($k^2=1-r^2$).

ЗАКЛЮЧЕНИЯ НА СЪВКУПНОСТТА

Полученият по първичните данни коэффициент на корелация носи информация за зависимостта между стойностите в извадката, т.е той е точкова оценка на параметъра на съвкупността, който се бележи с **p**. При малки извадки стойността му може да бъде повлияна от случайни фактори. Поради това се проверява дали извадковият корелационен коэффициент е **статистически значим** (т.е. дали статистически значимо се различава от 0).

Най-често използваните, в областта на спорта, коэффициенти на корелация са коэффициенти на Пирсън и Спирман. Те са разгледани в следващите два раздела.

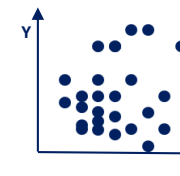
[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Пример 1

Няма зависимост между X и Y

Вариацията на Y не е свързана с тази на X

$$r^2=0 \quad k^2=1$$

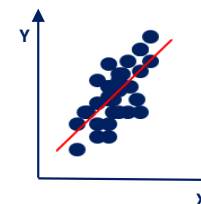


Пример 2

Зависимостта е корелационна

Част от вариацията на Y е свързана с тази на X

$$r^2 < 1 \quad k^2 > 0$$

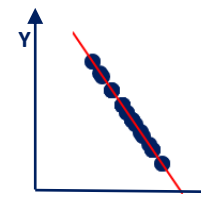


Пример 3

Зависимостта е функционална

Вариацията на Y е свързана напълно с вариацията на X

$$r^2=1 \quad k^2=0$$



КОЕФИЦИЕНТ НА ОБИКНОВЕНА ЛИНЕЙНА КОРЕЛАЦИЯ НА ПИРСЪН (r)

Коефициентът на Пирсън (r) се прилага за описание на силата и посоката на зависимостта, когато зависимостта е обикновена по вид, линейна по форма, а **X** и **Y** – количествени.

Формулата, по която се изчислява стойността на коефициента на Пирсън (**r**) по данни от извадката е:

$$r = \frac{P}{S_x S_y} \quad (4.1)$$

Където:

P – момент на произведенията (формула 4.2), **S_x** – стандартно отклонение на **X** и **S_y** – стандартно отклонение на **Y**

$$P = \frac{\sum XY}{n-1} - \frac{\sum X \sum Y}{n(n-1)} \quad (4.2)$$

Където:

ΣX – сумата на стойностите на **X**, **ΣY** – сумата на стойностите на **Y**, **ΣXY** – сумата на произведенията на **X** и **Y**, **n** – обем на извадката

Изчисленият по данни от извадката коефициент (**r**) е точкова оценка на параметъра на съвкупността, който се бележи с **ρ**.

Статистическата значимост на коефициента на Пирсън (**r**) се проверява като се изчислява **t**-статистиката по формула:

$$t_{emp} = r \sqrt{\frac{n-2}{1-r^2}} \quad (4.3)$$

Получената стойност се сравнява с отчетената от [приложение 4](#) стойност (**t_α**) в зависимост от равнището на значимост (**α=0,05** или **α=0,01**) и степени на свобода **df=n-2**. Ако **t_{emp} ≥ t_α** - коефициентът на корелация се различава статистически значимо от 0 (зависимостта е достоверна).

Тема 4

За удобство, в [приложение 6](#) са табулирани критичните стойности на коефициента на Пирсън при дадено **равнище на значимост** (0,05 или 0,01) и **степени на свобода df=n-2**. Ако извадковият коефициент по абсолютна стойност е равен или по-висок от критичната стойност – коефициентът е статистически значим. В противен случай се приема, че проявата на зависимостта се дължи на случайни фактори.

По данните от примера в началото на главата за връзката между постиженията в 20 м. гл. б. и вертикалния отскок на младежки футболни отбори **r=- 0.814**, **t_{emp}=4.43**, **t_{0,05}=2,23**, което означава, че зависимостта е голяма по сила, низходяща по посока и статистически значима.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

КОЕФИЦИЕНТ НА РАНГОВА КОРЕЛАЦИЯ НА СПИРМАН (r_s)

Коефициентът на рангова корелация на Спирман (r_s) се прилага, когато променливите са рангово скалирани величини. Неговото широко приложение в научните изследвания в спорта се определя от факта, че спортният резултат в редица спортове е рангово скалирана величина (класиране на отбори и състезатели, ранглисти и т.н.). Коефициентът може да се ползва и при количествени променливи, които имат известни отклонения от нормалното разпределение. За целта същите трябва да бъдат трансформирани в рангови. Това става, като конкретните количествени стойности се заменят с ранговите им номера¹, съответно X' и Y' . Ако в първичните данни има повтарящи се стойности – те трябва да получат един и същ, среден рангов номер. Стойността на коефициента се получава по формулата:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \quad (4.4)$$

Където:

d - разлики между X'_i и Y'_i ;

n – обем на извадката.

Интерпретацията на коефициента става по описания вече начин (може да си го припомните [от тук](#)).

Статистическата значимост на коефициента на Спирман (r_s) се определя, като се сравни стойността, изчислена по данни от извадката, с тази, отчетена от [приложение 7](#) в зависимост от степените на свобода $df=n-2$ и възприетото равнище на значимост ($\alpha=0,05$ или $\alpha=0,01$). Ако емпиричният коефициент по абсолютна стойност е равен или по-голям от отчетения от приложението – зависимостта е статистически значима.

ПРИМЕР

По данни от извадка с обем $n=32$ е установено, че коефициентът на Спирман между отскокливостта и класирането на отбори по волейбол е $r_s=-0,850$.

Как би следвало да се интерпретира този факт?

В извадката се проявява голяма низходяща зависимост, което означава, че отборите с по-добра отскокливост заемат по-предни места в класирането.

Критичната стойност при $df=30$ и $\alpha=0,05$ е $r_{0,05}=0,36$. Това означава, че зависимостта е статистически значима,

Тема 4

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

¹ SPSS извършва тази процедура автоматично.



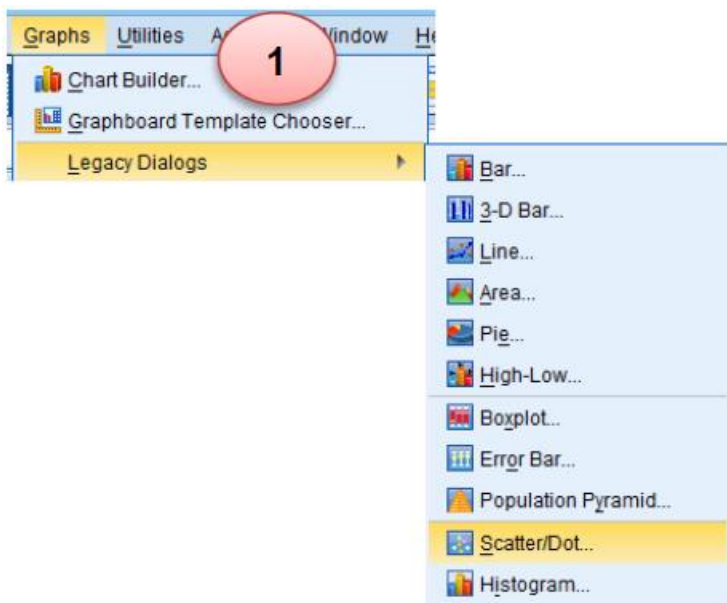
КОРЕЛАЦИОНЕН АНАЛИЗ С IBM SPSS STATISTICS

Преди да се осъществи корелационен анализ трябва да се построи диаграма на разсейване, за да се установи формата на зависимостта и евентуално да се установи има ли наличие на силно отклоняващи се стойности, които да се отразяват на изводите за зависимостта.

По данните от Soccer_teams.sav

Да направи диаграма на разсейване на зависимостите между тестовите за физическа дееспособност t1, t2, t3, t4

1 За построяване на диаграма на разсейване се избира менюто **Graphs>Legacy Dialogs>Scatter/Dot**.

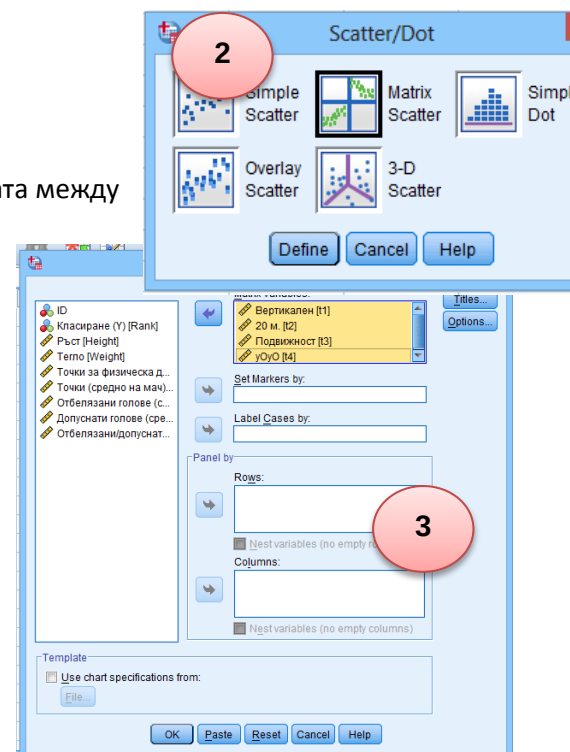


2 От отворения се диалогов прозорец се избира:

- **Simple Scatter**, ако се изследва само връзката между две променливи;

- **Matrix Scatter**, ако (както е в случая) се изследват обикновените зависимости между повече от две променливи. След това се потвърждава, като се натисне бутона **Define**.

3 Обработваните променливи се изпращат в полето **Matrix Variables** и се потвърждава с **OK**.

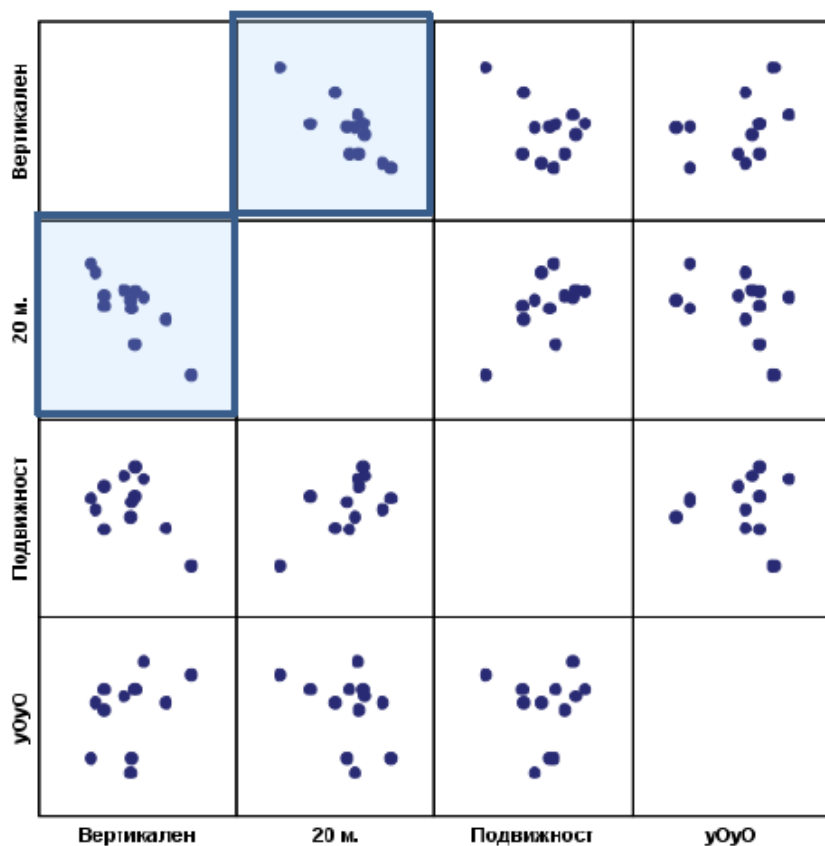


Тема 4

Фиг. 4.5

РЕЗУЛТАТИ ОТ ПОСТРОВАНЕ НА ДИАГРАМА НА РАЗСЕЙВАНЕ

В резултат на осъществената процедура се появява матрица от диаграми на разсейване (фиг. 4.6).



Фиг. 4.6

Вляво и отдолу на графиката са изписани зададените променливи величини.

В клетките, получени при пресичане на даден ред и колона се появява диаграма на разсейване между съответните променливи. По диагонала, на получената таблица се намират празни клетки. Там би трябвало да се появи диаграма за дадена променлива със себе си, което е излишно, защото зависимостта е функционална.

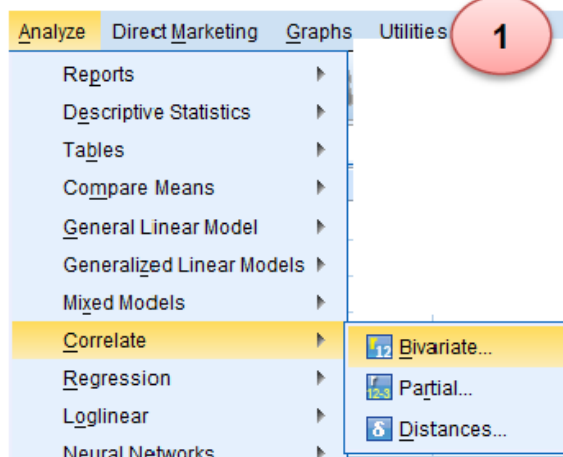
Матрицата е симетрична по отношение на този диагонал – тестове 20 м. гл. б. и вертикален отскок се пресичат на две места на матрицата – където са заштрихованите клетки на таблицата. Формата на облака на разсейване ясно показва, че между двете променливи има добре изразена линейна низходяща зависимост.

Както вече беше споменато, огледът на диаграмите на разсейване се прави, за да се установи формата на зависимостите (за да се прецени дали коефициентът на Пирсън е подходящ за описание на силата на зависимостта) и да се установи съществуват ли стойности, които грубо да нарушават очертаващата се конфигурация от точки и да променят заключението на зависимостта. В случая няма такива, поради което може да се премине към корелационния анализ.



ИЗЧИСЛЯВАНЕ НА КОЕФИЦИЕНТА НА ПИРСЪН (r)

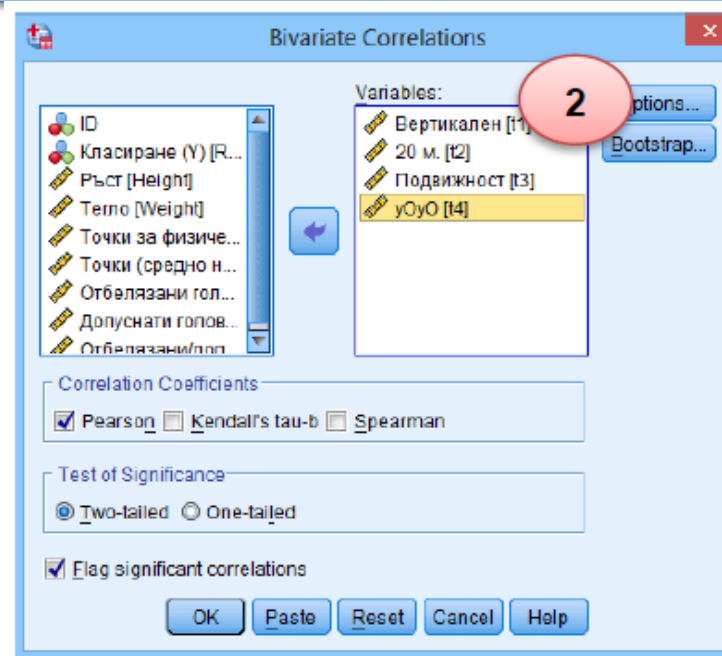
1 Обикновената линейна корелация се задава от менюто **Analyze>Correlate>Bivariate**. Отваря се диалоговият прозорец **Bivariate Correlations**.



2 Променливите величини се прехвърлят в полето **Variables**. По подразбиране е зададено изчисляване на коефициента на Пирсън (**Pearson**), така че, при така ползваните данни отметката не се променя. Процедурата стартира, когато се потвърди с **OK**.

По данните от [Soccer_teams.sav](#)

Да се направи корелационен анализ на зависимостите между показателите за физическа дееспособност t1, t2, t3 и t4.



Тема 4

Фиг. 4.7. Задаване на корелационен анализ

В резултат на осъществената процедура се появява корелационната матрица, представена на фиг. 4.8.

Correlations		Вертикален	20 м.	Подвижност	yoYo
Вертикален	Pearson Correlation	1	-,814**	-,457	,444
	Sig. (2-tailed)		,001	,135	,148
	N	12	12	12	12
20 м.	Pearson Correlation	-,814**	1	,615*	-,388
	Sig. (2-tailed)	,001		,033	,212
	N	12	12	12	12
Подвижност	Pearson Correlation	-,457	,615*	1	-,118
	Sig. (2-tailed)	,135	,033		,715
	N	12	12	12	12
yoYo	Pearson Correlation	,444	-,388	-,118	1
	Sig. (2-tailed)	,148	,212	,715	
	N	12	12	12	12

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Фиг. 4.8

Тя е аналогична на подробно описаната матрица на диаграмите на разсейване. В клетките на матрицата се появяват 3 стойности:

- **Pearson Correlation** – стойност на коефициента на корелация;
- **Sig. (2-tail)** – равнище на значимост, съответстващо на коефициента на корелация;
- **N** – брой на изследваните, на базата на които е изчислен коефициентът. В случай, че има липсващи стойности в някоя от колоните **N** ще е различно за корелациите, в които този показател участва.

Коефициентите на корелация, които са статистически значими при равнище на значимост $\alpha=0,05$ са маркирани с една (*), а тези които са съществени при $\alpha=0,01$ – с две звездички (**).

Таблица 4.2. Корелационен анализ между показателите за физическа дееспособност

Показатели	Вертикален отскок	20 м. гл.б.	Подвижност	YoYo
Вертикален отскок	1			
20 м. гл.б.	-0,814**	1		
Подвижност	-0,457	0,615*	1	
YoYo	0,444	-0,388	-0,118	1

защото, при този обем на извадката не корелира с останалите показатели.

Критични стойности на коефициента на Пирсън при $n=12$: $r_{0,05}=0,58$ и $r_{0,01}=0,71$

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Да се направи корелационен анализ, който да включва класирането и показателите за физическа дееспособност t1, t2, t3 и t4.

КОЕФИЦИЕНТ НА РАНГОВА КОРЕЛАЦИЯ НА СПИРМАН (r_s)

Correlations							
			Класиране (Y)	Вертикален	20 м.	Подвижност	yOyO
Spearman's rho	Класиране (Y)	Correlation Coefficient	1,000	-,396	,657 [*]	,455	-,085
		Sig. (2-tailed)		,203	,020	,138	,793
		N	12	12	12	12	12
Вертикален	Вертикален	Correlation Coefficient	-,396	1,000	-,711 ^{**}	-,095	,546
		Sig. (2-tailed)	,203		,010	,770	,066
		N	12	12	12	12	12
20 м.	20 м.	Correlation Coefficient	,657 [*]	-,711 ^{**}	1,000	,531	-,358
		Sig. (2-tailed)	,020	,010		,075	,254
		N	12	12	12	12	12
Подвижност	Подвижност	Correlation Coefficient	,455	-,095	,531	1,000	-,071
		Sig. (2-tailed)	,138	,770	,075		,827
		N	12	12	12	12	12
yOyO	yOyO	Correlation Coefficient	-,085	,546	-,358	-,071	1,000
		Sig. (2-tailed)	,793	,066	,254	,827	
		N	12	12	12	12	12

*, Correlation is significant at the 0.05 level (2-tailed).

**, Correlation is significant at the 0.01 level (2-tailed).

Фиг. 4.9

Таблица 4.2. Зависимости между двигателните способности и класирането на отборите

Показатели	Класиране	Вертикален отскок	20 м. гл.б.	Подвижност	YoYo
Класиране	1				
Вертикален отскок	-,396	1			
20 м. гл.б.	0,657*	-,711**	1		
Подвижност	0,455	-,095	0,531	1	
YoYo	-,085	0,546	-,358	-,071	1

Класирането на отборите е рангова променлива, поради което за изследване на зависимостите трябва да се ползва коефициентът на рангова корелация на Спирман. Алгоритмът за неговото изчисляване е аналогичен с този за коефициента на Пирсън, поради което няма да го описваме отново. Единствената разлика е, че в диалоговия прозорец **Bivariate Correlations** трябва да се постави отметка в полето **Spearman**. В резултат това се появява матрицата, представена на фиг. 4.9.

При представяне на корелационната матрица в научни разработки се редактира, по начина представен в таблица 4.2.

Коефициентите на рангова корелация между класирането и изследваните двигателни способности показват, че статистически значимо *значително* влияние има само **бързината** на състезателите ($r_s=0,657$). Положителният знак показва, че отборите, които имат по-високо равнище на скоростните възможности заемат по-предни места в класирането. **Подвижността** има *умерено* влияние ($r_s=0,455$), но поради твърде малкия обем на извадката ($n=12$), коефициентът не е статистически значим.

Критична стойност при $n=12$ и $\alpha=0,05$ - $rs_{0,05}=0,65$



ЗАДАЧИ ЗА УПРАЖНЕНИЕ

[EDesign2x2.sav](#)

[Soccer_teams.sav](#)

[Калкулатор](#)

Задача 4.1

По данните от файла **EDesign2x2.sav** направете корелационен анализ на 5 променливи по избор по данни от първо изследване.

Показатели					

Критичната стойност на r при $n=.....$ и $\alpha=0,05$ е равна на([прил. 6](#))

Анализирайте резултатите.

Задача 4.2

По данните от файла **Soccer_teams.sav** направете корелационен анализ с коефициента на Спирман, който да включва класирането на отборите, ръста, теглото, точките за физическа дееспособност и отношението вкарани/допуснати голове.

Показатели	Класиране	Ръст	Тегло	SPARQ	Отбелязани/допуснати голове
Класиране					
Ръст					
Тегло					
SPARQ					
Отбелязани/допуснати голове					

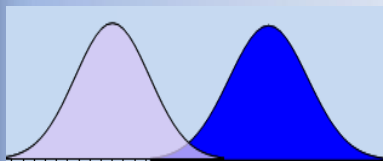
Критичната стойност на r_s при $n=.....$ и $\alpha=0,05$ е равна на([прил. 7](#)).

Тема 4

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ТЕМА 5. ПРОВЕРКА НА ХИПОТЕЗИ ПРИ НЕЗАВИСИМИ ИЗВАДКИ

$$H_0: \mu_1 - \mu_2 = 0$$



Изследвани се резултатите в тест **скок на дължина от място** на млади футболисти ($\bar{X}_1 = 196,48$ см.) и баскетболисти ($\bar{X}_2 = 203,9$ см.).

Средните стойности на извадките се различават. Това означава ли, че се различават и средните на генералните съвкупности?

В настоящата тема се разглежда начинът, по който се сравнява разпределението на променливи величини с IBM SPSS Statistics.

Теоретичен материал

- [Предназначение](#)
- [Понятия](#)
- [Алгоритъм за проверка на хипотези](#)
- [Критерии за проверка на хипотези](#)
 - [t-критерий за независими извадки](#)
 - [F-критерий на Фишер](#)

Проверка на хипотези при независими извадки с IBM SPSS

- [t-критерий на Стюдънт за независими извадки](#)
- [F-критерий на Фишер \(One Way ANOVA\)](#)
- [Задачи за упражнение](#)



Статистическата проверка на хипотези се прилага, когато е необходимо да се сравнят на статистическите характеристики на изучаваните явления в извадките и се провери дали наблюдаваната в точковите оценки разлика се проявява и в съвкупността.

Тази проблематика се ползва, когато се сравнява състоянието **характеристиките на извадката с тази на съвкупността**, при сравняване на изследваните признаци **при различни групи от изследвани** (състезатели от различен вид спорт, квалификация, пол, игрови пост, експериментална-контролна група) или **при поредни изследвания** (преди-след прилагане на различни тренировъчни методики, в хода на повишаване на квалификацията и др.).

ПРЕДНАЗНАЧЕНИЕ

ПРИМЕР

Изследователят се интересува различават ли се по физическите качества състезатели от различен вид спорт. За да реши тази изследователска задача, се провежда тестиране и установява състоянието на физическите качества в извадките. Съвсем естествено **точковите оценки** (резултатите в извадките) ще се различават. С помощта на статистическата проверка на хипотези се установява дали се различават **и параметрите на съвкупностите**, от които са селектирани извадките.

- Ако данните от извадката не дават основание да се счита, че параметрите на съвкупността се различават се приема, че разликата е случайна (несъществена, недоверна).
- Ако на базата на данните от извадката се приеме, че параметрите на съвкупността се различават – дадената разлика се нарича статистически значима (съществена, доверна).

Тема 5

ПОНЯТИЯ

Видове хипотези

Научната хипотеза е предположение, който изследователят желае да провери по емпиричен път. По данните от примера в началото на темата, тя гласи, че състезателите от различен вид спорт се различават по състоянието на физическите си качества. Отговорът трябва да се базира на емпирични данни (данни от извадки), а да се отнася за съответните генералните съвкупности. За целта, научната хипотеза се преформулира в някоя от следните две **статистически хипотези**:

Нулева (работна) хипотеза (H_0), която твърди, че параметрите на сравняваните съвкупности не се различават. Въз основа на констатираното в извадките се прави заключение дали емпиричните данни основание за нейното отхвърляне.

Алтернативна хипотеза (H_1), която твърди, че разликата, която се наблюдава по данни от извадката е статистически значима (съществена), което означава, че се проявява и по отношение на параметрите на съответните генерални съвкупности.

Вероятностен характер на заключенията

Заключенията, които се правят имат вероятностен характер. Степента на сигурност, с която се приема за вярна алтернативната хипотеза, се нарича **гаранционна вероятност (P)**. Вероятността да е вярна нулевата хипотеза се

нарича **равнище на значимост (α)**.

В практиката се използват следните гранични стойности за **гаранционната вероятност (P)** и **равнището на значимост (α)**:

P= 95%, на която съответства $\alpha = 0,05$

P= 99%, на която съответства $\alpha = 0,01$

Видове извадки

От гледна точка на проверката на хипотези извадките се делят на:

A

Б

I

II

независими, при които когато статистическите единици в извадките са различни;

зависими, при които два или повече пъти се изследват едни и същи статистически единици.

С подобен характер са заключенията, които се правят за:

- допустимите отклонения на коефициентите на асиметрия (**As**) и ексцес (**Ex**) от 0, разгледани в тема 3;
- статистическата значимост на коефициентите на корелация, разгледана в тема 4.

Тема 5

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

АЛГОРИТЪМ ЗА ПРОВЕРКА НА ХИПОТЕЗИ

Статистическата проверка на хипотези преминава през следните стъпки:

Нулевата (H_0) хипотеза твърди, че наблюдаваната в извадката разлика е случайна, а **алтернативната** – че тя е съществена.

Подходящият критерий се подбира в зависимост от проверяваната хипотеза, вида на променливите, вида и броя на извадките. Критериите са описани на следващата страница.

За всеки критерий има разработена процедура за изчисляването му по данни от извадката. Получената стойност се нарича емпирична и се означава със символа на съответния критерий и долен индекс **emp** (Kp_{emp}). Когато данните се обработват със специализиран софтуер, освен Kp_{emp} , се изчислява и съответстващото му равнище на значимост (α), от което може да се изчисли гаранционната вероятност (**P%**) по формулата $P(\%) = (1 - \alpha) * 100$.

Критичната стойност (Kp_{α}) се определя от приложения в зависимост от **степените на свобода (df)** и **равнището на значимост** ($\alpha=0,05$ или $0,01$).

Решението може да се вземе на базата на някои от следващите условия:

- Kp_{emp} се сравнява с критичната - Kp_{α} .
- Достигнатата α се сравнява с възприетата от **0,05**, или **0,01**;
- Достигната **P%** се сравнява с възприетата от **95%** или **99%**.

Хипотеза	Емпирична – критична стойност*	Равнище на значимост (α)	Гаранционна вероятност (P%)
Приемане на H_0	$Kp_{emp} < Kp_{\alpha}$	$\alpha > 0,05$ (0,01)	$P(\%) < 95\%$ (99%)
Отхвърляне на H_0 (приемане на H_1)	$Kp_{emp} \geq Kp_{\alpha}$	$\alpha \leq 0,05$ (0,01)	$P(\%) \geq 95\%$ (99%)

При някои критерии условието за проверка на нулевата хипотеза е обратно – H_0 се отхвърля, ако $Kp_{emp} \leq Kp_{\alpha}$. Това не важи за останалите начини за вземане на решение.

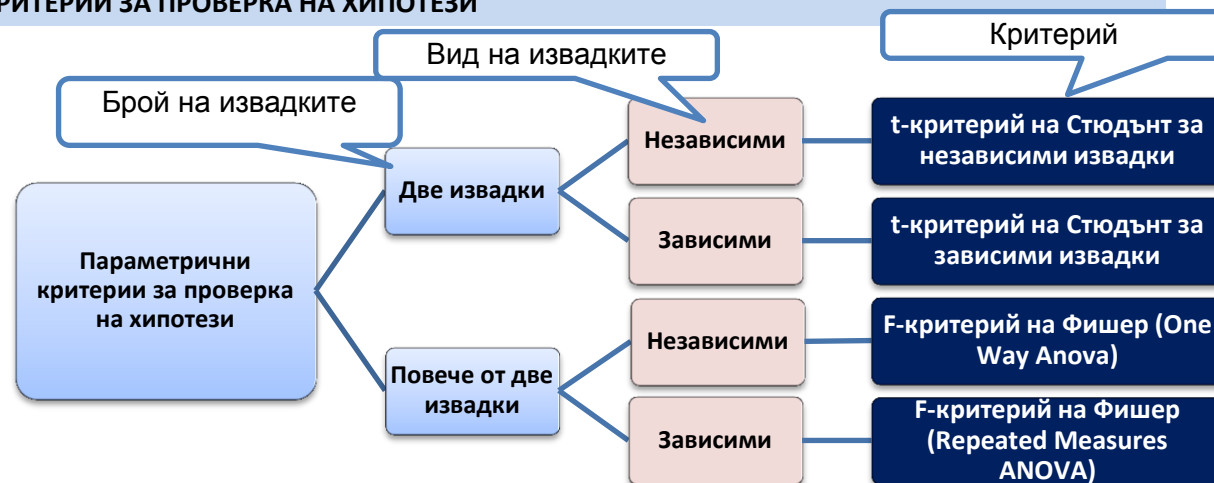
Тема 5

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

КРИТЕРИИ ЗА ПРОВЕРКА НА ХИПОТЕЗИ

Важно място в процеса на проверка на хипотези заема изборът на подходящ критерий, осъществяван в стъпка 2. Той се подбира в зависимост естеството на хипотезата, вида и разпределението на изучаваните променливи, броя и вида на извадките.. Най-общо критериите се делят на:

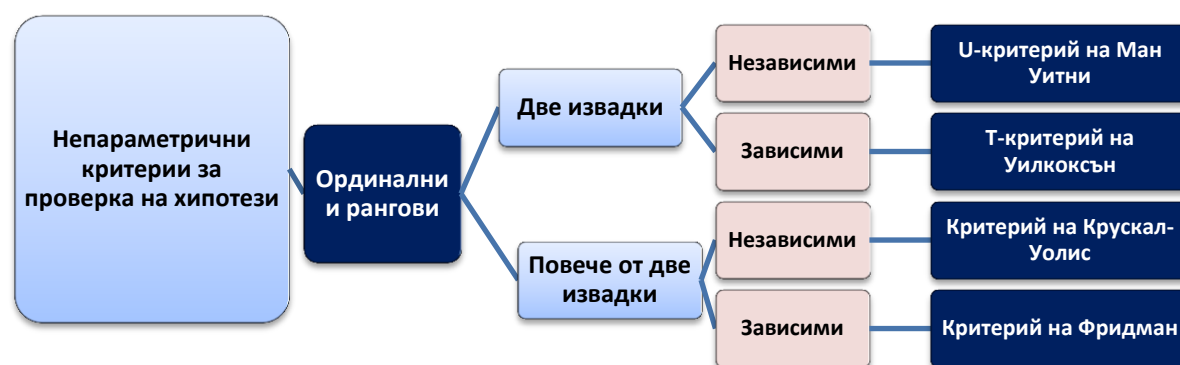
- **Параметрични**, които се ползват за сравняване на количествени променливи, които имат нормално разпределение;
- **Непараметрични**, които се ползват при качествени променливи и количествени, които имат различно от нормалното разпределение.



Фиг. 5.1

На фигури 5.1 и 5.2 са представени едни от най-често прилаганите критерии за проверка на хипотези – тези, които се ползват за сравняване на количествени, рангови и ординални променливи величини.

В настоящия модул се изучават само параметрични критерии за проверка на хипотези. В избираемия модул **MSP.2.4** се разглеждат и някои от непараметричните критерии за проверка на хипотези.



Фиг. 5.2

Тема 5

t-КРИТЕРИЙ НА СТЮДЪНТ ЗА НЕЗАВИСМИ ИЗВАДКИ

Но: $\mu_1 - \mu_2 = 0$

t-критерият на Стюдънт за независими извадки се ползва за сравняване на две независими извадки по признаци, които са количествени и имат нормално разпределение.

Емпиричната стойност на критерия (t_{emp}) се изчислява по формулата:

$$t_{emp} = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2} \cdot \frac{n_1+n_2}{n_1 n_2}}} \quad (5.1)$$

Където:

$\bar{X}_1$ и $\bar{X}_2$ – средни стойности в извадките;
 S_1 и S_2 – стандартни отклонения в извадките;
 n_1 и n_2 – обеми на извадките.

Критичната стойност (t_α) се определя от [приложение 4](#) при степени на свобода $df=n_1+n_2-2$ и възприетото равнище на значимост ($\alpha=0,05$ или $\alpha=0,01$).

Решение:

Ако $t_{emp} < t_\alpha$ - няма основание за отхвърляне на нулевата хипотеза (H_0),

Ако $t_{emp} \geq t_\alpha$ - за вярна се приема алтернативната хипотеза (H_1)

ПРИМЕР

Сравняват се резултатите в тест **скок на дължина от място** на футболисти ($\bar{X}_1 = 196,48$ см. и $n_1=32$) и баскетболисти ($\bar{X}_2 = 203,9$ см. и $n_2=50$).

Нулевата хипотеза (Но) твърди, че разликата в постиженията на представителите на двата вида спорт е несъществена. За проверка на хипотезата се ползва t-критерият на Стюдънт за независими извадки. Емпиричната стойност е равна на $t_{emp}=1.95$. [Критичната стойност](#) при $df=80$ и $\alpha=0,05$ е равна на 1,98.

$t_{emp} < t_{0.05}$, което означава, че разликата във взривната сила на долните крайници на двете групи състезатели е несъществена.

Тема 5

Ho: $\mu_1 - \mu_2 - \mu_3 = 0$

F-КРИТЕРИЙ НА ФИШЕР

Класическият F-критерий на Фишер се ползва за сравняване на повече от две независими извадки. Анализът е известен под наименованието **дисперсионен анализ (One Way ANOVA)**. Неговата логика се основава на разлагане на общата дисперсия на стойностите на изучаваната променлива на два основни компонента:

Таблица 5.1. Схема на дисперсионния анализ (One Way ANOVA)

Вътрешно групова дисперсия (MS_{within}), която се описва с отклоненията на индивидуалните стойности ($\bar{X}_{i,j}$) от средната на съответната извадка ($\bar{X}_j$);

Междугрупова дисперсия ($MS_{between}$), която се описва с отклоненията на груповите средни ($\bar{X}_j$) от общата средна стойност ($\bar{X}_{total}$).

Източник на дисперсията	SS (сума от квадратите на отклоненията)	df (степени на свобода)	MS (средна квадратична)	F - критерий
Between groups	$SS_{between} = \sum n_j (\bar{X}_j - \bar{X}_{total})^2$	$df_b = k - 1$	$MS_{between} = \frac{SS_{between}}{df_b}$	$\frac{MS_{between}}{MS_{within}}$
Within groups	$SS_{within} = \sum (X_{i,j} - \bar{X}_j)^2$	$df_w = n - k$	$MS_{within} = \frac{SS_{within}}{df_w}$	
Обща	$SS_{total} = \sum (X_i - \bar{X}_{total})^2$	$n - 1$		

Схемата за изчисляване на тези компоненти на дисперсията е описана в таблица 5.1.

Емпиричната стойност на F-критерия се изчислява по формулата:

$$F_{emp} = \frac{MS_{between}}{MS_{within}} \quad (5.2)$$

Където: **$MS_{between}$** – междугрупова дисперсия, **MS_{within}** – вътрешногрупова дисперсия, **n** - обем на извадката, **k**-брой на извадките
Критичната стойност на F-критерия на Фишер се определя от [приложение 5](#) при степени на свобода **$df_b=k-1$** **$df_w=n-k$** .

Ако **$F_{emp} < F_{0.05}$** – няма основание за отхвърляне на нулевата хипотеза, ако **$F_{emp} \geq F_{0.05}$** – за вярна се приема алтернативната хипотеза.

Ако прилагането на дисперсионния анализ покаже, че разликата между групите е статистически значима, трябва да се установи между кои от групите тя е съществена. Това става с различни критерии, един от които е **критерия на Тюки**, който е разгледан в практическата част на заниманието.

Тема 5

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)



ИЗЧИСЛЯВАНЕ НА t-КРИТЕРИЯ НА СТЮДЪНТ ЗА НЕЗАВИСИМИ ИЗВАДКИ С IBM SPSS

По данните [от Soccer_juniors.sav](#)

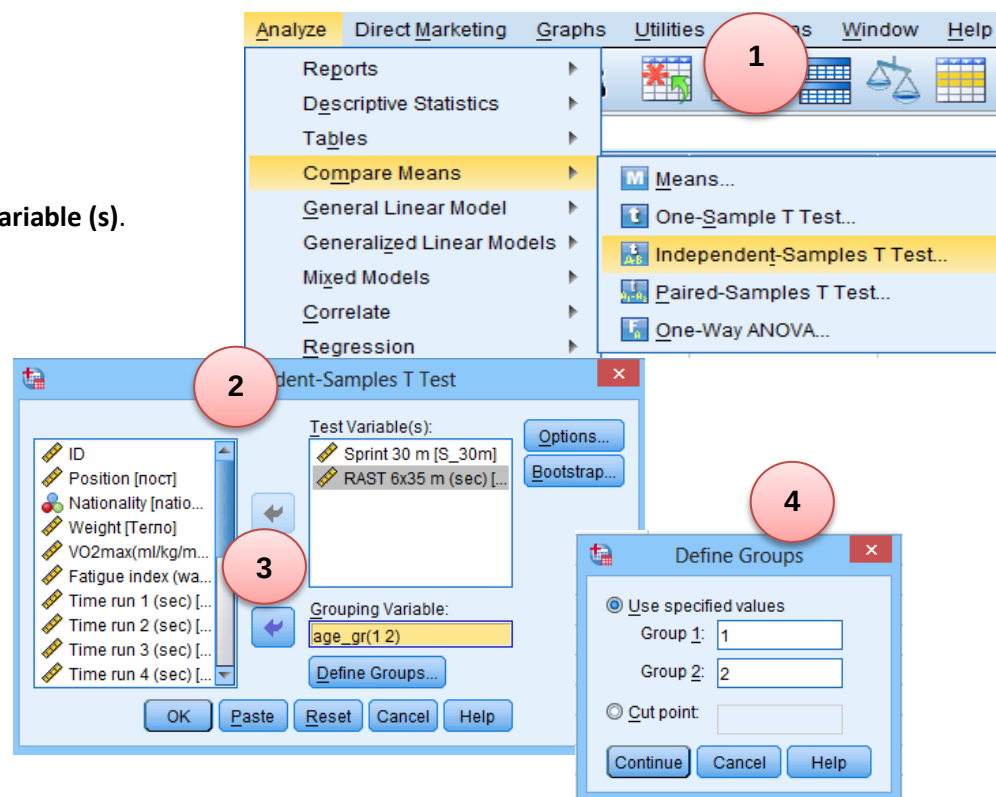
Да се установи статистически значима ли е разликата между бързината (**Sprint 30 m.**) и скоростната издръжливост (**RAST**) на състезатели на 16 и 18 години (променливата **Age_groups** – кодове 1 и 2).

1 Анализът се задава, като се избере **Analyze>Compare Means>Independent Samples t test.**

2 Обработваната(ните) променлива(и) се изпраща(т) в полето **Test Variable (s).**

3 Групиращата променлива – в полето **Grouping Variable.** Необходимо е да се зададат кодовете на сравняваните групи, поради което се натиска бутонът **Define Groups.**

4 В отворения се диалогов прозорец се задават кодовете на сравняваните групи и се потвърждава с **Continue>OK.**



Фиг. 5.3

Тема 5

t-КРИТЕРИЙ НА СТЮДЪНТ ЗА НЕЗАВИСИМИ ИЗВАДКИ – РЕЗУЛТАТ ОТ ОБРАБОТКАТА

Резултатът от обработката съдържа две части:

Group Statistics

	Age group	N	Mean	Std. Deviation	Std. Error Mean
Sprint 30 m	16 years	64	4,2447	,18189	,02274
	18 years	27	4,1467	,16618	,03198
RAST 6x35 m (sec)	16 years	64	32,1252	1,49315	,18664
	18 years	27	30,3919	1,19284	,22956

Фиг. 5.4

Group Statistics (фиг. 5.4). В нея е изписана информация за разпределението на стойностите в сравняваните извадки (групи) - обем (**N**), средни стойности (**Mean**), стандартни отклонения (**Std. Deviation**) и стандартните грешки (**Std. Error Mean**).

Independent Samples Test (фиг. 5.5). В тази част е публикувана информация за разликата между средните стойности на сравняваните групи (**Mean Difference**), емпиричната стойност на t-критерия на Стюдънт (**t**), степените на свобода (**df**) и достигнатото равнище на значимост (**Sig. 2-tailed**).

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Sprint 30 m	Equal variances assumed	,121	,729	2,407	89	,018	,09802	,04072	,01711	,17893
	Equal variances not assumed			2,498	53,803	,016	,09802	,03924	,01933	,17671
RAST 6x35 m (sec)	Equal variances assumed	,841	,362	5,349	89	,000	1,73330	,32404	1,08945	2,37716
	Equal variances not assumed			5,858	60,775	,000	1,73330	,29586	1,14165	2,32496

Фиг. 5.5

Разликата между групите е статистически значима, ако α (Sig. 2-tailed) $\leq 0,05$.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ПРЕДСТАВЯНЕ НА РЕЗУЛТАТИТЕ ОТ СРАВНЯВАНЕ НА ДВЕ НЕЗАВИСИМИ ИЗВАДКИ

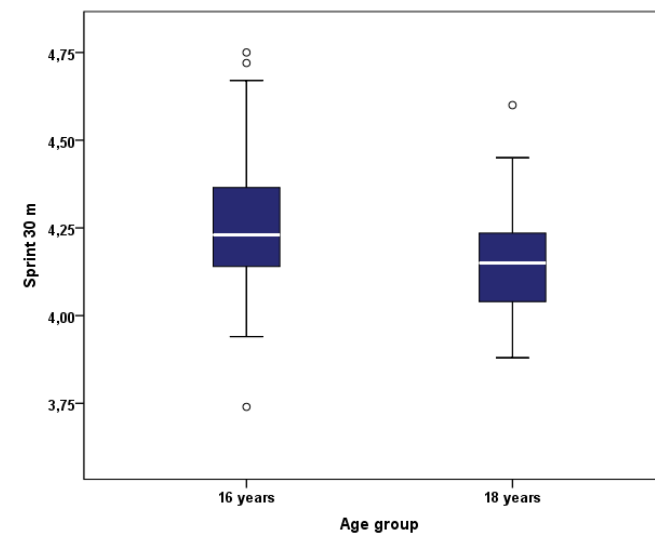
Таблица 5.2. Сравнителен анализ на бързината и скоростната издръжливост на 16 – 18 годишни футболисти

Показател	16 годишни			18 годишни			Разлика	Статистическа значимост на разликите	
	n_1	$\bar{X}_1$	S_1	n_2	$\bar{X}_2$	S_2	d	t_{emp}	α
30 м. гл. б.	64	4,24	0,18	27	4,15	0,17	-0,10	2,41	0.018
RAST	64	32,13	1,49	27	30,39	1,19	-1,73	5,35	0.000

Забележка: Критичната стойност на t-критерия при $df=n_1+n_2-2=89$ и $\alpha=0.05$ е приблизително равна на 1,98.

Резултатите от изследването показват, че средната стойност на постиженията в тест 30 м. гл. б. в групата на 16-годишните футболисти е 4,24 сек, а в горната възрастова група – 4,15 сек. Разликата от 0,1 сек. е статистически значима, защото достигнатото равнище на значимост е по-малко от 0,05 ($\alpha=0,018$).

Подобни изводи налага и сравняването на резултатите в теста за скоростна издръжливост (RAST). Средните стойности на групите са съответно 32,13 сек. и 30,39 сек., като разликата между тях е с висока статистическа значимост ($\alpha\approx 0,000$).



Фиг. 5.6

ИЗЧИСЛЯВАНЕ НА F-КРИТЕРИЯ НА ФИШЕРС IBM SPSS



1 Анализът се задава, като се избере **Analyze>Compare Means>One Way ANOVA**.

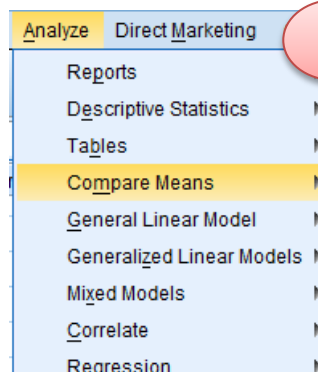
2 Изучаваната променлива се изпраща в полето **Dependent List**, а групиращата променлива - в полето **Factor**.

3 Натиска се бутонът **Options**, поставя се отметка в полето **Descriptives** и се потвърждава с **Continue**.

4 Натиска се бутонът **Post Hoc**, поставя се отметка в полето **Tukey** и се потвърждава с **Continue**.

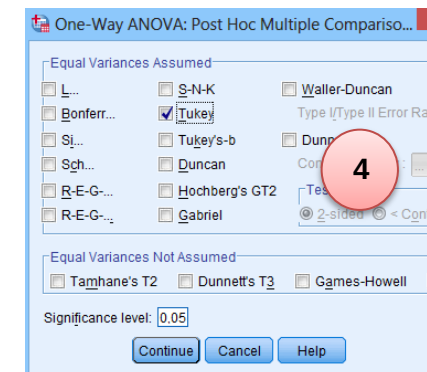
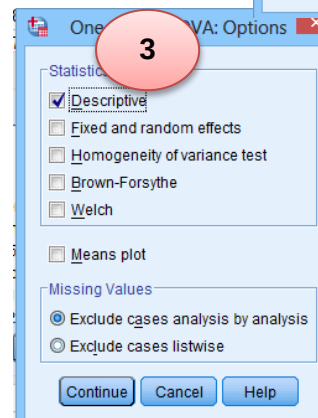
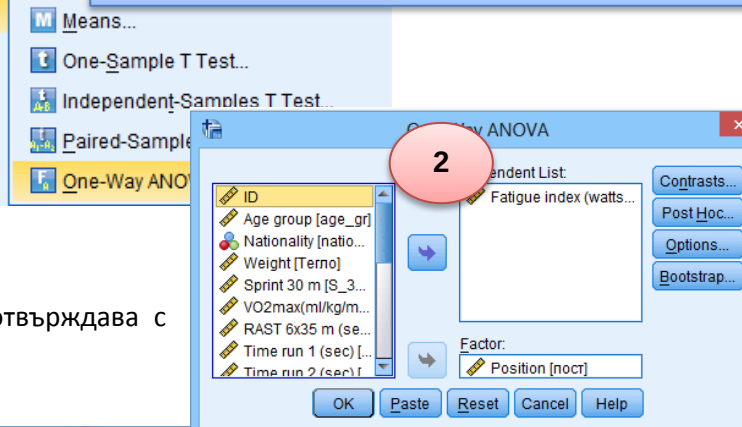
Изчислителната процедура стартира, като от основния прозорец се потвърди с **OK**.

Фиг. 5.7



По данните от [Soccer_juniors.sav](#)

Да се установи статистически значима ли е разликата между скоростната издръжливост (**fatigue_index**) на състезатели с различен игрови пост.



Тема 5

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

РЕЗУЛТАТИ ОТ ОБРАБОТКАТА

Разпечатката от дисперсионния анализ съдържа 3 основни части:

Descriptives. В тази част е описано разпределението на стойностите в извадките (групите) – брой на изследваните (**N**), средни стойности (**Mean**), стандартни отклонения (**Std. Deviation**) и стандартните грешки (**Std. Error Mean**), както и минималната и максималната стойност.

Descriptives

Fatigue index (watts/sec)						
	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
forwards	19	6,4527	2,32734	,53393	1,89	10,52
mildfielders	36	6,1234	2,35879	,39313	1,70	12,38
defenders	33	6,6887	1,95851	,34093	2,32	9,73
goalkeepers	3	10,3851	3,03850	1,75428	8,58	13,89
Total	91	6,5376	2,32120	,24333	1,70	13,89

Фиг. 5.8

ANOVA. В тази част са представени резултатите от разлагането на дисперсиите, степените на свобода (**df**), емпиричната стойност на **F**-критерия на Фишер (**F**) и съответстващото му равнище на значимост (**Sig.**)

ANOVA

Fatigue index (watts/sec)				
	Sum of Squares	df	Mean Square	F
Between Groups	51,475	3	17,158	3,444
Within Groups	433,443	87	4,982	
Total	484,917	90		

Фиг. 5.9

Multiple Comparisons. В таблицата са представени резултати от междугруповите сравнения. Разликата между **индекса на умората** на нападатели и играчи от средата на терена е -0,3293, като тя е подкрепена с равнище на значимост (**Sig.**) $\alpha=0,954$, което означава, че тя е несъществена.

За удобство на потребителите, статистически значимите разлики са обозначени със звездичка (*). Прегледът на стойностите в таблицата показва, че статистически значими разлики в индекса на умората се наблюдават само за групата на вратарите – те има статистически значимо по-висок индекс на умората от останалите игрови постове.

Multiple Comparisons

Fatigue index (watts/sec)
Tukey HSD

(I) Position	(J) Position	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
forwards	mildfielders	,32930	,63294	,954	-1,3286	1,9872
	defenders	-,23596	,64280	,983	-1,9197	1,4478
	goalkeepers	-3,93234*	1,38669	,029	-7,5646	-,3001
mildfielders	forwards	-,32930	,63294	,954	-1,9872	1,3286
	defenders	-,56526	,53793	,720	-1,9743	,8438
	goalkeepers	-4,26164*	1,34130	,011	-7,7750	-,7483
defenders	forwards	,23596	,64280	,983	-1,4478	1,9197
	mildfielders	,56526	,53793	,720	-,8438	1,9743
	goalkeepers	-3,69638*	1,34598	,036	-7,2220	-,1707
goalkeepers	forwards	3,93234*	1,38669	,029	,3001	7,5646
	mildfielders	4,26164*	1,34130	,011	,7483	7,7750
	defenders	3,69638*	1,34598	,036	,1707	7,2220

*. The mean difference is significant at the 0.05 level.

Фиг. 5.10

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

ПРЕДСТАВЯНЕ НА РЕЗУЛТАТИТЕ ОТ ДИСПЕРСИОНЕН АНАЛИЗ

Таблица 5.4. Сравнителен анализ на индекса на умората на състезатели с различен игрови пост

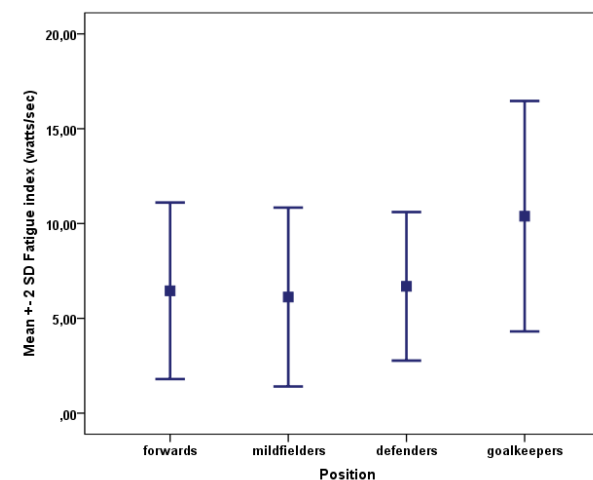
Игрови пост	n	$\bar{x}$	S	F	Sig.
Нападатели	19	6,45	2,33	3,444	0,020
Среда на терена	36	6,12	2,36		
Защитници	33	6,69	1,96		
Вратари	3	10,39*	3,04		
Общо	91	6,58	2,32		

Средните стойности на индекса на умората на полевите играчи са близки по стойност – 6,12 вата/мин. при играчите от средата на терена, следвани от 6,45 вата/мин. при нападателите и 6,69 вата/мин. при защитниците. Групата на вратарите има чувствително най-ниска издръжливост – 10,39 вата/мин. Резултатите от дисперсионния анализ показват, че наблюдаваните различия са статистически значими ($F_{\text{emp}}=3.444$, $\alpha=0.020$).

Критична стойност при $df_b=k-1=3$ и $df_w=n-k=89$ е $F_{0.05} \approx 2,7$

Множествените сравнения, направени с критерия на Тюки показват, че вратарите се отклоняват съществено по своята скоростна издръжливост от полевите играчи.

Фиг. 5.11



Тема 5



ЗАДАЧИ ЗА УПРАЖНЕНИЕ

[Soccer_juniors.sav](#)

[Калкулатор](#)

Задача 5.1

По данните от файла **Soccer_juniors.sav** сравнете резултатите в два теста по избор между футболист от двете възрастови групи.

Показатели	16-годишни			18-годишни			Разлика	Статистическа значимост	
	n_1	$\bar{X}_1$	S_1	n_2	$\bar{X}_2$	S_2	d	t_{emp}	P (t)

Критичната стойност на **t** при $df=n_1+n_2-2=$ и $\alpha=0,05$ е равна на([прил. 4](#))

Анализирайте резултатите.

Задача 5.2

По данните от файла **Soccer_juniors.sav** сравнете резултатите в един тест по избор на футболист с различен игрови пост.

	n	$\bar{X}$	S	As	Ex	F emp	P(F)
Нападатели							
Среда на терена							
Защитници							
Вратари							
Общо							

Критичната стойност на **F** при $df_b=$ и $df_w=$ е равна на $F_{0,05}=$([прил. 5](#)).

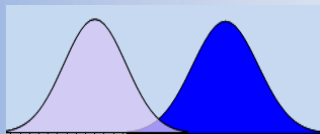
Анализирайте резултатите.

[Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Тема 5

ТЕМА 6. ПРОВЕРКА НА ХИПОТЕЗИ ПРИ ЗАВИСИМИ ИЗВАДКИ

$$H_0: \mu_1 - \mu_2 = 0$$



Изследвани са постиженията на ученици в тест **скок на дължина от място** в началото и края на учебната година.

Средните стойности са съответно $\bar{X}_1=174,3$ см. и $\bar{X}_2=181,7$ см.

Може ли да се счита, че през учебната година са настъпили съществени промени във взривната сила на долните крайници?

В темата се разглеждат начините за описание на величината на прираста на резултатите между две поредни изследвания и неговата статистическа значимост.

Теоретичен материал

- [Абсолютен и относителен прираст](#)
- [t-критерий на Стюдънт за зависими извадки](#)

Практическа част

- [Изчисляване на t-критерия на Стюдънт за зависими извадки с IBM SPSS](#)
- [Резултати от обработката](#)
- [Представяне на резултатите](#)

[Задача за упражнение](#)

ПОНЯТИЯ



Проверка на хипотези при зависими извадки се прилага, когато е необходимо да се проследи промяната на състоянието на изследваните показатели под влияние на естественото развитие на изследваните статистически единици или под влияние на дадено експериментално въздействие.

ЕМПИРИЧНАТА ВЕЛИЧИНА НА ПРОМЯНА на показателите се нарича **прираст на резултатите** и може да бъде изразена в абсолютни или относителни единици.

- **Абсолютният прираст (d)** показва с колко единици, в мерните единици на променливата величина, се променя изучаваното явление.

$$d = \bar{X}_2 - \bar{X}_1 \quad (6.1) \quad \text{Където: } \bar{X}_2 - \text{средна стойност на изучавания показател при крайното, } \bar{X}_1 - \text{при началното изследване.}$$

- **Относителният прираст (d%)** показва с колко процента се променя изучаваното явление .

$$d\% = \frac{d}{\bar{X}_1} \cdot 100 \quad (6.2)$$

СТАТИСТИЧЕСКАТА ЗНАЧИМОСТ НА ПРИРАСТА на резултатите се проверява с подходящ статистически критерий за зависими извадки (виж [фиг. 5.1.](#) и [5.2](#)).

Тема 6

t-КРИТЕРИЙ НА СТЮДЪНТ ЗА ЗАВИСИМИ ИЗВАДКИ

$$H_0: \mu_1 - \mu_2 = 0$$

t-критерият на Стюдънт за зависими извадки се прилага за сравняване на две зависими извадки при количествени признаци, които има нормално разпределение.

Емпиричната стойност на критерия (t_{emp}) се изчислява по формулата:

$$t_{emp} = \frac{|\bar{d}|}{\sqrt{\frac{\sum d^2 - n\bar{d}^2}{n^2 - n}}} \quad (6.3) \quad \text{Където: } d - \text{прираст на резултатите, } n - \text{обем на извадката.}$$

Критичната стойност на критерия (t_α) се определя от [приложение 4](#) при степени на свобода $df=n-1$ и възприетото равнище на значимост ($\alpha=0,05$ или $\alpha=0,01$).

Решение:

Ако $t_{emp} < t_\alpha$ – няма основание за отхвърляне на нулевата хипотеза;
Ако $t_{emp} \geq t_\alpha$ – нулевата хипотеза се отхвърля, което означава, че тестваната разлика е статистически значима.

ПРИМЕР

Изследвани са постиженията на ученици в тест **скок на дължина от място** в началото и края на учебната година.

n	I изследване		II изследване	
	$\bar{X}_1$	S_1	$\bar{X}_2$	S_2
35	174,31	23,08	181,74	25,21

Абсолютният прираст е $d=7,43$ см.

Относителният прираст е $d\%=4,26\%$

Нулевата хипотеза (H_0) гласи, че прирастът на резултатите е недостоверен. За проверка на хипотезата се ползва t-критерият на Стюдънт за зависими извадки. Емпиричната стойност е равна на $t_{emp}=5,78$. [Критичната стойност](#) при $df=34$ и $\alpha=0,05$ е равна на 2,03.

$t_{emp} > t_{0.05}$, което означава, че прирастът на резултатите в хода на учебната година е статистически значим.

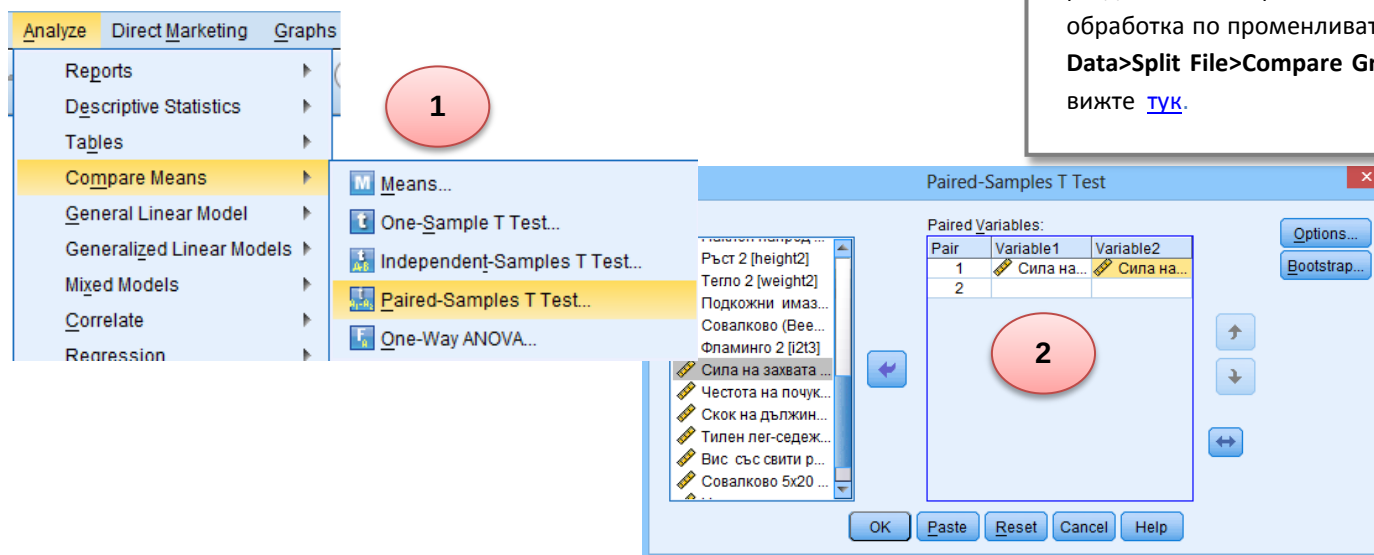
Тема 6



ИЗЧИСЛЯВАНЕ НА t -КРИТЕРИЯ НА СТЮДЪНТ ЗА ЗАВИСИМИ ИЗВАДКИ С IBM SPSS

1 Анализът се задава, като се избере **Analyze>Compare Means>Paired-Samples T Test**.

2 Резултатите от първото и второто изследване се прехвърлят последователно в полето **Paired Variables**, след което се потвърждава с **OK**.



По данните от [EDesign2x2.sav](#)

Да се опише величината и провери статистическата значимост на прираста резултатите в тест „Сила на захвата“ поотделно за спортуващи и неспортуващи ученици.

Необходимо е предварително да се зададе разделяне на файла на групи за самостоятелна обработка по променливата **Group**. За целта изберете **Data>Split File>Compare Groups**. Повече подробности вижте [ТУК](#).

Фиг. 6.1

Тема 6

РЕЗУЛТАТИ ОТ ОБРАБОТКАТА

Paired Samples Statistics

Група			Mean	N	Std. Deviation	Std. Error Mean
Спортуващи	Pair 1	Сила на захвата 1	27,56	16	5,501	1,375
		Сила на захвата 2	32,13	16	6,076	1,519
Неспортуващи	Pair 1	Сила на захвата 1	25,32	19	6,438	1,477
		Сила на захвата 2	26,74	19	5,848	1,342

Разпечатката съдържа 3 части:

Paired Sample Statistics. В първата част е направена характеристика на разпределението на стойностите в двете извадки (при двете изследвания), съответно: средна стойност (**Mean**), брой на изследваните (N), стандартно отклонение (**Std. Deviation**) и стандартна грешка (**Std. Error Mean**).

Фиг. 6.2

Paired Samples Correlations. Извадките са зависими, което означава, че между стойностите при двете изследвания трябва да има статистически значима зависимост. В тази част от разпечатката се проверява тази хипотеза. Публикува се емпиричният коефициент на корелация (**Correlation**) и неговата статистическа значимост (**Sig.**).

Paired Samples Correlations

Група			N	Correlation	Sig.
Спортуващи	Pair 1	Сила на захвата 1 & Сила на захвата 2	16	,971	,000
Неспортуващи	Pair 1	Сила на захвата 1 & Сила на захвата 2	19	,963	,000

Фиг. 6.3

Paired Samples Test. В третата част от разпечатката се проверява хипотезата за равенство на средните стойности. Прирастът на резултатите се публикува в колона **Mean**, емпиричната стойност на t-критерия на Стьудънт (t_{emp}) – в колона **t**, а съответстващото му равнище на значимост – в колона **Sig.(2-tailed)**.

Paired Samples Test

Група			Paired Differences					t	df	Sig. (2-tailed)
			Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
						Lower	Upper			
Спортуващи	Pair 1	Сила на захвата 1 - Сила на захвата 2	-4,563	1,504	,376	-5,364	-3,761	-12,133	15	,000
Неспортуващи	Pair 1	Сила на захвата 1 - Сила на захвата 2	-1,421	1,774	,407	-2,276	-,566	-3,492	18	,003

Фиг. 6.4

ПРЕДСТАВЯНЕ НА РЕЗУЛТАТИТЕ

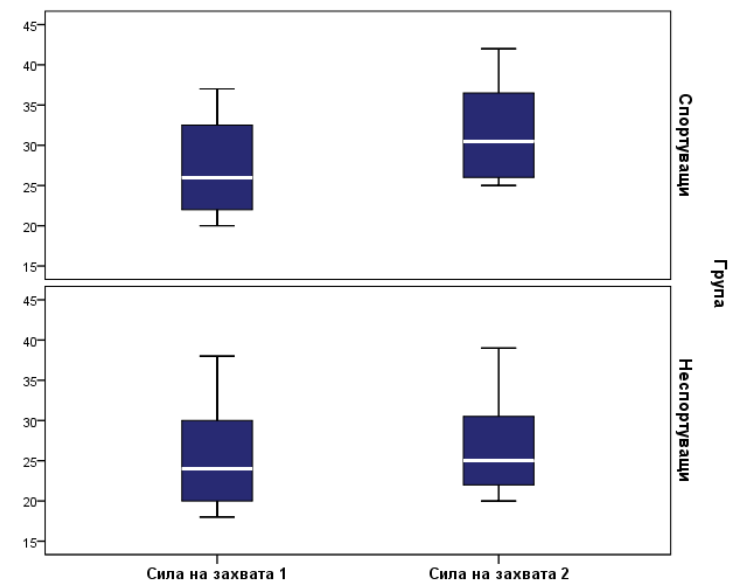
Таблица 6.1. Развитие на резултатите между двете изследвани в тест „Сила на захвата“

Показател	n	I изследване		II изследване		Прираст		Статистическа значимост	
		$\bar{X}_1$	S_1	$\bar{X}_2$	S_2	d	d%	t_{emp}	P (t)
Спортуващи	16	27,56	5,50	32,13	6,08	4,56	16,55	12,13	100,00
Неспортуващи	19	25,32	6,44	26,74	5,85	1,42	5,61	3,49	99,74

Прирастът на резултатите, който реализират спортуващите ученици е 4,56 кг. (16,55%), като проверката с t-критерия на Стюдънт за зависими извадки показва, че той е статистически значим (P=12,13, P(t)=100%).

С по-малка величина е прирастът на резултатите на неспортуващите - 1,42 кг. (5,61%), но също е статистически значим (P(t)=99,74%).

Това означава, че в изследвания възрастов период силата на горните крайници се увеличава, като в случай, че учениците спортуват – темповете на нарастване са по-високи.



Фиг. 6.5



ЗАДАЧИ ЗА УПРАЖНЕНИЕ

ДАННИ ЗА УПРАЖНЕНИЕТО:

[Edesign2x2.sav](#)

[Калкулатор](#)

Задача 6.1

По данните от файла **EDesign2x2.sav** установете статистическата значимост на прираста на резултатите между първо и второ изследване (отделно за спортуващи и неспортуващи) на променлива по избор.

УКАЗАНИЯ

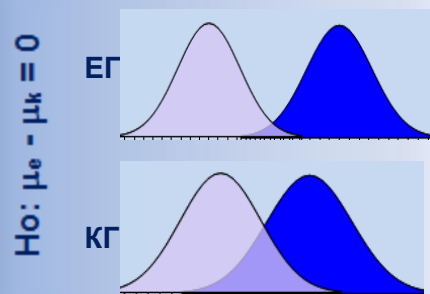
Разделете файла на групи за самостоятелна обработка (по подробно вижте [тук](#)) и след това приложете алгоритъма за проверка на хипотези с t-критерия за зависими извадки по начина, описан [тук](#).

Показател	n	I изследване		II изследване		Прираст на резултатите			
		$\bar{X}_1$	S_1	$\bar{X}_2$	S_2	d	d%	t	P (t)
Спортуващи									
Неспортуващи									

Анализирайте резултатите.

ТЕМА 7. ОБРАБОТКА НА ДАННИ ОТ ЕКСПЕРИМЕНТ С ДВЕ ГРУПИ И ДВЕ ИЗСЛЕДВАНИЯ

$$H_0: \mu_1 - \mu_2 = 0$$



В настоящата тема се описва как се обработват данни от експеримент с две групи и две изследвания.

Теоретичен материал

- [Описание на експерименталната схема](#)

Практическа част

- [Обработка на данните с IBM SPSS](#)
- [Задача за упражнение](#)



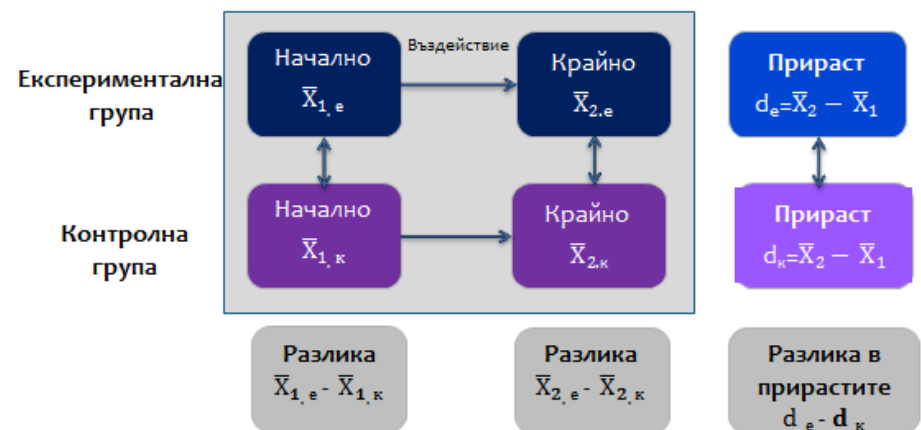
ОПИСАНИЕ НА ЕКСПЕРИМЕНТАЛНАТА СХЕМА

Широко застъпена проблематика в научните изследвания в областта на спорта е разработването на нови, посъвършени средства и методи за въздействие (тренировъчни методики, учебно съдържание във физическото възпитание, кинезитерапия и др.). За установяване на тяхната ефективност се провежда експеримент, който трябва да опише величината и статистическата значимост на промените, които настъпват в изучаваните показатели **при наличието** и **при отсъствието** на експериментираното въздействие. Най-простата и същевременно най-често застъпената схема на експерименти с такава насоченост включва формирането чрез случаен подбор на две групи:

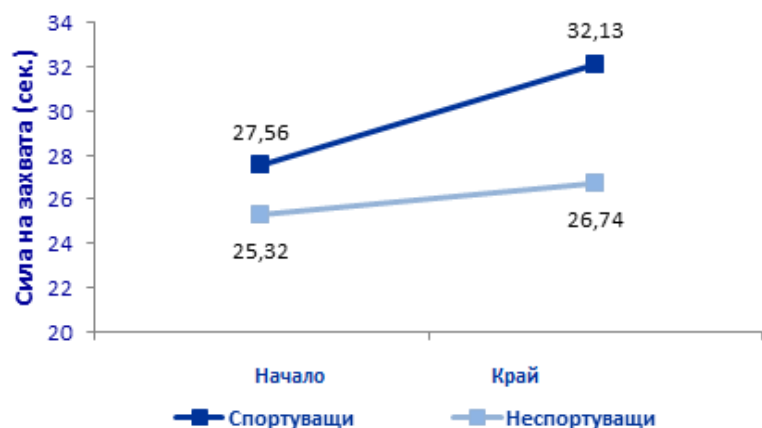
- **Експериментална**, в която се прилага новото въздействие;
- **Контролна**, която е идентична с експерименталната във всяко отношение, с изключение на това, че в нея не се прилага експериментираното въздействие.

В зависимост от насочеността на експеримента, изследователят дефинира показатели, по които може да се долови ефектът от приложените въздействия. За установяване на тяхното развитие на в хода на експеримента се провеждат поне две изследвания:

- **Начално** - в което се установява изходното състояние на показателите;
- **Крайно** – в което се регистрира заключителното състояние на изследваните показатели.



Фиг. 7.1



Фиг. 7.2

Частично резултати от подобен експеримент бяха разгледани в тема 6 и са представени на фиг. 7.2. Нека приемем, че спортуващите са от експерименталната (ЕГ), а неспортуващите – от контролната група (КГ).

От фигурата ясно се вижда, че в началото на експеримента двете групи имат близки резултати, в хода на експеримента и двете групи подобряват постижението си, но нарастването на теста при ЕГ е по-ясно изразено. При второто изследване спортуващите имат чувствително по високи резултати от неспортуващите. Всичко това налага предположението, че методиката, прилагана в ЕГ е по-ефективна.

Това е научна хипотеза, за проверката на която трябва да се премине през следните сравнения:

Изследване Група	n	Начало		Край		Прираст		Статистическа значимост	
		$\bar{X}_1$	S_1	$\bar{X}_2$	S_2	d	d%	t_{emp}	P(t)
Експериментална		Сравнение 1							
Контролна									
Разлика									
Статистическа значимост на разликата	t_{emp}								
	P(t)								

Фиг. 7.3

Сравнения 1 и 2. На първо място трябва да се установи величината и статистическата значимост на прираста на резултатите във всяка от групите.

- **Величината на прираста** се установява с показателите абсолютен (d) и относителен прираст (d%), описани в тема 6.
- **Статистическата значимост на прираста на резултатите** се установява с параметричен или непараметричен критерий за проверка на хипотези за зависими извадки, описани в тема 6.

Сравнения 3 и 4. На следващо място е необходимо да се сравни състоянието на изследваните признаци в двете групи **при всяко от изследванията**.

Като правило, за да се сравнява ефектът от въздействия при различни групи от изследвани е необходимо в началото на експеримента резултатите на **ЕГ** и **КГ** да бъдат на едно и също ниво. Чрез случайния подбор при формиране на групите се цели именно такова

сходство. Известно е, обаче, че случаен подбор се прилага рядко, поради което е необходимо да се сравни **началното** състояние на изследваните показатели (**сравнение 3**). Добре би било, ако разликата между групите в началото на експеримента е несъществена.

Изследване Група	n	Начало		Край		Прираст		Статистическа значимост	
		$\bar{X}_1$	S_1	$\bar{X}_2$	S_2	d	d%	t_{emp}	P(t)
Експериментална		Сравнение 3		Сравнение 4					
Контролна									
Разлика									
Статистическа значимост на разликата	t_{emp}								
	P(t)								

Фиг. 7.4

Един от възможните начини да се установи ефективността на приложените въздействия е да се сравни състоянието на изучаваните показатели в края на експеримента (**сравнение 4**). Величината на разликите се описва с аритметичната разлика между средните стойности на групите, а статистическата значимост – с някой от критериите за сравняване на две независими извадки.

Сравнение 5. Обобщаваща информация за развитието на показателите под влияние на експериментираното въздействие се получава, като се сравни **прирастът** на резултатите в двете групи. За целта:

- Изчислява се производна променлива с разликите между **второ** и **първо** изследване

$$d_i = X_{2,i} - X_{1,i}$$

- Проверява се статистическата значимост на разликите между прираста на резултатите в двете групи (**d**_{експериментална} и **d**_{контролна}) с някой от критериите за сравняване на две независими извадки.

Изследване Група	n	Начало		Край		Прираст		Статистическа значимост	
		$\bar{X}_1$	S ₁	$\bar{X}_2$	S ₂	d	d%	t _{emp}	P(t)
Експериментална									
Контролна									
Разлика									
Статистическа значимост на разликата	t _{emp}								
	P(t)								

Фиг. 7.5

На фиг. 7.6 са представени резултатите от обработката на данните от примера в тема 6. Обозначени са и петте сравнения.

Коментарът на първите две сравнения е направен в тема 6, като основният извод е, че прирастът на резултатите в **експерименталната група (ЕГ)** е 4,56 сек. (4,56%) е подкрепен с близка до 100% гаранционна вероятност. По малък по величина, но също с висока статистическа достоверност е прирастът на **контролната група (КГ)**. По този начин не може да се намери отговор на научната хипотеза, че въздействието, прилагано в експерименталната група е с по-висока ефективност от това в контролната група.

Показател	n	I изследване		II изследване		Прираст		Статистическа значимост	
		$\bar{X}_1$	S_1	$\bar{X}_2$	S_2	d	d%	t_{emp}	P (t)
Спортуващи	16	27,56	5,50	32,13	6,08	4,56	16,55	12,13	100,00
Неспортуващи	19	25,32	6,44	26,74	5,85	1,42	5,61	3,49	99,74
Разлика		2,25		5,39		3,14			
Статистическа значимост	t	1,10		2,67		5,59			
	P(t)	71,99		98,83		100,00			

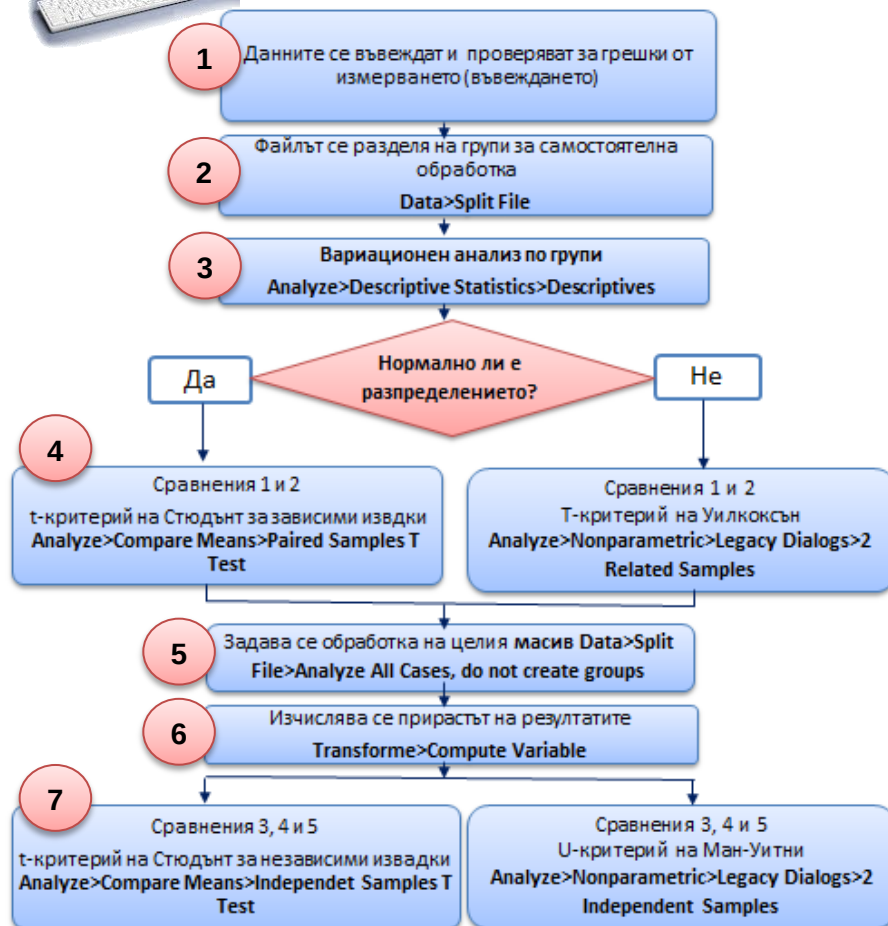
Фиг. 7.6

Сравнението на средните стойностите на групите в началото на експеримента (сравнение 3) показва, че разликата от 2,25 сек. е несъществена ($t_{emp}=1,1$, $P(t)=71,99\%$). За разлика от това, в края на експеримента **ЕГ** има достоверно ($t_{emp}=2,67$, $P(t)=98,83\%$) по-високо постижение в тест „сила на захвата“ (сравнение 4). Обобщаваща информация за проверка на научната хипотеза се получава от сравнение 5. В хода на експеримента **ЕГ** реализира съществено по-голям прираст ($d_e=4,56$ сек.) от този, показан от **КГ** ($d_k=1,42$ сек.). Разликата между прирастите от 3,14 сек. е статистически значима ($t_{emp}=5,59$, $P(t)\approx 100\%$).



ОБРАБОТКА НА ДАННИ ОТ ЕКСПЕРИМЕНТ С ДВЕ ГРУПИ И ДВЕ ИЗСЛЕДВАНИЯ С IBM SPSS

При обработката на количествени променливи, които имат нормално разпределение се ползват алгоритми, които са описани вече в помагалото, поради което ще се спрем на тях накратко.



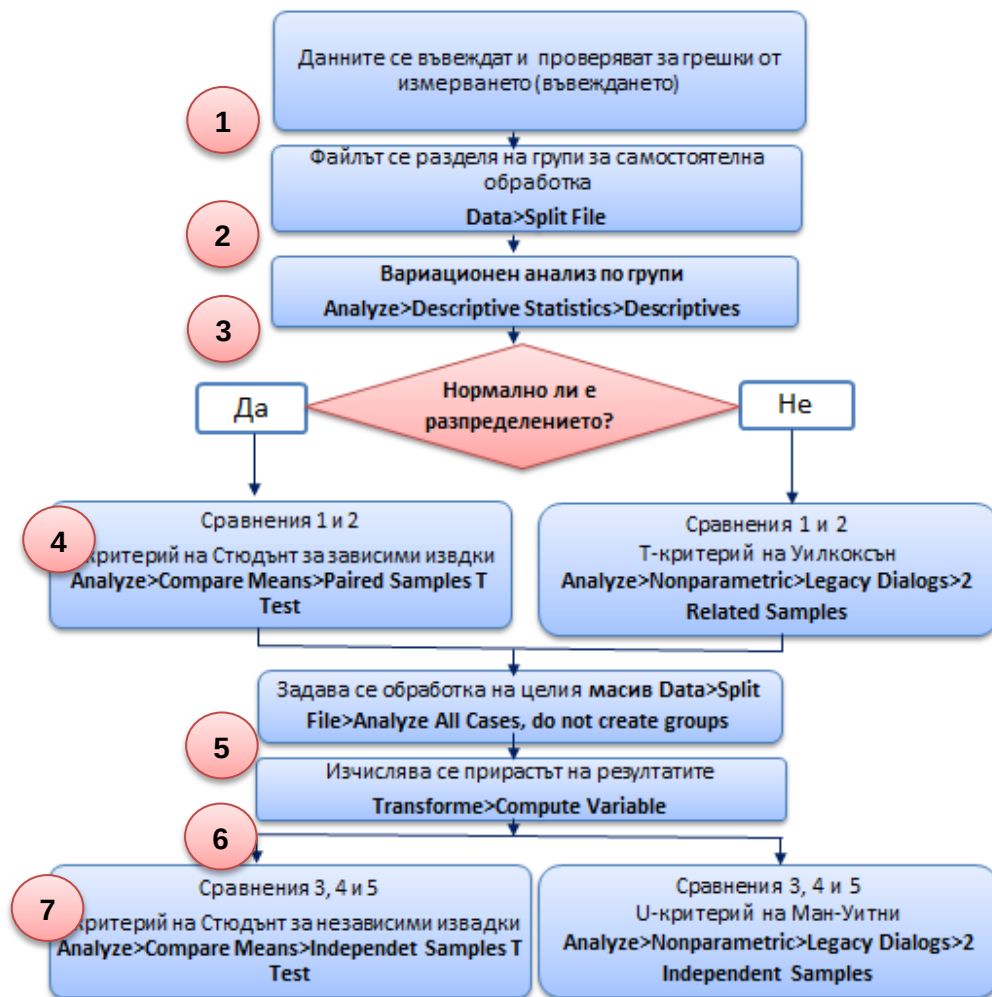
Фиг. 7.7

❶ Данните за принадлежността към групата се въвеждат в една колона, а резултатите от поредните изследвания – като две самостоятелни колони, като се внимава да не се разместят резултатите на изследваните лица. Проверката за грешки се прави на базата на логически критерии и статистическа обработка (начертава се бокс-плот диаграма).

❷ Файлът се разделя на групи за самостоятелна обработка, по начина, показан [тук](#). Уместно е да се зададе опцията **Compare Means**.

❸ Прилага се вариационен анализ на резултатите при двете изследвания поотделно за двете групи. Ако разпределението на стойностите е нормално, в по нататъшната обработка се ползват параметрични критерии за проверка на хипотези.

❹ Изчисляват се t-критерият на Стьюдънт за зависими извадки, и относителният прираст (d%). Тъй като файлът е разделен на групи за самостоятелна обработка, анализът ще бъде направен поотделно за експериментална и контролна група. Обработката до момента дава възможност да се попълнят първите два реда на таблицата, показана на фиг. 7.6.



Фиг. 7.7

5 Разделянето на файла на групи за самостоятелна обработка се маха, като се избере **Data>Split File**, в отворилия се прозорец се зададе **Analyze all cases, do not create groups** и се потвърди с **OK**.

6 За да може едновременно да се направят сравнения 3, 4 и 5 е необходимо предварително да се изчисли прирастът (d) на резултатите. Това става, като се избере **Transform>Compute Variable**. В полето **Numeric Expression** се задава разликата между двете изследвания, а в полето **Target Variable** – името на новата променлива.

7 Избира се **Analyze>Compare Means>Independent Samples T Test**. Резултатите от първото, второто изследване и прирастът на резултатите се изпращат в полето **Test Variable (s)**, а групиращата променлива – в полето **Grouping Variable**. Натиска се бутонът **Define Groups** и се задават кодовете, с които е въведена принадлежността към групите. След осъществяване на анализа могат да се попълнят последните три реда от таблицата, показана на фиг. 7.6.



ЗАДАЧИ ЗА УПРАЖНЕНИЕ

ДАННИ ЗА УПРАЖНЕНИЕТО:

[Edesign2x2.sav](#)

[Калкулатор](#)

Задача 7.1

Изследване Група	n	Начало		Край		Прираст		Статистическа значимост	
		$\bar{X}_1$	S_1	$\bar{X}_2$	S_2	d	d%	t_{amp}	P(t)
Експериментална									
Контролна									
Разлика									
Статистическа значимост на разликата	t_{amp}								
	P(t)								

Обработете резултатите за признак по избор по данните от файла Edesign2x2.sav.

Анализирайте резултатите

. [Начало на учебника](#) [Тема 1](#) [Тема 2](#) [Тема 3](#) [Тема 4](#) [Тема 5](#) [Тема 6](#) [Тема 7](#)

Приложение 1. Критични стойности на коефициентите на асиметрия (As) и ексцес (Ex) при равнище на значимост $\alpha=0,05$

Обем на извадката (n)	$As_{0,05}$	$Ex_{0,05}$	Обем на извадката (n)	$As_{0,05}$	$Ex_{0,05}$
10	1,374	2,668	150	0,396	0,787
15	1,160	2,242	200	0,344	0,684
20	1,024	1,985	250	0,308	0,614
25	0,927	1,803	300	0,281	0,561
30	0,854	1,665	350	0,261	0,520
35	0,795	1,556	400	0,244	0,487
40	0,748	1,465	450	0,230	0,459
45	0,707	1,389	500	0,218	0,436
50	0,673	1,324	600	0,200	0,398
55	0,643	1,267	700	0,185	0,369
60	0,617	1,217	800	0,173	0,345
65	0,594	1,172	900	0,163	0,326
70	0,574	1,133	1000	0,155	0,309
75	0,555	1,096	1100	0,148	0,295
80	0,538	1,064	1200	0,141	0,282
85	0,522	1,034	1300	0,136	0,271
90	0,508	1,006	1400	0,131	0,261
95	0,495	0,980	1500	0,126	0,253
100	0,483	0,957	2000	0,109	0,219

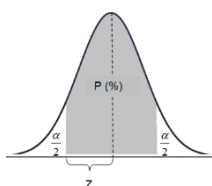
[Назад към теорията за \$As\$](#)

[Назад към теорията за \$Ex\$](#)

[Назад към практиката](#)

[Назад към задачата](#)

Приложение 2. Нормирано нормално разпределение в % (централен процент от случаите)



U (Z)	P%	α	U (Z)	P%	α	U (Z)	P%	α	U (Z)	P%	α
0,08	6,38	0,94	0,88	62,11	0,38	1,68	90,70	0,09	2,48	98,69	0,01
0,12	9,55	0,90	0,92	64,24	0,36	1,72	91,46	0,09	2,52	98,83	0,01
0,16	12,71	0,87	0,96	66,29	0,34	1,76	92,16	0,08	2,58	99,01	0,01
0,20	15,85	0,84	1,00	68,27	0,32	1,80	92,81	0,07	2,60	99,07	0,01
0,24	18,97	0,81	1,04	70,17	0,30	1,84	93,42	0,07	2,64	99,17	0,01
0,28	22,05	0,78	1,08	71,99	0,28	1,88	93,99	0,06	2,68	99,26	0,01
0,32	25,10	0,75	1,12	73,73	0,26	1,92	94,51	0,05	2,72	99,35	0,01
0,36	28,12	0,72	1,16	75,40	0,25	1,96	95,00	0,05	2,76	99,42	0,01
0,40	31,08	0,69	1,20	76,99	0,23	2,00	95,45	0,05	2,80	99,49	0,01
0,44	34,01	0,66	1,24	78,50	0,21	2,04	95,86	0,04	2,84	99,55	0,00
0,48	36,88	0,63	1,28	79,95	0,20	2,08	96,25	0,04	2,88	99,60	0,00
0,52	39,69	0,60	1,32	81,32	0,19	2,12	96,60	0,03	2,92	99,65	0,00
0,56	42,45	0,58	1,36	82,62	0,17	2,16	96,92	0,03	2,96	99,69	0,00
0,60	45,15	0,55	1,40	83,85	0,16	2,20	97,22	0,03	3,00	99,73	0,00
0,64	47,78	0,52	1,44	85,01	0,15	2,24	97,49	0,03	3,04	99,76	0,00
0,68	50,35	0,50	1,48	86,11	0,14	2,28	97,74	0,02	3,08	99,79	0,00
0,72	52,85	0,47	1,52	87,15	0,13	2,32	97,97	0,02	3,12	99,82	0,00
0,76	55,27	0,45	1,56	88,12	0,12	2,36	98,17	0,02	3,16	99,84	0,00
0,80	57,63	0,42	1,60	89,04	0,11	2,40	98,36	0,02	3,20	99,86	0,00
0,84	59,91	0,40	1,64	89,90	0,10	2,44	98,53	0,01	3,24	99,88	0,00

Приложение 3. t-разпределение на Стюдънт в %

Степени на свобода df=n-1																	
t	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2,0	88,4	89,8	90,8	91,4	91,9	92,3	92,7	92,9	93,1	93,3	93,5	93,6	93,7	93,8	93,9	94,0	94,1
2,1	89,6	91,0	92,0	92,6	93,1	93,5	93,8	94,0	94,2	94,4	94,6	94,7	94,8	94,9	95,0	95,1	95,1
2,2	90,7	92,1	93,0	93,6	94,1	94,5	94,8	95,0	95,2	95,4	95,5	95,6	95,7	95,8	95,9	96,0	96,0
2,3	91,7	93,0	93,9	94,5	95,0	95,3	95,6	95,8	96,0	96,1	96,3	96,4	96,5	96,6	96,6	96,7	96,8
2,4	92,6	93,8	94,7	95,3	95,7	96,0	96,3	96,5	96,6	96,8	96,9	97,0	97,1	97,2	97,3	97,3	97,4
2,5	93,3	94,6	95,3	95,9	96,3	96,6	96,9	97,0	97,2	97,3	97,5	97,5	97,6	97,7	97,8	97,8	97,9
2,6	94,0	95,2	95,9	96,5	96,8	97,1	97,4	97,5	97,7	97,8	97,9	98,0	98,1	98,1	98,2	98,2	98,3
2,7	94,6	95,7	96,4	96,9	97,3	97,6	97,8	97,9	98,1	98,2	98,3	98,4	98,4	98,5	98,5	98,6	98,6
2,8	95,1	96,2	96,9	97,3	97,7	97,9	98,1	98,3	98,4	98,5	98,6	98,7	98,7	98,8	98,8	98,9	98,9
2,9	95,6	96,6	97,3	97,7	98,0	98,2	98,4	98,6	98,7	98,8	98,8	98,9	99,0	99,0	99,0	99,1	99,1
3,0	96,0	97,0	97,6	98,0	98,3	98,5	98,7	98,8	98,9	99,0	99,0	99,1	99,2	99,2	99,2	99,3	99,3
3,1	96,4	97,3	97,9	98,3	98,5	98,7	98,9	99,0	99,1	99,2	99,2	99,3	99,3	99,3	99,4	99,4	99,4
3,2	96,7	97,6	98,1	98,5	98,7	98,9	99,1	99,2	99,2	99,3	99,4	99,4	99,4	99,5	99,5	99,5	99,6
3,3	97,0	97,9	98,4	98,7	98,9	99,1	99,2	99,3	99,4	99,4	99,5	99,5	99,5	99,6	99,6	99,6	99,6
3,4	97,3	98,1	98,6	98,9	99,1	99,2	99,3	99,4	99,5	99,5	99,6	99,6	99,6	99,7	99,7	99,7	99,7
3,5	97,5	98,3	98,7	99,0	99,2	99,3	99,4	99,5	99,6	99,6	99,6	99,7	99,7	99,7	99,7	99,8	99,8
3,6	97,7	98,4	98,9	99,1	99,3	99,4	99,5	99,6	99,6	99,7	99,7	99,7	99,8	99,8	99,8	99,8	99,8
3,7	97,9	98,6	99,0	99,2	99,4	99,5	99,6	99,6	99,7	99,7	99,8	99,8	99,8	99,8	99,8	99,8	99,9
3,8	98,1	98,7	99,1	99,3	99,5	99,6	99,7	99,7	99,7	99,8	99,8	99,8	99,8	99,9	99,9	99,9	99,9

[Назад към теорията за независими извадки](#)
[Назад към теорията за зависими извадки](#)

[Назад към практиката](#)
[Назад към практиката](#)

[Назад към задачите](#)
[Назад към задачите](#)

Приложение 4. Критични стойности на t-критерия на Стюдънт

Степени на свобода (df)	Равнище на значимост (α)		Степени на свобода (df)	Равнище на значимост (α)	
	0,05	0,01		0,05	0,01
4	2,78	4,60	21	2,08	2,83
5	2,57	4,03	22	2,07	2,82
6	2,45	3,71	23	2,07	2,81
7	2,36	3,50	24	2,06	2,80
8	2,31	3,36	25	2,06	2,79
9	2,26	3,25	26	2,06	2,78
10	2,23	3,17	27	2,05	2,77
11	2,20	3,11	28	2,05	2,76
12	2,18	3,05	29	2,05	2,76
13	2,16	3,01	30	2,04	2,75
14	2,14	2,98	40	2,02	2,70
15	2,13	2,95	50	2,01	2,68
16	2,12	2,92	60	2,00	2,66
17	2,11	2,90	80	1,99	2,64
18	2,10	2,88	100	1,98	2,63
19	2,09	2,86	120	1,98	2,62
20	2,09	2,85	200	1,97	2,60

- При зависими извадки $df=n-1$
- При независими извадки $df=n_1+n_2-2$

Приложение 5. Критични стойности на F-критерия на Фишер при равнище на значимост $\alpha=0,05$

Степени на свобода	Междугрупова дисперсия $df_w = k-1$									
df_2	1	2	3	4	5	6	8	12	24	∞
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,52
60	4,00	3,16	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Приложение 6. Критични стойности на коефициента на обикновена линейна корелация на Пирсън (r)

Степени на свобода $df=n-2$	Равнище на значимост (α)		Степени на свобода $df=n-2$	Равнище на значимост (α)	
	0,05	0,01		0,05	0,01
5	0,754	0,875	25	0,381	0,487
6	0,707	0,834	26	0,374	0,479
7	0,666	0,798	27	0,367	0,471
8	0,632	0,765	28	0,361	0,463
9	0,602	0,735	29	0,355	0,456
10	0,576	0,708	30	0,349	0,449
11	0,553	0,684	35	0,325	0,418
12	0,532	0,661	40	0,304	0,393
13	0,514	0,641	45	0,288	0,372
14	0,497	0,623	50	0,273	0,354
15	0,482	0,606	60	0,250	0,325
16	0,468	0,590	70	0,232	0,302
17	0,456	0,575	80	0,217	0,283
18	0,444	0,561	90	0,205	0,267
19	0,433	0,549	100	0,195	0,254
20	0,423	0,537	200	0,138	0,181
21	0,413	0,526			
22	0,404	0,515			
23	0,396	0,505			
24	0,388	0,496			

Приложение 7. Критични стойности на коефициента на рангова корелация на Спирмън

Степени на свобода df=n-2	Равнище на значимост (α)		Степени на свобода df=n-2	Равнище на значимост (α)	
	0,05	0,01		0,05	0,01
6	0,89	-	21	0,44	0,56
7	0,79	0,93	22	0,43	0,55
8	0,74	0,88	23	0,42	0,54
9	0,68	0,83	24	0,41	0,53
10	0,65	0,79	25	0,40	0,52
11	0,61	0,77	26	0,39	0,51
12	0,59	0,75	27	0,38	0,50
13	0,56	0,71	28	0,38	0,49
14	0,54	0,69	29	0,37	0,48
15	0,52	0,66	30	0,36	0,47
16	0,51	0,64	31	0,36	0,46
17	0,49	0,62	32	0,36	0,45
18	0,48	0,61	33	0,34	0,45
19	0,46	0,60	35	0,33	0,43
20	0,45	0,58	40	0,31	0,40

Приложение 8. Описание на файловете с данни

В процеса на обучение и самостоятелна работа се ползват следните файлове с данни:

EDesign2x2. Изследвани са 14 годишни момчета, обособени в две групи – експериментална, която се занимава с ветроходство и контролна, които не спортуват. Проведени са две изследвания – в началото и края на учебната година. Използвана е тестовата батерия Еврофит. Описанието на тестовите фигурира във файла с данни.

Survey 1.sav. Проведено е анкетно проучване със студенти от ОКС „Магистър“ на НСА по въпроси, разпределени в две основни направления - желание и възможности за включване в дистанционно обучение и предпочитани стилове на обучение. Въпросите и възможните отговори може да откриете във файла на SPSS или на съответната страница на екселовия калкулатор.

Soccer juniors.sav. Изследвани са млади футболисти от две възрастови групи (16 и 18 годишни) , с различна националност и игрови пост по 4 теста . спринтово бягане - 30 м., совалково бягане (бийп тест), на базата на което е изчислена кислородната консумация (VO2max), совалково бягане 6х35, при изпълнението на който са отчетени:

- времето за изпълнение на теста (RAST)
- индекс на умората (fatigue Index), който се получава като разликата между максималната и минималната мощност, реализирана при пробягване на отсечките, се раздели на общото време за изпълнение на теста
 - ако $FI < 10$ - скоростната издръжливост е на добро ниво;
 - ако $FI > 10$ - скоростната издръжливост е на ниско ниво;

Soccer teams .sav. Изследвани са физическата дееспособност, игровата ефективност и класирането в първенство то на 12 младежки футболни отбора, както следва:

- **Физическа дееспособност** – 20 м. гл. б. (**t1**), вертикален отскок (**t2**), подвижност (**t3** – тест „Стрела“, който описва способността за смяна на посоката), обща издръжливост (**t4** - YoYo тест, който представлява совалково бягане до отказ по зададено от звуков сигнал темпо) и оценка за физическа дееспособност (**SPARQ**).
- **Състезателна реализация** – брой точки средно на мач (**points_match**), вкарани голове средно на мач (**G_for**), допуснати голове средно на мач (**G_ag**), отношение вкарани-допуснати голове (**G_for/def**)
- **Класиране (Rank)** – класиране на отборите в първенството, в което участват.

Rowing.sav. Изследвани са елитни гребци (мъже и жени), които изпълняват тест 4х500 на гребен ергометър. Отчетени са темпа, мощността и скоростта на гребане, пулсът, лактатът и индексът на енергетичния оптимум, реализирани при всяка от отсечките. Проведени са две изследвания – в началото и края на експеримент за повишаване на тяхната функционална работоспособност.