# СТАТИСТИКА С SPSS

SPSS е специализиран софтуер за обработка и статистически анализ на данни.

## 1. ВЪВЕЖДАНЕ НА ДАННИТЕ

***Статистически признак и променлива***

Признакът е интересуваща ни характеристика (качество, свойство, проява) на изучаваната статистическа съвкупност.

Признакът присъства в инструментариума на статистическото изследване под формата на въпрос или твърдение, а те от своя страна като променливи в масива от данни изграждан при обработката на информацията.

За обработката на въпрос с един отговор се дефинира една променлива.

За въпроси с повече от един отговор (множествени отговори/въпроси), за всеки отговор се дефинира отделна променлива.

Въвеждането на данните става в меню Data View на SPSS, след като се създаде макет.

Създаването на макета се извършва в меню Variable View и включва въвеждане/определяне на:

- ***Name*** – кратко (кодирано) наименование на променливата. Обикновено то се състой от една буква и число, идентично с това на номера на въпроса. Имената не могат да се повтарят; не могат да започва с цифра; не могат да съдържат определени символи (-, % и др.);
- ***Type*** – определя характера на променливата. Основните типове променливи са:
    - ***Numeric*** – ако характерът на информацията е числов. За тази променлива трябва да се зададат: *Width* – позициите необходими за въвеждане на информацията (брой се и запетаята, ако числото е десетично), *Decimal Places* – знаците след запетаята;
    - ***String*** – ако се въвежда текст, като в *Characters* се задава очакваната максимална дължината на записа.
- ***Label*** – етикета на променливата и на практика това е текста на въпроса. При множествен въпрос, етикетите на променливите дефинирани за съответните отговори са еднакви.
- ***Values*** – етикети на отговорите на променлива на номинална и рангова скала (променливи чиито отговори се дефинират словесно, а цифрите им се присвояват само за да могат бъдат обработени). В полето Value се въвежда цифрата, която е присвоена на отговора, а в полето Value Label се изписва самия текст на отговора. С Add присвояваме на дадената цифра въведения текст и се записва в празното полето. Продължава се по същия начин за останалите използвани цифри. Ако променливата е количествена, етикети не се въвеждат. Ако променливата е стрингова тази команда не е достъпна.

Всяка от тези характеристики може да бъде лесно коригирана.

Създаването на макета за въвеждане на данните е възможно и чрез Syntax file, който се създава чрез SPSS > File >New > Syntax …

*Пример:* Синтаксис файлът за въпросите от Приложение 1 има следното съдържание (текстът в <….> е пояснение и не се пише в синтаксиса)

---

```
Data list/
Num 1-3                 <максималният отговор е трицифрено число>
q1 1                    <максималният отговор е едноцифрена число>
q2.1 to q2.9 1-9        <дефинират се 9 броя едноцифрени променливи>
q2.10 1-2               <дефинира се 1 двуцифрена променлива >
q2.9_txt (A200)         <дефинира се текстова/стрингова променлива с максимален брой на символите 300>
q3.1 to q3.5 1-5        <дефинират се 5 броя  едноцифрени променливи>
q4.1 q4.2 1-6.          <дефинират се 2 броя трицифрени променливи>

Begin data

End data.
List.

Var lab num 'Пореден номер на анкетната карта'.

Var lab q1 'Националност на фирмата'.
Val lab q1
1 'България'
2 'Македония'
3 'Румъния'.

Var lab q2.1 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.2 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.3 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.4 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.5 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.6 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.7 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.8 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.9 'В какви конкурентни отношения Вашата фирма участва?'.
Var lab q2.10 'В какви конкурентни отношения Вашата фирма участва?'.

Val lab q2.1 to q2.10
1 'за цената на определен продукт/услуга'
2 'за качеството на определен продукт/услуга'
3 'при лансирането на нови продукти/услуги'
4 'при възлагане на поръчки/проекти'
5 'в достъпа до определени суровини и компоненти'
6 'при наемането на квалифициран персонал'
7 'при получаване на субсидии от държавни фондове'
8 'в областта на развойна дейност'
9 'Друго'
10 'Не мога да преценя'.

var lab q2.9_txt 'Други конкурентни отношения'.

*3. Как оценявате влиянието на следните фактори върху перспективите за развитие на Вашия бизнес?. <* означава, че това не е команда, т.е. не се изпълнява>

Var lab q3.1 'Хармонизирането на българското законодателство с европейското'.
Var lab q3.2 'Либерализацията на националния пазар'.
```

## 2. ПРЕОБРАЗУВАНЕ НА ДАННИТЕ

***Подбор на отделни случаи*** – при определени обстоятелства се налага да се работи не с целия масив от данни, а само с определени случаи, отговарящи на някакви критерии. За да се подберат тези случаи се използва функцията/командата Select Cases от Data менюто. В рамките на тази функция се задават условията за селекция, както и режима на отстраняване (временно – филтриране или трайно-изтриване).

***Обединяване на файлове*** – чрез функцията Merge Files в меню Data. С командата Add cases се обединяват случаи (хоризонтално лепене на файлове), а с Add variables променливи (вертикално лепене на файлове)

***Преобразуване на скали*** – извършва се с функцията Recode от меню Transform, където има възможност преобразованията да станат в изходната променлива (Into Same Variable) или в нова променлива (Into Different Variable). За предпочитане е в нова, защото се запазва оригиналната и тя може да се подложи и на други преобразования.

*Пример:* Преобразуване на отговорите на въпрос 4.1 от Приложение 1, на скалата:

1. 0%
2. до 5%
3. над 5 до 20%
4. над 20%

- Въвежда се името на новата променлива в полето Output variable, например q4.1_rec
- Определят се правилата за рекодиране в Old and New Value:

| Old Value | New Value |
|---|---|
| Value 0 | Value 1 |
| Range 1 through 5 | Value 2 |
| Range 5 through 20 | Value 3 |
| Range 20 through highest | Value 4 |

Тази команда има следния синтаксис:

RECODE q4.1 (0=1) (1 thru 5=2) (5 thru 20=3) (20 thru 100=4) INTO q4.1_rec.

***Преобразуване на променливи*** - създаване на нова променлива въз основа на вече съществуващи променливи с функцията Compute от меню Transform.

*Пример:*

- ако променливата dohod съдържа информация за средномесечния доход на домакинството, а променливата sizedom за броя на членовете на домакинството, то може да се създаде нова променлива за дохода на член от домакинството:

  Target Variable dohod1 Numeric Expression dohod/sizedom

- Създаването на променлива, която да идентифицира случаите дали отговор 1 или 2 на въпрос 2 от Приложение 1, изисква в Compute да се зададе:

  Target Variable q2_12 Numeric Expression 1

  А в *If > Include if case satisfies condition* да се запише q2.1=1 or q2.2=2

## ПРЕДСТАВЯНЕ НА СТАТИСТИЧЕСКИ ДАННИ

**Едномерни разпределения** – показват честотата с която се срещат значенията на признака в изучаваната съвкупност. Честотата може да бъде абсолютна или относителна.

*Относителните честоти са проценти и осигуряват една по-добра сравнимост и прегледност на резултатите.*

***Едномерното разпределение на въпроси с един отговор*** се получава след като от главното меню се изберат последователно Analyze->Descriptive Statistics->Frequencies… В отворения прозорец се маркира променливата на която ще се формира едномерно разпределение и с кликване на клавиша „стрелка дясно" се прехвърля в полето Variable(s). По този начин могат да се изберат и други променливи и наведнъж да се формират едномерните им разпределения.

*Пример:* Едномерно разпределение на въпрос 3.2 от Приложение 1

**Либерализацията на националния пазар**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | Не оказва влияние | 46 | 10,5 | 10,6 | 10,6 |
| | Малко влияние | 65 | 14,9 | 15,0 | 25,7 |
| | Средно влияние | 122 | 27,9 | 28,2 | 53,9 |
| | Голямо влияние | 199 | 45,5 | 46,1 | 100,0 |
| | Total | 432 | 98,9 | 100,0 | |
| Missing | System | 5 | 1,1 | | |
| Total | | 437 | 100,0 | | |

Едномерното разпределение съдържа:

- *Frequency* – честотата с която се среща всеки отговор, т.е. броя на анкетираните посочили даден отговор
- *Percent* – процентът на посочилите даден отговор от всички анкетирани
- *Valid Percent* – процентът на посочилите даден отговор спрямо далите отговор на този въпрос, т.е. изключени са неотговорилите (за тях не е въведена стойност и имаме missing)
- *Cumulative Percent* – кумулативен процент, получен като сума на всички предходни проценти. Има съдържателна интерпретация, ако отговорите са ранжирани и разкрива процента на посочилите отговор до даден отговор.

***Едномерното разпределение на въпроси с повече от един отговор*** се формират след като от главното меню се изберат последователно Analyze -> Multiple Responds. В Define Sets се указват променливите, които се обработват заедно. Тези променливи се маркират и се прехвърлят в полето Variables in Set. В зависимост от начина на кодиране на отговорите на въпроса във Variables Are Coded As се избира:

- *Dichotomies Counted value*, ако на всеки от отговорите е преписана един и същи код (обикновено 1) и този код трябва да бъде указан в празното поле.
- *Categories*, ако на всеки от отговорите е преписано различен код. Въвежда се минималната стойност след Range и максималната стойност след through.

Въвежда се име (Name) на общата променлива, както и етикета и (Label).

Така дефинираната Multiple Response се добавя чрез Add в полето Mult response Sets.

След дефинирането на съставната променлива, едномерното и разпределение се получава от Analyze -> Multiple Responds -> Frequencies…

*Пример:* Едномерно разпределение на въпрос 2 от Приложение 1

| **$q2 Frequencies** | | Responses | | Percent of Cases |
|---|---|---|---|---|
| | | N | Percent | |
| Участие в конкурентни отношения(а) | за цената на определен продукт/услуга | 336 | 25,5% | 77,1% |
| | за качеството на определен продукт/услуги | 339 | 25,7% | 77,8% |
| | при лансирането на нови продукти/услуги | 204 | 15,5% | 46,8% |
| | при възлагане на поръчки/проекти | 199 | 15,1% | 45,6% |
| | в достъпа до определени суровини и компоненти | 47 | 3,6% | 10,8% |
| | при наемането на квалифициран персонал | 93 | 7,1% | 21,3% |
| | при получаване на субсидии от държавни ф | 18 | 1,4% | 4,1% |
| | в областта на развойна дейност | 54 | 4,1% | 12,4% |
| | друго | 17 | 1,3% | 3,9% |
| | не мога да преценя | 12 | ,9% | 2,8% |
| Total | | 1319 | 100,0% | 302,5% |

Едномерното разпределение съдържа два процента:

- *Responses Percent* – процентът на всеки отговор, спрямо всички отговори.
- *Percent of Cases* – процентът на анкетираните посочили даден отговор, спрямо отговорилите. Това е процента който се интерпретира, чрез който се описва изучаваната съвкупност.

Едномерното разпределение има *обобщаващи статистически характеристики*, изчисляването на които се задава в меню Statistics. Най-често използваните са:

- *Mean* (средно аритметично) – измерител на централната тенденция;
- *Median* (медиана) – значението, което има единицата заемаща централно положение в подредения възходящо или низходящо ред, т.е. 50% от значенията са до стойността на медианата. Могат да се изчисляват и други позиционни средни чрез Percentile Values;
- *Mode* (мода) – най-често срещаното значение, разпределението може да е многомодално;
- *Minimum* (минимална стойност) и *maximum* (максимална стойност);
- *Std.deviation* (стандартно отклонение) $\sigma$ - мярка за разсейването и показва доколко единодушна е получената средна оценка, т.е. доколко близо или далеч от средната оценка се намират отделните значения. Много често в практиката се използва втората степен на стандартното отклонение $\sigma^2$, наричана дисперсия.

**Двумерни разпределения (кростаблици)**

Кростаблиците описват едновременно разпределението по *две променливи*. Формират се с командата: Analyze -> Descriptive Statistic -> Crosstabs и включва:

- избор на променливите, които да се съдържат в таблицата и тяхното аранжиране по ред и колона;
- избор на съдържание на клетките в подфункцията Cells. Въпреки богатия избор, в практиката се използва основно процент спрямо положението на променливата дефинираща описваните подсъвкупности (фактора) - ред или колона. Например, за да опишем степента на удовлетвореност от организацията на работата (разположена в редовете на таблицата) по секторите във фирмата (разположени в колоните на таблицата), то за съдържание на клетките трябва да се зададе – Column Percentage;
- избор на съпътстващи таблицата статистически анализи от подфункцията Statistics, на което ще се спрем по късно;
- избор на формата, обикновено се оставят автоматичните настройки.

**Таблици**

За разлика от кростаблиците, таблиците дават възможност за едновременно представяте на кросове между повече от 2 променливи. Формират се със следния порядък от команди: Analyze -> Tables -> General Tables и включват:

- избор на променливи и тяхното аранжиране. Предимството на таблиците, е че могат да съдържат въпроси с един отговор и въпроси с повече от един отговор. Веднъж дефинирани в подфункцията Mult Response Sets множествените променливи се съхраняват във файла, за разлика от дефинираните множествени променливи във функцията Analyze -> Multiple Responds.
- избор на съдържанието на клетките, персонално за всяка от включените в таблицата променливи.
- вмъкване след последната променлива на едномерното разпределение с избор на Insert Total. Тъй като този тотал се взема от последната променлива в таблицата, то тя трябва да бъде избрана така, че да няма неотговорили по нея.

С функции: Analyze -> Tables -> Tables of Frequencies се формират таблица на въпроси, които имат едни и същи отговори (таблични въпроси, напр. въпрос 3 от Приложение 1).

**Графики**

Графичното представяне на данните улеснява описанието, тъй като не се налага да се цитират всички относителни дялове, а само тези който се открояват най-ярко. За автоматично построяване на графични изображения се използват функциите от меню Graphs в SPSS. Алтернативна възможност са графичните възможности на Excel.

Най-често използваните графични изображения са:

- секторни графики - Pie, за представяне на структура т.е. въпрос с един отговор
- правоъгълни графики - Bar/Column, за представяне на въпрос с повече от един отговор.

## СТАТИСТИЧЕСКО ОЦЕНЯВАНЕ

Генералната съвкупност се характеризира чрез обобщаващи числови характеристики (средни, относителни дялове и др.) наричани *параметри и* означавани с θ.

Характеристиките (средни, относителни дялове и др.) изчислени от емпиричните данни са характеристики на извадката.

Ако извадката е представителна, с определена вероятност, по характеристиките на извадката могат да се правят заключения за характеристиките на генералната съвкупност. Характеристики на извадката се наричат *оценки на параметрите* (бел.с ).

Разликата между оценката и параметъра се нарича *грешка на* извадката (оценката).

**Грешката има два компонента:**

- *Систематична (неслучайна) грешка*– получава се независимо дали изследването е извадково или изчерпателно. Резултат е от грешки във всеки етап на изследването – от регистрацията до обработката. Тази грешка *не може да бъде измерена*, но трябва да се вземат мерки да се сведе до минимум.
- *Стохастична (случайна) грешка* - поражда се от обстоятелството, че не са изследвани всички единици на генералната съвкупност. Ако извадката е излъчена чрез случаен подбор, то грешката *може да се измери* и контролирана.

**Оценките биват:**

- *Точкови* – оценките са фиксирани, точно определени числови характеристики и не съдържат отклонението от търсения параметър.
- *Интервални* – съдържат не само точковата оценка, но и размера на стохастичната грешка. Изчисляват се само за представителни извадки и важат с определена гаранция за сигурност.

Интервалната оценка задава интервала, наричан **доверителен интервал**, в който се намира търсения параметър.

Вероятността с която се гарантира заключението се нарича **гаранционна вероятност** и се бележи с P. В социално-икономическите изследвания се работи с P = 0,95 = 95%

Доверителният интервал се построява по следния начин:

- изчислява се средната стохастична грешка (стандартната грешка) по формулата:

μ = . – , където: n – обема на извадката, N – обема на генералната съвкупност

– оценката на стандартното отклонение, изчислено по формулата

= – за средно аритметично

= за относителен дял

*Забележка:* формула важи за прост случаен подбор.

- изчислява се максималната грешка по формулата:

Δ = z. μ където: z = 1 при P=68% , z се нарича гаранционен множител

z = 1,96 при P=95%

z = 2,58 при P=99%

- построява се доверителния интервал: - Δ < θ < + Δ

За да има познавателно значение *доверителния интервал не трябва да бъде твърде широк,* т.е. максималната грешка трябва да бъде малка. От формулата по която се изчислява, се вижда че *единственият начин по който можем да се влияе върху размера на максималната грешка е обема на извадката*. Например, за да се намали 2 пъти размера на максималната грешка, трябва да се увеличи 4 пъти обема на извадката.

*Пример:* На основата на представителна извадка с обем 437 от генерална съвкупност с обем 1800 е получено, че 26% от фирмите участват в конкуренцията за цената на продуктите. Да се определи доверителния интервал на оценявания параметър.

$$\Delta = z.\ \mu = z.\ \sqrt{\qquad\qquad} = 1{,}96\ .\ \sqrt{\qquad\qquad} = 3{,}58\ \%$$

Доверителният интервал е: 26% - 3,6% < p < 26% + 3,6% => с гаранционна вероятност 95% можем да твърдим, че удовлетворените от организацията на работата във фирмата са между 22,4% и 29,6%. 95% гаранционна вероятност означава, че ако направим 100 случайни извадки със същия обем (437), то само в 5 от тях изчисления процент (p) може да излиза извън интервала [22,4%; 29,6%]

Прието е при отношение n/N < 5%, максималната грешка да се изчислява по формулата

$\Delta = z.\ \sqrt{\qquad}$, защото множителят $\sqrt{\qquad}$ клони към 1.

Това дава възможност да се изчисли максималната грешка на различни относителни дялове, независимо от обема на генералната съвкупност.

Таблицата по-долу показва максималната грешка на различни относителни дялове при различен обем на извадката. Изследователят вземат решение за обема на извадката въз основа на грешката на 50% относителен дял, тъй като за него грешката е максимална.

**Максимална грешка при 95% гаранционна вероятност**

| **p** | **3** | **5** | **10** | **15** | **20** | **30** | **40** | **50** |
|---|---|---|---|---|---|---|---|---|
| **N** | **97** | **95** | **90** | **85** | **80** | **70** | **60** | **50** |
| **100** | 3,34 | 4,27 | 5,88 | 7,00 | 7,84 | 8,98 | 9,60 | 9,80 |
| **200** | 2,36 | 3,02 | 4,16 | 4,95 | 5,54 | 6,35 | 6,79 | 6,93 |
| **300** | 1,93 | 2,47 | 3,39 | 4,04 | 4,53 | 5,19 | 5,54 | 5,66 |
| **400** | 1,67 | 2,14 | 2,94 | 3,50 | 3,92 | 4,49 | 4,80 | 4,90 |
| **500** | 1,50 | 1,91 | 2,63 | 3,13 | 3,51 | 4,02 | 4,29 | 4,38 |
| **600** | 1,36 | 1,74 | 2,40 | 2,86 | 3,20 | 3,67 | 3,92 | 4,00 |
| **700** | 1,26 | 1,61 | 2,22 | 2,65 | 2,96 | 3,39 | 3,63 | 3,70 |
| **800** | 1,18 | 1,51 | 2,08 | 2,47 | 2,77 | 3,18 | 3,39 | 3,46 |
| **900** | 1,11 | 1,42 | 1,96 | 2,33 | 2,61 | 2,99 | 3,20 | 3,27 |
| **1000** | 1,06 | 1,35 | 1,86 | 2,21 | 2,48 | 2,84 | 3,04 | 3,10 |
| **1100** | 1,01 | 1,29 | 1,77 | 2,11 | 2,36 | 2,71 | 2,90 | 2,95 |
| **1200** | 0,97 | 1,23 | 1,70 | 2,02 | 2,26 | 2,59 | 2,77 | 2,83 |
| **1300** | 0,93 | 1,18 | 1,63 | 1,94 | 2,17 | 2,49 | 2,66 | 2,72 |
| **1400** | 0,89 | 1,14 | 1,57 | 1,87 | 2,10 | 2,40 | 2,57 | 2,62 |
| **1500** | 0,86 | 1,10 | 1,52 | 1,81 | 2,02 | 2,32 | 2,48 | 2,53 |
| **1600** | 0,84 | 1,07 | 1,47 | 1,75 | 1,96 | 2,25 | 2,40 | 2,45 |
| **1700** | 0,81 | 1,04 | 1,43 | 1,70 | 1,90 | 2,18 | 2,33 | 2,38 |
| **1800** | 0,79 | 1,01 | 1,39 | 1,65 | 1,85 | 2,12 | 2,26 | 2,31 |
| **1900** | 0,77 | 0,98 | 1,35 | 1,61 | 1,80 | 2,06 | 2,20 | 2,25 |
| **2000** | 0,75 | 0,96 | 1,31 | 1,56 | 1,75 | 2,01 | 2,15 | 2,19 |

## СТАТИСТИЧЕСКА ПРОВЕРКА НА ХИПОТЕЗИ

В практиката много често е необходимо не само да се опишат получените данни, но и тяхното съпоставяне. Методите с които могат да се отговори дали наблюдаваните разлики са значими, т.е. надхвърлят рамките на грешката, се наричат тестове за хипотези.

Хипотезата е *предположение за стойностите на изучаваните параметри* на съвкупността. Проверката на хипотезата преминава през следните **основни етапи**:

- **Дефиниране на хипотезата**, според която разликите между сравняваните стойности на параметрите е случайна. Тази хипотеза се нарича нулева и се записва $H_o$: $X_1 = X_2$.

  Хипотезата, която се противопоставя на нулевата хипотеза се нарича алтернативна. Според нея разликата е статистически значима. Бележи се $H_1$: $X_1 \neq X_2$ при сложна хипотеза и $X_1 < X_2$ или $X_1 > X_2$ при проста. Начина на дефиниране на алтернативната хипотеза, като сложна или проста, определя с каква критична зона се работи, съответно двустранна (2-tailed) или едностранна (1-tailed).
- **Дефиниране на равнището на значимост**. Проверката на хипотеза се основава на ограничена информация (тази в извадката) и следователно заключенията не могат да бъдат категорични, но теорията на статистическите заключения дава възможност рискът от грешка да бъде оценен. Възможните грешки са две:
  - да се отхвърли $H_o$, когато тя е вярна, наричана грешка от I род или $\alpha$ грешка.
  - да се приеме $H_o$, когато тя не е вярна, наричана грешка от II род или $\beta$ грешка.

Вероятността да бъде допусната $\alpha$ грешка се нарича *равнище на значимост*, бел. с $\alpha$ и *се задава от изследователя*. В социално-икономическите изследвания се работи основно с $\alpha$=0,05. В области където последиците от направените изводи са от решаващо значение с работи с $\alpha$=0,01.

- **Избор на тест за проверка на хипотезата в зависимост от:**
  - равнището на измерване (вида на скалата) - *непараметрични тестове* при слаби скали (номинални и ординални) и *параметрични тестове* при силни скали (интервални – количествени променливи);
  - формата на разпределение на съвкупността - *параметричните* методи имат изисквания към формата на разпределение (изискват нормално разпределение), докато при *непараметричните* тя е без значение;
  - зависими или независими са извадките от които са получават сравняваните параметри, т.е. едни и същи или различни са единиците в извадките. *Пример:* Извадките са *независими*, ако се сравняват оценката по признака Х на мъже и оценката по признака Х на жените. Независимите извадки се задават с групираща променлива, която да разбие съвкупността на подсъвкупности. Извадките са *зависими*, ако се сравняват оценката на цялата изследвана съвкупност по признака Х и оценката на цялата изследвана съвкупност по признака У.
  - колко са сравняваните параметъра (с колко извадки се работи) - с два или повече от две.

- **Статистическо заключение**, извършва се на основата на сравнение на изчисленото от SPSS гранично равнище на значимост (Asymp.Sig.) и фиксираното от изследователя, обикновено 0,05.
    - Ако Sig.>$\alpha$=0,05$\Rightarrow$ $H_o$, т.е. няма разлика между сравняваните характеристики (параметри).
    - Ако Sig.<$\alpha$=0,05$\Rightarrow$ $H_1$, т.е. има (статистически значима) разлика.

**ОСНОВНИ ТЕСТОВЕ**:

**One-Sample Kolmogorov-Smirnov** – непараметричен метод за проверка за съгласуваност между емпирично и теоретично разпределение (нормално, равномерно, поасоново, експоненциално). Тестът се изпълнява чрез функциите: *Analyze -> Nonparametric Test* -> 1 Sample K-S… Конкретизира се с какво теоретично разпределение (Нормално; Равномерно; Поасоново; Експоненциално) ще се сравнява емпиричното разпределение.

Проверяваната хипотеза е:

$H_0$: ЕР ~ НТР (няма статистически значима разлика между разпределението на променливата в генералната съвкупност и избраното теоретично разпределение, в случая нормално)

Ако Sig. $\geq$ $\alpha$ $\Rightarrow$ $H_0$, т.е. приемаме че емпиричното разпределение е но

**Хи квадрат тест –** непараметричен тест за проверка на връзка между две променливи на слабите скали. Тестът се изпълнява чрез функциите: *Analyze -> Descriptive Statistic -> Crosstabs-> Statistic*, където се избира *Chi square.*

*Пример:* Има ли връзка между оценката за ролята на хармонизирането на националното законодателство с европейското за развитие на бизнеса и националността на фирмата.

Нулевата хипотеза ($H_0$) гласи, че няма връзка между двете променливи.

Резултатите от теста са:

**Chi-Square Tests**

| | Value | Df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson **Chi-Square** | 19,752(a) | 6 | **,003** |
| Likelihood Ratio | 19,458 | 6 | ,003 |
| Linear-by-Linear Association | 9,974 | 1 | ,002 |
| N of Valid Cases | 433 | | |

a 0 cells (,0%) have expected count less than 5. The minimum expected count is 9,51.

Тъй като Asymp.Sig = 0,003 < 0,05 $\Rightarrow$ $H_1$, т.е. има връзка между националността и оценката. Силата на това връзка се измерва с коефициенти на контингенция (виж по-долу).

*Условия за коректното приложението* на теста са:

- Да няма очаквана честота по-малка от 5 в повече от 20% от клетките, т.е. процентът в „*0 (cells (,**0%**) have expected count less than 5*" да бъде по-малък от 20%
- Минималната очаквана честота да бъде 1, т.е. числото в „*The minimum expected count is **9,51***" да бъде по-голямо от 1.

**U-тест на Ман-Уетни -** непараметричен тест за сравняване на разпределенията на променливи на ординална скала при независими извадки.

Проверката се осъществява след последователно изпълнение на функциите:

*Analyze -> Nonparametric Test -> 2 Independent Samples,* където се задават:

- в *Test variable list* – променливата, чиито разпределения ще се сравняват. Могат да се зададат повече от една променливи за едновременно тестване.
- *Grouping Variable* - променливата определяща двете независими извадки, като в подфункцията *Define Groups* се въвеждат стойностите определящи групите.

*Пример:* Сравнението на разпределението на оценките за влиянието на хармонизирането на законодателство с европейското за развитие на бизнеса в България и Македония, изисква да се зададе q3.1 в *Test variable list,* q1 в *Grouping Variable,* стойност 1 за Group 1 и 2 за Group 2 в *Define Groups.*

Резултатът от теста е:

**Test Statistics(a)**

| | Хармонизирането на българското законодателство с европейското |
|---|---|
| Mann-Whitney U | 7407,000 |
| Wilcoxon W | 10977,000 |
| Z | -,127 |
| Asymp. Sig. (2-tailed) | **,899** |

a Grouping Variable: Националност на фирмата

От Asymp.Sig = 0,899 >0,05 $\Rightarrow H_0$, т.е. няма разлика в оценките за ролята на хармонизирането на националното законодателство с европейското в България и Македония.

**Тест на Уилкоксън -** непараметричен тест за сравняване на разпределенията на променливи на ординална скала при зависими извадки.

Проверката се осъществява след последователно изпълнение на функциите:

*Analyze -> Nonparametric Test -> 2 Related Samples,* където последователно се маркират двете променливи и се прехвърлят в *Test Pair(s) List.*

*Пример:* Сравнението на разпределението на оценките за влиянието на „разширяване на достъпа до пазарите на ЕС" и „либерализацията на националния пазар" за развитие на бизнеса, показва че тези два фактора се оценяват различно (Asymp.Sig = 0,000 <0,05 $\Rightarrow H_1$)

**Test Statistics(b)**

| | Разширяване на достъпа до пазара на ЕС - Либерализацията на националния пазар |
|---|---|
| Z | -6,591(a) |
| Asymp. Sig. (2-tailed) | ,000 |

**Т тест за разлика между две средни** – параметричен тест за независими извадки.

Проверката се осъществява с функциите: *Analyze -> Compare Means -> Independent-> Samples T Test.* Тестващата променлива се задава в *Test variable list,* а променливата дефинираща двете групи в *Grouping Variable.*

*Пример:* Сравнение на средния процент отделян за обучение и преквалификация на персонала от фирмите в изследвания сектор в България и Румъния.

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | |
|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) |
| Каква част от продажбите за 1999 г. сте отделили за обучение и преквали-фикация на персонала | Equal variances assumed | 11,151 | **,001** | 2,459 | 325 | ,014 |
| | Equal variances not assumed | | | 2,404 | 244,524 | **,017** |

Анализът преминава през две стъпки:

- проверка на тест за разлика на дисперсиите ($\sigma^2$), сравнявайки Sig. от таблицата Levene's Test for Equality of Variances с 0,05.
    - Ако Sig. > 0,05 анализът продължава по реда Equal variances assumed
    - Ако Sig. < 0,05 анализът продължава по реда Equal variances not assumed.
- проверка на теста за разлика на средните, сравнявайки Sig. от избрания на първата стъпка ред от таблицата t-test for Equality of Means с 0,05.

В конкретния пример Sig. за Levene's Test =0,001 < 0,05, от където Sig. за t-test =0,017 < 0,05 $\Rightarrow$ $H_1$, т.е. сравняваните относителни дяла са различни. От съпътстващата проверката Group Statistics се определя коя от двете сравнявани средни е по-голема.

**ANOVA** – параметричен тест за сравняване на повече от две независими извадки.

Този тест е известен като дисперсионен анализ.

Проверката се осъществява с функциите: *Analyze -> Compare Means ->One-Way ANOVA,* където в:

- *Dependent List* се задава *количествената променлива*, чиито средни стойности в различни групи ще се сравняват;
- *Factor* се задава *качествена променливата* дефинираща групите (подсъвкупностите);
- *Options* се маркира *Homogenety of variance test* за проверка на условието на анализа за равенство на дисперсиите в подсъвкупностите;
- *Post Hoc ...* се маркира теста на *Duncan* за проверка на конкретните разлики между сравняваните средни.

*Пример:* Сравнение на процента отделян за развойна дейност от фирмите в изследвания сектор за трите страни.

Резултатите от АНОВА показват, че между трите средни има разлика (Sig = 0,000 <0,05 ⇒ $H_1$).

**ANOVA**

**Каква част от продажбите за 1999 г. сте отделили за развойна дейност (R&D)**

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 25629,262 | 2 | 12814,631 | 26,260 | **,000** |
| Within Groups | 200561,453 | 411 | 487,984 | | |
| Total | 226190,715 | 413 | | | |

От теста на Дънкан, следва че за България и Македония процентът не се различава статистически, докато за Румъния той е по-малък.

**Duncan**

| Националност на фирмата | N | Subset for alpha = .05 | |
|---|---|---|---|
| | | 1 | 2 |
| Румъния | 171 | 2,93 | |
| България | 164 | | 17,33 |
| Македония | 79 | | 21,35 |
| Sig. | | 1,000 | ,156 |

Тестът за равенство на разсейването в трите страни показва, че условието не е изпълнено – дисперсиите не са равни (Sig = 0,000 <0,05 ⇒ $H_1$). Това прави ненадеждни заключенията и налага проверката да се направи с непараметричен тест за сравнение на три средни или със параметричен, като се сравняват две по две трите средни.

**Test of Homogeneity of Variances**

| Levene Statistic | df1 | df2 | Sig. |
|---|---|---|---|
| 59,304 | 2 | 411 | **,000** |

**Т тест за разлика между две зависими средни** – параметричен тест за зависими извадки.

Проверката се осъществява с функциите: *Analyze -> Compare Means -> Paired-Samples T Test.* Маркират се последователно двете количествени променливи и се прехвърлят в *Paired Variables.*

*Пример:* С този тест се установява, че средният процент отделян за развойна дейност от фирмите в изследвания сектор е по-голям от този за обучение и преквалификация на персонала. Изводите са на основата на Sig = 0,000 <0,05 ⇒ $H_1$ – сравняваните средни са различни и конкретните стойност на двата процента от придружаващата теста таблица - Paired Samples Statistics.

**Paired Samples Test**

| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| Pair 1 | Каква част от продажбите за 1999 г. сте отделили за развойна дейност (R&D) - Каква част от продажбите за 1999 г. сте отделили за обучение и преквалификация на персонала | 4,268 | 18,899 | ,945 | 2,410 | 6,125 | 4,516 | 399 | **,000** |

## СТАТИСТИЧЕСКИ ЗАВИСИМОСТИ

Степента на зависимост между две променливи се определя чрез:

**Коефициенти на контингенция** при качествени променливи (променливи чиито значения нямат количествено изражение, а се дефинират словесно напр. пол, семейно положение, националност).

Коефициентите на контингенция приемат стойности в интервала [0,1]. Колкото по-близо е стойността им до 1, толкова зависимостта е по-силна. Условно зависимостта се разглежда, като слаба, ако коефициента е до 0.3; умерена - от 0.3 до 0.5; значителна - от 0.5 до 0.7; голяма - от 0.7 до 0.9; много голяма - над 0.9.

Най-често използваните коефициенти на контингенция са Фи коефициент, V коефициент на Крамер и C коефициент на контингенция, които са включени в подфункцията Statistics на Crosstabs.

Пример: Стойностите на тези коефициенти за зависимостта между променливите q1 и q3.1 се съдържат в колоната Value.

**Symmetric Measures**

| | | Value | Approx. Sig. |
|---|---|---|---|
| Nominal by Nominal | Phi | ,214 | ,003 |
| | Cramer's V | ,151 | ,003 |
| | Contingency Coefficient | ,209 | ,003 |
| N of Valid Cases | | 433 | |

Важно е да се знае, че стойността на коефициента може да се тълкува (да се приеме за статистически значима), само ако изчисленото за него равнище на значимост (Approx. Sig) е по-малко от 0,05 отговарящо на гаранционна вероятност 95%.

Сравняването на тези коефициенти за зависимостта между q1 и всяко от променливите от q3.1 до q3.5, ще покаже по кои от твърденията националните различия са най-големи.

**Коефициенти на рангова корелация** на Кендал и Спирман при ординално скалирани променливи (променливи в които степента на различие се дефинира с разновидности от рода „по-голямо", „равно", „по-малко" *или* „напълно", „отчасти", „не" *или* „редовно, почти всеки ден", „2-3 пъти седмично", „няколко пъти месечно", „по-рядко").

**Коефициент на корелация** на Пирсън при количествени променливи (променливи чиито значения имат количествена определеност, напр. възраст, доходи).

Коефициентите на рангова корелация, както и коефициентите на корелация (често наричани с общото наименование „коефициенти на корелация") приемат стойности в интервала [-1,1]. Силата на зависимостта се определя от *абсолютната стойност* на коефициента, аналогично на коефициентите на контингенция. *Знакът на коефициента* показва характера на зависимостта. Положителният знак, идентифицира положителна връзка (на високи стойности на едната променлива отговарят високи стойности от другата променлива). Отрицателният знак, определя връзката като обратна/отрицателна (на високи стойности на едната променлива отговарят ниски стойности от другата променлива).

Коефициентите на корелация се формират със следния порядък от функции:

Analyze -> Correlate -> Bivariate, където се указва:

- променливите между които трябва да се определи зависимостта;
- желаният коефициент. За количествени променливи - Pearson. За рангови променлива има избор между коефициента на Kendall's и на Spearman, като се отчита, че по правило между двата коефициента е в сила връзката: Kendall's Coef. < Spearman Coef.

Получената корелационна таблица е симетрична със стойности по диагонална 1 (пълна зависимост). Тълкуват се само коефициентите над или под диагонала, но само ако изчисленото за тях равнище на значимост (Sig.) е по-малко от 0,05 отговарящо на гаранционна вероятност 95%. За улеснение SPSS маркира с * всяка стойност удовлетворяваща това условие, а с ** тези коефициенти които са статистически значими при 99% гаранционна вероятност. Отчитайки, че работим с 95% гаранционна вероятност е достатъчно да има *, независимо от броя им, за да може да се тълкува получения коефициент.

**Приложение 1**

# ВЪПРОСНИК

**Q1. Националност на фирмата**

1. България
2. Македония
3. Румъния

**Q2. Предлагаме ви списък с видовете конкурентни отношения между фирмите. В кои от тях Вашата фирма участва?**

1. Конкуренция за цената на определен продукт/услуга
2. Конкуренция за качеството на определен продукт/услуга
3. Конкуренция при лансирането на нови продукти/услуги
4. Конкуренция при възлагане на поръчки/проекти
5. Конкуренция в достъпа до определени суровини и компоненти
6. Конкуренция при наемането на квалифициран персонал
7. Конкуренция при получаване на субсидии от държавни фондове
8. Конкуренция в областта на развойна дейност
9. Друго ……
10. Не мога да преценя

**Q3. Как оценявате влиянието на следните фактори върху перспективите за развитие на Вашия бизнес?**

| | | Не оказва влияние | Малко влияние | Средно влияние | Голямо влияние |
|---|---|---|---|---|---|
| 3.1 | Хармонизирането на българското законодателство с европейското | 2 | 3 | 4 | 5 |
| 3.2 | Либерализацията на националния пазар | 2 | 3 | 4 | 5 |
| 3.3 | Разширяване на достъпа до пазара на ЕС | 2 | 3 | 4 | 5 |
| 3.4 | Участие на страната в международни инфраструктурни проекти | 2 | 3 | 4 | 5 |
| 3.5 | Присъединяването на страната към НАТО | 2 | 3 | 4 | 5 |

**Q4. Каква част от продажбите за 1999 г. сте отделили за:**

4.1 Развойна дейност (R&D) .....%

4.2 Обучение и преквалификация на персонала .....%