

### 3.1. Въведение

Единиците на статистическите съвкупности се различават (варират) по величината на наблюдаваните признаци. Вариацията е присъщо свойство на статистическите съвкупности. Ако тя не съществуваше, статистическите проучвания щяха да бъдат излишни. Тези различия са породени от различни по своето естество фактори, които най-общо се разделят на две категории:

- **закономерни**, които действат еднакво върху всички изучавани единици и определят типичното, закономерното състояние на признака (например типично, характерно за баскетболистите е, че имат висок ръст).
- **случайни**, които действат с различна сила върху изучаваните единици и водят до индивидуални отклонения във величината на признака (отделните баскетболисти се различават по своя ръст).

Задача на статистическото проучване е да даде обобщаващи количествени характеристики на състоянието на изследвания признак. За целта е необходимо, **на първо място**, да се установи какво е типичното за дадената съвкупност стойност на признака, за да се опише влиянието на закономерно действащите фактори. **На второ място**, от особена важност е характеризиране на разсейването, варирането на стойностите на признака, за да се отчете влиянието на случайните фактори. **Трета група** от показатели носят информация за формата на разпределението на емпиричните стойности и степента на съответствието му с някое от известните теоретични разпределения, най-често с това на нормалното разпределение.

-

### 3.2. Показатели за средно равнище

Показателите за средно равнище представляват обобщаващи количествени характеристики, с помощта на които се описва типичното, характерното състояние на изследвания признак. Най-често употребяваните показатели са **средна аритметична величина ( $\bar{X}$ ), мода ( $M_0$ ) и медиана ( $Me$ ).**

- **Средната аритметична величина** ( $\bar{X}$ ) е най-често използваният в научно-изследователската практика показател за средно равнище. Изчислява се по формулата:

$$\bar{X} = \frac{\sum X}{n} \quad 3.1 \quad \begin{array}{l} \text{Където:} \\ \Sigma X - \text{сума на стойностите} \\ n - \text{брой на наблюденията} \end{array}$$

- **Модата (Mo)** е онази стойност, която има най-голяма абсолютна честота, т.е. която се повтаря най-голям брой пъти.
- **Медианата (Me)** е онази стойност на променливата величина, която заема средно положение във вариационния ред и го дели на две равни части. Броят на стойностите, които са по-малки или равни на медианата, е равен на броя на тези, които са по-големи или съвпадат с нея.

### 3.3. Показатели за разсейване

Показателите за разсейване дават количествена характеристика на величината на отклоняване на стойностите от типичното, средното равнище. За измерване на статистическото разсейване най-често се използват показателите **размах (R)**, **стандартно отклонение (S)** и **коэффициент на вариация (V)**.

- **Размахът (R)** е най-елементарният измерител на статистическото разсейване. Той представлява разликата между най-голямата и най-малката стойност на променливата величина и носи информация за диапазона, в който варират значенията на признака. Изчислява се по формулата:

$$R = X_{\max} - X_{\min} \quad (3.2) \quad \begin{array}{l} \text{Където:} \\ X_{\max} - \text{най-голямата стойност на} \\ \text{променливата} \\ X_{\min} - \text{най-малката стойност на променливата} \end{array}$$

Тъй като в изчисляването на размаха участват само двете крайни стойности на вариационния ред, той е твърде **неустойчив измерител на разсейването**.

- **Стандартното отклонение (S)** е най-прецизният и често използван показател за разсейване. Той описва големината на отклонение на стойностите на променливата от средната аритметична величина. Формулата, която го дефинира е:

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}} \quad (3.3)$$

Където:  
 $X_i$  – всяка една стойност на променливата в извадката  
 $\bar{X}$  – средна аритметична величина  
 $n$  – обем на извадката

Изброените статистически показатели за разсейване изразяват степента на различията в същата мерна единица, в която е изследваният признак. Това създава затруднения при сравняване на вариацията на различни признаци.

Показателят, който дава възможност за сравняване на различията на признаци, изразени в различни мерни единици или относно различна средна стойност е **коэффициентът на вариация (V)**. Той се изчислява по формула:

$$V = \frac{S}{\bar{X}} \cdot 100 \quad (3.4)$$

Където:  
 $S$  – стандартно отклонение  
 $\bar{X}$  – средна аритметична величина

Освен за сравняване на разсейването на стойностите на различни променливи, коэффициентът на вариация се ползва и за оценяване на степента на разсейването (еднородността на извадката):

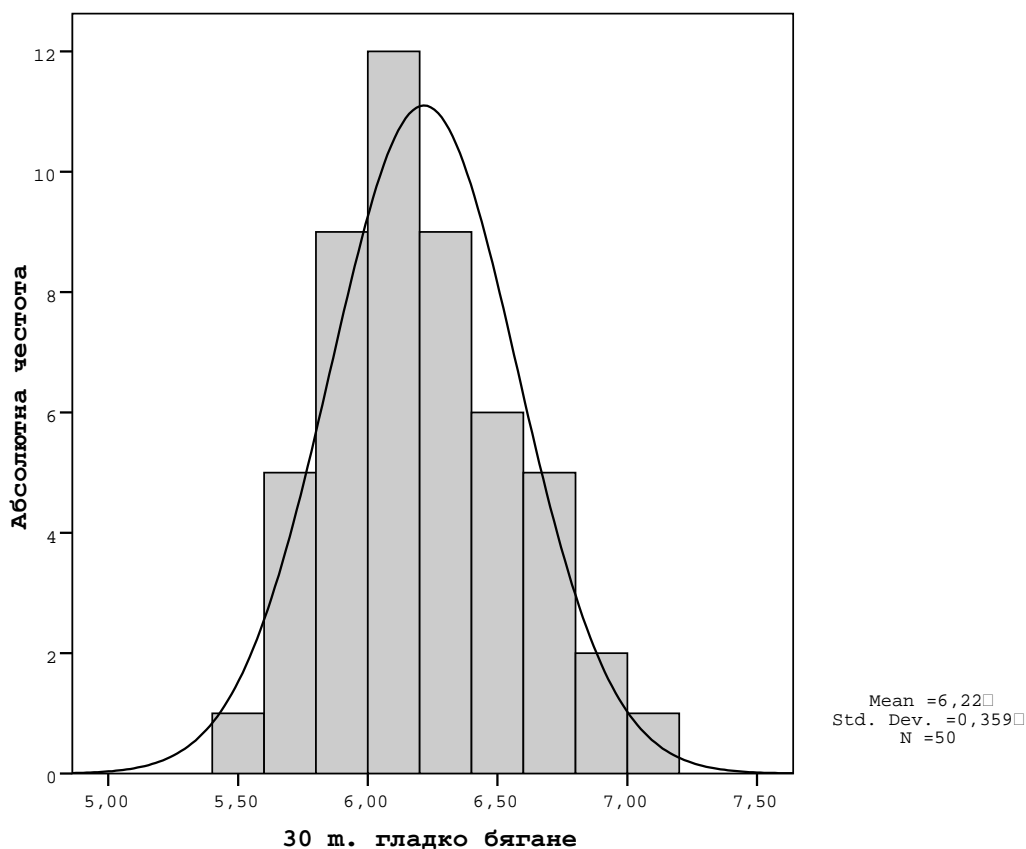
- Счита се, че разсейването на признака е малко (**извадката е еднородна**), когато стойността му е до 10-12%.
- Между 10 и 30% извадката е **приблизително еднородна**.
- Когато е над 30% разсейването на признака е голямо (**извадката е силно нееднородна**).

-

### 3.4. Форма на разпределението

В настоящото помагало многократно е използван терминът “разпределение на променливата величина”, с който се означава съответствието между стойностите на променливата и техните абсолютни или относителни честоти. Тъй като статистическите проучвания се основават на изследването на извадки, това разпределение се нарича емпирично. Ако то се онагледява графично, по конфигурацията на графичния образ може да се съди за закономерността, на която е подчинено изучаваното масово явление, и да се търси аналог сред т. нар. теоретични разпределения.

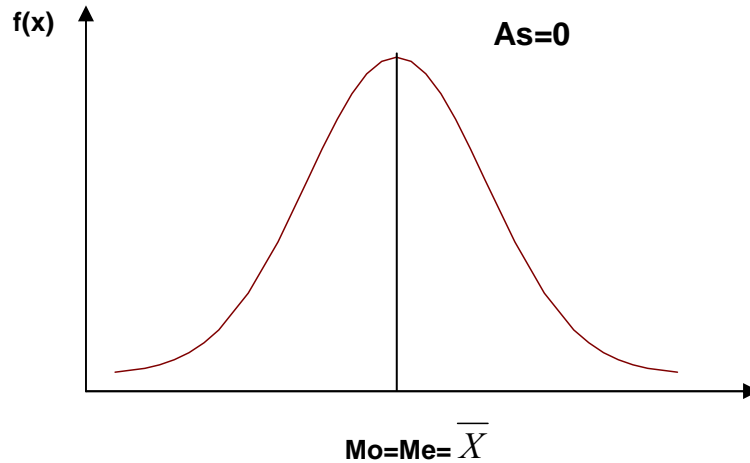
Изключително важно място в теорията на статистиката заема **нормалното разпределение**. То носи наименованието Гаус-Лапласово по имената на немския математик Карл Гаус (1777 – 1855) и френския математик Пиер Лаплас (1739 – 1827), които са го изследвали и описали теоретично. Нормалното разпределение стои в основата на изясняване на важни положения от репрезентативните статистически проучвания, интервалното оценяване, проверката на хипотези, методите за изготвяне на нормативи и др. Прилагането на параметрични методи за проверка на хипотези и за изследване на зависимости между променливите величини също изисква предварително изясняване на въпроса доколко разпределението на емпиричните данни следва закона на нормалното разпределение. Ще се спрем накратко на някои негови особености:



**Фиг. 3.1**

На фигура 3.1 е представена хистограма на резултатите в тест 30 м. гл. б., които се ползват за практическите упражнения в настоящото помагало. Както се вижда, разпределението на честотите следва една характерна тенденция – най-често се срещат стойностите, които се намират около средната, а онези, които са близки до минималната и максималната стойност се срещат най-рядко. На същата графика е представена и кривата на нормалното разпределение. Много ясно се вижда, че разпределението на емпиричните стойности (височината на правоъгълниците на хистограмата) в голяма степен наподобява това на нормалното разпределение. Преди да се спрем на статистическите показатели, които описват степента на приближение на емпиричното към теоретичното нормално разпределение ще разгледаме някои характерни негови особености:

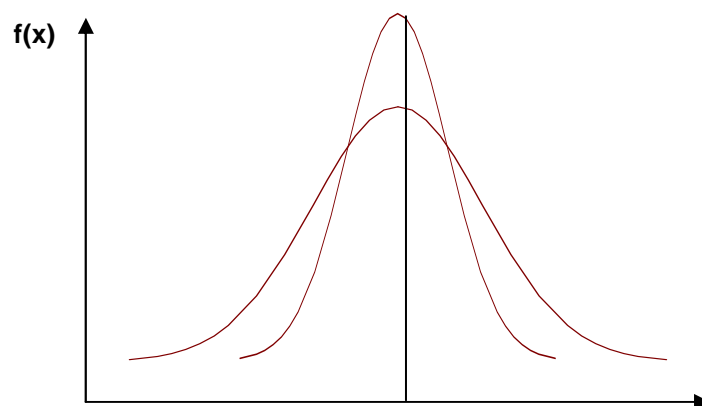
1. Кривата на разпределение на стойностите на променливата и съответните им абсолютни или относителни честоти е **камбановидна** (фигура 3.2).



Фиг. 3.2

2. Средната аритметична величина, модата и медианата съвпадат по стойност и определят центъра на разпределението.

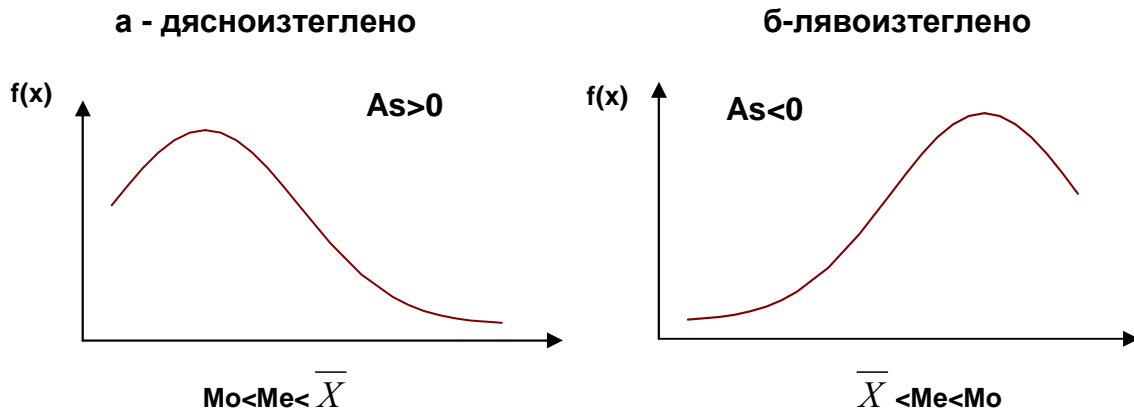
3. Стандартното отклонение (**S**) определя заостреността на кривата на разпределение. Колкото отклоненията от средното равнище са по-малки, толкова кривата е по-изострена и обратно. На фиг. 3.3 са представени две емпирични разпределения с еднакъв център, но с различно разсейване.



Фиг. 3.3

4. Кривата на нормалното разпределение е **симетрична** по отношение на перпендикуляра, спуснат от нейния връх към абсцисата, където се намира средната аритметична величина. На фигура 3.2 е представено симетрично

разпределение, а на фигура 3.4 две несиметрични разпределения..  
 Разпределение а е с дясно изтеглено, а б - с ляво изтеглено рамо.



Фиг. 3.4

Симетричността на разпределението се описва с показателя за **асиметрия (As)**, който при нормално разпределение е равен на нула (**As=0**). Когато коефициентът на асиметрия е с положителен знак, кривата на разпределение е с дясно изтеглено рамо (**а**), средната аритметична е по-голяма от модата и медианата. При разпределение с ляво изтеглено рамо (**б**) стойността на коефициента на асиметрия е с отрицателен знак, а средната аритметична е по-малка от модата и медианата.

За симетричността на емпиричното разпределение се съди по изчисления от извадката по формула 3.5 **коефициент на асиметрия (As)**.

$$As = \frac{m_3}{S^3} \quad (3.5)$$

Където:

**$m_3$**  – трети централен момент, който се изчислява по формула 3.6

**S** – стандартно отклонение

Третият централен момент се изчислява по формула:

$$m_3 = \frac{\sum (X_i - \bar{X})^3}{n} \quad (3.6)$$

Където:

**$X_i$**  – всяка една стойност на променливата в извадката

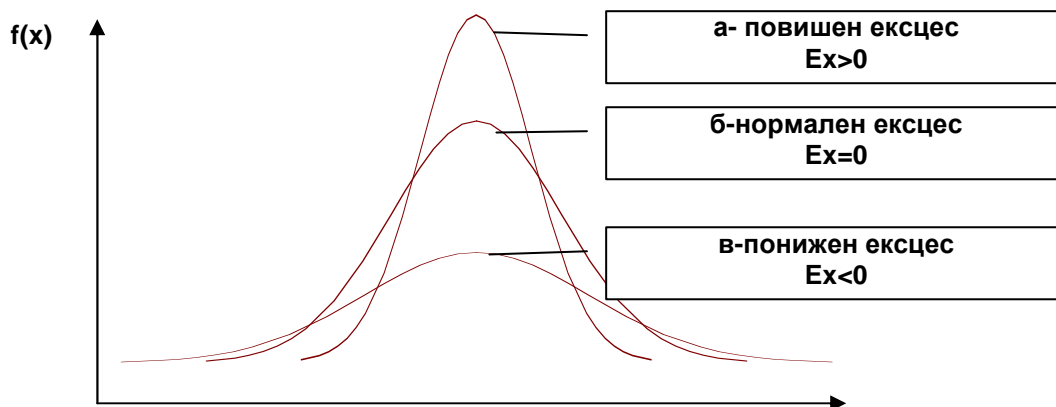
**$\bar{X}$**  – средна аритметична величина

**n** – обем на извадката

Абсолютната стойност на получения по данни от извадката коефициент се сравнява с табличната стойност на коефициента, подбрана от приложение 3 в от обема на извадката ( $n$ ) и равнището на значимост ( $\alpha$ ). В случай че:

- $|As| < As_\alpha$  се приема, че разпределението на емпиричните данни е симетрично.
- $|As| \geq As_\alpha$  се счита, че разпределението е несиметрично.

5. Височината на върха на нормалното разпределение се описва с показателя ексцес ( $Ex$ ), който при нормално разпределение е равен на нула. На фигура 3.5 са представени примерни разпределения с нормален (б), повишен - положителен (а) и понижен - отрицателен ексцес (в).



Фиг. 3.5

Информация за степента на съвпадение на височината на върха на емпиричното разпределение с приетия за еталон връх на нормалното разпределение ( $Ex=0$ ) дава изчисленият по формула 3.7 коефициент на ексцес ( $Ex$ ).

$$Ex = \frac{m_4}{S^4} - 3 \quad (3.7)$$

Където:

$m_4$  – четвърти централен момент, който се изчислява по аналогичен на третия централен момент начин, като разликите между всяка една стойност на променливата и средната аритметична се повдигат на четвърта степен

$S$  – стандартно отклонение



Абсолютната стойност на изчисления коефициент се сравнява с подбраната в зависимост от обема на извадката ( $n$ ) и равнището на значимост ( $\alpha$ ) таблична стойност (приложение 4). В случай че:

- $|E_x| < E_{x_\alpha}$  се приема, че разпределението на емпиричните данни има нормален ексцес.
- $|E_x| \geq E_{x_\alpha}$  се счита, че разпределението има ненормален ексцес.

Изчислените емпирични стойности на коефициентите на асимерия и ексцес на представеното на фиг. 3.1 разпределение на резултатите в тест 30 м. гл. б. са съответно  $A_s=0,331$  и  $E_x=-0,286$ . При обем на извадката  $n=50$  и  $\alpha=0,05$  критичните стойности са съответно  $A_{s_{0,05}}=0,533$  и  $E_{x_{0,05}}=0,848$ . Тъй като емпиричните стойности са по-малки от критичните може да се приеме, че разпределението на променливата е нормално.

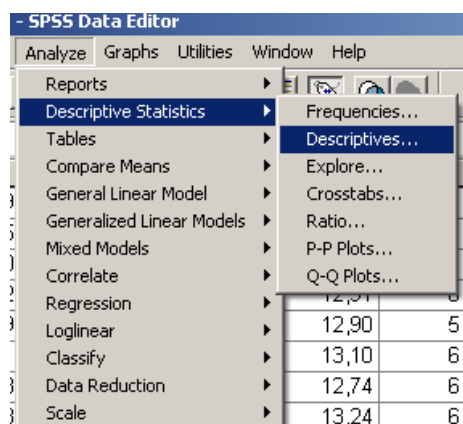
### 3.5. Вариационен анализ със SPSS

#### 3.5.1. Обработка на целия масив

##### Задача за разучаване 3.1.

По данните от файла Ph\_education.sav направете вариационен анализ на резултатите от първото тестиране.

1. Изберете менюто **Analyze > Descriptive statistics > Descriptives** (фиг. 3.6).



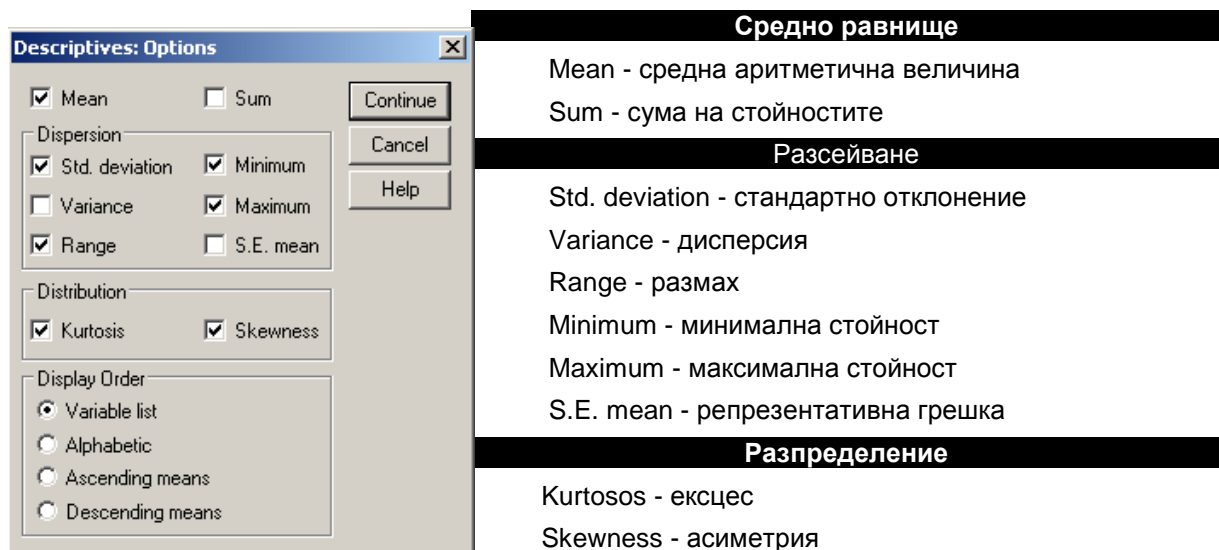
Фиг. 3.6

2. В диалоговия прозорец **Descriptives** (фиг.3.7) маркирайте променливите, натиснете бутона със стрелката и ги изпратете в полето Variable(s);



Фиг. 3.7

3. Натиснете бутона Options. В отворения се диалогов прозорец (фиг. 3.8) отбележете статистическите показатели, които да включи обработката.



Фиг. 3.8

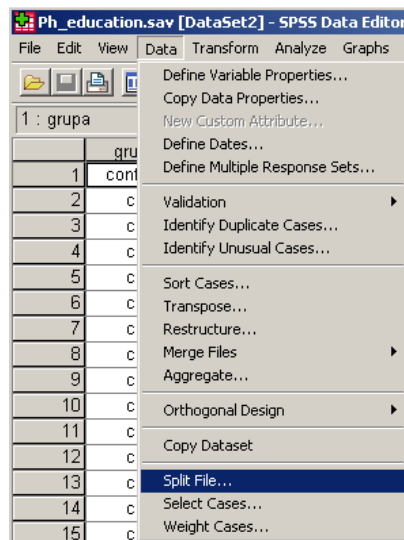
4. Натиснете бутона **Continue**, за да се върнете към основния диалогов прозорец и **OK**, за да стартирате обработката.

### 3.5.2. Обработка на подгрупи от базата данни

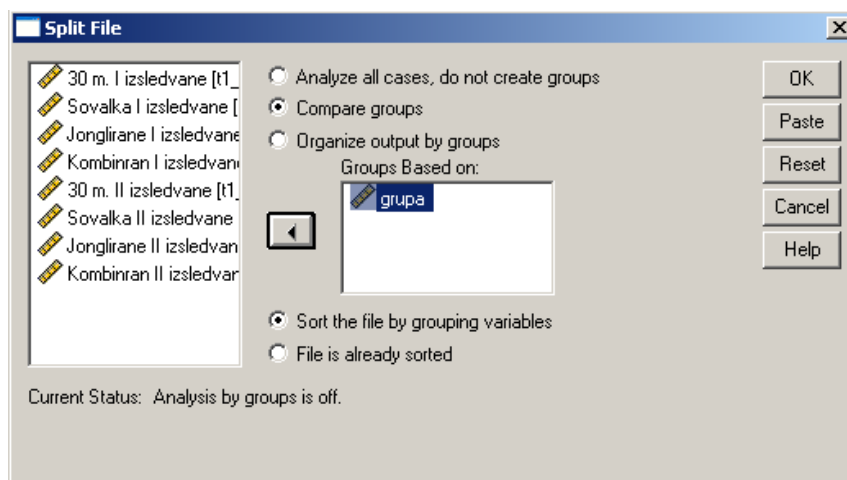
#### Задача за разучаване 3.2.

Разделете масива от данни на групи за самостоятелна обработка по променливата **grupa** и направете вариационен анализ за резултатите в тест «Совалка».

1. Изберете **Data>Split File** (фиг. 3.9). Появява се диалоговият прозорец, показан на фигура 3.10. От него личи, че по подразбиране програмата обработва всички случаи.



Фиг. 3.9



Фиг. 3.10

2. Задайте последователно:

- **Compare groups** (ако искате анализът за групите да се появи в една таблица) или **Organize output by groups** (ако искате анализът за всяка група да се появи в отделна таблица),
- Изпратете променливата, в която е записана принадлежността към групите, за които желаете да осъществите самостоятелна обработка в полето **Groups Based on**;
- Потвърдете с **ОК**.

3. Повторете вариационния анализ по алгоритъма, описан в 3.5.1. Като резултат обработката ще бъде осъществена за всяка от групите.

### 3.6. Представяне на резултати от вариационен анализ

Резултатът от обработката (задача за изучаване 3.1) се появява в разпечатката на SPSS по начина, представен на таблица 3.1:

Таблица 3.1

| Descriptive Statistics  |           |           |           |           |           |           |           |            |           |            |
|-------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|-----------|------------|
|                         | N         | Range     | Minimum   | Maximum   | Mean      | Std.      | Skewness  |            | Kurtosis  |            |
|                         | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| 30 m. I izsledvane      | 50        | 1,56      | 5,45      | 7,01      | 6,2164    | ,35932    | ,331      | ,337       | -,286     | ,662       |
| Sovalka I izsledvane    | 50        | 2,50      | 13,93     | 16,43     | 14,7460   | ,56767    | ,818      | ,337       | ,066      | ,662       |
| Jonglirane I izsledvane | 50        | 2,00      | 2,00      | 4,00      | 2,8200    | ,62890    | ,150      | ,337       | -,473     | ,662       |
| Kombinran I izsledvane  | 50        | 4,16      | 10,45     | 14,61     | 13,1036   | ,91901    | -,475     | ,337       | ,559      | ,662       |
| Valid N (listwise)      | 50        |           |           |           |           |           |           |            |           |            |

При публикуване в научни разработки таблицата се нуждае от редактиране, изписване на статистическите показатели с възприетите символи и допълнително изчисляване на коефициента на вариация ( $V\%$ ). Окончателно оформената таблица изглежда по начина, представен в таблица 3.2.

Таблица 3.2.

| № | Тест                    | Показател | n  | $X_{\min}$ | $X_{\max}$ | R    | $\bar{X}$ | S     | V     | As    | Ex    |
|---|-------------------------|-----------|----|------------|------------|------|-----------|-------|-------|-------|-------|
| 1 | 30 m. I izsledvane      |           | 50 | 5,45       | 7,01       | 1,56 | 6,22      | 0,359 | 5,78  | 0,33  | -0,29 |
| 2 | Sovalka I izsledvane    |           | 50 | 13,93      | 16,43      | 2,5  | 14,75     | 0,567 | 3,85  | 0,82  | 0,07  |
| 3 | Jonglirane I izsledvane |           | 50 | 2          | 4          | 2    | 2,82      | 0,628 | 22,30 | 0,15  | -0,47 |
| 4 | Kombinran I izsledvane  |           | 50 | 10,45      | 14,61      | 4,16 | 13,10     | 0,919 | 7,01  | -0,48 | 0,56  |

*Забележка:* Критичните стойности на коефициентите на асиметрия и ексцес при  $n=50$  и  $\alpha=0.05$  са равни на 0.673 и 1.324.

Резултатите от вариационния анализ на отделните групи (задача за изучаване 3.2.) се появява по начина, представен в таблица 3.3, а редактираният и вид – в таблица 3.4.

Таблица 3.3

| Descriptive Statistics |                      |           |           |           |           |           |           |           |            |           |            |
|------------------------|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|-----------|------------|
| grupa                  |                      | N         | Range     | Minimum   | Maximum   | Mean      | Std.      | Skewness  |            | Kurtosis  |            |
|                        |                      | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| experimental           | Sovalka I izsledvane | 25        | 2,50      | 13,93     | 16,43     | 14,7720   | ,63827    | ,990      | ,464       | ,309      | ,902       |
|                        | Valid N (listwise)   | 25        |           |           |           |           |           |           |            |           |            |
| control                | Sovalka I izsledvane | 25        | 1,52      | 14,09     | 15,61     | 14,7200   | ,49912    | ,386      | ,464       | -1,191    | ,902       |
|                        | Valid N (listwise)   | 25        |           |           |           |           |           |           |            |           |            |

Таблица 3.4

| № | Тест            | Показател | n  | $X_{\min}$ | $X_{\max}$ | R    | $\bar{X}$ | S    | V    | As   | Ex    |
|---|-----------------|-----------|----|------------|------------|------|-----------|------|------|------|-------|
| 1 | Експериментална |           | 25 | 13,93      | 16,43      | 2,5  | 14,77     | 0,63 | 4,32 | 0,99 | 0,31  |
| 3 | Контролна       |           | 25 | 14,09      | 15,61      | 1,52 | 14,72     | 0,49 | 3,39 | 0,39 | -1,19 |