

ПРЕДМЕТ

**СТАТИСТИЧКЕ МЕТОДЕ У БИОМЕДИЦИНСКИМ ИСТРАЖИВАЊИМА**

**ВЕЖБА БРОЈ 2**

**ПРЕЛИМИНАРНЕ АНАЛИЗЕ**

Copyright © 2018 – Факултет медицинских наука Универзитета у Крагујевцу. Сва права задржана. Без претходне писмене дозволе од стране Факултета медицинских наука забрањена је репродукција, трансфер, дистрибуција или меморисање неког дела или читавих садржаја овог документа, копирањем, снимањем, електронским путем, скенирањем или на било који други начин.

Copyright © 2018 – Faculty of Medical Sciences of University of Kragujevac. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying,, recording, scanning or otherwise, without the prior written permission of Faculty of Medical Sciences.

**САДРЖАЈ**

5.0 Употреба дијаграма за описивање и анализу података

5.1 Хистограм

5.2 Кружни графикон

5.3 Стубичасти дијаграм

5.4 Линијски дијаграм

5.5 Дијаграм растурања

5.6 Правоугаони дијаграм

5.7 Дорада дијаграма

5.8 Увоз дијаграма у Word-ове документе

6.0 Нормалнa расподела

6.2 Процена нормалности расподеле

6.2 Процена нормалности расподеле у програму PASW Statistics 18

6.3 Откривање нетипичних тачака

6.4 Пример за вежбу

**5.0 Употреба дијаграма за описивање и анализу података**

Разматраћемо следеће врсте дијаграма:

* хистограм (***histogram***),
* кружни графикон (***pie***)
* стубичасти дијаграм (***bar***),
* линијски дијаграм (***line***),
* дијаграм растурања (***scatter/dot***) и
* правоугаони (**кутијасти**) дијаграм (***boxplot***).

Користимо датотеку ***survey3ED.sav***

**5.1 Хистограм**

Хистограмима се приказује расподела једне непрекидне променљиве (на пример, старости или субјективно доживљеног стреса).

|  |  |
| --- | --- |
| Untitled | Untitled |
|  | |

Поступак цртања хистограма:

1. У главном менију кликне се на **Graphs**, затим на **Legacy Dialogs** и изабере се опција **Histogram**.

2. Бирамо променљиву која нас занима и пребацимо је у поље **Variables**. То треба да је непрекидна променљива (на пример, укупан оптимизам).

3. Ако желимо посебне хистограме разних група (на пример, мушкарци/жене), у одељак **Panel by:** треба ставити одговарајућу променљиву (на пример, пол). Сада бирамо **Rows** да бисмо те засебне дијаграме ставили један изнад другог, односно **Column** када желимо да их ставимо један поред другог.

4. Кликнути на **ОК**.

Пример за вежбу: Нацртати хистограм за променљиву старост, груписати према полу, хистограме нацртати један испод другог.

|  |  |
| --- | --- |
| Untitled | Untitled |
|  | |

Облик хистограма показује расподелу вредности непрекидне променљиве. Многе статистичке технике се заснивају на претпоставци да је расподела вредности свих променљивих нормална, тј. да има облик Гаусове криве. Расподела може задржати приближно звонаст облик, али бити асиметрична улево или удесно од средње вредности, или попримити правоугаони облик.

**5.2 Кружни графикон**

Кружни графикон служи за приказивање учесталости код категоријских променљивих.

Поступак цртања кружног графикона:

1. У главном менију кликне се на **Graphs**, затим на **Legacy Dialogs** и изабере се опција **Pie**.

2. У одељку **Data in chart are** кликнути на **Summaries for groups of cases**. Кликнути на дугме **Define**.

3. У пољe **Define Slices by** треба пребацити категоријску променљиву.

4. Кликнути на категоријску променљиву (пол) чије учесталости желимо да прикажемо помоћу кружног графикона (на пример, ***sex***).. Кликом на дугме са стрелицом пребацимо је у поље **Define Slices by**.

5. Кликнути на **ОК**.

|  |  |  |
| --- | --- | --- |
|  | |  |
|  |  | |

**5.3 Стубичасти дијаграм**

Стубичасти дијаграми могу бити једноставни или веома сложени, у зависности од броја укључених променљивих. У основи потребне су две главне променљиве: једна категоријска, а друга непрекидна. Њима се по жељи може додати још једна категоријска променљива.

Поступак цртања стубичастог дијаграма:

1. У главном менију кликне се на **Graphs**, затим на **Legacy Dialogs** и изабере се опција **Bar**. Кликнути на **Clustered**.

2. У одељку **Data in chart are** кликнути на **Summaries for groups of cases**. Кликнути на дугме **Define**.

3. У пољу **Bars represent** изабрати **Other statistics**.

4. Кликнути на непрекидну променљиву која нас занима (на пример, укупан оптимизам). Она би требало да се појави у пољу **Variable (Mean)**. То значи да ће на скали Оптимизам бити приказане средње вредности разних група.

5. Кликнути на прву категоријску променљиву (на пример, ***age***). Кликом на дугме са стрелицом пребацимо је у поље **Category axis**. Та променљива се приказује на апсциси (хоризонталној оси, X оси) стубичастог дијаграма.

6. Кликнути на другу категоријску променљиву (на пример, пол) и пребацити је у поље **Define Clusters by:**. Она ће бити наведена у легенди дијаграма.

7. Кликнути на **ОК**.

|  |  |
| --- | --- |
| Untitled | Untitled |
| Untitled | |

Резултат претходне процедуре је сажет приказ расподеле резултата по изабраним групама (у овом случају, њих чине мушкарци и жене из различитих старосних група). Горњи дијаграм наводи на помисао да жене показују већи степен субјективно доживљеног стреса од мушкараца и да је та разлика израженија у старијим добним групама. У групи особа старих између 18 и 29 година, разлика између одговарајућих вредности за мушкарце и жене врло је мала. Стубичасте дијаграме треба тумачити пажљиво. Увек треба погледати скалу Y (вертикалне) осе. Оно што визуелно изгледа као велика разлика, понекад износи тек неколико јединица мерне скале. Разлика између најмање (мушкарци старости 45 година и више) и највеће вредности (жене старости од 18 до 29 година) износи само три јединице скале.

Пример за вежбу: Нацртати стубичасти дијаграм за исти пример, **Category Axis**: ***agegr5***.

|  |
| --- |
| Untitled |

**5.4 Линијски дијаграм**

Линијски дијаграм приказује средњу вредност једне непрекидне променљиве за више различитих вредности неке категоријске променљиве (на пример, време 1, време 2, време 3).

Поступак цртања линијског дијаграма:

1. У главном менију кликне се на **Graphs**, затим на **Legacy Dialogs** и изабере се опција **Line**.

2. Потврдити поље **Multiple**. У одељку **Data in Chart Are** кликнути на **Summaries for groups of cases**. Кликнути на дугме **Define**.

3. У пољу **Lines represent** изабрати **Other summary function**. Кликнути на непрекидну променљиву која нас занима (на пример, укупан оптимизан) и затим на дугме са стрелицом. Тако ће се пребацити у поље **Mean**. То значи да ће на скали Оптимизма бити приказане средње вредности разних група.

4. Кликнути на прву категоријску променљиву (на пример, старост). Кликом на дугме са стрелицом пребацити је у поље **Category axis**. Та променљива се приказује на хоризонталној оси линијског дијаграма.

5. Кликнути на другу категоријску променљиву и пребацити је у поље **Define Lines by:**. Та променљива биће представљена у легенди дијаграма.

6. Ако дијаграму хоћемо да додамо стубиће сразмерне одступањима, треба кликнути на **Options**.

7. Кликнути на дугме **OK**.

|  |  |
| --- | --- |
| Untitled | Untitled |

|  |
| --- |
| Untitled |

Прво треба погледати утицај старости на субјективно доживљен стрес за сваки пол. Млађи мушкарци као да показују већи степен субјективно доживљеног стреса од средовечних и од старијих мушкараца. За жене, разлика између старосних група није толико изражена. Старије жене су под незнатно мањим стресом од млађих.

Може се размотрити и разлика између мушкараца и жена. Свеукупно, као да мушкарци субјективно доживљавају мање стреса од жена. Иако је разлика за млађу групу веома мала, она расте с повећањем старости испитаника.

Резултати наводе на то да би требало дубље истражити однос старости и субјективно доживљеног стреса за обе групе. Уместо три, посматрамо пет старосних група. Сада је слика знатно јаснија.

|  |  |
| --- | --- |
| Untitled |  |

**5.5 Дијаграм растурања**

Дијаграми растурања се обично користе за истраживање односа између две непрекидне променљиве (на пример, старости и самопоштовања). Препоручљиво је дијаграм растурања нацртати пре рачунања корелације. Дијаграм растурања предочава посматрачу да ли је однос променљивих линеаран или криволинијски.

Дијаграм растурања показује и да ли су променљиве корелиране позитивно или негативно. Код позитивних корелација линија расте (великим вредностима једне променљиве одговарају велике вредности друге), док код негативних опада.

Дијаграм растурања грубо показује и јачину корелације две променљиве. Када је корелација слаба, тачке су без реда расуте посвуда (не чине препознатљив геометријски облик). Када је корелација јака, тачке чине приближно ваљкаст облик, тј. види се гомилање тачака око замишљене праве линије.

Када желимо засебан дијаграм растурања за сваку групу, треба задати категоријску променљиву у одељку **Panel by:** уместо у долеприказаном пољу **Set Markers by**. Уколико нам треба дијаграм растурања за цео узорак (не по групама), треба занемарити упутства наведена у одељку **Set Markers by**.

Поступак цртања дијаграма растурања:

1. У главном менију кликне се на **Graphs**, затим на **Legacy Dialogs** и изабере се опција **Scatter/Dot**.

2. Потврдити поље **Simple Scatter** и затим кликнути на дугме **Define**.

3. Кликнути на прву променљиву која нас занима, то је обично она која се сматра зависном променљивом (на пример, укупан субјективно доживљен стрес).

4. Затим кликнути на дугме са стрелицом да би се изабрана променљива пребацила у поље **Y axis**. Та променљива ће бити приказана на вертикалној оси.

5. Пребацити другу променљиву (на пример, укупна субјективно доживљена самоконтрола унутрашњег стања) у поље **X axis**. Та променљива ће бити приказана на хоризонталној оси.

6. SPSS може да означи сваку тачку и по некој другој категоријској променљивој. Пребацити ту променљиву (на пример, пол) у поље Set Markers by:. Тиме ће тачке које одговарају мушкарцима односно женама бити представљене различитим маркерима (ознакама).

7. У поље **Label Cases by:** пребацити променљиву ID. Тако добијамо прилику да очитамо број анализираног случаја с дијаграма када наиђемо на нетипичну тачку.

8. Кликнути на дугме **ОК**.

|  |  |
| --- | --- |
| Untitled | Untitled |

|  |
| --- |
| Untitled |

Из претходног као да се назире умерена негативна корелација између две посматране променљиве у узорку као целини. Испитаници с високим нивоом субјективно доживљене самоконтроле доживљавају знатно виши ниво стреса.

**5.6 Правоугаони дијаграм**

Правоугаони (кутијасти) дијаграм се употребљава за поређење расподеле вредности променљиве. Може се употребити за испитивање расподеле једне непрекидне променљиве (на пример, позитивних осећања) или се резултати могу разделити на више група (на пример, старосних). За поређење се може додати још једна категоријска променљива (на пример, мушкарци и жене). У доњем примеру истражиће се расподела вредности на скали позитивних осећања за мушкарце и жене.

Поступак цртања правоугаоног дијаграма:

1. У главном менију кликне се на **Graphs**, затим на **Legacy Dialogs** и изабере се опција **Boxplot**.

2. Кликнути на дугме **Simple.** У одељку **Data in Chart Are** потврдити поље **Summaries for groups of cases**, и затим кликнути на дугме **Define**.

3. Кликнути на прву непрекидну променљиву која нас занима (на пример, укупно позитивних осећања). Кликом на дугме са стрелицом пребацује се у поље **Variable**.

4. Кликнути на категоријску променљиву која нас занима (на пример, пол). Кликом на дугме са стрелицом пребацити је у поље **Category axis**.

5. Кликнути на променљиву ID и пребацити је у поље **Label cases**. Тако ћемо добити прилику да с дијаграма очитамо бројеве случајева у којима су се јавиле екстремне вредности.

6. Кликнути на дугме **OK**.

|  |  |
| --- | --- |
| Untitled | Untitled |

|  |
| --- |
| Untitled |

Свака расподела вредности представљена је правоугаоником и линијама (реповима) које из њега излазе. Правоугаоник садржи 50% случајева, а његова дужина је сразмерна интерквартилном распону променљиве. Линија унутар правоугаоника је медијана. Репови иду до навећих и најмањих опажених вредности променљиве.

Пример за вежбу: Нацртати кутијасти дијаграм за променљиву ***tpstress***, Category Axis: ***smoke***.

|  |
| --- |
| Untitled |

Сви резултати које SPSS сматра нетипичним тачкама приказани су кружићима поред којих су бројеви. То су ID бројеви тих случајева. Нетипичне тачке су случајеви чије се вредности превише разликују од остатка узорка, било да су превелики или премали. SPSS сматра нетипичним све тачке које су од ивице правоугаоника удаљене више од 1,5 његових дужина. Екстремне вредности (означене \*) су оне које су од ивице правоугаоника удаљене више од 3 његове дужине. У горњем примеру приказане су за мушкарце и жене свеукупно три нетипичне тачке с малим вредностима променљиве Позитивна осећања.

**5.7 Дорада дијаграма**

Пре него што се дијаграм штампа или стави у извештај, понекад треба дорадити његове наслове, натписе, маркере итд. Дијаграми се дорађују у прозору **Chart Editor**, који се отвара двокликом на дијаграм који треба дорадити. Приказаће се нов прозор с тим дијаграмом, додатним опцијама менија и иконицама. Требало би да се појави и мањи прозор **Properties**, који служи за дораду дијаграма. Ако се не појави, треба на дијаграм кликнути десним кликом и у приручном менију изабрати **Properties Window**.

|  |
| --- |
| Untitled |

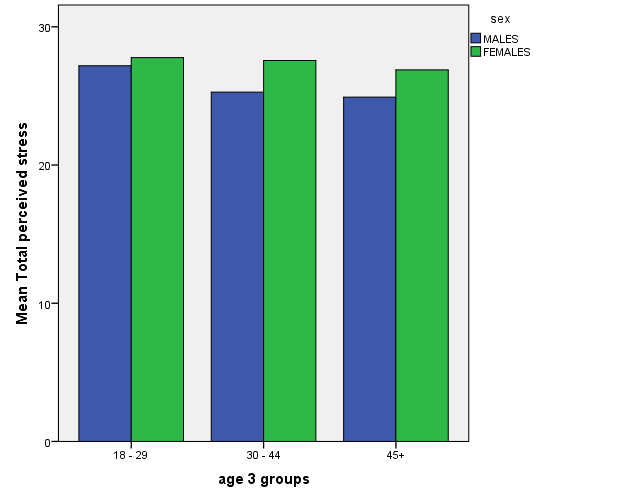
У Chart Editor-у могу се спровести разне измене и дораде:

- Речи у натписима се мењају тако што се једном кликне на натпис да би се изабрао (око текста би требало да се појави плава ивица). Кликне се још једном (требало би да се појави црвени курсор) и текст може да се мења. По завршетку измена треба притиснути тастер ENTER на тастатури.

- Положај натписа дуж X и Y осе се мења двокликом. У овиру **Properties** који ће се појавити, кликнути на језичак Text. У одељку Justify изабрати жељени положај (тачка означава централни положај, стрелица улево помера дијаграм улево, стрелица удесно помера дијаграм удесно).

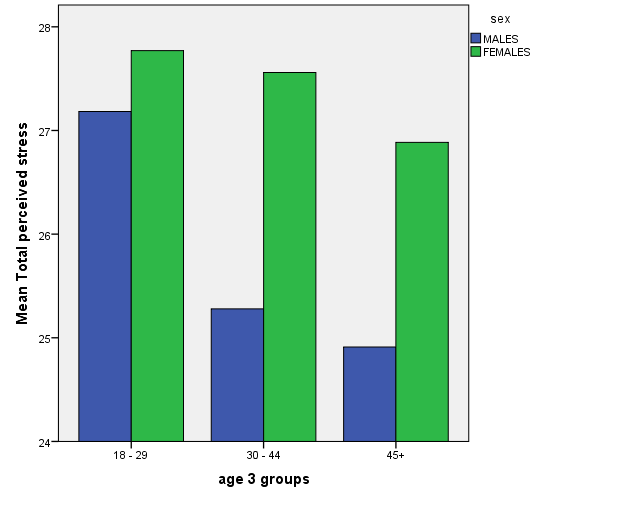
- Да бисмо изменили обележја текста, линије, маркере, боје, попуну позадине или скале, треба једном кликнути на оно својство текста чија се измена жели. У зависности од својства које се изабере, у прозору **Properties** ће се на одговарајући начин изменити опције. Ако се жели измена линија у вишелинијском дијаграму (промена једне у испрекидану и сл.) или маркера одређене групе, ту категорију треба изабрати у легенди а не на самом дијаграму.

Могућности програма показаћемо на примеру стубичастог графикона.

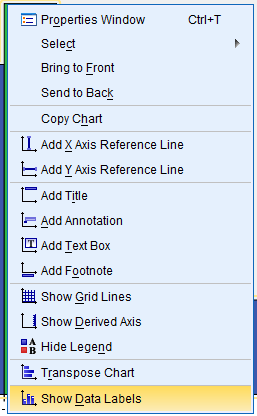


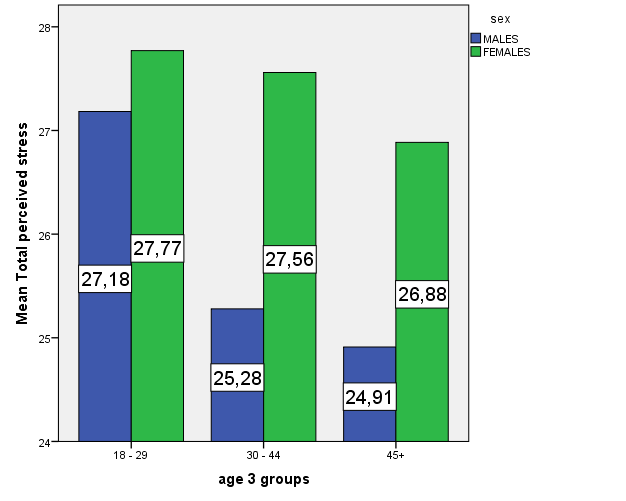
Подешавања у-осе

|  |  |
| --- | --- |
|  |  |



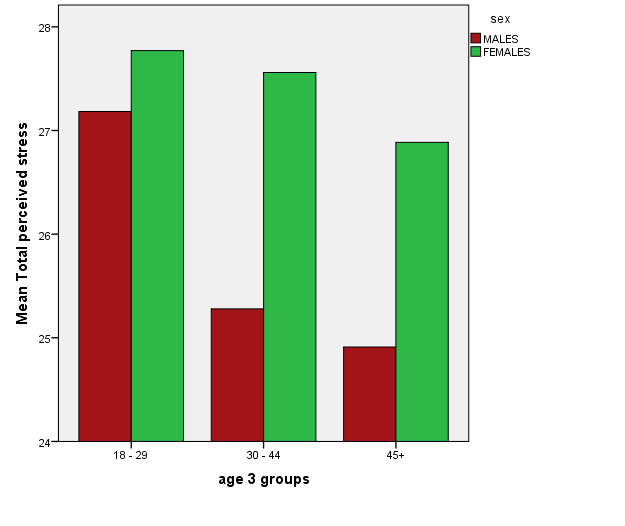
Додавање средњих вредности преко стубића





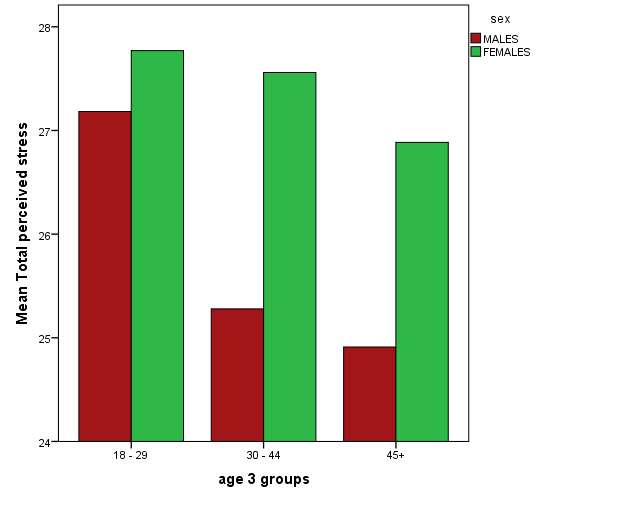
Промена боје

|  |  |
| --- | --- |
|  |  |

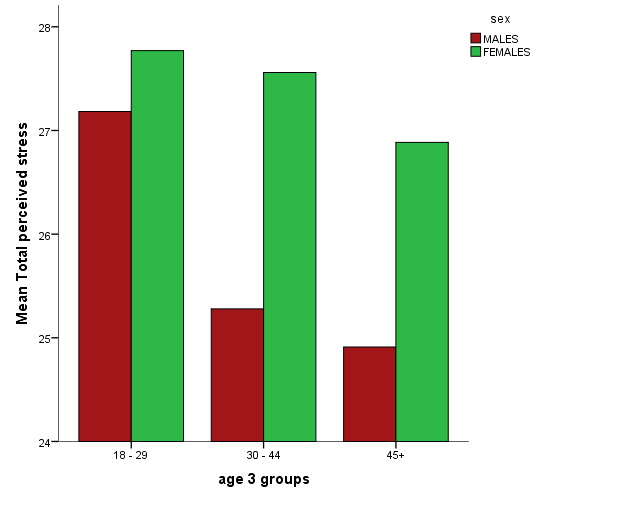


Промена позадине

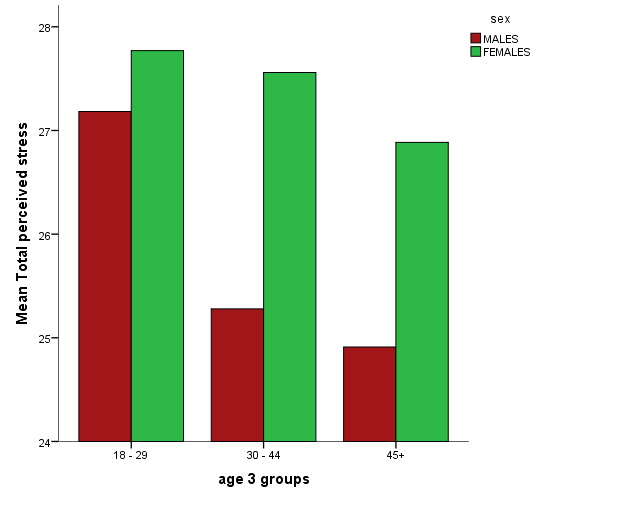
|  |  |
| --- | --- |
|  |  |

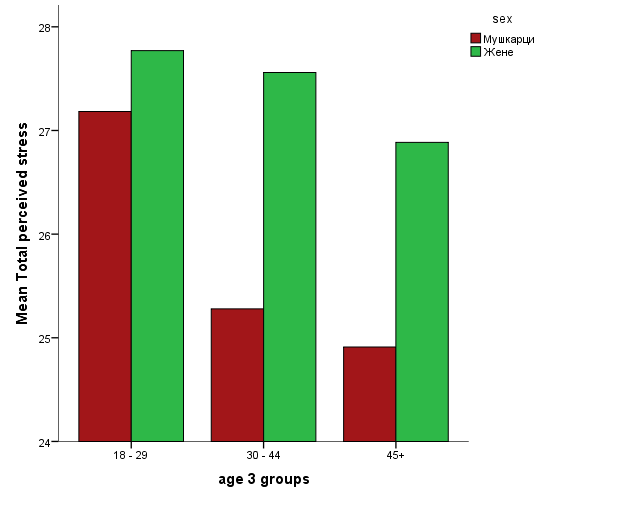


|  |  |
| --- | --- |
|  |  |



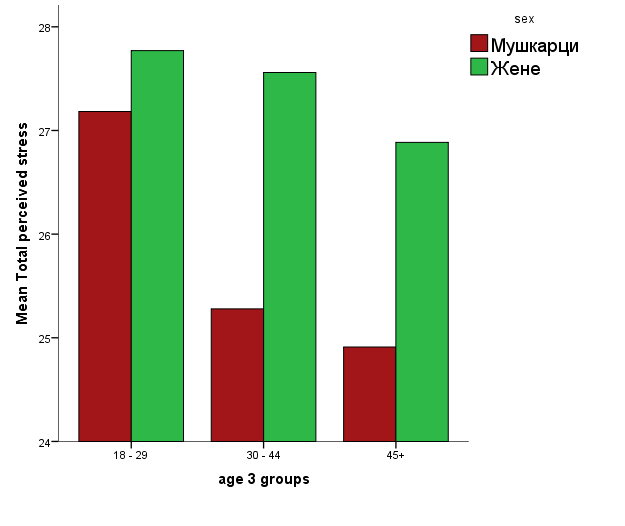
Промена текста





Промена величине фонта

|  |  |
| --- | --- |
|  |  |

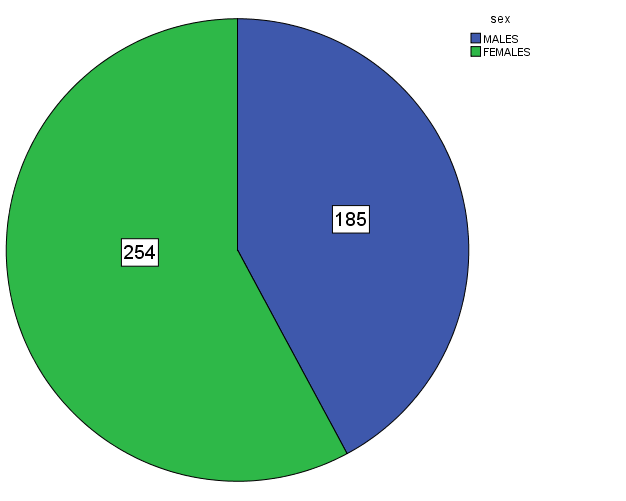
Изглед графикона након примењених измена

|  |  |
| --- | --- |
|  |  |

Додавање броја испитаника код кружног графикона

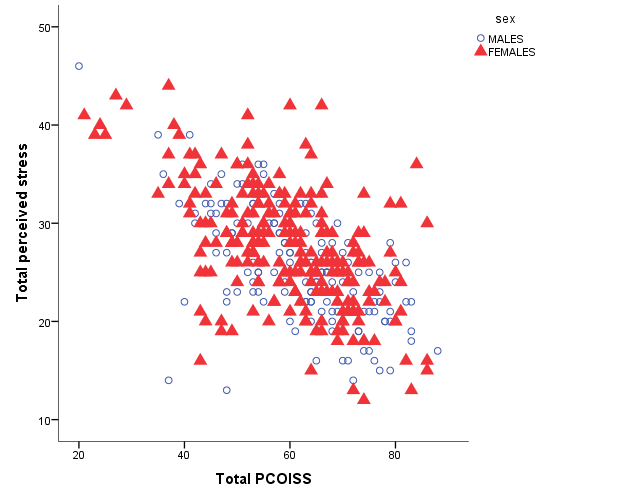
|  |  |
| --- | --- |
|  |  |

|  |  |
| --- | --- |
|  |  |



Промена маркера код дијаграма растурања

|  |  |
| --- | --- |
|  |  |



**5.8 Увоз дијаграма у Word-ове документе**

SPSS омогућава непосредно копирање дијаграма у програм за обраду текста (на пример, Microsoft-ов Word). То се ради када се припреми коначна верзија извештаја и део резултата се жели приказати графички.

Поступак увоза дијаграма у Word-ов документ:

1. Да би увоз дијаграма био могућ и **SPSS** и **Word** морају бити отворени.

2. У SPSS-у је потребно приказати прозор **Viewer**.

3. Кликне се на дијаграм који треба да се копира како би се означио.

4. У главном менију кликнути на **Edit** и изабрати опцију **Copy Objects**.

Или кликнути десним кликом миша на дијаграм и из менија изабрати опцију **Copy**.

5. Активирати Word (већ је отворен и минимизиран).

6. У Word-у поставити курсор на место где се жели увести дијаграм.

7. У главном менију кликнути на **Edit** и изабрати опцију **Paste**. Или на палети са алаткама кликнути на иконицу **Paste**. Или десни клик мишем и из менија изабрати опцију **Paste**.

8. Снимити измене у Word-овом документу. **File** → **Save**.

**6.0 Нормална расподела**

**Нормална (Гаусова) расподела** је важна фамилија непрекидних расподела вероватноћа. Чланови фамилије нормалне расподеле су дефинисани преко два параметра, средине и варијансе.



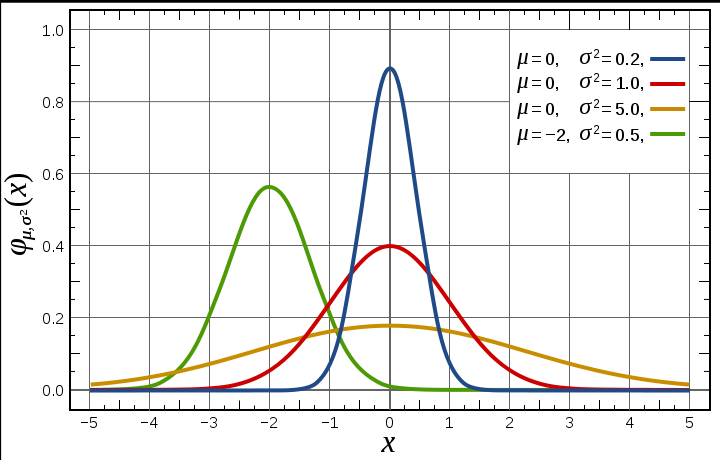
где је:

 - средина

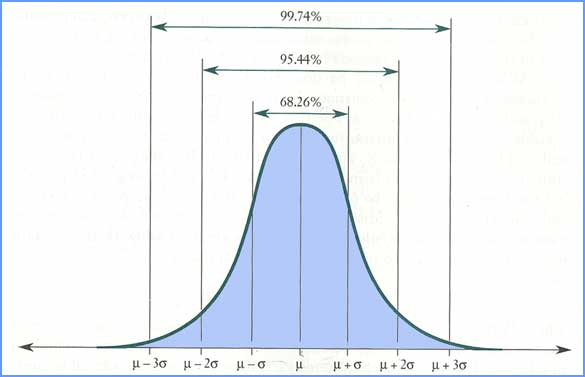
 - варијанса







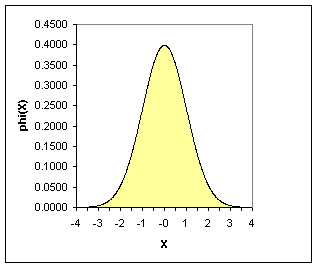
Интервали на растојању 1, 2 и 3 вредности стандардног одступања од средине заузимају 68,26%, 95,44% и 99,74%. Исти проценти важе за сваку нормалну расподелу, без обзира на средину и стандардно одступање.



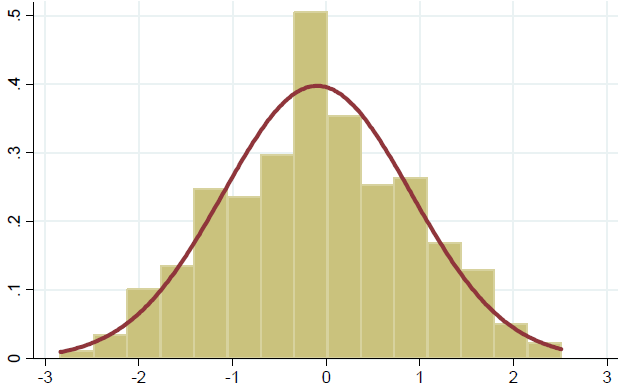
Стандардна нормална расподела је нормална расподела са средином која је једнака нули и варијансом једнаком један.



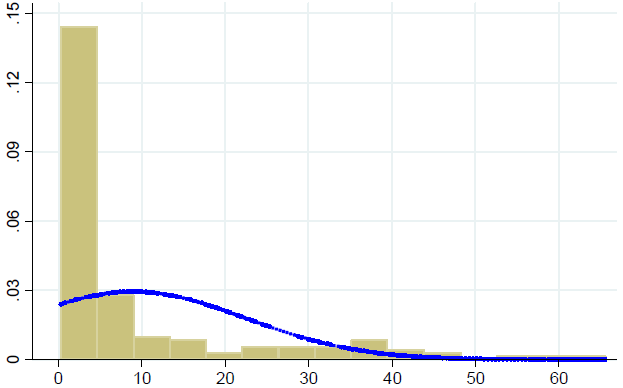
Све нормалне расподеле могу бити конвертоване у стандардну нормалну расподелу одузимањем аритметичке средине од сваке вредности и дељењем са стандардним одступањем.



Пример нормалне расподеле



Пример расподеле која није нормална:



**6.1** **Процена нормалности расподеле**

Нумеричке методе за оцену нормалности расподеле:

Нормална расподела треба да има вредности ***skewness*** (скјунис) и ***kurtosis*** (куртосис) између -1 и 1. Ако су ове вредности веће од 3 и мање од -3 онда то указује на одступање од нормалне расподеле.

Код статистичког тестирања нормалности користи се ***Kolmogorov-Smirnov*** тест и ***Shapiro-Wilk*** тест. Ако је у овим тестовима  онда емпиријска расподела значајно одступа од нормалне.

Графичке методе за оцену нормалности расподеле:

- ***Хистограм*** – визуелна процена да ли је емпиријска расподела слична звонастој симетричној расподели.

**- *Нормални Q-Q графикон* –** Ако је расподела нормална тачке ће бити на правој линији. Одступање тачака од праве линије указује на одступање расподеле од нормалне.

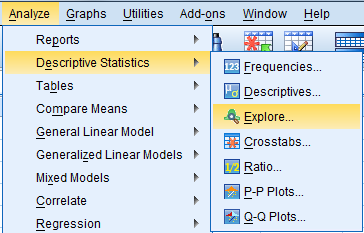
**- *Детрендовани нормални Q-Q графикон*** – Ако је расподела нормална тачке ће бити равномерно распоређене изнад и испод хоризонталне линије.

**- *Правоугаони дијаграм*** – Ако постоји неколико екстремних вредности или нетипичних тачака на било ком крају расподеле то указује на одступање од нормалне расподеле. Ако медијана није у центру правоугаоника, већ знатно ближа једном од крајева, то такође указује на одступање од нормалне расподеле.

**6.2 Процена нормалности расподеле у програму PASW Statistics 18**

Нормална расподела је симетрична, звонолика крива с највећим бројем резултата у средини и мањим бројем резултата према крајевима звона. Нормалност се може донекле проценити на основу израчунатих вредности асиметрије и спљоштености звона. Међутим, SPSS у менију **Descriptive Statistics** има и ставку **Explore** која нуди друге начине.

1. У главном менију кликнути на **Analyze**, затим на **Descriptive Statistics**, па на **Explore**.

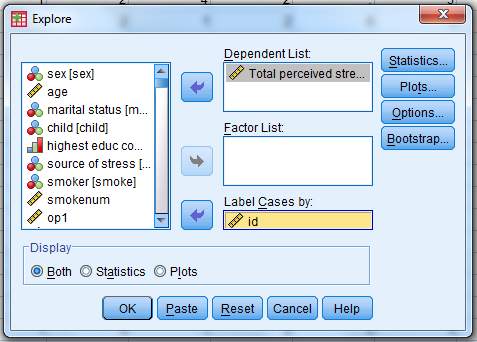


2. Кликнути на жељене променљиве (на пример, Укупан субјективно доживљен стрес) и затим на дугме са стрелицом како би их пребацили у поље **Dependent List**.

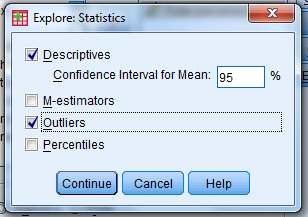
3. У поље **Label Cases by** ставити променљиву ID.

4. У одељку **Display** треба да је потврђено поље **Both**.

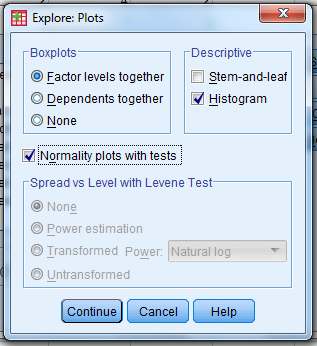
Ако се у прозор ***Factor List*** пребаци променљива која одређује групе, онда се добија резултат испитивања нормалности променљиве у свакој групи.



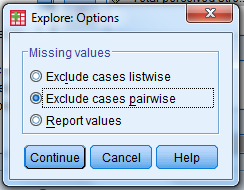
5. Кликнути на дугме **Statistics**, затим на **Descriptives**, па на **Outliers**.



6. Кликнути на дугме **Plots**. У одељку **Descriptive** кликнути на **Histogram**. Означити опцију **Normality plot with tests**. Кликнути на дугме **Continue**.



7. Кликнути на дугме **Options**. У одељку **Missing Values** означити **Exclude cases pairwise**.



8. Кликнути на дугме **Continue**, а затим и на дугме **OK**.

**Добијени резултати**:

У табели ***Case Processing Summary*** можемо очитати колико случајева је програм разматрао, односно колико случајева учествује у овој анализи.

| **Case Processing Summary** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Cases | | | | | |
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| Total perceived stress | 433 | 98,6% | 6 | 1,4% | 439 | 100,0% |

У табели **Descriptives** дати су описни показатељи и друге информације о анализираним променљивима. Део тих информација је: средња вредност, медијана, стандардно одступање, минимум, максимум итд. 5% Trimmed Mean се добија када SPSS занемари горњих и доњих 5% случајева и без њих израчуна нову средњу вредност. Када се упореде првобитна (26,73) и нова средња вредност (26,64), долази се до закључка да ли екстремне вредности много утичу на средњу вредност. Ако се ове две вредности много разликују, требало би додатно испитати екстремне вредности. ID бројеви најекстремнијих случајева приказани су у табели **Extreme Values**. Међу резултатима су и асиметрија и спљоштеност.

| **Descriptives** | | | | |
| --- | --- | --- | --- | --- |
|  | | | Statistic | Std. Error |
| Total perceived stress | Mean | | 26,73 | ,281 |
| 95% Confidence Interval for Mean | Lower Bound | 26,18 |  |
| Upper Bound | 27,28 |  |
| 5% Trimmed Mean | | 26,64 |  |
| Median | | 26,00 |  |
| Variance | | 34,194 |  |
| Std. Deviation | | 5,848 |  |
| Minimum | | 12 |  |
| Maximum | | 46 |  |
| Range | | 34 |  |
| Interquartile Range | | 8 |  |
| Skewness | | ,245 | ,117 |
| Kurtosis | | ,182 | ,234 |

| **Extreme Values** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | | Case Number | id | Value |
| Total perceived stress | Highest | 1 | 10 | 24 | 46 |
| 2 | 19 | 157 | 44 |
| 3 | 9 | 61 | 43 |
| 4 | 14 | 144 | 42 |
| 5 | 18 | 6 | 42a |
| Lowest | 1 | 45 | 5 | 12 |
| 2 | 42 | 404 | 12 |
| 3 | 69 | 119 | 13 |
| 4 | 56 | 301 | 13 |
| 5 | 53 | 127 | 13 |
| a. Only a partial list of cases with the value 42 are shown in the table of upper extremes. | | | | | |

У табели **Tests of Normality** дати су коначни резултати испитивања нормалности расподеле које су изумели Колмогоров и Смирнов. Нормалност се показује статистички незначајним одступањем од нормалности, тј. износом ***Sig.*** већим од 0,05. У овом случају ***Sig.*** износи 0,000 што показује да претпоставка о нормалности расподеле није потврђена и да је морамо одбацити. То је сасвим уобичајено за велике узорке. Када имамо да мање од 50 случајева учествује у анализи користе се подаци из десног дела табеле (***Shapiro-Wilk***).

| **Tests of Normality** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Total perceived stress | ,069 | 433 | ,000 | ,992 | 433 | ,021 |
| a. Lilliefors Significance Correction | | | | | | |

Стварни облик расподеле резултата сваке групе види се на њиховим **хистограмима**. У овом примеру изгледа као да су резултати прилично нормално расподељени. То потврђује и изглед криве нормалне вероватноће **Normal Q-Q Plot**. На њој се опажена вредност резултата црта заједно са очекиваном вредношћу коју би дала нормална расподела. Што је линија опажених резултата ближа правој линији, то је расподела ближа нормалној.





Крива **Detrended Normal Q-Q Plot** приказује стварно одступање опажених резултата од хоризонталне линије која представља нормалну расподелу. Требало би да већина тачака буде приближно симетрично распоређена око те хоризонталне линије и да се тачке нигде не гомилају.



***Правоугаони дијаграм*** – Правоугаоник представља 50% случајева, при чему репови (линије које излазе из правоугаоника) иду до највећих и најмањих вредности. Понекад се виде и изоловани кружићи изван тог распона и они представљају нетипичне тачке. Линија унутар правоугаоника је медијана.



**У овом примеру расподела је била прилично нормална.** То се не догађа баш често.

**6.3 Откривање нетипичних тачака**

Многе статистичке технике осетљиве су на нетипичне тачке (***outliers***) тј. резултате који су много изнад или испод већине осталих резултата.

* Најпре погледајмо **хистограм**, нарочито репове расподеле. Има ли ту усамљених тачака екстремних вредности? Ако има, то су потенцијалне нетипичне тачке.
* Друго, погледајмо **правоугаони дијаграм** (*boxplot*). Све вредности које PASW Statistics 18 сматра нетипичним тачкама, приказане су као кружићи означени бројевима (ID бројеви тих случајева). PASW Statistics 18 сматра нетипичним све тачке које су од ивице правоугаоника удаљене више од 1,5 његових дужина. Екстремне вредности (означене \*) су оне које су од ивице правоугаоника удаљене више од 3 његове дужине. У нашем примеру није било екстремних тачака, али су биле две нетипичне тачке (ID бројеви 24 и 157).
* Треба проверити да ли је вредност нетипичне тачке заиста толика или је у питању грешка. Ако се ради о грешци, треба је исправити и поново нацртати правоугаони дијаграм.
* Информације у табели **Descriptives** показују колики проблем ће представљати случајеви с нетипичним тачкама. Погледајте средњу вредност израчунату без горњих и доњих 5% случајева, тј. **5% Trimmed Mean**. Када се та и права средња вредност много разликују, те тачке треба још испитати. У овом примеру су те две средње вредности (26,73 и 26,64) веома блиске.
* Када затреба да се измене или уклоне неки подаци из датотеке, прелази се у **Data Editor** и сортирају се подаци по опадајућем или растућем редоследу.

**6.4 Пример за вежбу**

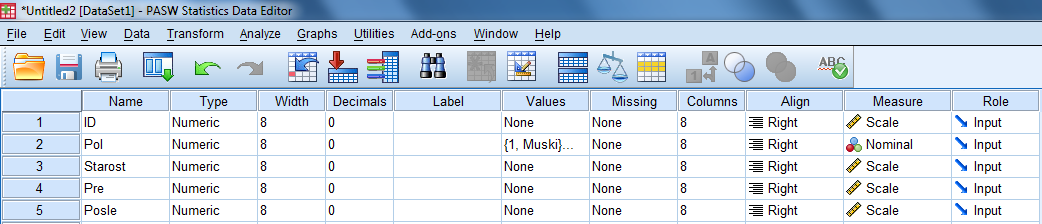
**Пример 1:**

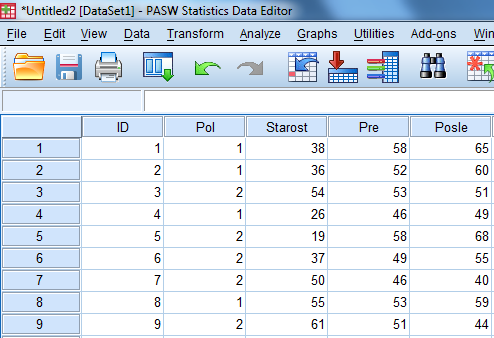
Код 9 испитаника измерене су вредности албумина пре и после третмана.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Испитаник | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Пол | М | М | Ж | М | Ж | Ж | Ж | М | Ж |
| Старост | 38 | 36 | 54 | 26 | 19 | 37 | 50 | 55 | 61 |
| Пре | 58 | 52 | 53 | 46 | 58 | 49 | 46 | 53 | 51 |
| После | 65 | 60 | 51 | 49 | 68 | 55 | 40 | 59 | 44 |

Одредити да ли вредности албумина пре и после прате нормалну расподелу.

**Решење:**





Analyze → Descriptive Statistics → Explore

|  |  |
| --- | --- |
| Untitled | Untitled |
|  | Untitled |

| **Tests of Normality** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Pre | ,169 | 9 | ,200\* | ,916 | 9 | **,362** |
| Posle | ,126 | 9 | ,200\* | ,974 | 9 | **,930** |
| a. Lilliefors Significance Correction | | | | | | |
| \*. This is a lower bound of the true significance. | | | | | | |

С обзиром да имамо мање од 50 податка посматрамо резултате Shapiro-Wilk-овог теста нормалности.

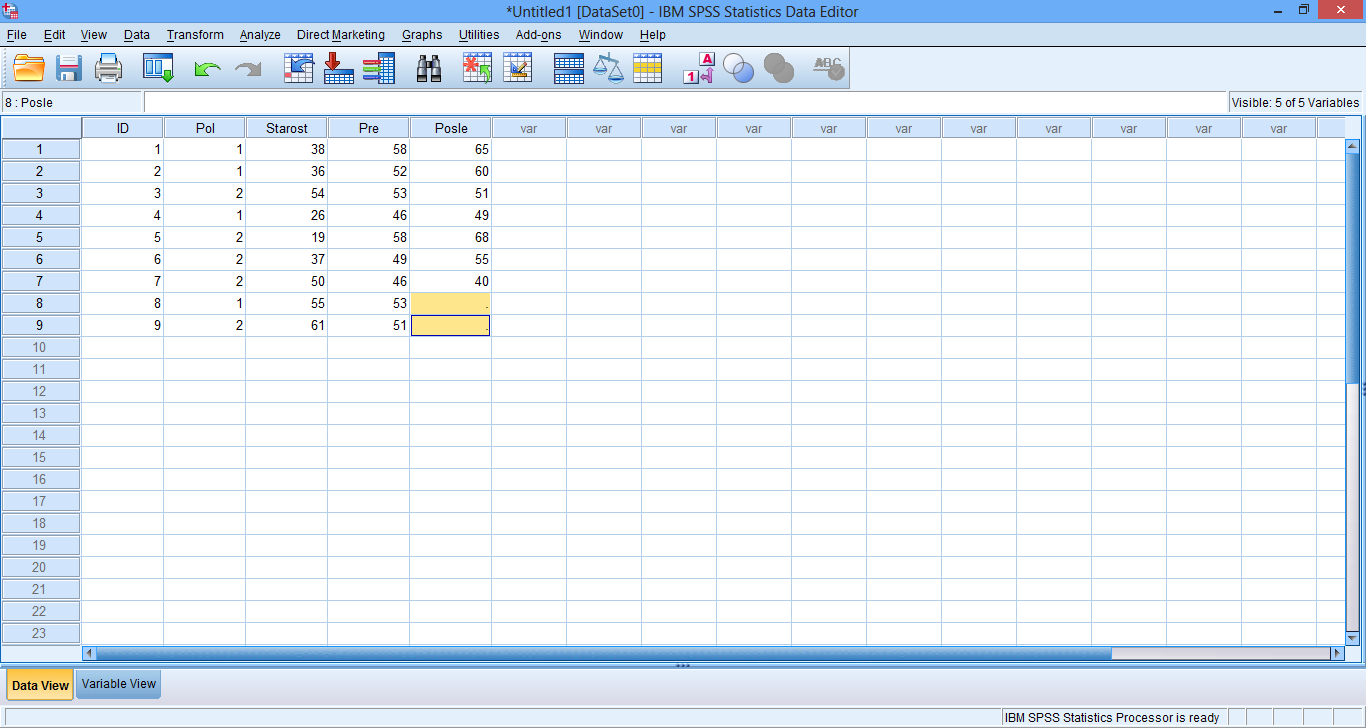
За вредности „пре“ – 0,362>0,05

За вредности „после“ – 0,93>0,05

**Значајност за обе променљиве је већа од 0,05 тако да закључујемо да вредности албумина пре и после третмана прате нормалну расподелу**.

**Пример 2:**

Из претходног примера обрисати последње две вредности из варијабле „После“.



|  |  |
| --- | --- |
| Untitled | Untitled |
|  | Untitled |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Pre | ,179 | 7 | ,200\* | ,889 | **7** | ,269 |
| Posle | ,123 | 7 | ,200\* | ,975 | **7** | ,930 |

Из поставке задатка се види да за варијаблу „Пре“ имамо девет података а у табели је приказано да је прорачун извршен само за 7 података. Грешка је што нисмо дефинисали шта програм да ради са недостајућим подацима.

Тачно решење:

|  |  |
| --- | --- |
| Untitled | Untitled |
| Untitled |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Pre | ,169 | 9 | ,200\* | ,916 | 9 | ,362 |
| Posle | ,123 | 7 | ,200\* | ,975 | 7 | ,930 |

С обзиром да имамо мање од 50 податка посматрамо резултате Shapiro-Wilk-овог теста нормалности.

За вредности „пре“ – 0,362>0,05

За вредности „после“ – 0,93>0,05

**Значајност за обе променљиве је већа од 0,05 тако да закључујемо да вредности албумина пре и после третмана прате нормалну расподелу**.