

# Статистички софтвер 4

## Други час

Марија Радичевић

Математички факултет, Београд

2015.

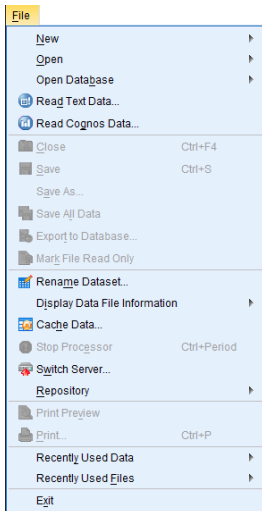
# Садржај

- 1 Главне команде
  - File
  - Edit
  - View
  - Data
  - Transform
- 2 Команде трансформације SPSS командног језика
- 3 Преузимање података из других програмских окружења
- 4 Примери

# Главни мени у прозору едитора података

- 1 File
- 2 Edit
- 3 View
- 4 Data
- 5 Transform
- 6 Analyze
- 7 Direct marketing
- 8 Graphics
- 9 Utilities
- 10 Add-ons
- 11 Window
- 12 Help

# File



File омогућава рад са датотекама података.

*Open database* учитавање података из система за рад са базама података

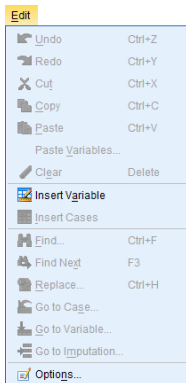
*Read Text data* учитавање података из текстуалне датотеке

*Export to Database* пренос података из SPSS датотеке у изабрану базу података

*Display Data File Info* приказ података о посебно наведеној датотеци података

- Working File
- External File

# Edit



Edit омогућава премештање, копирање, преношење текста и дијаграма из активног прозора у програм за привремено чување садржаја, као и преузимање садржаја из њега, брисање, претраживање, итд.

*Insert Variable* отвара се нова колона за унос нове променљиве

*Insert Cases* убацује се нови ред

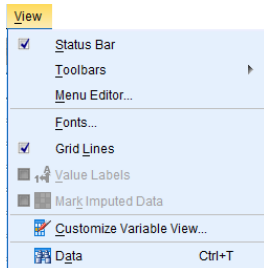
*Go to Variable* прелази се на колону (променљиву) са наведеним редним бројем

*Go to Case* прелази се на ред (опсервацију) са наведеним редним бројем

*Go to Imputation* активна команда само у случају рада са недостајућим подацима

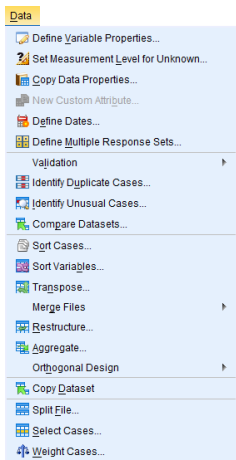
*Options* подешавања SPSS програмског окружења

# View



View се користи за подешавање радног окружења.

# Data



Data садржи скуп команди и њихов подсистем које пружају разноврсне могућности за рад са подацима.

*Define Variable Properties* отвара се комуникациони прозор за одређивање атрибута променљиве

*Copy Data Properties* копирање свих атрибута једне променљиве на другу променљиву

*Define Dates* генерисање датумеске променљиве

*Define Multiple Response Sets* формира се низ за вишеструке одговоре, тј. више променљивих

*Validation* контрола исправности података у складу са постављеним правилима

- 1 претходно дефинисана правила
- 2 дефинисање правила
  - *Single – Variable Rules*
  - *Cross – Variable Rules*
- 3 спровођење контроле

*Identify Duplicate Cases* проверава за задату променљиву да ли постоје опсервације са истом вредношћу

*Identify Unusual Cases* открива "необичне" опсервације које се значајно разликују у односу на друге опсервације

*Sort Cases* сортира редове (опсервације) активне датотеке података према вредностима изабране једне или више променљивих

*Sort Variables* сортирање колона (променљивих) и то по било ком атрибуту променљивих

*Transpose* креирање новог едитора података са транспонованим редовима и колонама

*Restructure* формирање нове структуре података на бази већ постојеће табеле података

- превођење изабране опсервације у преоменљиве
- превођење изабране променљиве у опсервације
- транспоновање свих података



*Merge Files* удруживање података из две датотеке: активне датотеке података и датотеке података сачуване у *SPSS* формату

- удруживање датотека података које садрже исту променљиву а различите опсервације
- удруживање датотека података које садрже исте опсервације а различите променљиве

*Aggregate* формирање нове датотеке агрегираних података тако што се врши агрегирање свих опсервација изабраних променљивих у активној датотеци података, а према вредности променљиве која је изабрана за променљиву прелома

*Split File* подела активне датотеке података на више делова

*Select Cases* издвајање опсервација активне датотеке података које ће се даље процесуирати

- дефинисање логичког услова
- избор случајног узорка
- одређена област опсервација
- филтер променљива

*Weight Cases* додељује пондере опсервацијама, да би тако пондерисане вредности опсервација ушле у следећи поступак обраде и анализе

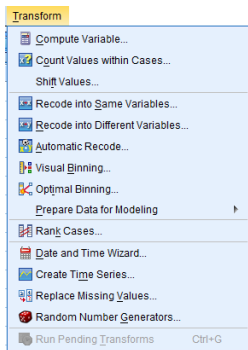
# Transform

Transform садржи систем команди које пружају различите могућности за трансформисање променљивих и података.

*Compute* рачуна вредности нове или већ постојеће променљиве применом нумеричких израза у којима се могу појавити променљиве из активне датотеке података, неке од уграђених функција, аритметички и логички оператори, као и константе

*Count Values Within Cases* формирање нове променљиве у активној датотеци података која броји појављивање одређене вредности за изабране променљиве посебно за сваку опсервацију

*Shift Values* формирање нове променљиве која садржи вредности већ постојеће променљиве, али померене у односу на првобитну променљиву



*Recode into Same Variables* промена вредности (прекодирање) изабране(их) променљиве(их) према наведеној спецификацији

*Record into Different Variables* слично као претходна команда, с том разликом што се вредности изабране променљиве прекодирају и додељују новој променљивој

*Automatic Recode* активирање процедуре којом се за изабрану променљиву (променљиве) формира нова променљива (променљиве) чије су вредности цели бројеви

*Visual Binning* формирање категоричке променљиве на бази изабране нумеричке променљиве

*Optimal Binning* подела (једне или више) непрекидне променљиве на групу

*Prepere Data for Modeling* извршавање алгоритма аутоматизоване припреме података

- Interactive
- Automatic
- Backtransform Score

*Rank Cases* активирање процедуре која за наведену(е) нумеричку(е) променљиву(е) формира нову променљиву која садржи ранг, нормалне скорове или неке друге ранг величине

- Rank
- Savage score
- Fractional rank
- Sum of case weights
- Ntiles  $g_i = \text{truncate}(1 + \frac{R^n_i}{W+1})$
- Proportional estimates
- Normal Score

*Data and Time Wizard* спровођење операција које се најчешће користе у раду са датумским и временским величинама

*Create Time Series* покретање процедуре за креирање нових променљивих од изабраних променљивих применом неке функције временске серије

*Replace Missing Values* формирање нове променљиве (променљивих) код које је извршена замена недостајућих података за изабрану променљиву (променљиве)

*Random Number Generators* избор једног од два понуђена генератора случајних бројева

# Садржај

- 1 Главне команде
  - File
  - Edit
  - View
  - Data
  - Transform
- 2 Команде трансформације SPSS командног језика
- 3 Преузимање података из других програмских окружења
- 4 Примери

## Команде трансформације

*IF*: команда за условну трансформацију

IF [логички израз] циљна променљива=израз.

*DO IF-END IF*: омогућава спровођење више трансформација

на групи опсервација

DO IF[логички израз].

команде трансформације.

ELSE IF [логички израз].

команде трансформације.

...

ELSE.

команде трансформације.

END IF.

*LOOP-END LOOP*: трансформације се понављају на истој

опсервацији док се не постигне услов за излазак

из структуре

LOOP.

команде трансформације.

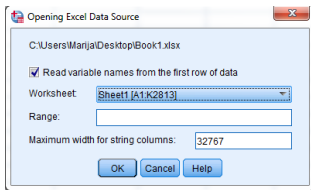
END LOOP.

# Садржај

- 1 Главне команде
  - File
  - Edit
  - View
  - Data
  - Transform
- 2 Команде трансформације SPSS командног језика
- 3 Преузимање података из других програмских окружења
- 4 Примери

## Преузимање података из:

- Excel: File⇒Open⇒Data



- Текстуални документ File⇒Read Text File
- програма за управљање базама података File⇒Open Database



# Преузимање података из текстуалног документа

**Step 1 of 6**

How do you wish to import the data?

This wizard will help you read data from your textfile and specify information about the variables.

Does your text file match a predefined format?

Yes

No

Text file: D:\statsoft\SPSS\examprg\Zdrava 2008-2009\data.txt

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																																																																																																																																	
Belgians:Belgium	Spain	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100

**Step 2 of 6**

How are your variables separated?

Delimited - Variables are determined by a specific character (i.e., comma tab).

Fixed width - Variables are aligned in fixed-width columns.

How variable names included at the top of your text?

Yes

No

Text file: D:\statsoft\SPSS\examprg\Zdrava 2008-2009\data.txt

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																																																																																																																																	
Belgians:Belgium	Spain	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100

**Step 3 of 6**

Which variables should be imported into the data view?

All

Exclude

Selection

Single page

Exclude page

Other

Data preview:

Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20	Y21	Y22	Y23	Y24	Y25	Y26	Y27	Y28	Y29	Y30	Y31	Y32	Y33	Y34	Y35	Y36	Y37	Y38	Y39	Y40	Y41	Y42	Y43	Y44	Y45	Y46	Y47	Y48	Y49	Y50	Y51	Y52	Y53	Y54	Y55	Y56	Y57	Y58	Y59	Y60	Y61	Y62	Y63	Y64	Y65	Y66	Y67	Y68	Y69	Y70	Y71	Y72	Y73	Y74	Y75	Y76	Y77	Y78	Y79	Y80	Y81	Y82	Y83	Y84	Y85	Y86	Y87	Y88	Y89	Y90	Y91	Y92	Y93	Y94	Y95	Y96	Y97	Y98	Y99	Y100																																																																																																																																	
Belgians	Spain	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100

**Step 4 of 6**

How many cases do you wish to import?

All of the cases

The first cases

A specific percentage of the cases (representable %)

Data preview:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																																																																																																																																	
Belgians:Belgium	Spain	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100

**Step 5 of 6**

How do you wish to handle missing values?

Exclude

Replace

Single page

Exclude page

Other

Data preview:

Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20	Y21	Y22	Y23	Y24	Y25	Y26	Y27	Y28	Y29	Y30	Y31	Y32	Y33	Y34	Y35	Y36	Y37	Y38	Y39	Y40	Y41	Y42	Y43	Y44	Y45	Y46	Y47	Y48	Y49
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

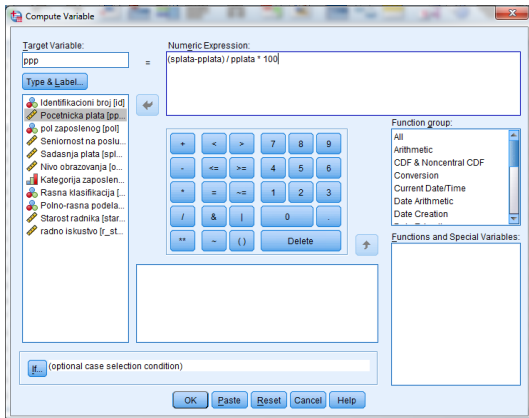
# Садржај

- 1 Главне команде
  - File
  - Edit
  - View
  - Data
  - Transform
- 2 Команде трансформације SPSS командног језика
- 3 Преузимање података из других програмских окружења
- 4 Примери

## Пример 1

У датотеци *baza.sav* формирати променљиву *ppp* (процент пораста плате) која показује колико је процената порасла садашња у односу на почетну плату за сваког запосленог. Затим формирати категоријску променљиву *kpp* (категорија пораста плате) којом се службеници групишу у 4 категорије у зависности од процента пораста плате (до 50%, од 50.01 – 100%, од 100.01 – 150%, преко 150%)

- ppp: Transform  $\Rightarrow$  Compute Variable



- kpp:Transform⇒Visual Binning

Visual Binning

Scanned Variable List: ppp

Name: Current Variable: ppp Label: ppp (Binned)

Binned Variable: kpp

Minimum: 24,76 Nonmissing Values Maximum: 407,94

Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

Grid:	Value	Label
1		<= 50,00
2		100,00 50,01 - 100,00
3		150,00 100,01 - 150,00
4		HIGH 150,01+
5		

Cases Scanned: 474

Missing Values: 0

Copy Bins: From Another Variable, To Other Variables...

Upper Endpoints:  Included (<=),  Excluded (<)

Make Cutpoints, Make Labels, Reverse scale

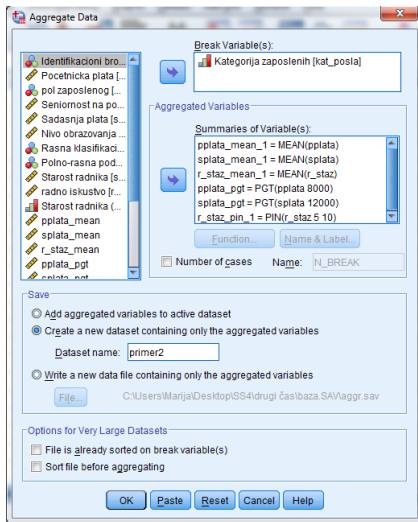
OK, Paste, Reset, Cancel, Help

## Пример 2

За сваку категорију посла запослених приказати следеће:

- Просечну вредност почетне плате
- Просечну вредност садашње плате
- Просечну вредност радног искуства
- Процент запослених чија је садашња плата већа од 12000
- Процент запослених чија је почетна плата већа од 8000
- Процент запослених са радним искуством између 5 и 10 година
- Број запослених у свакој категорији

## Data ⇒ Aggregate



# Хвала на пажњи!