



# Statistika 2 z računalniško analizo podatkov



# Statistika 2 z računalniško analizo podatkov

1. Spoznavanje programa SPSS: osnove, statistične analize (snov predmeta Statistika).
2. Napredne statistične analize z SPSS-jem:
  - nadgradnja preverjanja domnev o razliki aritmetičnih sredin (t testi, analize variance);
  - nadgradnja regresijske analize: multipla in polinomska linearna regresija;
  - nadgradnja preverjanja domnev: neparametrični testi.



# Prvo predavanje

- I. Statistični programski paketi
- II. Osnove SPSS-ja
  1. Uvod
  2. Okna
  3. Potek reševanja
- III. Priprava podatkov v SPSS-ju
  1. Osnovne definicije
  2. Data Editor in delovna sistemska SPSS datoteka
  3. Vnos ali prenos podatkov
  4. Definiranje spremenljivk



## I.1 Splošni statistični prog. paketi

[http://en.wikipedia.org/wiki/Comparison\\_of\\_statistical\\_packages](http://en.wikipedia.org/wiki/Comparison_of_statistical_packages)

- SPSS: Najpogosteje uporabljen v družboslovju.
- SAS: Najprestižnejši in najdražji.
- Stata: Prevladuje v ekonomiji in poslovnih vedah, pa tudi v akademskem svetu in politologiji.
- R: Odprtokodni program, najpogostejši program za razvoj statističnih metod.
- S-Plus: Komercialni “brat” R-ja, prevladuje v biologiji in medicini.

## I.1 Splošni statistični prog. paketi

Paket	Cene licence na leto v USD	
	Komercialna	Akademsko
SAS	~ 6000 na sedež (vezija za PC) ~ 28000 na procesor (Windows server)	?
S-Plus	2500	?
SPSS	1800	700 (Študentska: 200)
Stata	1250	600
R	Brezplačna	



## II. Osnove SPSS-ja

### II.1 Uvod

- SPSS – paket za statistično obdelavo podatkov z osebnimi računalniki.
- V zadnjem času se je preimenoval v *PASW Statistics* in nato “nazaj” v *IBM SPSS Statistics*.
- Sestavljen iz modulov, namenjenih reševanju različnih nalog (pomembno, če kupujete licenco – katere module vključuje).
- Moduli vsebujejo (pod)programe/procedure za individualne naloge.
- Uporabnik pripravi navodila za izvajanje (pod)programov s pomočjo menijev ali z zapisom programskih stavkov.
- Delo podobno kot v drugih programskih paketih za Okna.



## II.2 Okna

- Več oken.
- Vsebina vsakega okna tvori samostojno datoteko.
- Osnovna tri okna (datoteke):
  1. Okno s podatki (Data Editor, končnica *sav*).
  2. Okno z rezultati analiz (Viewer, končnica *spo/spv* – odvisno od verzije programa).
  3. Okno z navodili za izvajanje analiz – programski stavki (Syntax Editor, končnica *sps*).



## II.3 Potek

- *Priprava – opis raziskovalnega problema (vsebina raziskovanja in podatki, ki jih bomo analizirali).*
- Klic SPSS-a in prilagoditev začetnih nastavitev.
- Prenos ali vnos izbranih podatkov v SPSS, pregled podatkov in prilagoditve načinu dela ter raziskovalnemu problemu.
- Izbira enot, spremenljivk in metod za rešitev posameznih nalog (delov raziskovalnega problema).
- Pregled, interpretiranje, urejanje in izpis rezultatov.
- *Zaključek – rešitev raziskovalnega problema.*





# Prvo predavanje

- I. Statistični programski paketi
- II. Osnove SPSS-ja
  1. Uvod
  2. Okna
  3. Potek reševanja
- III. Priprava podatkov v SPSS-ju
  1. Osnovne definicije
  2. Data Editor in delovna sistemska SPSS datoteka
  3. Vnos ali prenos podatkov
  4. Definiranje spremenljivk



## III.1 Osnovne definicije

- Enota – posamezni preučevani element – 1 vrstica v SPSS datoteki s podatki.
- Spremenljivka – lastnost enot – 1 stolpec v SPSS datoteki s podatki.

Npr. Zbrali smo podatke o spolu, starosti, letniku študija ter oddaljenosti doma od kraja študija za 30 študentov. Koliko vrstic in koliko stolpev bo obsegala SPSS podatkovna datoteka?



## III.2 *Data Editor* in delovna sistemska datoteka

- Okno, v katerem vpisujemo, urejamo oz. pregledujemo podatke = *Data Editor*.
- 2 načina pogleda okna *Data Editor*.
  1. *Data View*: baza podatkov (vrednosti spremenljivk po enotah): 1 vrstica = 1 enota, 1 stolpec = 1 spremenljivka
  2. *Variable View*: podatki o spremenljivkah: 1 vrstica = 1 spremenljivka, 1 stolpec = 1 opis spremenljivke (npr. ime, format, labela, merska lestvica, itd.)
- Datoteka s podatki in njihovimi opisi = delovna sistemska datoteka, končnica *sav*.



### III.3 Vnos ali prenos podatkov

3 možnosti vnosa, prenosa podatkov v okno *Data Editor*.

1. Uporabimo SPSS sistemsko datoteko s končnico *sav*.
2. Podatke prenesemo iz drugih programskih paketov (npr. Excela).
3. Podatke vnesemo (vtipkamo, kopiramo) v oknu *Data Editor*.



## II.4 Definiranje spremenljivk

Spremenljivka je definirana z naslednjimi opisi (določamo in vidimo v oknu *Data Editor*, v pogledu *Variable View*) :

- Ime (*Name*): krajše ime
- Tip glede na zapis vrednosti (*Type*): opisna s. (*String*), številka spremenljivka (*Numeric*), itd.
- Število znakov (*Width*): potrebno število znakov za zapis vrednosti
- Število decimalnih mest (*Decimals*), če gre za številsko spremenljivko (sicer ima privzeto vrednost 0)
- Daljši opis spremenljivke (*Label*)
- Opis vrednosti spremenljivke (*Values*)
- Uporabniške manjkajoče vrednosti (*Missing*)
- Širina stolpca (*Columns*): širina stolpca v pogledu *Data View*
- Poravnava niza v stolpcu (*Align*)
- Tip glede na način merjenja (*Measure*): nominalna (*nominal*), ordinalna (*ordinal*), intervalna ali razmernostna (*scale*)



## II.4 Definiranje spremenljivk

- Ime (*Name*): krajše ime

- Ime se mora začeti s črko ali enim od sledečih znakov: @, #, or \$, ostali znaki pa so lahko črke, številke, pika (.) in večino ostalih znakov, razen ločil (presledek, vejica,...).
- Dolžina imena je v verziji 18 omejena na 32 oz. 64 znakov (64 znakov je možnih, če uporabljamo le ANSI znake). V zgodnjih verzijah je omejitev 8 znakov.

- Tip glede na zapis vrednosti (*Type*) – mi bomo uporabljali predvsem sledeče:

- Opisna spremenljivka (*String*) – vnašamo črke ali besedilo (s temi ne moremo računati, tudi če v tako spremenljivko vpisujemo samo številke)
- Številska spremenljivka (*Numeric*) – vnašamo samo številke



## III.4 Definiranje spremenljivk

### ■ Število znakov (*Width*):

- Potrebno število znakov za zapis vrednosti.
- Pri številskih spremenljivkah to vključuje tudi decimalne številke in ločilo.

### ■ Število decimalnih mest (*Decimals*):

- Uporabno samo, če gre za številsko spremenljivko (sicer ima privzeto vrednost 0).
- Število mora biti obvezno manjše od števila znakov (*Width*).

### ■ Daljši opis spremenljivke (*Label*):

- Daljše ime/opis spremenljivke.
- Vsebuje lahko 256 oz. 128 znakov (glej opis za ime).

### ■ Opis vrednosti spremenljivke (*Values*):

- Nastavimo jih lahko pri številskih in opisnih spremenljivkah, npr.: 1 = moški, 2 = ženska, ...



## III.4 Definiranje spremenljivk

- Uporabniške manjkajoče vrednosti (*Missing*):
  - Tu povemo SPSS-u, katere vrednosti označujejo manjkajoče podatke (npr., 99 pogosto uporabljamo za odgovor “ne vem”).
  - To lahko povemo na 2 načina:
    1. Navedemo do 3 vrednosti, ki označujejo (uporabniške) manjkajoče vrednosti, npr: 99, 88, 77.
    2. Navedemo interval, na katerem vse vrednosti označujejo manjkajoče vrednosti (npr.: od vključno -99 do -1 – se pravi vse negativne vrednosti). Zraven lahko opredelimo še eno dodatno vrednost.
- Širina stolpca (*Columns*): širina stolpca v pogledu *Data View*.
- Poravnava niza v stolpcu (*Align*).





## III.4 Definiranje spremenljivk

- Tip glede na način merjenja oz. merska lestvica (*Measure*) – možne so sledeče vrednosti:
  - nominalna (*nominal*)
  - ordinalna (*ordinal*)
  - intervalna ali razmernostna (*scale*)
  
  - OPOZORILO: Merska lestvica pri novejših procedurah v SPSS-u vpliva tudi na to, kaj lahko in česa ne morete početi z določeno spremenljivko.
  
- Vloga spremenljivke (*Role*) – Nekatera (naprednejša) pogovorna okna prikažejo seznam le tistih spremenljivk, ki imajo definirano potrebno vlogo. Zaenkrat tega ne bomo definirali.



## II.5 Prilagoditev SPSS-ovih nastavitev

SPSS-ove nastavitve lahko spreminjamo v oknu *Edit – Options*. Nekaj pomembnejših nastavitev:

- Zavihek *General*:
  - *Variable lists* – izberemo način prikaza spremenljivk v pogovornih oknih (ime ali opis) ter njihov vrstni red (kot v datoteki ali po abecedi)
  
- Zavihek *File Locations*:
  - *Session Journal* – Lahko vidimo in spremenimo lokacijo “dnevnika”, kamor se shranjujejo vsi ukazi, ki jih izvedemo.



## III.5 Prilagoditev SPSS-ovih nastavitev

### ■ Zavihek *Viewer*:

- *Initial Output State* – Nastavimo, kaj je vidno v oknu z rezultati. Priporočljivo je vklopiti *Display commands in the log* (v oknu z rezultati se nam izpiše pred rezultati tudi koda, uporabljena za pridobitev rezultatov).
- Desna stran – nastavitve pisav

## III.5 Prilagoditev SPSS-ovih nastavitev

### ■ Zavihek *Output Labes*.

- Tu nastavimo, kako se izpisujejo spremenljivke in njihove vrednosti.
- Izbiramo ločeno za izpis v stranskem “kazalu” (*Outline Labeling*) in ločeno za glavni izpis (*Pivot Table Labeling*).
- Za spremenljivke lahko izberemo:
  - Izpis samo imena (kratko ime v polju *Name*)
  - Izpis samo opisa (*Label*)
  - Oboje
- Za vrednosti spremenljivk je podbono:
  - Samo vrednost
  - Samo opis vrednosti (*Value Labels*)
  - Oboje