



UNIVERZITET SINGIDUNUM

DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE

SAVREMENE INFORMACIONE TEHNOLOGIJE

MASTER RAD

Statistical Package for the Social Sciences

**Statistički paketi za analizu podataka sa osvrtom na
primenu IBM SPSS programa**

Mentor

Prof. dr Aleksandar Jevremović

Student

Marko Draganović
410289/2014

Beograd, 2016.

SADRŽAJ

UVOD	5
Predmet istraživanja.....	6
Metodologija i ciljevi istraživačkog rada.....	6
Hipoteze istraživanja.....	7
1. R-Programiranje.....	8
1.1. Šta je R?	8
1.2. Istorija R-a.....	8
1.3. R okruženje i njegove osobine	9
1.4. R je jezik.....	10
1.5. Široko dostupan.....	10
1.6. Editori R koda	10
1.7. Problemi	11
2. Kako R funkcioniše	11
3. MATLAB	14
3.1. Uvod u matlab	14
3.2. Šta je Matlab?.....	15
3.3. Primena Matlab-a	16
3.4. Alati (Toolbox-ovi)	16
3.5. Računarske platforme.....	16
3.6. Sistem MATLAB-a	17
3.7. Radno okruženje Matlab-a	17
3.8. Matrice u Matlab-u.....	17
4. SPSS	19
4.1. Uvod u SPSS	19
4.2. Istorija SPSS-a.....	20
4.3. Šta je SPSS?	20
4.4. Analitički proces	20
4.5. Širina funkcionalnosti	21
4.6. Primena SPSS-a.....	22

5. IBM SPSS Statističke opcije	22
Statistika baza	22
Napredna statistika.....	22
Bootstrapping.....	22
Kategorije.....	22
Složene uzorke	22
Udružena analiza.....	22
Custom Tabele	23
Priprema podataka	23
Decision Trees	23
Direktni marketing.....	23
Tačni testovi.....	23
Forecasting.....	23
Missing Values (nedostajajuće vrednosti)	23
Neuronske mreže	23
Regresija	23
6. Osnovni elementi interfejsa i početak rada sa programom.....	24
6.1. Data View.....	24
6.2. Variable View	25
6.3. Outout View	26
6.4. Pivot Table Editor	26
6.5. Chart Editor	27
6.6. Sintax Editor i rad u njemu	27
7. Glavni meniji i palete alati u programu SPSS	28
7.1. File.....	28
7.2. Edit	28
7.3. View	28
7.4. Data	28
7.5. Transform	29
7.6. Analyze.....	29
7.7. Direct Marketing	30
7.8. Graphs	30
7.9. Utilities	30
7.10. Add-ons	30

7.11.	Window	30
7.12.	Help	31
7.13.	Paleta alatki i statusna traka u programu SPSS	31
8.	Primena SPSS-a	31
8.1.	Definisanje podataka	31
	Filtriranje podataka	32
8.2.	Deskriptivna statistika	33
8.3.	Tabeliranje podataka	35
8.4.	Testovi značajnosti razlika	37
8.5.	Procedura One-way ANOVA	37
9.	ZAKLJUČAK	43
10.	LITERATURA	44

UVOD

Danas nam primena računara i programskih paketa u velikoj meri olakšava statističku obradu i može smanjiti mogućnost grešaka na minimum. Ako koristimo računar sa paketima za statističku obradu podataka, unošenje tačnih podataka koje smo dobili posmatranjem predstavlja prvi zadatak koji treba da uradimo. Za statističku obradu podataka danas možemo naći veliki broj alata. Postoje oni koji su jednostavniji za korišćenje od drugih i koji su namenjeni široj populaciji, dok možemo naći i one pakete koji su namenjeni profesionalnom bavljenju statistikom.

Neki od alata za statističku obradu podataka su: **R**, **MATLAB**, **SPSS**, **STATS**, **STATISTICA**, itsl.

R i Matlab u svojim bibliotekama raspolazu velikim kolekcijama alata, koji se uvećavaju iz dana u dan. Zbog relativno jednostavne sintakse programskih jezika R i MATLAB, ako želite specifičan model koji nije dobijen u standardnom paketu, možete ga implementirati sami. Što se tiče savremenijih statističkih alata, biblioteka R-a nudi veći izbor istih, ali MATLAB ima daleko bolju grafiku od R programskog jezika.

Na drugoj strani spektra je paket poput SPSS-a. SPSS je prilično uzak u svojim mogućnostima i omogućava vam da uradite samo oko polovinu glavne statistike. Ipak, SPSS je veoma popularan među praktičarima jer ne zahteva gotovo nikakav trening, jer program poseduje širok spektar funkcija koje su lake za korišćenje, a opet veoma korisne u analizi. SPSS je veoma detaljan program, koji pruža odlične vizuelne rezultate. Program sadrži sve glavne testove i dijagnostičke alate u vezi sa postupkom i omogućava da se napiše informativni statistički deo empirijske analize. SPSS se često koristi za standardne zahteve klijenata, kao što su linearne regresije, ANOVA ili analiza glavnih komponenti.

U prikazu statističkih metoda i tehnika koje se koriste u SPSS-u, za potrebe pisanja ovog rada bili su potrebni podaci na kojim bi se praktično pokazalo kako se program upotrebljava.

Set podataka preuzet je sa sajta Centra za istraživanje po imenu Pew, sa sedištem u Vašingtonu. Podaci su prikupljeni u periodu od 10. juna do 12.jula 2015.godine, u nacionalnom uzorku koji sačinjava 2.001 odrasla osoba (18 godina ili stariji) koji žive u svih 50 američkih država i distriktu Kolumbija.

Podaci su prikupljeni kombinacijom nasumičnog biranja cifara pomoću fiksne i mobilne telefonije. Razgovori su vođeni na engleskom i španskom jeziku. Ispitanici iz grupe fiksne telefonije su izabrani slučajno, tražeći najmlađeg člana porodice koji je u tom trenutku bio kod kuće. Intervjui iz grupe mobilne telefonije, su obavljani sa osobom koja se javila, ako je ta osoba odrasla osoba u tom trenutku imala 18 ili više godina.

Napomena: Set podataka koji je korišćen za istraživanje i analizu je preuzet i obrađivan u originalnom formatu na engleskom jeziku.

Predmet istraživanja

Predmet istraživanja u ovom radu je upotreba statističkih programa za analizu podataka. Kroz upotrebu SPSS programskog paketa urađena je analiza nad podacima koji su prikupljeni da bi se ispitalo pitanje o automatizaciji radne snage u budućnosti.

Sa svakodnevnim razvojem tehnologije, mašine su sve više u stanju da obavljaju širok spektar poslova koji su već odavno u ljudskim domenima, od samostalne vožnje vozila i polu-autonomnih robota do inteligentnih algoritama i prediktivnih analitičkih alata. Prema podacima koji su prikupljeni kroz anketu nasumično biranih ljudi na području Amerike, urađena je analiza kako bi se došlo do rezultata o pitanju da li će mašine, odnosno napredak tehnologije i veštačke inteligencije, preuzeti sadašnje poslove koji obavljaju ljudi.

Kako većina Amerikanaca u ovom istraživanju predviđa da će u roku od 50 godina, roboti i kompjuteri raditi više poslova nego što trenutno obavljaju ljudi – nekoliko radnika očekuje da će njihovi poslovi ili zanimanja imati značajne uticaje.

U ovom radu, zaključeno je da krajnji stepen u kojem će roboti i algoritmi ugroziti ljudsku radnu snagu zavisi od mnoštva faktora. Mnogi Amerikanci očekuju da će ova promena postati stvarnost u narednih 50 godina.

Metodologija i ciljevi istraživačkog rada

Metode istraživanja su sprovedene kroz određene analize nad setom podataka koji je preuzet sa sajta takozvanog “Pew” istraživačkog centra. Korišćene su određene funkcije u analizi podataka kako bi se došlo do željenih rezultata kao i da bi se prikazala sama primena SPSS programa, postepeno kroz faze.

Cilj istraživanja je da se ispita da li će automatizacija radne snage preuzeti neke od današnjih poslova koji obavljaju ljudi, kao i da li postoji statistički značajna razlika u prosečnoj vrednosti dostignutog nivoa obrazovanja između ispitanika koji su trenutno zaposleni na pozicijama koje su predstavljene kao nivo Eksperta, i poziciji koja je predstavljena kao Korisnička podrška.

Ovaj rad mogu koristiti i manje stručni korisnici programa za statističku analizu podataka, ali i napredniji korisnici kao skraćeni prikaz obrade podataka korišćenjem SPSS programskog paketa.

Hipoteze istraživanja

- Opšta hipoteza

Cilj razvoja alata za statističku obradu podataka i njihovu implementaciju je da korisnicima omogući lakšu analizu i obradu istih. Korišćenjem nekih od ovih alata i njihovih funkcija, može se rešiti određeni statistički problem i doći do željenog rezultata.

- Posebna hipoteza

SPSS softverski paket raspolaže alatima koji su se pokazali kao adekvatno rešenje u obradi, analizi i donošenju preciznih rezultata nad korišćenim setom podataka. Odabirom ispravnih funkcija, istraživanje se svodi na svega nekoliko koraka, pri čemu je dobijeni rezultat prikazan i deskriptivno i grafički, što nije omogućeno u nekim drugim programima.

1. R-Programiranje



Slika 1.1 - R logo

1.1. Šta je R?

R je programski jezik namenjen za statističku obradu podataka i grafičko prikazivanje. Nastao je implementacijom otvorenog koda S programskog jezika, a delimično je inspirisan i Scheme programskim jezikom. Danas se R smatra svojevrsnim standardom za analizu podataka i koristi ga veliki broj profesionalaca koji se bave statistikom. Kao deo GNU projekta, izvorni kod R-a izdat je pod GNU GPL licencom i dostupan je korisnicima različitih operativnih sistema.¹

Džon Čejmbers, razvio je S programski jezik u Bell Labs-u (današnji AT&T) sredinom sedamdesetih godina 20. veka. Profesori Univerziteta u Oklandu (Novi Zeland), Robert Džentlmen i Ros Ihaka, 1993. godine, razvijaju novi S jezik koji ubrzo postaje jedan od popularnijih programskih jezika za statističku analizu u svetu. Uz pomoć Martina Mehlera (profesora sa instituta u Cirihi), Džentlmen i Ihaka razvijaju R sve do 1997. godine, nakon čega se formira veća grupa statističara koja dalje implementira i razvija program. Ova grupa nazvana je *R Development Core Team*.

Brojni dodaci, odnosno paketi, danas se besplatno mogu preuzeti sa “CRAN” mreže. Sa više od 2000 paketa, koji pružaju dodatnu funkcionalnost R-u, gotovo da ne postoji statistička tehnika koju R ne može da izvrši.

1.2. Istorija R-a

Ros Ihaka i Robert Džentlmen razvili su R kao besplatno softversko okruženje za svoje časove nastave kada su bili kolege na Univerzitetu Okland na Novom Zelandu. Pošto su obojica bili upoznati sa S-om, komercijalni programskim jezikom za statistiku, bilo je prirodno da upotrebe sličnu sintaksu u svome radu. Nakon što su Ihaka i Džentlmen objavili njihov softver na

¹ <http://savtajr.blogspot.rs/p/sta-je-r.html>, stranica posećena 14. marta 2015.

S-news mejling listi, nekoliko ljudi je postalo zainteresovano i započelo saradnju sa njima, značajno Martin Mehler.

Trenutno, grupa od 18 ljudi ima pravo da modifikuje centralnu arhivu izvornog koda. Ta grupa se naziva R Development Core Team. Pored toga, mnogi ljudi su doprineli novom kodu i ispravljanju “bagova” u projektu.

Evo nekih prekretnica u datumima u razvoju R-a:

- **Početak 90-ih:** Počeo je razvoj R-a.
- **Avgust 1993:** Softver je objavljen u S-news mejling listi. Od onda, set aktivne R mejling liste je kreiran. Veb stranica www.r-project.org/mail.html obezbeđuje opis ove liste i instrukcije za pretplatu.
- **Jun 1995:** Nakon nekih ubedljivih argumenata Martina Mehlera (između ostalih) da napravi kod dostupan kao “besplatan softver”, kod je postao dostupan kao Free Software Foundation’s GNU General Public Licence (GPL), Version 2.
- **Sredinom 1997:** Početni R Development Core Team je formiran.
- **Februar 2000:** Prva verzija R-a, verzija 1.0.0, je realizovana.

Ros Ikaha je napisao sveobuhvatan pregled razvoja R-a. Veb stranica <http://cran.r-project.org/doc/html/interface98-paper/paper.html> obezbeđuje fascinantnu istoriju.²

1.3. R okruženje i njegove osobine

R predstavlja integrisani paket softvera programskih objekata za upravljanje podacima, računanje i grafički prikaz. Između ostalog uključuje:

- mogućnost za efektivno upravljanje podacima i njihovo skladištenje,
- niz operatora za računanje sa poljima podataka, a posebno matricama
- veliku, koherentnu, integrisanu zbirku programskih alata za analizu podataka
- grafičke mogućnosti za analizu podataka, te za njihovo prikazivanje direktno na računaru ili na papiru, i
- dobro razvijen, jednostavan i efikasan programski jezik koji uključuje uslovne tvrdnje, petlje, rekurzivne funkcije definisane od strane korisnika te postupke za učitavanje i spremanje podataka. (Većina samih funkcija koje omogućuje sistem je pisana u S jeziku.)

Izraz "okruženje" upotrebljen je da se istakne da je R dobro planiran i konzistentan sistem, a ne sistem koji se postepeno dopunjavanja sa specifičnim i nefleksibilnim programskim alatima, što je često slučaj kod drugih programa za analizu podataka.

R, kao što je S, je dizajniran oko pravog računarskog jezika, a to omogućava korisnicima dodatne funkcionalnosti definisanjem novih funkcija. Veliki deo R sistema napisan je u S programskom dijalektu, što korisnicima znatno olakšava praćenje algoritamskih izvora.

² R for dummies, Andrie de Vries Joris Meys Andrie de Vries and Joris Meys, *R for dummies*, John Wiley and Sons, Ltd, England, 2012. Str. 22

R se može jednostavno proširiti preko paketa. Postoji oko osam takvih paketa koji pokrivaju vrlo širok raspon moderne statistike. R raspolaže širokim spektrom statističkih (linearnih i nelinearnih modeliranja, klasičnih statističkih testova, analizom vremenskih serija, grupisanjem, klasifikacijom, itd.) i grafičkih tehnika. S jezik je često alat izbora za istraživanja u statističkoj metodologiji, i R pruža Open Source put do učešća u toj delatnosti.

Jedna od prednosti R-a je lakoća sa kojom se mogu dizajnirati i proizvesti kvalitetni dijagrami, uključujući matematičke simbole i formule, gde su potrebne. Veliku pažnju preuzimaju osnovna podešavanja i rešenja manjih dizajna u grafici, ali korisnik zadržava potpunu kontrolu.

R je dostupan kao slobodan softver pod uslovima Free Software Foundation GNU General Public License u obliku izvornog koda. Može se prevoditi i izvršavati na najrazličitijim UNIXS platformama i sličnim sistemima (uključujući FreeBSD i Linuks), Windows i MacOS.

Veći komercijalni softverski sistemi podržavaju veze ili integraciju sa R-om, uključujući: JMP, Mathematica, MATLAB, Spotfire, SPSS, STATISTICA, Platforma Simphoni, SAS i Tableau.

1.4. R je jezik

Ovo je veoma važno. To znači da možete učiniti da računar radi umesto vas one zadatke koji se ponavljaju. Sa otvorenosti S jezika, to znači da je lako napraviti promene u funkciji koje ne rade baš ono što želite.

1.5. Široko dostupan

R radi na Windows, Macintosh, Linux i Unix platformama. Isti R program koji radi na vašem office Unix server-u, radiće na vašem kućnom računaru i Mac lap-top-u. Binarna datoteka podataka generalno može da se pročita na svim platformama.

1.6. Editori R koda

Nekoliko odličnih editora koda su na raspolaganju da obezbede funkcije kao što su podvlačenje sintakse R koda, auto izjednačavanje koda i komande koje šalju kod ili funkcije u R konzolu:

- Osnovni editor koda od Rguis
- RStudio: Integrated Disk Electronic, bazirana na grafičkom korisničkom interfejsu za R
- Vim-R-Tmux: radno okruženje R-a zasnovano na vim-u i tmux-u
- Emacs (ESS dodati paket)
- gedit i Rgedit
- RKward
- Eclipse
- Tinn-R
- Notepad++ (NppToR)³

³ <http://science.webhostinggeeks.com/programiranje-u-r>, stranica posećena 16. marta 2015.

1.7. Problemi

Dok je jezik jedna od najvećih snaga R-a, on može biti teži za učenje onima koji nemaju nikakva iskustva u programiranju. Ono što može da bude problematično za karakteristike R-a je to da svi objekti moraju da budu u memoriji. Očigledno je da to ograničava veličinu seta podataka sa kojima može da se radi. Ovo je cena koja mora da se plati za ne-pravougaone podatke. Mnogi korisnici R programa se nikada ne sretnu sa problemima vezano za memoriju, dok drugi moraju stalno da budu svesni veličine njihovih objekata. R ima funkcije za pristup bazama podataka. Uobičajeno rešenje kada se radi sa velikim skupovima podataka je smeštanje podataka u bazu podataka i donošenje delova R-a po potrebi. Ovo je uglavnom efikasno, ali ne daje kompleksnost. Pojava 64-bit-nih mašina znači da sada memorija može biti prilično velika. U mnogim oblastima skupovi podataka će verovatno rasti sporije nego pristupačne memorije mašina. U nekoliko oblasti skupovi podataka mogu rasti brže nego mašine. Ali ovi poslednji skupovi podataka ne mogu biti pravougaoni. Još jedna slabost je nedostatak sveobuhvatne dokumentacije koji pokriva ceo R.⁴

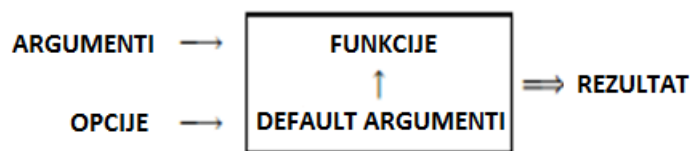
2. Kako R funkcioniše

Činjenica je da R jezik može odvratiti neke korisnike koji misle “ja ne mogu da programiram”. Ovo ne bi trebao da bude slučaj iz dva razloga. Prvo, R je jezik tumačenja, nije jedan od sastavljenih, što znači da su sve komande otkucane na tastaturi direktno obavljene bez potrebe da se izgradi kompletan program kao u većini računarskih jezika (C, Fortran, Pascal, ...).

Drugo, Sintaksa R-a je veoma jednostavna i intuitivna. Na primer, linearna regresija može da se uradi komandom `lm(y ~ x)` što znači “postavljanje linearnog modela sa y kao odgovor i x kao predskazivač”.

U R-u da bi naredba bila izvršena, funkcija uvek mora biti napisana sa zagradama čak i ako nema ništa unutar njih (e.g., `ls()`). Ako se ime funkcije otkuca bez zagrade, R će prikazati sadržaj funkcije. U ovom dokumentu, imena funkcija su generalno pisana sa zagradama u naredbi da se razlikuju od drugih objekata, osim ako tekst ne znači jasno tako. Kada je R u pokretu, promenljive, podaci, funkcije, rezultati i slično, se čuvaju u aktivnoj memoriji računara u formi predmeta koji imaju ime. Korisnik može da uradi radnje na ovim objektima sa operacijama (aritmetikom, logikom, poređenjem,...) I funkcijom (koji su sami objekti). R funkcija može da se skicira na sledeći način:

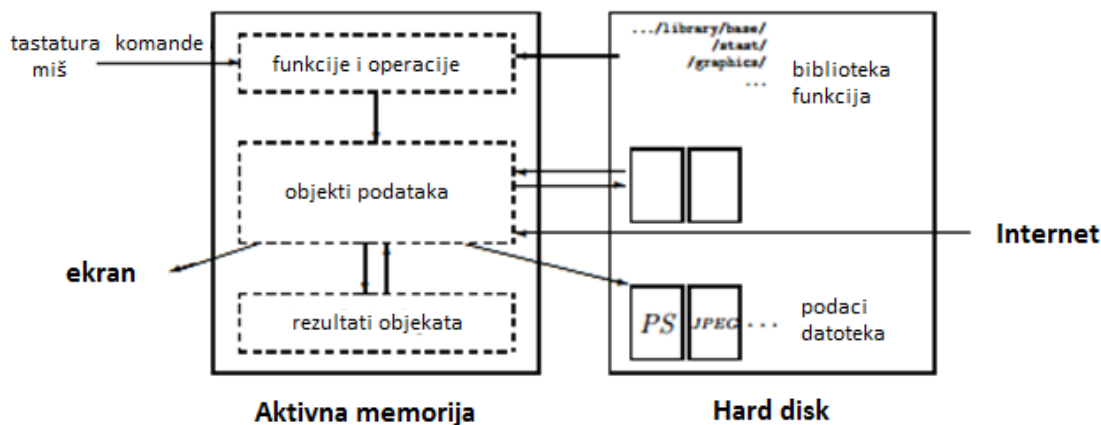
⁴R Relative to Statistical Packages: Comment 1 on Technical Report Number 1 (Version 1.0) Strategically using General Purpose Statistics Packages: A Look at Stata, SAS and SPSS Patrick Burns, Burns Statistics, patrick@burns-stat.com



Slika 1.2 - Skica R funkcije

Argumenti mogu biti objekti (“data”, formule, izrazi...), neki od njih bi mogli biti podrazumevano definisani u funkciji; te podrazumevane funkcije mogu biti modifikovane od strane korisnika određenim opcijama. R funkcija ne mora zahtevati nikakav argument; ili su svi argumenti podrazumevano definisani (i njihove vrednosti mogu biti promenjene u opcijama) ili ni jedan argument nije definisan u funkciji. Ovaj opis je dovoljan da se za trenutak razume kako R funkcioniše.

Sve akcije R-a koje su urađene na objektima se čuvaju u aktivnoj memoriji kompjutera; ne privremeni fajlovi se koriste (Slika 1.) Očitavanja i spisi fajlova se koriste za izlaz i ulaz podataka i rezultata (grafika,...). Korisnik izvršava funkcije preko nekih komandi. Prikazuju se rezultati direktno na ekranu, skladište u objektu, ili su zapisani na disku (naročito za grafiku). Kako su rezultati samo objekti, mogu se smatrati kao podaci i kao takvi se analizirati. Datoteke sa podacima se mogu čitati sa lokalnog diska ili sa udaljenog servera preko interneta.



Slika 1.3 - Šematski prikaz kako R funkcioniše

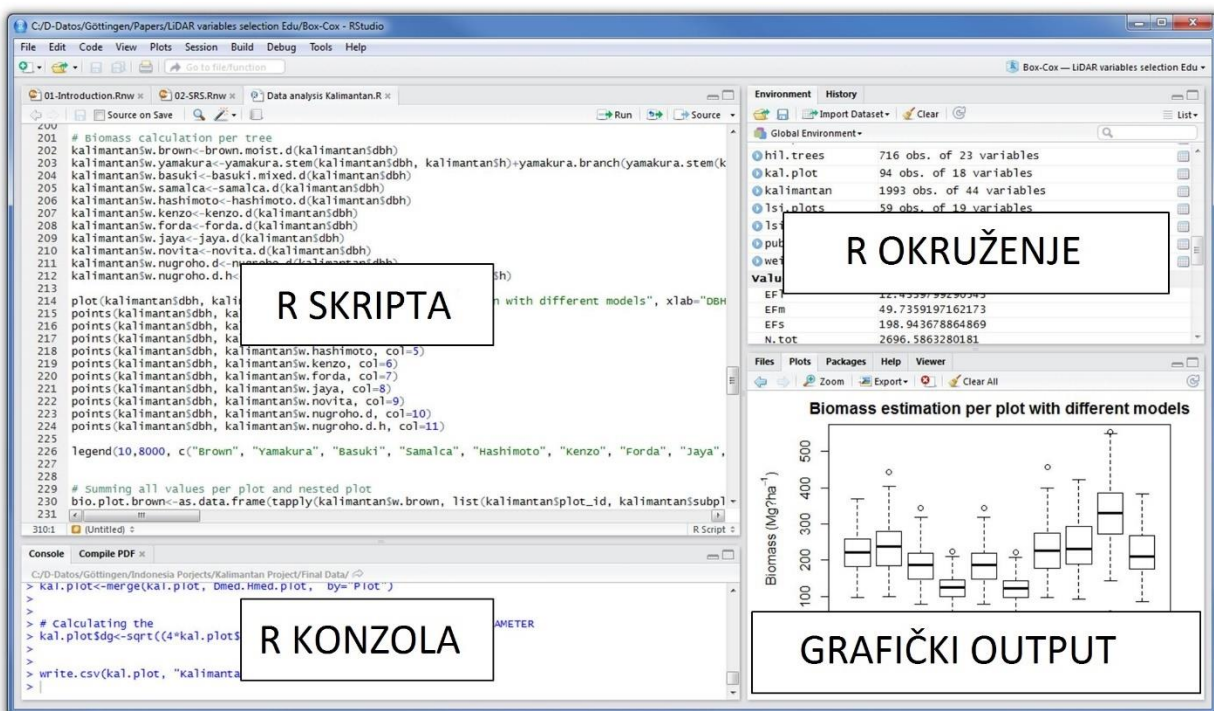
Funkcije dostupne korisniku se čuvaju u biblioteci koja je lokalizovana na disku u direktorijumu nazvanom “R HOME/library” (R HOME je direktorijum gde je instaliran R). Ovaj direktorijum sadrži paket funkcija koji su sami struktuirani u direktorijumima. Paket nazvan baza (“base”) je na neki način srž R-a i sadrži osnovne funkcije jezika, naročito, za čitanje i manipulisanje podacima. Svaki paket sadrži direktorijum pod nazivom R sa fajlom koji je nazvan kao i paket (na primer, za bazu paketa, ovo je fajl: R HOME/library/base/R/base). Ovaj fajl sadrži sve funkcije paketa.

Jedan od najjednostavnijih komandi je da ukucate ime nekog objekta za prikaz njegovog sadržaja. Na primer, ako objekat n sadrži vrednost 10:

```
> n
[1] 10
```

Cifra 1 u zagradama pokazuje da ekran počinje kao prvi element n. Ova komanda implicitno koristi funkciju “print” i primer iznad je sličan za “print” (n) (u nekim slučajevima, funkcija “print” se mora koristiti eksplicitno, kao što je unutrašnjost funkcije ili petlje).

Ime objekta mora početi sa slovom (A-Z i a-z) i može uključivati slova, cifre (0-9), tačke (.) i razmake (). R diskriminiše između velikih slova i onih malih u nazivima objekata tako da se x i X mogu imenovati kao dva različita predmeta (čak i u Windows-u).⁵



Slika 1.4 - Primer R prozora

⁵ Emmanuel Paradis, *R for Beginners*, France, 2005, Str. 38

3. MATLAB



Slika 2.1 - Matlab logo

3.1. Uvod u matlab

Kompanija MathWorks Inc, razvila je program za računanje tehničkim znanostima koji je nazvan Matlab. Program je nastao potrebom za računanjem zasnovanom na matričnoj aritmetici kao i potrebom za brzu obradu laboratorijskih mernih rezultata. Danas postoji čitav niz grupa programa za računanje u tehničkim naukama. Od najpoznatijih se mogu izdvojiti Matlab, Matrix i Easy. Razlikuju se u načinu izvršavanja i strukturi, dok im je korisnički interfejs sličan. U *Scripti* su grupisane naredbe za izvršavanje funkcija izgrađenih na osnovu matričnog kalkulatora koje i čine osnovno jezgro Matlaba. Danas Matlab čine čitav niz alata sa osnovnim jezgrom i korisničkim interfejsom, koji su omogućeni pisanjem, odnosno izgradnjom funkcija. U odnosu na druge slične programe kod kojih je nadogradnja moguća samo od proizvođača i koji imaju zatvoreniju strukturu, Matlab poseduje prednost baš zbog modularnosti i otvorenosti za definisanje novih alata. To omogućava svim korisnicima Matlaba da na neki način postanu razvojni tim ovog programa. Alati za većinu delatnosti iz područja tehničkih znanosti su razvijeni tokom razvoja Matlaba, a postoje i neki alati koji se primenjuju iznad područja tehničkih znanosti.⁶

Svojstva Matlaba danas u dalekoj meri prevazilaze originalnu “matričnu laboratoriju”. Matlab se može posmatrati kao programski jezik i intraktivni sistem za opšta i tehnička istraživanja. Pored osnovnog paketa, postoje i razni programski paketi koji se mogu primenjivati skoro u svim inženjerskim delatnostima: za identifikaciju sistema, automatsko upravljanje, 2D i 3D grafičke prikaze, slike, obradu signala, statističke obrade, za analizu u vremenskom i frekvencijskom domenu i brojne druge.

Uz pomoć dijagrama stanja i funkcije blokova dijagrama, simulacija kontinuiranih i diskretnih sistema moguća je u Matlab-u uz paket SIMULINK.

⁶ Željko Ban, *Osnove Matlab-a* doc.dr.sc.

3.2. Šta je Matlab?

Firma MatWorks je pe proizvela Matlab kao programski jezik koji služi za numeričke proračune. Sa Matlab-om je moguće manipulirati matricama, implementirati algoritme, prikazati funkcije i fitovati, stvarati grafički korisnički interfejs kao i povezati se sa programima.

Pun naziv za MATLAB je „MATrix LABoratory“ („laboratorija za matrice“). Kliv Moler, šef katedre za informatiku na Univerzitetu Novi Meksiko, izumeo je Matlab kasnih sedamdesetih godina 20-og veka.

Danas Matlab koristi preko milion zaposlenih na univerzitetima i u ostalim industrijama.

MATLAB je moguće nabaviti zajedno sa Simulinkom kao i brojnim dodacima.

Dodaci su svrstani po sledećim kategorijama:

- Obrada slike,
- Komunikacija i procesiranje signala,
- Distribuirano računanje
- Analiza i finansijsko modelovanje,
- Analiza podataka i statistika,
- Optimizacija i matematika,
- Analiza i kontrola sistema,
- Izveštavanje i povezivanje baza podataka,
- Povezivanje sa drugim aplikacijama.⁷

Matlab možemo definisati na više načina;

- kao program za računanje u tehnici koji je u osnovi matematički alat i koji se u velikoj meri primenjuje u primenjenim disciplinama kao što je inženjerstvo,
- kao program sa visokim performansama kada je u pitanju nemeričko izračunavanje i vizuelizacija podataka,
- kao interpreter koji može da pruži intraktivni rad i batch obradu.

Za razliku od programskih jezika kao što su Fortran, Pascal i C, u Matlabu se veliki broj problema može lakše rešiti jer je osnovni element matrica koja ne zahteva dimenzionisanje.

⁷ <https://sr.wikipedia.org/sr-el/MATLAB>, stranica posećena 18. marta 2015.

3.3. Primena Matlab-a

Matlab se danas primenjuje na Univerzitetima i služi kao alat za obuku. Industrije ga upotrebljavaju za razna istraživanja, za rešavanja praktičnih inženjerskih i raznih vrsta problema u matematici. Najčešće se upotrebljava u

- matematici i numerici,
- u razvoju algoritama,
- modeliranju,
- simulaciji,
- razvoju prototipova,
- analizi podataka,
- istraživanju i vizuelizaciji,
- naučnoj i inženjerskoj grafici,
- razvoju aplikacija i grafičkim korisničkim interfejsom.

3.4. Alati (Toolbox-ovi)

Kolekcija alata sadrži više funkcija koje su namenjene za rešavanje problema iz određenih oblasti koje proširuju mogućnosti Matlab programa. Postoje toolboxovi za:

- analizu signala,
- automatsko upravljanje,
- simulaciju dinamičkog Sistema ponašanja,
- identifikaciju sistema,
- veštačke neuronske mreže,
- fuzzy sisteme,
- optimizaciju,
- analizu robustnosti sistema.

3.5. Računarske platforme

- Podržane su brojne računarske platforme
 - PC,
 - Mackintosh, Map Power PC,
 - Sun SPARC,
 - SGI Indy,
 - IBM RS6000,
 - DEC Alpha.
- i operativni sistemi
 - MS Windows,
 - Unix,
 - Mac OS,
 - VMS.
- kod pisan u Matlab-u je prenosiv, uključujući i aplikacije sa korisničkim interfejsom u obliku prozora.⁸

⁸ <https://nadicakusic.files.wordpress.com/2012/03/matlab-osnove.pdf>, stranica posećena 18. marta 2017.

3.6. Sistem MATLAB-a

Matlab se sastoji od:

- **Razvojnog okruženja** – odnosno niza alata i mogućnosti koji olakšavaju upotrebu Matlab funkcija i datoteka.
- **Biblioteke matematičkih funkcija** – kompleksna i elementarna aritmetika, kompleksne funkcije, ...
- **Programskog jezika** - Jezik visokog nivoa, sa kontrolom toka programa, funkcijama, strukturama podataka, ulazom/izlazom, objektnoorijentisano programiranje. Omogućava brzo pisanje kratkih programa i kao složenih aplikacija.
- **Upotrebe grafike** - 2D i 3D grafika visokog nivoa, analiza slike, animacije, prezentaciona grafika. Omogućava i rad sa grafikom na niskom nivou .
- **API** (Application Program Interface) - omogućava pisanje C i Fortran programa koji koriste Matlab: pozivanje rutina iz Matlab-a, upotreba Matlab engine-a, čitanje i pisanje u Matlab formatima.

3.7. Radno okruženje Matlab-a

Sastoji se od:

- Komandnog prozora – koji je mesto gde se daju, odnosno gde interpreter ispisuje naredbe i gde su ispisani rezultat, kao što su:
 - sekvencijalno prihvatanje i izvršavanje naredbe
 - posedovanje linijskog editora
- Grafičkog prozora
 - obično su prikazani 2D i 3D dijagrami
 - dijagrami mogu biti prikazani u različitim prozorima, a takođe možemo podeliti jedan prozor u više delova,
 - mogu biti prikazani kada se automatski izvršavaju grafičke naredbe
 - na više načina se može odštampati sadržaj prozora
- Korisničkog editora za pisanje M-skript datoteka

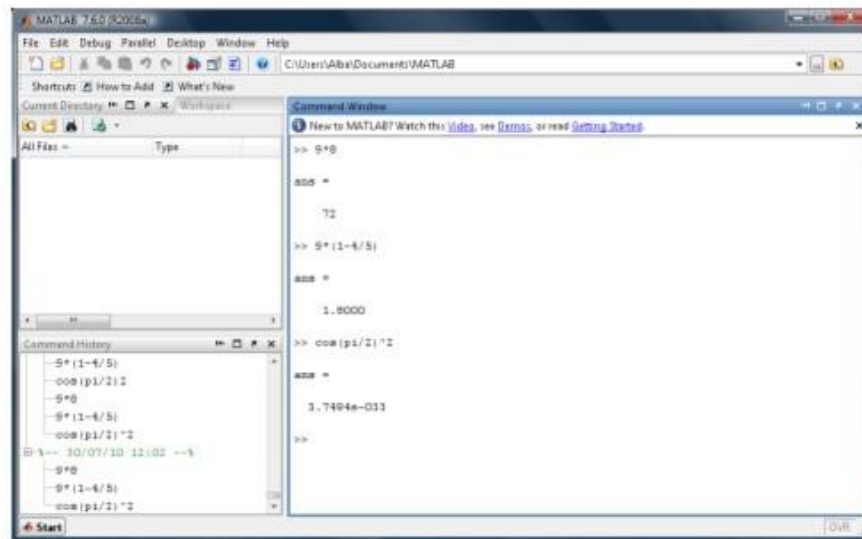
Ostali alati:

- Unapređen uvoz i izvoz podataka
- Profiler – pomaže u otkrivanju uskih grla u aplikaciji
- Podrška Source Control System-ima
- Notebook – pristup Matlab-u iz programa za obradu teksta (MS Word)

3.8. Matrice u Matlab-u

- osnovni tipovi podataka su matrice (brojevi)
- skalari su matrice dimenzija 1x1 i slične su sa vektorima u oblicima kolona ili vrsta
- elementi matrice mogu biti kompleksni brojevi

- matrice se mogu tumačiti kao
 - niz koeficijenata polinoma
 - niz slova u string-u
 - bit-mapirana slika
 - zvučni zapis
- podržan je rad sa retkim matricama
- dimenzije matrica se ne moraju unositi
- potreban memorijski prostor za matrice se automatski zauzima (alocira)
- veličina matrice nije ograničena i može prevazići raspoloživi RAM računara (zahvaljujući postojanju Windows swap datoteke)



Slika 2.2 - Matlab okruženje

4. SPSS



Slika 3.1 - Logo SPSS programa

4.1. Uvod u SPSS

IBM SPSS je sveobuhvatan sistem za analizu podataka. SPSS može uzeti podatke iz gotovo bilo koje vrste datoteke i koristiti ih za generisanje tabeliranih izveštaja, grafikone i grafičke podatke distribucije i trendova, deskriptivne statistike, kao i složene statističke analize.⁹

SPSS obezbeđuje informacije korišćenjem statističkih procedura koje će služiti u kasnijem donošenju odluka.

U tabelarnim i grafičkim autputima su prezentovani rezultati analiza, koji doprinose bržem i efikasnijem odlučivanju kroz otkrivanje elementarnih činjenica i obrazaca.

SPSS je moguće koristiti u:

- istraživanju tržišta i direktnom marketingu,
- visokom školskom obrazovanju,
- administrativnom istraživanju, ljudskim resursima i planiranju resursa,
- oblasti medicinskog, kliničkog i socijalnog istraživanja,
- planiranju i predviđanju,
- unapređenju kvaliteta,
- izveštavanju i ad hoc odlučivanju,
- kao i u mnogim drugim oblastima.

⁹ <https://www.csun.edu/sites/default/files/statistics20-core-system-guide-64bit.pdf>, stranica posećena 19.marta 2015.

4.2. Istorija SPSS-a

Originalna verzija SPSS-a je napisana je kasnih 70-ih, od strane dva postdiplomska studenta političkih nauka koji su očajnički hteli da koriste programski jezik Fortran za preradu, manipulisanje i analiziranje podataka iz istraživačkih anketa. Dovedi su programera da im pomogne da proizvedu nešto što će pomoći korisnicima u pisanju komandi koje bi izgledale manje ili više prihvatljivo za društvene nauke u Engleskoj, i što bi po prvi put dalo naučnicima kontrolu nad svojim istraživanjima. Nepotrebno je reći, SPSS se proširio kad divlja vatra; sociolozi i naučnici političkih nauka su ga obožavali; statističari i programeri su ga mrzeli.

SPSS je prvi put objavljen u SAD-u 1968. godine po preporuci Toni Kokstona, prvi put je instaliran u Velikoj Britaniji 1970. Godine na Univerzitetu u Edinburgu (jedini sa IBM računarom) od strane Dejvid Maksvortija i Mardžori Berita. Bio je integrisani skup programa za upravljanje i statističku analizu podataka društvene nauke, razvijana posebno za analizu i obradu podataka istraživačkih anketa. Zbog svog jasnog Engleskog-kao komandnog jezika, impresivnog uputstva, brzo se proširila širom društvene (ankete) istraživačke zajednice. Bio je toliko uspešan da, radi zaštite dobrotvornog statusa na Univerzitetu u Čikagu, SPSS Inc je formiran kao poseban entitet. Brzo postaje svetski standard za društvenu nauku u računarstvu. Kasnije se SPSS okrenuo biznisu više nego društvenim istraživačkim aplikacijama i razvio je grafički korisnički interfejs (GUI) na osnovu padajućeg menija, a ne sintakse.

4.3. Šta je SPSS?

SPSS Statistics je softverski paket korišćen za statističku analizu. Proizveden je od strane SPSS Inc., te je spojen sa IBM-om 2009. godine. Sadašnja verzija (2015) je zvanično nazvana *IBM SPSS Statistics*. Partnerski proizvodi u istoj porodici se koriste za istraživanje autorizacije i upoređivanje (IBM SPSS Data Collection), rudarenje podataka (data mining) - (IBM SPSS Modeler), tekst analitiku, saradnju i raspoređivanje (serija i atomatsko osvojenih usluga).

Ime softvera originalno važi kao *Statistički paket za društvenu nauku (Statistical Package for the Social Sciences)* koji odražava originalno tržište, i ako je softver sada popularan i u drugim oblastima, uključujući sada i medicinske nauke i marketing.

4.4. Analitički proces

Analitički proces podataka, sam po sebi, nije tako jednostavan.

Pre svega, potrebno je da se sakupe podaci, da se sprovede analiza i nakon toga objaviti rezultate u upotrebljivom i razumljivom formatu. Ako bi se koristio softver sa ograničenim mogućnostima, okončanje posla bi zahtevalo primenu rezultata iz različitih izvora. SPSS sadrži sve što je potrebno za analitički proces, kao što je planiranje, obrada podataka, analiziranje, prikaz izlaznih podataka, i sve to na jednom mestu. SPSS se sastoji od nekoliko faza za izgradnju sistema procesa podataka, a uz pomoć svojih alata, olakšan je rad u svakom od njih. Sastoji se od:

- **Planiranja** (engl. Planning) štedi se vreme i novac korišćenjem SamplePower, SPSS Complex Samples i SPSS Conjoint-a;
- **Prikupljanja podataka** (engl. Data collection) - efikasnije se prikupljanju i unose podaci uz pomoć Dimensions and SPSS Data Entry;

- **Pristupa podacima** (engl. Data access) -brža analiza uz jednostavniji pristup podacima: SPSS Base;
- **Pripreme i upravljanja podacima** (engl. Data management and data preparation) - kroz efikasnije upravljanje i pripremu podataka, može se brže stići i do procesa analize: SPSS Data Preparation, SPSS Missing Value Analysis, SPSS Base, SPSS Complex Samples, SPSS Text Analysis for Surveys and Dimensions;
- **Analize podataka** (engl. Data analysis) – analiziraju se podaci kroz korišćenje različitih statističkih procedura sa ciljem da se dobiju precizniji modeli: SPSS Tables, SPSS Advanced Models, SPSS Base, SPSS Trends and Amos SPSS Regression Models, SPSS Complex Samples, SPSS Data Preparation, SPSS Classification Trees, SPSS Categories, SPSS Exact Tests;
- **Izveštavanje** (engl. Reporting) – prezentuju se rezultati korisnicima: , SPSS Maps and Dimensions, SPSS Base, SPSS Tables.
- **Razvoja** (engl. Deployment) razmenjuju se rezultati preko Web ili drugih prezentacija i publikacija: SmartScore i SmartViewer Web Server.

4.5. Širina funkcionalnosti

SPSS ima mnoštvo statističkih i matematičkih funkcija, mnoštvo statističkih procedura i veoma fleksibilnu sposobnost upravljanja podacima. Može pročitati podatke u skoro svakom formatu (npr. numerički, alfanumerički, binarni, dolar, datum-vreme formati), a verzija 6 pa nadalje može pročitati fajlove kreirane pomoću spread sheet/data base softvera. Ona takodje ima odličnu manipulaciju komunalnim podacima.

U nastavku je kratak pregled nekih funkcionalnosti SPSS-a:

- Transformacija podataka
- Deskriptivna statistika
- Nepredviđene tablice
- Testovi pouzdanosti
- Korelacije
- T-tests
- ANOVA
- MANOVA
- General Linear Model (Release 7.0 and higher)
- Regresija
- Nelinerana regresija
- Logistička regresija
- Loglinearna regresija
- Diskriminitivna analiza
- Analiza faktora
- Klaster analiza
- Multidimenzionalno skaliranje
- *Forecasting/Time Series*
- Analiza opstanka
- Neparametarska naliza
- Grafika i grafički interfejs

4.6. Primena SPSS-a

SPSS je široko upotrebljiv program za statističku analizu u društvenim naukama. Takođe se koristi za istraživanje tržišta, medicinska istraživanja, ankete kompanija, vlade, edukativna istraživanja, marketing organizacije, rudarenje podataka, i ostalo. Originalni priručnik SPSS-a (Nie, Bent & Hull, 1970) je opisan kao jedan od “najuticajnijih knjiga sociologije” za dozvoljavanje prosečnim naučnicima da rade svoje sopstvene analize. Pored statističke analize, upravljanje podacima (Izbor slučaja, preoblikovanje fajlova, kreiranje izvedenih odataka) i dokumentacija podataka, su odlike osnovnog softvera.¹⁰

5. IBM SPSS Statističke opcije

Sledeće opcije su dostupne kao dodatak na poboljšanjima celog IBM SPSS Statističkog Intel Core sistema:

Statistika baza daje širok spektar statističkih procedura za osnovne analize i izveštaje, uključujući datoteke, krostabove (crosstabs) i deskriptivne statistike, OLAP kocke i šifarnik izveštaja. Takođe nudi širok spektar tehnika za smanjenje dimenzija klasifikacije i segmentacije kao što su analize faktora, klaster analize, najbliže komšijske analize i diskriminativne analize funkcije. Osim toga, SPSS Statistics baza nudi širok spektar algoritama za poređenje sredstava i prediktivnih tehnika kao što su t-test, analiza varijanse, linearne regresije i redna regresija.

Napredna statistika se fokusira na tehnike koje se često koriste u sofisticiranim eksperimentalnim i biomedicinskim istraživanjima. To uključuje procedure za opšte linearne modele, linearne mešovite modele, varijanse komponenti analiza, loglinearne analize, redne regresije, aktuarske životne stolovi, Kaplan-Meier analize opstanka i osnovne i proširene Cox regresije.

Bootstrapping je metoda za dobijanje snažne procene standardne greške i poverene intervale za procene kao što su sredina, medijana, proporcije, odnos šansi, koeficijenta korelacije ili koeficijent regresije.

Kategorije obavljaju optimalne procedure skaliranja, uključujući analizu korespodencije.

Složene uzorke primenjuju razna istraživanja, tržišta, zdravlje, i istraživači javnog mnjenja, kao i naučnici društvenih nauka koji koriste metodologiju istraživanja da uključe svoje složene dizajne uzoraka u analizi podataka.

Udružena analiza daje realni način da se izmeri koliko osobine pojedinog proizvoda utiču na potrošača i prioritet građana. Sa udruženom analizom može lako da se izmeri trade-off

¹⁰ <http://en.wikipedia.org/wiki/SPSS>, stranica posećena 22. marta 2015.

efekat svake osobine proizvoda u kontekstu niza osobina proizvoda, kao što potrošači rade prilikom svake odluke u kupovini.

Custom Tabele stvaraju niz prezentacija – tabele kvaliteta izveštaja, uključujući kompleks *Stub-i-banner* tabele i prikaze više odzivnih podataka.

Priprema podataka omogućava brz vizuelni snimak vaših podataka. Ona pruža mogućnost da se primene pravila validacije koje identifikuju nevažeće vrednosti podataka. Možete da kreirate pravila koja označavaju vrednosti van dometa (out-of-range values), nedostajuće vrednosti, ili prazne vrednosti. Takođe možete sačuvati promenljive koje beleže individualna kršenja pravila i ukupan broj pravila kršenja po predmetu. Ograničen skup definisanih pravila koja možete da kopirate ili modifikujete je obezbeđen.

Decision Trees kreira tree-based model za klasifikaciju. Ona klasifikuje opservacije u grupama ili predviđa vrednosti podređenog (ciljne) promenljive zasnovane na vrednostima nezavisnih (prediktorskih) varijabli. Postupak obezbeđuje validacione alate za istraživačku i potvrdnu analizu klasifikacije.

Direktni marketing omogućava organizacijama da osiguraju njihove marketinške programe da budu efikasni koliko je moguće, kroz tehnike posebno dizajnirane za direktni marketing.

Tačni testovi izračunavaju tačne p vrednosti za statističke testove kada mali ili veoma nejednako distribuirani uzorci mogu učiniti uobičajene testove netačne. Ova opcija je dostupna samo na Windows operativnim sistemima.

Forecasting vrši sveobuhvatna predviđanja i analizu vremenskih serija sa više modela krive (curve-fitting models), „smoothing” modele i metode za procenu autoregresivnih funkcija.

Missing Values (nedostajajuće vrednosti) opisuje obrasce nedostajućih podataka, procenjuje sredstva i druge statistike, a pripisuje vrednosti za nestala zapažanja.

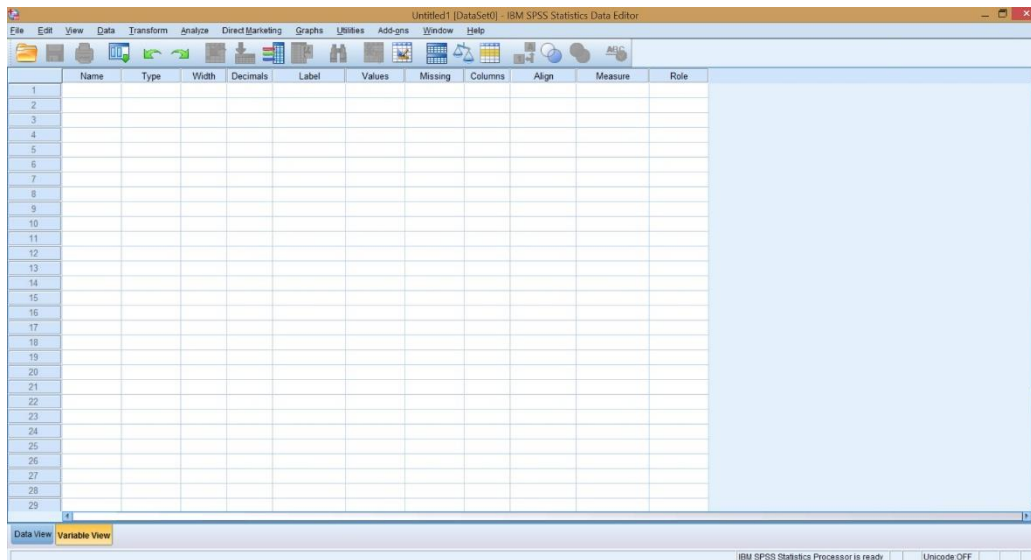
Neuronske mreže se mogu koristiti za pravljenje poslovnih odluka tako što će predviđati tražnju za proizvod kao funkciju cene i ostalih varijabli, ili kategorizaciju kupaca na osnovu kupovne navike i demografske karakteristike. Neuronske mreže su nelinearni alati za modeliranje podataka. Oni mogu da se koriste za modeliranje kompleksnih odnosa između ulaza i izlaza, ili da pronađu obrasce u podacima.

Regresija raspolaže tehnikama za analizu podataka koji se ne uklapaju u tradicionalne linearne statističke modele. To uključuje procedure za *probit* analize, logističke regresije, procenu težine, dvostepenu regresiju najmanjih kvadrata i opštu nelinearnu regresiju.

Amos™ (analiza trenutnih struktura) koristi modeliranje strukturnih jednačina da potvrdi i objasni konceptualne modele koji uključuju stavove, shvatanja i druge faktore koji utiču na ponašanje.

6. Osnovni elementi interfejsa i početak rada sa programom

Kada se otvori program SPSS, prvo što možemo videti je početni ekran koji možemo videti na slici 4.1. Ovo je takozvani “Data View”, odnosno “Data Editor”, koji služi za manipulisanje i unošenje podataka. U donjem levom uglu mogu se videti opcije “Data View” i “Variable View”. Ovo je drugi važni deo interfejsa, koji služi za definisanje varijabli. Rad u SPSS-u možemo podeliti u četiri faze: definisanje promenljivih, unošenje podataka, analiza i interpretacija podataka.



Slika 4.1 - Izgled Variable view prozora u SPSS-u

Kao i neki drugi programi u Windows-u, SPSS ima liniju naslova, set padajućih menija, toolbar (alati), status bar, kao i centralni radni deo. SPSS sadrži jedan toolbar, ali je on izmenjiv i moguće ga je korigovati po potrebi korisnika. Radni deo je tabela slična onoj u Excel-u, ali koja za redove ima ispitanike, a za kolone promenljive. U ovu tabelu se unose svi podaci koji su se dobili istraživanjem. Prva tri i poslednja dva padajuća menija su manje-više ista kao i na drugim Windows programima.

6.1. Data View

U prikazu podataka (Data View), podaci su postavljeni u standardnom pravougaonom formatu statističkog softvera. Svaki red predstavlja jedinicu posmatranja, ponekad se takođe naziva kao zapis ili u SPSS-u kao opservacija (“case”). Opservacijski broj u koloni sa leve strane se automatski dodeljuje i ne čuva kao podatak. Svaka kolona predstavlja promenljivu. Svi podaci u koloni moraju biti istog tipa, ili numerički ili string (takođe poznat kao karakter).¹¹

¹¹Julie Pallant, *SPSS Survival manual, Third edition, Mikro knjiga, Beograd, 2009. Str. 33*

Svaka ćelija podataka ima vrednost podataka. Ako podaci nedostaju, oni se prikazuju kao period (".") ili kao praznina (" "). Vrednosti podataka se mogu prikazati ili kao stvarne vrednosti ili kao formatirane vrednosti. Na primer, vrednost podataka o prihodima neke osobe može biti 20000, dok njegova formatirana vrednost može biti prikazana kao 20.000 \$. Formatirani podaci mogu imati oblik oznaka vrednosti, na primer, podaci koji se registruju kao prvi (1.) i drugi (2.) mogu biti označeni kao "muški" i "ženski". Dok formatiranje olakšava da se tumače rezultati, važno je zapamtiti da su vrednosti podataka ono što SPSS zapravo procesuirao. Konkretno, kada uspostavite komandu koja zahteva da odredite jednu ili više vrednosti podataka, koristite vrednosti i ne formatirane vrednosti.

Možete prebaciti prikaz podataka (Data View) između formatiranih i neformatiranih podataka tako što ćete kliknuti na dugme Value Labels.

6.2. Variable View

U Variable View možete videti i izmeniti informacije koje definišu svaku promenljivu (ponekad se nazivaju meta-data) u vašem setu podataka: svaka kolona Data View-a je opisana nizom Variable View-a.

Prvi atribut svake varijable je njegovo ime. Ime promenljive označava kako je kolona podataka identifikovana u programskom jeziku, i kako bi programski jezik radio dostojanstveno, imena promenljivih moraju da se pridržavaju određenih ograničenja: imena moraju početi slovom, a može biti sastavljena od znakova, brojeva, bez-interpunkcionih znakova i perioda. Kapitalizacija se ignoriše. Nazivi promenljivih mogu biti dužine do 64 karaktera. Nazivi promenljivih mogu da se dodaju ili menjaju jednostavno tako što ćete ih ukucati.

Osnovni tipovi promenljivih su ili numerički ili string. Međutim, da bi stvari izgledale zbunjujuće, SPSS vam omogućava da izaberete između nekoliko različitih standardnih formata za prikazivanje numeričkih podataka (npr. naučna notacija, zarez (comma) formatiranje, valute) i da nazovete ovo tip. Podesite tip promenljive tako što ćete kliknuti na kolonu, a zatim kliknite na sivo dugme koje se pojavljuje i radi u dijalog okviru.

Oznaka Atribut će vam omogućiti da date svakoj promenljivoj duži opis koji se prikazuje u mestu imenu promenljive, analogne vrednosti oznake za vrednosti podataka. I oznaka promenljive i oznaka vrednosti su korisni za davanje razumljivijeg izlaza.

Vrednosti atributa vam omogućavaju da kreirate listu oznaka vrednosti. Često nekoliko varijabli će deliti zajednički skup oznaka vrednosti, i u tom prozoru možete da kopirate i pasteujete vrednost setova oznaka. Oznake promenljivih se postavljaju jednostavnim ukucavanjem. Oznake vrednosti rade preko okvira za dijalog.

Nedostajući atribut je mesto za vas da odredite određene vrednosti podataka koje želite SPSS da ignoriše kada se izračunava statistika. Na primer, u podacima iz ankete je uobičajena praksa da se snimi vrednost podataka kao broj 8 kada ispitanik kaže "ne znam" kao odgovor na pitanje, a možete učiniti da SPSS tretira 8-ce u promenljivoj kao da su nedostajući podaci.

6.3. Outout View

The screenshot displays the IBM SPSS Statistics Processor interface. The top menu bar includes File, Edit, View, Data, Transform, Insert, Format, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The left sidebar shows a project tree with folders for Output, Case Processing, and Frequencies. The main window is divided into two panes. The top pane, titled 'Crosstabs', shows a 'Case Processing Summary' table and a 'Crosstabulation' table. The bottom pane, titled 'Bar Chart', shows a bar chart and a legend for 'Final State code after zip/ips merges'.

Case Processing Summary

	Cases				Total	
	Valid		Missing		N	Percent
	N	Percent	N	Percent		
emphw = 1 OR emphw = 2 OR emphw = 5 (FILTER) * Final State code after zip/ips merges	1100	100.0%	0	0.0%	1100	100.0%

emphw = 1 OR emphw = 2 OR emphw = 5 (FILTER) * Final State code after zip/ips merges Crosstabulation

		Final State code after zip/ips merges															
		Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delaware	District of Columbia	Florida	Georgia	Idaho	Illinois			
emphw = 1 OR emphw = 2 OR emphw = 5 (FILTER)	Selected	Count	20	6	17	8	115	20	12	3	4	61	36	3	7		
	% of Total		1.8%	0.5%	1.5%	0.7%	10.5%	1.8%	1.1%	0.3%	0.4%	5.5%	3.3%	0.3%	0.6%		
Total	Count	20	6	17	8	115	20	12	3	4	61	36	3	7			
	% of Total		1.8%	0.5%	1.5%	0.7%	10.5%	1.8%	1.1%	0.3%	0.4%	5.5%	3.3%	0.3%	0.6%		

Bar Chart

The bar chart displays the frequency of the 'Final State code after zip/ips merges' variable. The Y-axis represents the frequency, ranging from 0 to 120. The X-axis represents the state codes. The legend indicates the following states: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming.

Output Viewer pokazuje tabelu statističkih učinaka i sve grafikone koje ste kreirali. Podrazumeva se takođe da vam se prikaže programski jezik za komande koje izdajete (naziva sintaksa u SPSS žargonu), i većina poruka o greškama će se pojaviti ovde. Output Viewer vam takođe omogućava da izmenite i štampate rezultate. Tabele Output Viewer-a se čuvaju (kliknite na File - Save ili Save as) u obliku fajla “.spv”, koji može da se otvori samo u SPSS programu.¹²

6.4. Pivot Table Editor

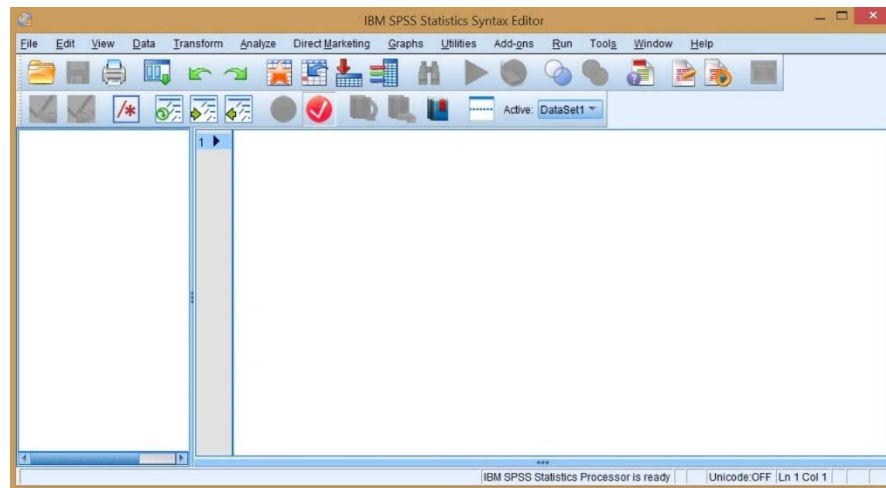
¹² http://www.ssc.wisc.edu/sscc/pubs/spss/classintro/spss_students1.html#data-view, strana posećena 10. 03. 2017.

6.5. Chart Editor

Možete menjati grafikone visoke rezolucije i grafičke podatke u grafičkim prozorima. Možete da promenite boju, izaberite različite tipove fontova ili veličina, zamenite horizontalne i vertikalne ose, rotirati tačkaste 3-D grafikone, pa čak i promenite tip grafikona.

6.6. Syntax Editor i rad u njemu

Sintaksa Editor nam omogućava da pišemo, menjamo, i pokrećemo komande u SPSS programskom jeziku. Ako takođe koristimo menije i dijalog prozor, taster *paste* automatski piše sintaksu za komandu koju ste naveli u aktivni Syntax Editor. Ovi fajlovi se čuvaju kao običan tekst i skoro svaki tekstualni editor može da ih otvori, ali sa ekstenzijom fajla *.sps*.



Slika 4.3 - Syntax prozor

Kao i kod ostalih vrsta prozora, možete imati otvoren više od jedan Syntax Editor. Aktivan prozor je označen sa malim narandžastim znakom plus. Kada *paste*-ujete sintaksu iz dijalog prozora, ona ide u aktivan Syntax Editor. Ako zatvorite sve svoje Syntax prozore i *paste*-ujete komandu, novi Syntax prozor će biti otvoren.

Osnovna jedinica rada u SPSS jeziku je komanda. U ovom jeziku, komande počinju sa ključnim rečima i na završavaju sa tačkom. Komande bi trebalo da počnu u levoj koloni u editoru. Ako su umotani u više od jedne linije, stalna linija bi trebalo da počne sa praznim prostorom. Kapitalizacija nije bitna. Syntax Editor prikazuje sintaksu koju SPSS ne može da tumači u crvenom obliku.

7. Glavni meniji i palete alati u programu SPSS

Meniji su ključni u radu sa bilo kojom Windows ili Mac aplikacijom, a tako je i u slučaju sa programom SPSS. Njegovih 12 glavnih menija, uključujući standardne Windows i Help menije za Windows verziju i 11 glavnih menija za Mac verziju, uključujući standardne Windows i Help menije, nude pristup do svih alati i mogućnosti koje SPSS ima u ponudi. U ovom radu pregledaćemo sadržaj svih tih menija i predstavimo paletu alati, skup ikona koje menjaju neke od komandi iz menija. Ikone ubrzavaju ili olakšavaju da se mnogo toga uradi, od snimanja datoteke do štampanja dijagrama.¹³

7.1. File

File meni je namenjen radu sa datotekama. Iz ovog menija možemo koristiti opcije kao što su, pravljenje novih datoteka, otvaranje postojećih ili snimanje datoteka u različitim formatima, prikazivanje informacija o datoteci, štampanje datoteka i izlazak iz SPSS programa. Meni File sadrži spisak datoteka koje su korišćene poslednje sa podacima kao i preostalih u skorije vreme korišćenih fajlova, što omogućava brzo vraćanje na prethodnu datoteku.

7.2. Edit

Meni Edit se koristi kada hoćemo da isečemo ili kopiramo neke od podataka i umetnemo ih na drugo mesto u trenutnoj ili nekoj drugoj datoteci sa podacima. Takođe se koristi da bismo pronašli neke podatke ili tekstove, zamenili tekstove ili prolagodili podešavanja SPSS programa.

7.3. View

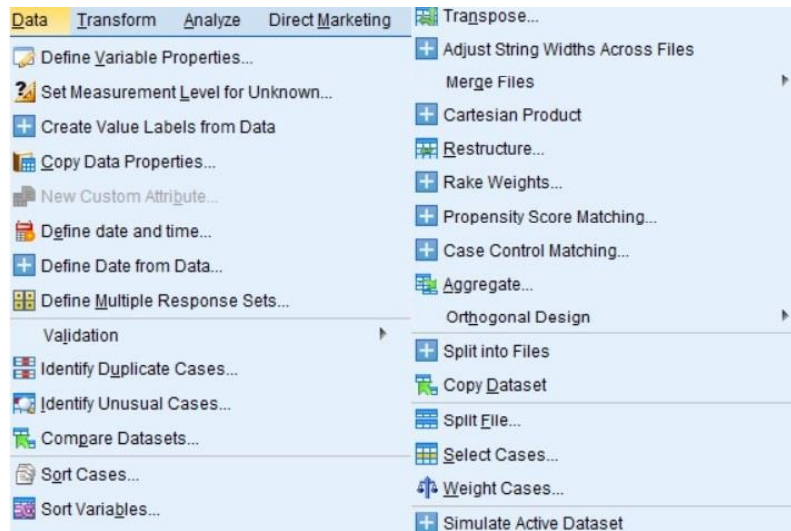
Meni View omogućava da izaberete koju paletu alati želite, da izaberete veličinu font-a, da dodate ili uklonite linije pomoćne mreže koje razdvajaju svaki deo podataka, i da izaberete da li da budu ili ne budu prikazani vaši neobrađeni podaci (raw data) ili oznake podataka.¹⁴

7.4. Data

Glavni elementi u bilo kojoj SPSS analizi su varijable i njihove vrednosti. Potrebni su moćni alati za rad sa tim varijablama i njih imate u SPSS programu. Data meni sadrži komande koje dozvoljavaju definisanje svojstva promenljivih, spajanje i grupisanje datoteka, sortiranje slučajeva kao i da njima dodeljujete značaj slučajevima onako kako vama to odgovara.

¹³ <http://www.ftn.kg.ac.rs/download/SIR/SIR%20Cuslovic%20Milos.pdf>, stranica posećena 12. marta. 2017.

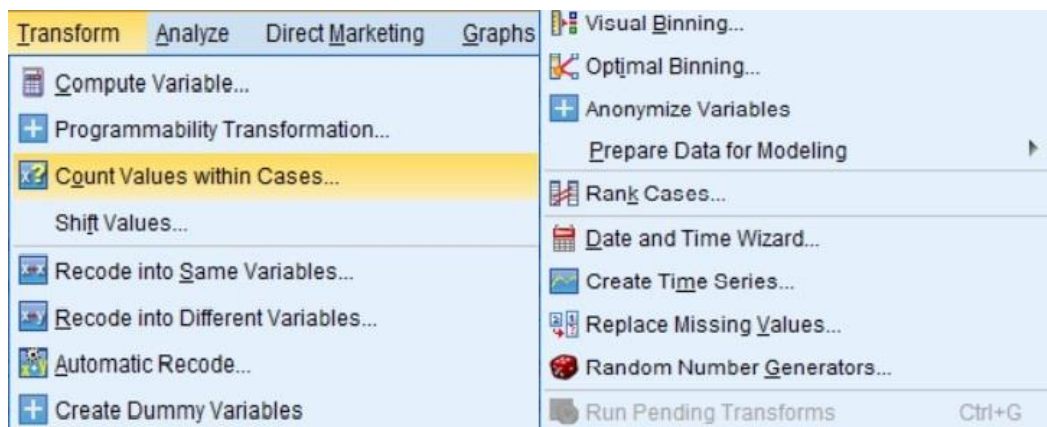
¹⁴ <http://www.uvm.edu/~dhowell/fundamentals7/SPSSManual/SPSSLongerManual/SPSSChapter1.pdf>, stranica posećena 12. marta 2017.



Slika 5.1 - Data meni

7.5. Transform

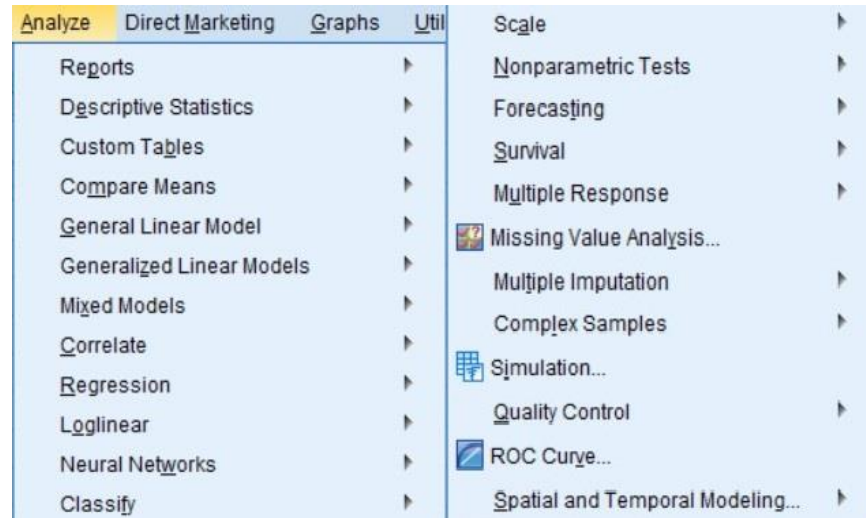
Meni Transform uključuje nekoliko opcija za menjanje promenljivih. Na ovom meniju ćete moći da pronađete komande koje omogućavaju da izračunate nove vrednosti, tumačite vrednosti, napravite skup slučajnih vrednosti, zamenite vrednosti koje nedostaju i još mnogo toga.



Slika 5.2 - Transform meni

7.6. Analyze

Postoji 24 različitih opcija na Analyze meniju (i brojni podmeniji) koji vode do raznih statističkih tehnika za analizu koje biste poželili da primenite. Analyze se koristi za izračunavanje deskriptivne statistike, za jednostavno izračunavanje srednje vrednosti i standardne devijacije, za analize vremenskih nizova i višestruke regresije, kao i mnogih drugih složenih analiza.



Slika 5.3 - Analyze meni

7.7. Direct Marketing

Meni Direct Marketing je usko specijalizovani meni koji korisnicima SPSS-a omogućava da naprave opcije menija za neke veoma posebne slučajeve.

7.8. Graphs

Graphs meni uključuje komande za kreiranje različitih grafika uključujući “box plotes”, histogram, linijski grafik, kao i “bar chart”.

7.9. Utilities

Ovde možemo pronaći informacije o datotekama i promenljivima, a može i da se koristi i definiše skup promenljivih. Utilities vam omogućava da listate informacije datoteka koje su lista svih promenljivih, oznaka, vrednosti, lokacija u datotekama podataka, i tipova.

7.10. Add-ons

Add-ons su programi koji mogu biti dodani u osnovni SPSS paket.

7.11. Window

Window meni se koristi da prelazite sa jednog prozora na drugi i da umanjite prozore Viewer ili Data Editor.

7.12. Help

Help meni ima mnogo korisnih opcija, uključujući i link za SPSS homepage, statističkog trenera i syntax vodiča. Koristeći „topics, možete koristiti opciju indeks da ukucate bilo koji taster, reč i dobiti spisak opcija, ili možete da vidite kategorije i podkategorije dostupne pod sadržajem. Ovo je odličan alat i može se koristiti da biste rešili većinu problema.

7.13. Paleta alatki i statusna traka u programu SPSS

Najlakši način korišćenja programa SPSS je korišćenjem palete alatki i skupa ikona koje se nalaze ispod menija. Na slici 5.4. se može videti paleta alatki Data View, koja predstavlja neku od komandi na SPSS meniju. Pritisnite na neku ikonu i odgovarajuća komanda će se izvršiti. Tako umesto da idete na meni Utilities da biste izabrali promenljive, jednostavno pritisnete ikonu Variables (promenljive) na paleti alatki.¹⁵



Slika 5.4 - Paleta alatki prozora Data Editor

8. Primena SPSS-a

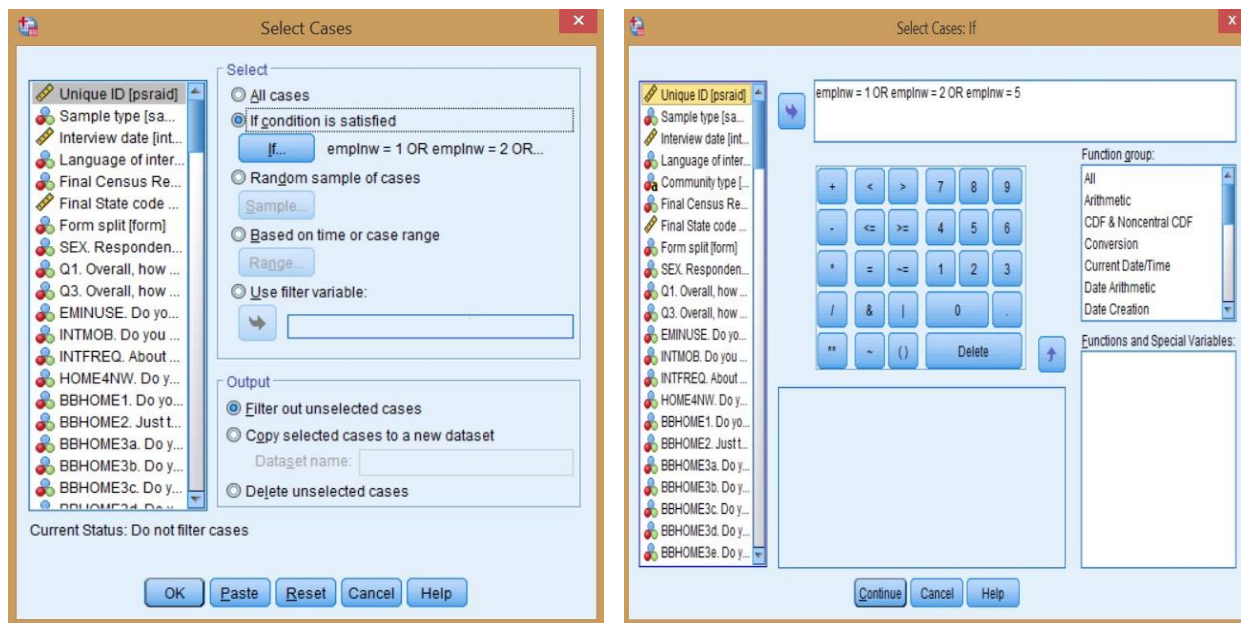
8.1. Definisanje podataka

Podaci koji su uneti u program, nisu nužno uvek podobni za dalju analizu. Ponekad se desi da varijabla nije ispravno numerički kodirana, ili je potrebno dodatno je definisati, izračunati i slično. Prema tome, nakon unošenja podataka u program, a pre statističkih analiza i dodatnih kalkulacija, poželjno je izvršiti selekciju podataka. Nju je moguće izvršiti pomoću komande Data koja sadrži skup komandi koje pružaju razne mogućnosti za rad sa podacima: definisanje promenljivih, kopiranje atributa promenljivih, definisanje grupe promenljivih, kontrolu ispravnosti podataka, sortiranje i transponovanje podataka, filtriranje podataka, ponderisanje i ostalo.

¹⁵ http://cet.rs/media/k2/attachments/spss_pog01.pdf, stranica posećena 13. marta 2017.

Filtriranje podataka

Ovom komandom se izdvajaju (filtriraju) opservacije aktivne datoteke podataka koja će se dalje procesuirati. Na slici 6.1 prikazan je primer izdvajanja opservacija.



Slika 6.1 - primer komande Select Cases i funkcije IF

Nakon što smo štiklirali opciju *If condition is satisfied*, pritiskom na dugme IF otvara nam se novi prozor u kome zadajemo i definišemo tačne uslove koji se odnose na datu promenljivu „*emplnw*“ (pokazuje odgovore ispitanika na pitanje da li ste trenutno zaposleni u kategoriji: 1. punog radnog vremena; 2. parcijalnog radnog vremena; 3. kao volonter (niste plaćeni za posao koji obavljate); 4. niste zaposleni; 5. u svojoj firmi; 6. student; 7. onesposobljen za rad; 8. ostalo?). Uslove smo definisali pomoću logične funkcije OR koja daje tačan izlazni rezultat kad je bar jedna od pretpostavki tačna. U našem slučaju, zadali smo uslove da promenljiva „*emplnw*“ bude jednaka 1, 2 ili 5, pošto su to tri jedina odgovora koja podrazumevaju plaćen posao i posao pod ugovorom. Kao izlazni rezultat, u prozoru Editora podataka može se primetiti da su precrtane kosom linijom one opservacije koje ne ulaze u dalju analizu, tačnije, ne ispunjavaju zadate uslove funkcije IF.

	psraid	sample	inf_date	lang	usr	region	state	form	sex	q1	q3	emmuse	intmob	intfreq	home4nw	bbhome1	bbhome2	bbhome3a
1	100006	1	61015	1 S		3	51	1	2	1	2	1	1	1	1	2		3
2	100009	1	61015	1 U		3	24	2	2	2	3	1	1	5	1	1	1	3
3	100011	1	61015	1 S		2	39	2	2	1	2	2	2	2				3
4	100039	1	61015	1 S		3	37	2	2	1	2	1	1	2	1	2		3
5	100040	1	61015	1 U		3	21	1	2	1	2	1	2	3	1	2		3
6	100043	1	62215	1 R		2	20	1	2	1	2	1	1	4	1	2		1
7	100049	1	61015	1 R		1	42	1	1	1	4	2	2					3
8	100051	1	61915	1 R		2	26	1	1	1	2	2	2					2
9	100058	1	61015	1 R		3	40	2	1	2	2	1	1	3	1	2		3
10	100059	1	61015	1 U		1	23	1	1	3	4	2	2					2
11	100060	1	61015	1 R		2	20	1	2	1	2	1	1	4	1	2		3
12	100065	1	61915	1 R		3	40	1	2	3	2	1	2	2	1	2		2
13	100067	1	61015	1 R		2	27	1	1	1	3	1	1	1	1	2		3
14	100079	1	62315	1 U		3	12	1	2	3	2	1	1	2	1	2		3
15	100080	1	62015	1 S		3	45	2	2	4	4	1	1	2	1	3		2
16	100084	1	61015	1 R		3	1	2	2	2	2	1	1	2	1	2		3
17	100091	1	61915	1 S		3	21	2	1	1	3	2	2					3
18	100096	1	61015	1 R		3	5	2	1	3	2	1	1	2	1	2		3
19	100102	1	61015	1 U		1	36	2	2	1	2	1	2	2	1	2		2
20	100118	1	61015	1 U		3	1	2	1	1	2	1	1	2	1	2		3
21	100129	1	62315	1 R		4	30	1	2	4	4	1	1	1	1	2		2
22	100134	1	61015	1 S		2	20	1	1	2	3	2	2					3
23	100141	1	61015	1 R		3	48	2	2	2	2	2	2					3
24	100142	1	61015	1 S		3	45	2	1	4	4	2	2					3
25	100143	1	61015	1 U		1	9	2	1	1	3	1	2	1	1	2		1
26	100151	1	61015	1 U		3	51	1	2	2	3	1	2	2	1	2		3
27	100152	1	62415	1 S		4	49	1	2	2	4	1	1	2	1	2		3
28	100156	1	63115	1 S		2	27	2	2	1	2	1	2	2	1	2		1

Slika 6.2 - Rezultat komande Select Cases

8.2. Deskriptivna statistika

Deskriptivna statistička analiza predstavlja skup metoda kojima se vrši izračunavanje, prikazivanje i opisivanje osnovnih karakteristika statističkih serija. Deskriptivna statistička analiza ima sledeće zadatke:

1. grupisanje i sređivanje statističkih podataka,
2. prikazivanje statističkih podataka,
3. određivanje osnovnih pokazatelja statističkih serija.¹⁶

Pritiskom na opciju *Analyze* pojavljuje se padajući meni gde možemo pronaći opciju za deskriptivnu statističku analizu. Kako bismo saznali mišljenje zaposlenih o tome da li misle da će struka u kojoj su trenutno zaposleni istrajati i postojati za 50 godina, od sada, biramo opciju *Analyze/Descriptive Statistics/Frequencies*. U polje *Variable(s)* unosimo promenljivu *AUTO2* („Thinking about the job or occupation that you work in now, how likely do you think it is that job will exist in its current form in 50 years?“), i podešavamo željene statistike i dijagrame koje želimo da analiza prikaže, u delu *Statistic* i *Charts*. Nakon što podesimo sve potrebne statističke parametre, puštamo program u obradu podataka. Na slici 7.1, u tabeli možemo pronaći tačan broj ispitanika opredeljen za svaku ponuđenu opciju, kao i procentualni deo. Za lakše tumačenje analize, u donjem delu slike nalazi se grafički prikaz funkcije, odnosno dijagram stubića, poznatiji kao Histogram, gde su na Y-osi prikazane kategorije promenljive *AUTO2*, a na X-osi frekvencija ispitanika opredeljenih za svaku kategoriju posebno.

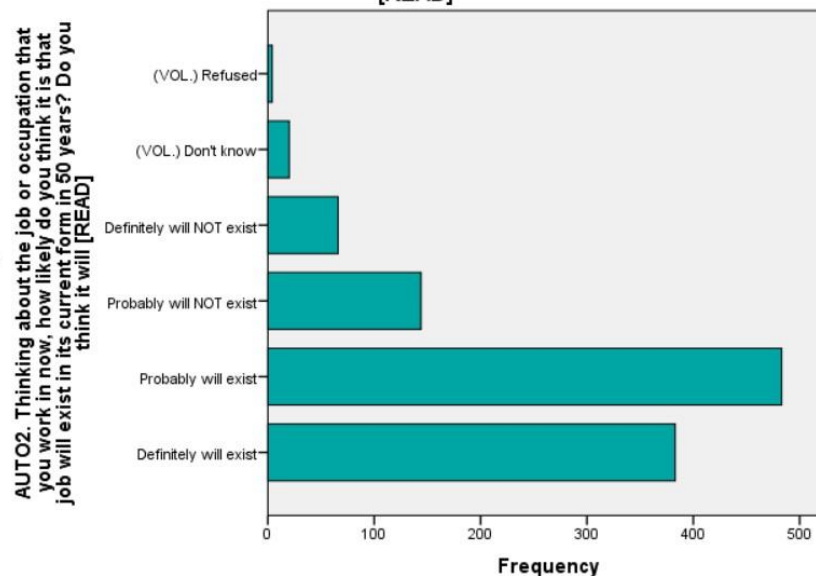
¹⁶ www.ef.uns.ac.rs, stranica posećena 15. marta 2017.

Kao što se može videti sa Histograma, najveći broj zaposlenih ljudi deli mišljenje da će njihova struka *Najverovatnije opstati* narednih 50 godina.

AUTO2. Thinking about the job or occupation that you work in now, how likely do you think it is that job will exist in its current form in 50 years? Do you think it will [READ]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Definitely will exist	383	34,8	34,8	34,8
	Probably will exist	483	43,9	43,9	78,7
	Probably will NOT exist	144	13,1	13,1	91,8
	Definitely will NOT exist	66	6,0	6,0	97,8
	(VOL.) Don't know	20	1,8	1,8	99,6
	(VOL.) Refused	4	,4	,4	100,0
	Total	1100	100,0	100,0	

AUTO2. Thinking about the job or occupation that you work in now, how likely do you think it is that job will exist in its current form in 50 years? Do you think it will [READ]

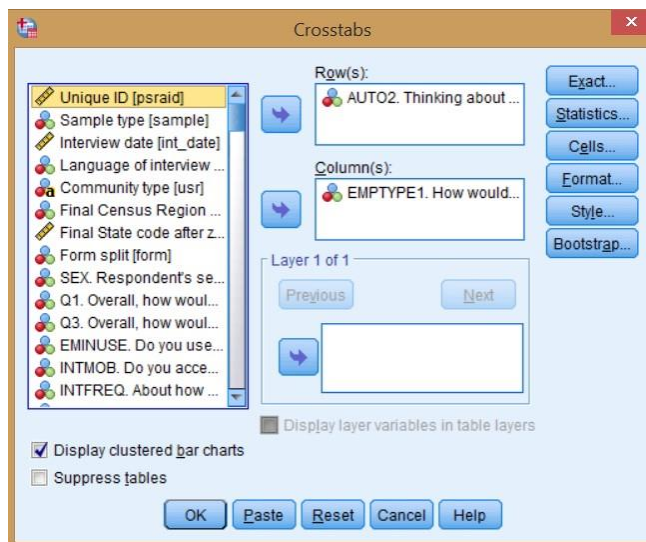


Slika 7.1 - Rezultat Deskriptivne statistike i funkcije *Frequencies*

8.3. Tabeliranje podataka

Krostabulacija (tabela ukrštanja/kontingencije) je postupak kojim ispitujeemo odnose između varijabli. Tabeliranje podataka je jednostavan proces računanja broja opservacija koje pripadaju određenoj kategoriji (klasi ili grupi) posmatrane promenljive.¹⁷ Tabeliranje se koristi kao prva preliminarna faza analize podataka u mnogim istraživanjima i predstavlja pogodan alat za prvo sumarno prikazivanje podataka. Jedna od najčešćih formi tabeliranja jeste i unakrsno tabeliranje (*Cross-tabulations*) koju koristimo u našem istraživanju. Unakrsno tabeliranje se primenjuje kada analitičar želi da ispita vezu-odnos između dve ili više kategorijskih promenljivih.

Kako bismo produbili analizu i saznali čime se bave ispitanici koji su u prethodnoj fazi ispitivanja odabrali opciju *Najverovatnije opstati* kao odgovor na pitanje „Da li misle da će struka u kojoj su trenutno zaposleni istrajati i postojati za 50 godina?“, biramo promenljivu „*AUTO2*“ koju smeštamo u deo *Row(s)* i promenljivu „*EMPTYE1*“ koju premeštamo u deo *Column(s)*.



Slika 8.1 - definisanje analize pomoću funkcije *Crosstabs*

U donjem levom uglu, na slici 8.1, može se primetiti da je štiklirana opcija *Display clustered bar charts*, koja će nam grafički predstaviti vezu između mesta na kome su ispitanici zaposleni i odgovora na pitanje koje definiše varijabla „*AUTO2*”.

Izlazni rezultat pojavljuje se u prozoru Output, i sastoji se iz dva dela: unakrsne tabele koja prikazuje odnos dve promenljive i gore spomenutog grafika. U tabeli je dat sumarni prikaz svih promenljivih koje su učestvovala u analizi kao i detaljnije prikazano mišljenje ispitanika. U našem slučaju, koncentrišemo se na grafik i na deo gde je odgovor na pitanje glasio

¹⁷ Stevan Živković, *Analiza podataka u SPSS-u, priručnik iz statistike*, De Facto Consultancy, Podgorica 2015. str. 37

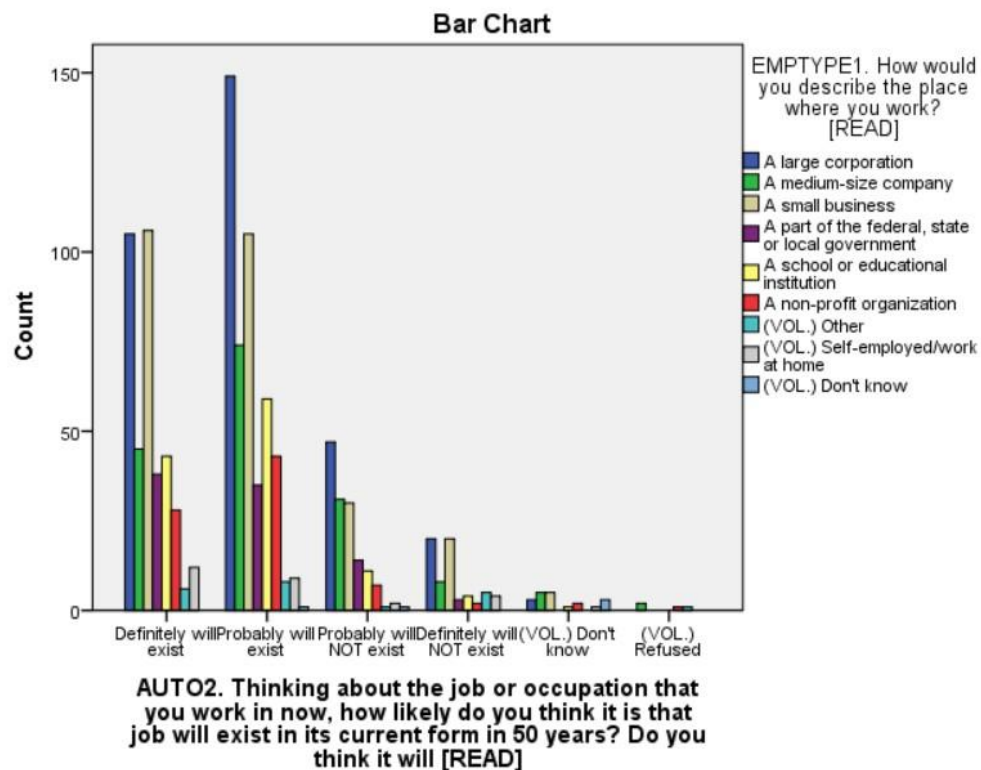
Najverovatnije opstati. Plavom bojom prikazani su ispitanici koji su deo Velikih korporacija, i oni zauzimaju najveći procenat zaposlenih sa pozitivnim odgovorom na pitanje. Posle njih, slede zaposleni koji su deo Malih biznisa, pa Srednjih kompanija, zatim, Škola ili edukacijskih ustanova, Neprofitnih organizacija, Vlade ili Države, samostalni radnici i ostali.

Prema tome, možemo zaključiti, na osnovu ispitivanja ljudi koji su trenutno zaposleni u Velikim korporacijama, da 79% deli mišljenje da će njihova struka opstati u narednih 50 godina.

AUTO2. Thinking about the job or occupation that you work in now, how likely do you think it is that job will exist in its current form in 50 years? Do you think it will [READ] * EMPTY1. How would you describe the place where you work? [READ] Crosstabulation

			EMPTY1. How would you describe the place where you work? [READ]										
			A large corporation	A medium-size company	A small business	A part of the federal, state or local government	A school or educational institution	A non-profit organization	(VOL.) Other	(VOL.) Self-employed/work at home	(VOL.) Don't know	Total	
AUTO2. Thinking about the job or occupation that you work in now, how likely do you think it is that job will exist in its current form in 50 years? Do you think it will [READ]	Definitely will exist	Count	105	45	106	38	43	28	6	12	0	383	
		% of Total	9,5%	4,1%	9,6%	3,5%	3,9%	2,5%	0,5%	1,1%	0,0%	34,8%	
	Probably will exist	Count	149	74	105	35	59	43	8	9	1	483	
		% of Total	13,5%	6,7%	9,5%	3,2%	5,4%	3,9%	0,7%	0,8%	0,1%	43,9%	
	Probably will NOT exist	Count	47	31	30	14	11	7	1	2	1	144	
		% of Total	4,3%	2,8%	2,7%	1,3%	1,0%	0,6%	0,1%	0,2%	0,1%	13,1%	
	Definitely will NOT exist	Count	20	8	20	3	4	2	5	4	0	66	
		% of Total	1,8%	0,7%	1,8%	0,3%	0,4%	0,2%	0,5%	0,4%	0,0%	6,0%	
	(VOL.) Don't know	Count	3	5	5	0	1	2	0	1	3	20	
		% of Total	0,3%	0,5%	0,5%	0,0%	0,1%	0,2%	0,0%	0,1%	0,3%	1,8%	
(VOL.) Refused	Count	0	2	0	0	0	1	1	0	0	4		
	% of Total	0,0%	0,2%	0,0%	0,0%	0,0%	0,1%	0,1%	0,0%	0,0%	0,4%		
Total	Count	324	165	266	90	118	83	21	28	5	1100		
	% of Total	29,5%	15,0%	24,2%	8,2%	10,7%	7,5%	1,9%	2,5%	0,5%	100,0%		

Slika 8.2 - rezultat funkcije *Crosstabs*



Slika 8.3 - grafički prikaz funkcije *Crosstabs* pod nazivom Bar Chart

8.4. Testovi značajnosti razlika

Ako organizujemo statistička posmatranja sa ciljem da se analiziraju uzorci, dolazimo do toga da se opisi upoređuju ili sa nekim zamišljenim teorijskim veličinama, ili sa nekim drugim uzorcima koji su organizovani na isti način. Istovetnosti ili razlike mogu da budu organizovane ili slučajne. Slučajne razlike su rezultat slučajnih variranja po pravcu, tačnije smislu i ne mogu da se predvide. Organizovane promene su nastale delovanjem faktora koji se mogu, ili koje su poželjne da budu identifikovane.

Poređenjem testova pripadnosti uzorka skupa, odnosno međusobne razlike, istraživač postavlja radnu i nultu hipotezu: H_0 – za nultu hipotezu se uzima pretpostavka da ne postoji statistički velika razlika između populacije i uzorka (ili između dva uzorka); H_1 – alternativna ili radna hipoteza: u testiranjima se obično pretpostavlja da postoje statistički velike razlike između populacije i uzorka (ili između dva uzorka). Može se postaviti i nivo značajnosti sa kojim će se raditi testiranje, odnosno nivo sigurnosti sa kojom možemo tvrditi da je nešto onako kako tvrdimo. Nivo standardnog nivoa pouzdanosti se uglavnom uzima kao nivo $p=0.05$, što znači da se sa sigurnošću od 95% može tvrditi da je nešto onako kako tvrdimo. Pouzdanost većeg stepena se predstavlja kao $p=0.01$, što znači da smo 99% sigurni u ono što tvrdimo.

8.5. Procedura *One-way ANOVA*

Iz statističke literature je poznato da se za testiranje jednakosti prosečne vrednosti jednog numeričkog obeležja za više od dve grupe ispitanika koristi procedura analize varijansi. Ako pretpostavimo da ispitujemo jednakost prosečnih vrednosti neke pojave za c grupu onda imamo sledeću nultu hipotezu:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_c$$

Alternativna hipoteza je da postoji razlika između prosečnih vrednosti posmatranog obeležja za bar dve grupe ispitanika.

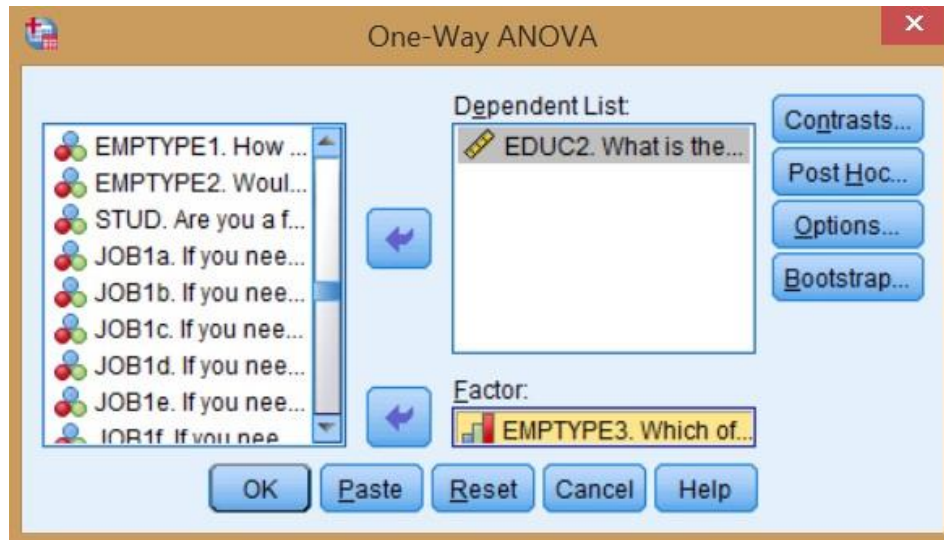
Osnovne pretpostavke primene procedure *One-way ANOVA* su: opservacije su nezavisne, izvučene su iz populacija sa normalnom raspodelom posmatranog obeležja i jednakim varijansama. Naravno, postoji još jedna bitna pretpostavka koja se odnosi na tip podataka, a to je da je zavisna promenljiva (pojava koja se posmatra) metrička promenljiva, što znači da je merena bar na intervalnoj skali.

Pretpostavimo da želimo da ispitamo postoje li značajne razlike između najvišeg dostignutog nivoa edukacija i trenutnog radnog mesta koji ispitanici obavljaju. Prvi korak ove procedure je definisanje zavisne i nezavisne (jedne ili više) promenljive. Zavisna promenljiva je metrička promenljiva (merena bar na intervalnoj skali), a nezavisna promenljiva (jedna ili više) je kategorijska. U SPSS proceduri ***One-way ANOVA*** postoji samo jedna nezavisna promenljiva. U našem primeru, zavisna promenljiva je „EDUC2“ (najviši dostignut nivo obrazovanja), a nezavisna promenljiva je „EMPTYE3“ (trenutna pozicija na mestu zaposlenja). Osnovna hipoteza koju testiramo je:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = \mu_9 = \mu_{10}^{18}$$

protiv alternativne hipoteze da postoji razlika između bar dve prosečne vrednosti.

Izbor procedure analize varijansi vrši se na sledeći način: *Analyze/Compare Means/One-way ANOVA*.



Slika 9.1 – procedura analize varijansi u One-way ANOVA

U polje *Dependent List*, ubacujemo zavisnu promenljivu „EDUC2“, a u polje *Factor* nezavisnu promenljivu „EMPTYTYPE3“. Glavni rezultati primene procedure prikazani su na sledećim slikama:

Descriptives

EDUC2. What is the highest level of school you have completed or the highest degree you have received? [DO NOT READ]

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Professional	349	6,86	5,175	,277	6,31	7,40	1	99
Manager or executive	117	6,20	1,549	,143	5,91	6,48	3	8
Government official	9	5,44	2,128	,709	3,81	7,08	3	8
Administrative or clerical	69	5,29	1,664	,200	4,89	5,69	2	8
Customer service	87	4,83	1,623	,174	4,48	5,17	1	8
(VOL.) Other (SPECIFY)	17	4,59	2,123	,515	3,50	5,68	1	8
(VOL.) Don't know	6	4,17	2,927	1,195	1,10	7,24	1	8
Total	654	6,20	4,031	,158	5,89	6,51	1	99

Slika 9.2 – rezultat deskriptivne statističke analize

¹⁸ Zato sto imamo 10 kategorija dostignutog nivoa obrazovanja

Na slici 9.2 prikazan je rezultat deskriptivne statističke analize koja je deo procedure One-way ANOVA. Na osnovu te tabele može se zaključiti da raspodela za posmatranu promenljivu ne odstupa značajno od normalne raspodele¹⁹, a slika 9.3 daje uvid u rezultate testiranja hipoteze o homogenosti varijanse posmatranih obeležja za različite kategorije ispitanika prema dostignutom nivou obrazovanja. Za testiranje ove hipoteze koristimo *Levene test* koji je, od svih testova koji za pretpostavku imaju normalnu raspodelu pojave, najmanje osetljiv na narušenost ove pretpostavke. *Levene test* se zasniva na računanju apsolutnih razlika vrednosti promenljive za svaku opservaciju od grupne sredine, da bi se zatim primenila analiza varijanse na tako dobijene rezultate.

U prvoj tabeli „Test of Homogeneity of Variances“, na slici 9.3, poslednja kolona Sig. predstavlja realizovan nivo značajnosti Levene testa i on je jednak =0,946. Kako je realizovan nivo značajnosti veći od 0,05 zaključujemo da je za ove promenljive reč o homogenovarijabilitetu po grupama.

Test of Homogeneity of Variances			
EDUC2. What is the highest level of school you have completed or the highest degree you have received? [DO NOT READ]			
Levene Statistic	df1	df2	Sig.
,282	6	647	,946

ANOVA					
EDUC2. What is the highest level of school you have completed or the highest degree you have received? [DO NOT READ]					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	445,655	6	74,276	4,729	,000
Within Groups	10163,105	647	15,708		
Total	10608,760	653			

Slika 9.3 – rezultat procedure *One-way ANOVA*

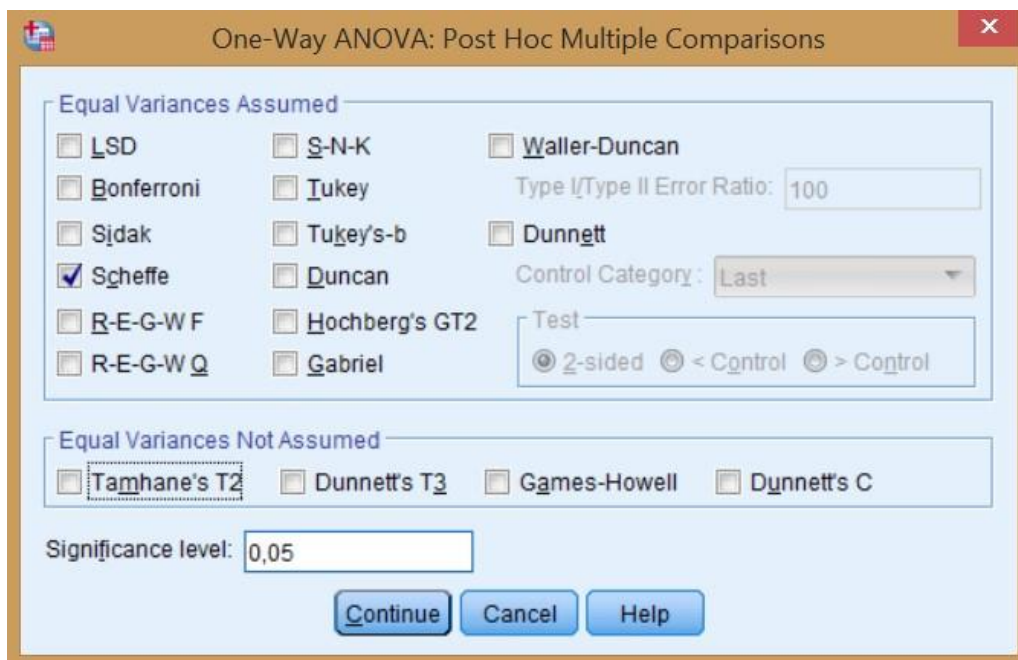
¹⁹ Normalna normirana raspodela je normalna raspodela sa očekivanjem jednakim nuli, i varijansom jednakom jedan.

U drugoj tabeli „ANOVA“, na slici 9.3, prikazani su rezultati primene *one-way* ANOVA testa za promenljivu EDUC2, za koju ne možemo da prihvatimo tvrđenje da postoji jednakost prosečnih vrednosti dostignutog nivoa obrazovanja u deset kategorija trenutnih radnih pozicija (Sig. 0,000).

Kada se ANOVA testom odbaci nulta hipoteza (H_0) o jednakosti prosečnih vrednosti posmatranog obeležja za više grupa, dalje se sprovodi *post-hoc* analiza, koja ukazuje na parove grupa gde postoje statistički značajne razlike u prosečnim vrednostima (slika 9.4). *Post-hoc* analiza može se odabrati kao opcija u prvobitnom prozoru ANOVA procedure, kao što se može primetiti na 9.1.

Kako se za promenljivu „EDUC2“ ne odbacuje hipoteza o jednakosti varijabiliteta po grupama, za *post-hoc* analizu primenjen je jedan od testova iz grupe *Equal Variances Assumed* (*Scheffe* test) – slika 9.5. U slučaju da se za promenljivu „EDUC2“ odbacivala nulta hipoteza, za višestruko poređenje primenili bismo jedan od testova iz grupe *Equal Variances Not Assumed* (*Tamhane's* test).

Na sledećim slikama dati su rezultati *post-hoc* analize.



Slika 9.4 – One way ANOVA prozor

Post Hoc Tests

Multiple Comparisons

Dependent Variable: EDUC2. What is the highest level of school you have completed or the highest degree you have received? [DO NC Scheffe

(I) EMPTY3. Which of the following best describes the type of work that you do? [READ]	(J) EMPTY3. Which of the following best describes the type of work that you do? [READ]	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Professional	Manager or executive	,660	,423	,876	-,85	2,17
	Government official	1,412	1,338	,981	-3,35	6,18
	Administrative or clerical	1,567	,522	,175	-,29	3,43
	Customer service	2,029*	,475	,006	,34	3,72
	(VOL.) Other (SPECIFY)	2,268	,984	,505	-1,24	5,77
	(VOL.) Don't know	2,690	1,632	,843	-3,12	8,50
Manager or executive	Professional	-,660	,423	,876	-2,17	,85
	Government official	,752	1,371	,999	-4,13	5,63
	Administrative or clerical	,907	,602	,893	-1,24	3,05
	Customer service	1,369	,561	,429	-,63	3,37
	(VOL.) Other (SPECIFY)	1,608	1,029	,874	-2,05	5,27
	(VOL.) Don't know	2,030	1,659	,959	-3,88	7,94
Government official	Professional	-1,412	1,338	,981	-6,18	3,35
	Manager or executive	-,752	1,371	,999	-5,63	4,13
	Administrative or clerical	,155	1,405	1,000	-4,85	5,16
	Customer service	,617	1,388	1,000	-4,32	5,56
	(VOL.) Other (SPECIFY)	,856	1,634	1,000	-4,96	6,67
	(VOL.) Don't know	1,278	2,089	,999	-6,16	8,71
Administrative or clerical	Professional	-1,567	,522	,175	-3,43	,29
	Manager or executive	-,907	,602	,893	-3,05	1,24
	Government official	-,155	1,405	1,000	-5,16	4,85
	Customer service	,462	,639	,998	-1,81	2,74
	(VOL.) Other (SPECIFY)	,702	1,073	,999	-3,12	4,52
	(VOL.) Don't know	1,123	1,687	,998	-4,88	7,13
Customer service	Professional	-2,029*	,475	,006	-3,72	-,34
	Manager or executive	-1,369	,561	,429	-3,37	,63
	Government official	-,617	1,388	1,000	-5,56	4,32
	Administrative or clerical	-,462	,639	,998	-2,74	1,81
	(VOL.) Other (SPECIFY)	,239	1,051	1,000	-3,50	3,98
	(VOL.) Don't know	,661	1,673	1,000	-5,30	6,62
(VOL.) Other (SPECIFY)	Professional	-2,268	,984	,505	-5,77	1,24
	Manager or executive	-1,608	1,029	,874	-5,27	2,05
	Government official	-,856	1,634	1,000	-6,67	4,96
	Administrative or clerical	-,702	1,073	,999	-4,52	3,12
	Customer service	-,239	1,051	1,000	-3,98	3,50
	(VOL.) Don't know	,422	1,882	1,000	-6,28	7,12
(VOL.) Don't know	Professional	-2,690	1,632	,843	-8,50	3,12
	Manager or executive	-2,030	1,659	,959	-7,94	3,88
	Government official	-1,278	2,089	,999	-8,71	6,16
	Administrative or clerical	-1,123	1,687	,998	-7,13	4,88
	Customer service	-,661	1,673	1,000	-6,62	5,30
	(VOL.) Other (SPECIFY)	-,422	1,882	1,000	-7,12	6,28

*. The mean difference is significant at the 0.05 level.

Slika 9.5 Post Hoc analiza

Postoji sedam kategorija trenutnih radnih mesta, definisanih nakon sakupljanja svih odgovora ispitanika. To su :

1. Eksperti
2. Menadžeri
3. Predstavnici Vlade
4. Administrativni radnici
5. Korisnička podrška
6. Ostalo
7. Ne znam/Nisam siguran

Kada svaku od ovih kategorija uporedimo međusobno, dobijamo tabelu koja je prikazana na slici 9.5. Ta tabela se zove **Multiple Comparisons** i ona je rezultat *post-hoc* analize.

U koloni *Mean* (prosečna vrednost) data je razlika između srednjih vrednosti dve kategorije radnih pozicija. Pored nje, nalazi se kolona *Std.Error* (Standardna greška) koja pokazuje u kojoj meri data statistika uzorka predstavlja tačnu procenu populacijskog parametra. U koloni *Sig.*, koja nam je najvažnija za ovu analizu i koja predstavlja realizovani nivo značajnosti testa, tražimo najmanju vrednost rezultata. U ovom slučaju, to je 0.006, pa prema tome, možemo zaključiti sledeće :

na nivou značajnosti od 5%, može se zaključiti da postoji statistički značajna razlika u prosečnoj vrednosti dostignutog nivoa obrazovanja između ispitanika koji su trenutno zaposleni na poziciji koja odgovara nivou Eksperta i poziciji koja je predstavljena kao Korisnička podrška.

9. ZAKLJUČAK

Tehnologija zasnovana na statističkoj analizi podataka omogućava brzo uočavanje trendova i zakonitosti u velikim količinama podataka i predviđanje budućnosti na osnovu njih. Poslovni programi koji koriste tehnologiju zasnovanu na statističkoj analizi podataka daju veću sigurnost u poslovanju, jer istovremeno postaju esencijalno značajan alat koji se koristi pri donošenju poslovnih odluka.

Dobra i precizna analiza podataka je od velikog značaja za većinu organizacija. Međutim, da bi se dobila kvalitetna analiza, kao i izveštaji potrebni za donošenje poslovnih odluka, neophodno je mnogo znanja i veština kako bi se isplanirao sam postupak analize, prikupili podaci, upravljalo i pristupilo podacima, izvršila analiza i, na kraju, putem raznih izveštaja konačni rezultati prezentovali svim zainteresovanim subjektima. Dobro rešenje je izbor jednog analitičkog alata koji se sastoji od potrebnih modula ili samostalnog softvera koji ima različite specijalizovane funkcije.

U godinama koje dolaze, može se očekivati da će se softverski paketi sve više usavršavati kako bi održali korak sa savremenim svetom, istovremeno udovoljavajući potrebama modernog tržišnog poslovanja.

Od samostalne vožnje vozila i polu-autonomnih robota do inteligentnih algoritama i prediktivnih analitičkih alata, jasno je da su mašine sve više u stanju da obavljaju širok spektar poslova koji su već odavno u ljudskim domenima.

Iako većina Amerikanaca očekuju da će se značajni nivoi radne snage i automatizacija poslova dogoditi u narednih 50 godina, većina današnjih zaposlenih izražava uverenje da njihovi poslovi ili zanimanja neće biti ugroženi u značajnoj meri. 36% radnika predviđa da će njihovi trenutni poslovi ili zanimanja "definitivno" postojati u njihovim sadašnjim oblicima pet decenija od sada, dok dodatnih 44% očekuje da će njihovi poslovi "verovatno" postojati u 50 godina. Otprilike jedan od pet radnika očekuje da njihovi sadašnji poslovi "verovatno neće" (12%) ili "definitivno neće" (6%) postojati tako daleko u budućnosti u njihovim sadašnjim oblicima.

Uprkos očekivanjima da će tehnologija zadreti u ljudska zaposlenja uopšte, zaključujemo da većina radnika smatra da će njihovi poslovi ili zanimanja i dalje postojati u narednih 50 godina.

10. LITERATURA

- [1] <http://savtajr.blogspot.rs/p/sta-je-r.html>
- [2] Andrie de Vries and Joris Meys, *R for dummies*, John Wiley and Sons, Ltd, England, 2012. Str. 22
- [3] <http://science.webhostinggeeks.com/programiranje-u-r>
- [4] R Relative to Statistical Packages: Comment 1 on Technical Report Number 1 (Version 1.0) Strategically using General Purpose Statistics Packages: A Look at Stata, SAS and SPSS Patrick Burns, Burns Statistics, patrick@burns-stat.com
- [5] Emmanuel Paradis, *R for Beginners*, France, 2005, Str. 38
- [6] Osnove Matlab-a doc.dr.sc. Željko Ban
- [7] <https://sr.wikipedia.org/sr-el/MATLAB>
- [8] <https://nadicakusic.files.wordpress.com/2012/03/matlab-osnove.pdf>
- [9] <https://www.csun.edu/sites/default/files/statistics20-core-system-guide-64bit.pdf>
- [10] <http://en.wikipedia.org/wiki/SPSS>
- [11] Julie Pallant, *SPSS Survival manual, Third edition, Mikro knjiga, Beograd, 2009. Str. 33*
- [12] http://www.ssc.wisc.edu/sscc/pubs/spss/classintro/spss_students1.html#data-view
- [13]] <http://www.ftn.kg.ac.rs/download/SIR/SIR%20Cuslovic%20Milos.pdf>
- [14] <http://www.uvm.edu/~dhowell/fundamentals7/SPSSManual/SPSSLongerManual/SPSSChapter1.pdf>
- [15] http://cet.rs/media/k2/attachments/spss_pog01.pdf
- [16] www.ef.uns.ac.rs
- [17] Stevan Živković, *Analiza podataka u SPSS-u, priručnik iz statistike*, De Facto Consultancy, Podgorica 2015. str. 37