

# ANALITIČKE KARAKTERISTIKE STATISTIČKOG SOFTVERA SPSS KROZ PRIMER

Nataša Papić-Blagojević\*

**Sažetak:** Statistički paket SPSS for Windows pripada redu najprimenjivijih programa za statističku analizu. Osim toga što značajno doprinosi rešavanju poslovnih i istraživačkih problema, ujedno nudi i širok spektar mogućnosti za celokupan analitički proces. U radu su objašnjene neke od procedura za sumarno prikazivanje podataka i poređenje sredina i njihova primenjivost na konkretnom primeru.

**Ključne reči:** statistički softver, proces analize, procedure

**Summary:** Statistical package SPSS for Windows is among the most widely used programmes for statistical analysis. It is very usefull in solving business and research problems and, also, in providing broad range or opportunities for the entire analytical process. This paper describes some procedures of Descriptive and Bivariate statistics on the example.

**Key words:** statistics software, analytical process, procedures

## Uvod

Korišćenje statističkih procedura SPSS-a obezbeđuje informacije koje će poslužiti u kasnijem donošenju odluka. Rezultati analize, prezentovani u tabelarnim i grafičkim autputima, doprinose efikasnijem i bržem odlučivanju kroz otkrivanje elementarnih činjenica i obrazaca.

SPSS for Windows, moguće je koristiti u različitim oblastima, na primer:

- Istraživanje tržišta i direktni marketing;
- Visokoškolsko obrazovanje;
- Administrativna istraživanja, ljudski resursi i planiranje resursa;
- Oblasti medicinskog, kliničkog i socijalnog istraživanja;
- Planiranje i predviđanje;
- Unapređenje kvaliteta;
- Izveštavanje i ad hok odlučivanje .

## Analitički proces

Proces analize podataka, sam po sebi, nije jednostavan. Prvo je potrebno prikupiti podatke, sprovesti analizu i potom rezultate objaviti u upotrebljivom i razumljivom formatu. U slučaju korišćenja softvera ograničenih mogućnosti, okončanje posla bi zahtevalo korišćenje rezultata iz različitih izvora. SPSS nudi sve što je potrebno za proces analize, od planiranja, obrade podataka, analiziranja do izlaznih rezultata, i to na jednom mestu.

Izgradnja sistema za proces analize podataka obuhvata nekoliko faza, gde SPSS, sa svojim proizvodima, olakšava rad u svakoj od njih:

1. **Planiranje** (engl. *Planning*) – ušteda vremena i novca korišćenjem SamplePower, SPSS Complex Samples i SPSS Conjoint;

---

\* Nataša Papić-Blagojević (saradnik u nastavi), Visoka poslovna škola strukovnih studija, Novi Sad

2. **Prikupljanje podataka** (engl. *Data collection*) – efikasnije prikupljanje i unošenje podataka uz pomoć Dimensions and SPSS Data Entry;
3. **Pristup podacima** (engl. *Data access*) – brža analiza uz lakši pristup podacima: SPSS Base;
4. **Priprema i upravljanje podacima** (engl. *Data management and data preparation*) – kroz efikasniju pripremu i upravljanje podacima, brže se stiže i do procesa analize: SPSS Base, SPSS Complex Samples, SPSS Missing Value Analysis, SPSS Data Preparation, SPSS Text Analysis for Surveys and Dimensions;
5. **Analiza podataka** (engl. *Data analysis*) – analiza podataka kroz korišćenje različitih statističkih procedura u cilju dobijanja preciznijih modela: SPSS Base, SPSS Regression Models, SPSS Advanced Models, SPSS Complex Samples, SPSS Data Preparation, SPSS Classification Trees, SPSS Categories, SPSS Tables, SPSS Exact Tests, SPSS Trends and Amos;
6. **Izveštavanje** (engl. *Reporting*) – prezentovanje rezultata korisnicima: SPSS Base, SPSS Tables, SPSS Maps and Dimensions;
7. **Razvoj** (engl. *Deployment*) – razmena rezultata preko Web ili drugih prezentacija i publikacija: SmartViewer Web Server and SmartScore.



Slika 1. Analitički proces<sup>14</sup>

<sup>14</sup> <http://www.spss.com> – Analitički proces prikazuje korake kroz koje treba proći kako bi se podaci pripremili za analizu, sproveda sama analiza, podneli izveštaji i prezentovali rezultati.

## SPSS for Windows

SPSS for Windows pripada redu aplikativnih softvera, pod kojima se podrazumevaju skupovi programa orijentisanih na najrazličitija područja. Aplikativne programe, u principu, izrađuje sam korisnik prema svojim potrebama i njihova glavna svrha je rešavanje konkretnog problema, što je, ujedno, i osnovni kriterijum koji ih svrstava ili ne svrstava u kategoriju aplikativnog softvera.

Uzimajući u obzir sva saznanja do kojih se došlo prilikom izučavanja statističkog paketa SPSS for Windows, u nastavku je ispoljena jasna namera da se, kroz adekvatan primer, prikaže njegova celovitost i upotrebljivost uz kratko objašnjenje svake od primenjenih procedura.

### Editor podataka

Ukoliko se opredeli za rad u ovom statističkom paketu, glavni deo posla će se obaviti u nekom od njegovih prozora: aplikacionom prozoru, editoru podataka ili prozoru podataka.

**Editor podataka-Data Editor** ima izgled pravougaone tabele, u čijim se kolonama prikazuju varijable, a u redovima opservacije. Svi postupci počev od unošenja podataka, korigovanja, brisanja, definisanja varijabli i sl., vrše se preko linije editora, a podatak se smešta u trenutno aktivnu ćeliju. U tabeli 1 prikazan je deo tabele u editoru podataka koja sadrži pet promenljivih i 21 opservaciju.

Tabela 1. SPSS Editor podataka

	nazivus	pozpovi	proizv	vtuseva	prinos				
1	pšenica	563269	2007060	z	prinos po ha u tonama				
2	raž	7168	15778	z	2,20				
3	grašak	11926	32807	pb	2,70				
4	jecam	104917	310850	z	3,00				
5	ovas	45749	90480	z	2,00				
6	kukuruz	1220174	7085366	z	5,80				
7	šećerna repa	64326	3101176	ib	48,20				
8	stocna repa	5368	64677	kb	12,00				

Izvor: Statistički godišnjak Srbije (2006), Republički zavod za statistiku Srbije, Beograd

Dimenzije datoteke podataka određene su brojem promenljivih i brojem opservacija, tako da SPSS datoteka uvek ima pravougaoni oblik. Unošenje podataka u ćeliju van trenutne granice aktivne datoteke podataka automatski proširuje datoteku kako bi ona očuvala pravougaonost.

Početak rada podrazumeva i jasno definisanje varijabli koje se vrši iz glavnog menija preko komande **Data** i komande iz njenog podsistema **Define Variables**. Za svaku promenljivu se određuje ime, tip, nedostajući podaci, oznaka vrednosti, kao i format kolone na koju se ona odnosi.

### Procedure za sumarno prikazivanje podataka

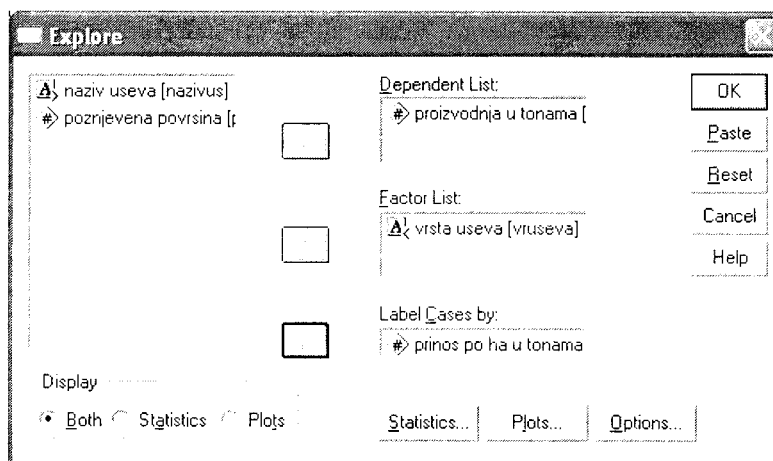
Statističke metode istraživanja masovnih pojava mogu se podeliti u dve osnovne grupe. Jedna obuhvata metode prikupljanja, sređivanja i prikazivanja podataka i metode određivanja parametara skupova. Ona spada u domen *deskriptivne statistike*.

Drugu grupu sačinjavaju metode statističke analize, čiji je osnovni zadatak objašnjenje varijabiliteta pomoću klasifikacionih, korelacionih i drugih statističkih pokazatelja, kao i statističko zaključivanje na osnovu uzorka. Ovim metodama se bavi *analitička statistika* koja se ne može strogo razdvojiti od deskriptivne statistike.

## Procedura Explore

Procedura *Explore* se često definiše kao preliminarna statistička analiza, jer pruža niz važnih informacija o raspodeli podataka, robustnim ocenama lokacije i testovima ispunjenosti pretpostavki osnovnih statističkih modela analize.

Pozivanjem procedure **Statistics-Summarize-Explore** otvara nam se komunikacioni prozor prikazan na slici 2. U listu zavisnih promenljivih (engl. *Dependent list*) potrebno je uneti najmanje jednu numeričku promenljivu i to tako što se iz izvorne liste promenljivih vrši premeštanje onih za koje se želi sprovođenje ove procedure. U primeru je odabrana numerička promenljiva *proizvodnja u tonama*. U listu faktora (engl. *Factor list*), iz izvorne liste se prenose promenljive koje će formirati grupe opservacija (npr. *vrsta useva*), a u listu označavanja opservacija (engl. *Label Cases by*) unosi se promenljiva čiji će modaliteti obeležiti nestandardne i ekstremne opservacije. Faktorska promenljiva *vrsta useva* ima 4 modaliteta, tako da će se sve opservacije zavisnih promenljivih grupisati u 4 grupe, dok će nestandardni i ekstremni slučajevi biti označeni vrednostima promenljive *prinos po ha u tonama*.



Slika 2. Komunikacioni prozor procedure *Explore*

Kao izlazni rezultati ove procedure javljaju se statistike i dijagrami. Izbor statističkih mera se ostvaruje preko opcije **Statistics**. U obrađenom primeru prvo je odabrana opcija **Descriptives** u okviru koje je izvršeno izračunavanje aritmetičke sredine, intervala poverenja za aritmetičku sredinu, medijane, varijanse, standardne devijacije, minimalne i maksimalne vrednosti, intervala varijacije, interkvartilne razlike, mera asimetričnosti i spljoštenosti. Sva izračunavanja su izvršena za odabranu zavisnu promenljivu *proizvodnja u tonama* za svaku vrstu useva posebno.

Druga odabrana opcija **M-estimators** u izlazu daje robustne ocene maksimalne verodostojnosti za lokaciju. Pod robustnim ocenama se podrazumevaju one koje zavise od jednostavnih pretpostavki osnove distribucije i nisu previše osetljive na njihovo narušavanje. Suština je u određivanju „potkresane“ sredine (engl. *Trimmed Mean*) kojom se isključuju one vrednosti koje su veoma udaljene od centralnog dela. U obilju robustnih ocena, najčešće korišćene su: *Huber's*, *Tukey's*, *Hampel's* i *Andrew's* M-ocene koje se, u suštini, razlikuju u načinu formiranja pondera.

Treća opcija, **Outliers**, u izlazu daje 5 najvećih i 5 najmanjih vrednosti opservacija. Dešava se da je taj broj manji ukoliko zahtevani broj ekstremnih vrednosti prevazilazi broj unetih podataka.

Kao izlazni rezultati procedure *Explore* javljaju se i dijagrami, koji se biraju preko opcije **Plots**. Grafički prikaz podataka omogućen je kroz korišćenje *histograma*, *dijagrama stabljike-i-listova* (engl. *Stem-and-Leaf Plot*) i *dijagrama pravougaonika* (engl. *Boxplot*).

*Histogram frekvencija* se sastoji iz niza pravougaonika. Osnovu svakog pravougaonika čini veličina grupnog intervala, a njegovu visinu odgovarajuća frekvencija intervala. Površina svakog pravougaonika je proporcionalna frekvenciji odgovarajućeg grupnog intervala, a ukupna površina svih pravougaonika histograma prikazuje ukupnu frekvenciju, odnosno ceo statistički skup. Histogram frekvencija se često

koristi za grafičko prikazivanje numeričkih serija u primenjenoj statistici i kao takav je sveprisutan u široj literaturi, pa će mu ovog puta biti posvećena manja pažnja.

*Dijagram stabljike-i-listova* daje detaljnije informacije od histograma. Svaka opservacija je prikazana numerički i vrednost svih opservacija se deli na vodeće cifre koje predstavljaju stabljiku (engl. *Stem*) i poslednje cifre koje predstavljaju listove (engl. *leaf*). Slika 3. prikazuje jedan od dijagrama iz ove grupe.

proizvodnja u tonama Stem-and-Leaf Plot for

VRUSEVA= industrijsko bilje

Frequency Stem & Leaf

3,00 0 . 011

.00 1 .

.00 2 .

2,00 3 . 56

1,00 Extremes (>=3101176)

Stem width: 100000

Each leaf: 1 case(s)

**Slika 3. Dijagram stabljike i listova**

Na slici, prvi red prikazuje opservacije čije numeričke vrednosti počinju cifrom 0 (širina stabljike je 100 000), s tim što je druga cifra 0 ili 1 i njihova frekvencija je 3. Konkretno, proizvedena količina uljane repice, koja pripada kategoriji industrijskog bilja, iznosi 3333 t, proizvedena količina duvana je 11336 t i proizvedena količina industrijske paprike je 12004 t. Listovi drugog i trećeg reda govore o tome da nema opservacija čije numeričke vrednosti počinju ciframa 1 i 2, a dok četvrti red prikazuje opservacije čija je vodeća cifra 3, a sledeće cifre se 5 ili 6. U datom primeru postoje 2 takve opservacije. Peti, poslednji red prikazuje ekstremnu vrednost čija je frekvencija 1. Njena numerička vrednost data je u zagradi.

*Dijagram pravougaonika* nam, za razliku od histograma i dijagrama stabljike-i-listova koji grafički prikazuju raspored individualnih opservacija, prezentuje raspored sumarnih statističkih mera. Na ovom dijagramu, jasno se mogu očitati medijana, prvi i treći kvartil, interkvartilna razlika i nestandardne i ekstremne opservacije.

### **Procedure za poređenje sredina**

Postoji veliki broj situacija u kojima se nastoji da se izvedu zaključci o razlikama ili jednakostima parametara između skupova. Na primer, ukoliko se želi da se uporede rezultati dva različita proizvodna procesa ili ako se teži spoznaji da li određena marketing strategija doprinosi boljoj prodaji nego druga. Ova i slična pitanja efikasno se rešavaju preko odgovarajućih procedura o testiranju hipoteza. Pored testiranja hipoteze o vrednosti jednog parametra skupa, oblast testiranja obuhvata i metode kojima poredimo dva skupa, tj. ispituje se jednakost ili razliku vrednosti istog parametra u skupovima.

### **Procedura Independent-Samples T Test**

Poređenje aritmetičkih sredina dva skupa u praksi je veoma često. T test se upravo i primenjuje za testiranje hipoteze o jednakosti aritmetičkih sredina i, kao takav, spada u red parametarskih statističkih testova. Ovaj test se primenjuje kada je nepoznata varijansa osnovnih skupova, tako da se ona ocenjuje na osnovu uzoračke varijanse.

Nulta hipoteza glasi:  $H_0: \mu_1 = \mu_2$

Nakon što se varijansa osnovnih skupova oceni varijansama uzorka, ove ocene se primenjuju za određivanje standardne greške. Pri tome se razlikuju dva slučaja: (a) varijanse skupova su među sobom jednake; (b) varijanse skupova su različite.

Ukoliko su varijanse jednake onda njihovu jedinstvenu vrednost  $\sigma^2$  ocenjujemo na osnovu varijansi dva slučajna uzorka koje smo izabrali iz dva osnovna skupa. Hipotezu o jednakosti aritmetičkih sredina dva skupa se testira primenom  $t$  testa, čija je statistika:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{Sp\sqrt{(1/N_1 + 1/N_2)}} \quad (1)$$

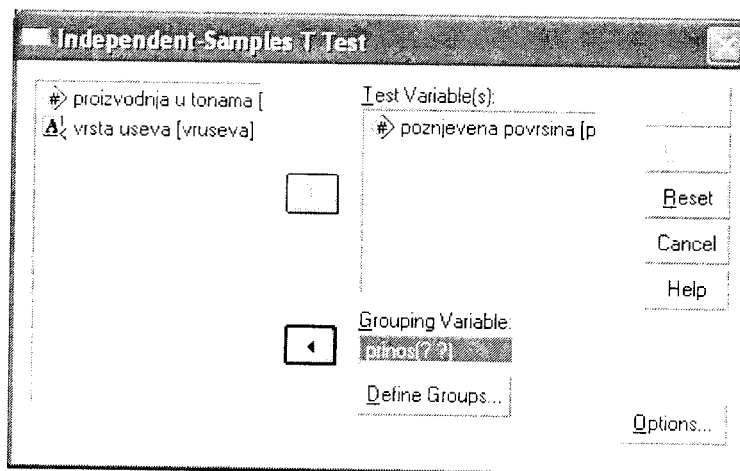
Pod pretpostavkom da je hipoteza  $H_0$  istinita, statistika testa (1) ima studentov raspored sa  $n_1+n_2-2$  stepeni slobode.

Ukoliko varijanse nisu jednake koristi se sledeća  $t$  statistika:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \quad (2)$$

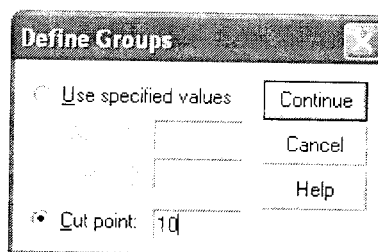
Za testiranje hipoteze o jednakosti varijansi koristi se *Levene test* koji je najmanje osetljiv na pretpostavku o normalnosti rasporeda osnovnih skupova. Primenom ovog testa se za svaku opeservaciju računa njeno apsolutno odstupanje od grupne aritmetičke sredine i zatim se primenjuje analiza varijanse na te razlike.

Pozivanjem procedure *Statistics-Compare Means-Independent Samples T-test* otvara se sledeći komunikacioni prozor:



Slika 4. Komunikacioni prozor procedure *Independent Samples T-test*

U listu promenljivih za testiranje (engl. *Test Variable(s)*), iz liste svih promenljivih se unosi jedna ili više numeričkih promenljivih (u primeru, *poznjevena površina*), dok se u listu promenljivih koje formiraju grupe (engl. *Grouping Variable*) unosi jedna numerička ili alfanumerička promenljiva. Za promenljivu koja formira grupe moraju se definisati dve grupe (slika 5).



### Slika 5. Opcija Define Groups

Ukoliko je promenljiva numerička, onda se upisuje jedna numerička vrednost za tu promenljivu koja deli uzorak na dva dela. To je tačka preseka (engl. *Cut Point*). U primeru, promenljiva *prinos po ha u tonama* definiše grupe, pa se kao tačka preseka navodi vrednost 10. Na taj način su opservacije u uzorku podeljene na grupu čiji je prinos jednak ili veći od 10 i grupu čija je prinos manji od 10 t/ha. Ako je promenljiva alfanumerička navode se njene dve vrednosti, a ne nazivi vrednosti.

Ovim primerom se testira hipoteza o jednakosti aritmetičkih sredina promenljive *požnjevena površina* za dva skupa. U izlaznim rezultatima korišćene procedure, nivo signifikantnosti od 0,240 ( $>0,05$ ) za Levene test ukazuje da treba prihvatiti nultu hipotezu o jednakosti varijansi dva skupa, tako da će se u daljoj analizi koristiti rezultati koji pretpostavljaju jednakost varijansi. Realizovani nivo značajnosti *t* testa od 0,474 ( $>0,05$ ) ukazuje da ne postoji značajna razlika između aritmetičkih sredina dva skupa.

Statistički softver SPSS for Windows nudi i mnoštvo drugih, isto tako korisnih procedura, koje korisniku daju mogućnost da vešto i lako manipuliše podacima i, na taj način, dolazi do željenih rezultata i saznanja. SPSS for Windows je modularni, čvrsto integrisani paket za proces analize – planiranje, prikupljanje podataka, pristup podacima, pripremu i upravljanje podacima, analizu podataka, izveštavanje i razvoj. Intenzivne promene savremenih tokova poslovanja sve više od zaposlenih zahtevaju posedovanje praktičnih znanja i veština kako bi bili uspešni u oblastima svog delovanja. Istovremeno, i informaciona tehnologija je postala jedna od osnovnih komponenti poslovanja, alat za planiranje, kreiranje i implementaciju različitih poslovnih programa.

### 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.

Statistički softver SPSS for Windows znatno povećava produktivnost jednog analitičkog procesa, jer se njegovom primenom značajno smanjuje obim rada i olakšava se prolazak kroz sve etape procesa analize korišćenjem jedinstvenog softverskog rešenja. Prezentovanjem svoje moći na jednom mestu, lako se zaključuje da se rad sa velikim brojem podataka i ne bi mogao zamisliti bez ovog ili nekog sličnog softvera.

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.

### Literatura

- [1] Hanić H., Krsmanović S., (1999), *Upravljački informacioni sistemi*, Beograd, Ekonomski fakultet Beograd
- [2] Newbold P., Carlson W.L., Thorne B., *Statistics for business and economics*, New Jersey, Prentice Hall
- [3] Petrović Lj., (2006), *Teorijska statistika*, Beograd, Ekonomski fakultet Beograd
- [4] Soldić-Aleksić J., (2004), *Aplikativni softver za statističku analizu i tabelarna izračunavanja*, Beograd, Ekonomski fakultet Beograd
- [5] Žižić M., Lovrić M., Pavličić D., (1999), *Metodi statističke analize*, Beograd, Ekonomski fakultet Beograd
- [6] <http://www.spss.com>