



Procedura One - Sample t - test

Sofija Čarkilović 12/2011

Sonja Jovanović 122/2011

Matematički fakultet, Beograd
maj, 2015.

Uvod

Postoji nekoliko statističkih testova koji koriste t – raspodelu, koji se jednim imenom mogu nazvati t - testovi.

Jedan od njih je Studentov t - test za jedan uzorak (One Sample t - test), nazvan po "studentu", pseudonimom koji je William Gosset koristio da sakrije svoj posao u Ginisovoj pivari u ranim 1900-im. (Oni su imali pravilo da njihovi zaposleni ne smeju da budu objavljeni, a sam Ginis nije želeo da ostali zaposleni znaju da oni prave izuzetak za Gosseta).

Ginis je regrutovao studente sa najboljim ocenama sa Oxforda i Kembridža da se prijave za biohemiju i statistiku u njegovim industrijskim procesima.

Studentov t - test

Studentov t - test je najčešće upotrebljavan parametarski test značajnosti za testiranje nulte hipoteze.

One - Sample t - test (t - test za jedan uzorak) se koristi kada želimo da znamo da li naš uzorak dolazi iz određene populacije, ali nemamo potpunu informaciju o toj populaciji.

On testira da li uzorak dolazi iz populacije sa specifičnom srednjom vrednošću. Zbog toga se ovaj test naziva još i t – test razlike srednje vrednosti osnovnog skupa i uzorka.

Studentov t - test

Računamo test statistiku t_s koristeći formulu: $t_s = \frac{\bar{x} - \mu_0}{s/\sqrt{n-1}}$

Testiramo nultu hipotezu:

Ho: srednja vrednost populacije jednaka određenoj vrednosti μ_0 ,
tj. između posmatrane i određene vrednosti ne postoji druga razlika do one koja
može da se pripiše slučajnim greškama

\bar{x} je uzoračka sredina

s je standardna devijacija uzorka

n je veličina uzorka

Hipoteze

Statističke hipoteze za One – Sample t-test uzimaju jedan od sledećih oblika, u zavisnosti od toga da li su naše hipoteze u istraživanju usmerene ili neusmerene.

U jednačinama desno μ_1 se odnosi na populaciju iz koje je uzorak; μ se menja sa stvarnom srednjom vrednošću populacije.

$$H_0 : \mu_1 = \mu$$

$$H_A : \mu_1 \neq \mu$$

$$H_0 : \mu_1 \leq \mu$$

$$H_A : \mu_1 > \mu$$

$$H_0 : \mu_1 \geq \mu$$

$$H_A : \mu_1 < \mu$$

Hipoteze

Ako je realizovana t - vrednost manja od granične tablične vrednosti za odgovarajući broj stepeni slobode i prag značajnosti, nulta hipoteza se prihvata kao tačna, a odbacuje alternativna hipoteza.

Obrnuto, ako je realizovana t - vrednost jednaka ili veća od granične tablične vrednosti za odgovarajući broj stepeni slobode i prag značajnosti, nulta hipoteza se odbacuje kao netačna, a prihvata se alternativna hipoteza.

Po default-u SPSS koristi 95% nivo poverenja, što odgovara pragu značajnosti manjem od 0.05.

Pretpostavke

Kada odlučimo da analiziramo podatke koristeći One - Sample t - test, deo procesa podrazumeva proveru da li se podaci koje želimo da analiziramo zaista mogu analizirati korišćenjem One - Sample t - testa.

Ovo je neophodno uraditi, jer je jedino prikladno koristiti One - Sample t - test, ako naši podaci zadovoljavaju četiri pretpostavke koje su potrebne kako bi one-sample t - test dao validan rezultat. U praksi, za proveru ove četiri pretpostavke je potrebno malo vremena, zahteva da kliknemo na još nekoliko dugmadi u SPSS pri obavljanju naše analize, ali nije težak zadatak.

Čak i ako naši podaci ne prođu određene pretpostavke, često postoji rešenje za njihovo prevazilaženje.

Pretpostavke

- **Pretpostavka #1:** Naša zavisna promenljiva treba da se sastoji od podataka koji su izmereni u intervalu ili nekom srazmernom nivou (tj., kontinuirano). Primeri promenljivih koje ispunjavaju ovaj kriterijum su, recimo, vreme revizije (dato u satima), inteligencija (merena pomoću IQ rezultata), test performansi (meren od 0 do 100), težina (izražena u kg), itd.
- **Pretpostavka #2:** Podaci su nezavisni (tj. nisu korelisani/povezani), što znači da ne postoji veza između opservacija. Ovo je više problem izučavanja dizajna, nego nešto što može da testira, ali je važna pretpostavka za One - Sample t - test.

Pretpostavke

- **Pretpostavka #3:** Promenljiva ne treba da ima neke značajne autlajere. Autlajeri (outliers) su tačke podataka u okviru naših podataka koje ne prate uobičajene šeme (na primer, u studiji od IQ bodova 100 učenika, gde je srednji rezultat 108 sa samo malim varijacijama među učenicima, jedan student je imao rezultat od 156, koji je veoma neuobičajen, pa se čak može staviti u vrh 1% IQ globalnih rezultata). Problem sa autlajerima je da mogu imati negativan efekat na One - Sample t - test, smanjujući tačnost naših rezultata. Srećom, kada se koristi SPSS za pokretanje One - Sample t - test na našim podacima, možemo lako detektovati moguće autlajere.

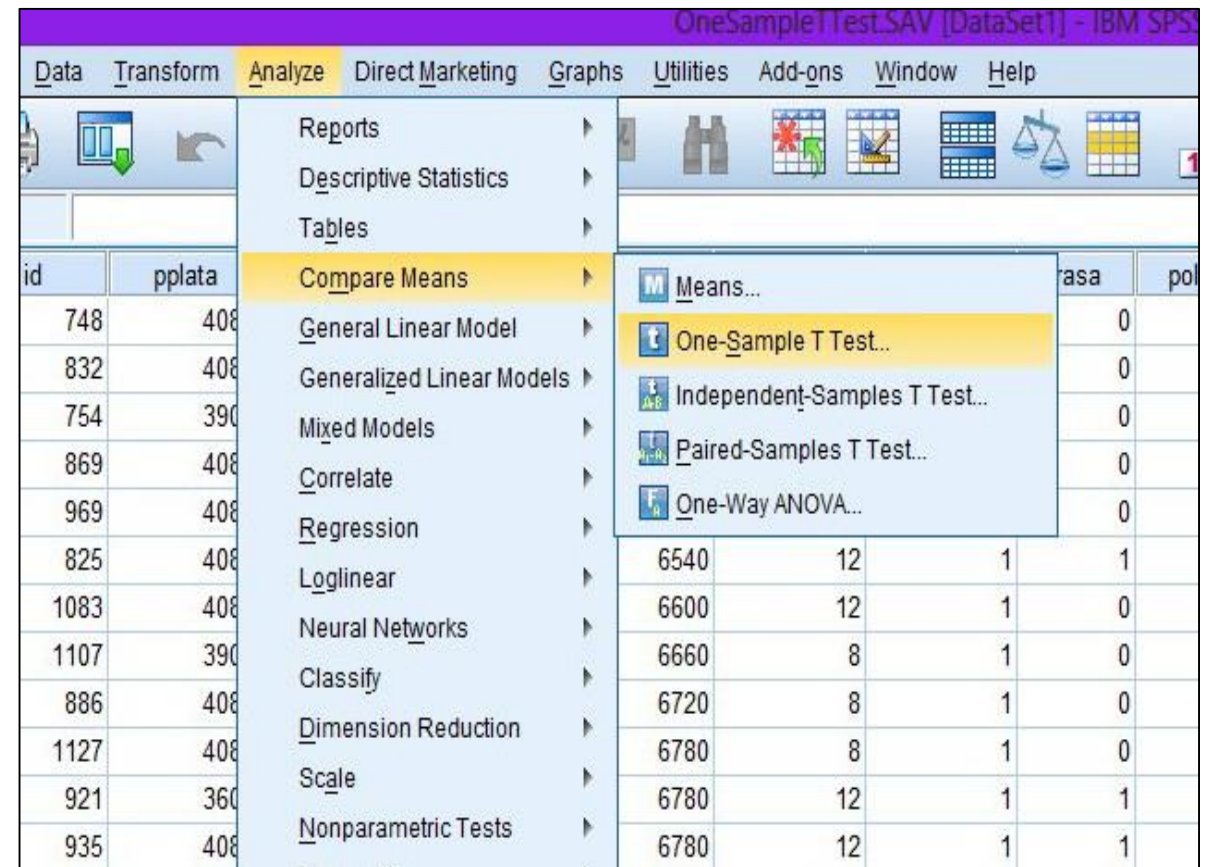
Pretpostavke

- **Pretpostavka #4:** Naša zavisna promenljiva treba da bude aproksimirana normalnom raspodelom. One - sample t -test zahteva samo normalno aproksimirane podatke, jer je prilično "robustan" za kršenje normalnosti, što znači da pretpostavka može biti malo narušena, a da i dalje daje valjane rezultate. Možemo testirati normalnost koristeći Shapiro - Wilk test normalnosti, koji se lako testira koristeći SPSS.

Primena t-testa u okviru SPSS-a

U SPSS-u do One Sample t-testa se može doći na sledeći način:

Analyze->Compare Means->One-Sample T Test...

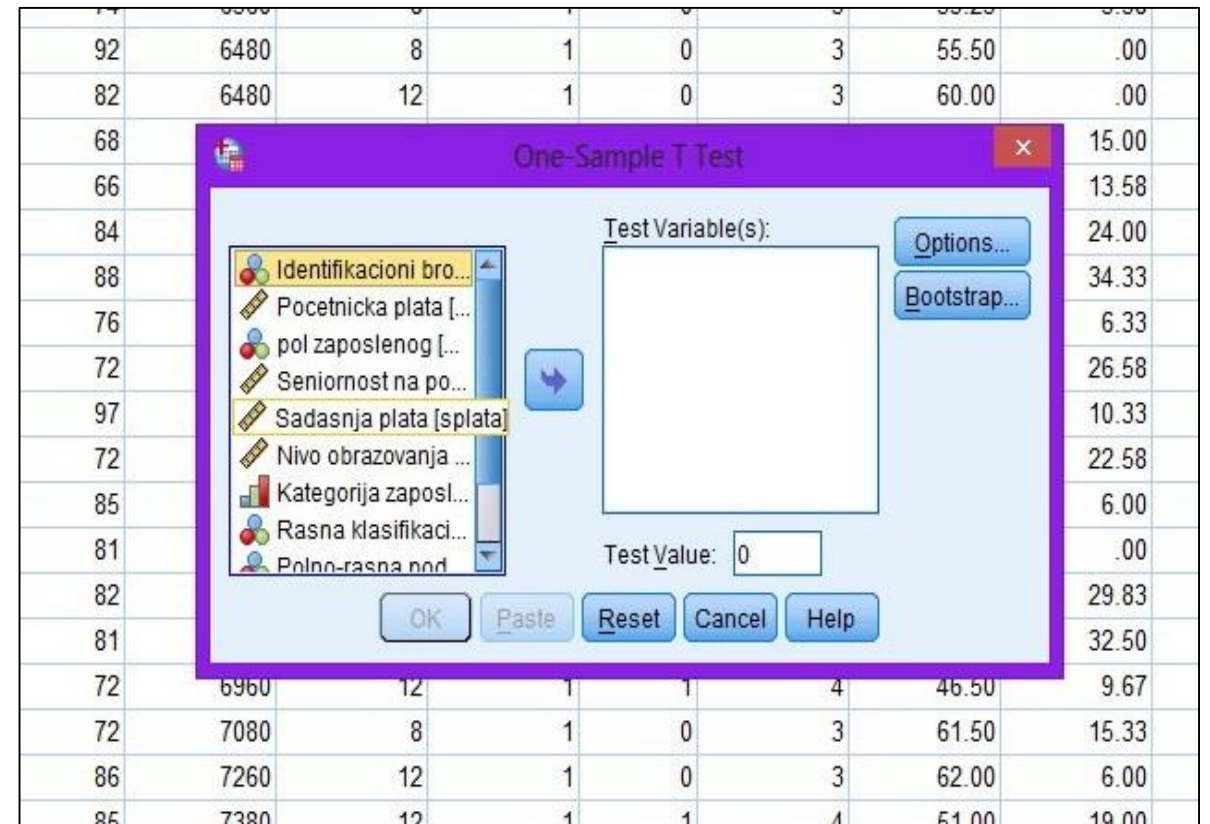


The screenshot shows the SPSS software interface with the 'Analyze' menu open. The 'Compare Means' option is selected, and the 'One-Sample T Test...' option is highlighted. The background shows a data table with columns 'id' and 'pplata'.

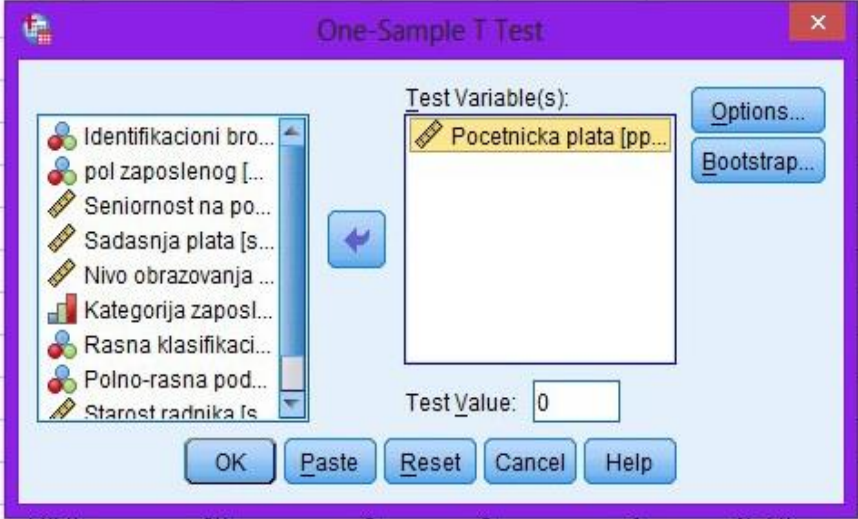
id	pplata
748	408
832	408
754	390
869	408
969	408
825	408
1083	408
1107	390
886	408
1127	408
921	360
935	408

Primena t-testa u okviru SPSS-a

Nakon pritiska na opciju **One Sample T Test**, pojaviće se sledeći prozor - dijalog:



Primena t-testa u okviru SPSS-a

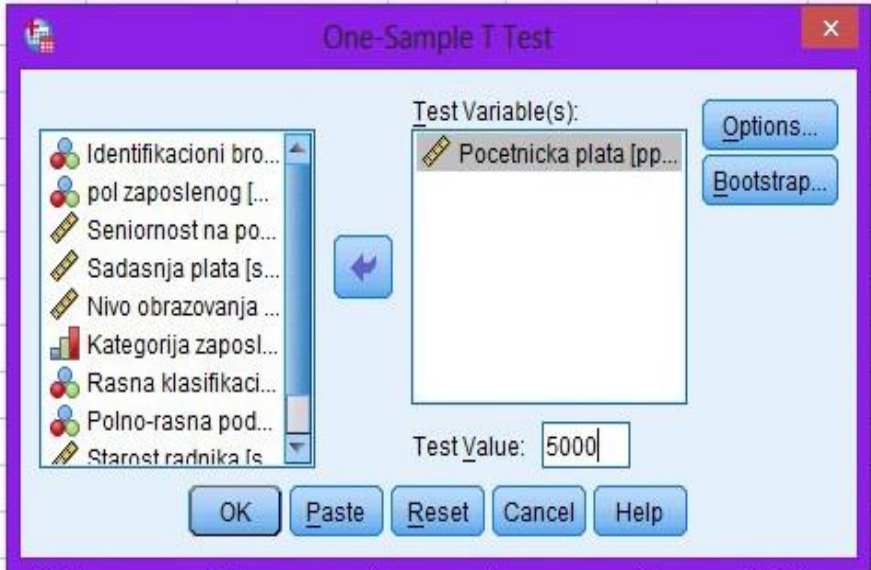


92	6480	8	1	0	3	55.50	.00
82	6480	12	1	0	3	60.00	.00
68						15.00	
66						13.58	
84						24.00	
88						34.33	
76						6.33	
72						26.58	
97						10.33	
72						22.58	
85						6.00	
81						.00	
82						29.83	
81						32.50	
72	6960	12	1	1	4	46.50	9.67
72	7080	8	1	0	3	61.50	15.33
86	7260	12	1	0	3	62.00	6.00
85	7380	12	1	1	4	51.00	19.00

U otvorenom prozor - dijalogu potrebno je odabrati odgovarajuću promenljivu (numeričku) koju želimo da testiramo klikom na nju samu, a zatim pritiskom na strelicu između kolona preneti je u **Test Variable(s)** box.

Primena t-testa u okviru SPSS-a

Ostaje još da se u **Test Value** box-u unese vrednost sa kojom želimo da uporedimo srednju vrednost odgovarajuće promenljive (u ovom primeru, to je vrednost 5000).



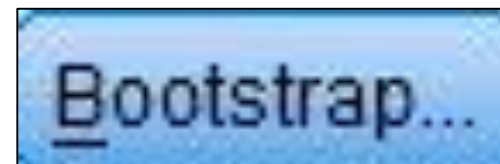
The screenshot shows the SPSS One-Sample T Test dialog box overlaid on a data spreadsheet. The dialog box has a purple title bar and contains the following elements:

- Test Variable(s):** A list box containing "Pocetnicka plata [pp...".
- Test Value:** A text input field containing the value "5000".
- Buttons:** "Options...", "Bootstrap...", "OK", "Paste", "Reset", "Cancel", and "Help".
- Variable List:** A scrollable list on the left showing variables such as "Identifikacioni bro...", "pol zaposlenog [...", "Seniornost na po...", "Sadasnja plata [s...", "Nivo obrazovanja ...", "Kategorija zaposl...", "Rasna klasifikaci...", "Polno-rasna pod...", and "Starost radnika [s".

The background spreadsheet shows data for various variables, including "Pocetnicka plata [pp..." with values ranging from 55.50 to 62.00.

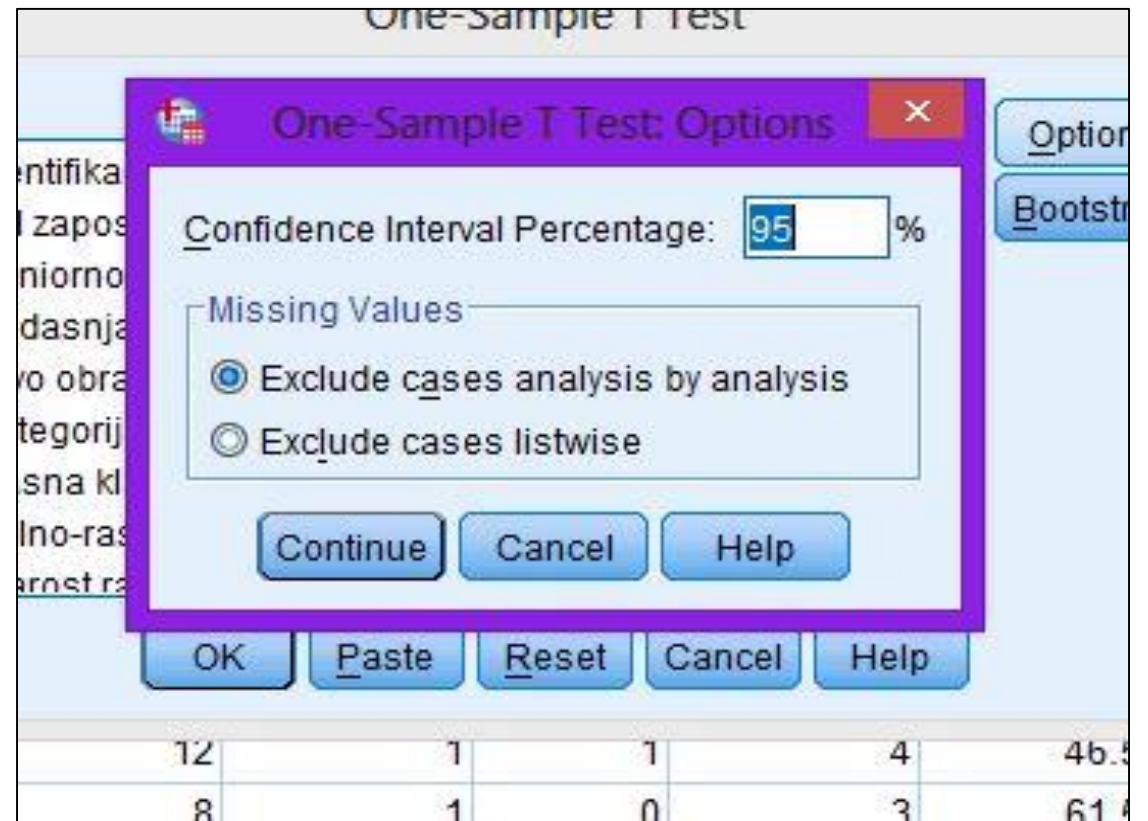
Primena t-testa u okviru SPSS-a

Pored **Test Variable(s)** i **Test Value** box-a, u otvorenom prozor - dijalogu nalaze se i dva dugmeta:



Primena t-testa u okviru SPSS-a

Pritiskom na dugme Options otvara se **One-Sample T Test: Options** prozor - dijalog prikazan na sledećoj slici:



Primena t-testa u okviru SPSS-a

U ovom prozor - dijalogu unosi se vrednost procenta intervala poverenja (vrednost od 1 do 99). Po default-u, SPSS koristi 95% intervale poverenja, tj. nivo značajnosti manji od 0.05.

Pored intervala poverenja, u ovom dijalog-prozoru nalazi se podešavanje vezano za nedostajuće vrednosti u okviru **Missing Values**.



Ako se testira više od jedne promenljive i pri tom baza podataka sadrži nedostajuće vrednosti, treba voditi računa o odabiru neke od sledećih opcija. Pogrešan izbor opcije može dovesti do toga da SPSS ukloni neke podatke koje ste želeli da uključite pri analizi.

Primena t-testa u okviru SPSS-a

Pritiskom na dugme Bootstrap otvara se prozor – dijalog u kome se čekiranjem **Perform bootstrapping** omogućuje primena Bootstrap metoda. Ostala podešavanja u ovom prozor – dijalogu vezana su za postupak uzorkovanja i intervale poverenja. Uzorak koji se koristi sa ocenjivanjem može biti prost (opcija **Simple**) ili stratifikovan (opcija **Stratified**). Broj uzoraka se određuje unošenjem vrednosti u **Number of samples** box. Takođe, moguće je izvršiti uzorkovanje pomoću Mersenne-Twister generatora, čekiranjem **Set seed for Mersenne Twister**, nakog čega se u **Seed** box unosi početna vrednost.

Primena t-testa u okviru SPSS-a

Na kraju, klikom na dugme Ok, na izabranu promenljivu se primenjuje One Sample t-test, pri čemu se na prozoru Viewer prikazuju sledeći rezultati:

SPSS generiše dve tabele kao izlazne podatke, i one sadrže sve informacije koje su neophodne za tumačenje rezultata primene One Sample t-testa.

```
T-TEST
  /TESTVAL=5000
  /MISSING=ANALYSIS
  /VARIABLES=pplata
  /CRITERIA=CI (.95).
```

T-Test

[DataSet1] D:\Fakultet\4. godina\SS4\Sonja Sofija\OneSampleTTest.SAV

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Pocetnicka plata	474	6806.43	3148.255	144.604

One-Sample Test

	Test Value = 5000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Pocetnicka plata	12.492	473	.000	1806.435	1522.29	2090.58

Primena t-testa u okviru SPSS-a

Prva tabela, tj. tabela **One - Sample Statistics**, sadrži deskriptivne statistike promenljive koju smo testirali. U koloni **N** se nalazi broj opservacija testirane promenljive, a kolone **Mean**, **Std. Deviation**, **Std. Error Mean** sadrže redom srednju vrednost, standardnu devijaciju i standardnu grešku srednje vrednosti (količnik srednje vrednosti i kvadrata broja opservacija) testirane promenljive.

	N	Mean	Std. Deviation	Std. Error Mean
Pocetnicka plata	474	6806.43	3148.255	144.604

Primena t-testa u okviru SPSS-a

Druga tabela, tj. tabela **One - Sample Test**, sadrži najbitnije informacije o rezultatu primenjenog testa. U zaglavlju tabele nalazi se poznata ili pretpostavljena vrednost srednje vrednosti za koju testiramo datu promenljivu (**Test Value**), tj. vrednost koju smo unosili u prvom prozor – dijalogu (u ovom primeru to je 5000). Dalje, u prve tri kolone ispod zaglavlja Test Value nalaze se informacije o tome da li je uzorak iz populacije sa srednjom vrednošću koja je u zaglavlju tabele (u ovom primeru to je 5000) ili, drugačije formulisano, da li su te dve srednje vrednosti statistički značajno različite. U prvoj koloni **t** nalazi se t-vrednost (vrednost t-statistike) koja se dobija na osnovu već spomenutog izraza. Naredna kolona **df** sadrži vrednost koja predstavlja broj stepeni slobode. Nakon nje, kolona **Sig.(2-tailed)** sadrži vrednost nivoa dvostrane značajnosti. Iako smo u postavljenoj hipotezi naveli da nam treba jednostrani interval (dobili smo dvostrani interval), ne postoji opcija u okviru SPSS-a za to podešavanje, tako da smatrajte da je reč o jednostranom intervalu.

Primena t-testa u okviru SPSS-a

U ovom našem slučaju u Sig.(2-tailed) koloni dobili smo vrednost 0 (napomena: nije u pitanju tačno vrednost 0, već SPSS zaokružuje sve vrednosti manje od 0.005 na 0) što je sigurno manje od pretpostavljenog nivoa značajnosti (0.05), pa je hipoteza o jednakosti srednjih vrednosti odbačena. Drugi način za testiranje hipoteze jeste (konkretno u ovom primeru) da iz tabele za studentovu raspodelu sa 473 stepeni slobode pročita vrednost za nivo značanosti 0.05 jedностранog intervala, pa da se ta vrednost uporedi sa t - vrednošću iz tabele. Ukoliko je t - vrednost manja od kritične vrednosti, onda se prihvata nulta hipoteza o jednakosti srednjih vrednosti.

Primena t-testa u okviru SPSS-a

Na kraju, poslednje kolone sadrže informacije o razlikama između pretpostavljene i stvarne srednje vrednosti promenljive. **Mean Difference** sadrži vrednost razlike između stvarne i pretpostavljene srednje vrednosti promenljive, a naredne dve kolone (**Lower** i **Upper** u okviru 95% Confidence Interval of the Difference) predstavljaju gornju i donju granicu 95% intervala poverenja u kom se nalazi data razlika.

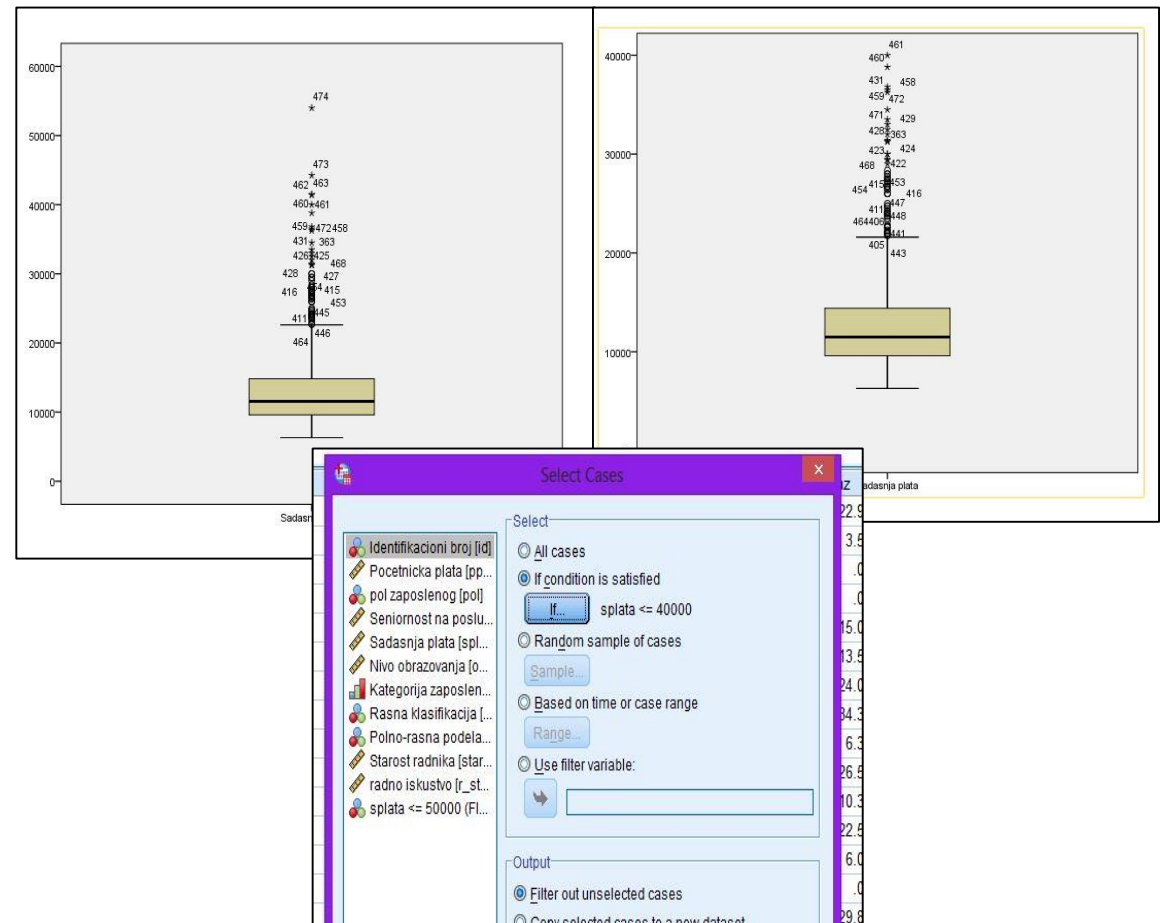
PRIMER 1.

U datoteci ***OneSampleTTest.sav*** na promenljivu *splata* primeniti One Sample t-test, ukoliko je to moguće.

PRIMER 1.

Provera pretpostavke #3:

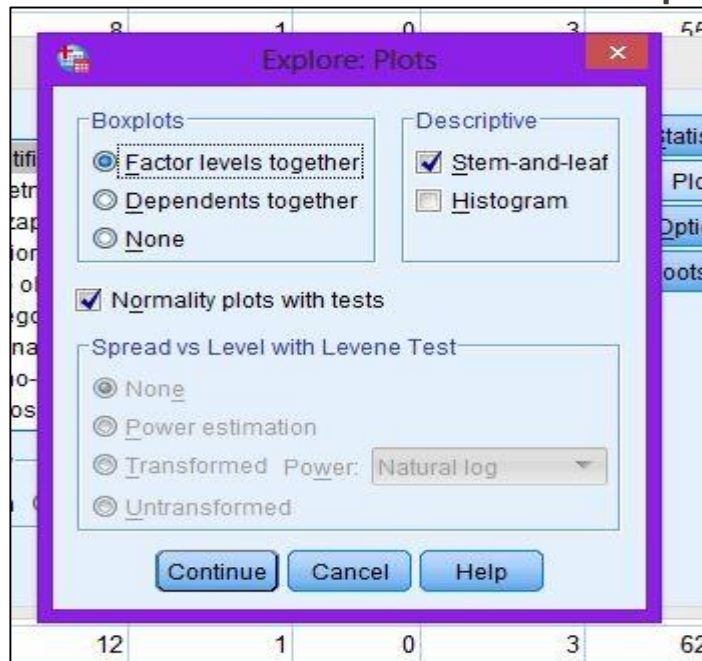
Na formiranom boxplotu uočava se postojanje autlajera. Uklonićemo neke od ekstreminskih vrednosti iz promenljive.



PRIMER 1.

Provera pretpostavke #4:

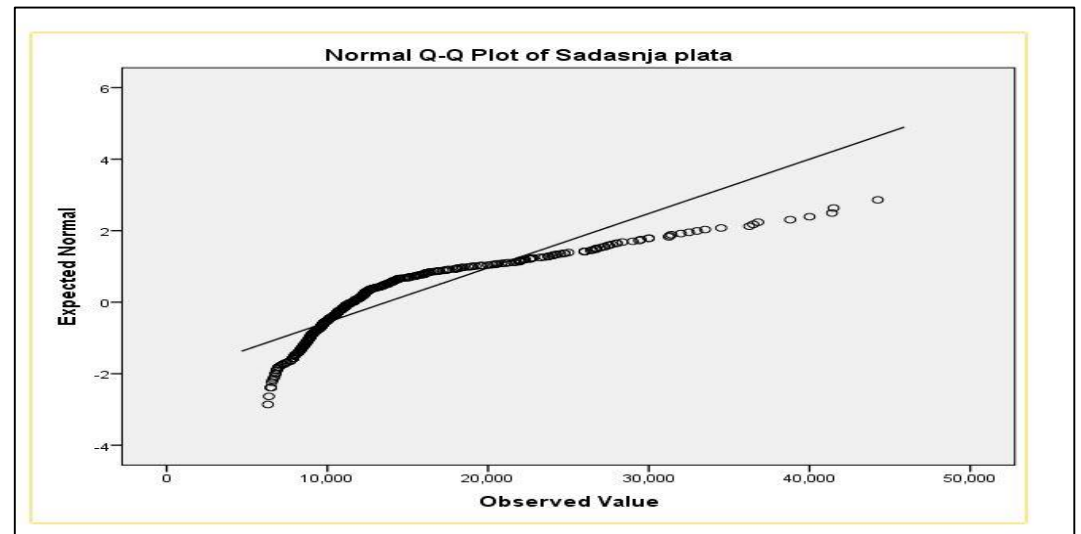
Testiranje normalnosti koristeći Shapiro -
Wilk test normalnosti i Q-Q plot:



U prozoru izlaznih podataka dobija se:

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Sadasnja plata	.202	470	.000	.800	470	.000

a. Lilliefors Significance Correction.



PRIMER 1.

Sig. vrednost iz Shapiro - Wilk testa normalnosti je manja od 0.05 pa možemo zaključiti da data promenljiva nije normalno raspodeljena. Takođe, posmatranjem Q-Q plot, uočava se da linija koja je sastavljena od vrednosti iz date promenljive ne nalazi blizu linije normalne raspodele, pa i grafičkim putem zaključujemo da promenljiva nije normalno raspoređena. Dakle, na ovu promenljivu ne možemo primeniti One Sample T – test.

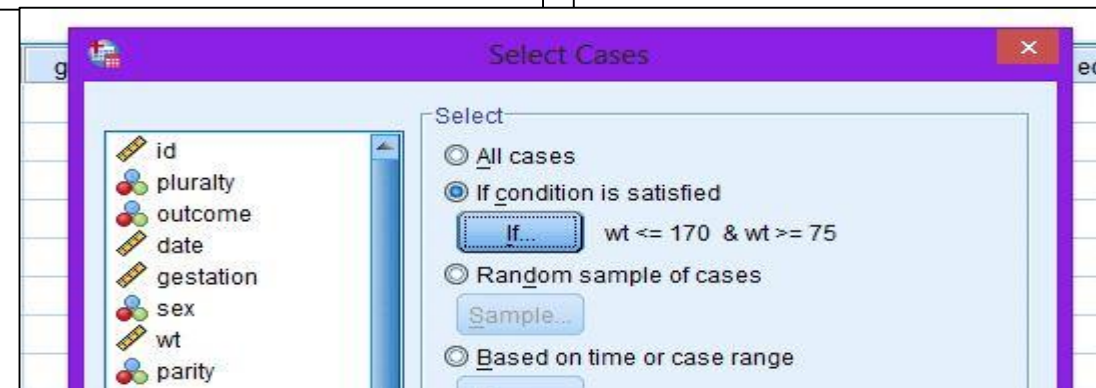
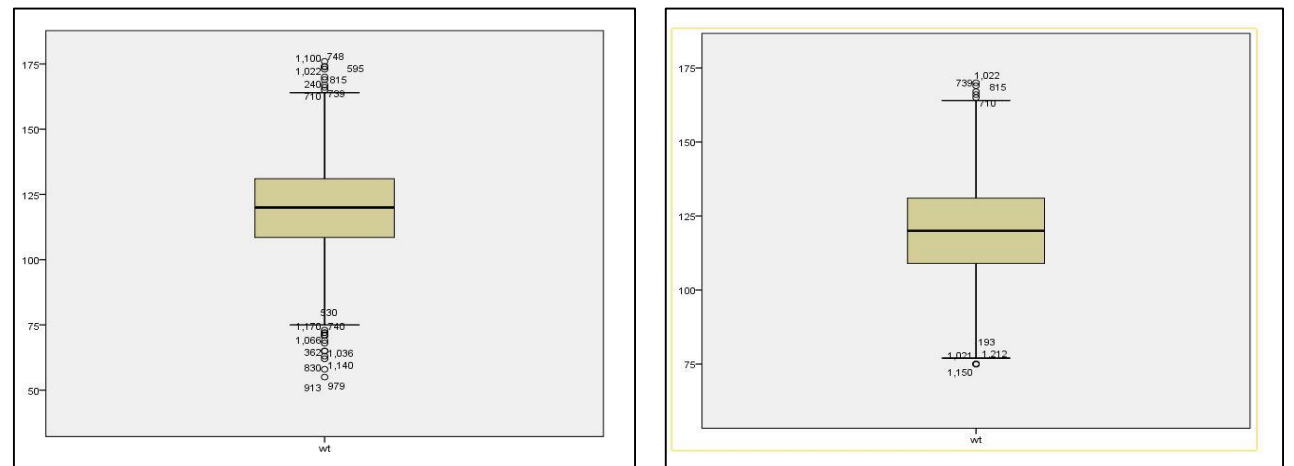
PRIMER 2.

U datoteci ***babes.sav*** na promenljivu *wt* primeniti One Sample t-test, ukoliko je to moguće.

PRIMER 2.

Provera pretpostavke #3 *wt*:

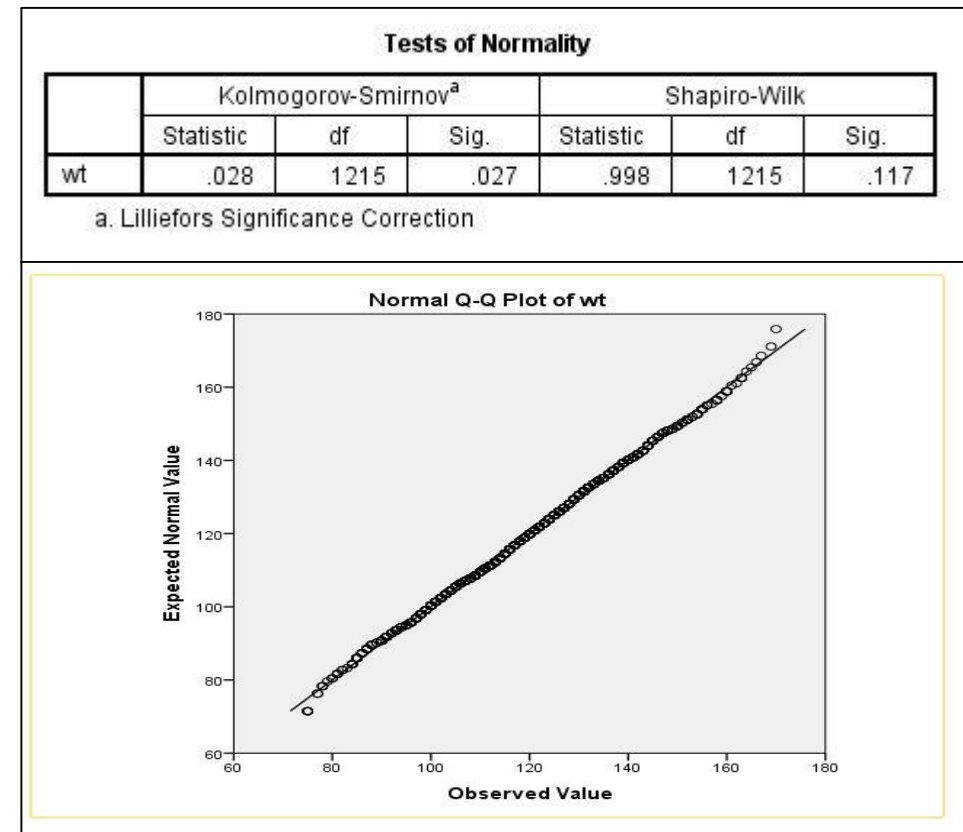
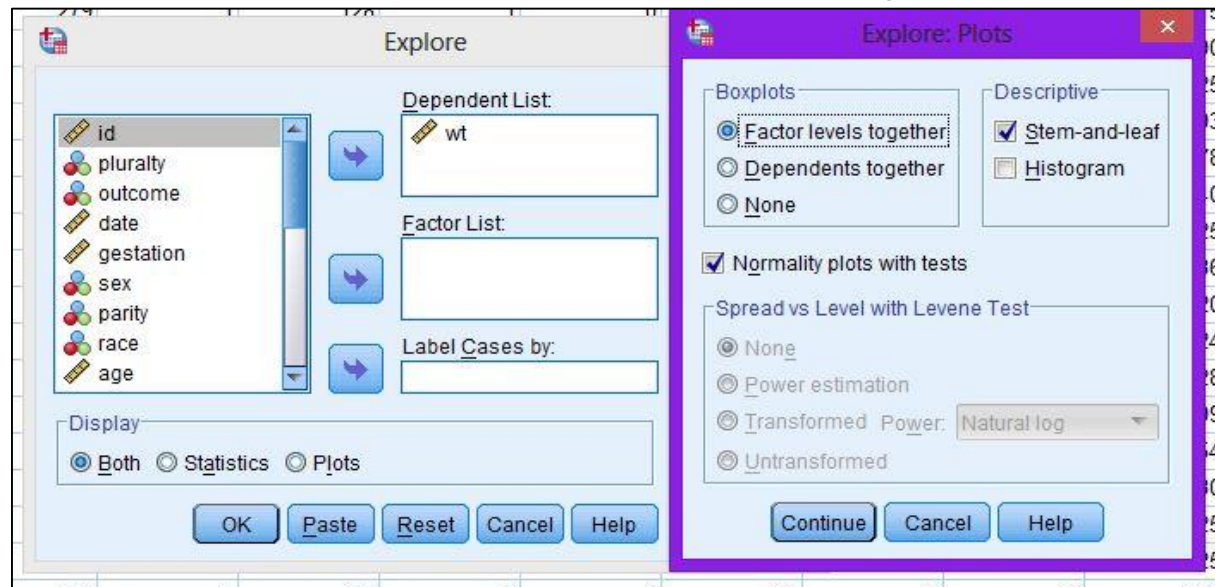
Na formiranom boxplotu uočava se postojanje autlajera. Uklonićemo neke od ekstreminskih vrednosti iz promenljive *wt*. Na osnovu tabela na izlaznom prozoru može se videti da je eliminisano 20 vrednosti.



PRIMER 2.

Provera pretpostavke #4:

Testiranje normalnosti koristeći Shapiro -
Wilk test normalnosti i Q-Q plot:



PRIMER 2.

Sig. vrednost iz Shapiro - Wilk testa normalnosti je večja od 0.05 (0.117) pa možemo zaključiti da data promenljiva jeste normalno raspodeljena. Takođe, posmatranjem Q-Q plot, uočava se da linija koja je sastavljena od vrednosti iz date promenljive prati liniju normalne raspodele, pa i grafičkim putem zaključujemo da promenljiva jeste normalno raspoređena. Dakle, ova promenljiva zadovoljava uslove #3 i #4, pa možemo primeniti One Sample T – test.

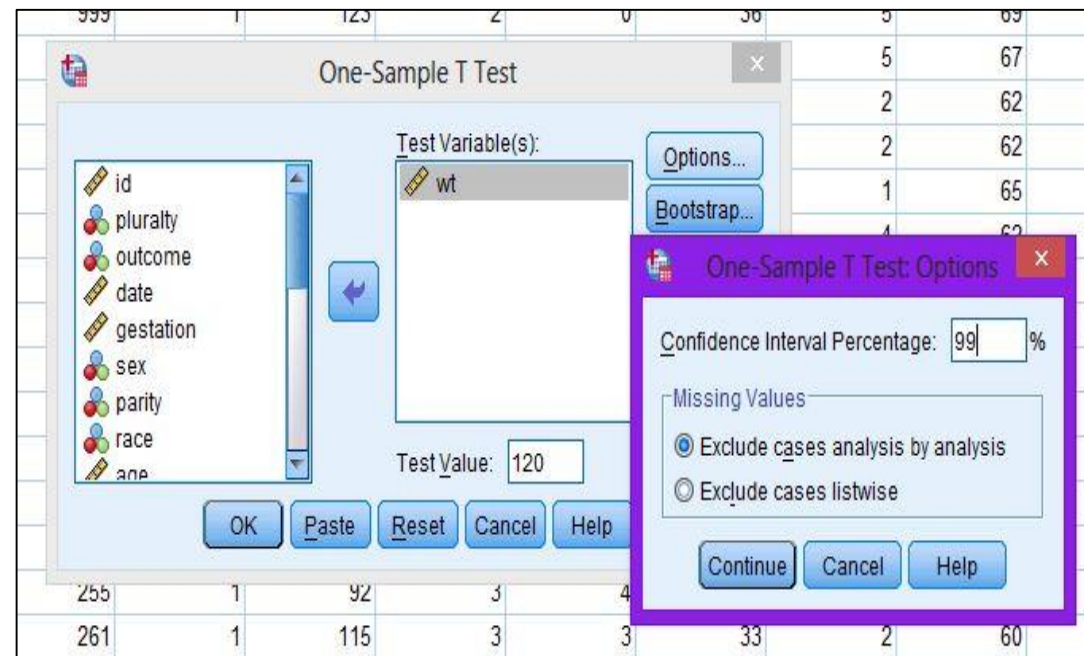
PRIMER 2.

Testiramo promenljivu *wt* (težina bebe):

Ho: srednja vrednost promenljive wt jednaka je 120

120-prosečna težina bebe u uncama

Koristićemo 99% interval poverenja, pa u Options treba da promenimo vrednost procenta intervala poverenja



PRIMER 2.

U Sig.(2-tailed) koloni dobili smo vrednost 0.934 što je veće od pretpostavljenog nivoa značajnosti (0.01), pa je hipoteza o jednakosti srednjih vrednosti prihvaćena.

Konačni zaključak: Prihvata se nulta hipoteza sa greškom manjom od 0.01 i sa sigurnošću većom od 99% tvrdimo da je srednja vrednost promenljive wt jednaka pretpostavljenoj, tj. 120 unci.

T-Test

[DataSet1] D:\Fakultet\4. godina\SS4\Sonja Sofija\babies.sav

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
wt	1215	120.04	17.013	.488

One-Sample Test

	Test Value = 120					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference	
					Lower	Upper
wt	.083	1214	.934	.040	-1.22	1.30



HVALA NA PAŽNJI!

