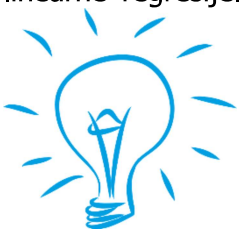
7. Statistička obrada podataka SPSS

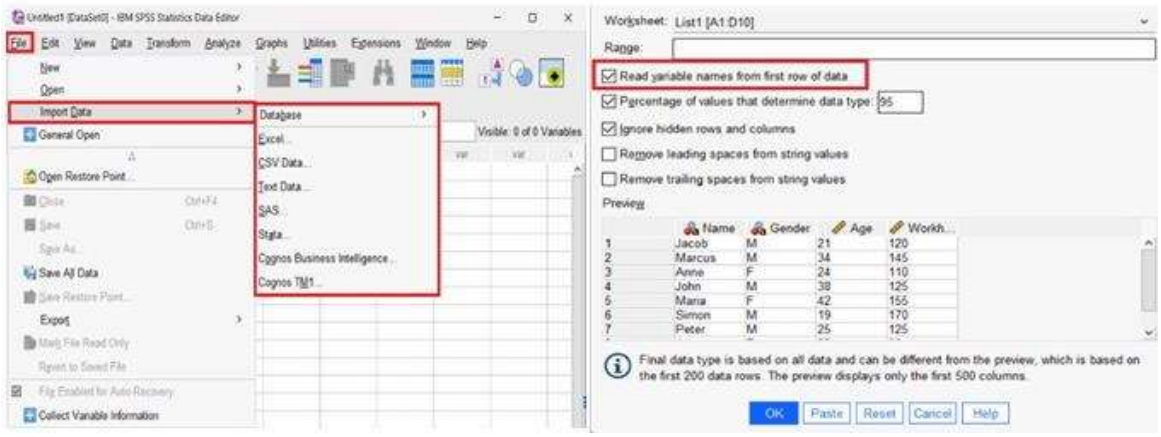
Do sada ste već stekli osnovno razumevanje statistike, manipulacije podacima, simulacije, modeliranja i analize unutar logističkih lanaca snabdevanja, zajedno s direktnim metodama linearne regresije. Dok statistika nudi širok raspon modela i tehnika za poboljšanje vaših napora optimizacije, sprovođenje analize i prepoznavanje potencijalnih poboljšanja, možda ste primetili da kako složenost analiziranih podataka i proračuna raste, tradicionalni pristupi mogu postati sve zamršeniji i izazovniji za izračunavanje. Kako zamršenost podataka i izračunavanja raste, konvencionalne metode mogu biti nedovoljne i, u nekim slučajevima, ugroziti pouzdanost rezultata. Kako bi se to prepoznalo, statistika koristi različite softverske programe koji automatizuju analizu i interpretaciju prikupljenih podataka, a istovremeno pružaju mnoštvo modela i funkcija za osiguranje pouzdanih rezultata. Jedan takav softver je IBM-ov SPSS, koji će biti ključni alat u ovom poglavlju. U ovom poglavlju pružićemo sažeti uvod u primarnu upotrebu softvera SPSS, istražujući njegove funkcionalnosti i praktične primene. Nakon početnog uvoda sledi praktična primena programa kroz četiri bazna testa za izračunavanje rezultata: T-test, korelacije, Hi-kvadrat i ANOVA. Kako bismo vam olakšali učenje, predstavićemo jednostavne probleme i njihova rešenja kako biste se lakše upoznali s ovim testovima.



7.1 Osnove IBM-ovog SPSS-a

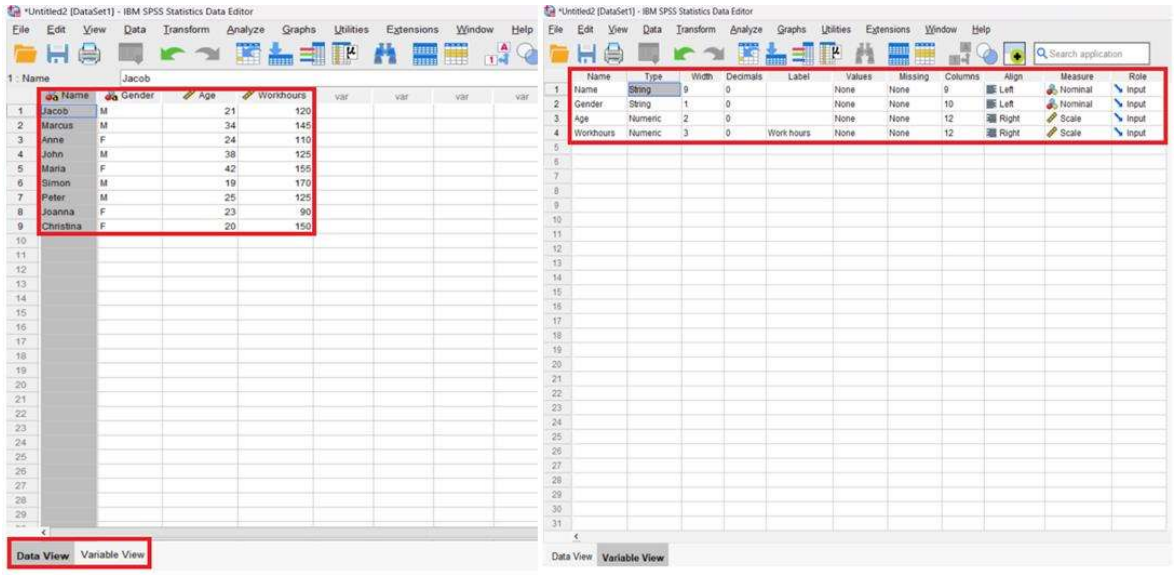
Možda ste već imali iskustva sa SPSS softverom. Međutim, ako vam ipak treba, dopustite da vam ponudimo kratki uvod. SPSS, poput svog opšte priznatijeg pandana Excela, olakšava manipulaciju podacima, analizu i vizualizaciju. Ipak, za razliku od Excela, koji ponekad može biti naporan i složen za programiranje funkcija, SPSS nudi korisnički interfejs za statističku analizu (IBM, 2021.). Nudi niz funkcija i metodologija za učinkovito rukovanje vašim podacima. Dok se SPSS softver ističe u pružanju opsežnih mogućnosti statističke analize, upravljanje manipulacijom podacima i konfigurisanje početnih postavki za analizu ponekad može biti izazovno (IBM, 2021.). Stoga ćemo se pozabaviti osnovama uvoza podataka i pripreme podataka za naknadne statističke testove.

S obzirom na široku upotrebu Excela za rukovanje numeričkim podacima, vaši se podaci mogu dobiti ili pripremiti u proračunskoj tabeli programa Excel. Srećom, SPSS može uvesti podatke iz različitih formata datoteka u svoje proračunske tabele. Nakon što pripremite finalne Excel proračunske tabele, otvorite softver SPSS. Na početnom ekranu idite na ikonu "Datoteka" i odaberite "Uvezi podatke". U sedećem prozoru možete odabrati format podataka koji nameravate uvesti (pogledajte sliku 7.1). Sledeći ovaj korak, pronađite pripremljenu datoteku, odaberite je i pređite na sedeći prozor. Ovaj prozor će od vas tražiti da konfigurišete dodatne postavke. Ako ste već uključili nazive kolona u prvi red vaših podataka, odaberite opciju "Pročitaj nazive varijabli iz prvog reda podataka" (pogledajte sliku 7.1), a zatim kliknite "Završi" da bi se podaci pojavili u proračunskoj tabeli (IBM, 2021).



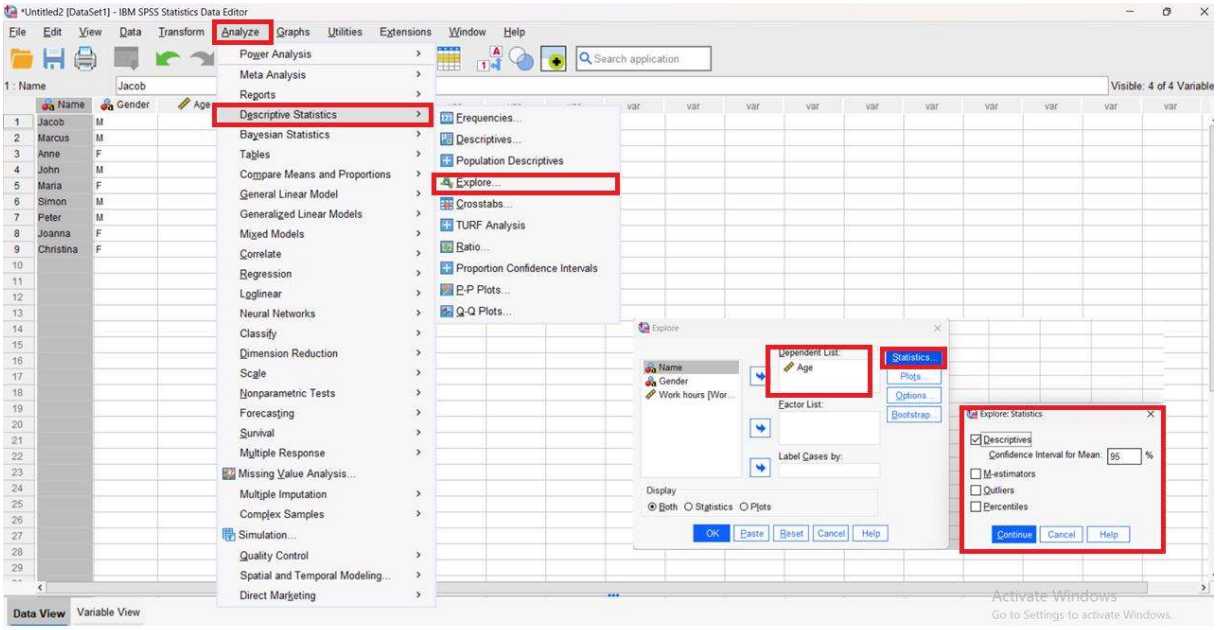
Slika 4.29SPSS postavke uvoza podataka.

Sada kada imamo podatke u našoj proračunskoj tabeli, primetićete jasnu razliku u prezentaciji u poređenju s Excelom. SPSS kategorizuje podatke u dve primarne vrste, svaka s dve dodatne podvrste. Kao što je prikazano na slici 7.2, podaci se mogu klasifikovati kao numerički ili kategorički. Numerički podaci sastoje se od brojeva i mogu se kategorizovati kao diskretni (s ograničenim opcijama) ili kontinuirani (nude beskonačne mogućnosti). S druge strane, kategorički podaci sastoje se od reči i mogu se dalje razlikovati kao redni (imaju hijerarhiju) ili nominalni (bez hijerarhije). Zavisno od prirode vaših podataka, možda ćete morati konfigurisati varijable kako bi se uskladile sa željenom analizom. U većini slučajeva, SPSS će automatski prikladno kategorizovati varijable. Pretpostavimo da želite izvršiti dalju manipulaciju tipovima podataka. U tom slučaju možete pristupiti opciji "Prikaz" i pod "Prikaz varijable" prilagoditi varijabilne informacije kao što su naziv, vrsta, širina, mera i više (IBM, 2021.).



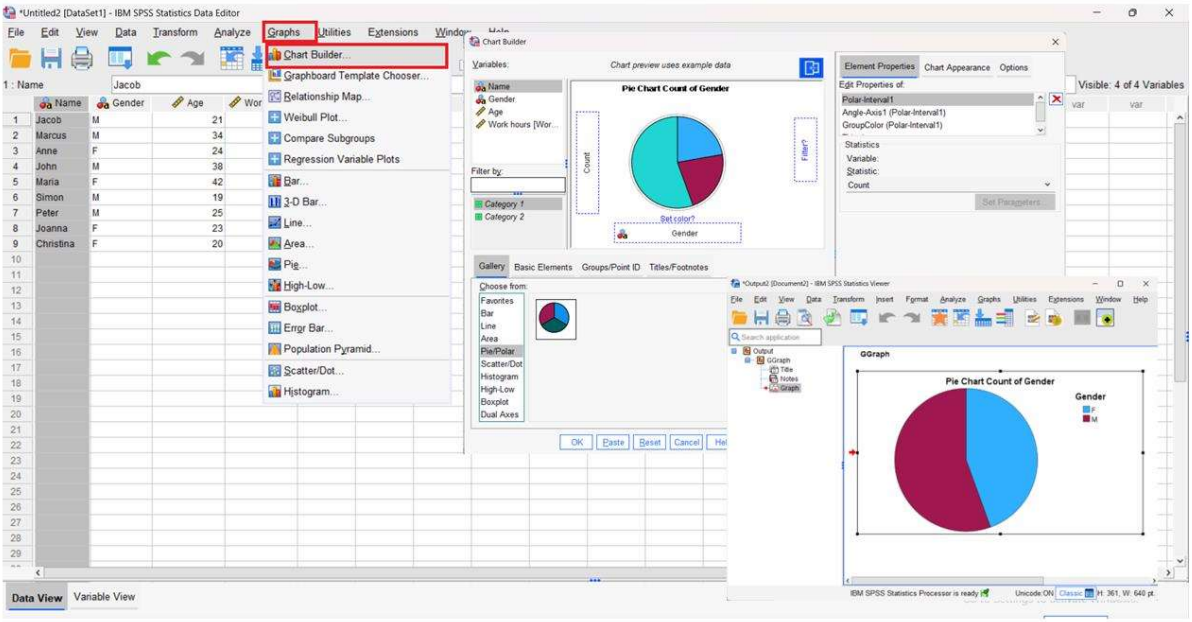
Slika 4.30Prozori za prikaz podataka i varijabli.

Nakon što ispravno postavite svoje podatke, možete ih istraživati unutar SPSS-a. SPSS omogućava korisnicima izvođenje detaljne statističke analize bez oslanjanja na unapred definisane funkcije. Na početnom ekranu (pogledajte sliku 7.3), idite na "Analiziraj", nakon čega sledi "Deskriptivna statistika", a zatim odaberite "Istraži". U odjeljku "Istraži" pronaći ćete različite opcije zavisno od karakteristika podataka koje ste uneli. U ovom načinu rada SPSS će vam pružiti informacije o "deskriptivnoj statistici" o vašim podacima. Iako je ovo važno za početnu analizu podataka, nudi samo osnovne uvide i ne ulazi u detaljniju statističku analizu, koja će biti obrađena u narednim poglavljima. Pre nego što nastavimo dalje, takođe ćemo istražiti još jednu funkciju u SPSS-u — vizualizaciju grafikona (IBM, 2021).



Slika 4.31Postavke deskriptivne statistike.

SPSS nudi niz opcija za vizualizaciju podataka, uključujući histograme, kutijaste dijagrame, stubičaste grafikone, dijagrame rasipanja, linijske grafikone, pita grafikone i još mnogo toga. Do ove tačke trebali biste imati osnovno razumevanje o tome što svaka vrsta grafikona predstavlja i kako tumačiti rezultate koje oni pružaju. Stoga ćemo se fokusirati na to kako izraditi te grafikone unutar softvera SPSS. Da biste izradili grafikone, odaberite karticu "Grafikoni" na početnom ekranu, nakon čega sledi "Izrada grafikona". U novom prozoru možete odabrati vrstu grafikona koju želite izraditi i odabrati varijable koje želite uključiti. Nakon izbora "Završi", pojavi će se novi prozor s rezultatima vizualiziranim u odabranom formatu grafikona. U ovom novom prozoru možete aktivno komunicirati s grafikonom, što vam omogućava izmenu varijabilnih boja i fontova, istraživanje distribucija varijabli na grafikonu, i više (IBM, 2021).



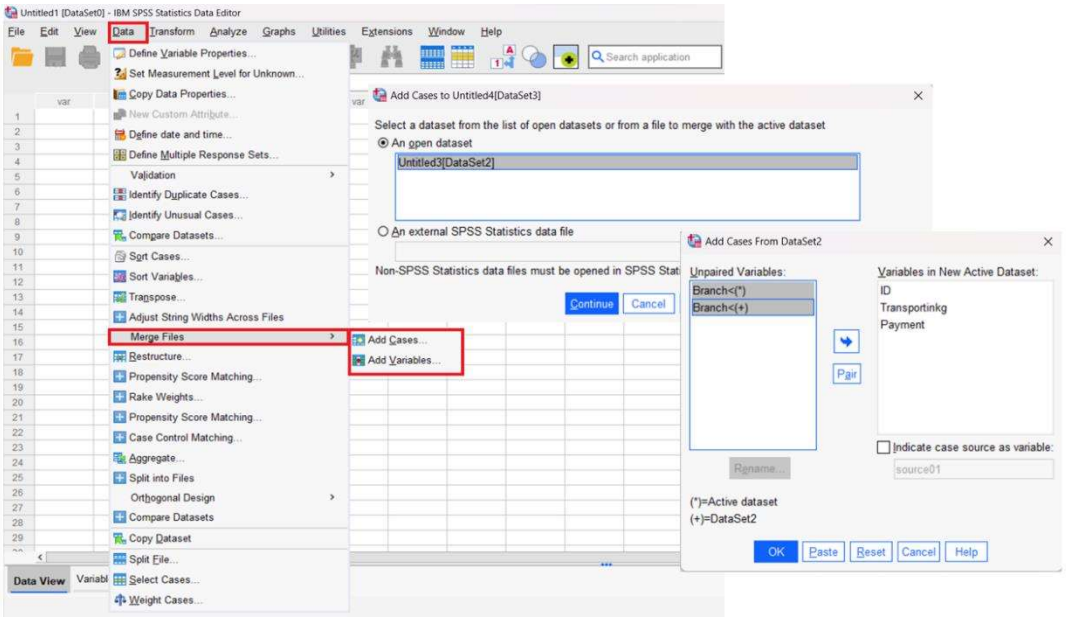
Slika 4.32Postavke izrade grafikona u SPSS-u.

Do ove tačke pokrili smo tri od četiri pravila za "Istraživanje podataka", koja uključuju posmatranje podataka (istraživanje neobrađenih podataka), identifikaciju podataka (određivanje tipova podataka) i, u određenoj meri, grafičko prikazivanje i opisivanje podataka putem deskriptivne statistike i izrade grafikona. Poslednje pravilo je "Formulacija pitanja", gde se pitamo što želimo postići analizom podataka i u skladu s tim postavljamo grafikone i deskriptivnu statistiku kako bismo dobili odgovore na naša specifična pitanja. Na primer, u našem posmatranom primeru, pitanje bi moglo biti: "Da li je naša analizirana populacija pretežno ženska?" Korišćenjem i grafikona i deskriptivne statistike možemo zaključiti da se naša populacija sastoji uglavnom od muških osoba. Kada formulišete svoja pitanja, uvek uzmite u obzir dostupne podatke i varijable koje ste identifikovali (Garth, 2008). Ovime je završen prvi deo SPSS analize podataka, a sada ćemo nastaviti s pripremom testa.

7.2 Upravljanje podacima

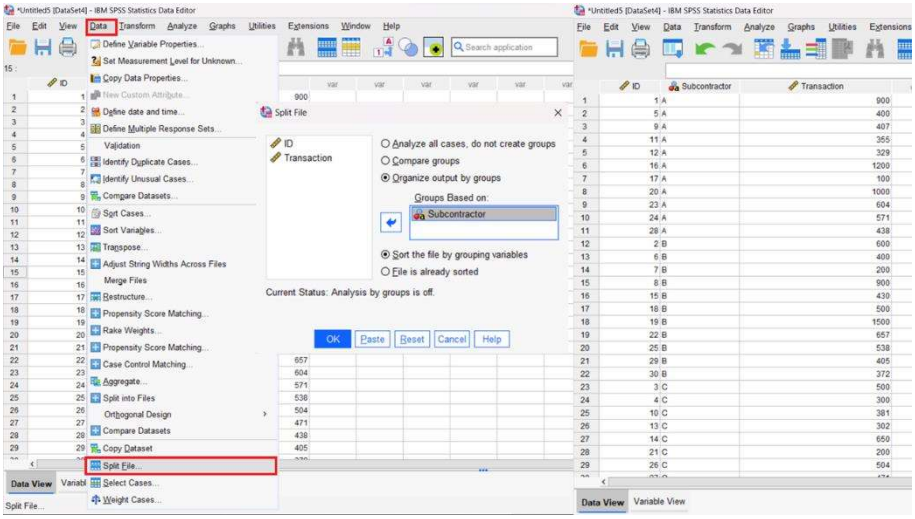
Kada se bavite ključnim podacima u SPSS softveru, postaje ključno razumeti tehnike za manipulisanje informacijama u pojedinačnim aktivnim skupovima podataka. SPSS pruža funkcionalnosti koje olakšavaju manipulaciju postojećim podacima sadržanim u aktivnim skupovima podataka. Povremeno možete naići na dve baze podataka odvojeno uvezene u

skupove podataka, no prednost je da se spoje radi poboljšane analize. Razmotrimo logističku kompaniju s dve podružnice, od kojih svaka daje podatke o troškovima i prevozu tereta u kilogramima. Cilj menadžera je analizirati ukupnu učinkovitost poduzeća. U SPSS-u to uključuje navigaciju "Podaci", izbor "Spoji datoteke" i dve različite opcije. Jedan uključuje izbor "Slučajevi" i određivanje varijable za spajanje, uklanjanje te varijable dok spaja ostale. Alternativno, izborom opcije "Varijabla" zadržava se varijabla u novom skupu podataka. Praktična primena očita je u našem scenariju logistike, gde spajanje skupova podataka pojednostavljuje sveobuhvatnu analizu učinka kompanije.



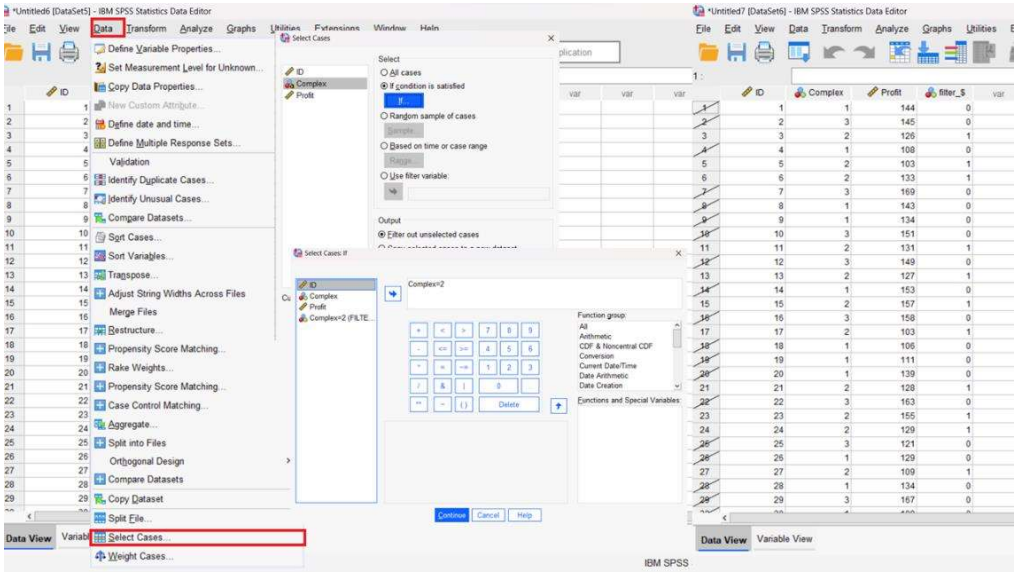
Slika 4.33Prozor za spajanje datoteka.

Dok funkcije spajanja i razdvajanja omogućavaju određenu manipulaciju podacima, opcija "Odaberi slučajeve" nudi različite prednosti. Zamislite da imate podatke za prodavnice B, C i D u jednoj bazi podataka, a fokus je isključivo na poređenju prodavnice A i prodavnice C. Izborom "Podataka" i "Odaberi slučajeve" možete odrediti varijable od interesa, uspešno filtrirajući izbaciti neželjene podatke. Na primer, postavljanje Prodavnice C kao 2 upućuje softver da se koncentriše isključivo na Prodavnicu C, generišući izlaz koji je zatim dostupan za naknadne analize, kao što je deskriptivna statistika, fokusirajući se isključivo na odabrane slučajeve. Takav pristup takođe omogućava komparativnu analizu samo između vrednosti Prodavnica A i Prodavnica C.



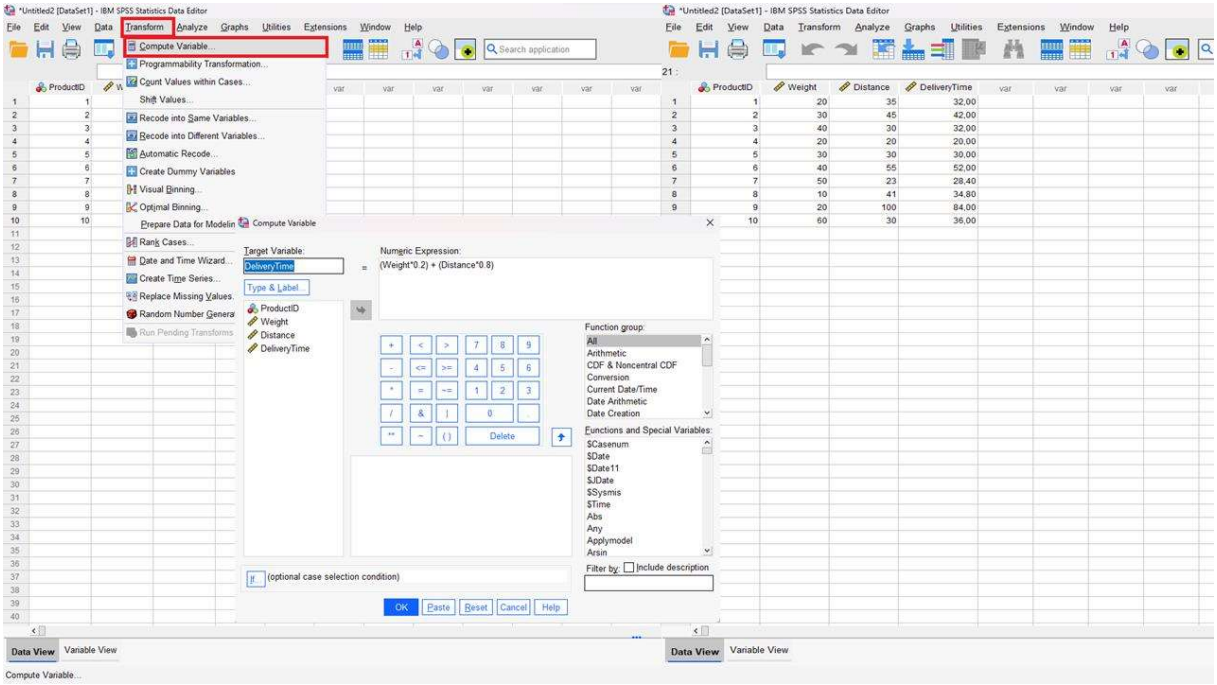
Slika 4.34Prozor za deljenje datoteke.

Dok funkcije spajanja i razdvajanja omogućavaju određene manipulacije podacima, postoji i opcija "Odaberi slučajeve". Zamislite da pouzdano znamo da prodavnica A ima u proseku 120 € dobiti i želimo to uporediti s prodavnicom C. Nažalost, u našoj bazi podataka imamo podatke za prodavnice B, C i D u jednoj bazi podataka i analiza bi uključivala podatke iz sve tri prodavnice. Klikom na "Podaci" i "Odaberi slučajeve" možemo odabrati varijablu na koju se želimo fokusirati. U našim slučajevima smo definisali da prodavnica C treba biti postavljena kao 2, a zatim smo kreirali funkciju za softver da se fokusira samo na prodavnicu C. Izlaz se zatim može koristiti za naknadnu analizu izborom ove nove kolone (npr. deskriptivna statistika).



Slika 4.35Izbor slučaja.

Povremeno skupovi podataka mogu već sadržavati varijable, ali ipak postoji potreba za uvođenjem novih varijabli na osnovu postojećih. Uzmimo, na primer, menadžera logističke kompanije koji poseduje podatke o težini i pređenoj udaljenosti za razne proizvode, ali zahteva vreme isporuke za optimizaciju ruta. U SPSS-u, da bi se to postiglo, potreban je klik na "Transform", a zatim na "Compute Variables". Nova varijabla, DeliveryTime, stvara se unutar novog prozora postavljanjem numeričkih izraza. U ovom slučaju, dodjeljivanje lestvice od 0,8 za udaljenost i 0,2 za težinu rezultuje novom varijablom koja predstavlja vreme isporuke, što je ključni dodatak skupu podataka. Postoji fleksibilnost izračunavanja dodatnih varijabli, kreiranih za potrebe statističkih testova.



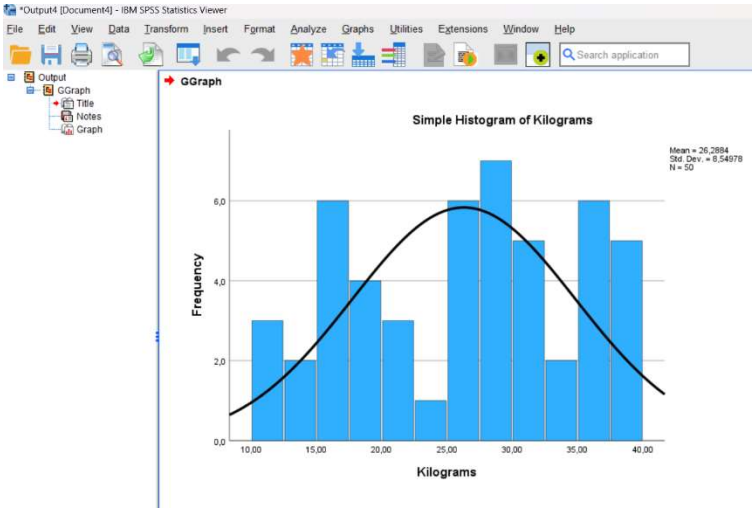
Slika 4.36Postupak izračunavanja varijabli.

Na ovaj način se zaključuje mali pregled funkcija upravljanja podacima koje pokriva SPSS, a koje bi mogle biti korisne tokom sledećih testova modela koji su obuhvaćeni u ovom poglavlju. Nastavićemo s fazama koje su potrebne pre nego što možemo sprovesti statistički test u softveru SPSS.

7.3 Priprema testa

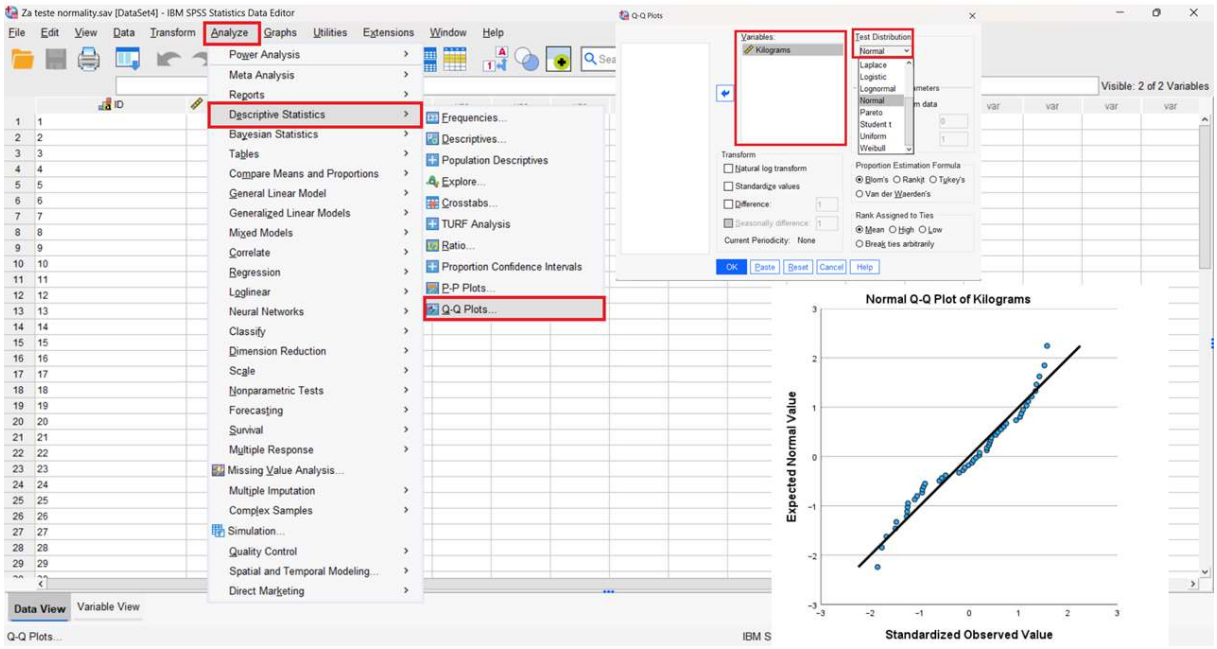
Pre nego što nastavite sa statističkim testovima, bitno je pridržavati se standardnog toka procesa analize podataka, koji uključuje istraživanje podataka (kao što je objašnjeno u poglavljima 7.1 i 7.2), analizu podataka i interpretaciju rezultata (Garth, 2008; George i Mallery, 2022). U ovom poglavlju naš fokus je na analizi podataka pomoću softvera SPSS. Budući da smo hipoteze već obradili u prethodnim poglavljima, naš primarni fokus biće na sprovođenju testova normalnosti unutar SPSS-a. Postoje tri metode za procenu normalnosti: histogram, QQ-grafikon i test normalnosti. Preporučljivo je upotrebiti najmanje dve, ako ne i sve tri ove opcije, budući da svaka pruža različite informacije (Ghasemi i Zadesiasl, 2012.). Za izradu histograma idite na " Graphs ", a zatim na "Chart Builder". U novom prozoru odaberite "Histogram". Ako imate više varijabli, morate ponoviti ovaj postupak za svaku kako biste dobili rezultate. Histogram potvrđuje test za normalnu distribuciju ako stupci koji predstavljaju

vrednosti varijable liče na zvonastu liniju. Ako su stupci više nagnuti u levu ili desnu stranu, to može značiti eksponencijalnu distribuciju. Na primer, generisali smo bazu podataka od 100 ID-ova, svaki s varijablom koja predstavlja težinu u kilogramima. Prateći uputstva, izradili smo histogram, kao što je prikazano na slici 7.9. Kao što se vidi sa slike, stupci su raspoređeni po grafikonu i iako možda ne odražavaju savršenu krivu liniju, ipak sugerišu normalnu distribuciju i pozitivan rezultat testa (George i Mallery, 2022; Goeman i Solari, 2021).



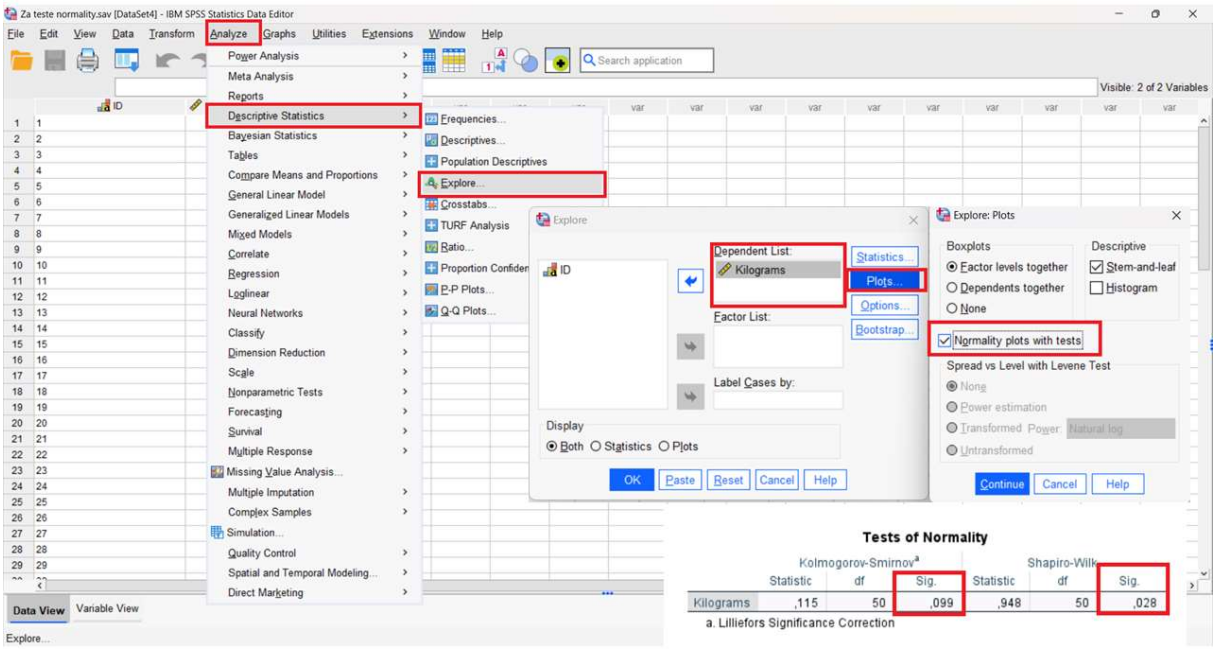
Slika 4.37Histogram rezultata testa normalnosti.

Još jedna opcija za sprovođenje testova normalnosti je QQ-grafikon, koji se može pokrenuti klikom na "Analyze" (hrv. Analiziraj), nakon čega sledi "Descriptive Statistics", a zatim izborom "Q-Q Plots". Prednost ovog pristupa je što omogućava procenu više varijabli istovremeno (Williamson, bd). Test se smatra uspešnim kada se tačke na dijagramu grupišu usko oko prave linije, što predstavlja normalnu distribuciju. Ako tačke formiraju "repove", to ukazuje na neuspešan test normalnosti (Andersen i Dennison, 2018). Koristeći istu bazu podataka iz testa histogramskog grafikona, sproveli smo QQ grafikon test. Na slici 7.10 u nastavku možete primetiti da je većina tačaka klastera za našu varijablu poravnata s pravom linijom, što ukazuje na normalnu distribuciju naših podataka. Iako smo već u ovoj fazi mogli zaključiti da je test normalnosti pozitivan, odlučili smo tražiti potvrdu iz sva tri testa.



Slika 4.38QQ grafikon test normalnosti - postavke i rezultati.

Konačna opcija za sprovođenje testa normalnosti je takozvani Test normalnosti, koji se smatra statističkim testom. Obično se koristi Kolmogorov-Smirnov test, ali za male veličine uzorka može se koristiti Shapiro-Wilkov test (Goeaman i Solari, 2021.). U SPSS-u možete izvršiti ovaj test klikom na "Analyze", nakon čega sledi "Descriptive Statistics", a zatim "Explore". Morate postaviti varijable koje želite proveriti ispod okvira "Dependent List" (hrv. Zavisna lista). Zatim pod "Plots" odaberite "Normality Plots with Tests" (hrv. Grafike normalnosti s testovima). Test se smatra uspešnim ako je stupac Sig (p-vrednost) u rezultatima veći od 0,05, što ukazuje na normalnu distribuciju. Ako je p-vrednost manja od 0,05, to ukazuje da distribucija nije normalna i test se smatra neuspešnim. Ovaj test smo još jednom proveli koristeći istu bazu podataka kao i u prethodnim testovima. Iz rezultata možemo zaključiti da je prema standardu Kolmogorov-Smirnov test pozitivan jer je p-vrednost veća od 0,05. Međutim, za Shapiro-Wilkov test, p-vrednost je niža, što ukazuje na negativan rezultat testa. Do ovih različitih rezultata dolazi jer oba pristupa imaju različite postavke osetljivosti i snagu u otkrivanju odstupanja (Ghasemi i Zahediasl, 2012.). Budući da smo već sproveli testove QQ dijagrama i histogramskog grafikona, Test normalnosti može se generalno smatrati pozitivnim. Uz potvrđene testove normalnosti, možemo sprovesti glavne testove, kao što je test jednog uzorka.



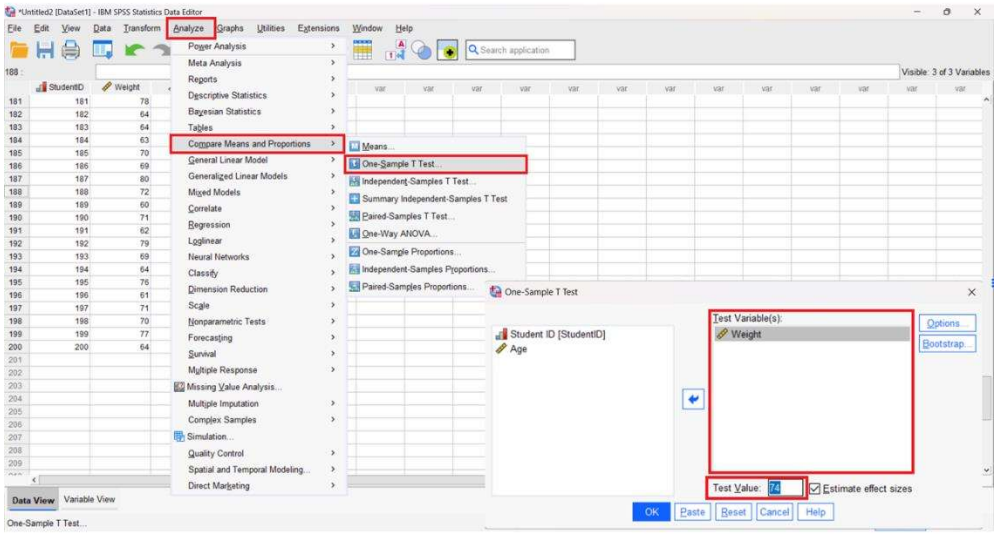
Slika 4.39Postavke i rezultati testa normalnosti.

7.4 T-test jednog uzorka

Već je obrađena teorija za T-test jednog uzorka u prethodnim poglavljima, stoga ćemo se prvenstveno fokusirati na sprovođenje testa sa softverom SPSS. Za naš T-test jednog uzorka pripremili smo bazu podataka s uzorkom od 200 ispitanika, koji uključuje 1 kategoričku varijablu (ID studenta) i 2 numeričke varijable (težinu i starost) (Kim, 2015.). Prateći uputstva iz prethodnih potpoglavlja sprovodimo sledeće korake:

- Istražite podatke, tačnije naše **varijable** i **deskriptivnu statistiku** i postavite naše **pitanje**.
- Proverite **normalnost**, budući da bi samo jedan varijabilni histogram i QQ grafikon trebali biti dovoljni.
- Postavite hipotezu, gde se za **nultu -** varijabla ne razlikuje od određene vrednosti i **alternativu** gde je drugačija.
- Uradite **Studentov T-test** .
- Tumačite rezultate, fokusirajući se na to da li je **nulta hipoteza odbijena** ili **ne** , odgovorite na pitanje i napišite izveštaj o našem testu.

U našem slučaju odlučili smo da naše pitanje bude: Da li je prosečna težina učenika veća od 74 kilograma? Nakon pitanja postavljamo našu hipotezu za pitanje, a to je "Nulta = nema razlike" i "Alternativa = postoji razlika". Sproveli smo histogram i QQ grafikone kako bismo proverili testove normalnosti, a nakon njihovog završetka, sledio je T-test. Da bismo pokrenuli T-test, kliknemo "Analyze" i nastavimo s "Compare Means" (Uporedi srednje vrednosti) i "One-Sample T-test" (T-test jednog uzorka). U okvir s varijablama testa stavljamo studentski ID, postavljamo vrednost testa na 74 i započinjemo test (pogledajte sliku 7.12).



Slika 4.40Postavke T-testa jednog uzorka.

Nakon potvrde testa, pojaviće se drugi prozor s rezultatima naše analize (pogledajte sliku 7.13). Ovaj prozor pruža nekoliko informacija u vezi s našom analizom. U ovom slučaju su obe *p* -vrednosti niže od 0,05, što ukazuje na značajnost testa. Dodatno, proveravamo vrednosti *t* i *df* , koje su u našem slučaju -9,806 odnosno 199. Iz ovih rezultata možemo zaključiti da je naša nulta hipoteza odbačena. Stoga, izveštaj o rezultatima glasi: "Prosečna težina studenta značajno je niža (srednja vrednost = 69,63) od vrednosti od 74 kg (t-test jednog uzorka, t = -9,806, df = 199, p -vrednost < 0,001)".

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Weight	200	69,63	6,303	,446

One-Sample Test

Test Value = 74

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Weight	-9,806	199	<,001	<,001	-4,370	-5,25	-3,49

One-Sample Effect Sizes

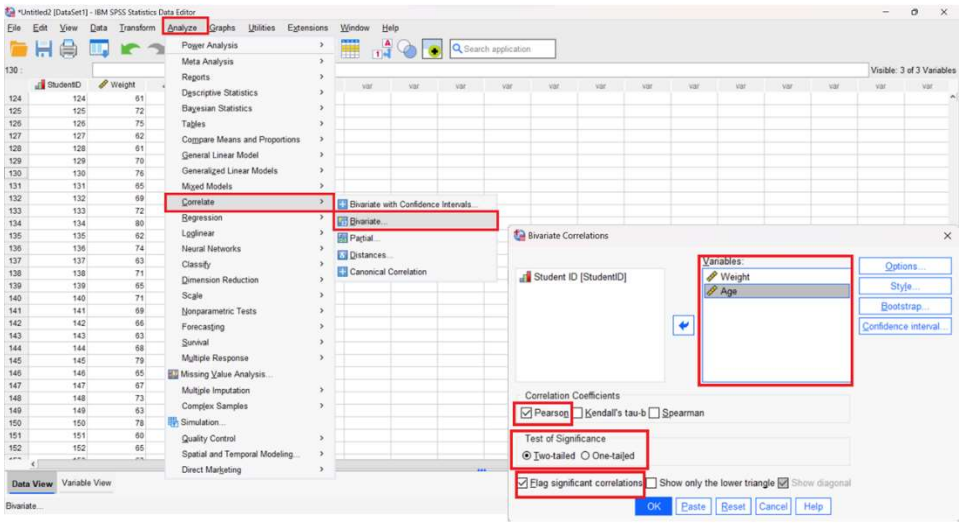
		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Weight	Cohen's d	6,303	-,693	-,847	-,538
	Hedges' correction	6,326	-,691	-,844	-,536

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Slika 4.41Rezultati T-testa jednog uzorka.

7.5 Korelacija

Pređimo sada na drugi test, a to je test korelacije. Uradićemo ga koristeći istu bazu podataka kao u primeru t-testa s jednim uzorkom. Slično t-testu s jednim uzorkom, pratićemo postupak uz nekoliko izmena. Kada se vrši korelacija između dve varijable, važno je odrediti koja je zavisna, a koja nezavisna varijabla (Janse i dr., 2021.; Mishra i dr., 2019). Ovaj izbor možete napraviti na osnovu vašeg istraživačkog pitanja. U našem slučaju želimo istražiti "Postoji li korelacija između starosti studenta i njegove težine?". Nakon pitanja, težinu smatramo zavisnom varijablom, a starost nezavisnom varijablom, jer želimo istražiti jesu li varijacije u godinama povezane s varijacijama u težini. Definišemo naše nulte i alternativne hipoteze (vidi 7.3 i 7.4), a zatim pokrećemo test klikom na "Analyze", nakon čega slede "Correlate" (hrv. Koreliraj) i "Bivariate" (hrv. Bivarijantno). Obe varijable treba staviti u polje "Variable". Proverite jesu li odabrani ili postavljeni "Pearson", "Two-Tailed" i "Flag Significant" (pogledajte sliku 7.14). U ovom slučaju smo odabrali "Pearson" jer naši podaci pokazuju normalnu distribuciju i mogu se analizirati pomoću parametarskih metoda. Ako normalna distribucija nije naznačena, treba koristiti neparametarske metode (u ovom slučaju, odabrali biste Spearmana umesto Pearsona) (George i Mallery, 2022; McClure, 2005).



Slika 4.42Postavke testa korelacije.

Još jednom dobijamo rezultate u novom prozoru (pogledajte sliku 7.15). Iz rezultata možemo videti da je naša Pearsonova korelacija -0,038, a *p* -vrednost 0,596. U korelacionoj analizi, što je vrednost korelacije bliža nuli, to je korelacija između varijabli slabija. U našem slučaju, korelacija je vrlo blizu nule, što ukazuje da nema značajne korelacije između dve varijable (McClure, 2005). Dodatno, visoka *p* -vrednost (0,596) sugeriše da nema značajnih dokaza za zaključak da postoji značajna korelacija između dve odabrane varijable (Williamson, bd). Kao rezultat, naša nulta hipoteza nije odbačena. Na tom osnovu možemo izvestiti da "nije bilo korelacije između starosti i težine studenta".

Correlations

Correlations

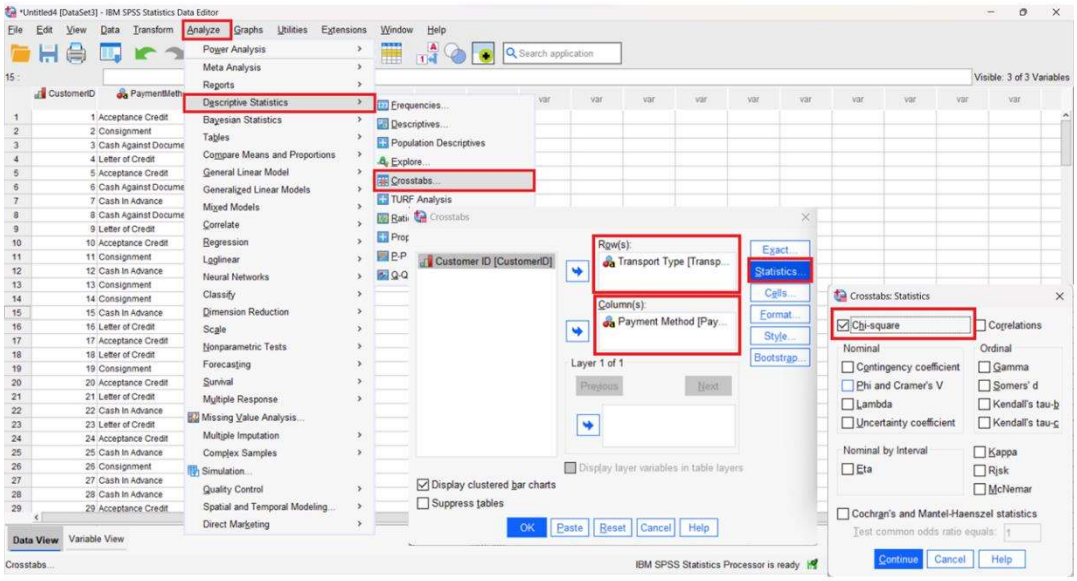
		Weight	Age
Weight	Pearson Correlation	1	-,038
	Sig. (2-tailed)		,596
	N	200	200
Age	Pearson Correlation	-,038	1
	Sig. (2-tailed)	,596	
	N	200	200

Slika 4.43Rezultati testa korelacije.

7.6 Hi-kvadrat

Treći test koji ćemo izvesti u SPSS softveru je Hi-kvadrat test. Za razliku od prethodna dva testa, Hi-kvadrat test upoređuje dve kategoričke varijable, a ne numeričke varijable (Turhan, 2020). Kao i postupak u odeljcima 7.4 i 7.5, počinjemo istraživanjem podataka i formulisanjem

istraživačkog pitanja. U našem primeru imamo logističku kompaniju s 200 kupaca, i podatke o vrsti plaćanja i vrsti prevoza koju je svaki kupac odabrao. Pitanje na koje želimo odgovoriti je: "Pokazuju li različite vrste plaćanja različite preferencije za vrste prevoza?" Budući da se radi samo o kategoričkim varijablama, nema potrebe za testom normalnosti. Postavljamo našu nultu hipotezu (preferencije za vrste prevoza iste su za sve vrste plaćanja) i alternativnu hipotezu. Za sprovođenje hi-kvadrat analize kliknite na "Analyze", nakon čega sledi "Descriptive Statistics", i odaberite "Crosstabs" (Unakrsne analize). Ključno je smestiti varijable na osnovu vašeg istraživačkog pitanja u kolone ili redove (pogledajte sliku 7.16) (Garth, 2008.).



Slika 4.44Postavke Hi-kvadrat testa.

Nakon analize, novi prozor prikazuje rezultate (pogledajte sliku 7.17). U ovom prozoru možete primijetiti da Pearsonova hi-kvadrat vrednost iznosi 11,614, *df* vrednost 12, a *p* -vrednost (asimptotska značajnost) 0,477. Na osnovu ovih rezultata možemo zaključiti da ne postoji značajna povezanost između dve varijable, a nulta hipoteza nije odbačena. Stoga sledi izveštaj: "Nema otkrivenih značajnih preferencija između različitih vrsta plaćanja za različite vrste prevoza (dvostrani Chi-Square test, chi-sq = 11,614, *df* = 12, *p* -vrednost = 0,477)."

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Transport Type * Payment Method	200	100,0%	0	0,0%	200	100,0%

Transport Type * Payment Method Crosstabulation

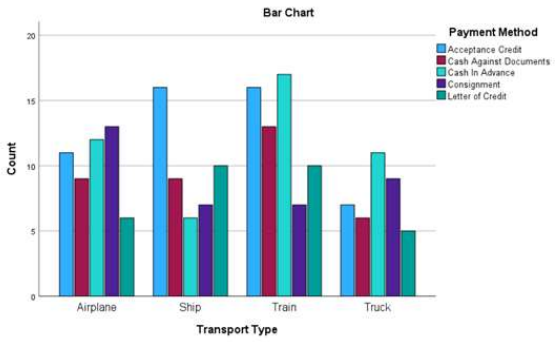
Count

		Payment Method					
		Acceptance Credit	Cash Against Documents	Cash In Advance	Consignment	Letter of Credit	Total
Transport Type	Airplane	11	9	12	13	6	51
	Ship	16	9	6	7	10	48
	Train	16	13	17	7	10	63
	Truck	7	6	11	9	5	38
Total		50	37	46	36	31	200

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	11,614[a]	12	,477
Likelihood Ratio	11,965	12	,448
N of Valid Cases	200		

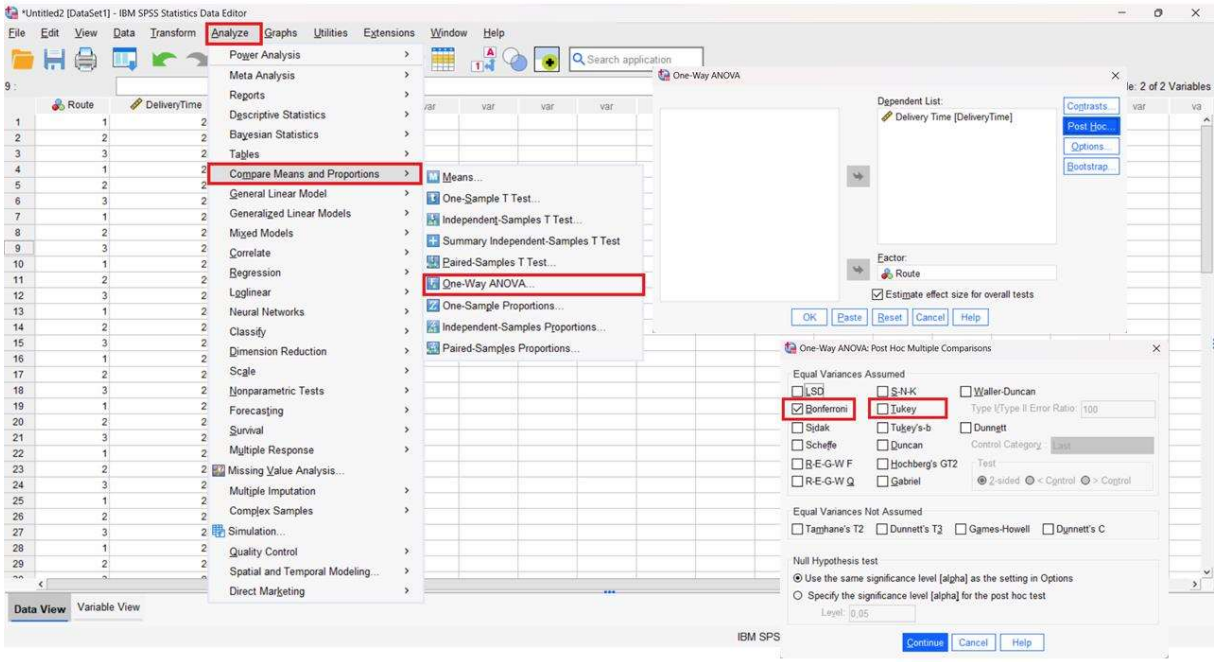
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,89.



Slika 4.45Rezultati Hi-kvadrat testa.

7.7 ANOVA

Poslednji test koji ćemo pokriti je ANOVA test, koji se posebno fokusira na jednostavniji model poznat kao jednosmerna ANOVA, koji uključuje kategoričku varijablu i numeričku varijablu (Goeman i Solari, 2021). Kao i kod T-testa, sledićemo isti postupak: istražiti podatke, formulisati istraživačko pitanje, sprovesti test normalnosti i postaviti hipoteze. Razmotrimo studiju slučaja transportnog dispečera koji radi za logističku kompaniju. Dispečer blisko sarađuje s partnerskom kompanijom i redovno planira tri različite rute kamionima za isporuku robe. Zbog politike "Just-in-time" koja naglašava brže isporuke, postavlja se pitanje: "Utiče li izbor rute dostave na vreme isporuke za kompaniju?" Da biste pokrenuli ANOVA test u SPSS-u, idite na "Analyze" nakon čega sledi "Compare Means..." i zatim "One-way ANOVA". Postavite zavisnu varijablu u okvir "Dependent List" (Popis zavisnih), a varijablu Faktor u okvir "Factor" (pogledajte sliku 7.18). Za detaljnu analizu uključili smo i Post Hoc postavku. Važno je napomenuti da se Post Hoc analiza treba sprovesti samo ako je početni ANOVA test pozitivan. Primenom Post Hoc analize možemo identifikovati optimalan izbor (u našem slučaju rutu). Najpouzdanije metode koje se koriste za Post Hoc analizu su ili Bonferronijeva korekcija ili Tukeyjeva HSD metoda (Goeman i Solari, 2021.).



Slika 4.46Postavke za ANOVA analizu.

Rezultati naše analize pokazuju da je naša *F*-statistička vrednost 11,173 (više vrednosti ukazuju na više varijacija između grupa) i *p*-vrednost <0,001, što znači da je naša nulta hipoteza odbačena (vidi sliku 7.19). Budući da postoji značajna razlika između tri rute (<0,001), post hoc test u našem slučaju je takođe valjan (George & Mallery, 2022). Nakon sprovođenja Bonferronijevog testa korekcije, možemo videti da su najbolje *p*-vrednosti zabeležene u slučaju rute 2 (pogledajte sliku 7.19). U izveštaju možemo zaključiti da je „postojala značajna razlika u izboru rute isporuke u korelaciji s vremenom isporuke (1-way ANOVA, $F = 11,173$, $df = 47$, *p*-vrednost = <0,001). Ruta 2 imala je najbolje rezultate vremena isporuke."

ANOVA

Delivery Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16,527	2	8,263	11,173	<,001
Within Groups	33,280	45	,740		
Total	49,807	47			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Delivery Time

Bonferroni

(I) Route	(J) Route	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-1,4250*	,3040	<,001	-2,181	-,669
	3	-,5500	,3040	,231	-1,306	,206
2	1	1,4250*	,3040	<,001	,669	2,181
	3	,8750*	,3040	,018	,119	1,631
3	1	,5500	,3040	,231	-,206	1,306
	2	-,8750*	,3040	,018	-1,631	-,119

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

Slika 4.47Početni rezultati ANOVA analize i rezultati post hoc testa.

Zaključujemo ovo poglavlje knjige uz razumevanje da smo u ovom poglavlju pokrili neke od uobičajenih testova. Postoje i drugi testovi, kao što je ANOVA ponovljenih merenja, testovi pouzdanosti i testovi osetljivosti, koji se takođe mogu modelirati i analizirati pomoću softvera SPSS. Ovi dodatni testovi pružaju širi raspon alata za analizu podataka i daju korisne uvide u različita istraživanja i praktične primene.

LITERATURA 7. POGLAVLJA

- Andersen, A.J. & Dennison, J.R. (2018). An Introduction to Quantile-Quantile Plots for the Experimental Physicist. Journal Articles, 51.
- Garth, A. (2008). Analysing data using SPSS [available at: https://students.shu.ac.uk/lits/it/documents/pdf/analysing_data_using_spss.pdf, access October 26, 2023]
- George, D. & Mallery, P. (2022). IBM SPSS Statistics 27 Stet by Step: A Simple Guide and Reference, 17TH edition, Abingdon: Routledge
- Ghasemi, A. & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. International Journal of Endocrinology and Metabolism, 10(2), pp. 486-489.
- Goeman, J.J. & Solari, A. (2021). Comparing Three Groups. The American Statistician, 76(2), pp. 168-176
- IBM (2021). IBM SPSS Statistics 28 Brief [available at: https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Brief_Guide.pdf, access October 26, 2023]
- Janse, R.J., Hoekstra, T., Jager, K.J., Zoccali, C., Tripepi, G., Dekker, F.W. & van Diepen, M. (2021). Conducting correlation analysis: important limitations and pitfalls, 14(11), pp. 2332-2337.
- Kim, T.K. (2015). T test as a parametric statistic. Korean Journal of Anesthesiology, 68(6), pp. 540-546
- Landau, S. & Everitt, B.S. (2004). A Handbook of Statistical Analyses using SPSS, 1st edition, London: Chapman & Hall/CRC

- McClure, P. (2005). Correlation Statistics Review of the Basics and Some Common Pitfalls. Journal of Hand Therapy, 18(3), pp. 378-380
- Mishra, P., Singh, U., Pandey, C.M., Mishra, P. & Pandey, G. (2019). Application of Student's t-test, Analysis of Variance, and Covariance. Annals of Cardiac Anesthesia, 22(4), pp. 407-411
- Turhan, N.S. (2020). Karl Pearson's chi-square tests. Educational Research and Reviews, 15(9), pp. 575-580
- Williamson, M. (b.d.). Data Analysis using SPSS [available at: https://med.und.edu/research/daccota/_files/pdfs/berdc_resource_pdfs/data_analysis_using_spss.pdf, access October 26, 2023]