7. Statistična obdelava podatkov SPSS

Do zdaj ste že pridobili temeljno razumevanje statistike, ravnanja s podatki, vzpostavitve simulacije, modeliranja in analize v logističnih oskrbovalnih verigah ter preprostih metod linearne regresije. Čeprav statistika ponuja raznoliko paleto modelov in tehnik za izboljšanje vaših optimizacijskih

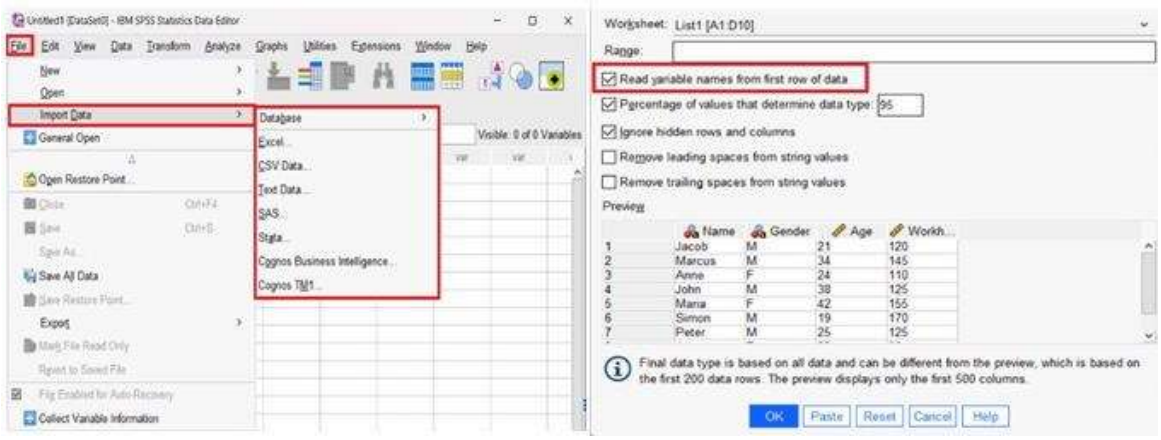
prizadevanj, izvajanje analiz in ugotavljanje morebitnih izboljšav, ste morda opazili, da lahko z naraščanjem kompleksnosti analiziranih podatkov in izračunov tradicionalni pristopi postanejo vse bolj zapleteni in zahtevni za izračunavanje. S stopnjevanjem zapletenosti podatkov in izračunov lahko običajne metode postanejo zastarele in v nekaterih primerih ogrozijo zanesljivost vaših rezultatov. Za reševanje tega problema statistika uporablja različne računalniške programe, ki avtomatizirajo analizo in razlago zbranih podatkov, hkrati pa zagotavljajo številne modele in funkcije za zagotavljanje zanesljivih rezultatov. Eden od takih programov je IBM-ov SPSS, ki bo ključno orodje v tem poglavju. V tem poglavju bomo na kratko predstavili osnovno uporabo programske opreme SPSS ter raziskali njene funkcionalnosti in praktično uporabo. Začetnemu uvodu bo sledila praktična uporaba programa s štirimi temeljnimi testi za izračun rezultatov: T-test, korelacije, Chi-kvadrat in ANOVA. Za lažje učenje bomo predstavili preproste probleme in njihove rešitve, ki vam bodo pomagali pri spoznavanju teh testov.

7.1 Osnove IBM-ovega programa SPSS

Morda ste že imeli nekaj izkušenj s programsko opremo SPSS. Če pa jo še vedno potrebujete, naj vam ponudimo kratko predstavitev. SPSS, podobno kot njegov bolj znan ekvivalent Excel, omogoča manipulacijo, analizo in vizualizacijo podatkov. Kljub temu pa za razliko od programa Excel, ki je lahko včasih naporen in zapleten pri programiranju funkcij, SPSS ponuja uporabniku prijazen vmesnik za statistično analizo (IBM, 2021). SPSS ponuja različne funkcije in metodologije za učinkovito obdelavo zagotovljenih podatkov. Čeprav se programska oprema SPSS odlikuje z zagotavljanjem obsežnih zmožnosti statistične analize, je navigacija pri manipulaciji s podatki in konfiguracija začetnih nastavitev za analizo včasih lahko zahtevna (IBM, 2021). Zato se bomo poglobili v osnove uvoza podatkov in priprave podatkov za poznejše statistične teste.

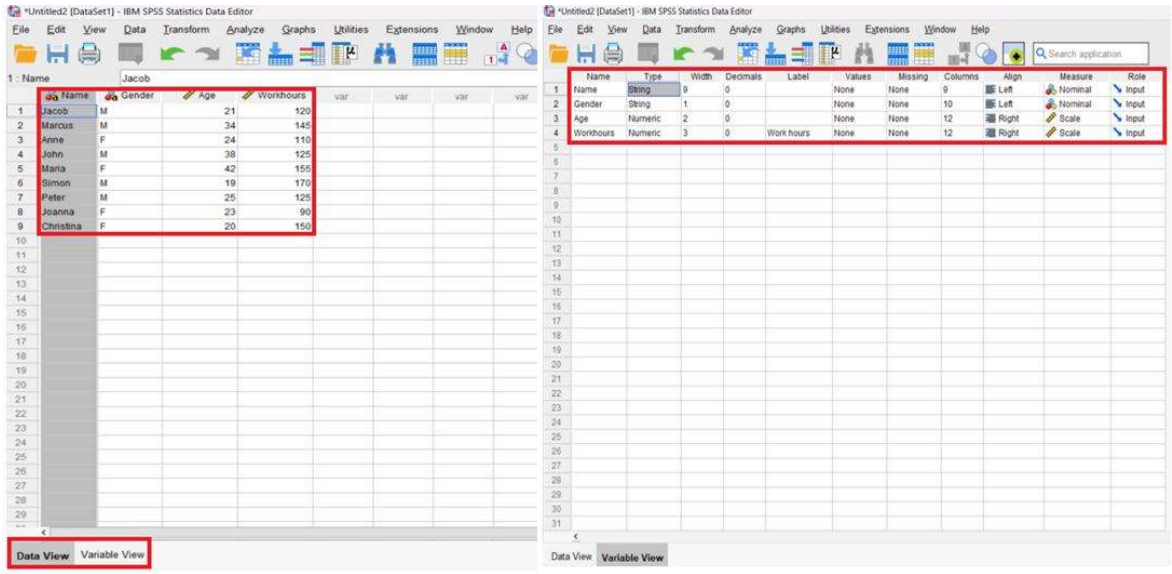
Glede na razširjeno uporabo programa Excel za obdelavo številčnih podatkov so lahko vaši podatki pridobljeni ali pripravljeni v Excelovi preglednici. Na srečo lahko SPSS v svoje preglednice uvozi podatke iz različnih oblik datotek. Ko imate pripravljene končne preglednice Excel, odprite

programsko opremo SPSS. Na začetnem zaslonu pojdite na zavihek "File" (Datoteka) in izberite "Import Data" (Uvozi podatke). V naslednjem oknu lahko izberete obliko podatkov, ki jo nameravate uvoziti (glejte sliko 7.1). Po tem koraku poiščite svojo pripravljeno datoteko, jo izberite in nadaljujte z naslednjim oknom. V tem oknu boste morali konfigurirati dodatne nastavitve. Če ste imena stolpcev že vključili v prvo vrstico podatkov, izberite možnost "Preberi imena spremenljivk iz prve vrstice podatkov" (glejte sliko 7.1) in nato kliknite "Dokončaj", da se podatki prikažejo v preglednici (IBM, 2021).



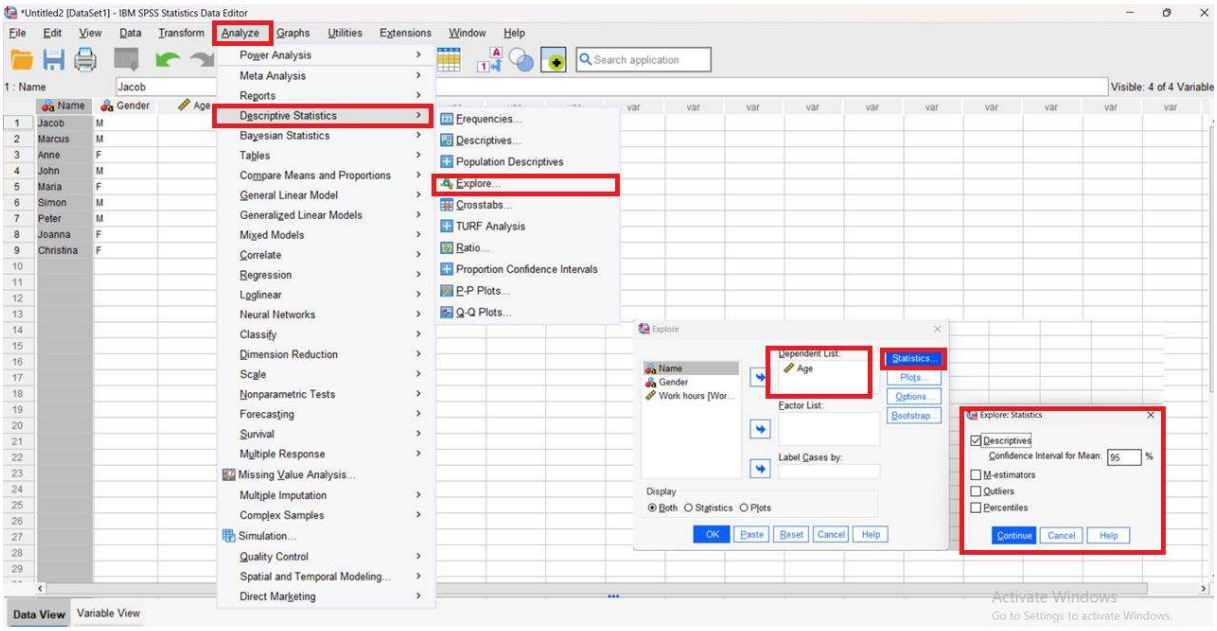
Slika 7.1 Nastavitve uvoza SPSS.

Zdaj, ko imamo podatke v preglednici, boste opazili jasno razliko v predstavitvi v primerjavi z Excelom. SPSS razvršča podatke v dve osnovni vrsti, vsaka pa ima še dve dodatni podvrsti. Kot je prikazano na sliki 7.2, lahko podatke razvrstimo kot numerične ali kategorične. Numerični podatki so sestavljeni iz števil in jih lahko razvrstimo kot diskretne (s končnimi možnostmi) ali zvezne (ponujajo neskončno možnosti). Po drugi strani so kategorični podatki sestavljeni iz besed in se lahko nadalje ločijo na ordinalne (s hierarhijo) ali nominalne (brez hierarhije). Glede na naravo vaših podatkov boste morda morali nastaviti spremenljivke tako, da bodo ustrezale želeni analizi. V večini primerov SPSS samodejno ustrezno razvrsti spremenljivke. Predpostavimo, da želite dodatno manipulirati s podatkovnimi vrstami. V tem primeru lahko dostopate do možnosti "Pogled" in v razdelku "Pogled spremenljivke" prilagodite informacije o spremenljivkah, kot so ime, vrsta, širina, mera in drugo (IBM, 2021).



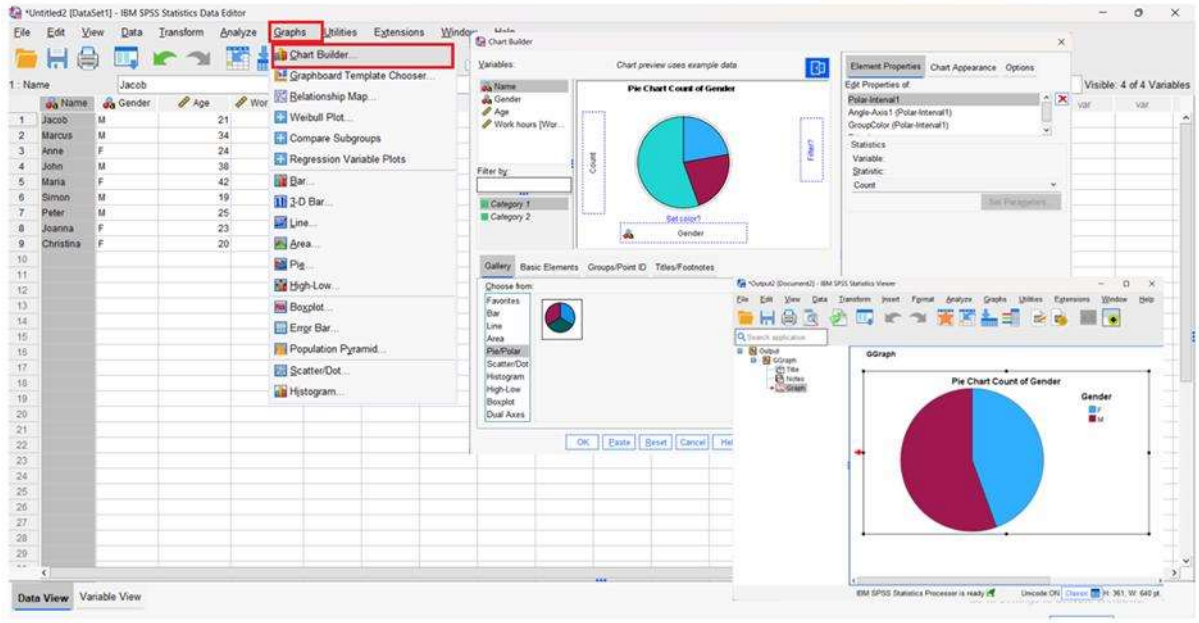
Slika 7.2 Okna za prikaz podatkov in spremenljivk.

Ko pravilno nastavite podatke, jih lahko raziskujete v programu SPSS. SPSS uporabnikom omogoča izvajanje temeljnih statističnih analiz, ne da bi se pri tem zanašali na vnaprej določene funkcije. Na začetnem zaslonu (glejte sliko 7.3) pojdite na "Analyze" (Analiza), nato na "Descriptive Statistics" (Opisna statistika) in izberite "Explore" (Raziskovanje). V razdelku "Explore" (Raziskovanje) boste našli različne možnosti, ki so odvisne od značilnosti predloženih podatkov. V tem načinu vam bo SPSS zagotovil informacije "opisne statistike" o vaših podatkih. To je sicer dragoceno za začetno analizo podatkov, vendar ponuja le temeljni vpogled in se ne spušča v podrobnejšo statistično analizo, ki bo obravnavana v naslednjih poglavjih. Preden nadaljujemo, bomo raziskali še eno funkcijo v SPSS - vizualizacijo grafov (IBM, 2021).



Slika 7.3 Nastavitve opisne statistike.

SPSS ponuja številne možnosti vizualizacije podatkov, vključno s histogrami, škatlastimi diagrami, stolpčnimi diagrami, diagrami razpršitve, linijskimi diagrami, krožnimi diagrami in drugimi. Do te točke bi morali imeti osnovno znanje o tem, kaj predstavljajo posamezne vrste grafov in kako razlagati rezultate, ki jih zagotavljajo. Zato se bomo osredotočili na to, kako ustvariti te grafe v programu SPSS. Za ustvarjanje grafov na začetnem zaslonu izberite zavihek "Graphs" (Grafi) in nato "Chart Builder" (Graditelj grafov). V novem oknu lahko izberete vrsto grafa, ki ga želite ustvariti, in izberete spremenljivke, ki jih želite vključiti. Po izbiri možnosti "Finish" (Končaj) se bo prikazalo novo okno z rezultati, vizualiziranimi v izbrani obliki grafa. V tem novem oknu lahko aktivno sodelujete z grafom, kar vam omogoča spreminjanje barv in pisav spremenljivk, raziskovanje porazdelitve spremenljivk po grafu in drugo (IBM, 2021).



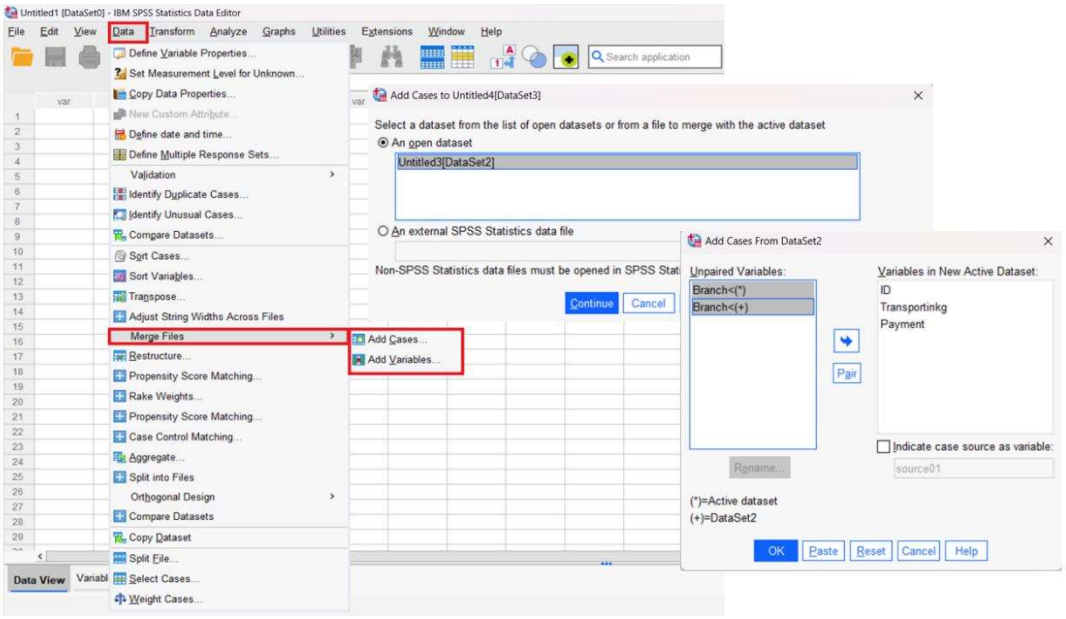
Slika 7.4 Nastavitve programa Chart Builder v SPSS.

Do te točke smo obravnavali tri od štirih pravil za "raziskovanje podatkov", ki vključujejo pregledovanje podatkov (raziskovanje surovih podatkov), prepoznavanje podatkov (določanje vrst podatkov) ter do neke mere grafično prikazovanje in opisovanje podatkov z opisno statistiko in izdelavo grafov. Zadnje pravilo je "oblikovanje vprašanj", pri katerem se vprašamo, kaj želimo doseči z analizo podatkov, ter ustrezno nastavimo grafe in opisno statistiko, da dobimo odgovore na naša specifična vprašanja. V našem trenutnem primeru bi se lahko na primer vprašali: "Ali v naši analizirani populaciji prevladujejo ženske?" Z uporabo grafov in opisne statistike lahko ugotovimo, da našo populacijo sestavljajo predvsem moški posamezniki. Pri oblikovanju vprašanj vedno upoštevajte razpoložljive podatke in opredeljene spremenljivke (Garth, 2008). S tem smo zaključili prvi del analize podatkov v programu SPSS, zdaj pa bomo nadaljevali pripravo testa.

7.2 Upravljanje podatkov

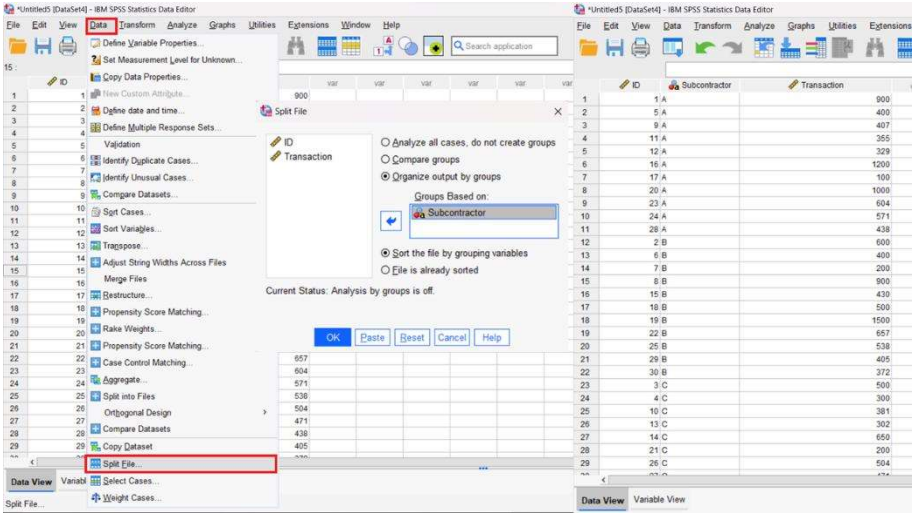
Pri delu s ključnimi podatki v programski opremi SPSS je ključno razumeti tehnike za manipulacijo informacij v posameznih aktivnih podatkovnih nizih. Program SPSS zagotavlja funkcije, ki omogočajo manipulacijo obstoječih podatkov, vsebovanih v aktivnih naborih podatkov. Občasno se lahko zgodi, da sta dve podatkovni zbirki ločeno uvoženi v podatkovne nize, vendar ju je treba zaradi boljše analize združiti. Razmislite o logističnem podjetju z dvema podružnicama, od katerih vsaka prispeva podatke o stroških in prevozu tovora v kilogramih. Vodstveni cilj je analizirati splošno

učinkovitost podjetja. V programu SPSS to dosežemo tako, da se pomaknemo na "Podatki", izberemo "Združiti datoteke" in imamo na voljo dve različni možnosti. Ena vključuje izbiro "Cases" (Primeri) in določitev spremenljivke za združevanje ter odstranitev te spremenljivke ob združevanju drugih. Druga možnost je izbira "Variable" (Spremenljivka), pri kateri se spremenljivka ohrani v novem podatkovnem nizu. Praktična uporaba je očitna v našem logističnem scenariju, kjer združevanje podatkovnih nizov poenostavi celovito analizo uspešnosti podjetja.



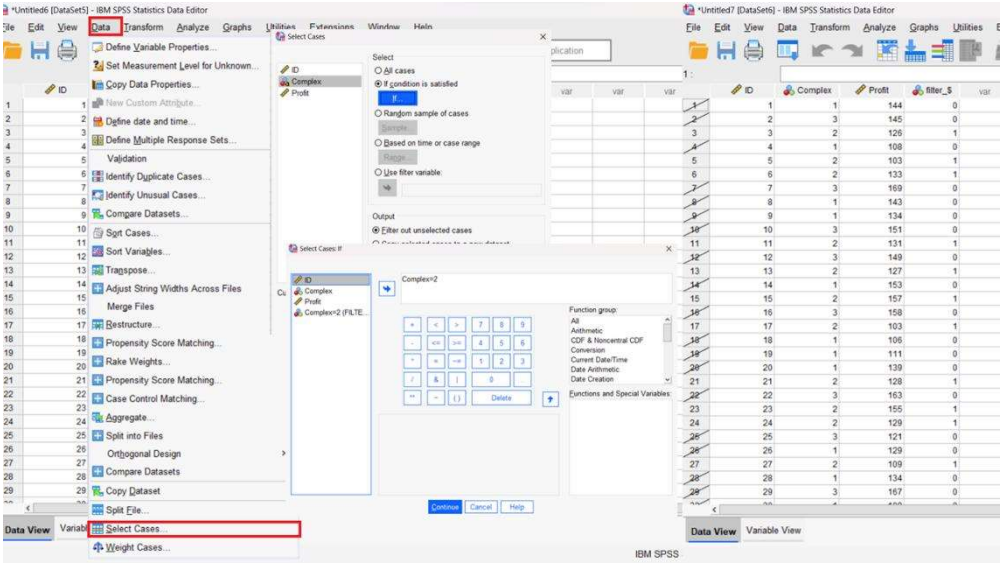
Slika 7.5 Okno za združevanje datotek.

Funkciji združevanja in delitve omogočata posebno ravnanje s podatki, vendar ima možnost "Izberi primere" posebne prednosti. Predstavljajte si, da imate podatke za trgovine B, C in D v eni podatkovni zbirki in se osredotočate izključno na primerjavo trgovine A in trgovine C. Z izbiro "Podatki" in "Izberi primere" lahko določite spremenljivke, ki vas zanimajo, in tako učinkovito filtrirate neželene podatke. Če na primer določite trgovino C kot 2, se programska oprema osredotoči izključno na trgovino C, pri čemer se ustvarijo rezultati, ki so na voljo za nadaljnje analize, kot je opisna statistika, ki se osredotoča izključno na izbrane primere. Takšen pristop omogoča tudi primerjalno analizo samo med vrednostmi v trgovini A in trgovini C.



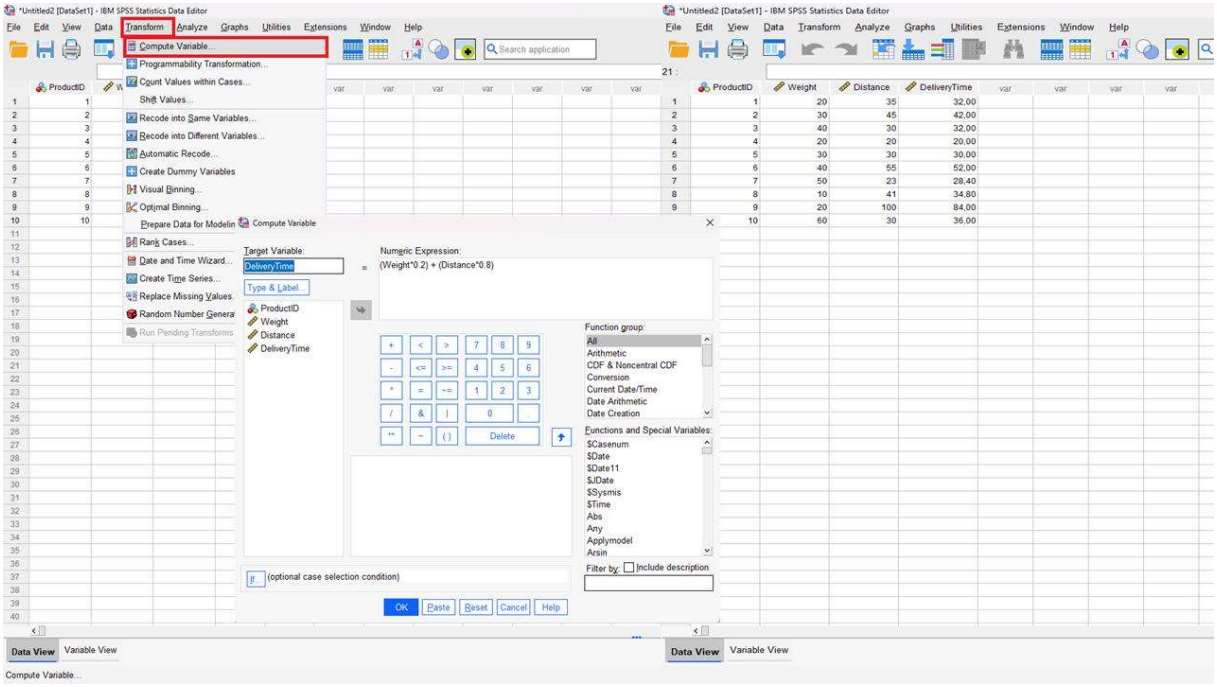
Slika 7.6 Okno za delitev datoteke.

Funkciji združevanja in delitve omogočata določeno ravnanje s podatki, na voljo pa je tudi možnost "Izberite primere". Predstavljajte si, da zagotovo vemo, da ima trgovina A v povprečju 120 EUR dobička, in to želimo primerjati s trgovino C. Žal imamo v naši zbirki podatkov podatke za trgovine B, C in D v eni zbirki podatkov in analiza bi vključevala podatke vseh treh trgovin. S klikom na "Podatki" in "Izberi primere" lahko izberemo, na katero spremenljivko se želimo osredotočiti. V naših primerih smo določili, da mora biti trgovina C nastavljena kot 2, in nato ustvarili funkcijo za programsko opremo, da se osredotoči samo na trgovino C. Rezultat lahko nato z izbiro tega novega stolpca (npr. opisna statistika) uporabimo za poznejšo analizo.



Slika 7.7 Izbira postopka za primer.

Včasih lahko podatkovni nizi že vsebujejo spremenljivke, vendar je treba uvesti nove spremenljivke na podlagi obstoječih. Vzemimo na primer vodjo logističnega podjetja, ki ima podatke o teži in prevoženi razdalji različnih izdelkov, vendar za optimizacijo poti potrebuje čas dostave. V programu SPSS to dosežemo tako, da kliknemo "Transform" in nato "Compute Variables". V novem oknu se z nastavitvijo številskih izrazov ustvari nova spremenljivka, DeliveryTime. V tem primeru se z dodelitvijo lestvice 0,8 razdalji in 0,2 teži ustvari nova spremenljivka, ki predstavlja čas dostave, kar je ključni dodatek k podatkovni zbirki. Obstaja možnost izračunavanja dodatnih spremenljivk, ustvarjenih za potrebe statističnih testov.



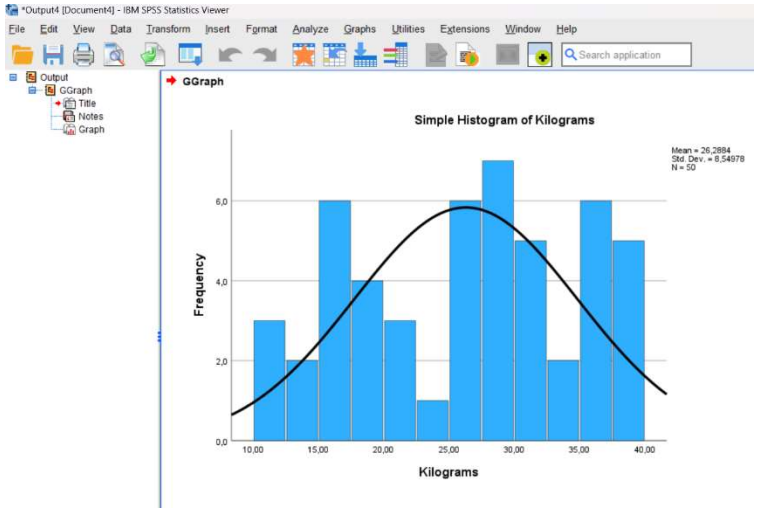
Slika 7.8 Postopek izračuna spremenljivk.

S tem zaključujemo kratek pregled funkcij za upravljanje podatkov, ki jih pokriva SPSS in bi lahko bili koristni pri nadaljnjih testih modelov, ki jih obravnavamo v tem poglavju. Nadaljevali bomo s fazami, ki so potrebne, preden lahko izvedemo statistični test v programu SPSS.

7.3 Priprava na testiranje

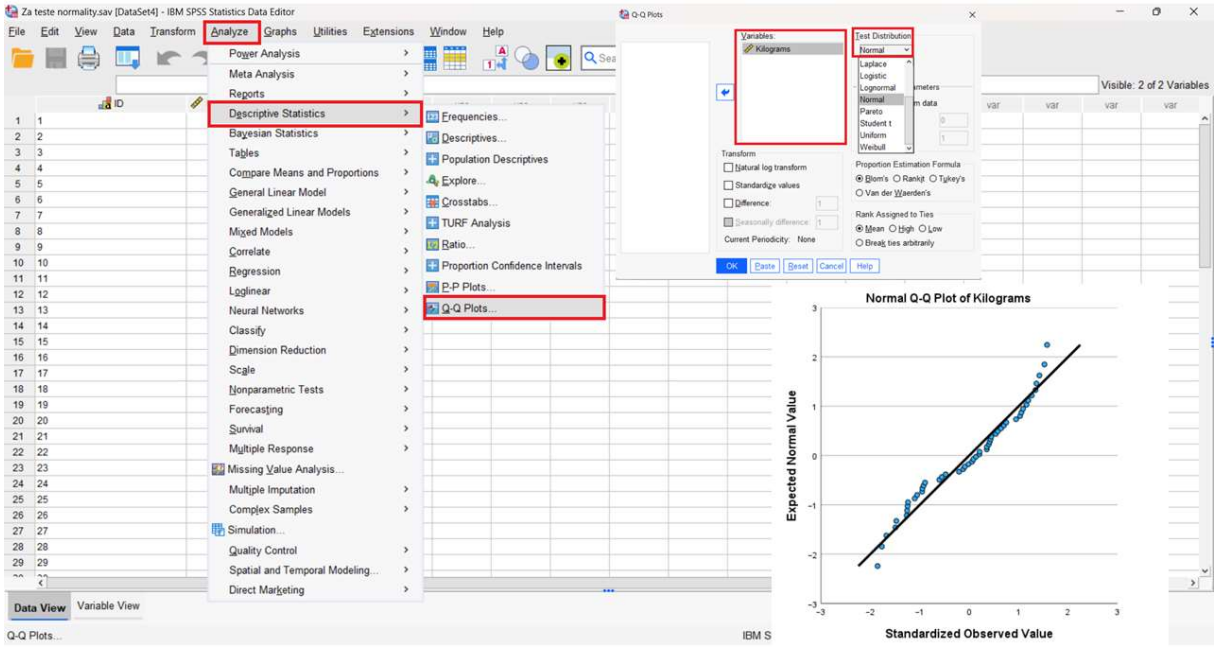
Preden se lotimo statističnih testov, se je treba držati standardnega postopka analize podatkov, ki vključuje raziskovanje podatkov (kot je opisano v poglavjih 7.1 in 7.2), analizo podatkov in interpretacijo rezultatov (Garth, 2008; George in Mallery, 2022). V tem poglavju se osredotočamo na analizo podatkov z uporabo programske opreme SPSS. Ker so bile hipoteze obravnavane že v prejšnjih poglavjih, se bomo osredotočili predvsem na izvajanje testov normalnosti v programu

SPSS. Obstajajo tri metode za ocenjevanje normalnosti: histogram, QQ-plot in test normalnosti. Priporočljivo je uporabiti vsaj dve, če ne vseh treh možnosti, saj vsaka od njih zagotavlja različne informacije (Ghasemi in Zadesiasl, 2012). Če želite ustvariti histogram, pojdite na "Grafi" in nato na "Grafični gradnik". V novem oknu izberite "Histogram". Če imate več spremenljivk, morate ta postopek ponoviti za vsako spremenljivko posebej, da dobite rezultate. Histogram potrdi preskus normalne porazdelitve, če so stolpci, ki predstavljajo vrednosti spremenljivk, podobni zvonovi krivulji. Če so palice bolj nagnjene na levo ali desno stran, to lahko pomeni eksponentno porazdelitev. Na primer, ustvarili smo zbirko podatkov s 100 identifikacijskimi podatki, vsak s spremenljivko, ki predstavlja težo v kilogramih. Po navodilih smo ustvarili histogram, kot je prikazano na sliki 7.9. Kot je razvidno iz slike, so stolpci razporejeni po grafu, in čeprav morda ne odražajo popolnoma krivulje, vseeno kažejo na normalno porazdelitev in pozitiven rezultat testa (George & Mallery, 2022; Goeman & Solari, 2021).



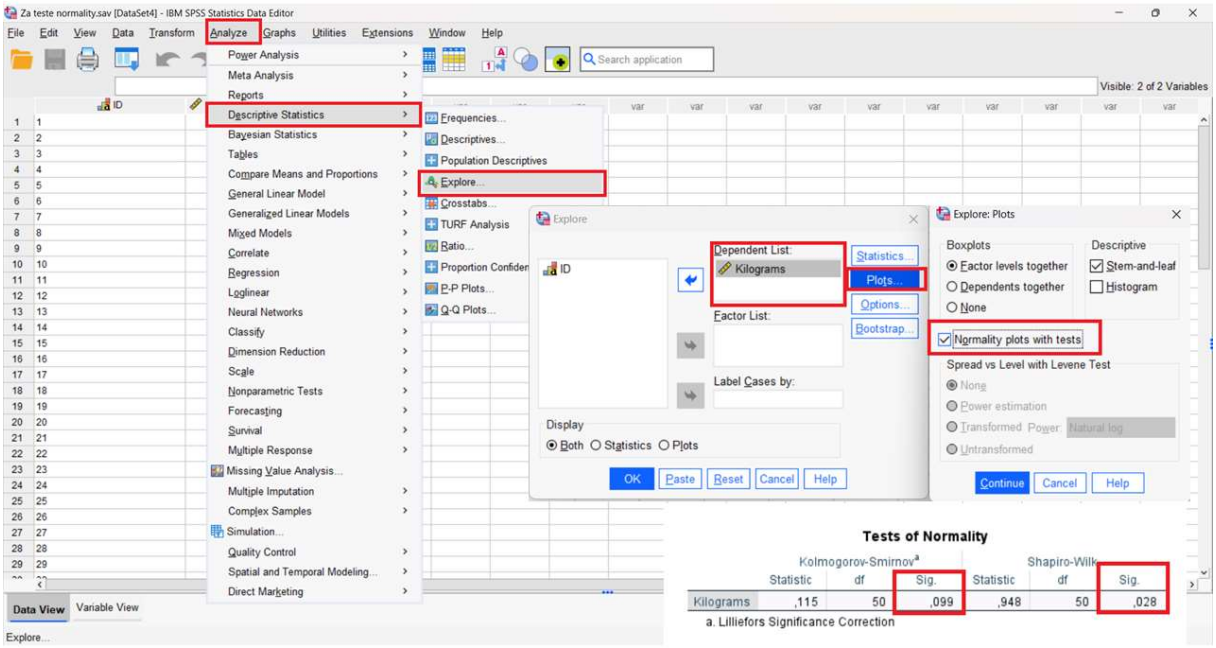
Slika 7.9 Histogram rezultatov testa normalnosti.

Druga možnost za izvajanje testov normalnosti je diagram QQ, ki ga lahko sprožite tako, da kliknete "Analyze", nato "Descriptive Statistics" in izberete "Q-Q Plots". Prednost tega pristopa je, da omogoča hkratno ocenjevanje več spremenljivk (Williamson, b.d.). Test se šteje za uspešnega, če se točke na grafu tesno združijo okoli ravne črte, ki predstavlja normalno porazdelitev. Če točke tvorijo "repe", to pomeni, da test normalnosti ni uspel (Andersen in Dennison, 2018). Z uporabo iste podatkovne zbirke iz testa histogramskega grafa smo izvedli test Q-Q Plot. Na spodnji sliki 7.10 lahko opazite, da se večina točk grozda za našo spremenljivko ujema z ravno črto, kar kaže na normalno porazdelitev naših podatkov. Čeprav bi že na tej stopnji lahko sklepali, da je test normalnosti pozitiven, smo se odločili, da poiščemo potrditev z vsemi tremi testi.



Slika 7.10 Nastavitve in rezultati preskusa normalnosti Q-Q ploskve.

Zadnja možnost za izvedbo testa normalnosti je tako imenovani test normalnosti, ki velja za statistični test. Običajno se uporablja test Kolmogorov-Smirnov, pri majhnih vzorcih pa se lahko uporabi test Shapiro-Wilk (Goeaman in Solari, 2021). V programu SPSS lahko ta test izvedete tako, da kliknete "Analyze" (Analiza), nato "Descriptive Statistics" (Opisna statistika) in nato "Explore" (Raziskovanje). V polju "Dependent List" (Seznam odvisnih spremenljivk) morate določiti spremenljivke, ki jih želite preveriti. Nato v razdelku "Plots" (Grafi) izberite "Normality Plots with Tests" (Grafi normalnosti s testi). Test se šteje za uspešnega, če je stolpec Sig (*p-vrednost*) v rezultatih večji od 0,05, kar kaže na normalno porazdelitev. Če je *p-vrednost* manjša od 0,05, to kaže na nenormalno porazdelitev in test se šteje za neuspešnega. Ta test smo izvedli še enkrat, pri čemer smo uporabili isto podatkovno zbirko kot pri prejšnjih testih. Iz rezultatov lahko sklepamo, da je po standardu Kolmogorova-Smirnova test pozitiven, saj je *p-vrednost* večja od 0,05. Pri Shapiro-Wilkovem testu pa je *p-vrednost* nižja, kar kaže na negativen rezultat testa. Do teh različnih rezultatov pride, ker imata oba pristopa različne nastavitve občutljivosti in moč pri odkrivanju odstopanj (Ghasemi in Zahediasl, 2012). Ker smo že izvedli teste Q-Q Plots in teste histogramskega grafa, lahko test normalnosti na splošno štejemo za pozitivnega. S potrditvijo testov normalnosti lahko izvedemo glavne teste, kot je test enega vzorca.



Slika 7.11 Nastavitve in rezultati testa normalnosti.

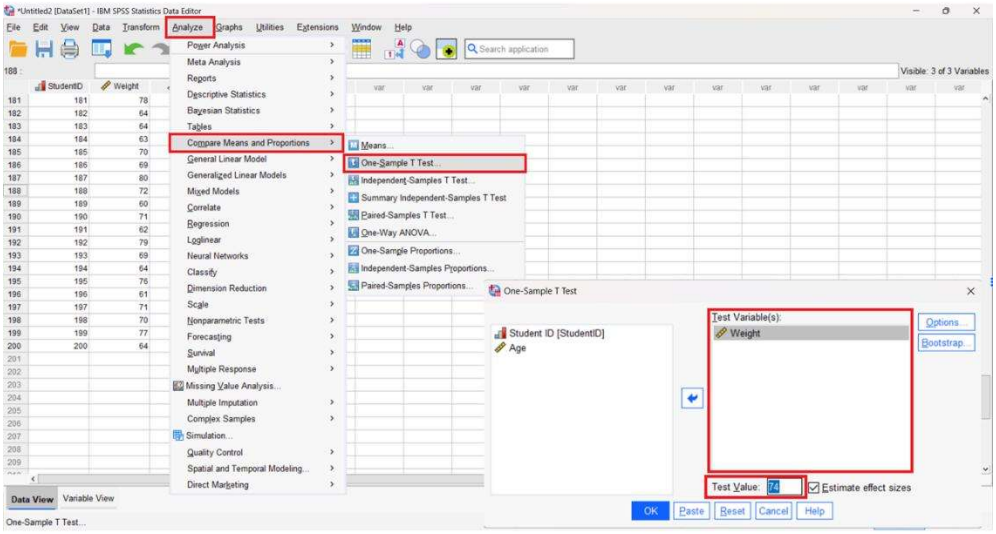
7.4 T-test z enim vzorcem

V prejšnjih poglavjih ste že obravnavali teorijo, ki stoji za T-testom enega vzorca, zato se bomo osredotočili predvsem na izvedbo testa s programsko opremo SPSS. Za naš Eno-vzorčni T-test smo pripravili podatkovno zbirko z vzorcem 200 vnosov, ki vključuje 1 kategorično spremenljivko (ID študenta) in 2 številčni spremenljivki (teža in starost) (Kim, 2015). Po navodilih iz prejšnjih podpoglavij izvedemo:

- Raziščite podatke, in sicer naše **spremenljivke** in **opisno statistiko**, ter določite naše **vprašanje**.
- Preverite **normalnost**, saj bi morala zadostovati le histogram ene spremenljivke in graf Q-Q.
- Postavite hipotezo, pri kateri se pri **ničelni** spremenljivki spremenljivka ne razlikuje od določene vrednosti, pri **alternativni** pa se razlikuje.
- Izvedite **študentski T-test**.
- Interpretirajte rezultate, osredotočite se na to, ali je **ničelni test zavrnjen** ali **ne**, odgovorite na vprašanje in napišite poročilo o našem testu.

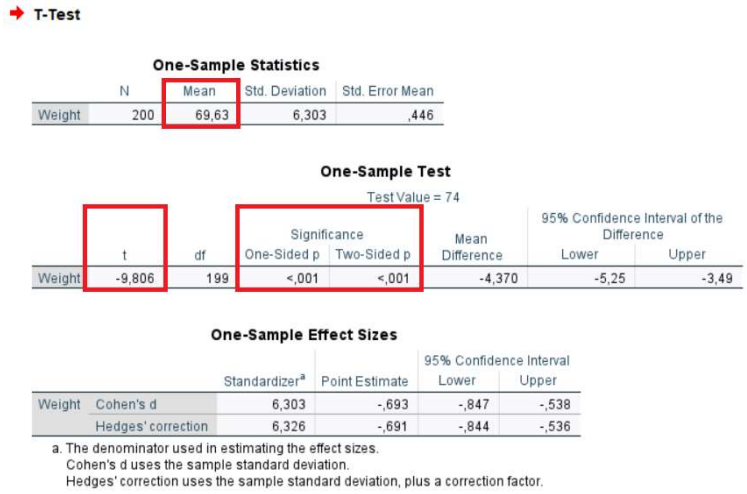
V našem primeru smo se odločili, da se bo naše vprašanje glasilo: "Ali je povprečna teža učencev večja od 74 kilogramov?". Po vprašanju določimo hipotezo za vprašanje, ki se glasi: "Ničelna =

razlike ni" in "Alternativna = razlika obstaja". Izvedli smo histogram in Q-Q grafe, da bi preverili teste normalnosti, po njihovem zaključku pa je sledil T-test. T-test izvedemo tako, da kliknemo "Analyze" (Analiziraj) in nadaljujemo s "Compare Means" (Primerjaj sredine) in "One-Sample T-test" (Eno-vzorčni T-test). V polje za spremenljivko testa vpišemo ID našega študenta, vrednost testa nastavimo na 74 in zaženemo test (glej sliko 7.12).



Slika 7.12 Nastavitve T-testa z enim vzorcem.

Po potrditvi testa se prikaže drugo okno z rezultati analize (glejte sliko 7.13). To okno vsebuje več informacij o naši analizi. V tem primeru sta obe *p-vrednosti* nižji od 0,05, kar kaže na pomembnost testa. Poleg tega preverimo vrednosti *t* in *df*, ki sta v našem primeru -9,806 oziroma 199. Iz teh rezultatov lahko sklepamo, da je naša ničelna hipoteza zavrnjena. Zato je celotno poročilo o rezultatih naslednje: "Povprečna teža učencev je bistveno nižja (povprečje = 69,63) od vrednosti 74 kg (t-test za 1 vzorec, 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

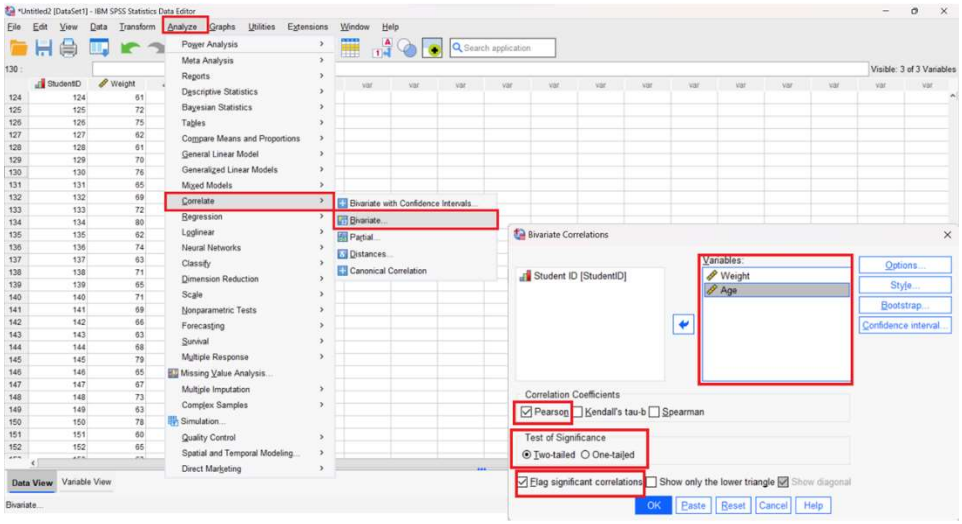
a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Slika 7.13 Rezultati T-testa z enim vzorcem.

7.5 Korelacija

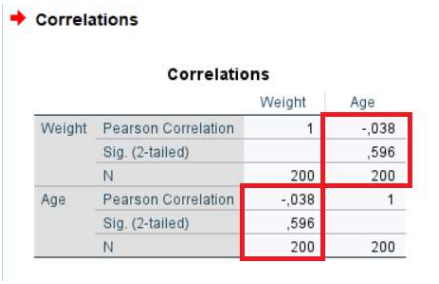
Preidimo na drugi test, to je test korelacije. Izvedli ga bomo z isto podatkovno bazo kot v primeru t-testa. Podobno kot pri t-testu bomo upoštevali postopek z nekaj spremembami. Pri izvajanju korelacije med dvema spremenljivkama je pomembno določiti, katera je odvisna in katera neodvisna spremenljivka (Janse *et al.*, 2021; Mishra *et al.*, 2019). To izbiro lahko opravite na podlagi svojega raziskovalnega vprašanja. V našem primeru želimo raziskati "Ali obstaja povezanost med starostjo učenca in njegovo telesno težo?". V skladu z vprašanjem upoštevamo težo kot odvisno spremenljivko in starost kot neodvisno spremenljivko, saj želimo raziskati, ali so razlike v starosti povezane z razlikami v teži. Opredelimo ničelno in alternativno hipotezo (glej 7.3 in 7.4), nato pa izvedemo test s klikom na "Analyze", ki mu sledita "Correlate" in "Bivariate".

Obe spremenljivki je treba postaviti v polje "Variable" (Spremenljivka). Prepričajte se, da so izbrane ali nastavljene možnosti "Pearson", "Two-Tailed" in "Flag Significant" (glejte sliko 7.14). V tem primeru smo izbrali "Pearson", ker so naši podatki kazali normalno porazdelitev in jih je bilo mogoče analizirati s parametričnimi metodami. Če normalna porazdelitev ni nakazana, je treba uporabiti neparametrične metode (v tem primeru bi namesto Pearsonove izbrali Spearmanovo) (George in Mallery, 2022; McClure, 2005).



Slika 7.14 Nastavitve korelacijskega testa.

Rezultate ponovno dobimo v novem oknu (glej sliko 7.15). Iz rezultatov je razvidno, da je naša Pearsonova korelacija -0,038, *p-vrednost* pa 0,596. Pri korelacijski analizi velja, da je vrednost korelacije tem bližje ničli, čim šibkejša je korelacija med spremenljivkama. V našem primeru je korelacija zelo blizu nič, kar kaže na to, da med spremenljivkama ni pomembne korelacije (McClure, 2005). Poleg tega visoka *p-vrednost* (0,596) nakazuje, da ni bistvenih dokazov za sklepanje, da obstaja pomembna korelacija med izbranima spremenljivkama (Williamson, b. d.). Zato naša ničelna hipoteza ni zavrnjena. Na podlagi tega lahko poročamo, da "med starostjo in telesno težo učencev ni bilo korelacije".



Correlations

Correlations

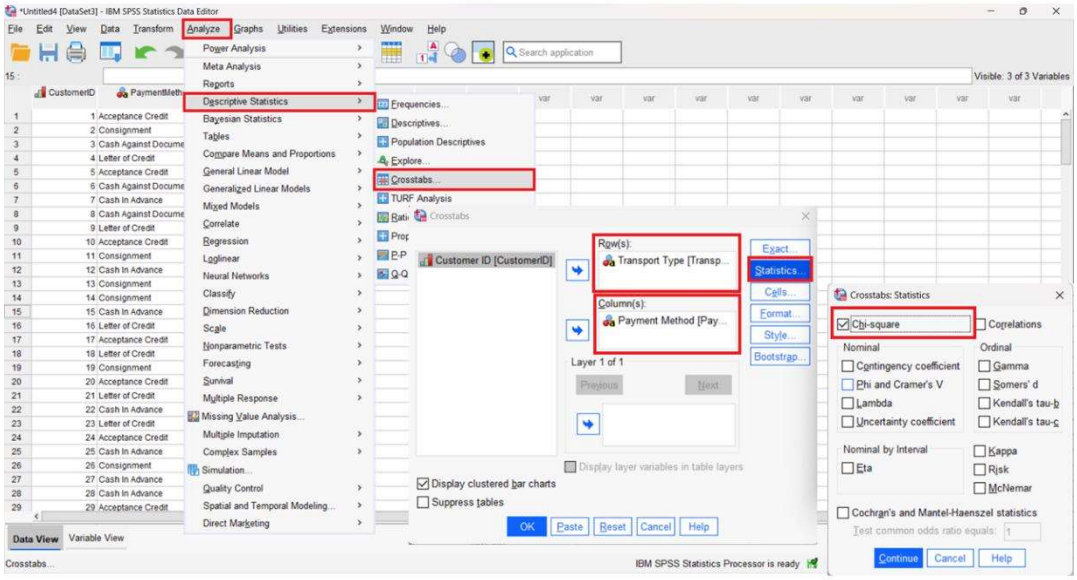
		Weight	Age
Weight	Pearson Correlation	1	-,038
	Sig. (2-tailed)		,596
	N	200	200
Age	Pearson Correlation	-,038	1
	Sig. (2-tailed)	,596	
	N	200	200

Slika 7.15 Rezultati korelacijskega preskusa.

7.6 Chi-kvadrat

Tretji test, ki ga bomo izvedli v programu SPSS, je test Chi-Square. Za razliko od prejšnjih dveh testov test Chi-Square primerja dve kategorični spremenljivki in ne številčnih spremenljivk (Turhan, 2020). Tako kot postopek v razdelkih 7.4 in 7.5 začnemo raziskovanje podatkov in oblikovanje raziskovalnega vprašanja. V našem primeru imamo logistično podjetje z 200 strankami in podatke

o vrsti plačila in vrsti prevoza, ki ga je izbrala vsaka stranka. Vprašanje, na katero želimo odgovoriti, se glasi: "Ali različne vrste plačil kažejo različne preference glede vrste prevoza?" Ker imamo opravka samo s kategoričnimi spremenljivkami, test normalnosti ni potreben. Določimo našo ničelno hipotezo (Preference za vrste prevoza so enake med vsemi vrstami plačil) in alternativno hipotezo. Za izvedbo analize Chi-Square kliknite "Analyze", nato "Descriptive Statistics" in izberite "Crosstabs". Ključno je, da spremenljivke, ki temeljijo na vašem raziskovalnem vprašanju, postavite v polje stolpcev ali vrstic (glejte sliko 7.16) (Garth, 2008).



Slika 7.16 Nastavitve testa Chi-Square.

Po analizi se v novem oknu prikažejo rezultati (glejte sliko 7.17). V tem oknu lahko opazite, da je vrednost Pearsonovega chi-kvadrata 11,614, vrednost *df* 12 in *p-vrednost* (asimptotska pomembnost) 0,477. Na podlagi teh rezultatov lahko sklepamo, da med spremenljivkama ni pomembne povezave in da ničelna hipoteza ni zavrnjena. Zato poročilo sledi: "Med različnimi vrstami plačil za različne vrste prevoza ni zaznati pomembnih preferenc (dvostranski test Chi-Square, 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

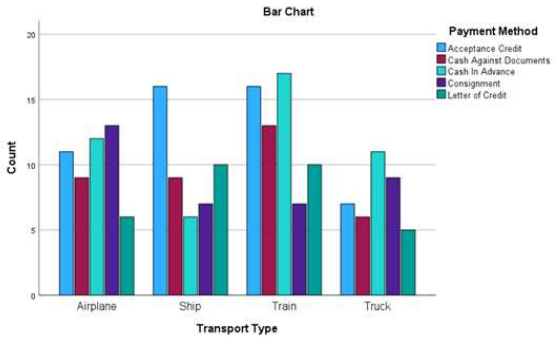
Count

		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		

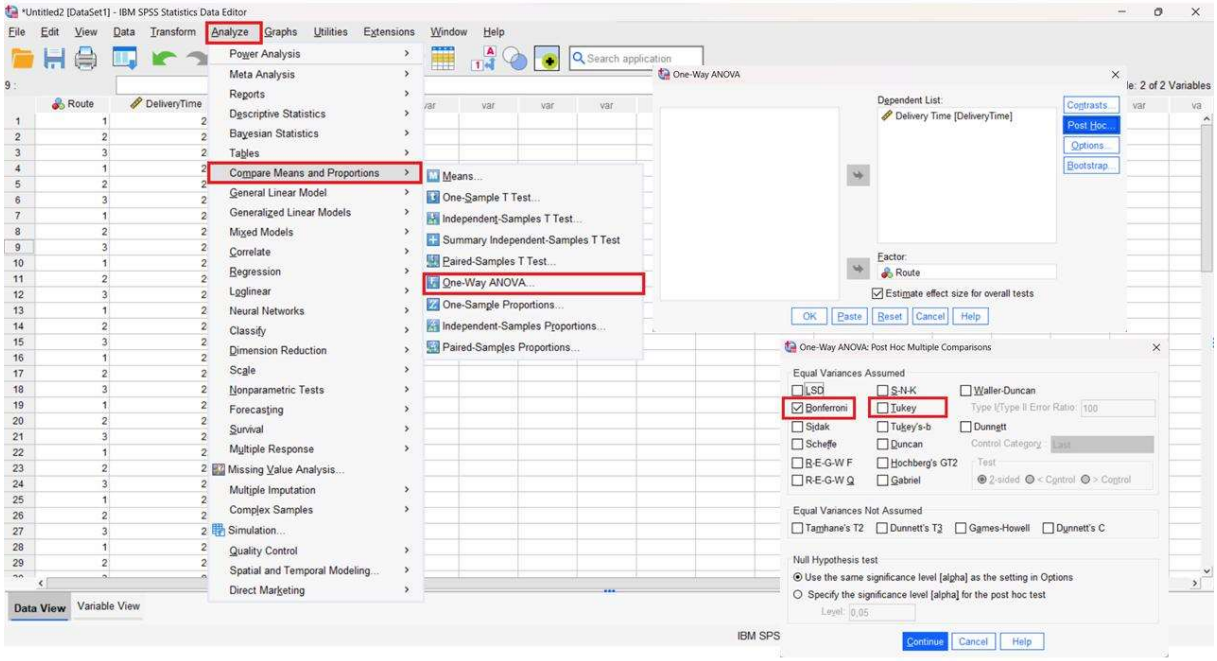
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,89.



Slika 7.17 Rezultati testa Chi-Square.

7.7 ANOVA

Zadnji test, ki ga bomo obravnavali, je test ANOVA, pri čemer se bomo osredotočili na preprostejši model, znan kot enosmerna ANOVA, ki vključuje kategorično spremenljivko in številčno spremenljivko (Goeman in Solari, 2021). Tako kot pri T-testu bomo uporabili enak postopek: raziskali bomo podatke, oblikovali raziskovalno vprašanje, izvedli test normalnosti in postavili hipoteze. Obravnavajmo študijo primera prometnega dispečerja, zaposlenega v logističnem podjetju. Dispečer tesno sodeluje s partnerskim podjetjem in redno načrtuje tri različne poti za tovornjake za dostavo blaga. Zaradi politike "Just-in-time", ki poudarja hitrejše dobave, se postavlja vprašanje: "Ali izbira poti dostave vpliva na čas dostave za podjetje?" Če želite v programu SPSS izvesti test ANOVA, pojdite na "Analyze", ki mu sledi "Compare Means..." in nato "One-way ANOVA". Odvisno spremenljivko postavite v polje "Seznam odvisnih", spremenljivko dejavnik pa v polje "Dejavnik" (glejte sliko 7.18). Za temeljito analizo smo vključili tudi nastavitev Post Hoc. Pomembno je opozoriti, da je treba analizo Post Hoc izvesti le, če je začetni test ANOVA pozitiven. Z uporabo analize Post Hoc lahko določimo najbolj optimalno izbiro (v našem primeru pot). Najzanesljivejši metodi, ki ju je treba uporabiti za analizo Post Hoc, sta Bonferronijeva metoda ali metoda Tukey HSD (Goeman in Solari, 2021).



Slika 7.18 Nastavitve ANOVA.

Rezultati naše analize kažejo, da je vrednost *F-statistike* 11,173 (višje vrednosti kažejo na večje razlike med skupinami) in *p-vrednost* <0,001, kar pomeni, da je naša ničelna hipoteza zavrnjena (glej sliko 7.19). Ker med tremi potmi obstaja pomembna razlika (<0,001), je post hoc test veljaven tudi v našem primeru (George in Mallery, 2022). Po izvedbi Bonferronijevega korekcijskega testa lahko vidimo, da so najboljše *p-vrednosti* zabeležene v primeru poti 2 (glej sliko 7.19). V poročilu lahko ugotovimo, da "je bila ugotovljena pomembna razlika pri izbiri poti dostave v korelaciji s časom dostave (1-way ANOVA, *F = 11*,173, *df* = 47, *p-vrednost* = <0,001). Pot 2 je imela najboljše rezultate glede časa dostave."

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 7.19 Začetni rezultati ANOVA in rezultati testa Post Hoc.

To poglavje knjige zaključimo z zavedanjem, da smo v tem poglavju obravnavali nekaj najpogostejših testov. Obstajajo še drugi testi, kot so ANOVA s ponovljenimi meritvami, testi zanesljivosti in testi občutljivosti, ki jih je prav tako mogoče modelirati in analizirati s programsko opremo SPSS. Ti dodatni testi zagotavljajo širši nabor orodij za analizo podatkov in prinašajo smiseln vpogled v različne raziskave in praktične aplikacije.

LITERATURA POGLAVJE 7

- 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]