Maja Biljan-August ◆ Snježana Pivac ◆ Ana Štambuk

UPORABA STATISTIKE U EKONOMIJI

2. izdanje

Rijeka, 2009.

UDŽBENICI SVEUČILIŠTA U RIJECI

MANUALIA UNIVERSITATIS STUDIORUM FLUMINENSIS

Prof. dr. sc. Maja Biljan-August
Prof. dr. sc. Snježana Pivac
Doc. dr. sc. Ana Štambuk

UPORABA STATISTIKE U EKONOMIJI
2. IZDANJE

Izdavač:
Ekonomski fakultet Sveučilišta u Rijeci

Recenzentice:
Prof. dr. sc. Jasna Horvat

Doc. dr. sc. Suzana Marković

Doc. dr. sc. Alemka Šegota

Lektorica:
Kerol Musul-Perić, prof.

Autor naslovnice:
Luka Mičetić, dipl. oec.

Pri izradi naslovnice korišteni su materijali objavljeni na:
www.free-stockphotos.com, www.sxc.hu, www.hnb.hr.

Objavljivanje ovog sveučilišnog udžbenika odobrilo je Povjerenstvo za izdavačku djelatnost Sveučilišta u Rijeci Odlukom – klasa: 602-09/09-01/29, ur. broj: 2170-57-05-09-3 od 25. rujna 2009.

Objavljeno na URL: http://www.efri.hr/prikaz.asp?txt_id=6326
i http://oliver.efri.hr/~statist/biljan-pivac-stambuk-uporaba2.pdf.

ISBN: 978-953-6148-86-8

Rijeka, rujan, 2009.

Prof. dr. sc. Maja Biljan-August
Prof. dr. sc. Snježana Pivac
Doc. dr. sc. Ana Štambuk

UPORABA STATISTIKE U EKONOMIJI

EKONOMSKI FAKULTET U RIJECI
RIJEKA, 2009.

MAJA BILJAN-AUGUST
SNJEŽANA PIVAC
ANA ŠTAMBUK

ISBN: 978-953-6148-86-8

PREDGOVOR

Ovaj udžbenik namijenjen je prvenstveno studentima Ekonomskog fakulteta u Rijeci, ali i svim drugim zainteresiranim korisnicima koji u svom stručnom i znanstvenom radu, baveći se društvenim istraživanjima, primjenjuju statističke metode i tehnike.

Rad obuhvaća teorijske osnove i objašnjenja za svako, u ovaj rad, uključeno područje statistike. Kroz rješavanje konkretnih primjera daju se objašnjenja dobivenih rezultata i njihovo kritičko vrednovanje.

Na kraju udžbenika u privitku nalaze se detaljne upute za upotrebu statističkog programa za računala *Statistica*. Ovaj program pruža mnoštvo mogućnosti za provođenje statičkih metoda i tehnika na konkretnim analizama. Naime upotrebom statističkih paketa, počevši već od pripremne faze statističkog istraživanja, znatno se skraćuje i pojednostavljuje vrijeme potrebno za primjenu statističkih metoda i tehnika. Na taj se način statistički postupci približavaju mnogim korisnicima. Svaki primjer prezentiran u poglavljima i u privitku knjige sadrži i rješenja u svrhu kontrole valjanosti usvojenoga gradiva. Na taj se način studentima ne ostavlja dvojbenim ni način pismene provjere znanja, a ujedno ih se osposobljava da samostalno statistički analiziraju određene pojave na stručno zadovoljavajući način.

Studenti koji nastave obrazovanje na poslijediplomskim studijima bit će pripremljeni za stručni i znanstveni rad u svojim istraživanjima, gdje se ekonomski problemi statistički rješavaju upotrebom računala. Dio studenata koji će se nakon diplomiranja neposredno uključiti u poslovnu praksu, imat će koristi zbog mogućnosti upotrebe usvojenog znanja statističke teorijske i programske potpore pri konkretnim poslovnim i ekonomskim analizama.

Potrebno je napomenuti da je program *Statistica* kompatibilan s Microsoft Excelom, stoga se već postojeći podaci iz jednog mogu jednostavno kopirati u drugi program. To je važna činjenica, s obzirom da je poznato da je Microsoft Excel jedan od najraširenijih programa za tablične izračune te je lako dostupan većini korisnika.

U radu su rabljene oznake i simboli koji su preuzeti iz standardne statističke literature. Ako se, pak, u literaturi rabe različite oznake, upotrijebljena je češće spominjana verzija.

Na kraju treba napomenuti da je ovaj udžbenik nastao kao rezultat višegodišnjeg iskustva autorica kod primjene statističkih metoda u izradi brojnih znanstvenih i stručnih radova, studija i analiza te kao rezultat predavačkog iskustva pri prenošenju znanja iz područja statistike na mnoge generacije studenata ekonomskog fakulteta.

Za suradnju i korisne sugestije zahvaljujemo recenzenticama prof. dr. sc. Jasni Horvat, doc. dr. sc. Suzani Marković i doc. dr. sc. Alemki Šegoti.

Rijeka, Split, rujan 2009.

Autorice

UPORABA STATISTIKE U EKONOMIJI

2. IZDANJE

Maja Biljan-August, Snježana Pivac, Ana Štambuk

SADRŽAJ

1. UVOD S DESKRIPTIVNOM STATISTIČKOM ANALIZOM

1.1. Temeljni pojmovi

Pojam statistike mijenjao se kroz povijest. Negdje do 19. stoljeća on je podrazumijevao brojčane i nebrojčane podatke koji su bili od izričite važnosti za jednu državu. Danas s razvojem medija (radio, TV, Internet) postaju dostupne mnoge informacija, stoga vlada potreba za njihovom selekcijom kao i odvajanjem bitnih od onih koje to nisu. U skladu s jasno postavljenim ciljem pri komunikaciji vezanoj za različita područja društvenih aktivnosti: od ekonomije, politike, medicine, sporta i sl., vrši se odabir i analiza prikupljenih podataka.

Statistika je posebna znanstvena disciplina koja u svrhu realizacije postavljenih ciljeva istraživanja na organiziran način prikuplja, odabire, grupira, prezentira i vrši analizu informacija ili podataka, te interpretira rezultate provedene analize.

Da bi ostvarila postavljene ciljeve statistika koristi posebne metode i tehnike. Uz njihovu primjenu u raznim segmentima društva u ekonomiji se statističke metode i tehnike koriste na razini poduzeća i na makroekonomskoj razini. Na razini poduzeća u poslovnoj ekonomiji primjena statistike obuhvaća sve faze poslovnog sustava kao npr. proizvodnju, financije, marketing, planiranje poslovanja. U makroekonomskoj analizi pri kreiranju gospodarske politike statistika se primjenjuje u regionalnoj, nacionalnoj i međunarodnoj ekonomiji.

Statistika se kao znanstvena disciplina može podijeliti na deskriptivnu i inferencijalnu statistiku.

Deskriptivna ili opisna statistika temelji se na potpunom obuhvatu statističkog skupa, čiju masu podataka organizirano prikuplja, odabire, grupira, prezentira i interpretira dobivene rezultate analize. Na taj se način izračunavanjem različitih karakteristika statističkog skupa, sirova statistička građa svodi na lakše razumljivu i jednostavniju formu. Ako se statističke metode i tehnike primjenjuju na čitav statistički skup, dakle ako su istraživanjem obuhvaćeni svi elementi skupa oni tvore populaciju.

Inferencijalna statistika temelji se na dijelu (uzorku) jedinica izabranih iz cjelovitog statističkog skupa, pomoću kojeg se uz primjenu odgovarajućih statističkih metoda i tehnika donose zaključci o čitavom statističkom skupu. Uvijek je prisutan odgovarajući stupanj rizika kada se koriste rezultati iz uzorka, za kojeg je poželjno da bude izabran na slučajan način i da bude reprezentativan. Inferencijalna statistika pripada skupini induktivnih metoda, kojima se izvode zaključci polazeći od općega prema posebnome.

Postoji još podjela statistike na teorijsku i primijenjenu.

Teorijska statistika se ne bavi stvarnim podacima, već definira i nadograđuje opće pojmove i znanstvene okvire.

Primijenjena statistika koristi teorijske i znanstvene statističke pojmove u analizi stvarnih podataka iz različitih područja.

Statističke metode i tehnike temelj su za provođenje statističke analize prirodnih i društvenih pojava.

Predmet proučavanja statistike su određene zakonitosti koje se javljaju u masovnim pojavama. Zadaća statistike je da uoči zakonitosti u masovnim i slučajnim pojavama, te da ih iskaže brojčano.

Masovne pojave su skupine istovrsnih elemenata koji imaju jedno ili više zajedničkih svojstava. Takvu skupinu nazivamo statističkom masom ili statističkim skupom.

Pri definiranju statističkog skupa potrebna je velika preciznost da bi se na temelju takve definicije moglo jednoznačno utvrditi da li neki element pripada ili ne pripada tom skupu.

Statistički skup potrebno je definirati pojmovno, prostorno i vremenski.

Pojmovno odrediti statistički skup podrazumijeva odrediti pojam ili svojstvo svakog elementa promatranog skupa.

Prostorno odrediti statistički skup znači odrediti prostor na koji se odnosi ili kojemu pripadaju elementi statističkog skupa.

Vremenski odrediti statistički skup znači odrediti vremenski trenutak ili razdoblje kojim će se obuhvatiti svi elementi koji ulaze u statistički skup.

Primjer 1.1.1.

Statistički skup, "studenti I. godine Ekonomskog fakulteta u Rijeci u Republici Hrvatskoj u akademskoj godini 2007./2008., na dan 30.10.2008. godine", je vrlo precizno određen i iz takve njegove definicije mogu se odrediti svi njegovi elementi.

Svojstvo svakog elementa odnosno studenta definirano je najprije **pojmovno,** tj. jasno je da je riječ o studentima I. godine Ekonomskog fakulteta.

Elementi ovog skupa određeni su i **prostorno**, tj. nalaze se u Rijeci na području Republike Hrvatske.

Vremenska definicija ovog statističkog skupa upućuje na vremenski interval s kojim su elementi skupa obuhvaćeni tj. akademsku godinu 2006./2007. i vremenski trenutak u kojem je vršena selekcija studenata tj. 30.10.2006. godine.

Statistička obilježja su opće karakteristike elemenata statističkog skupa, po kojima su ti elementi međusobno slični i po kojima se međusobno razlikuju.

Primjer 1.1.2.

Statistički skup, "studenti I godine Ekonomskog fakulteta u Rijeci u Republici Hrvatskoj u akademskoj godini 2007./2008., na dan 30.10.2008. godine", promatran je prema **obilježju** "spol".

Sasvim je jasno da je spol jedna od karakteristika svakog studenta. Postoji muški i ženski spol. Neki studenti su jednakog spola, a neki se razlikuju po spolu.

Općenito se statistička obilježja mogu podijeliti na:

1. **kvalitativna statistička obilježja** koja mogu poprimiti različite oblike, ali se izražavaju opisno. Ako se modalitetima ovog obilježja slučajno pridruže brojevi, s njima nisu dopuštene nikakve računske operacije.

2. **kvantitativna statistička obilježja** koja se izražavaju brojčano.

1. Kvalitativna statistička obilježja se mogu podijeliti na: a) **nominalna statistička obilježja** i b) **redoslijedna statistička obilježja.**

a) **Nominalna statistička obilježja** se izražavaju opisno, a nazivaju se još i **atributivna** statistička obilježja. Primjeri ovog obilježja su: spol, nacionalnost, bračno stanje, i slično. Nominalna obilježja koja mogu poprimiti samo 2 modaliteta nazivaju se alternativna obilježja. Na primjer obilježje spol može poprimiti oblike: muški i ženski.

Posebnu grupu nominalnih statističkih obilježja predstavljaju **zemljopisna statistička obilježja** koja označavaju prostor s kojim su elementi statističkog skupa u vezi. Na primjer: mjesto rođenja, mjesto stanovanja, lokacija podružnica tvrtke prema regijama i slično.

b) **Redoslijedna statistička obilježja** predstavljaju takvu vrstu obilježja koja se mijenjaju prema intenzitetu ili rangu. Primjer takvih obilježja su: uspjeh na testu iz predmeta Statistika (ovdje se javljaju modaliteti ovog obilježja od 1 do 5, ali treba voditi računa o tome da se s njima ne vrše nikakve računske operacije, već se pomoću njih elementi skupa, odnosno učenici, mogu rangirati), stručna sprema zaposlenih u nekom poduzeću i slično.

2. Kvantitativna statistička obilježja mogu se izraziti brojčano. Nazivaju se još i **numerička** statistička obilježja, a mogu se podijeliti na: a) **neprekidna ili kontinuirana statistička obilježja** i b) **prekidna ili diskontinuirana statistička obilježja.**

a) **Neprekidna ili kontinuirana statistička obilježja** su takva numerička obilježja koja mogu poprimiti neprebrojivo beskonačno mnogo vrijednosti (npr. u skupu realnih brojeva, koji je beskonačan, na zatvorenom intervalu od 1 do 2 ima neprebrojivo mnogo elemenata tog skupa). Primjeri takvog obilježja su: visina, težina, duljina, starost itd.

b) **Prekidna ili diskontinuirana statistička obilježja** su takva numerička obilježja koja mogu poprimiti prebrojivo beskonačno mnogo vrijednosti (npr. u skupu cijelih brojeva koji je beskonačan na zatvorenom intervalu od 1 do 2 ima prebrojivo mnogo elemenata tog skupa, tj. 2). Primjeri takvog obilježja su: broj djece, broj učenika u razredu, starost u godinama (godine se mogu prebrojati), visina plaće u kunama itd.

Kvantitativna statistička obilježja su vezana za intervalnu i omjernu skalu. Kod **intervalne skale** položaj nule je unaprijed dogovoren. Nula u ovom slučaju ne znači nepostojanje promatrane pojave. Na primjeru numeričkog obilježja "temperatura zraka" nula (0^0C) ne upućuje na nepostojanje temperature, već upućuje da je hladno. Vrijednosti ove skale se ne mogu dijeliti jer temperatura od 5^0C na nekom području u odnosu na temperaturu od 15^0C na nekom drugom području ne znači da je na jednom mjestu bilo tri puta hladnije u odnosu na drugo mjesto. Kod **omjerne skale** nula podrazumijeva nepostojanje pojave. Na primjeru numeričkog obilježja "visina ušteđevine na računu u banci" nula (0) upućuje na nepostojanje ušteđevine. Ako netko ima ušteđevinu od 20 000 kn, može se reći da on ima dva puta veću štednju u odnosu na nekoga tko ima na računu 10 000 kn.

Postoje još i **vremenska statistička obilježja** koja označavaju trenutak ili vremenski interval s kojim su elementi statističkog skupa u svezi.

1.2. Statistički podaci

Osnovne **faze statističkog istraživanja** su:

a) statističko promatranje

b) grupiranje (tabelarno i grafičko prikazivanje statističkih podataka)

c) statistička analiza i interpretacija rezultata provedene analize.

Statističko promatranje je organizirano prikupljanje statističkih podataka.

Nakon precizne definicije zadatka, cilja i predmeta istraživanja tj. statističkog skupa pristupa se organiziranom prikupljanju statističkih podataka. Uspješnost i objektivnost ovog prvog koraka uvjetuje kvalitetu rezultata ostalih faza statističkog istraživanja. Nepotpune i neistinite prikupljene informacije do kojih bi se došlo u ovoj fazi značile bi da konačan rezultat statističkog istraživanja sadrži pogrješku. Pri tom po**grješka može biti sistematska i slučajna**. Sistematsku pogrješku je lakše uočiti (npr. neispravnost određenog mjernog instrumenta, neistinito izjašnjavanje ispitanika). Slučajnu pogrješku je teško precizno identificirati jer se ona ne javlja kod svakog mjerenja i ne javlja se istim intenzitetom, stoga se kod slučajne pogrješke često veže pretpostavka o poništavanju njenog utjecaja na globalnoj razini promatranja.

U ovisnosti o **karakteru izvora podataka**, statistički podaci se dijele na:

a) sekundarne podatke

b) primarne podatke.

Sekundarni podaci su oni koji se pribavljaju iz već postojećih baza podataka različitih državnih ustanova. Takvi se podaci prikupljaju sustavno na odgovarajući način, a njihov opseg ne ovisi o donošenju neke poslovne odluke ili zadanom cilju nekakvog istraživanja.

Takve podatke u Hrvatskoj prikupljaju: Državni zavod za statistiku, Hrvatska narodna banka, Hrvatska gospodarska komora, te neke druge specijalizirane agencije. Jedan od najčešće korištenih sekundarnih izvora podataka u Hrvatskoj je Statistički ljetopis Hrvatske u izdanju Hrvatskog zavoda za statistiku. U svjetskim okvirima poznat je World Statistical Yearbook, a putem Internet-a su danas dostupne mnoge baze sekundarnih podataka (na primjer: Eurostat, U.S: Census Bureau i slično). Na razini poduzeća, raznovrsna specifična izvješća o poslovanju imaju sekundarni karakter.

Sekundarni podaci su uglavnom brojčani. Predočeni su tablicama, a vrlo često i grafičkim prikazima.

Primarni podaci prikupljaju se neposrednim promatranjem svojstava elemenata statističkog skupa u skladu s unaprijed definiranim ciljevima statističkog istraživanja. Prikupljanje ovih podataka zahtjeva definiranje statističkog skupa, izbor obilježja koja se žele istražiti, određivanje modaliteta promatranog obilježja, pripremanje anketnih upitnika i/ili pratećih formulara te organiziranje i provođenje samog prikupljanja podataka.

Vrlo često istraživanja koja se odnose na svaki član statističkog skupa zahtijevaju velike troškove, stoga se podaci prikupljaju za podskup osnovnog skupa, odnosno za uzorak. Takvo promatranje se naziva **reprezentativno promatranje**.

Prema vremenu promatranje se može podijeliti na **jednokratno, periodično i tekuće**. Jednokratno promatranje provodi se jednom i nema ponavljanja. Periodično promatranje se ponavlja nakon jednakih vremenskih razdoblja (npr. bilanca uspjeha u poduzeću). Za tekuće promatranje mjerenje je kontinuirano (npr. proizvodnja mlijeka).

Prikupljanje primarnih podataka može se vršiti i pomoću **statističkih pokusa**. Statističkim pokusom se vrši mjerenje vrijednosti obilježja nastalih u kontroliranim uvjetima. Vrlo čestu primjenu statistički pokus ima u marketinškom istraživanju tržišta. Na primjer, želi se istražiti kako boja pakiranja deterdženta za rublje ima utjecaja na kupnju. U takvom slučaju istraživač će u pokusu odrediti različite boje pakiranja (vrijednosti kontroliranog obilježja) i u njima će izložiti proizvod. U određenom vremenu vršiti će se mjerenje opsega prodaje. Na taj način dobiveni statistički podaci su primarni.

Pojedinačne metode statističkog promatranja su:

a) mjerenje
b) brojanje
c) ocjenjivanje
d) evidentiranje
e) anketiranje.

Jedan od načina statističkog promatranja i prikupljanja podataka je **mjerenje**. Mjeri se na primjer urod pšenice po jedinici poljoprivredne površine. Može se mjeriti težina proizvoda kao i visina stanovništva nekog područja.

Brojanjem se može doći do podataka o broju zaposlenih u pojedinim organizacijskim jedinicama poduzeća, broju upisanih učenika u srednje škole itd. Broje se i noćenja turista u turističkoj sezoni i na taj se način dobivaju podaci relevantni za analizu uspješnosti sezone.

Ocjenjivanjem kao metodom prikupljanja statističkih podataka određuje se kvaliteta provođenja određenih radnji, stoga se taj postupak obično veže za redoslijedna obilježja statističkog skupa. Ocjenjuju se i rangiraju učenici na testu iz statistike, ocjenjuje se usluga nekog hotelskog poduzeća i slično.

Evidentiranje podataka podrazumijeva kontinuirano praćenje kretanja neke pojave u duljem ili kraćem vremenskom razdoblju. Za potrebe evidencije često se uređuju odgovarajući obrasci. Na primjer pri promatranju izostanaka i broja ostvarenih radnih sati djelatnika u nekom poduzeću postoji obrazac za evidenciju gdje zaposlenici upisuju: ime i prezime, vrijeme dolaska i odlaska s radnog mjesta, vrijeme izostanka, razlog izostanka. Evidentiranje se može vršiti i tehničkim uređajima. To su umrežena računala, optički čitači, uređaji za brojenje npr. putnika i slično.

Anketa ili intervju je metoda kojom se prikupljaju podaci uz pomoć unaprijed pripremljenih upitnika, na kojima ispitanici svojim odgovorima daju informacije o promatranim obilježjima statističkog skupa. Da bi anketa uspjela potrebno je veliku pozornost obratiti sastavljanju upitnika. Sastavlja je statističar, a može se konzultirati i psiholog. Upiti moraju biti kratki, precizni i jasni. Moraju biti postavljeni tako da ne sugeriraju odgovor. Broj pitanja ne smije biti velik da ne zamara one koji odgovaraju. Pri provođenju ankete pristup ispitaniku, odnosno jedinici statističkog skupa može biti izravan i neizravan.

Primjer 1.2.1.

Tijekom svibnja 2001. godine studenti četvrte godine Ekonomskog fakulteta u Splitu željeli su **anketnim upitnikom** doći do podataka koji će pokazati kolika je zainteresiranost učenika i studenata za poduke iz različitih predmeta, o zadovoljstvu postojećim uslugama, frekvencijama pohađanja i slično.

Način komuniciranja bio je osobno i telefonom, a upitnik se sastojao od 10 pitanja:

ANKETNI UPITNIK

Poštovani,

Studenti četvrte godine Ekonomskog fakulteta u Splitu obvezni su u okviru predmeta "Istraživanje tržišta" i "Promocija" izraditi poduzetnički projekt. U svrhu tog projekta provodimo istraživanje o zainteresiranosti učenika i studenata za poduke iz različitih predmeta. Molimo Vas pažljivo pročitajte pitanja i na njih iskreno odgovorite. Ovaj anketni upitnik je u potpunosti anoniman. Hvala na suradnji!

1. Jeste li do sada koristili usluge poduka? (zaokružiti)
   a) Da (prijeći na pitanje broj 3)
   b) Ne

2. Zašto niste koristili usluge poduka?
   a) Nisam imala/imao potrebe za takvim uslugama
   b) Previsoka cijena poduka
   c) Nezadovoljstvo postojećom kvalitetom usluga
   d) Ostalo

   Prijeći na pitanje broj 8.

3. Iz kojih predmeta ste pohađali poduke?
   a) Matematika
   b) Fizika
   c) Engleski jezik
   d) Hrvatski jezik
   e) Ostalo

4. Jeste li zadovoljni postojećom ponudom poduka?
   a) Da (prijeći na pitanje broj 6)
   b) Donekle
   c) Ne

5. Koji je razlog vašem nezadovoljstvu?
   a) Kvaliteta
   b) Cijena
   c) Uslužnost
   d) Lokacija

6. Koliko sati tjedno biste željeli pohađati poduke?
   a) 1 - 2 sata
   b) 3 - 4 sata
   c) 5 i više sati

7. Željeli biste pohađati:
   a) Individualne poduke
   b) Grupne poduke
   c) Svejedno mi je

8. Vaša dob:
   a) 10 - 14 godina
   b) 15 - 19 godina
   c) 20 - 24 godine

9. Spol
   a) žensko
   b) muško

10. Ukupni mjesečni prihodi vašeg kućanstva:
   a) do 2 000 kn
   b) 2 001 – 4 000 kn
   c) 4 001 – 6 000 kn
   d) 6 001 – 8 000 kn
   e) 8 001 i više kn

Opis terenskog rada: Istraživanje je obavljeno tijekom svibnja 2001. godine, i to u poslijepodnevnim satima. Tijekom anketiranja nije bilo nikakvih problema, osim što je manji broj ispitanika odbilo anketiranje pravdajući se žurbom. Ispitanici su bili iznimno susretljivi, stoga smatramo kako su odgovori iskreni i mogu biti reprezentativni.

Izvor: Katedra za marketing, Ekonomski fakultet Split, 2005. godine

Izravan pristup ostvaruje se kada osoba ili tim koji provodi anketu izlaze na teren i u direktnom kontaktu s ispitanicima prikupljaju odgovore na pitanja iz upitnika. **Neizravan pristup** ostvaruje se putem pošte, telefonom i elektroničkom poštom. Na ovaj način smanjuju se troškovi prikupljanja podataka (npr. putni troškovi osoba koje provode anketu tj. anketara). Iako se i na ovaj način ispitanicima prezentira tko provodi i koja je svrha istraživanja, praksa je pokazala da je ovim neizravnim pristupom anketiranja prisutan velik postotak neodaziva kao i često velik postotak nevaljano i nepotpuno popunjenih upitnika. Razloge treba tražiti u činjenici da ispitaniku nije na raspolaganju osoba koja će mu pojasniti nejasna pitanja.

Ovisno o obimu istraživanja **organizaciju prikupljanja podataka može provoditi jedna osoba, skupina istraživača ili čitavo osoblje neke specijalizirane ustanove** kojoj je to osnovna djelatnost. Ako se radi o manjem istraživanju u prikupljanju podataka će sudjelovati manji broj istraživača, dok će veće istraživanje zahtijevati rad većeg broja istraživača.

Prije početka obrade prikupljenih podataka potrebno je izvršiti kontrolu sirove statističke građe. Kontrola se može vršiti tijekom ili na kraju postupka prikupljanja podataka, što ovisi i o različitim metodama prikupljanja.

Preventivna kontrola obavlja se već tijekom samog postupka prikupljanja statističkih podataka. Pri provođenju ankete to podrazumijeva kontrolu upitnika pri njegovom preuzimanju od ispitanika. Kontrolira se da li su dani odgovori na sva pitanja i da li su pravilno popunjena predviđena mjesta za tražene odgovore.

Naknadna kontrola prikupljenih podataka obavlja se nakon postupka prikupljanja. **Formalnom kontrolom** se uspoređuje realizirani broj prikupljenih podataka s onim planiranim obuhvatom statističkog skupa. Ako je anketa vršena neizravnim putem, na primjer poštom, uspoređuje se broj odaslanih upitnika s brojem vraćenih upitnika. **Materijalnom kontrolom** ispituje se potpunost i točnost sadržaja prikupljenih podataka. Na primjeru ankete to podrazumijeva kontrolu potpunosti i logičnosti danih odgovora.

1.3. Uređivanje podataka i tabeliranje

1.3.1. Formiranje statističkih nizova

Prikupljeni statistički podaci u svom izvornom obliku često nisu pregledni, stoga ih je potrebno na odgovarajući način urediti, odnosno grupirati.

Ako se urede prikupljeni statistički podaci prema nekom obilježju ili karakteristici dobiva se statistički niz.

Jedna od najvažnijih metoda uređivanja podataka je metoda grupiranja.

Grupiranje statističkih podataka je postupak diobe statističkog skupa na određeni broj podskupova prema prethodno utvrđenim modalitetima promatranog obilježja i uz poštivanje načela isključivosti i iscrpnosti.

Načelo isključivosti podrazumijeva da svaki element statističkog skupa istovremeno može pripadati samo jednoj grupi tj. podskupu. Načelo iscrpnosti podrazumijeva da postupkom grupiranja trebaju biti obuhvaćeni svi elementi statističkog skupa.

Primjer 1.3.1.

Statistički skup, "studenti I. godine Ekonomskog fakulteta u Rijeci u Republici Hrvatskoj u akademskoj godini 2007./2008., na dan 30.10.2007. godine" **podijeljen je na dvije grupe prema obilježju spol.**

Nakon što su na taj način **uređeni podaci** dobiven je **statistički niz**. Jasno je da jedan student istovremeno može pripadati samo jednoj grupi jer je muškog ili ženskog spola. Takvim uređivanjem elemenata statističkog skupa obuhvaćeni su svi njegovi elementi, odnosno promatrani studenti.

Statistička obilježja, odnosno opće karakteristike elemenata statističkog skupa općenito se označavaju s

X_i, i = 1, 2, …, N. (1.3.1)

Prema tome postoji N modaliteta, tj. pojavnih oblika promatranih obilježja.

Broj elemenata statističkog skupa koji pripadaju određenoj grupi, tj. jednom modalitetu ili pojavnom obliku promatranog obilježja naziva se apsolutna frekvencija.

Oznaka za **apsolutnu frekvenciju** je

f_i, i = 1, 2, …,N. (1.3.2)

Zbroj svih elemenata statističkog skupa naziva se opseg statističkog skupa. Zbog principa isključivosti i iscrpnosti taj broj odgovara i **zbroju svih apsolutnih frekvencija** što je i prikazano izrazom (1.3.3).[1]

$$f_1 + f_2 + f_3 + \ldots + f_N = \sum_{i=1}^{N} f_i \quad (1.3.3)$$

Može se definirati da je **skup uređenih parova modaliteta promatranog obilježja i njima pripadajućih apsolutnih frekvencija statistički niz.**

---

[1] U izrazu (1.3.3) dan je znak zbrajanja $\sum_{i=1}^{N} f_i$ koji se čita: zbroj apsolutnih frekvencija f_i, gdje i ide od 1 do N. Neke od karakteristika znaka zbrajanja ili znaka sume su sljedeće:

1. $\sum_{i=1}^{N} X_i = X_1 + X_2 + \ldots + X_N$, gdje je i = indeks zbrajanja
2. $\sum_{i=1}^{N} (X_i \pm Y_i \pm Z_i \pm \ldots) = \sum_{i=1}^{N} X_i \pm \sum_{i=1}^{N} Y_i \pm \sum_{i=1}^{N} Z_i \pm \ldots$
3. $\sum_{i=1}^{N} X_i Y_i = X_1 Y_1 + X_2 Y_2 + \ldots + X_N Y_N \neq \sum_{i=1}^{N} X_i \cdot \sum_{i=1}^{N} Y_i$ (ovdje se za desni dio nejednakosti, primjenom izraza 1, može zaključiti da suma produkta nije jednaka produktu suma)
4. $\sum_{i=1}^{N} \frac{X_i}{Y_i} = \frac{X_1}{Y_1} + \frac{X_2}{Y_2} + \ldots + \frac{X_N}{Y_N} \neq \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} Y_i}$ (ovdje se može zaključiti da je suma kvocijenta različita od kvocijenta suma)
5. $\sum_{i=1}^{N} aX_i = aX_1 + aX_2 + \ldots + aX_N = a(X_1 + X_2 + \ldots + X_N) = a\sum_{i=1}^{N} X_i$, gdje je "a" konstanta jer ne ovisi o indeksu zbrajanja i.
6. $\sum_{i=1}^{N} (aX_i \pm bY_i \pm cZ_i \pm \ldots) = a\sum_{i=1}^{N} X_i \pm b\sum_{i=1}^{N} Y_i \pm c\sum_{i=1}^{N} Z_i \pm \ldots$, gdje su a, b, c, … konstante.

Grupiranjem statističkih podataka veliki se broj pojedinačnih podataka razvrstava u manji broj, ovisno o broju grupa obilježja. Ako se podaci grupiraju prema modalitetima samo jednog obilježja, tada je riječ o **jednodimenzionalnom grupiranju,** a ako se statistički podaci grupiraju prema modalitetima dvaju ili više obilježja radi se o **dvodimenzionalnom i višedimenzionalnom grupiranju.**

Primjer 1.3.2.

Statistički skup, "studenti I. godine Ekonomskog fakulteta u Rijeci u Republici Hrvatskoj u akademskoj godini 2007./2008., na dan 30.10.2007. godine" grupira se **prema dva statistička obilježja: spol i visina.**

Riječ je o dvodimenzionalnom grupiranju statističkog skupa prema dva statistička obilježja.

Grupiranjem elemenata statističkog skupa gube se pojedinačne informacije o njima, stoga pri grupiranju, tj. formiranju, statističkih nizova istraživač treba biti odgovoran i pažljivo voditi računa o postavljenim ciljevima istraživanja.

1.3.2. Nizovi kvalitativnih podataka

Kvalitativni statistički nizovi su **nominalni** (atributivni i zemljopisni) i **redoslijedni**.

Nominalni statistički nizovi formiraju se grupiranjem elemenata statističkog skupa prema modalitetima odgovarajućeg nominalnog obilježja.

Modaliteti nominalnog obilježja izražavaju se pomoću atributa, kategorija i slovnih oznaka, a mogu se navoditi abecednim redom, prema veličini apsolutnih frekvencija, dogovorno ili zakonski utvrđenim popisima. Na primjer službena statistika koristi: SMTK ili Standardnu međunarodnu trgovinsku klasifikaciju i NKD ili Nacionalnu klasifikaciju djelatnosti.

Zemljopisni statistički nizovi formiraju se grupiranjem elemenata statističkog skupa prema modalitetima odgovarajućeg zemljopisnog ili prostornog obilježja.

Modaliteti prostornog obilježja najčešće odgovaraju teritorijalno-administrativnoj podjeli određenog geografskog prostora, na primjer: gradovi, županije, regije, države, kontinenti. Grupe ovog obilježja mogu biti poredane po abecednom redu, po veličinama apsolutnih frekvencija ili po geografskom položaju. Ako se zbog veće preglednosti nekoliko modaliteta stavi zajedno u grupu "ostalo", ona obično dolazi na kraju statističkog niza.

Geografski statistički nizovi se vrlo često pojavljuju u službenim statističkim publikacijama. Na primjer, u Statističkom ljetopisu Republike Hrvatske sadržana je cjelina "Međunarodni pregled" u kojem je prikazan veliki broj geografskih statističkih nizova prema prostornim obilježjima: države, specifične svjetske regije, kontinenti i slično.

Primjer 1.3.3.

Statistički skup, "Zaposleni u poslovnim subjektima u Republici Hrvatskoj, stanje 31.03.2007. godine" **podijeljen je na grupe prema prostornom obilježju,** odnosno po "Županijama".

Tablica 1.1.

Modaliteti obilježja "Županije"

Županije u Republici Hrvatskoj
Zagrebačka županija
Krapinsko-zagorska županija
Sisačko-moslavačka županija
Karlovačka županija
Varaždinska županija
Koprivničko-križevačka županija
Bjelovarsko-bilogorska županija
Primorsko-goranska županija
Ličko-senjska županija
Virovitičko-podravska županija
Požeško-slavonska županija
Brodsko-posavska županija
Zadarska županija
Osječko-baranjska županija
Šibensko-kninska županija
Vukovarsko-srijemska županija
Splitsko-dalmatinska županija
Istarska županija
Dubrovačko-neretvanska županija
Međimurska županija
Grad Zagreb

Izvor: Statistički ljetopis Republike Hrvatske, 2008. godine, str. 652.

U tablici 1.1. dani su modaliteti, odnosno grupe prostornog ili zemljopisnog obilježja "Županije u Republici Hrvatskoj". Ovo obilježje ima 21 modalitet.

Redoslijedni statistički nizovi formiraju se grupiranjem elemenata statističkog skupa prema modalitetima odgovarajućeg redoslijednog obilježja.

Grupiranje podataka statističkog skupa prema redoslijednom obilježju vrši se slično kao kod nominalnih obilježja, ali je redoslijed grupa ovdje određen rangom koji pojedina grupa obilježja predstavlja. Rangiranje se može vršiti polazeći od najnižeg ranga prema najvišem ili obrnuto.

Primjer 1.3.4.

Statistički skup, "Diplomirani studenti Ekonomskog fakulteta u Rijeci akademske godine 2007./2008." prema obilježju "prosječni uspjeh tijekom studiranja" **podijeljen je na grupe prema redoslijednom obilježju.**

Tablica 1.2.

Modaliteti obilježja "prosječan uspjeh tijekom studiranja"

Prosječan uspjeh
dovoljan
dobar
vrlo dobar
odličan

Izvor: Knjiga matične evidencije Ekonomskog fakulteta u Rijeci 2007./2008. godine

U tablici 1.2. dane su grupe, odnosno modaliteti redoslijednog obilježja "prosječan uspjeh tijekom studiranja" od najnižeg prema najvišem rangu. Ovim grupama obuhvaćeni su svi pojavni oblici promatranog obilježja.

1.3.3. Tabeliranje

Tabeliranje je postupak svrstavanja grupiranih prikupljenih statističkih podataka u tablice.

Statističke tablice kao jedan od oblika prikazivanja statističkih podataka prisutne su u literaturi svuda oko nas.

Tablica nastaje crtanjem okomitih i vodoravnih linija prema određenim pravilima. Svaka **statistička tablica mora imati: naslov, broj tablice (ako ih ima više), tekstualni dio, numerički ili brojčani dio i izvor podataka.**

Tablica 1.3.

Naslov tablice

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="4">Zaglavlje</td><td rowspan="2">Zbirni stupac: $\sum_{j=1}^{n}$</td></tr>
<tr><td>Oznake stupca</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">0</td><td>1</td><td>2</td><td>…</td><td>n</td><td></td></tr>
<tr><td rowspan="4">Pred stu- pac</td><td>Oznake</td><td></td><td>redak</td><td></td><td></td><td></td></tr>
<tr><td>retka</td><td>stu-</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>pac</td><td></td><td>polje</td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td>polje</td><td></td></tr>
<tr><td colspan="2">Zbirni redak: $\sum_{i=1}^{m}$</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Izvor podataka

Tablica 1.3. prikazuje opći oblik statističke tablice.

Naslov tablice mora biti jasan i kratak, a istovremeno u sebi mora sadržavati pojmovnu, prostornu i vremensku definiciju statističkog skupa, da bi onaj tko je čita mogao precizno odrediti njezine elemente.

Tekstualni dio statističke tablice sastoji se od dva dijela: zaglavlja i predstupca. U zaglavlju ili tumaču stupaca opisuje se i objašnjava sadržaj stupaca. U predstupcu ili tumaču redaka opisuje se i objašnjava sadržaj redaka. To su najčešće oblici statističkog obilježja po kojemu je promatran statistički niz. Uz tekstualni dio svaki stupac može biti označen i odgovarajućim brojem. Kako se vidi u tablici 1.3., predstupac se često označava s nulom, a numeriranje dalje ide po redu: 1, 2, …, n.

Brojčani ili numerički dio tablice sastoji se od polja u koja se unose frekvencije, odnosno rezultati grupiranja statističkih podataka. Zbirni ili marginalni stupac sadrži zbrojeve pojedinih redaka. Njegova suma $\sum_{j=1}^{n}$ zbraja elemente svakog retka po j - stupcima. Zbirni ili marginalni redak sadrži zbrojeve pojedinih stupaca. Njegova suma $\sum_{i=1}^{m}$ zbraja elemente svakog stupca po i - redcima.

Izvor podataka se navodi ispod tablice. On omogućuje provjeru ispravnosti prikupljenih podataka u tablici kao i eventualnu dopunu podataka, naravno, ako to zahtjeva statističko istraživanje.

Zbog veće preglednosti tablice, često se zaglavlje, predstupac, zbirni redak i zbirni stupac odvajaju debljim crtama od brojčanog dijela, kao što se može vidjeti u tablici 1.3..

Dakle, **statistička tablica mora poštivati 3 načela: preglednost, potpunost i jasnoću.**

U skladu s tim načelima pri sastavljanju tablice treba voditi računa o tome da ona sadrži što manje redaka i stupaca. Pri upisivanju brojčanih podataka, kod decimalnih brojeva potrebno je smanjiti decimalna mjesta, a kod velikih brojeva odvojiti svake 3 znamenke počevši od decimalnog zareza u lijevo (tisuće, milijune, ...). U tablici treba biti popunjeno svako polje, da onaj koji je čita ne bi bio u dilemi je li neki podatak možda zaboravljen. U statističkim publikacijama postoje posebne oznake za neke kategorije posebnih ili nedostupnih podataka:

*	ispravljen podatak
-	nema pojave
…	ne raspolaže se podatkom
()	nepotpun, odnosno nedovoljno provjeren podatak
0	podatak je manji od 0,5 upotrijebljene jedinice mjere
0,0	podatak je manji od 0,05 upotrijebljene jedinice mjere
1)	oznaka za napomenu ispod tablice

Ovi znakovi upotrebljavaju se i u izvješćima i publikacijama Državnog zavoda za statistiku Republike Hrvatske (Stat. ljetop. Repub. Hrvat. 2008., str. 33).

Statističke tablice mogu se **podijeliti na:**

1) **opće ili izvještajne statističke tablice** i
2) **analitičke ili sumarne statističke tablice**.

Opće statističke tablice prikazuju ogroman broj statističkih podataka o nekome promatranom statističkom skupu.

Analitičke statističke tablice se konstruiraju za neku posebnu analizu i one su u pravilu preglednije.

U ovisnosti o tome prikazuju li se prikupljeni statistički podaci u tablici prema jednom ili više statističkih obilježja i prikazuje li se jedan ili više statističkih skupova **razlikuju se:**

1) **jednostavne statističke tablice** i
2) **složene statističke tablice** koje se mogu podijeliti na:
   - **skupne statističke tablice**, i
   - **kombinirane statističke tablice**.

Jednostavne statističke tablice prikazuju jedan statistički skup, grupiran u jedan statistički niz, prema jednom obilježju.

Skupne statističke tablice prikazuju dva ili više statistička skupa, grupiran u dva ili više statistička niza prema jednom obilježju.

Kombinirane statističke tablice prikazuju jedan statistički skup, grupiran u dva ili više statističkih nizova prema dva ili više obilježja. Ako kod kombinirane tablice s dva obilježja, jedno obilježje ulazi u tablicu iz predstupca, a drugo iz zaglavlja za nju se kaže da je tablica s dva ulaza. U skladu s tim postoje kombinirane tablice s više ulaza, ali pri njihovom kreiranju gubi se jasnoća i preglednost.

Primjer 1.3.5.

Statistički skup, "Diplomirani studenti Fakulteta "E" u Rijeci akademske godine 2007./2008." prema obilježju "prosječni uspjeh tijekom studiranja" **podijeljen je na grupe prema redoslijednom obilježju.**

Tablica 1.4.

Diplomirani studenti Fakulteta "E" u Rijeci godine 2007./2008. prema prosječnom uspjehu tijekom studiranja

Prosječan uspjeh	**Broj studenata**
X_i	f_i
dovoljan	42
dobar	65
vrlo dobar	34
odličan	6
Ukupno: $\sum_{i=1}^{4}$	**147**

Izvor: Knjiga matične evidencije Fakulteta "E" u Rijeci 2007./2008. godine, (simulirani podaci)

U tablici 1.4. prikazan je statistički skup "Diplomirani studenti Fakulteta "E" u Rijeci godine 2007./2008." grupiran prema **redoslijednom obilježju** "prosječni uspjeh tijekom studiranja". U predstupcu su navedene grupe obilježja. Zaglavlje opisuje sadržaj stupaca. U drugom stupcu su dane apsolutne frekvencije, odnosno broj elemenata skupa koji pripadaju pojedinoj grupi obilježja. Suma 2. stupca je zbroj apsolutnih frekvencija i odgovara opsegu statističkog skupa: $\sum_{i=1}^{4} f_i = 147$. Ovo je primjer jednostavne statističke tablice.

1.3.4. Numerički niz

Grupiranjem elemenata statističkog skupa prema numeričkom ili kvantitativnom obilježju nastaje numerički statistički niz.

Grupiranje i uređivanje elemenata numeričkog statističkog skupa ovisi o broju podataka i o njihovoj vrsti, odnosno o tome radi li se o neprekidnim (kontinuiranim) ili prekidnim (diskontinuiranim) numeričkim statističkim obilježjima.

Ako numeričko obilježje može poprimiti samo mali broj numeričkih vrijednosti, tada svaka vrijednost može biti posebna grupa u nizu. To je najčešće slučaj kod prekidnog ili diskontinuiranog numeričkog obilježja.

Uređen negrupiran prekidni numerički statistički niz ima karakteristiku da su njegove vrijednosti složene po veličini počevši, najčešće, od najmanje:

$$X_1, X_2, ..., X_N;\ X_{i-1} < X_i,\ i = 2,3,...,N. \qquad (1.3.4)$$

Općenito je pojedina vrijednost obilježja: $X_i,\ i = 1,2,...,k$, a pripadajuća apsolutna frekvencija: $f_i,\ i = 1,2,...,k$.

Skup uređenih parova:

$$(X_i, f_i),\ i = 1,2,3,...,k \qquad (1.3.5)$$

je distribucija frekvencija promatranog prekidnog numeričkog obilježja.

Također vrijedi da je **suma svih apsolutnih frekvencija jednaka opsegu statističkog skupa:** $\sum_{i=1}^{k} f_i = N$.

Za vrijednosti koje se pojavljuju manje puta formiraju se zajedničke grupe ili "razredi". U razredima se nalaze elementi skupa s **vrijednostima obilježja koje se nalaze između donje i gornje granice razreda**.

Na taj način formiranjem razreda postiže se veća preglednost, ali se gubi na preciznosti i potpunosti informacija. U slučaju **diskontinuiranog numeričkog obilježja donja granica slijedećeg (i+1) razreda razlikuje se od gornje granice prethodnog (i) razreda za jedinicu mjere promatranog obilježja.**

Ove originalne granice razreda nazivaju se još i **nominalne granice**. Prije provođenja statističke analize potrebno je nominalne granice zamijeniti **preciznim ili pravim granicama**. To se najčešće radi na način da se donja granica slijedećeg (i+1) nominalnog razreda umanji, a gornja granica prethodnog (i) nominalnog razreda uveća za polovinu razlike između tih granica. Ponekad se precizne granice formiraju jednostavno na način da se sve gornje granice uvećaju za jednu jedinicu obilježja i tako izjednače sa svojom sljedećom donjom granicom ili da se sve donje granice umanje za jednu jedinicu obilježja i tako izjednače sa svojom prethodnom gornjom granicom.

Kontinuirano ili neprekidno statističko obilježje može poprimiti vrijednosti iz nekog intervala, pa **se njegovi modaliteti formiraju kao razredi**. **Skup uređenih parova razreda obilježja i njima pripadajućih frekvencija**

$$(DG_i \leq X_i < GG_i, f_i),\ i = 1,2,3,...,k \tag{1.3.6}$$

je distribucija frekvencija promatranog neprekidnog numeričkog obilježja. U slučaju **kontinuiranog numeričkog obilježja donja granica sljedećeg (i+1) razreda jednaka je gornjoj granici prethodnog (i) razreda,** pa ovdje nije potrebno računati precizne ili prave granice.

Ako nije definirana najmanja vrijednost obilježja za neki skup, **donja granica početnog razreda** pri grupiranju može se ostaviti **otvorena**. Isto tako ako nije definirana najveća vrijednost obilježja za neki skup **gornja granica posljednjeg razreda** može se ostaviti **otvorena**. Međutim prije početka same statističke analize, ove otvorene granice potrebno je zatvoriti takozvanim **procijenjenim granicama**, koje se stavljaju u zagrade, jer njima istraživač procjenjuje da niti jedan element skupa nema vrijednost manju od prve donje ili veću od zadnje gornje procijenjene granice.

Veličina razreda (i) je razlika između gornje i donje granice razreda (kada je riječ o diskontinuiranom numeričkom obilježju radi se s preciznim ili pravim granicama).

$$i = GG_i - DG_i,\quad i = 1,2,...,k. \tag{1.3.7}$$

Sredina razreda ($sredX_i$) je prosjek, odnosno jednostavna aritmetička sredina, između donje i gornje granice razreda.:

$$sredX_i = \frac{DG_i + GG_i}{2},\ i = 1,2,...,k. \tag{1.3.8}$$

Ako vrijednosti promatranog numeričkog niza nisu ravnomjerno raspoređene, veličine koje se izračunavaju pomoću sredine razreda sadrže tzv. **grešku grupiranja.**

Ponekad se u statističkoj analizi statističkih nizova želi utvrditi koliko elemenata (apsolutno i/ili relativno) promatranog statističkog numeričkog niza ima vrijednost obilježja manju ili veću od neke vrijednosti. Odgovor na to pitanje može se dobiti formiranjem odgovarajućih nizova **kumulativnih apsolutnih i relativnih frekvencija.** Postoje:

a) Kumulativni niz apsolutnih frekvencija "manje od"

b) Kumulativni niz apsolutnih frekvencija "više od"

c) Kumulativni niz relativnih frekvencija "manje od"

d) **Kumulativni niz relativnih frekvencija "više od".**

Kumulativni niz apsolutnih frekvencija "manje od" dobije se postupnim ili sukcesivnim zbrajanjem vrijednosti apsolutnih frekvencija počevši od prve u nizu prema posljednjoj. **Frekvencije** kumulativnog niza apsolutnih frekvencija "manje od" pokazuju koliki broj elemenata promatranog statističkog skupa ima vrijednost obilježja **manju od gornje granice pripadajućeg razreda.**

Kumulativni niz apsolutnih frekvencija "više od" dobije se postupnim ili sukcesivnim zbrajanjem vrijednosti apsolutnih frekvencija počevši od posljednje u nizu prema prvoj. **Frekvencije** kumulativnog niza apsolutnih frekvencija "više od" pokazuju koliki broj elemenata promatranog statističkog skupa ima vrijednost obilježja **višu ili jednaku od donje granice pripadajućeg razreda.**

Kumulativni niz relativnih frekvencija "manje od" dobije se postupnim ili sukcesivnim zbrajanjem vrijednosti relativnih frekvencija počevši od prve u nizu prema posljednjoj. **Frekvencije** kumulativnog niza relativnih frekvencija "manje od" pokazuju koliki udio (ili postotak ako su izražene u %) elemenata promatranog statističkog skupa ima vrijednost obilježja **manju od gornje granice pripadajućeg razreda.**

Kumulativni niz relativnih frekvencija "više od" dobije se postupnim ili sukcesivnim zbrajanjem vrijednosti relativnih frekvencija počevši od posljednje u nizu prema prvoj. **Frekvencije** kumulativnog niza relativnih frekvencija "više od" pokazuju koliki udio (ili postotak ako su izražene u %) elemenata promatranog statističkog skupa ima vrijednost obilježja **višu ili jednaku od donje granice pripadajućeg razreda.**

1.4. Statistička grafika

Uz statističke tablice, pomoćno sredstvo u analizi statističkih nizova su grafički prikazi.

Grafikonima se na jednostavan i pregledan način uz pomoć različitih geometrijskih likova prezentiraju osnovne karakteristike statističkih nizova.

Grafički prikazi statističkih podataka su pregledniji i razumljiviji u odnosu na njihovo prikazivanje statističkom tablicom. Grafikoni omogućuju jednostavnije uočavanje glavnih karakteristika promatranih pojava, ali vrlo često ta preglednost ide na štetu preciznosti statističkih informacija. Stoga je poželjno uz grafički prikaz prezentirati i tablicu s originalnim vrijednostima statističkog niza. Suvremeni statistički programski paketi, naravno u skladu sa statističkom teorijom, imaju mnoštvo mogućnosti kreiranja grafičkih prikaza. Pomoću njih se mogu odabrati različite boje, oblici i linije na grafikonu, što omogućuje još zorniji prikaz promatrane pojave.

Oznake na grafikonu moraju biti takve da onaj tko čita sliku može jasno raspoznati koji su elementi i koja je pojava prikana. Stoga i grafikon mora imati **naslov, jedinice mjere promatranog obilježja, oznake modaliteta obilježja, izvor podataka i po potrebi kazalo ili tumač oznaka.**

Postoje **tri skupine grafičkih prikaza:**

1) **površinski grafikoni**
2) **linijski grafikoni**
3) **kartogrami.**

1.4.1. Grafičko prikazivanje kvalitativnih nizova

Nominalni atributivni statistički nizovi grafički se prikazuju površinskim grafikonima. Površinski grafikoni mogu biti: **jednostavni stupci, dvostruki i razdijeljeni stupci, strukturni stupci, proporcionalni strukturni krugovi i polukrugovi.** Osim nabrojenih mogu se koristiti i neki drugi geometrijski likovi (trokuti, kvadrati i slično), raznih veličina i strukture u skladu s frekvencijama promatranog statističkog niza.

Grafikon ima naslov i izvor podataka, koji odgovara izvoru podataka u tablici po kojoj je konstruiran. Ako se grafikon nalazi na istoj stranici gdje je i njegova tablica, na njemu se može izostaviti izvor podataka.

Pri konstrukciji **jednostavnih stupaca** na apscisi se označavaju svi modaliteti obilježja, a na ordinati je skala apsolutnih frekvencija s aritmetičkim mjerilom, dakle s originalnim jedinicama. Jednostavni stupci imaju jednake baze i jednako su udaljeni jedan od drugog i razlikuju se samo po visini koja odgovara veličini apsolutnih frekvencija. Stoga vrijedi da je i površina stupaca proporcionalna apsolutnim frekvencijama (baza $\times$ visina).

Dvostruki i razdijeljeni stupci upotrebljavaju se za grafičko prikazivanje dvaju ili više statističkih skupova koji su grupirani prema modalitetima istog obilježja. I ovdje vrijedi da se na apscisi označavaju svi modaliteti obilježja, a na ordinati skala apsolutnih frekvencija.

Strukturni stupci također služe za prikazivanje više statističkih skupova podijeljenih na iste grupe prema jednom obilježju. Oni su jednake veličine, a razlikuju se po strukturi.

Proporcionalni strukturni krugovi i polukrugovi također prikazuju dva ili više skupova podijeljena na jednake grupe. Krugovi, odnosno polukrugovi, mogu biti iste veličine i razlikovati se samo po strukturi, a mogu biti i različitih polumjera, tj. proporcionalni veličinama promatranih statističkih skupova.

Izraz za površinu kruga je: $P = r^2\pi$, a ona je proporcionalna opsegu skupa, tj. sumi svih apsolutnih frekvencija: $\sum_{i=1}^{n} f_i$.

U skladu s tim računa se polumjer ili radijus kruga:

$$r = \sqrt{\frac{\sum_{i=1}^{n} f_i}{\pi}} \tag{1.4.1}$$

Na taj način se statistički skupovi mogu uspoređivati po obujmu. Strukturni krugovi razlikuju se po strukturi koja odgovara frekvencijama modaliteta obilježja skupa. Dijelovi strukturnog kruga mogu se izraziti u postotcima (%) ili u stupnjevima (^{0}). Postotci odgovaraju **relativnim frekvencijama**:

$$fr_i = \frac{dio}{cjelina} = \frac{f_i}{\sum_{i=1}^{n} f_i} \quad \text{ili} \quad fr_i = \frac{dio}{cjelina} \cdot 100 = \frac{f_i}{\sum_{i=1}^{n} f_i} \cdot 100 \ (u\%) \tag{1.4.2}$$

Relativne frekvencije imaju svojstva:

a) $0 \le fr_i \le 1, \quad \sum_{i=1}^{n} fr_i = 1$ ili

b) $0 \le fr_i \le 100, \quad \sum_{i=1}^{n} fr_i = 100$. (1.4.3)

Stupnjevi strukturnog kruga ($S^0{}_i$) računaju se:

$$S^0{}_i = \frac{dio}{cjelina} \cdot 360^0 = \frac{f_i}{\sum_{i=1}^{n} f_i} \cdot 360^0 \tag{1.4.4}$$

Ako se konstruiraju **strukturni polukrugovi**, njegovi se stupnjevi računaju prema (1.4.5):

$$S^0{}_i = \frac{dio}{cjelina} \cdot 180^0 = \frac{f_i}{\sum_{i=1}^{n} f_i} \cdot 180^0 \tag{1.4.5}$$

Nominalni prostorni statistički nizovi, uz navedene grafikone, mogu se prikazivati **kartogramima**.

Kartogrami su zemljovidi na kojima se na različite načine (točkama, geometrijskim likovima, slikama i bojama) prikazuje prostorna rasprostranjenost elemenata statističkog skupa.

Postoje **tri vrste kartograma:**

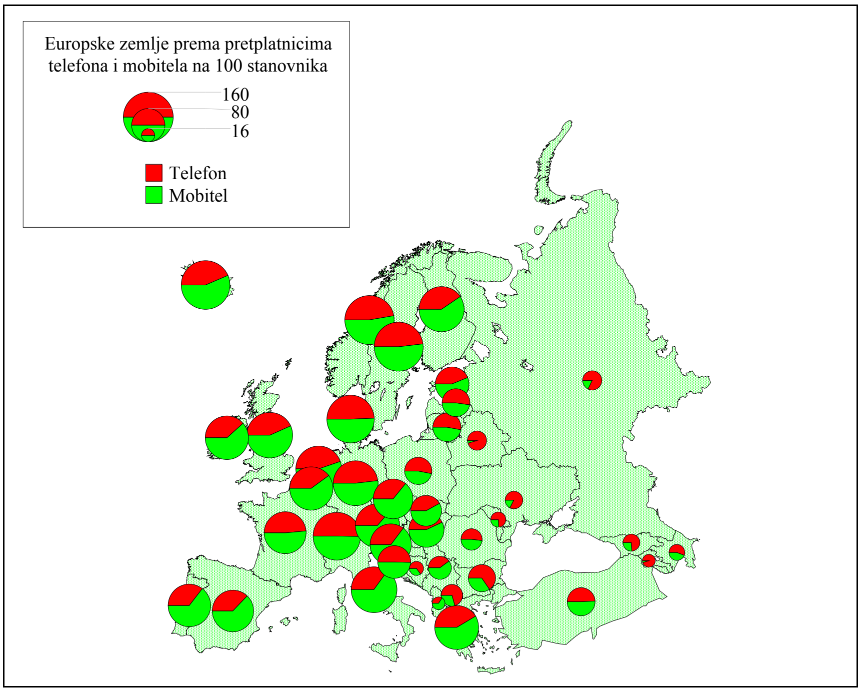
a) dijagramske karte

b) **statističke karte**

c) piktogrami.

Dijagramske karte crtaju se spajanjem zemljovida i površinskih grafikona, na primjer, kvadrata, trokuta, krugova, i slično. Površinski grafikoni, odnosno likovi, moraju biti proporcionalni apsolutnim frekvencijama skupa, čime se izražava intenzitet promatranog obilježja ili pojave. Likovi se ucrtavaju unutar granica površine na zemljovidu koja predočava odgovarajući modalitet prostornog statističkog obilježja, čime se izražava prostorna rasprostranjenost elemenata statističkog skupa. Primjer dijagramske karte:

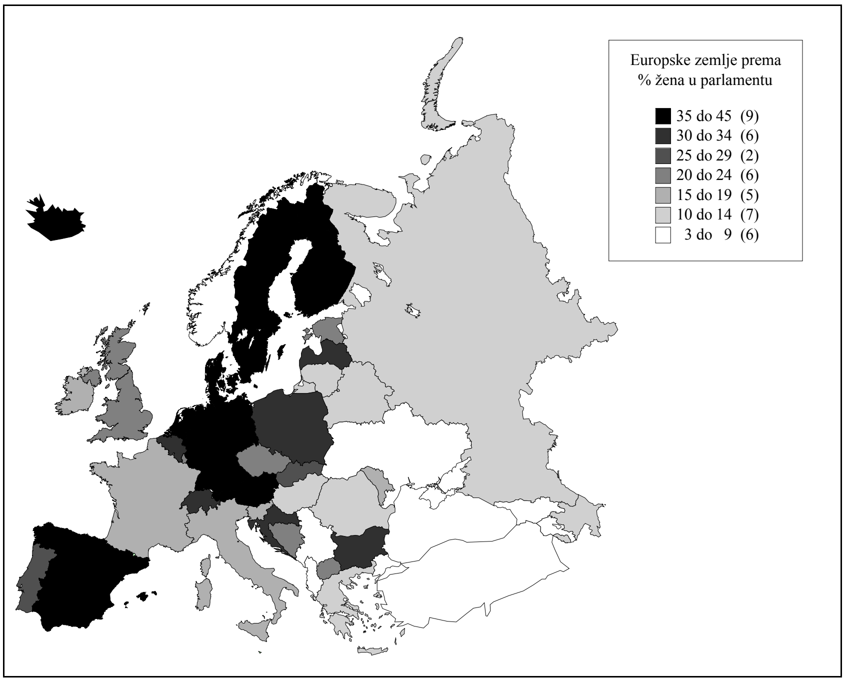
Grafikon 1.



Izvor: Human Developmemnt Report, 2003., str. 274.-277.

Statističke karte crtaju se tako da se na zemljovidu različitim bojama ili sjenčanjem po pojedinim dijelovima nekog područja pokazuje intenzitet neke pojave koji je najčešće izražen relativnim brojevima. Primjer statističke karte:

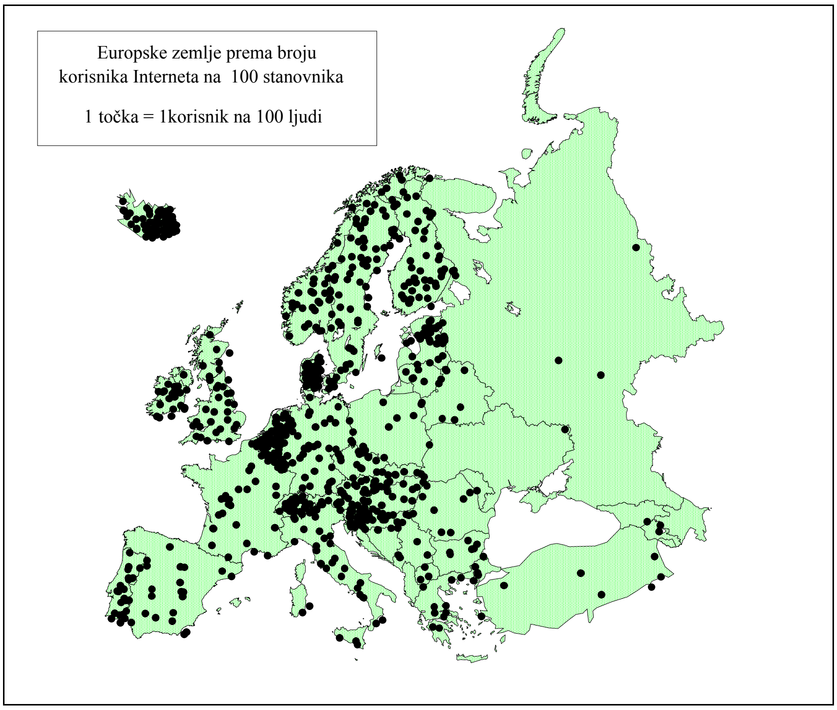
Grafikon 2.



Izvor: Human Developmemnt Report, 2003.god., str. 314.-317.

Piktogrami prostornu rasprostranjenost i intenzitet elemenata statističkog skupa prikazuju gušće ili rjeđe raspoređenim točkama (ili nekim drugim znakovima) na odgovarajućem zemljovidu. Primjer piktograma:

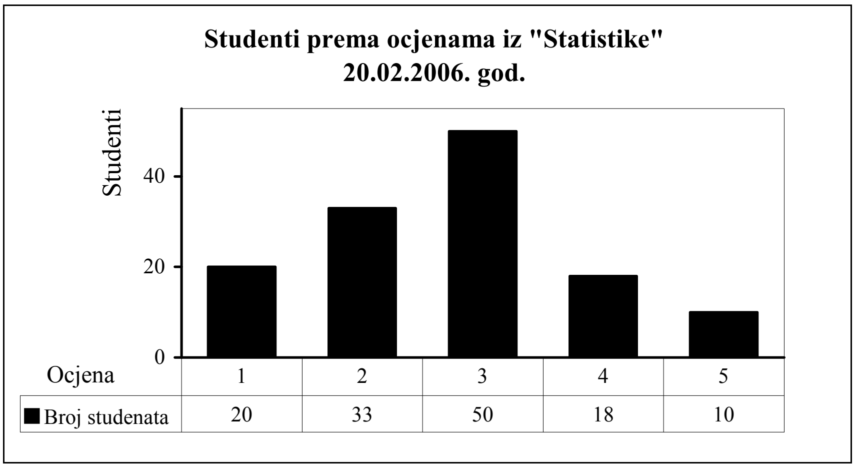
Grafikon 3.



Izvor: Human Developmemnt Report, 2003.god., str. 274.-277.

Redoslijedni statistički nizovi grafički se najčešće prikazuju **jednostavnim stupcima**, a u obzir dolaze i **ostali površinski grafikoni** (strukturni stupci, strukturni polukrugovi, krugovi i slično). Primjer jednostavnih stupaca:

Grafikon 4.



Izvor: Podaci su simulirani.

1.4.2. Grafičko prikazivanje numeričkih nizova

Ako je **prekidni numerički statistički skup negrupiran**, može se grafički prikazati pomoću dvije vrste **grafikona:**

a) **dijagram s točkama**

b) **dijagram "stablo-list"** (ili "Stem and Leaf" dijagram).

Primjer 1.4.1.

Na području "Z" anketiranjem 20 stanova dobiveni su podaci o broju djece po obitelji na dan 31.12.2008. godine:

2 2 0 3 1 1 0 2 1 4 2 1 1 2 5 0 2 1 2 1

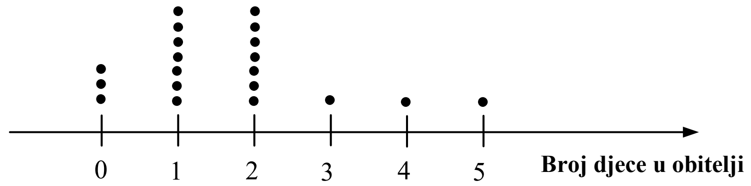
Ovaj skup potrebno je urediti, ali zbog malog broja numeričkih vrijednosti neće se grupirati, već će se podaci složiti od najmanjih vrijednosti obilježja "broj djece" prema najvećima. Dakle vrijednosti obilježja X_i su sljedeće:

0 0 0 1 1 1 1 1 1 1 2 2 2 2 2 2 2 3 4 5

Zadatak je ovaj **negrupirani diskontinuirani numerički niz** prikazati na **dijagramu s točkama:**

Grafikon 5:

Obitelji na području "Z" prema broju djece, stanje 31. 12. 2008. god.



Izvor: Podaci su simulirani.

Dijagram s točkama na grafikonu 10 zorno prikazuje raspored podataka, u ovom slučaju broj djece, u ovih 10 promatranih obitelji.

Primjer 1.4.2.

U poduzeću "Z" pri donošenju organizacijskih odluka 31.01.2009. god. izvršena je analiza djelatnika prema navršenim godinama radnog staža. Prema matičnoj evidenciji svakog od 25 zaposlenika dobiveni su sljedeći podaci:

2 5 14 1 7 10 2 3 4 5 1 3 10 6 3 2 14 9 12 22 5 25 23 2 18

Skup je potrebno urediti, počevši od najmanjih vrijednosti obilježja "navršene godine radnog staža".

1 1 2 2 2 2 3 3 3 4 5 5 5 6 7 9 10 10 12 14 14 18 22 23 25

Zadatak je ovaj **negrupirani diskontinuirani numerički niz** prikazati na **grafikonu "stablo-list":**

Grafikon 6.

Djelatnici poduzeća "Z" prema navršenim godinama radnog staža, stanje 31. 01. 2009. god.

Stablo (Stem)	List (Leaf)
0	1 1 2 2 2 2 3 3 3 4 5 5 5 6 7 9
1	0 0 2 4 4 8
2	2 3 5

Izvor: Podaci su simulirani.

Grafikon "stablo-list" nastaje tako da se svaka znamenka podijeli na dva dijela. Prvi dio predstavlja "stablo", a drugi dio "list". U ovom primjeru znamenke na mjestu desetica označavaju "stablo", dok "list" predstavljaju znamenke na mjestu jedinica. U prvom retku grafikona 11 prikazano je 16 brojeva i to: 01, 01, 02, 02, 02, …, 09. U drugom retku prikazano je 6 podataka: 10, 10, 12, …, 18., a u trećem 3 podatka: 22, 23, 25.

Grupirani numerički (prekidni i neprekidni) statistički nizovi, odnosno njihove distribucije frekvencija **grafički se prikazuju pomoću:**

a) **histograma (površinski grafikoni)**

b) **linijski grafikoni ili poligoni frekvencija.**

Histogram se konstruira pomoću stupaca koji su naslonjeni jedan na drugi, a njihova visina je proporcionalna s odgovarajućim frekvencijama numeričkog niza. Baze stupaca naslonjene su na os apscisu na kojoj je aritmetičko mjerilo promatranog obilježja X.

Pri konstrukciji histograma moguća su dva slučaja. Prvi, kada su **razredi** ili grupe promatranog obilježja **jednake veličine**, tada su i **baze stupaca grafikona** također **jednake, a visine stupaca su proporcionalne originalnim apsolutnim frekvencijama**. U drugom slučaju, **veličine razreda obilježja nisu jednake,** stoga ni **baze stupaca** histograma **nisu jednake**, a **za visinu stupaca potrebno je korigirati originalne apsolutne frekvencije,** dakle **izračunati korigirane frekvencije** fc_i.

$$fc_i = \frac{f_i}{i} \tag{1.4.6}$$

Korigirane frekvencije (1.4.6) **su omjer apsolutnih frekvencija i veličine razreda.** Pri tom veličina razreda (i) može biti originalna, ili toj veličini proporcionalna vrijednost.

Linijski grafikon ili poligon frekvencija konstruira se spajanjem sredina vrhova histograma.

Grafički prikaz kumulativnih nizova jest primjena linijskog grafikona, tj. poligona kumulativnih frekvencija.

Primjer 1.4.3.

Numerički niz «Zaposleni u trgovačkom centru Z prema starosti» prikazan je u tablici 1.5.

Tablica 1.5.

Broj zaposlenih u trgovačkom centru «Z» prema starosti, 2008. god.

Godine starosti (X_i)	**Broj zaposlenih (f_i)**	**Veličine razreda (i)**	**Sredine razreda (X_i)**
20-25	5	5	22,5
25-30	7	5	27,5
30-35	24	5	32,5
35-40	18	5	37,5
40-45	6	5	42,5
45-50	4	5	47,5

50-55	3	5	52,5

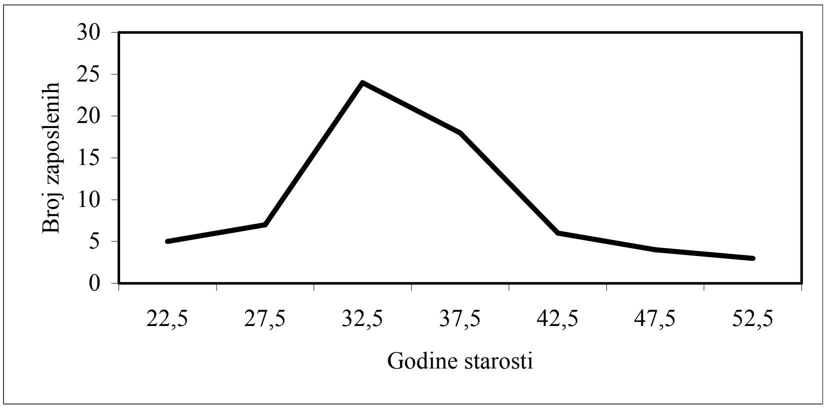
Izvor: Podaci su simulirani.

Zadatak je podatke iz tablice 1.5. prikazati grafički.

S obzirom da su veličine razreda (i) jednake grafički se prikazuju originalne frekvencije (f_i), koje se ucrtavaju iznad sredine razreda:

Grafikon 7.

Broj zaposlenih u trgovačkom centru «Z» prema starosti, 2008. god.



Izvor: Podaci su simulirani.

1.5. Relativni brojevi i njihova primjena

Relativni brojevi su neimenovani, stoga se pomoću njih mogu uspoređivati i analizirati pojave koje imaju različitu jedinicu mjere ili različit broj elemenata. Na taj način dobije se relativna važnost dijela ili cjeline statističkog niza.

Relativni brojevi nastaju dijeljenjem dviju veličina. Veličina s kojom se dijeli zove se osnova relativnog broja i po njoj se relativni brojevi međusobno razlikuju.

Općenito se **relativni brojevi mogu podijeliti na:**

a) **relativne brojeve strukture**
b) **relativne brojeve koordinacije**
c) **indekse** (niza kvalitativnih podataka).

Relativni brojevi strukture pokazuju odnos dijela prema cjelini, i njima se olakšava analiza rasporeda podataka prema modalitetima obilježja u jednom statističkom nizu, odnosno njihova struktura. Najčešće se izražavaju u postotcima, a mogu i u promilima.

Ako relativni brojevi strukture (P_i) pokazuju odnos apsolutnih frekvencija prema opsegu statističkog skupa (tj. ukupnom broju elemenata statističkog skupa) tada se zovu relativne frekvencije (fr_i).

$$P_i = \frac{dio}{cjelina}; \quad fr_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{1.5.1}$$

$$P_i = \frac{dio}{cjelina} \cdot 100 \ (u\%); \quad fr_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \cdot 100 \ (u\%) \tag{1.5.2}$$

$$P_i = \frac{dio}{cjelina} \cdot 1000 \ (u\ prom.); \quad fr_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \cdot 1000 \ (u\ prom.) \tag{1.5.3}$$

Zbroj svih relativnih brojeva u jednom statističkom nizu je 1 ili 100 ili 1000, ovisno o tome na koji je način relativni broj izražen.

Relativni brojevi strukture **grafički** se mogu prikazivati pomoću strukturnih stupaca, strukturnih krugova, polukrugova, ili nekim drugim geometrijskim likom. Pri tom se konstruiraju geometrijski likovi jednakih površina jer je zbroj relativnih frekvencija uvijek isti. Likovi će se razlikovati samo po strukturi.

Primjer 1.5.1.

Statistički niz «Stanovništvo RH prema aktivnosti» prikazan je u tablici 1.6.

Tablica 1.6.

Stanovništvo RH prema aktivnosti, popis 2001. god.

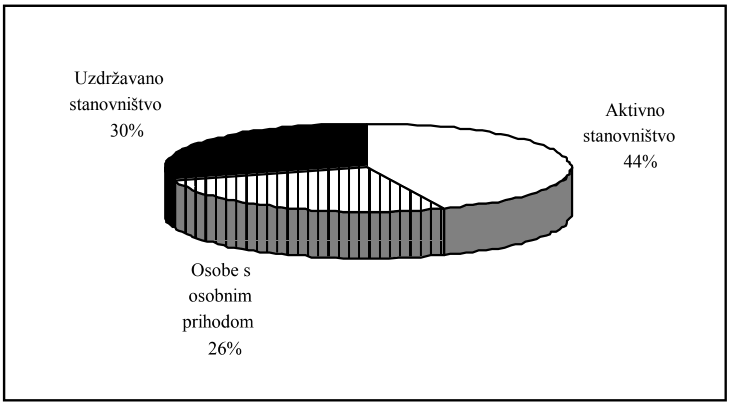
Aktivnost	**Broj stanovnika**	**Postoci (P_i)**
Aktivno stanovništvo	1.952.619	44%
Osobe s osobnim prihodom	1.147.554	26%
Uzdržavano stanovništvo	1.337.287	30%
Ukupno	4.437.460	100%

Izvor: www.dzs.hr.

Zadatak je izračunati postotak stanovništva RH prema aktivnosti i dobivene relativne brojeve strukture prikazati grafički strukturnim krugom:

Grafikon 8.

Struktura stanovništva RH prema aktivnosti, popis 2001. god.



Izvor: www.dzs.hr.

Relativni brojevi koordinacije pokazuju odnos dviju pojava ili frekvencija u različitim statističkim nizovima koje mogu imati različitu jedinicu mjere, a koje ima smisla uspoređivati. Na primjer, to je broj stanovnika po km^2.

Računaju se:

$$R_i = \frac{A_i}{B_i}, \qquad (1.5.4)$$

gdje je:

A_i - veličina koja se uspoređuje

B_i - veličina s kojom se A_i uspoređuje.

Grafički prikaz relativnih brojeva koordinacije je tzv. Varzarov znak. To su stupci, odnosno pravokutnici, čija baza na apscisi odgovara nazivniku (B_i), a visina samom omjeru relativnog broja koordinacije (R_i). Ako se zna da je površina stupca (pravokutnika) baza × visina, može se zaključiti da površina odgovara brojniku (A_i), tj. veličini koja se uspoređuje.

Primjer 1.5.2.

Statistički nizovi "površine i stanovništvo u odabranim zemljama 2001. godine" **prema nominalnom prostornom obilježju** "zemlje", prikazani su u tablici 1.7.

Tablica 1.7.

Površine i stanovništvo u odabranim zemljama 2001. godine

Zemlje	Stanovništvo u 000	Površina u km²	Br.stan./km²
	A_i	B_i	R_i
Hrvatska	4437	56594	78,40054
Slovenija	1930	20273	95,20051
Austrija	8151	83871	97,18496
Mađarska	10106	93032	108,6293
Bosna i Hercegovina	3922	51197	76,60605

Izvor: Statistički ljetopis Republike Hrvatske, 2003. godine, str. 743.

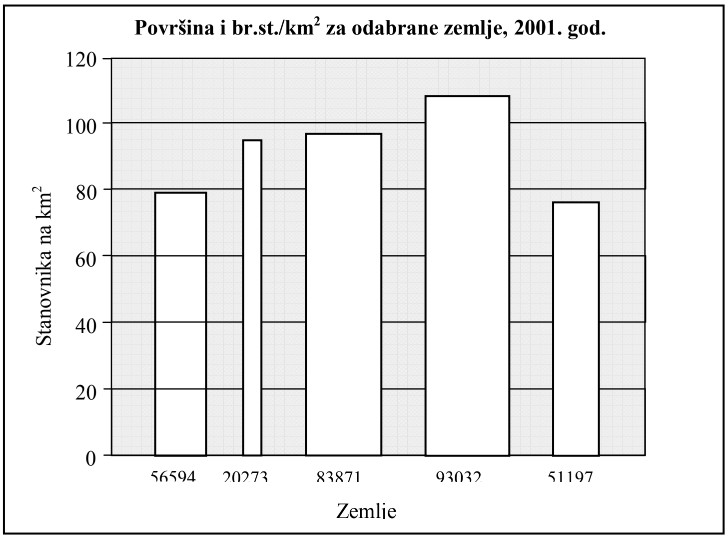
Tablica 1.7. prikazuje 2 različita statistička skupa; površine i stanovništvo u nekim odabranim europskim zemljama. Pomoću tih podataka željela se prikazati gustoća naseljenosti u tim zemljama, odnosno broj stanovnika po km^2. Te vrijednosti su prikazane u posljednjem stupcu tablice. Na primjer, za Republiku Hrvatsku:

$$R_1 = \frac{A_1}{B_1} = \frac{4437000}{56594} = 78{,}40 \text{ stan/km}^2$$

Rezultat se može komentirati: u Republici Hrvatskoj je po km^2 bilo prosječno 78,40 stanovnika u 2001. godini. Najveću gustoću naseljenosti od promatranih zemalja imala je Mađarska (108,63 stan/km^2), a najnižu Bosna i Hercegovina (76,61 stan/km^2).

Ovdje je potrebno napomenuti da je stanovništvo u tablici dano u tisućama, stoga su kod računanja za vrijednosti A_i dodane tisuće.

Grafikon 9.



Izvor: Statistički ljetopis Republike Hrvatske, 2003. godine, str. 743.

Pri konstrukciji Varzarovog znaka na grafikonu 9 na apscisi se baza stupca crta proporcionalno površinama odabranih zemalja, odnosno nazivniku (B_i) relativnog broja koordinacije. U skladu s tim može se uzeti mjerilo da je 1cm=40000km^2. Visina stupaca odgovara broju stanovnika po km^2, tj. vrijednosti (R_i). Površina stupaca je proporcionalna broju stanovnika (A_i).

Indeksima niza kvalitativnih podataka uspoređuje se smjer i intenzitet varijacija frekvencija nekog statističkog niza s takvim varijacijama drugog statističkog niza.

Računaju se tako da se svaki član promatranog niza stavi u odnos prema odabranoj bazi koja može biti jedan od članova niza ili neka druga zadana veličina.

$$I_i = \frac{f_i}{B} \cdot 100 \qquad (1.5.5)$$

Dakle, prema izrazu (1.5.5), indeksi se računaju kao omjer frekvencije promatranog niza i odabrane baze (B).

Grafički prikaz indeksa kvalitativnih nizova su jednostavni stupci koji imaju jednake baze jer se ovdje članovi niza uspoređuju uvijek s istom veličinom. Specifičnost ovog grafikona jest da se stupci naslanjaju na bazu koja je jednaka 100.

Primjer 1.5.3.

Statistički niz «GDP po stanovniku u zemljama EU i RH» prikazan je u tablici 1.8.

Tablica 1.8.

GDP po stanovniku prema kupovnoj moći (*PPP*) u zemljama EU i RH, 2003. god.

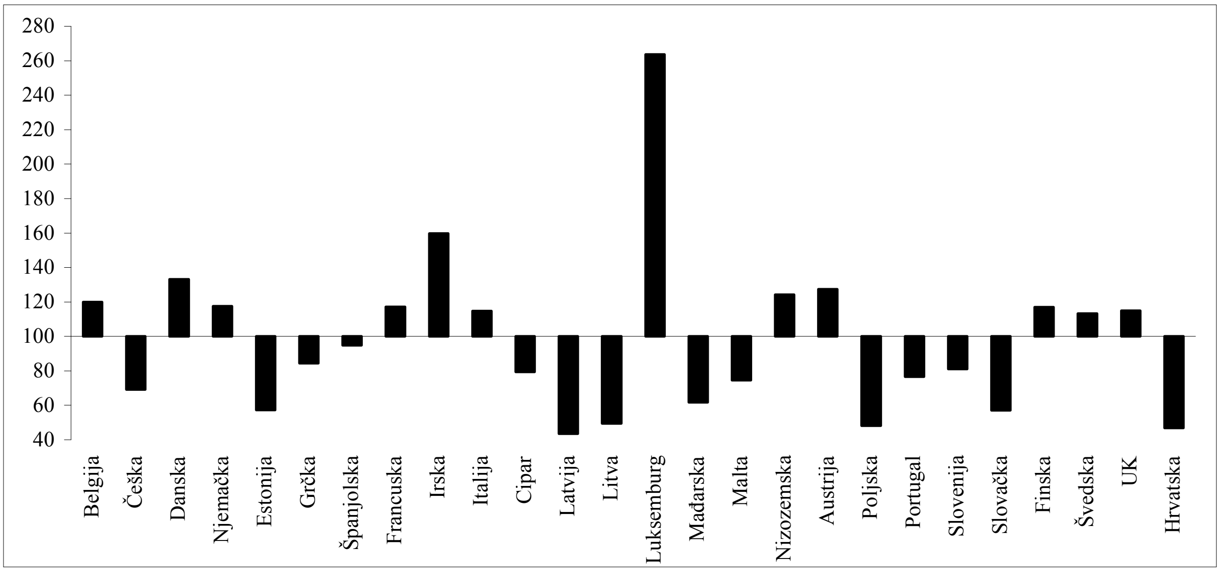
Zemlja (***Geonomenklatura EU***)	**GDP po stanovniku** ***(PPP US$)***	**Indeksi** (***ØEU=100***)
Belgija (BE)	28.335	119,9
Češka (CZ)	16.357	69,2
Danska (DK)	31.465	133,2
Njemačka (GE)	27.756	117,5
Estonija (EE)	13.539	57,3
Grčka (EL)	19.954	84,4
Španjolska (ES)	22.391	94,8
Francuska (FR)	27.677	117,1
Irska (IE)	37.738	159,7
Italija (IT)	27.119	114,8
Cipar ((CY)	18.776	79,5
Latvija (LV)	10.270	43,5
Litva (LT)	11.702	49,5
Luksemburg (LU)	62.298	263,7
Mađarska (HU)	14.584	61,7
Malta (MT)	17.633	74,6
Nizozemska (NL)	29.371	124,3
Austrija (AT)	30.094	127,4
Poljska (PL)	11.379	48,2
Portugal (PT)	18.126	76,7
Slovenija (SI)	19.150	81,0
Slovačka (SK)	13.494	57,1
Finska (FI)	27.619	116,9
Švedska (SE)	26.750	113,2
Ujedinjeno Kraljevstvo (UK)	27.147	114,9
ØEU25	23.629	100,0
Hrvatska (HR)	11.080	46,9

Izvor: Human Development Report, 2005. god., UN, str. 219.-222.; http://epp.eurostat.ec.

Zadatak je izračunati indekse GDP po stanovniku na bazi prosjeka EU i dobivene indekse prikazati grafički stupcima:

Grafikon 10.

Indeksi GDP po stanovniku prema kupovnoj moći u zemljama EU i RH, 2003. god. (ØEU=100)



Izvor: Human Development Report, 2005., UN, str. 219.-222.; http://epp.eurostat.ec.

1.6. Programska potpora za statističku analizu

Prikupljeni statistički podaci nekom od navedenih metoda predstavljaju "sirovu" statističku građu koju je potrebno na odgovarajući način urediti i pripremiti za analizu. Danas se statistički podaci uglavnom obrađuju prikladnim programima za računalo. Postoje suvremeni statistički programski paketi koji omogućavaju, uz vrlo jednostavno rukovanje, unos, pripremu i obradu prikupljenih statističkih podataka. Primjeri suvremenih statističkih paketa su: STATISTICA, SAS, SPSS (Statistical Package for the Social Sciences). Oni se koriste za složenije statističke analize.

Jedan od najpopularnijih softwarea (programskih jezika) za tablične proračune u provođenju različitih aspekata statističke analize je Microsoft Excel. U području programa koji su namijenjeni prvenstveno tabličnim proračunima su Lotus 1-2-3 i QuatroPro, ali upotreba MS Excela je najraširenija.

MS Excel je u funkciji svih bitnih faza procesa statističke analize od formiranja baze podataka, sređivanja i grupiranja podataka, grafičkog prikazivanja statističkih nizova, izračuna temeljnih karakteristika statističkog niza pa sve do složenijih statističkih analiza i procedura i analize vremenskih nizova. Upotrebom statističkih paketa, počevši već od pripremne faze statističkog istraživanja, znatno se skraćuje i pojednostavljuje vrijeme potrebno za primjenu statističkih metoda i tehnika. Na taj se način statistički postupci približavaju mnogim korisnicima.

Prikupljene statističke podatke potrebno je unijeti i pohraniti u odgovarajuću datoteku odabranog statističkog paketa. Kada se unose podaci kvalitativnog ili opisnog karaktera, potrebno je izvršiti šifriranje ili kodiranje vrijednosti takvih obilježja.

Primjer 1.6.1.

Iz ankete u Primjeru 1.2.1. kodirana su pitanja 3., 7. i 9:

3. Iz kojih ste predmeta pohađali poduke?
   a) Matematika
   b) Fizika
   c) Engleski jezik
   d) Hrvatski jezik
   e) Ostalo

7. Željeli biste pohađati:
   a) Individualne poduke
   b) Grupne poduke
   c) Svejedno mi je

9. Spol
   a) žensko
   b) muško

Tablica 1.9.

Kodna lista

OBILJEŽJE	OBLICI OBILJEŽJA	KOD
poduke iz željenog predmeta(V1)	Matematika Fizika Engleski jezik Hrvatski jezik Ostalo	A B C D E
željena vrsta poduka (V2)	Individualne poduke Grupne poduke Nije bitno	A B C
spol (V3)	ženski	1
	muški	2

U tablici 1.9. vidi se da obilježje "poduka iz željenog predmeta" može poprimiti 5 oblika: Matematika, Fizika, Engleski jezik, Hrvatski jezik i ostalo. Svakom obliku obilježja dodijeljena je šifra ili kod: A, B, C, D i E. Na taj način, pri unosu podataka, umjesto da se ispisuju čitave riječi, upisuje se šifra čime se postiže velika ušteda vremena.

Kodovi ne moraju biti slova, stoga se kod obilježja spol za "ženski" unosi 1, a za "muški" 2.

Moderni statistički paketi imaju mogućnosti kreiranja i grupiranja unesenih podataka u različitim željenim varijantama u tablice, što opet ovisi o vrsti istraživanja i postavljenim ciljevima statističkog istraživanja.

Primjer 1.6.2.

Kodirana pitanja iz Primjera 1.6.1. potrebno je za 10 ispitanika unijeti u tablicu:

Tablica 1.10.

Tablica s kodiranim podacima:

Ispitanici	V1	V2	V3
1	A	A	1
2	C	A	2
3	A	B	1
4	B	C	1
5	A	A	2
6	A	B	1
7	D	A	2
8	E	C	1
9	B	A	2
10	A	A	2

U tablici 1.10. vrste statističkih obilježja smještene su u stupce. U drugom stupcu tablice obilježje V1, prema kodnoj listi u tablici 1, znači "poduke iz željenog predmeta". Stupac V2 podrazumijeva "željenu vrstu poduke", a V3 znači "spol".

Ako se pogledaju podaci za 3 ispitanika u tablici 2, A u stupcu obilježja V1 (poduke iz željenog predmeta) znači da taj ispitanik želi poduku iz Matematike. Oznaka B u stupcu V2 (željena vrsta poduka) istog ispitanika podrazumijeva da on preferira grupnu poduku, a njegova oznaka 1 u stupcu V3 (spol) pokazuje da je taj ispitanik ženskog spola.

Takav način obrade podataka olakšava rad s masom statističkih podataka i omogućava veliku uštedu vremena.

Upotreba paketa «Statistica» u statističkoj analizi prikazana je u Privitku.

1.7. Analiza podataka metodama deskriptivne statistike

Kako je već definirano **deskriptivna ili opisna statistika** temelji se na potpunom obuhvatu statističkog skupa, čiju masu podataka organizirano prikuplja, odabire, grupira, prezentira i interpretira dobivene rezultate analize.

Ako se podaci prikupljaju za sve članove nekog skupa, takovo promatranje naziva se **census**. Primjer takvog prikupljanja je popis stanovništva.

Prema samoj definiciji statistike, kao posebne znanstvene discipline koja u svrhu realizacije postavljenih ciljeva istraživanja na organiziran način koristi svoje metode i tehnike, mogu se definirati koraci pri statističkom istraživanju:

1. Definiranje zadatka, cilja i predmeta istraživanja, tj. statističkog skupa

U ovoj početnoj fazi istraživanja, u skladu s postavljenim ciljem, definira se statistički skup, njegove karakteristike ili obilježja. Donosi se i odluka hoće li se koristiti primarni ili sekundarni izvori podataka, tj. hoće li se vršiti neposredno promatranje svojstava elemenata statističkog skupa ili će se pribavljati iz već postojećih baza podataka.

2. Promatranje i analiza prikupljenih podataka

U ovoj fazi vrši se konkretno prikupljanje podataka iz odabranih izvora, te se ocjenjuje kvaliteta takve "sirove" statističke građe. Nakon toga potrebno je prikupljene statističke podatke unijeti i pohraniti u odgovarajuću datoteku odabranog statističkog paketa.

3. Grupiranje, tabelarno i grafičko prikazivanje podataka

U ovom koraku statističkog istraživanja vrši se grupiranje prikupljenih statističkih podataka. Tabelarnim i grafičkim prikazivanjem postiže se jasnija i preglednija prezentacija grupiranih podataka. Uz vizualnu prezentaciju statističkih informacija, tablice i grafikoni služe i kao pomoćno sredstvo istraživanja i analize prikupljenih podataka.

4. a Za podatke koji potpuno obuhvaćaju statistički skup koriste se metode i tehnike deskriptivne ili opisne statistike

U ovoj fazi kada su prikupljeni svi podaci čitavog statističkog skupa, računaju se apsolutni i relativni pokazatelji odnosa unutar jedne ili više pojava.

4. b Za podatke koji obuhvaćaju dio tj. slučajni uzorak statističkog skupa koriste se metode i tehnike inferencijalne statistike

Kada se pri statističkom istraživanju raspolaže samo podacima iz uzorka, računaju se procjene parametara iz čitavog osnovnog statističkog skupa te se vrši testiranje hipoteza o tim parametrima.

5. Interpretacija rezultata i zaključci provedene analize

Nakon deskriptivne analize čitavog statističkog skupa vrši se tumačenje dobivenih brojčanih rezultata i statističkih tablica te se donose zaključci u skladu s postavljenim ciljem statističkog istraživanja.

Nakon inferencijalne analize statističkog skupa na bazi uzorka, opet u skladu s postavljenim ciljevima istraživanja, donosi se izvješće o procijenjenim parametrima iz čitavog osnovnog statističkog skupa i rezultatima postavljenih hipoteza o tim parametrima.

1.8. Mjere centralne tendencije, disperzije, koncentracije, asimetrije i zaobljenost

1.8.1. Mjere centralne tendencije

Računanjem srednjih vrijednosti dolazi se do informacija o vrijednostima statističkog obilježja oko kojih se raspoređuju elementi statističkog niza.

Srednja vrijednost je vrijednost statističkog obilježja oko koje se grupiraju podaci statističkog niza. Još se zove i "mjera centralne tendencije".

Srednje vrijednosti mogu se podijeliti na:

1. **Potpune srednje vrijednosti** računaju se upotrebom svih podataka u statističkom nizu. Potpune srednje vrijednost su:
   a) **aritmetička sredina**
   b) **harmonijska sredina**
   c) **geometrijska sredina.**
2. **Položajne srednje vrijednosti** određuju se položajem podataka u nizu. Najvažnije položajne srednje vrijednosti su :
   a) **medijan**
   b) **mod.**

1.8.1.1. Aritmetička sredina

Aritmetička sredina spada u potpune srednje vrijednosti i računa se upotrebom svih podataka u statističkom nizu.

Aritmetička sredina je omjer zbroja svih vrijednosti numeričkog obilježja jednog niza i broja elemenata tog niza.

Ako se radi o **negrupiranom statističkom numeričkom nizu računa se jednostavna aritmetička sredina**:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1 + X_2 + ... + X_N}{N} \qquad (1.8.1)$$

Primjer 1.8.1.

Statistički skup "odabrani studenti II. godine Ekonomskog fakulteta Rijeka akademske godine 2006./2007." promatrani su prema **prekidnom numeričkom obilježju** "završna ocjena iz kolegija Statistika"[2]. Dobiveni su sljedeći podaci:

2 3 4 3 2 3 3 5 4 3 3 4 2 3 4

2 3 5 3 3 2 2 4 3 4 3 2 2 3 4

Zadatak je **izračunati aritmetičku sredinu**.

U izabranom skupu ima 30 studenata. Prosjek se računa na poznati način, tako da se zbroje ocjene svih učenika, te da se njihov zbroj podijeli s brojem koliko ih ima, odnosno s 30.

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{2+3+4+3+...+3+4}{30} = \frac{93}{30} = 3{,}1$$

Dakle, prosječna završna ocjena iz predmeta Statistika za izabranih 30 studenata je 3,1.

Ako se radi o **grupiranom statističkom numeričkom nizu računa se složena, vagana ili ponderirana aritmetička sredina.**

Poznato je da apsolutne frekvencije označavaju broj elemenata skupa koji pripadaju određenom modalitetu promatranog statističkog obilježja. Iz toga se može zaključiti da je važnost svake vrijednosti obilježja ovisna o frekvenciji

---

[2] Statističko obilježje "završna ocjena iz predmeta Statistika" izražava se brojevima, u ovom slučaju: 2,3,4 i 5, stoga je ovo prekidno numeričko obilježje, iako se može svrstati i u redoslijedna obilježja jer se elementi skupa (tj. učenici) po njemu mogu rangirati.

jer veća frekvencija znači veći udio tog oblika obilježja u promatranom skupu. Zato se frekvencija još naziva težinski faktor ili ponder kod grupiranih nizova.

Vagana ili ponderirana aritmetička sredina računa se prema izrazu:

$$\overline{X} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_k x_k}{f_1 + f_2 + \ldots + f_k} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_k x_k}{N} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} \qquad (1.8.2)$$

U jednadžbi (1.8.2), k predstavlja broj grupa, odnosno modaliteta promatranog numeričkog obilježja. Nazivnik ovog izraza, kao suma svih apsolutnih frekvencija, odgovara opsegu statističkog skupa.

S obzirom da su **relativne frekvencije** upravno proporcionalne apsolutnim frekvencijama, **aritmetička sredina se može izračunati i pomoću njih:**

$$\overline{X} = \frac{fr_1 x_1 + fr_2 x_2 + \ldots + fr_k x_k}{fr_1 + fr_2 + \ldots + fr_k} = \frac{\sum_{i=1}^{k} fr_i x_i}{\sum_{i=1}^{k} fr_i} \qquad (1.8.3)$$

Nazivnik jednadžbe (1.8.3) može biti 1 ili 100, što ovisi o tome jesu li relativne frekvencije dane u decimalnom obliku ili u postotcima.

Ako su veličine razreda grupiranog numeričkog niza različite od 1, za računanje vagane ili ponderirane aritmetičke sredine, potrebno je izvršiti aproksimaciju pomoću sredine razreda. Sredina razreda, kako je već rečeno, dobije se kao jednostavan prosjek donje i gornje granice razreda. Na taj način ona predstavlja sve vrijednosti obilježja koje se javljaju u jednoj grupi. Stoga je, u ovom slučaju, aritmetička sredina procjena stvarne aritmetičke sredine numeričkog niza.

Aritmetička sredina izražava se u originalnim jedinicama mjere promatranog numeričkog obilježja, obuhvaća sve elemente nekog skupa te se pomoću nje mogu uspoređivati nizovi koji su grupirani po jednakom obilježju. Uz ove, aritmetička sredina zadovoljava i sljedeće kriterije:

a) aritmetička sredina nalazi su između najveće i najmanje vrijednosti promatranog numeričkog obilježja:

$$x_{\min} \leq \overline{X} \leq x_{\max} \qquad (1.8.4)$$

b) zbroj odstupanja vrijednosti numeričkog obilježja od aritmetičke sredine u jednoj distribuciji je uvijek nula.

$\sum_{i=1}^{N}(x_i - \overline{X}) = 0$, za negrupirani niz (1.8.5)

$$\sum_{i=1}^{k} f_i(x_i - \overline{X}) = 0 \text{, za grupirani niz} \tag{1.8.6}$$

c) zbroj kvadrata odstupanja vrijednosti numeričkog obilježja od aritmetičke sredine u jednoj distribuciji je manji ili jednak zbroju kvadrata odstupanja vrijednosti numeričkog obilježja u istoj distribuciji od bilo koje druge vrijednosti (a).

$$\sum_{i=1}^{N}(x_i - \overline{X})^2 \le \sum_{i=1}^{N}(x_i - a)^2 \text{, za negrupirani niz} \tag{1.8.7}$$

$$\sum_{i=1}^{k} f_i(x_i - \overline{X})^2 \le \sum_{i=1}^{k} f_i(x_i - a)^2 \text{, za grupirani niz} \tag{1.8.8}$$

Ako neki numerički niz sadrži ekstremno male ili velike vrijednosti obilježja, aritmetička sredina kao prosječna vrijednost gubi na svojoj reprezentativnosti. Taj problem je dodatno izražen kada u distribuciji postoje razredi s otvorenom donjom, odnosno gornjom granicom obilježja, i kada nije moguće te granice objektivno procijeniti.

1.8.1.2. Harmonijska sredina

Harmonijska sredina spada u potpune srednje vrijednosti i računa se upotrebom svih podataka u statističkom nizu.

Harmonijska sredina je recipročna vrijednost aritmetičke sredine recipročnih vrijednosti numeričkog obilježja u jednom nizu.

Ako se radi o **negrupiranom statističkom numeričkom nizu računa se jednostavna harmonijska sredina**:

$$H = \frac{1}{\frac{\sum_{i=1}^{N}\frac{1}{x_i}}{N}} = \frac{1}{\frac{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_N}}{N}} = \frac{N}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_N}} = \frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}} \tag{1.8.9}$$

Ako se radi o **grupiranom statističkom numeričkom nizu računa se složena ili vagana harmonijska sredina**:

$$H = \frac{1}{\frac{\sum_{i=1}^{k}\frac{f_i}{x_i}}{\sum_{i=1}^{k} f_i}} = \frac{1}{\frac{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_k}{x_k}}{\sum_{i=1}^{k} f_i}} = \frac{\sum_{i=1}^{k} f_i}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_k}{x_k}} = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k}\frac{f_i}{x_i}} \tag{1.8.10}$$

Ponderi su i ovdje apsolutne frekvencije. Ako su veličine razreda različite od 1, i za harmonijsku sredinu se radi aproksimacija na način da se računaju sredine razreda x_i.

Harmonijska sredina rijetko se primjenjuje u praksi. Specifično područje njene primjene je izračunavanje srednjih vrijednosti relativnih brojeva koordinacije.

Relativni brojevi koordinacije (R_i) pokazuju odnos dviju pojava koje ima smisla uspoređivati (A_i i B_i).

$$R_i = \frac{A_i}{B_i} \tag{1.8.11}$$

Općenito **srednja vrijednost relativnih brojeva koordinacije je omjer zbroja svih vrijednosti veličine koja se uspoređuje ($\sum_{i=1}^{k} A_i$) i zbroja svih vrijednosti veličine s kojom se uspoređuje veličina iz brojnika ($\sum_{i=1}^{k} B_i$)**:

$$\overline{R} = \frac{\sum_{i=1}^{k} A_i}{\sum_{i=1}^{k} B_i} \tag{1.8.12}$$

Ponekad u konkretnom statističkom istraživanju **nisu dostupni podaci o veličinama sadržanim u relativnom broju koordinacije, stoga su pri izračunavanju njegove srednje vrijednosti moguće 3 situacije:**

a) ako su dostupni podaci o relativnom broju koordinacije (R_i) i o veličini s kojom se uspoređuje veličina iz brojnika (B_i), srednja vrijednost ($\overline{R}$) izračunat će se pomoću izraza za **složenu ili vaganu aritmetičku sredinu:**

$$\overline{R} = \frac{\sum_{i=1}^{k} R_i B_i}{\sum_{i=1}^{k} B_i} \tag{1.8.13}$$

Na temelju osnovne relacije (1.8.12) jasno je da vrijedi, da je veličina iz brojnika: $\sum_{i=1}^{k} A_i = \sum_{i=1}^{k} R_i B_i$. Ponderi su ovdje vrijednosti veličine s kojom se uspoređuje veličina iz brojnika: B_i

b) ako su dostupni podaci o relativnom broju koordinacije (R_i), bez informacija o veličinama u brojniku (A_i) i nazivniku (B_i), srednju vrijednost $\overline{R}$ nije moguće izračunati. Pokušaj aproksimacije ove srednje vrijednosti ne daje vjerodostojan rezultat

c) ako su dostupni podaci o relativnom broju koordinacije (R_i) i o veličini koja se uspoređuje (A_i), srednja vrijednost ($\overline{R}$) izračunat će se pomoću izraza za **složenu ili vaganu harmonijsku sredinu:**

$$\overline{R} = \frac{\sum\limits_{i=1}^{k} A_i}{\sum\limits_{i=1}^{k} \frac{A_i}{R_i}} \qquad (1.8.14)$$

Na temelju osnovne relacije (1.8.12) jasno je da vrijedi, da je veličina iz nazivnika: $\sum\limits_{i=1}^{k} B_i = \sum\limits_{i=1}^{k} A_i / R_i$. Ponderi su ovdje vrijednosti veličine koja se uspoređuje iz brojnika: A_i.

Primjer 1.8.2.

Za nekoliko odabranih fakulteta u Republici Hrvatskoj dani su podaci o broju studenata i broju studenata po nastavniku u akademskoj godini 2002./2003.

Tablica 1.11.

Broj studenata i broj studenata po nastavniku u nekim odabranim fakultetima, u RH u akademskoj godini 2002./2003.

Fakulteti	**Broj studenata (A_i)**	**Broj studenata po nastavniku (R_i)**	**Broj nastavnika (A_i/R_i)=B_i**
PMF	3238	11,64748200	278
Građevinski fakulteti	2480	8,70175439	285
Fakult. el. stroj. i brod.	7603	10,20536910	745
Medicinski fakulteti	3116	3,72281959	837
Ekonomski fakulteti	12217	36,03834810	339
Ukupno:	28654	70,3157732	2484

Izvor: Statistički ljetopis Republike Hrvatske, 2003. godine, str. 447., 448., 450.

Zadatak je izračunati prosječan broj studenata po nastavniku za sve promatrane fakultete zajedno, odnosno **prosjek relativnih brojeva koordinacije ($\overline{R}$).**

S obzirom da je u ovom primjeru poznata (A_i) veličina koja se uspoređuje i relativni brojevi koordinacije (R_i), njihov se prosjek računa pomoću izraza za složenu ili vaganu harmonijsku sredinu, gdje su ponderi (A_i):

$$H = \frac{\sum\limits_{i=1}^{k} f_i}{\sum\limits_{i=1}^{k} \frac{f_i}{x_i}} \Rightarrow \overline{R} = \frac{\sum\limits_{i=1}^{k} A_i}{\sum\limits_{i=1}^{k} \frac{A_i}{R_i}} = \frac{28654}{2484} = 11{,}5354$$

Prosječan broj studenata po nastavniku za sve promatrane fakultete zajedno iznosi 11,5354 studenata po nastavniku.

1.8.1.3. Geometrijska sredina

Geometrijska sredina spada u potpune srednje vrijednosti i računa se upotrebom svih podataka u statističkom nizu.

Geometrijska sredina je N-ti korijen umnoška[3] **svih vrijednosti negrupiranog numeričkog obilježja jednog niza.**

$$G = \sqrt[N]{\prod_{i=1}^{N} x_i} = \sqrt[N]{x_1 \cdot x_2 \cdot ... \cdot x_N} \qquad (1.8.15)$$

U izrazu (1.8.15) dana je **jednostavna geometrijska sredina** koja se koristi za negrupirane statističke numeričke nizove. Ako su nizovi grupirani po razredima računa se **složena ili vagana geometrijska sredina**:

$$G = \sqrt[N]{\prod_{i=1}^{N} f_i x_i} = \sqrt[N]{f_1 x_1 \cdot f_2 x_2 \cdot ... \cdot f_N x_N} \qquad (1.8.16)$$

Ponderi su i ovdje apsolutne frekvencije. Ako su veličine razreda različite od 1, i za geometrijsku sredinu se radi aproksimacija na način da se računaju sredine razreda x_i.

Geometrijska sredina kod numeričkih nizova nema neku logičnu interpretaciju. U **poslovnoj i makroekonomskoj statistici geometrijska sredina se najčešće upotrebljava u analizi vremenskih nizova**.

1.8.1.4. Medijan

Medijan je vrijednost statističkog obilježja koja statistički niz dijeli na dva jednaka dijela.

Medijan se može primijeniti na redoslijedne i kvantitativne statističke nizove, a spada u položajne srednje vrijednosti. Medijan se ne primjenjuje kod nominalnih nizova jer poredak oblika ovog obilježja može biti proizvoljan.

Kod **negrupiranog, a uređenog niza (po veličini vrijednosti obilježja)**, medijan je vrijednost obilježja koja pripada elementu statističkog niza koji se nalazi u sredini niza. Ako je broj elemenata statističkog niza paran, onda se za medijan uzima jednostavan prosjek vrijednosti obilježja dvaju članova koji se nalaze na sredini statističkog niza. (Npr. za niz s 10 podataka,

[3] Znak produkta ili umnoška niza vrijednosti je: $\prod_{i=1}^{N} x_i = x_1 \cdot x_2 \cdot ... \cdot x_N$

nakon što se vrijednosti obilježja poredaju po veličini medijan je prosjek 5. i 6. vrijednosti obilježja.)

Ako je **statistički niz grupiran u razrede**, prije računanja medijana potrebno je izračunati frekvencije kumulativnog niza "manje od" ili "više od", te se u takvom nizu traži središnji član. **Ako se računaju frekvencije kumulativnog niza "manje od" traži se prva frekvencija niza koja sadrži** $N/2$ **. Razred koji odgovara ovoj frekvenciji kumulativnog niza "manje od" je medijalni razred.**

Ako su veličine razreda obilježja jednake 1, vrijednost obilježja koja odgovara odabranoj frekvenciji kumulativnog niza je medijan.

Ako su veličine razreda obilježja različite od 1, medijan se računa po izrazu:

$$Me = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^{m} f_i}{f_{med}} \cdot i \,, \qquad (1.8.17)$$

gdje je:

L_1 - donja prava ili precizna granica medijalnog razreda

$N/2$ - polovina elemenata statističkog niza

$\sum_{i=1}^{m} f_i$ - zbroj svih apsolutnih frekvencija do medijalnog razreda, ne uključujući medijalni razred (tj. kumulativna frekvencija ispred kumulativne frekvencije medijalnog razreda)

f_{med} - apsolutna frekvencija medijalnog razreda

i - originalna veličina medijalnog razreda.

Ako je statistički niz zadan pomoću relativnih frekvencija, medijalna vrijednost ili razred se određuje pomoću prve frekvencije kumulativnog niza koja u sebi sadrži 0,5 ili 50%.

U statističkim distribucijama postoji samo jedan medijan i on se nalazi između najveće i najmanje vrijednosti obilježja. Prednost medijana jest da na njega ne utječu ekstremno male ili velike vrijednosti obilježja, stoga je primjerena srednja vrijednost i kod izrazito asimetričnih distribucija.

1.8.1.4.1. Kvantili

Kvantili su vrijednosti statističkog obilježja koje statistički niz dijele na q jednakih dijelova.

Kvartili su vrijednosti statističkog obilježja koje statistički niz dijele na 4 jednaka dijela. Kvartili se mogu podijeliti na:

a) **donji kvartil (Q_1)**

b) **gornji kvartil (Q_3) .**

Donji kvartil dijeli statistički niz na četiri jednaka dijela u omjeru 1:3, odnosno 25% elemenata statističkog skupa ima vrijednost obilježja manju od donjeg kvartila, a 75% elemenata statističkog skupa ima vrijednost obilježja veću od donjeg kvartila.

Gornji kvartil dijeli statistički niz na četiri jednaka dijela u omjeru 3:1, odnosno 75% elemenata statističkog skupa ima vrijednost obilježja manju od donjeg kvartila, a 25% elemenata statističkog skupa ima vrijednost obilježja veću od gornjeg kvartila.

Kvartili se, slično kao i medijan, mogu primijeniti na redoslijedne i kvantitativne statističke nizove, a određuju se položajem u nizu. Ni oni se ne primjenjuju kod nominalnih nizova jer kako je već naglašeno poredak oblika ovog obilježja može biti proizvoljan.

Kod **negrupiranog, a uređenog niza (po veličini vrijednosti obilježja), donji kvartil** je vrijednost obilježja koja pripada elementu statističkog niza čiji rang (r) tj. mjesto u nizu odgovara:

$Q_1 = x_r$, gdje je $r = INT(N/4) + 1$, ako N nije djeljiv s 4;[4]

$Q_1 = \frac{x_r + x_{r+1}}{2}$, gdje je $r = N/4$, $(r+1) = N/4 + 1$, ako je N djeljivo s 4, računa se prosjek obilježja koji se nalaze na mjestu u nizu: $r \; i \; (r+1)$.

Kod **negrupiranog, a uređenog niza (po veličini vrijednosti obilježja), gornji kvartil** je vrijednost obilježja koja pripada elementu statističkog niza čiji rang tj. mjesto u nizu odgovara:

$Q_3 = x_r$, gdje je $r = INT(3N/4) + 1$, ako N nije djeljiv s 4;

[4] INT predstavlja cijeli dio decimalnog broja dobivenog dijeljenjem (eng. integer). npr.:

$INT(5{,}3) = 5,$

$INT(5{,}7) = 5.$

$Q_3 = \frac{x_r + x_{r+1}}{2}$, gdje je $r = 3N/4$, $(r+1) = 3N/4+1$, ako je N djeljivo s 4, računa se prosjek obilježja koji se nalaze na mjestu u nizu: r i $(r+1)$.

Ako je **statistički niz grupiran u razrede**, prije računanja kvartila potrebno je izračunati frekvencije kumulativnog niza "manje od" ili "više od", te se u takvom nizu traži odgovarajući član. **Ako se računaju frekvencije kumulativnog niza "manje od" za donji kvartil se traži prva frekvencija niza koja sadrži $N/4$. Za gornji kvartil se traži prva frekvencija niza koja sadrži $3N/4$. Razred koji odgovara odabranoj frekvenciji kumulativnog niza "manje od" je razred donjeg, odnosno gornjeg kvartila.**

Ako su veličine razreda obilježja jednake 1, vrijednost obilježja koja odgovara odabranoj frekvenciji kumulativnog niza je donji, odnosno gornji kvartil.

Ako su veličine razreda obilježja različite od 1, donji kvartil se računa po izrazu:

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum_{i=1}^{q} f_i}{f_{k\,\mathrm{var}\,t}} \cdot i, \tag{1.8.18}$$

gdje je:

L_1 - donja prava ili precizna granica razreda donjeg kvartila

$N/4$ - četvrtina elemenata statističkog niza

$\sum_{i=1}^{q} f_i$ - zbroj svih apsolutnih frekvencija do razreda donjeg kvartila, ne uključujući taj razred (tj. kumulativna frekvencija ispred kumulativne frekvencije razreda donjeg kvartila)

$f_{k\,\mathrm{var}\,t}$ - apsolutna frekvencija razreda donjeg kvartila

i - originalna veličina kvartilnog razreda.

Ako su veličine razreda obilježja različite od 1, gornji kvartil se računa po izrazu:

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum_{i=1}^{q} f_i}{f_{k\,\mathrm{var}\,t}} \cdot i, \tag{1.8.19}$$

gdje je:

L_1 - donja prava ili precizna granica razreda gornjeg kvartila

$3N/4$ - tri četvrtine elemenata statističkog niza

$\sum_{i=1}^{q} f_i$ - zbroj svih apsolutnih frekvencija do razreda gornjeg kvartila, ne uključujući taj razred (tj. kumulativna frekvencija ispred kumulativne frekvencije razreda gornjeg kvartila)

$f_{k\,\text{var}\,t}$ - apsolutna frekvencija razreda gornjeg kvartila

i - originalna veličina kvartilnog razreda.

1.8.1.5. Mod

Mod je vrijednost statističkog obilježja koja se najčešće javlja u nekom nizu, tj. vrijednost obilježja kojoj pripada najveća frekvencija.

Mod se može primijeniti na kvalitativne i kvantitativne statističke nizove, a spada u položajne srednje vrijednosti.

Kod nominalnih obilježja mod se određuje brojanjem, traži se vrijednost obilježja u nizu koja se najčešće javlja. Ako je niz grupiran, traži se najveća apsolutna frekvencija. Vrijednost obilježja kojoj pripada ta najveća apsolutna frekvencija je mod.

Primjer 1.8.3.

Statistički skup "prodane cipele u prodavaonici "K" na dan 30.05. 2006. god." grupiran je prema **nominalnom obilježju** "boja". Modaliteti ovog obilježja su: crna (C), bijela (B), ružičasta (R), smeđa (S).

Zadatak je **odrediti mod.**

Uređeni podaci dobiveni evidencijom na određeni dan su:

C C C C B R R R S S

Dakle na dan 30.05.2006. god. u prodavaonici "K" je prodano 10 cipela. Od toga 4 crne, 1 bijele, 3 ružičaste i 2 smeđe. Vrijednost obilježja "boja" koja se u ovom nizu najčešće javlja je crna, pa se može zaključiti da je mod, $Mo = crna\ boja$.

Statistički niz prikazan je u tablici 1.12.

Tablica 1.12.

Prodane cipele u prodavaonici "K" na dan 30.05. 2008. god. prema boji

Boja cipela	**Broj prodanih cipela**
X_i	**f_i**
crna	**4**
bijela	1
ružičasta	3
smeđa	2
Ukupno:	**10**

Izvor: Evidencija prodavaonice "K", svibanj 2008. godine

Kod negrupiranih nizova mod se određuje prema najvećoj apsolutnoj frekvenciji: $f_{\max} = 4$. Modalitet obilježja koji odgovara ovoj frekvenciji je: $Mo = crna\ boja$.

Kod kvantitativnih statističkih obilježja vrijednost moda odgovara vrijednosti obilježja kojoj pripada najveća korigirana frekvencija (fc_i). Ako su veličine razreda ili grupa numeričkog obilježja jednake nije potrebno korigirati frekvencije, već se radi s originalnim apsolutnim frekvencijama. Ukoliko se vrijednost moda ne može točno odrediti primjenjuje se sljedeći izraz za **izračunavanje moda:**

$$Mo = L_1 + \frac{(b-a)}{(b-a)+(b-c)} \cdot i \ , \tag{1.8.20}$$

gdje je:

L_1 - donja prava ili precizna granica modalnog razreda

b - najveća korigirana frekvencija (tj. frekvencija modalnog razreda)

a - korigirana frekvencija ispred frekvencije modalnog razreda

c - korigirana frekvencija iza frekvencije modalnog razreda

i - veličina modalnog razreda.

U izrazu (1.8.20) frekvencije ispred (a) i iza (c) modalnog razreda (b) služe kao ponderi koji pomiču mod od sredine odgovarajućeg modalnog razreda prema njegovoj donjoj ili gornjoj granici.

Osim **unimodalnih nizova** koji imaju samo jedan mod, postoje statistički nizovi u kojima se dvije ili više vrijednosti obilježja mogu pojavljivati češće u odnosu na ostale modalitete obilježja. U tom slučaju kaže se da su to **bimodalni ili multimodalni nizovi**. Kod bimodalne distribucije koja ima dva vrha postoji glavni mod i lokalni mod. U takvom slučaju kada je u nizu prisutno više od jednog moda potrebno je statistički skup podijeliti na više podskupova, od kojih će svaki imati svoja karakteristična svojstva, te izvršiti analizu svakog podskupa posebno.

1.8.1.6. Odnosi među srednjim vrijednostima

Srednje vrijednosti oko kojih se grupiraju podaci statističkog niza ili mjere centralne tendencije određuju se položajem podataka u statističkom nizu kao i obuhvatom svih podataka niza. **Izbor srednje vrijednosti u nekom skupu ovisi o svojstvima elemenata skupa i o vrsti statističkog obilježja po kojem se ti elementi promatraju:**

a) za **nominalna obilježja** može se računati: **mod (Mo)**

b) za **redoslijedna obilježja** može se računati: **mod (Mo) i medijan (Me)**

c) za **numerička obilježja** može se računati: **mod (Mo), medijan (Me) aritmetička sredina ($\overline{X}$), geometrijska sredina (G) i harmonijska sredina (H).**

Elementi numeričkog statističkog niza mogu biti i negativni. U tom slučaju može se računati samo geometrijska (G) i harmonijska (H) sredina.

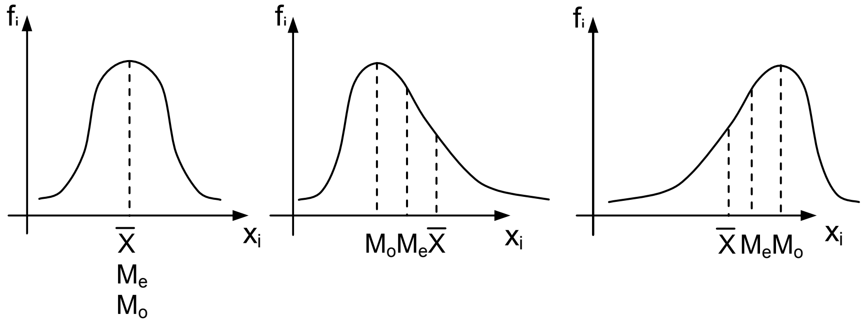
Kod **numeričkih statističkih nizova sa strogo pozitivnim vrijednostima vrijedi sljedeći odnos** između harmonijske, geometrijske i aritmetičke sredine:

$$x_{\min} \leq H \leq G \leq \overline{X} \leq x_{\max} \qquad (1.8.21)$$

Dakle, harmonijska, geometrijska i aritmetička sredina se nalaze između najveće i najmanje vrijednosti numeričkog obilježja, a aritmetička sredina ima najveću srednju vrijednost. Njihove veličine se poklapaju samo onda kada u jednom nizu svi elementi imaju jednake vrijednosti obilježja.

Prema rasporedu po veličini srednjih vrijednosti u jednom nizu može se zaključiti kakav je raspored elemenata skupa. Naime elementi skupa mogu biti ravnomjerno ili simetrično raspoređeni oko srednjih vrijednosti, a mogu biti raspršeni pozitivno tj. na desnu stranu ili negativno tj. na lijevu stranu.

Slika 1.1.



a)	b)	c)
Simetrična distribucija	**Pozitivna ili desnostrana asimetrija**	**Negativna ili lijevostrana asimetrija**

Na slici 1.1. prikazana su 3 slučaja raspršenosti podataka oko srednjih vrijednosti. Na dijelu slike pod (a) prikazana je jedna savršeno **simetrična distribucija u kojoj su aritmetička sredina** (prosjek), **medijan** (vrijednost obilježja koja skup dijeli na dva jednaka dijela) i **mod** (najčešća vrijednost obilježja, koja odgovara vrhu krivulje) u istoj točki, odnosno **jednaki:**

$$\overline{X} = Me = Mo \quad . \qquad (1.8.22)$$

Na dijelu slike pod (b) prikazana je jedna **pozitivno ili desnostrano asimetrična distribucija** (vidi se rasipanje vrijednosti obilježja na desnu stranu) **u kojoj je aritmetička sredina veća od medijana koji je veći od moda**.

$$\overline{X} > Me > Mo \quad . \qquad (1.8.23)$$

Na dijelu slike pod (c) prikazana je jedna **negativno ili lijevostrano** a**simetrična distribucija** (vidi se rasipanje vrijednosti obilježja na lijevu stranu) **u kojoj je aritmetička sredina manja od medijana koji je manji od moda**.

$$\overline{X} < Me < Mo \quad . \qquad (1.8.24)$$

Ako promatrani numerički niz sadrži neke ekstremno male i/ili velike vrijednosti, razumljivo ja da aritmetička sredina kao srednja vrijednost nije dobar pokazatelj. Aritmetička sredina je potpuna srednja vrijednost koja obuhvaća sve elemente skupa, stoga ekstremno velike ili male vrijednosti obilježja mogu znatno utjecati na njenu veličinu. U takvom slučaju, da bi se utvrdile karakteristike skupa, treba odrediti i položajne srednje vrijednosti mod i medijan.

1.8.2. Mjere disperzije

Može se dogoditi da neki statistički skupovi imaju jednake npr. aritmetičke sredine, a da su njihovi elementi potpuno različiti. To znači da je raspored elemenata u tim skupovima različit. **Informaciju o rasporedu elemenata daju mjere raspršenosti ili disperzije elemenata numeričkog statističkog skupa.**

Postoje apsolutne i relativne mjere raspršenosti. Apsolutni pokazatelji izraženi su u originalnim jedinicama mjere i omogućavaju usporedbu nizova prema istom obilježju (**apsolutni pokazatelji su: raspon varijacije, interkvartil, varijanca i standardna devijacija**).

Usporedbu raspršenosti elemenata nizova s različitom mjernom jedinicom omogućuju relativni pokazatelji, koji su najčešće izraženi u postocima (**relativni pokazatelji su: koeficijent kvartilne devijacije i koeficijent varijacije**).

Neke mjere raspršenosti temelje se na dijelu podataka, a neke obuhvaćaju sve elemente promatranog statističkog niza. Stoga se razlikuju:

a) **nepotpune mjere raspršenosti: raspon varijacije, interkvartil i koeficijent kvartilne devijacije**

b) **potpune mjere raspršenosti: varijanca, standardna devijacija i koeficijent varijacije.**

1.8.2.1. Raspon varijacije, interkvartilni raspon i koeficijent kvartilne devijacije

Raspon varijacije je najjednostavnija mjera disperzije, a predstavlja razliku između najveće i najmanje vrijednosti numeričkog obilježja promatranog niza.

$$R = x_{\max} - x_{\min} \tag{1.8.25}$$

Ovaj apsolutni pokazatelj raspršenosti izražen je u originalnim jedinicama mjere numeričkog obilježja. Može poprimiti vrijednost 0. To se događa u slučaju kada svi elementi niza imaju jednaku vrijednost obilježja. Najveća vrijednost ovog pokazatelja nije ograničena jer ona ovisi o konkretnoj raspršenosti promatranih vrijednosti obilježja.

Raspon varijacije je nepotpuna mjera disperzije jer se računa samo na temelju dvije vrijednosti obilježja, odnosno na temelju najveće i najmanje vrijednosti. Može se reći da ovo nije precizna mjera raspršenosti elemenata niza, osobito u slučaju postojanja ekstremno malih i/ili ekstremno velikih vrijednosti obilježja. Tada se dobije veliki raspon varijacije, a možda je većina elemenata skupa raspršena usko oko srednjih vrijednosti.

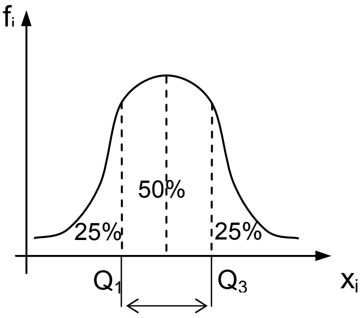
Taj problem preciznosti rješava interkvartilni raspon ili interkvartil.

Interkvartil je apsolutna, nepotpuna mjera raspršenosti, koja pokazuje disperziju srednjih 50% elemenata uređenog numeričkog niza.

$$I_q = Q_3 - Q_1 \tag{1.8.26}$$

Interkvartil predstavlja razliku gornjeg i donjeg kvartila. Na taj način se eliminira 25% ekstremno malih i 25% ekstremno velikih vrijednosti obilježja u nizu.

Slika 1.2.



Slika 1.2. prikazuje jednu simetričnu distribuciju, gdje su elementi skupa ravnomjerno raspoređeni oko srednjih vrijednosti. Poznato je da donji kvartil (Q_1) dijeli distribuciju u omjeru 1:3, tj. da 25% elemenata skupa ima vrijednost obilježja manju od donjeg kvartila, a 75% elemenata skupa ima vrijednost obilježja veću od donjeg kvartila. Isto tako gornji kvartil (Q_3) dijeli distribuciju u omjeru 3:1, tj. da 75% elemenata skupa ima vrijednost obilježja manju od gornjeg kvartila, a 25% elemenata skupa ima vrijednost obilježja veću od gornjeg kvartila. Na taj način interkvartil pokazuje disperziju srednjih 50% elemenata skupa.

Koeficijent kvartilne devijacije je relativna nepotpuna mjera raspršenosti. Predstavlja relativnu disperziju srednjih 50% elemenata numeričkog niza. Računa se na temelju samo dvije vrijednosti obilježja, a to su donji i gornji kvartil.

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} \tag{1.8.27}$$

Koeficijent kvartilne devijacije se računa kao omjer interkvartila (razlika kvartila) i zbroja donjeg i gornjeg kvartila. Vrijedi da je:

$$0 \leq V_q < 1 \tag{1.8.28}$$

Ovaj pokazatelj je jednak 0 samo u slučaju kada nema disperzije. Ako se njegova vrijednost približava 1, to znači da je raspršenost vrijednosti obilježja veća.

S obzirom da je ovo relativni pokazatelj, pomoću njega je moguće uspoređivati raspršenost srednjih 50% elemenata različitih numeričkih distribucija s različitim jedinicama mjere.

1.8.2.2. Varijanca, standardna devijacija i koeficijent varijacije

Varijanca spada u potpune mjere raspršenosti jer obuhvaća sve elemente odabranog numeričkog statističkog niza. Ovaj pokazatelj mjeri odstupanja, tj. raspršenost elemenata skupa od aritmetičke sredine.

S obzirom na poznato svojstvo aritmetičke sredine da je zbroj odstupanja vrijednosti obilježja od aritmetičke sredine u jednoj distribuciji uvijek nula, tj. da je: $\sum_{i=1}^{N}(x_i - \overline{X}) = 0$ (za negrupirani niz) i $\sum_{i=1}^{k} f_i(x_i - \overline{X}) = 0$ (za grupirani niz), varijanca se izražava preko kvadrata ovih odstupanja.

Varijanca je prosječno kvadratno odstupanje vrijednosti numeričkog obilježja od aritmetičke sredine.

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2}{N} = \frac{\sum_{i=1}^{N} x_i^2}{N} - \overline{X}^2 \text{, za negrupirani niz} \tag{1.8.29}$$

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i\left(x_i - \overline{X}\right)^2}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} - \overline{X}^2 \text{, za grupirani niz} \tag{1.8.30}$$

Varijanca je mjera raspršenosti izražena u drugom stupnju, zato kao rezultat daje jedinice mjere numeričkog obilježja na kvadrat, stoga je otežana njena interpretacija.

Standardna devijacija je pozitivan korijen iz varijance i izražena je u originalnim jedinicama mjere. Može se definirati kao **prosječno odstupanje vrijednosti numeričkog obilježja od aritmetičke sredine.**

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2}{N}} = +\sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N} - \overline{X}^2}\text{ , za negrupirani niz} \qquad (1.8.31)$$

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\frac{\sum_{i=1}^{k} f_i\left(x_i - \overline{X}\right)^2}{\sum_{i=1}^{k} f_i}} = +\sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} - \overline{X}^2}\text{ , za grupirani niz} \qquad (1.8.32)$$

Pomoću standardne devijacije u originalnim mjernim jedinicama obilježja može se uspoređivati raspršenost oko aritmetičke sredine nizova koji su grupirani po jednakom obilježju.

Varijanca i standardna devijacija kao apsolutne mjere disperzije ne omogućuju usporedbu disperzije vrijednosti obilježja koje imaju različitu jedinicu mjere. Također mogu uputiti na pogrešan zaključak pri usporedbi disperzije obilježja nizova s različitim pojedinačnim vrijednostima obilježja. U tom slučaju se može računati:

a) standardizirano obilježje

b) relativni pokazatelji disperzije

Standardizirano obilježje (z_i) je linearna transformacija originalnih vrijednosti numeričkog obilježja x_i, a pokazuje odstupanje vrijednosti obilježja od aritmetičke sredine u standardnim devijacijama.

$$z_i = \frac{x_i - \overline{X}}{\sigma},\ i = 1,2,3,\ldots,N. \qquad (1.8.33)$$

Vrijede sljedeća svojstva standardiziranog obilježja:

a) aritmetička sredina standardiziranog obilježja je uvijek jednaka nuli: $\overline{Z} = 0$

b) standardna devijacija standardiziranog obilježja je jednaka jedinici: $\sigma_z = 1$.

Ova svojstva omogućuju usporedbu relativnih položaja elemenata numeričkih nizova.

Ruski matematičar P. L. Čebišev (1821.-1894.) utvrdio je teorem koji definira najmanju proporciju vrijednosti numeričke varijable koja pripada određenom intervalu oko aritmetičke sredine.

Teorem Čebiševa: Najmanja proporcija vrijednosti numeričke varijable koje pripadaju intervalu $(\overline{X} \pm k\sigma)$, gdje je $k > 1$, iznosi: $\left[1 - (1/k)^2\right]$.

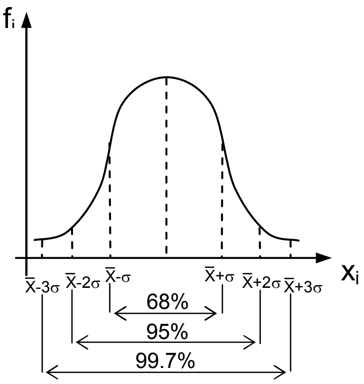
Prema ovom Teoremu vrijedi da se:

- **u intervalu $(\overline{X} \pm \sigma)$ nalazi najmanje 68% svih vrijednosti numeričke varijable**

- **u intervalu** $(\overline{X} \pm 2\sigma)$ **nalazi najmanje 95% svih vrijednosti numeričke varijable**
- **u intervalu** $(\overline{X} \pm 3\sigma)$ **nalazi najmanje 99,7% svih vrijednosti numeričke varijable**

Ako se definira jedna savršeno simetrična distribucija zvonastog oblika ovi odnosi aritmetičke sredine i standardne devijacije mogu se prikazati kao na slici 1.3..

Slika 1.3.



Koeficijent varijacije spada u potpune relativne mjere raspršenosti jer obuhvaća sve elemente odabranog numeričkog statističkog niza, a izražava se u postotcima (%).

Koeficijent varijacije je postotak standardne devijacije od aritmetičke sredine.

$$V = \frac{\sigma}{\overline{X}} \cdot 100 \qquad (1.8.34)$$

Vrijednost koeficijenta varijacije se kreće u intervalu $0 \leq 0 < +\infty$. Vrijednost od 0% će poprimiti samo u slučaju kada su sve vrijednosti numeričkog obilježja u jednom nizu jednake, odnosno kada nema disperzije. Veća vrijednost ovog pokazatelja upućuje na veću disperziju elemenata promatranog niza.

Ovaj pokazatelj raspršenosti je izražen u postotcima, pa omogućuje usporedbu disperzije numeričkih nizova s različitim jedinicama mjere.

1.8.3. Mjere koncentracije

Pomoću pokazatelja koncentracije mjeri se način raspodjele zbroja vrijednosti numeričke varijable (ili neke druge veličine) po elementima statističkog skupa ili po modalitetima promatranog statističkog obilježja.

Mjere koncentracije **se mogu podijeliti na:**

a) **apsolutne mjere koncentracije**

b) **relativne mjere koncentracije**.

Apsolutne mjere koncentracije su:

- **koncentracijski omjer** i
- **Herfindahlov (Hischmanov) indeks.**

Relativna mjera koncentracije je:

- **Ginijev koeficijent koncentracije.**

Koncentracijski omjer računa se za uređen statistički niz po veličini vrijednosti promatranog obilježja počevši od najveće vrijednosti prema najmanjoj:

$$X_1 \geq X_2 \geq ... \geq X_i \geq ... \geq X_N \,, \tag{1.8.35}$$

uz pretpostavku da su sve vrijednosti obilježja pozitivne.

Koncentracijski omjer reda r je:

$$C_r = \frac{\sum_{i=1}^{r} X_i}{\sum_{i=1}^{N} X_i} \,, \tag{1.8.36}$$

ili

$$C_r = \frac{1}{N} \sum_{i=1}^{r} \frac{X_i}{\overline{X}} \,, \tag{1.8.37}$$

gdje je:

X_i - vrijednosti promatranog obilježja

r - red koncentracijskog omjera

N - broj modaliteta promatranog obilježja

$\overline{X}$ - (jednostavna) aritmetička sredina promatranog obilježja.

Vrijednost koncentracijskog omjera kreće se na intervalu: $\frac{1}{N} \leq C_r \leq 1$. Ako su sve vrijednosti promatranog obilježja jednake, omjer iznosi: $1/N$. Ako su sve vrijednosti obilježja, osim posljednje u nizu jednake 0, omjer iznosi: 1.

Koncentracijski omjer r-tog reda se tumači na način da r prvih elemenata skupa zauzima odgovarajući postotak u cijelom statističkom skupu. U skladu s tim u ekonomiji se ovaj omjer često koristi u određivanju stupnja monopola na nekom tržištu.

Herfindahlov indeks je:

$$H = \frac{\sum_{i=1}^{N} X_i^2}{\left(\sum_{i=1}^{N} X_i\right)^2}, \tag{1.8.38}$$

a vrijedi i da je:

$$H = \frac{1}{N}\left(V^2 + 1\right) = \frac{1}{N}\left(\frac{\sigma^2}{\overline{X}^2} + 1\right) \tag{1.8.39}$$

gdje je:

X_i - vrijednosti promatranog obilježja

N - broj modaliteta promatranog obilježja

V - koeficijent varijacije zadanog niza (nepomnožen sa 100)

σ^2 - varijanca zadanog niza

$\overline{X}$ - aritmetička sredina promatranog obilježja.

Vrijednost Herfindahlov indeksa kreće se na intervalu: $\frac{1}{N} \leq H \leq 1$. Ako su sve vrijednosti promatranog obilježja jednake omjer iznosi: $1/N$. Ako su sve vrijednosti obilježja, osim posljednje u nizu jednake 0, omjer iznosi: 1 (koncentracija je najveća).

Herfindahlov indeks ne može se numerički uspoređivati s koncentracijskim omjerom jer se temelje na različitim polazištima.

Ginijev koeficijent koncentracije je relativna mjera koncentracije. U konačnoj interpretaciji ovaj pokazatelj se spaja s **Lorenzovom krivuljom** koja nastaje u prvom kvadrantu koordinatnog sustava spajanjem točaka od točke $(0{,}0)$ do točke $(1{,}1)$ na sljedeći način:

$$(0,0);\left(\frac{1}{N},\frac{X_1}{\sum\limits_{i=1}^{N}X_i}\right);\ ...\ ;\left(\frac{I}{N},\frac{\sum\limits_{i=1}^{I}X_i}{\sum\limits_{i=1}^{N}X_i}\right);\ ...\ ;(1,1). \tag{1.8.40}$$

Dakle druga koordinata svake točke dobije se tako da se svaki član kumulativnog niza «manje od» podijeli s totalom.

Pretpostavlja se da su vrijednosti obilježja za koje se mjeri koncentracija pomoću ovog pokazatelja nenegativne i poredane po veličini na način:

$$X_1 \le X_2 \le ... \le X_i \le ... \le X_N, \qquad X_i \ge 0,\ \forall i. \tag{1.8.41}$$

Na istu sliku se ucrtava i pravac jednolike distribucije koji prolazi točkama $(0,0)$ i $(1,1)$ i koji ustvari predstavlja dijagonalu spomenutog kvadrata. Taj pravac dijeli površinu kvadrata na dva jednaka dijela, stoga se može reći da je polovina te površine ispod pravca jednaka 0,5.

Što je koncentracija u promatranom statističkom skupu veća, to je Lorenzova krivulja udaljenija od pravca. Ako se od 0,5 oduzme površina ispod Lorenzove krivulje dobiva se temelj ove mjere koncentracije.

Površina ispod Lorenzove krivulje može se razdijeliti na 1 trokut i *N-1* trapez, pa se ukupna površina ispod krivulje računa kao zbroj površina trokuta i trapeza.

Površina između pravca jednolike distribucije i Lorenzove krivulje varira između $[0,0.5]$. Ako se dobiveni rezultat pomnoži s 2, tada on varira u intervalu $[0,1]$. Za zadani niz Ginijev koeficijent se računa:

$$G=\frac{2\sum\limits_{i=1}^{N}i\cdot X_i-(N+1)\sum\limits_{i=1}^{N}X_i}{N\sum\limits_{i=1}^{N}X_i}, \tag{1.8.42}$$

a nakon normiranja vrijedi da je:

$$G^*=G\frac{N}{N-1}, \qquad 0\le G^*\le 1, \tag{1.8.43}$$

gdje je:

X_i - vrijednosti promatranog obilježja

N - broj modaliteta promatranog obilježja.

1.8.4. Momenti numeričkih nizova

Momenti numeričkih nizova mogu biti **glavni i pomoćni.**

1.8.4.1. Glavni momenti numeričkih nizova

Glavni ili centralni momenti numeričkih nizova su momenti oko aritmetičke sredine:

$$\mu_r = \frac{\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^r}{N}, \quad r = 0,1,2,... \qquad \text{- za negrupirani niz.} \qquad (1.8.44)$$

$$\mu_r = \frac{\sum_{i=1}^{k} f_i\left(x_i - \overline{X}\right)^r}{\sum_{i=1}^{k} f_i}, \quad r = 0,1,2,... \qquad \text{- za grupirani niz.} \qquad (1.8.45)$$

Centralni moment nultog reda: $\mu_0 = 1$.

Centralni moment prvog reda: $\mu_1 = 0$.

Centralni momenti oko aritmetičke sredine mogu se izraziti **preko pomoćnih momenata.**

Centralni moment drugog reda:

$$\mu_2 = m_2 - m_1^2 . \qquad (1.8.46)$$

Centralni moment trećeg reda:

$$\mu_3 = m_3 - 3m_1 m_2 + 2m_1^3 . \qquad (1.8.47)$$

Centralni moment četvrtog reda:

$$\mu_4 = m_4 - 4m_3 m_1 + 6m_2 m_1^2 - 3m_1^4 . \qquad (1.8.48)$$

1.8.4.2. Pomoćni momenti numeričkih nizova

Pomoćni momenti numeričkih nizova su momenti oko nule:

$$m_r = \frac{\sum_{i=1}^{N} x_i^r}{N}, \quad r = 0,1,2,... \qquad \text{- za negrupirani niz.} \qquad (1.8.49)$$

$$m_r = \frac{\sum_{i=1}^{k} f_i x_i^r}{\sum_{i=1}^{k} f_i}, \quad r = 0,1,2,... \qquad \text{- za grupirani niz.} \qquad (1.8.50)$$

Pomoćni moment nultog reda: $m_0 = 1$.

Pomoćni moment prvog reda: $m_1 = \overline{X}$.

1.8.5. Mjere asimetrije

Asimetrija distribucije podrazumijeva nagnutost distribucije na lijevu ili desnu stranu.

1.8.5.1. Pearsonov koeficijent asimetrije

Pearsonov koeficijent asimetrije je:

$$\alpha_3 = \frac{\mu_3}{\sigma^3}, \qquad (1.8.51)$$

gdje je:

μ_3 - centralni moment trećeg reda

σ^3 - standardna devijacija na treću potenciju.

Interval u kojem se kreće vrijednost ovog koeficijenta:

$$-2 \leq \alpha_3 \leq 2. \qquad (1.8.52)$$

U slučaju izrazito asimetričnih distribucija ovaj koeficijent može poprimiti i vrijednosti izvan intervala: $[-2,2]$.

Ako je:

$\alpha_3 = 0$	$\Rightarrow$	simetrična distribucija
$0 < \alpha_3 \leq 2$	$\Rightarrow$	pozitivna ili desnostrana asimetrija
$-2 \leq \alpha_3 < 0$	$\Rightarrow$	negativna ili ljevostrana asimetrija.

1.8.5.2. Pearsonova mjera asimetrije

Pearsonova mjera asimetrije je:

$$S_k = \frac{(\overline{X} - Mo)}{\sigma} \quad , \tag{1.8.53}$$

gdje je:

$\overline{X}$ - aritmetička sredina

Mo - mod

σ - standardna devijacija.

Za umjereno asimetrične distribucije vrijedi da je **Pearsonova mjera asimetrije:**

$$S_k = \frac{3 \cdot (\overline{X} - Me)}{\sigma}, \tag{1.8.54}$$

gdje je:

$\overline{X}$ - aritmetička sredina

Me - medijan

σ - standardna devijacija.

Ako je:

$S_k = 0$	$\Rightarrow$	simetrična distribucija
$S_k > 0$	$\Rightarrow$	pozitivna ili desnostrana asimetrija
$S_k < 0$	$\Rightarrow$	negativna ili ljevostrana asimetrija.

Odnosi među srednjim vrijednostima kod različitih distribucija:

a) **simetrična distribucija** ima jednaku aritmetičku sredinu, medijan i mod:

$$\overline{X} = Me = Mo \tag{1.8.55}$$

b) **pozitivno ili desnostrano** a**simetrična distribucija** ima aritmetičku sredinu veću od medijana koji je veći od moda.

$$\overline{X} > Me > Mo \tag{1.8.56}$$

c) **negativno ili ljevostrano** a**simetrična distribucija** ima aritmetičku sredinu manju od medijana, koji je manji od moda.

$$\overline{X} < Me < Mo\ . \tag{1.8.57}$$

1.8.5.3. Bowleyeva mjera asimetrije

Bowleyeva mjera asimetrije zasnovana je na odnosu medijana i kvartila:

$$S_{kq} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1}, \tag{1.8.58}$$

gdje je:

Me - medijan

Q_1 - donji kvartil

Q_3 - gornji kvartil.

Ako je:

$S_{kq} = 0 \Rightarrow \quad (Q_3 + Q_1) = 2Me \Rightarrow$

$(Q_3 - Me) = (Me - Q_1) \Rightarrow$ simetrična distribucija

$S_k > 0 \Rightarrow \quad (Q_3 + Q_1) > 2Me \Rightarrow$

$(Q_3 - Me) > (Me - Q_1) \Rightarrow$ pozitivna ili desnostrana asimetrija

$S_k < 0 \Rightarrow \quad (Q_3 + Q_1) < 2Me \Rightarrow$

$(Q_3 - Me) < (Me - Q_1) \Rightarrow$ negativna ili ljevostrana asimetrija.

1.8.6. Mjera zaobljenosti

Mjera zaobljenosti predstavlja zaobljenost vrha krivulje distribucije frekvencija:

$$\alpha_4 = \frac{\mu_4}{\sigma^4}, \tag{1.8.59}$$

gdje je:

μ_4 - centralni moment četvrtog reda

σ^4 - standardna devijacija na četvrtu potenciju.

Ako je:

$\alpha_4 = 3$	$\Rightarrow$	normalno zaobljena distribucija
$\alpha_4 > 3$	$\Rightarrow$	šiljatiji vrh od normalno zaobljene distribucije
$1{,}8 < \alpha_4 < 3$	$\Rightarrow$	tupi oblik distribucije
$\alpha_4 = 1{,}8$	$\Rightarrow$	pravokutni oblik distribucije
$\alpha_4 < 1{,}8$	$\Rightarrow$	U - distribucija.

Primjer 1.8.4.

Prodaja kruha u trgovini «Z» tijekom radnih dana jednog tjedna u srpnju 2008. god. bila je sljedeća:

Tablica 1.13.

Prodaja kruha u trgovini «Z»

Radni dani	**Prodano kruha u kg (x_i)**	$x_i - \overline{X}$	$\left(x_i - \overline{X}\right)^2$
Ponedjeljak	56	9,9	98,01
Utorak	39	-7,1	50,41
Srijeda	61	14,9	222,01
Četvrtak	44	-2,1	4,41
Petak	52	5,9	34,81
Subota	48	1,9	3,61
Nedjelja	23	-23,1	533,61
Ukupno	323	-	946,87

Izvor: Evidencija trgovine «Z», srpanj 2008.god.

Zadatak je izračunati:
a) prosječnu prodaju kruha u tom tjednu i disperziju od prosjeka
b) medijalnu prodaju kruha
c) raspon varijacije, interkvartil i koeficijent kvartilne devijacije.

Rješenja:

a) S obzirom da se radi o negrupiranom nizu računat će se jednostavna aritmetička sredina:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{323}{7} = 46,1$$

Prosječna dnevna prodaja kruha u analiziranom tjednu iznosila je 46,1 kg.

Mjere disperzije od aritmetičke sredine su varijanca, standardna devijacija i koeficijent varijacije.

Varijanca za negrupirani niz računa se na sljedeći način:

$$\sigma^2 = \frac{\sum_{i=1}^{N} \left(x_i - \overline{X}\right)^2}{N} = \frac{946,87}{7} = 135,27$$

Standardna devijacija pokazuje da prosječna odstupanja prodanih količina kruha od aritmetičke sredine iznose 11,63 kg:

$$\sigma = +\sqrt{\sigma^2} = \sqrt{135,27} = 11,63$$

Koeficijent varijacije pokazuje da relativna odstupanja prodanih količina kruha od aritmetičke sredine u prosjeku iznose 25,23%, a to znači da je disperzija umjerena:

$$V = \frac{\sigma}{\overline{X}} \cdot 100 = \frac{11,63}{46,1} \cdot 100 = 25,23\%$$

b) Da bi se mogla odrediti vrijednost medijana, niz treba prvo urediti po veličini vrijednosti obilježja (od x_{min} do x_{max}):

Prodano kruha u kg (X_i)
23
39
44
48
52
56
61

Kod negrupiranog, a uređenog niza, **medijan** je vrijednost obilježja koja pripada jedinici koja se nalazi u sredini niza. Stoga medijalna prodaja kruha iznosi:

$Me = 48\ kg.$

c) **Raspon varijacije** je:

$R = x_{max} - x_{min} = 61 - 23 = 38\ kg.$

Interkvartil predstavlja razliku gornjeg i donjeg kvartila:

$I_q = Q_3 - Q_1 = 56 - 39 = 17\ kg.$

Vrijednosti kvartila:

S obzirom da N nije djeljiv s 4 ($N/4=7/4=1,75$) donji kvartil iznosi:

$Q_1 = x_r = x_2 = 39$

$r=INT(N/4)+1=1+1=2$

S obzirom da N nije djeljiv s 4 ($3N/4=3\cdot 7/4=5,25$) gornji kvartil iznosi:

$Q_3 = x_r = x_6 = 56$

$r=INT(N/4)+1=5+1=6$

Koeficijent kvartilne devijacije relativna je mjera disperzije srednjih 50% elemenata u nizu:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{56-39}{56+39} = 0{,}18$$

stoga je disperzija umjerena.

Primjer 1.8.5.

Prodaja vozila u autosalonu «Z» tijekom ranih dana u rujnu 2008. god. bila je sljedeća:

Tablica 1.14.

Prodaja vozila u autosalonu «Z»

Broj prodanih vozila (x_i)	**Broj radnih dana (f_i)**	$f_i \cdot x_i$	$f_i \cdot x_i^2$
0	4	0	0
1	6	6	6
2	9	18	36
3	4	12	36
4	2	8	32
5	1	5	25
Ukupno	26	49	135

Izvor: Evidencija autosalona «Z», studeni 2008.god.

Zadatak je izračunati:
a) prosječnu prodaju vozila u tom mjesecu i disperziju od prosjeka
b) medijalnu prodaju vozila
c) raspon varijacije, interkvartil i koeficijent kvartilne devijacije
d) mod
e) Pearsonov koeficijent asimetrije
f) Pearsonovu mjeru asimetrije
g) Bowleyevu mjeru asimetrije
h) mjeru zaobljenosti.

Rješenja:

a) S obzirom da se radi o grupiranom nizu računat će se vagana aritmetička sredina:

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{49}{26} = 1,88 \approx 2$$

Prosječna dnevna prodaja vozila u analiziranom mjesecu iznosila je oko 2 automobila.

Mjere disperzije od aritmetičke sredine su varijanca, standardna devijacija i koeficijent varijacije.

Varijanca za grupirani niz računa se na sljedeći način:

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} - \overline{X}^2 = \frac{135}{26} - 1,88^2 = 1,66$$

Standardna devijacija pokazuje da prosječna odstupanja prodanih vozila od aritmetičke sredine iznose 1,29.

$$\sigma = +\sqrt{\sigma^2} = \sqrt{1,77} = 1,29$$

Koeficijent varijacije pokazuje da relativna odstupanja statističkih obilježja od aritmetičke sredine u prosjeku iznose 68,62%, znači da je stupanj disperzije prodaje vozila velik:

$$V = \frac{\sigma}{\overline{X}} \cdot 100 = \frac{1,29}{1,88} \cdot 100 = 68,62\,\%$$

b) Da bi se mogla odrediti vrijednost medijana potrebno je prvo izračunati frekvencije kumultivnog niza:

Broj prodanih vozila (x_i)	Broj radnih dana (f_i)	Kumulativni niz «manje od»
0	4	4
1	6	10
2	9	19
3	4	23
4	2	25
5	1	26
Ukupno	26	-

Medijan je vrijednost obilježja koja pripada jedinici koja se nalazi u sredini niza, stoga se traži središnji član:

$N/2 = 26/2 = 13$

Prva frekvencija kumulativnog niza «manje od» koja sadrži $N/2$ određuje položaj medijana, u ovom primjeru to je 19, pa je medijalna vrijednost numeričkog obilježja uz tu frekvenciju:

$Me = 2$ *vozila*

c) **Raspon varijacije** je:

$R = x_{max} – x_{min} = 5 – 0 = 5$ *vozila.*

Interkvartil predstavlja razliku gornjeg i donjeg kvartila:

$I_q = Q_3 – Q_1 = 56 – 39 = 17$

Vrijednosti kvartila:

Prva frekvencija kumulativnog niza «manje od» koja sadrži $N/4$ određuje položaj donjeg kvartila ($N/4=26/4=6,5$), stoga je donji kvartil vrijednost numeričkog obilježja uz tu frekvenciju:

$Q_1 = 1$

Prva frekvencija kumulativnog niza «manje od» koja sadrži $3N/4$ određuje položaj gornjeg kvartila ($3N/4=3·26/4=19,5$), stoga je gornji kvartil vrijednost numeričkog obilježja uz tu frekvenciju:

$Q_3 =3$

Koeficijent kvartilne devijacije relativna je mjera disperzije srednjih 50% elemenata u nizu i iznosi:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{3-1}{3+1} = 0{,}5$$

d) **Mod** je vrijednost statističkog obilježja kojem pripada najveća frekvencija, stoga mod iznosi 2 prodana vozila dnevno:

$M_o = 2$ *vozila*

e) **Pearsonov koeficijent asimetrije** je:

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{0{,}898}{1{,}29^3} = 0{,}42$$

Centralni moment trećeg reda izračunan preko pomoćnih momenata iznosi:

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 = 16{,}88 - 3 \cdot 1{,}88 \cdot 5{,}19 + 2 \cdot 1{,}88^3 = 0{,}898$$

Za grupirani niz izraz za pomoćne momente oko nule je:

$$m_r = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i}$$

Za pomoćne momente potrebni su sljedeći izračuni:

Broj prodanih vozila (x_i)	**Broj radnih dana (f_i)**	$f_i \cdot x_i$	$f_i x_i^2$	$f_i x_i^3$	$f_i x_i^4$
0	4	0	0	0	0
1	6	6	6	6	6
2	9	18	36	72	144
3	4	12	36	108	324
4	2	8	32	128	512
5	1	5	25	125	625
Ukupno	26	49	135	439	1611

$$m_1 = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{49}{26} = 1{,}88$$

$$m_2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} = \frac{135}{26} = 5{,}19$$

$$m_3 = \frac{\sum_{i=1}^{k} f_i x_i^3}{\sum_{i=1}^{k} f_i} = \frac{439}{26} = 16{,}88$$

$$m_4 = \frac{\sum_{i=1}^{k} f_i x_i^4}{\sum_{i=1}^{k} f_i} = \frac{1611}{26} = 61{,}96$$

e) **Pearsonova mjera asimetrije** pomoću moda je:

$$S_k = \frac{\left(\overline{X} - M_o\right)}{\sigma} = \frac{(2-2)}{1{,}29} = 0$$

Pearsonova mjera asimetrije pomoću medijana iznosi:

$$S_k = \frac{3\left(\overline{X} - M_e\right)}{\sigma} = \frac{3(2-2)}{1{,}29} = 0$$

f) **Bowleyeva mjera asimetrije** iznosi:

$$S_{kq} = \frac{Q_3 - 2M_e + Q_1}{Q_3 - Q_1} = \frac{3 - 2 \cdot 2 + 1}{3 - 1} = 0$$

g) **Mjera zaobljenosti** je:

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{7{,}6}{1{,}29^4} = 2{,}75$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4 = 61{,}96 - 4 \cdot 1{,}88 \cdot 16{,}88 + 6 \cdot 5{,}19 \cdot 1{,}88^2 - 3 \cdot 1{,}88^4 = 7{,}6$$

Distribucija je zaobljenija od normalne.

Primjer 1.8.6.

Broj zaposlenih prema godinama starosti u trgovačkom društvu «Z» prikazan je u tablici 1.15.:

Tablica 1.15.

Zaposleni prema godinama starosti u trgovačkom društvu «Z», stanje 31.03.2009. god.

Godine starosti (x_i)	Broj zaposlenih (f_i)	Raz. sred. x_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
18-25	8	21,5	172,0	3698,0
25-30	12	27,5	330,0	9075,0
30-35	19	32,5	617,5	20068,7
35-40	30	37,5	1125,0	42187,5
40-50	18	45,0	810,0	36450,0
50-60	10	55,0	550,0	30250,0
60-65	3	62,5	187,5	11718,8
Ukupno	100	-	3792,0	153448,0

Izvor: Evidencija trgovačkog društva «Z», travanj 2009.god.

Zadatak je izračunati:

a) prosječnu starost zaposlenih i disperziju od prosjeka
b) medijalnu starost zaposlenih
c) raspon varijacije, interkvartil i koeficijent kvartilne devijacije
d) mod
e) Pearsonov koeficijent asimetrije
f) Pearsonovu mjeru asimetrije
g) Bowleyevu mjeru asimetrije
h) mjeru zaobljenosti.

Rješenja:

a) Veličine razreda grupiranog numeričkog niza različite su od 1 pa je za izračunavanje prosjeka potrebno izvršiti aproksimaciju pomoću sredina razreda. Razredna sredina jednostavan je prosjek donje i gornje granice razreda (vidi treći stupac gornje tablice). S obzirom da se radi o grupiranom statističkom nizu računat će se vagana aritmetička sredina:

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{3792}{100} = 37{,}92$$

Prosječna starost zaposlenih u trgovačkom društvu «Z» bila je oko 38 godina.

Mjere disperzije od aritmetičke sredine su varijanca, standardna devijacija i koeficijent varijacije.

Varijanca za grupirani niz računa se na sljedeći način:

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} - \overline{X}^2 = \frac{153448}{100} - 37{,}92^2 = 96{,}55$$

Standardna devijacija pokazuje da prosječna odstupanja starosti zaposlenih od aritmetičke sredine iznose 9,83 godina:

$$\sigma = +\sqrt{\sigma^2} = \sqrt{96{,}55} = 9{,}83$$

Koeficijent varijacije pokazuje da relativna odstupanja statističkih obilježja od aritmetičke sredine u prosjeku iznose 25,9%, a to nisu značajna odstupanja:

$$V = \frac{\sigma}{\overline{X}} \cdot 100 = \frac{9{,}83}{37{,}92} \cdot 100 = 25{,}9\%$$

b) Da bi se mogla odrediti vrijednost medijana potrebno je prvo izračunati frekvencije kumulativnog niza:

Godine starosti (x_i)	Broj zaposlenih (f_i)	Kumulativni niz «manje od»	Veličine razreda (i)
18-25	8	8	7
25-30	12	20	5
30-35	19	39	5
35-40	30	69	5
40-50	18	87	10
50-60	10	97	10
60-65	3	100	5
Ukupno	100	-	-

Medijan je vrijednost obilježja koja pripada jedinici koja se nalazi u sredini niza, stoga se prvo traži središnji član:

N/2 = 100/2 = 50

Prva frekvencija kumulativnog niza «manje od» koja sadrži *N/2* određuje položaj medijan, u ovom primjeru to je 69 i ona određuje medijalni razred. Vrijednost medijana određuje se pomoću sljedeće formule:

$$M_e = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^{m} f_i}{f_{med}} \cdot i = 35 + \frac{50 - 39}{30} \cdot 5 = 36{,}83$$

Me =36,83 godina.

c) **Raspon varijacije** je:

$R = x_{max} - x_{min} = 65 - 18 = 47$ *godina.*

Interkvartil predstavlja razliku gornjeg i donjeg kvartila:

$I_q = Q_3 - Q_1 = 43{,}33 - 31{,}32 = 12{,}01$

Vrijednosti kvartila:

Prva frekvencija kumulativnog niza «manje od» koja sadrži *N/4* određuje položaj donjeg kvartila (*N/4=100/4=25*), u ovom primjeru to je 39 i ona određuje razred donjeg kvartila. Vrijednost donjeg kvartila je sljedeća:

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum_{i=1}^{q} f_i}{f_{k\,\mathrm{var}\,t}} \cdot i = 30 + \frac{25 - 20}{19} \cdot 5 = 31{,}32$$

Prva frekvencija kumulativnog niza «manje od» koja sadrži *N/4* određuje položaj gornjeg kvartila (3*N/4=3·100/4=75*), u ovom primjeru to je 87 i ona određuje razred donjeg kvartila. Vrijednost gornjeg kvartila je sljedeća:

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum_{i=1}^{q} f_i}{f_{k\,\mathrm{var}\,t}} \cdot i = 40 + \frac{75 - 69}{18} \cdot 10 = 43{,}33$$

Koeficijent kvartilne devijacije relativna je mjera disperzije srednjih 50% elemenata u nizu i pokazuje da disperzija nije velika:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{43{,}33 - 31{,}32}{43{,}33 + 31{,}32} = 0{,}16$$

d) **Mod** je vrijednost statističkog obilježja kojem pripada najveća frekvencija. Ako veličine razreda nisu jednake uzima se najveća korigirana frekvencija (fc_i), stoga je potrebno frekvencije korigirati:

Godine starosti (x_i)	**Broj zaposlenih (f_i)**	**Veličine razreda (i)**	**Korigirane frekvencije (fc_i)**
18-25	8	7	1,1
25-30	12	5	2,4
30-35	19	5	3,8
35-40	30	5	6,0
40-50	18	10	1,8
50-60	10	10	1,0
60-65	3	5	0,6
Ukupno	100	-	-

Najviša korigirana frekvencija iznosi 6 i određuje modalni razred, a vrijednost moda iznosi:

$$M_o = L_1 + \frac{(b-a)}{(b-a)+(b-c)} \cdot i = 35 + \frac{(6-3{,}8)}{(6-3{,}8)+(6-1{,}8)} \cdot 5 = 36{,}72$$

M_o = 36,72 godine.

Modalna, najčešća starost iznosila je oko 37 godina.

e) **Pearsonov koeficijent asimetrije** je:

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{478{,}46}{9{,}83^3} = 0{,}51$$

i pokazuje da je distribucija pozitivno asimetrična.

Centralni moment trećeg reda izračunan preko pomoćnih momenata iznosi:

$$\mu_3 = m_3 - 3m_1 m_2 + 2m_1^3 = 65997{,}6 - 3 \cdot 37{,}92 \cdot 1534{,}48 + 2 \cdot 37{,}92^3 = 487{,}46$$

Za grupirani niz izraz za pomoćne momente oko nule je:

$$m_r = \frac{\sum_{i=1}^{k} f_i x_i^r}{\sum_{i=1}^{k} f_i}$$

Za pomoćne momente potrebno je numerička obilježja potencirati i pomnožiti s frekvencijama:

Raz. sred. x_i	**Broj zaposlenih (f_i)**	$f_i \cdot x_i$	$f_i x_i^2$	$f_i x_i^3$	$f_i x_i^4$
21,5	8	172,0	3698,0	79507,0	1709400,5
27,5	12	330,0	9075,0	249562,5	6862968,8
32,5	19	617,5	20068,7	652234,4	21197617,2
37,5	30	1125,0	42187,5	1582031,3	59326171,9
45,0	18	810,0	36450,0	1640250,0	73811250,0
55,0	10	550,0	30250,0	1663750,0	91506250,0
62,5	3	187,5	11718,8	732421,8	45776367,2
Ukupno	100	3792,0	153448,0	6599757,0	300190026,6

$$m_1 = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{3792}{100} = 37,92$$

$$m_2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} = \frac{153448}{100} = 1534,48$$

$$m_3 = \frac{\sum_{i=1}^{k} f_i x_i^3}{\sum_{i=1}^{k} f_i} = \frac{6599757}{100} = 65997,6$$

$$m_4 = \frac{\sum_{i=1}^{k} f_i x_i^4}{\sum_{i=1}^{k} f_i} = \frac{300190026}{100} = 3001900$$

e) **Pearsonova mjera asimetrije** pomoću moda je:

$$S_k = \frac{\left(\overline{X} - M_o\right)}{\sigma} = \frac{\left(37{,}92 - 36{,}72\right)}{9{,}83} = 0{,}12$$

Pearsonova mjera asimetrije pomoću medijana iznosi:

$$S_k = \frac{3\left(\overline{X} - M_e\right)}{\sigma} = \frac{3\left(37{,}92 - 36{,}83\right)}{9{,}83} = 0{,}33$$

f) **Bowleyeva mjera asimetrije** iznosi:

$$S_{kq} = \frac{Q_3 - 2M_e + Q_1}{Q_3 - Q_1} = \frac{43{,}33 - 2 \cdot 36{,}83 + 31{,}32}{43{,}33 - 31{,}32} = 0{,}08$$

Sve mjere asimetrija pokazuju isto: distribucija je pozitivno asimetrična.

g) **Mjera zaobljenosti** je:

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{27308{,}7}{9{,}83^4} = 2{,}93$$

$$\mu_4 = m_4 - 4m_3 m_1 + 6m_2 m_1^2 - 3m_1^4$$

$$\mu_4 = 3001900 - 4 \cdot 65997{,}6 \cdot 37{,}92 + 6 \cdot 1534{,}48 \cdot 37{,}92^2 - 3 \cdot 37{,}92^4 = 27308{,}7$$

Distribucija je neznatno zaobljenija od normalne.

2. REGRESIJSKA I KORELACIJSKA ANALIZA

2.1. Pojam regresijske i korelacijske analize

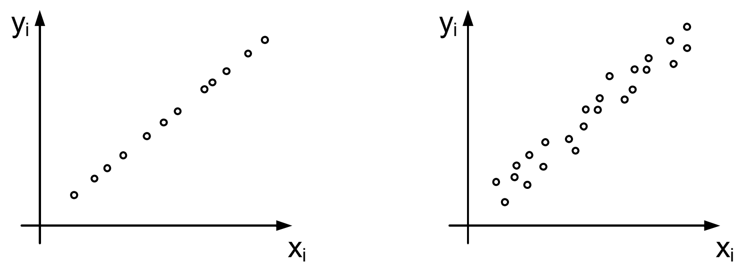
Poslovna i makroekonomska statistika često, uz analizu kretanja jedne ekonomske pojave, imaju potrebu istražiti ovisnosti dviju ili više pojava, odnosno numeričkih nizova, zajedno.

Prvi korak u istraživanju ovisnosti varijabli jeste crtanje grafičkog prikaza koji se naziva dijagram rasipanja.

Dijagram rasipanja u pravokutnom koordinatnom sustavu točkama (x_i, y_i) **prikazuje parove vrijednosti dviju promatranih numeričkih varijabli.**

Na osnovi takve slike mogu se odmah uočiti osnovne veze među promatranim varijablama.

Slika 2.1.



(a) pozitivna funkcionalna veza

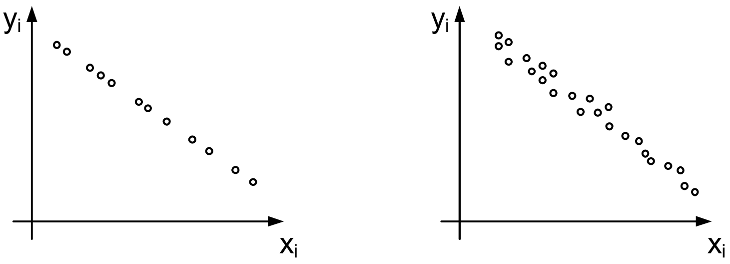
(b) pozitivna statistička veza

Na slici 2.1. prikazana su 2 dijagrama rasipanja. Slika (a) prikazuje funkcionalnu vezu između 2 varijable X i Y. Zamišljena linija koja povezuje sve točke na slici je pravac. Matematički oblik veze, ovih dviju promatranih varijabli, je jednadžba pravca. Od te linije nema nikakvog odstupanja, stoga se kaže da je ova veza strogo funkcionalna. Zamišljeni pravac je rastući, odnosno porast vrijednosti jedne varijable prati porast vrijednosti druge promatrane varijable zato je ova veza pozitivna.

Čest slučaj u praksi prikazan je na slici (b). Ako se između točaka ovog dijagrama zamisli krivulja, to bi opet bio pravac. Međutim ovdje su prisutna pozitivna i negativna odstupanja od linije pravca, a to se tumači raznim utjecajima drugih varijabli iz prakse. Stoga ova veza više nije strogo funkcionalna, već se za nju kaže da je statistička (stohastička ili slučajna) veza.

Porast vrijednosti jedne varijable u prosjeku prati porast druge varijable, pa je i ova veza pozitivna.

Slika 2.2.



(a) negativna funkcionalna veza

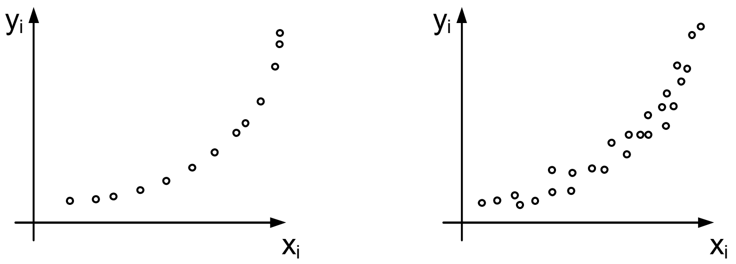
(b) negativna statistička veza

Na slici 2.2. prikazana su opet 2 dijagrama rasipanja. Slika (a) prikazuje funkcionalnu vezu između 2 varijable X i Y, a zamišljena linija koja povezuje sve točke na slici je opet pravac. Matematički oblik veze ovih dviju promatranih varijabli je jednadžba pravca od čije linije nema nikakvog odstupanja, pa je i ova veza strogo funkcionalna. Zamišljeni pravac je padajući, odnosno porast vrijednosti jedne varijable prati pad vrijednosti druge promatrane varijable, pa je ova veza negativna.

Čest slučaj u praksi prikazan je na slici (b). Ako se između točaka ovog dijagrama zamisli krivulja, to bi opet bio pravac. Međutim ovdje su prisutna pozitivna i negativna odstupanja od linije pravca, a to se, kako je već rečeno, tumači raznim utjecajima drugih varijabli iz prakse. Stoga ova veza više nije strogo funkcionalna, već se za nju kaže da je statistička (stohastička ili slučajna) veza. Porast vrijednosti jedne varijable u prosjeku prati pad druge varijable, stoga je i ova veza negativna.

Veza između promatranih varijabli ne mora uvijek odgovarati jednadžbi pravca.

Slika 2.3.



(a) pozitivna funkcionalna krivolinijska veza

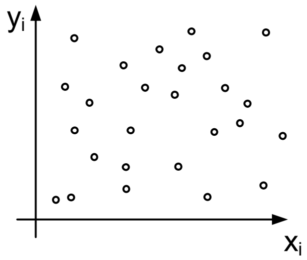
(b) pozitivna statistička krivolinijska veza

Na slici 2.3. su 2 dijagrama rasipanja. Slika (a) prikazuje funkcionalnu krivolinijsku pozitivnu vezu između 2 varijable X i Y. Zamišljena linija koja povezuje sve točke na slici je krivulja. Matematički oblik veze ovih dviju promatranih varijabli je neka eksponencijalna jednadžba od čije linije nema nikakvog odstupanja, pa je ova veza strogo funkcionalna. I ovdje vrijedi da porast vrijednosti jedne varijable prati porast vrijednosti druge promatrane varijable zato je ova veza pozitivna.[5]

U praksi se češće događa slučaj prikazan na slici (b). Ako se između točaka ovog dijagrama zamisli linija to bi opet bila krivulja. Međutim ovdje su prisutna pozitivna i negativna odstupanja zbog utjecaja drugih varijabli iz prakse. Ova veza je statistička (stohastička ili slučajna). I ovdje porast vrijednosti jedne varijable u prosjeku prati porast druge varijable, stoga je i ova veza pozitivna.

[5] U poslovnoj i makroekonomskoj statistici promatra se samo prvi kvadrant koordinatnog sustava jer su u ekonomiji varijable uglavnom pozitivne.

Slika 2.4.



(a) nema veze među pojavama

Na slici 2.4. prikazan je dijagram rasipanja koji upućuje na zaključak da nema povezanosti među promatranim pojavama. Naime zamišljena krivulja koja prolazi između točaka na ovom grafikonu ne postoji i ne može se definirati prati li porast jedne pojave rast ili pad druge promatrane pojave, jer se pri jednoj vrijednosti varijable x_i može dogoditi više različitih vrijednosti druge varijable y_i.

Pod pojmom **korelacija podrazumijeva se međuzavisnost ili povezanost slučajnih varijabli.** Po smjeru korelacija može biti pozitivna i negativna. **Pozitivna korelacija je prisutna kada rast jedne varijable prati rast druge promatrane varijable, odnosno kada pad jedne prati pad druge varijable. Negativna korelacija prisutna je kada rast jedne varijable prati pad druge varijable i obratno.**

Za razliku od korelacijske analize **zadaća regresijske analize je da pronađe analitičko-matematički oblik veze između jedne ovisne ili regresand varijable i jedne ili više neovisnih ili regresorskih varijabli.**

Osim objašnjavanja prirode ovisnosti promatranih pojava na temelju tog analitičkog oblika može se vršiti predviđanje vrijednosti ovisne varijable pri određenim vrijednostima neovisne-ih varijabli.

2.2. Regresijski model

U slučaju postojanja samo **jedne ovisne ili regresand, i samo jedne neovisne ili regresorske varijable** kaže se da je to **jednostavni, jednostruki ili jednodimenzionalni regresijski model.** Ako se uz **jednu ovisnu ili regresand**

varijablu u analizi jave **dvije ili više neovisnih ili regresorskih varijabli kaže se da je to složeni, višestruki ili višedimenzionalni model.**

Regresijska analiza se može postaviti na sljedeći način:

1. **Potpuno, precizno i koncizno definiranje predmeta i ciljeva istraživanja**, te postavljanje osnovnih pretpostavki
2. **Crtanje dijagrama rasipanja, izbor modela i definiranje varijabli** (na primjer **aditivni model**[6])

$$Y = f(X) + e, \tag{2.2.1}$$

gdje je: Y- ovisna ili regresand varijabla

X - neovisna ili regresorska varijabla

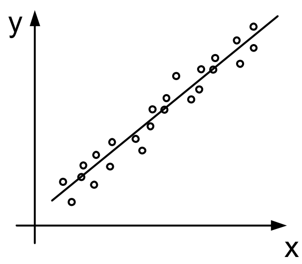
e - slučajna komponenta.

Svaki model ima slučajnu komponentu e, koja upućuje da veze između pojava u praksi nisu funkcionalne nego su statističke ili stohastičke, odnosno oko linije konkretnog aditivnog modela postoje pozitivna i/ili negativna odstupanja originalnih vrijednosti.

U ovom koraku bitno je formirati statističko-dokumentacijsku osnovu iz primarnog (direktno) i/ili sekundarnog (literatura) izvora vodeći računa da promatrani podaci budu usporedivi i da njihova usporedba zadovoljava ekonomske kriterije.

3. **Odabir konkretnog regresijskog modela, njegova specifikacija i pretpostavke** (na primjer **linearni model**: $Y = \beta_0 + \beta_1 X + e$.)

Slika 2.5.



[6] Modeli u praksi ne moraju biti aditivni: na primjer, multiplikativni model

$Y = f(X) \cdot e$,

gdje je: Y- ovisna ili regresand varijabla, X - neovisna ili regresorska varijabla , e - slučajna komponenta.

Na slici 2.5. prikazan je dijagram rasipanja koji upućuje na postojanje pozitivne statističke veze između dviju varijabli X i Y. Povlačenjem linije pravca između točaka dijagrama rasipanja pretpostavlja se aditivna linearna veza među varijablama.

4. **Statistička analiza modela: ocjena parametara i pokazatelja reprezentativnosti modela**

U ovoj fazi regresijske analize ocjenjuju se parametri konkretnog izabranog regresijskog modela, te se računaju odgovarajući pokazatelji reprezentativnosti modela, koji ukazuju na to zadovoljava li model statističke kriterije.

5. **Testiranje hipoteza o modelu i statističko teorijskih pretpostavki**

a) DA - ako su ispunjene pretpostavke, vrši se sinteza rezultata i donose se sudovi o predmetu istraživanja

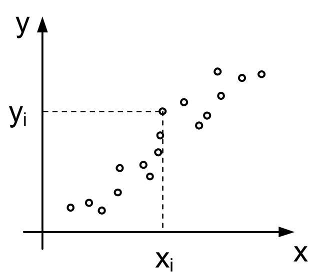
b) NE -ako nisu ispunjene pretpostavke: vrši se modifikacija modela i vraća se na korak 2., tj. na izbor novog modela i definiranje varijabli.

Regresijskom analizom traže se i ocjenjuju parametri funkcije koja na najbolji mogući način opisuje vezu između varijabli X i Y.

2.3. Model jednostavne linearne regresije

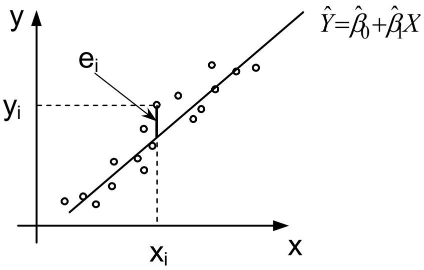
Ako su u analizi prisutne samo dvije varijable tada se radi o jednostavnoj regresiji. Na temelju uzorka parova vrijednosti varijabli X i Y: $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ crta se dijagram rasipanja koji je prikazan na slici 2.6..

Slika 2.6.



Dijagram rasipanja pokazuje pozitivnu statističku vezu između pojava X i Y.

Slika 2.7.



Ako se na dijagramu rasipanja povuče **pravac** on **je općenito oblika:**

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \qquad (2.3.1)$$

Svaka točka dijagrama rasipanja zadovoljava jednadžbu:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i , \qquad (2.3.2)$$

odnosno **svaka točka Y_i odstupa od linije pravca za pozitivnu ili negativnu razliku** e_i **.**

Regresijska analiza traži parametre $\hat{\beta}_0 \ i \ \hat{\beta}_1$, tako da pravac $\hat{Y}$ prolazi između stvarnih točaka promatranih varijabli i da najbolje tumači vezu između njih, odnosno pravac mora biti takav da odstupanja e_i budu najmanja.

Postoji više različitih metoda za ocjenu ovih parametara, a najčešće rabljena metoda je **metoda najmanjih kvadrata** koja upravo procjenjuje parametre $\hat{\beta}_0 \ i \ \hat{\beta}_1$ tako da odstupanja e_i budu najmanja.[7] Ona daje najbolje

---

[7] Pri formiranju modela postavljaju se i pretpostavke slučajne greške e_i (tzv. Gauss-Markovljevi uvjeti):

I. $E(e_i) = 0, \ \forall i$ (očekivanje slučajne pogrješke je nula za svaku opservaciju)

II. $E(e_i, e_j) = \sigma_e^2 < +\infty \ \ za \ \ i = j$ $= const.$ (homoskedastičnost varijance reziduala, tj. pretpostavlja se da je varijanca reziduala konačna i čvrsta)

III. $E(e_i, e_j) = 0, \ \forall i \neq j, \ \ tj.$ $Cov(e_i, e_j) = 0, \ \forall i \neq j$ (pogrješka je slučajna i nema korelacije između varijabli s pomakom od e_i)

linearne nepristrane ocjene i vrlo je često primjenjivana metoda za ocjenu parametara.

$$Y_i = \hat{Y}_i + e_i \tag{2.3.3}$$

Odstupanja originalnih vrijednosti od ocijenjenih vrijednosti e_i mogu biti pozitivna i negativna, stoga da se ne bi međusobno poništile te pozitivne i negativne vrijednosti, ova **metoda minimizira sumu kvadrata od e_i.**

$$\min \sum_{i=1}^{n} e_i^2 = \min \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \min \sum_{i=1}^{n} \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \right]^2 \tag{2.3.4}$$

Dakle, traži se minimum kvadrata odstupanja empirijskih (stvarnih) u odnosu na regresijske vrijednosti:

$$f(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \right]^2 \tag{2.3.5}$$

Nakon primjene matematičkog postupka traženja minimuma dobiju se dvije jednadžbe s dvije nepoznanice tj. parametri regresijskog modela $\hat{\beta}_0$ i $\hat{\beta}_1$.

$$n \cdot \hat{\beta}_0 + \hat{\beta}_1 \cdot \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \tag{2.3.6}$$

$$\hat{\beta}_0 \cdot \sum_{i=1}^{n} X_i + \hat{\beta}_1 \cdot \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i \tag{2.3.7}$$

Sustav (2.3.6), (2.3.7) uvijek ima rješenja i **vrijedi da je:**

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \quad i \quad \hat{\beta}_0 = \overline{Y} - \hat{b}\overline{X} \tag{2.3.8}$$

gdje su:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad i \quad \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \tag{2.3.9}$$

jednostavne aritmetičke sredine varijabli X i Y.

Konačno je ocijenjeni model:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \tag{2.3.10}$$

IV. $E(e_i, X_i) = 0$ (slučajna pogrješka je distribuirana nezavisno od regresorske varijable X)

Vrijedi da je slučajna pogrješka e distribuirana normalnom distribucijom: $N(0, \sigma^2 < \infty)$.

gdje je $\hat{\beta}_0$ **konstantni član, tj. očekivana vrijednost zavisne varijable kada je nezavisna varijabla jednaka nuli**: ($\hat{\beta}_0 = \overline{Y}$ kada je X=0). Ovaj parametar interpretira se i kao odsječak na osi koordinata u kojoj regresijski pravac siječe os, uz pretpostavku da je apscisa te točke X=0.

Regresijski koeficijent $\hat{\beta}_1$ pokazuje prosječnu promjenu zavisne varijable kada nezavisna varijabla poraste za jedinicu. Ovaj parametar interpretira se i kao koeficijent smjera, odnosno nagiba regresijskog pravca koji može imati pozitivni i negativni predznak, ovisno o smjeru veze između promatranih varijabli.

Može se postaviti i **suprotna ovisnost u modelu**, na način da je **varijabla X sada ovisna ili regresorska varijabla**:

$$X_i = \hat{\alpha}_0 + \hat{\alpha}_1 Y_i + e_i \qquad (2.3.11)$$

Ocjena parametara u ovom slučaju vrši se na jednak način kao kod početnog modela $\hat{Y}$, samo što je sada X ovisna varijabla, pa **u izrazima za izračunavanje parametara, X i Y mijenjaju mjesta**.

$$\hat{\alpha}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2} \quad i \quad \hat{\alpha}_0 = \overline{X} - \hat{a}_1 \overline{Y}$$

(2.3.12)

Matričnim putem regresijska jednadžba može se napisati:

$$Y = X\hat{\beta} \ ; \qquad (2.3.13)$$

gdje su matrice:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ .. \\ Y_n \end{bmatrix}; \ X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ .. & .. \\ 1 & X_n \end{bmatrix}; \ \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \qquad (2.3.14)$$

$$\hat{\beta} = (X^T X)^{-1} \cdot (X^T Y) \qquad (2.3.15)$$

gdje su:

$$(X^T X) = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} \ i \ (X^T Y) = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}. \qquad (2.3.16)$$

2.4. Linearna korelacija i procjena koeficijenta korelacije

2.4.1 Linearna korelacija

Najpoznatija mjera linearne korelacije između slučajnih varijabli je **Pearsonov koeficijent linearne korelacije (r):**

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})\cdot(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}, \text{ ili } r = \frac{\sum_{i=1}^{n} X_i Y_i - n \cdot \overline{X} \cdot \overline{Y}}{n \cdot \sigma_x \cdot \sigma_y} \qquad (2.4.1)$$

gdje su σ_x i σ_y jednostavne standardne devijacije promatranih varijabli:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{X}^2} \qquad \text{i } \sigma_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - \overline{Y}^2} \qquad (2.4.2)$$

Vrijednost koeficijenta linearne korelacije kreće se u intervalu:

$$-1 \leq r \leq 1 \qquad (2.4.3)$$

U skladu s veličinom ovog koeficijenta može se zaključiti smjer i intenzitet linearne korelacije među promatranim varijablama:

$r = -1$; $r = 1$ $\Rightarrow$ funkcionalna negativna/pozitivna korelacija

$-1 < r \leq -0{,}8$; $0{,}8 \leq r < 1$ $\Rightarrow$ jaka negativna/pozitivna korelacija

$-0{,}8 < r \leq -0{,}5$; $0{,}5 \leq r < 0{,}8$ $\Rightarrow$ srednje jaka negativna/pozitivna korelacija

$-0{,}5 < r < 0$; $0 < r < 0{,}5$ $\Rightarrow$ slaba negativna/pozitivna korelacija

$r = 0$ $\Rightarrow$ nema korelacije.

Koeficijent parcijalne korelacije je pokazatelj korelacije između dvije varijable uz istodobno isključenje utjecaja drugih varijabli.

Ako se računa **parcijalna korelacija između triju varijabli** u kombinaciji vrijedi da je:

- korelacija između 1. i 2. varijable uz isključenje utjecaja 3. varijable:

$$\rho_{12.3} = \frac{r_{12} - (r_{13} \cdot r_{23})}{\sqrt{(1 - r_{13}^2)\cdot(1 - r_{23}^2)}} \qquad (2.4.4)$$

- korelacija između 1. i 3. varijable uz isključenje utjecaja 2. varijable:

$$\rho_{13.2} = \frac{r_{13} - (r_{12} \cdot r_{32})}{\sqrt{(1 - r_{12}^2) \cdot (1 - r_{32}^2)}} \tag{2.4.5}$$

- korelacija između 2. i 3. varijable uz isključenje utjecaja 1. varijable:

$$\rho_{23.1} = \frac{r_{23} - (r_{21} \cdot r_{31})}{\sqrt{(1 - r_{21}^2) \cdot (1 - r_{31}^2)}}, \tag{2.4.6}$$

gdje su r_{ij} odgovarajući koeficijenti korelacije promatranih varijabli.

Matrica koeficijenata korelacije općenito je:

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & .. & r_{1k} \\ r_{21} & 1 & r_{23} & .. & r_{2k} \\ r_{31} & r_{32} & 1 & .. & r_{3k} \\ .. & .. & .. & .. & .. \\ r_{k1} & r_{k2} & r_{k3} & .. & 1 \end{bmatrix},$$

gdje je k broj promatranih varijabli. S obzirom da za korelaciju vrijedi da je $r_{ij} = r_{jk}$, matrica R je simetrična matrica.

Koeficijent parcijalne korelacije između 2. varijable, uz isključenje utjecaja ostalih promatranih varijabli je općenito:

$$\rho_{ij.klm....} = -\frac{R_{ij}}{\sqrt{(R_{ii} \cdot R_{jj})}}, \tag{2.4.7}$$

gdje je R_{ij} kofaktor, tj. algebarski komplement matrice koeficijenata korelacije R:

$$R_{ij} = (-1)^{i+j} \cdot M_{ij}, \tag{2.4.8}$$

a M_{ij} je odgovarajući minor (subdeterminanta) matrice koeficijenata korelacije R.

2.4.2. Procjena koeficijenta korelacije

Ako je vrijednost **koeficijenta korelacije bliža 1**, njegova sampling distribucija (distribucija iz uzoraka) je asimetrična i nema normalni oblik. Stoga se procjena vrši **Fisherovom transformacijom** (r u Z) - pomoću odgovarajućih tablica ili računom:

$$\hat{Z} = ar\,tgh\,\hat{r} \tag{2.4.9}$$

Interval povjerenja procjene za Z je:

$$\Pr\left\{\hat{Z} - Z \cdot Se(Z) < Z < \hat{Z} + Z \cdot Se(Z)\right\} = 1 - \alpha \tag{2.4.10}$$

gdje je:

Z - odgovarajuća vrijednost iz tablica normalne distribucije

$1-\alpha$ - odgovarajući nivo pouzdanosti procjene (najčešće 95%)

$$Se(Z) = \sqrt{\frac{1}{n-3}}\,. \tag{2.4.11}$$

Nakon izračunavanja intervala pouzdanosti za Z potrebno je donju i gornju granicu za Z transformirati natrag u r (Z u r) - pomoću odgovarajućih tablica ili računom:

$$r = tgh\,Z \tag{2.4.12}$$

Kod negativne korelacije prilikom transformiranja treba voditi računa o negativnom predznaku koeficijenta korelacije r.

Napomena: Fisherova transformacija se ne koristi za male uzorke.

2.5. Spearmanov koeficijent korelacije

Ako se želi istražiti međuovisnost pojava koje su izražene modalitetima redoslijednog obilježja, odnosno ako su im modaliteti pridruženi na temelju ordinalne skale računa se korelacija ranga.

Najpoznatija **mjera korelacije ranga između dviju varijabli je Spearmanov koeficijent korelacije ranga (r_S):**

$$r_S = 1 - \frac{6 \cdot \sum_{i=1}^{N} d_i^{\,2}}{N^3 - N}\,, \tag{2.5.1}$$

gdje je:

N - broj parova vrijednosti varijabli X i Y,

$d_i = r(x_i) - r(y_i)$ - razlika rangova vrijednosti varijabli X i Y.

Svakoj vrijednosti varijabli X i Y dodjeljuje se rang iskazan prvim N prirodnim brojevima. Pri tome se rangiranje može započeti rangom 1, počevši od najmanje vrijednosti obilježja ili počevši od najveće vrijednosti obilježja. Pri tom se rangiranje mora provesti na jednak način za obje promatrane varijable. Ako se javi **više jednakih vrijednosti jedne varijable** mora im se dodijeliti jednak rang na način da **se izračuna aritmetička sredina njihovih rangova**.

Spearmanov koeficijent korelacije ranga može poprimiti vrijednosti u intervalu:

$$-1 \leq r_S \leq 1 \tag{2.5.2}$$

Kada ovaj koeficijent poprimi vrijednosti -1 i 1, riječ je o potpunoj korelaciji ranga među varijablama. Vrijednost ovog koeficijenta 0 znači da nema nikakve korelacije ranga među pojavama. Najčešće se vrijednost Spearmanovog koeficijenta kreće u rasponu $-1 < r_s < 1$. Koeficijent bliži rubovima ovog intervala, tj. -1 i 1 upućuje na veću korelaciju ranga promatranih dviju varijabli.

Primjer 2.5.1.

Vlasnik velikog salona automobila «Z» želi utvrditi odnos između postignutih bodova na testu koji su prodavači ispunjavali prilikom prijema na posao i prodanih automobila, koje su ti prodavači uspjeli prodati tijekom svoje prve godine rada u tom salonu. Slučajni uzorak od 10 prodavača dao je sljedeće rezultate:

Tablica 2.1.

Bodovi postignuti na testu i broj prodanih automobila 10 prodavača autosalona «Z», 2008. god.

Prodavač	Bodovi na testu (x_i)	Broj prodanih automobila (y_i)	Rangirane varijable		$d_i=r(x_i)-r(y_i)$	d_i^2
			$r(x_i)$	$r(y_i)$		
A	51	35	8	8	0	0
B	65	46	6	5	1	1
C	49	33	10	9	1	1
D	66	45	5	6	-1	1
E	50	29	9	10	-1	1
F	64	42	7	7	0	0
G	68	47	4	4	0	0
H	72	50	3	3	0	0
I	77	52	1	2	-1	1
J	75	53	2	1	1	1

Izvor: Podaci autosalona «Z», 2009.god.

Zadatak je izračunati Spearmanov koeficijent korelacije ranga.

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{N} d_i^2}{N^3 - N} = 1 - \frac{6 \cdot 6}{10^3 - 10} = 0{,}96$$

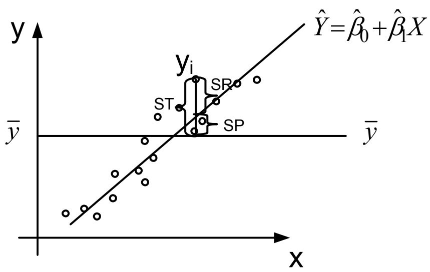
Očita je jaka veza između postignutih bodova na testu i broja prodanih automobila.

2.6. Regresijska dijagnostika

Nakon ocjene parametara regresijskog modela postavlja se pitanje **reprezentativnosti**, odnosno sposobnosti **modela da objasni kretanje ovisne varijable Y uz pomoć odabrane neovisne varijable X.**

U tu svrhu koriste se neki **apsolutni i relativni pokazatelji.** Ovi pokazatelji temelje se na raspodjeli odstupanja vrijednosti ovisne varijable Y_i u regresijskom modelu od njene aritmetičke sredine $\overline{Y}$ i njenih očekivanih vrijednosti $\hat{Y}_i$.

Slika 2.8.



Na slici 2.8. prikazan je dijagram rasipanja varijabli X i Y s ucrtanim ocijenjenim modelom pravca $\hat{Y}$. Na slici je označena i aritmetička sredina varijable $\overline{Y}$. Pri formiranju suma odgovarajućih odstupanja, zbog već ranije navedenog razloga, da se međusobno ne bi poništile pozitivne i negativne razlike računaju se njihovi kvadrati:

$$SP = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2 = \hat{\beta}_0 \cdot \sum_{i=1}^{n} Y_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i Y_i - n \cdot \overline{Y}^2 \qquad (2.6.1)$$

Dakle, **SP je suma kvadrata protumačenog dijela** odstupanja vrijednosti varijable Y od aritmetičke sredine, **odnosno suma kvadrata odstupanja ocijenjenih vrijednosti varijable Y od aritmetičke sredine**.

$$SR = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\beta}_0 \cdot \sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i Y_i \qquad (2.6.2)$$

Dakle, **SR je suma kvadrata neprotumačenog dijela** odstupanja vrijednosti varijable Y od aritmetičke sredine, **odnosno suma kvadrata odstupanja originalnih ili empirijskih vrijednosti varijable Y od ocijenjenih vrijednosti**. Ova odstupanja su u stvari slučajne pogrješke e_i.

$$ST = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n} Y_i^2 - n \cdot \overline{Y}^2 \qquad (2.6.3)$$

ST je suma kvadrata ukupnih odstupanja vrijednosti varijable Y od aritmetičke sredine.

Vrijedi da je:

$$SP + SR = ST\text{,} \qquad (2.6.4)$$

što se vidi i na slici 2.8. Ovaj izraz koji je u skraćenom obliku dan pomoću **(2.6.4) zove se jednadžba analize varijance i predstavlja temelj analize reprezentativnosti regresijskog modela.**

Standardna pogrješka regresije je apsolutni pokazatelj reprezentativnosti regresijskog modela, a pokazuje prosječni stupanj varijacije stvarnih vrijednosti ovisne varijable u odnosu na očekivane regresijske vrijednosti.

$$\hat{\sigma}_{\hat{Y}} = \sqrt{\frac{SR}{n-2}} \qquad (2.6.5)$$

Izraz (2.6.5) je standardna pogrješka regresije jednostrukog modela. Ovaj pokazatelj izražen je u originalnim jedinicama mjere ovisne varijable Y. Stoga je na temelju standardne pogrješke regresije teško uspoređivati reprezentativnost modela s različitim mjernim jedinicama.

Taj problem eliminira **relativni pokazatelj koeficijent varijacije regresije, koji predstavlja postotak standardne pogrješke regresije od aritmetičke sredine varijable Y.**

$$\hat{V}_{\hat{Y}} = \frac{\hat{\sigma}_{\hat{Y}}}{\overline{Y}} \cdot 100 \qquad (2.6.6)$$

Najmanja vrijednost koeficijenta varijacije je 0%, a najveća nije definirana. **Što je koeficijent varijacije regresijskog modela bliži nuli, to je model reprezentativniji.** Često se uzima dogovorena granica reprezentativnosti od 10%. Dakle ako je koeficijent varijacije manji od 10% kaže se da je model dobar.

Koeficijent determinacije je pokazatelj reprezentativnosti regresijskog modela koji se temelji na analizi varijance. On se definira kao **omjer sume**

kvadrata odstupanja protumačenih regresijom i sume kvadrata ukupnih odstupanja.

$$r^2 = \frac{SP}{ST} \qquad (2.6.7)$$

Koeficijent determinacije kaže koliko % je sume kvadrata odstupanja vrijednosti varijable Y od aritmetičke sredine protumačeno regresijskim modelom.

Prema (2.6.4), vrijedi da je:

$$r^2 = 1 - \frac{SR}{ST} \qquad (2.6.8)$$

Vrijednost koeficijenta determinacije kreće se u intervalu $0 \le r^2 \le 1$**. Regresijski model je reprezentativniji ako je ovaj pokazatelj bliži 1.** Teorijska granica reprezentativnosti modela je 0,9. U praksi je ponekad vrlo teško pronaći varijablu koja dobro objašnjava ovisnu pojavu, pa se ta granica reprezentativnosti spušta i do 0,6.

Korigirani koeficijent determinacije:

$$\bar{r}^2 = 1 - \frac{n-1}{n-(k+1)} \cdot (1 - r^2) \qquad (2.6.9)$$

je asimptotski nepristrana ocjena koeficijenta determinacije.

Za jednostruku linearnu regresiju vrijedi da je koeficijent linearne korelacije:

$$r = \pm\sqrt{r^2}\,,$$

gdje predznak koeficijenta linearne korelacije odgovara predznaku parametara ocijenjenog jednostrukog linearnog modela: $\hat{\alpha}_1$ i $\hat{\beta}_1$.

Još vrijedi da je:

$$r = \hat{\alpha}_1 \cdot \frac{\sigma_Y}{\sigma_X} \quad i \quad r = \hat{\beta}_1 \cdot \frac{\sigma_X}{\sigma_Y}\,,$$

gdje su $\sigma_X \; i \; \sigma_Y$ standardne devijacije varijabli X i Y.

Primjer 2.6.1.

Ispituje se veza između obrazovanja i visina plaća u trgovini «Z» u kojoj je zaposleno 10 djelatnika:

Tablica 2.2.

Godine obrazovanja i prosječne plaće zaposlenika u trgovini «Z» u 2008. god.

Obrazovanje - godine (x_i)	Prosječna neto mjesečna plaća u kn (y_i)	$x_i \cdot y_i$	x_i^2	y_i^2
6	3.500	21000	36	12250000
8	3.600	28800	64	12960000
10	3.600	36000	100	12960000
12	4.100	49200	144	16810000
12	4.200	50400	144	17640000
14	4.900	68600	196	24010000
14	4.700	65800	196	22090000
15	4.900	73500	225	24010000
16	5.800	92800	256	33640000
18	6.500	117000	324	42250000
125	45.800	603100	1685	218620000

Izvor: Podaci trgovine «Z», 2009.god.

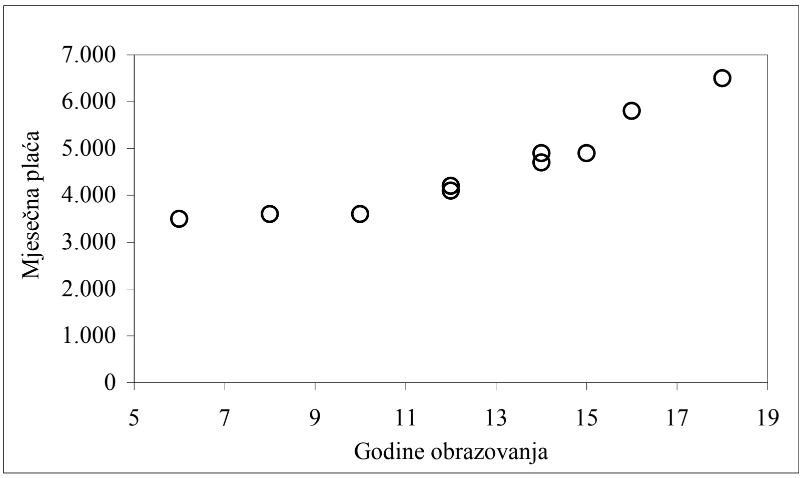
Zadatak je:

a) nacrtati dijagram rasipanja
b) ocijeniti parametre jednadžbi pravaca linearne regresije
c) izračunati Pearsonov koeficijent linearne korelacije i koeficijent determinacije
d) izračunati koeficijent varijacije regresije

Rješenja:

a) **Grafikon 11.**

Dijagram rasipanja



Izvor: Podaci trgovine «Z», 2009.god.

b) Jednadžba prvog pravca regresije glasi:

$$\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X = 1457{,}5 + 249{,}8X$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} = \frac{603100 - 10 \cdot 12{,}5 \cdot 4580}{1685 - 10 \cdot 12{,}5^2} = 249{,}8$$

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X} = 4580 - 249{,}8 \cdot 12{,}5 = 1457{,}6$$

Regresijski koeficijent ($\widehat{\beta}_1$) pokazuje da se mjesečna neto plaća povećava u prosjeku za 249,8 kn kada se dužina obrazovanje produži za 1 godinu.

Jednadžba drugog pravca regresije glasi:

$$\widehat{X} = \widehat{\alpha}_0 + \widehat{\alpha}_1 Y = -3{,}53 + 0{,}0035Y$$

$$\widehat{\alpha}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2} = \frac{603100 - 10 \cdot 12{,}5 \cdot 4580}{218620000 - 10 \cdot 4580^2} = 0{,}0035$$

$$\widehat{\alpha}_0 = \overline{X} - \widehat{\alpha}_1 \overline{Y} = 12{,}5 - 0{,}0035 \cdot 4580 = -3{,}53$$

Regresijski koeficijent ($\widehat{\alpha}_1$) pokazuje da se obrazovanje produžilo u prosjeku za 0,0035 godine ukoliko se mjesečna neto plaća povećala za 1 kn.

c) **Pearsonov koeficijent korelacije** (r) iznosi:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n \cdot \overline{X} \cdot \overline{Y}}{n \cdot \sigma_x \cdot \sigma_y} = \frac{603100 - 10 \cdot 12{,}5 \cdot 4580}{10 \cdot 3{,}5 \cdot 941} = 0{,}93$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{X}^2} = \sqrt{\frac{1685}{10} - 12{,}5^2} = 3{,}5$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - \overline{Y}^2} = \sqrt{\frac{218620000}{10} - 4580^2} = 941$$

Između obrazovanja i plaća postoji jaka i pozitivna korelacija.

Koeficijent determinacije glasi:

$$r^2 = \frac{SP}{ST} = \frac{7648460}{8856000} = 0{,}86$$

$$SP = \widehat{\beta}_0 \cdot \sum_{i=1}^{n} Y_i + \widehat{\beta}_1 \sum_{i=1}^{n} X_i Y_i - n \cdot \overline{Y}^2 = 1457{,}6 \cdot 45800 + 249{,}8 \cdot 603100 - 10 \cdot 4580^2 = 7648460$$

$$ST = \sum_{i=1}^{n} Y_i^2 - n \cdot \overline{Y}^2 = 218620000 - 10 \cdot 4580^2 = 8856000$$

Koeficijent determinacije pokazuje da je 86% sume kvadrata odstupanja vrijednosti varijable Y od aritmetičke sredine protumačeno regresijskim modelom.

Koeficijent linearne korelacije pomoću koeficijenta determinacije iznosi:

$$r = \sqrt{r^2} = \sqrt{0{,}86} = 0{,}93$$

d) Koeficijent varijacije regresije glasi:

$$\widehat{V}_{\widehat{Y}} = \frac{\widehat{\sigma}_{\widehat{Y}}}{\overline{Y}} \cdot 100 = \frac{388{,}5}{4580} \cdot 100 = 8{,}48$$

Koeficijent varijacije regresije manji je od 10% pa je ocijenjeni model regresije reprezentativan.

Standardna pogrješka regresije je:

$$\widehat{\sigma}_{\widehat{Y}} = \sqrt{\frac{SR}{n-2}} = \sqrt{\frac{1207540}{10-2}} = 388{,}5$$

$$SR = \sum_{i=1}^{n} Y_i^2 - \widehat{\beta}_0 \sum_{i=1}^{n} Y_i - \widehat{\beta}_1 \sum_{i=1}^{n} X_i Y_i = 218620000 - 1457{,}6 \cdot 45800 - 249{,}8 \cdot 603100 = 1207540$$

3. ANALIZA VREMENSKIH SERIJA

3.1. Definicija vremenskog niza

Praćenje različitih ekonomskih pojava u vremenu veoma je važno za poslovnu i gospodarsku politiku, stoga u poslovnoj i makroekonomskoj statistici obrada vremenskih nizova zauzima važno mjesto.

Vremenski niz je skup kronološki uređenih vrijednosti pojave. Vrijednosti promatrane varijable vremenskog niza označavaju se:

$$\{Y_t\},\ t = 1,2,...,N\ ,^{8} \qquad (3.1.1)$$

i zovu se frekvencije.

3.2. Vrste nizova

S obzirom na vrijeme opažanja vrijednosti pojave postoje dvije vrste vremenskog niza:

1. **intervalni vremenski niz**
2. **trenutačni vremenski niz.**

Kod **intervalnog vremenskog niza** vrijednost pojave mjeri se u vremenskom intervalu.

Kod **trenutačnog vremenskog niza** vrijednost pojave mjeri se u trenutku vremena.

3.3. Grafičko prikazivanje i uspoređivanje vremenskih nizova

Grafičkim prikazom vremenskih nizova postiže se jasnija i preglednija slika kretanja vrijednosti promatrane pojave kroz vrijeme.

Slika grafikona, kao i svi prikazi u statistici, mora biti jasna i potpuna. Kod vremenskih nizova vrijeme se uvijek nanosi na apscisu dok se na ordinatu nanose vrijednosti pojave Y_t. Grafički prikaz treba imati sve potrebne oznake:

---

[8] Indeks "t" ovdje znači: **time** (eng.) - vrijeme.

naslov grafikona, izvor grafikona, oznake na ordinati, oznake na apscisi i kazalo ako se na istom grafikonu prikazuje više različitih vremenskih nizova.

Na jednom grafičkom prikazu može se uspoređivati više vremenskih nizova s istom jedinicom mjere. Broj vremenskih nizova na jednom grafikonu ograničen je samo zbog tehničkih mogućnosti jer kod velikog broja nizova uspoređivanje postaje nepregledno. To su **grafikoni s aritmetičkim mjerilom na osi ordinata.**

Postoje **grafikoni s logaritamskim mjerilom na osi ordinata**. Oni omogućuju prikazivanje i uspoređivanje vremenskih nizova s različitom jedinicom mjere na istom grafikonu. Ovim grafikonom mogu se prikazivati i uspoređivati i vremenski nizovi s istom mjernom jedinicom, ali koji imaju velike razlike među vrijednostima pojave.

Njihova konstrukcija danas je olakšana u raznim statističkim paketima za računalo (Statistica, SPSS i slično).

Ako je omjer najveće i najmanje frekvencije vremenskog niza:

$\frac{Y_{\max}}{Y_{\min}} \leq 10$, $\Rightarrow$ dovoljan je 1 logaritamski ciklus

$10 < \frac{Y_{\max}}{Y_{\min}} \leq 100$, $\Rightarrow$ dovoljna su 2 logaritamska ciklusa

$100 < \frac{Y_{\max}}{Y_{\min}} \leq 1000$, $\Rightarrow$ dovoljna su 3 logaritamska ciklusa... itd.

Karakteristika logaritamske skale na osi ordinata je da se za vrijednost s kojom počinje jedan logaritamski ciklus uvećava svaka sljedeća vrijednost u ciklusu. S vrijednošću kojom završava prvi logaritamski ciklus počinje drugi i uvećavanje na skali se sada nastavlja s tom veličinom i tako redom dalje.

Intervalni vremenski nizovi mogu se prikazivati:

- linijskim grafikonom
- površinskim grafikonom (obično su to stupci koji su naslonjeni jedan na drugi jer vrijeme teče kontinuirano).

Kod linijskog grafikona za intervalne vremenske nizove (koji prikazuju vrijednost pojave u vremenskim intervalima) vrijednost pojave se nanosi na sredinu svakog promatranog razdoblja (mjesec, kvartal, godina i slično.)

Stupci površinskog grafikona naslonjeni su na os apscisu, imaju jednake baze, a visina im odgovara vrijednostima vremenskog niza za određeno razdoblje. Dakle površine ovih stupaca su proporcionalne vrijednostima niza, a razlike u njihovoj veličini upućuju na apsolutne razlike vrijednosti promatranih razdoblja.

Ako intervalni vremenski niz prikazuje vrijednosti pojave za različite jedinice vremena, kod grafičkog prikaza radi se aproksimacija kojom se računaju prosječne vrijednosti pojave za manju ili veću vremensku jedinicu. Na primjer, ako jedan niz tvore godišnji i mjesečni podaci, onda se sve svodi ili na godišnje ili na mjesečne vrijednosti pojave ($Y_{mjes.} = Y_{god.} / 12$ jer godina ima 12 mjeseci).

Trenutačni vremenski nizovi mogu se prikazivati:

- linijskim grafikonom.

Kod linijskog grafikona za trenutačne vremenske nizove (koji prikazuju vrijednost pojave u trenutku vremena) vrijednost pojave se nanosi u trenutku mjerenja vrijednosti promatrane pojave (početak, sredina, kraj ili neki drugi precizan datum izabranih vremenskih razdoblja) i ovdje se za nizove čije točke promatranja u vremenu nisu jednako udaljene ne treba vršiti korekcija vrijednosti pojave.

Ako su na ordinati sve vrijednosti (Y_t) promatranog vremenskog niza velike, dopušteno je napraviti **vodoravan prekid grafikona na ordinati, na način, da se apscisa približi tim velikim vrijednostima niza prema gore** da bi se izbjegao prazni prostor na grafu. Ovaj vodoravan prekid grafikona može se napraviti samo za linijski grafikon.

Okomiti prekid grafikona radi se na apscisi, ako nisu poznate vrijednosti vremenskog niza za sva promatrana vremenska razdoblja (na primjer, ratne godine).

Primjer 3.3.1.

Tablica 3.1.

Odobreni krediti i štedni ulozi banke «Z» u razdoblju od 1995. do 2005. god.

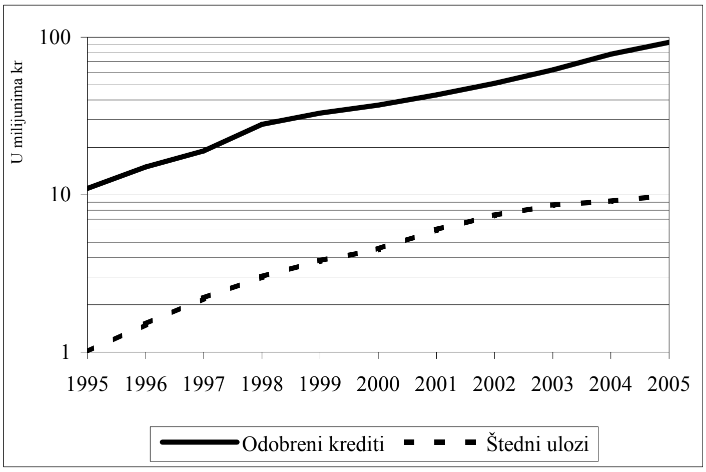
Godina	Odobreni krediti u milijunima kn	Štedni ulozi u milijunima kn
1995.	11	1
1996.	15	1,5
1997.	19	2,2
1998.	28	3
1999.	33	3,8
2000.	37	4,5
2001.	43	6
2002.	51	7,4
2003.	62	8,6
2004.	78	9,1
2005.	93	9,9

Izvor: Podaci banke «Z», travanj 2006. god.

Zadatak je zadane vremenske nizove prikazati grafički s 2 logaritamska ciklusa na os ordinate.

Grafikon 12.

Odobreni krediti i štedni ulozi banke «Z» u razdoblju od 1995. do 2005. god.



Izvor: Podaci banke «Z», travanj 2006. god.

3.4. Pokazatelji dinamike

Kretanje vrijednosti pojave vremenskog niza jasno se može vidjeti iz grafičkog prikaza. Međutim u statističkoj analizi često se javlja potreba preciznijeg definiranja kretanja vrijednosti neke pojave u vremenu. U tu svrhu služe apsolutni i relativni pokazatelji.

Apsolutni pokazatelji računaju se običnim oduzimanjem vrijednosti pojave u jednom vremenskom razdoblju od vrijednosti iste pojave u drugom razdoblju i izražavaju se u originalnim jedinicama mjere.

Pojedinačne apsolutne promjene od razdoblja do razdoblja računaju se tako da se od vrijednosti pojave u tekućem razdoblju oduzme vrijednost pojave u prethodnom razdoblju:

$$\Delta Y_t = Y_t - Y_{t-1}, \qquad t = 1,2,...,n \tag{3.4.1}$$

Tumače se kao **promjena vrijednosti pojave promatranog vremenskog niza u originalnim jedinicama mjere u tekućem razdoblju u odnosu na prethodno razdoblje**.

Pojedinačne apsolutne promjene u tekućem razdoblju u odnosu prema nekom baznom razdoblju računaju se tako da se od vrijednosti pojave u tekućem razdoblju oduzme vrijednost pojave u odabranom baznom razdoblju:

$$\Delta Y_t = Y_t - Y_b, \quad t = 1,2,...,n \tag{3.4.2}$$

Tumače se kao **promjena vrijednosti pojave promatranog vremenskog niza u originalnim jedinicama mjere u tekućem razdoblju u odnosu na odabrano bazno razdoblje**.

Relativni pokazatelji, za razliku od apsolutnih, omogućuju usporedbu kretanja pojava s različitim jedinicama mjere.

Individualni indeksi su relativni pokazatelji dinamike kretanja vrijednosti pojave vremenskog niza i njima se uspoređuje stanje jedne pojave u različitim vremenskim intervalima ili momentima.

Individualni indeksi dijele se na:

- **verižne indekse**
- **bazne indekse.**

3.5. Verižni indeksi i indeksi na stalnoj bazi

3.5.1. Verižni indeksi

Verižni indeksi pokazuju relativne promjene (u %) pojave u tekućem razdoblju u odnosu na prethodno razdoblje, odnosno pokazuju za koliko % se vrijednost pojave u jednom razdoblju promijenila u odnosu na prethodno razdoblje.

Verižni indeksi se računaju:

$$V_t = \frac{Y_t}{Y_{t-1}} \cdot 100, \quad t = 2,3,....N. \tag{3.5.1}$$

Iz izraza (3.5.1) vidi se da se verižni indeks računa tako da se stavi u odnos vrijednost pojave iz tekućeg razdoblja s vrijednošću pojave iz prethodnog razdoblja i sve se množi sa 100. S obzirom da vrijednost vremenskog niza za prethodno razdoblje od prvog nije poznata, ne može se izračunati prvi verižni indeks u jednom nizu. **Verižni indeksi se još nazivaju i lančani indeksi** jer pokazuju promjene pojave u uzastopnim vremenskim razdobljima i nadovezuju se jedan na drugi.

Verižni indeksi tumače se u postotcima preko stope promjene tako da se od njih oduzme 100.

$$S_t = V_t - 100 \qquad (3.5.2)$$

Grafički prikaz verižnih indeksa je specifičan. Crta se u pravokutnom sustavu gdje je apscisa pomaknuta do veličine 100 na osi ordinata. Ordinata se povlači za svako novo razdoblje. Za svaki novi vremenski period ucrtava se poseban linijski grafikon s ishodištem uvijek u prethodnom razdoblju. Tako se na grafikonu odmah može uočiti promjena vrijednosti promatrane pojave u jednom u odnosu na prethodno razdoblje.

Kako je već naglašeno pomoću verižnih indeksa računa se stopa promjene vrijednosti promatrane pojave iz razdoblja u razdoblje.

Stopa promjene je relativna (postotna) promjena vrijednosti neke pojave u tekućem u odnosu na prethodno razdoblje.

Računa se prema izrazu (3.5.3):

$$S_t = \left(\frac{Y_t}{Y_{t-1}} - 1 \right) \cdot 100, \quad t = 2,3,\ldots,N. \qquad (3.5.3)$$

Ako se u gornjoj jednadžbi vrijednosti u zagradama pomnože sa 100, vrijedi da je:

$$S_t = V_t - 100, \quad t = 2,3,\ldots,N., \qquad (3.5.4)$$

što je već dano izrazom (3.5.2).

Jednako kao kod verižnih indeksa stopa promjene ne računa se za prvo razdoblje, u nizu, jer vrijednost promatrane pojave ispred prve obično nije poznata.

Osim ovakve pojedinačne stope promjene iz razdoblja u razdoblje, statistička analiza često zahtijeva računanje prosječne stope promjene za čitavo promatrano razdoblje.

Prosječna stopa promjene je prosječna relativna (u %) promjena vrijednosti neke pojave kroz razdoblja u ukupnom promatranom vremenskom periodu.

Pretpostavka je da se vrijednost promatrane pojave u svakom razdoblju mijenja (raste ili pada) za jednak postotak kroz neki izabrani, dulji vremenski period.

Prosječna stopa promjene za neki dulji period računa se pomoću geometrijske sredine verižnih indeksa. Kako verižnih indeksa ima (N-1), jer se prvi u nizu ne računa, geometrijska sredina bit će (N-1) korijen od njihovih umnožaka:

$$G = \sqrt[N-1]{V_2 \cdot V_3 \cdot V_4 \ldots \cdot V_{N-1} \cdot V_N} = \sqrt[N-1]{\frac{Y_2}{Y_1} \cdot \frac{Y_3}{Y_2} \cdot \frac{Y_4}{Y_3} \cdot \ldots \cdot \frac{Y_{N-1}}{Y_{N-2}} \cdot \frac{Y_N}{Y_{N-1}}} \,. \quad (3.5.5)$$

Iz desnog dijela jednakosti (3.5.5) može se vidjeti da se mogu kratiti sve vrijednosti pojave osim Y_1 i Y_N, stoga vrijedi da je geometrijska sredina verižnih indeksa:

$$G = \sqrt[N-1]{\frac{Y_N}{Y_1}}\,, \qquad (3.5.6)$$

dakle, pod korijenom ostaje omjer posljednje i prve frekvencije vremenskog niza.

Prosječna stopa promjene računa se prema (3.5.7):

$$\bar{S} = (G-1) \cdot 100 \,. \qquad (3.5.7)$$

Ako su zadani godišnji podaci onda je to prosječna godišnja stopa promjene, ako su podaci dani po mjesecima, riječ je o prosječnoj mjesečnoj stopi promjene i slično.

Može se vršiti i preračunavanje prosječne stope promjene s duljeg na kraće i s kraćeg na dulje vremensko razdoblje. Na primjer, ako se želi izračunati prosječna mjesečna stopa promjene od prosječne godišnje stope bit će:

$$G_{mj} = \sqrt[12]{G_{god}}\,, \quad \bar{S}_{mj} = (G_{mj} - 1) \cdot 100\,, \qquad (3.5.8)$$

dakle, računa se dvanaesti korijen od godišnje geometrijske sredine jer godina ima 12 mjeseci.

Na primjer, ako se želi izračunati prosječna godišnja stopa promjene od prosječne polugodišnje stope bit će:

$$G_{god.} = G_{polug.}^2\,, \quad \bar{S}_{polug.} = (G_{polug} - 1) \cdot 100\,, \qquad (3.5.9)$$

odnosno računa se na drugu potenciju od polugodišnje geometrijske sredine jer godina ima 2 polugodišta.

Uz pretpostavku da će se vrijednosti neke pojave nastaviti kretati i u budućnosti na isti način, odnosno **prema izračunatoj prosječnoj stopi promjene kao i u promatranom razdoblju** preko geometrijske sredine **može se, počevši od posljednjeg elementa (Y_N) u nizu, vršiti prognoza njenog kretanja**:

$$\hat{Y}_{N+t} = Y_N \cdot G^t\,, \qquad (3.5.10)$$

gdje je:

$\hat{Y}_{N+t}$ - prognostička vrijednost pojave uz pretpostavku neizmijenjenog G u (N+t) razdoblju

Y_N - posljednja vrijednost pojave u nizu

G - izračunata ili pretpostavljena geometrijska sredina verižnih indeksa

t - broj vremenskih razdoblja nakon posljednjeg u nizu, za koje se vrši prognoza.

Primjer 3.5.1.

Tablica 3.2.

Proizvodnja vina u vinariji «Z» u razdoblju od 1996. do 2005. god.

Godina	Proizvedene količine vina u *l*	Verižni indeksi (V_t)	Stope promjene (S_t)
1996.	1538	-	-
1997.	1709	111,1	11,1
1998.	1552	90,8	-9,2
1999.	1860	119,9	19,9
2000.	1962	105,5	5,5
2001.	2176	110,9	10,9
2002.	2097	96,4	-3,6
2003.	2263	107,9	7,9
2004.	2481	109,6	9,6
2005.	2315	93,3	-6,7

Izvor: Podaci vinarije «Z», 2006. god.

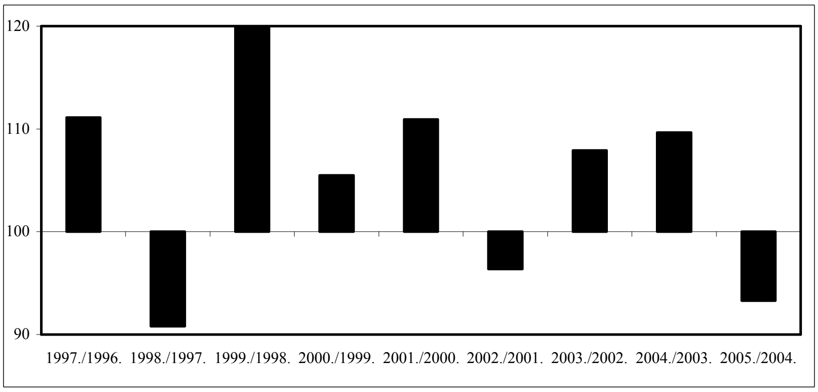
Zadatak je izračunati verižne indekse i pripadajuće stope promjene te indekse prikazati grafički stupcima.

Vrijednosti verižnih indeksa navedene su u trećem stupcu tablice, a pripadajuće stope promjene u četvrtom stupcu. Prvi izračunani verižni indeks iznosi 111,1 i pokazuje da je na svakih 100 litara vina proizvedenih u 1996. godini dolazilo 111, 1 litara proizvedenih u 1997. godini, ili za 11,1% više.

Grafički prikaz verižnih indeksa stupcima je sljedeći:

Grafikon 13.

Verižni indeksi proizvodnje vina u vinariji «Z»



Izvor: Podaci vinarije «Z», 2006. god.

3.5.2 Indeksi na stalnoj bazi

Indeksi na stalnoj bazi ili bazni indeksi pokazuju relativne promjene (u %) pojave u tekućem razdoblju u odnosu na neko odabrano bazno razdoblje, odnosno pokazuju za koliko % se vrijednost pojave u jednom razdoblju promijenila u odnosu na odabrano bazno razdoblje.

Bazni indeksi se računaju:

$$I_t = \frac{Y_t}{Y_b} \cdot 100, \quad t = 1,2,....N., \tag{3.5.11}$$

gdje je Y_b vrijednost pojave u nekom izabranom baznom razdoblju. Kod baznih indeksa **uobičajen je zapis za bazno razdoblje** da je jednako 100, **(b=100).**

Iz izraza (3.5.11) se vidi, da se bazni indeks računa tako da se postavi u odnos vrijednost pojave iz tekućeg razdoblja s vrijednošću pojave iz odabranog razdoblja i pomnoži sa 100. **Bazni indeksi se tumače u postotcima** tako da se od njih oduzme 100.

$$I_t - 100 = (\%)\,. \tag{3.5.12}$$

Pri provođenju statističke analize u praksi treba pažljivo birati bazno razdoblje jer se pogrešnim izborom mogu dobiti iskrivljene predodžbe o dinamici pojave. Ako se na primjer za bazno razdoblje odabere takvo razdoblje u kojemu je vrijednost pojave najmanja u nizu, izračunati bazni indeksi će pokazivati porast u odnosu na bazu. Ako se, suprotno, za bazno razdoblje odabere ono u kojemu je vrijednost pojave najveća u nizu, bazni indeksi će pokazivati stalan pad u odnosu na izabranu bazu. Na taj način se u praksi može manipulirati podacima.

Grafički prikaz baznih indeksa je jednostavan linijski grafikon. Crta se u pravokutnom sustavu, i na njemu mora biti naznačeno koje je razdoblje uzeto za bazno, uz sve ostale oznake koje statistički grafikon mora imati (naslov, izvor, oznake na ordinati i oznake na apscisi).

Kako je već rečeno, bazni indeksi se računaju dijeljenjem svakog člana niza istim brojem (bazom) i množenjem istim faktorom (sa 100). Prema tome se može zaključiti da su **bazni indeksi upravno proporcionalni originalnim vrijednostima vremenskog niza, pa sve što se može izračunati dijeljenjem originalnih vrijednosti pojave, može se dobiti i dijeljenjem baznih indeksa** (naravno po istoj bazi).

Stoga vrijedi da je **geometrijska sredina verižnih indeksa**:

$$G = \sqrt[N-1]{\frac{Y_N}{Y_1}} \Rightarrow G = \sqrt[N-1]{\frac{I_N}{I_1}}\,, \tag{3.5.13}$$

gdje su I_1 i I_N, prvi i posljednji bazni indeks u nizu.

Preračunavanje baznih indeksa po jednoj bazi u bazne indekse po drugoj bazi se vrši na sljedeći način:

$$I_t{*} = \frac{I_t}{I_b} \cdot 100\,, \tag{3.5.14}$$

gdje je b novo bazno razdoblje za indekse $I_t{*}$.

Računanje verižnih indeksa preko baznih indeksa je:

$$V_t = \frac{Y_t}{Y_{t-1}} \cdot 100 \Rightarrow V_t = \frac{I_t}{I_{t-1}} \cdot 100, \quad t = 2,3,....N\,, \tag{3.5.15}$$

gdje su I_{t-1} i I_t bazni indeksi jednake baze.

Preračunavanje verižnih indeksa u bazne po nekoj bazi b vrši se preko sljedećih izraza izvedenih iz (3.5.15):

- **za razdoblja prije baznog (b=100), računa se unatrag:**

$$I_{t-1} = \frac{I_t}{V_t} \cdot 100\,, \tag{3.5.16}$$

- **za razdoblja poslije baznog (b=100), računa se unaprijed:**

$$I_t = \frac{V_t \cdot I_{t-1}}{100}\,. \tag{3.5.17}$$

Primjer 3.5.2.

Tablica 3.3.

Uvoz banana na području «Z» u razdoblju od 1996. do 2005. god.

Godina	Uvezene količine banana u tisućama t	Bazni indeksi 1996.=100 (I_t)	Stope promjene (S_t)
1996.	21	100,0	0,0
1997.	19	90,5	-9,5
1998.	25	119,0	19,0
1999.	20	95,2	-4,8
2000.	18	85,7	-14,3
2001.	23	109,5	9,5
2002.	25	119,0	19,0
2003.	28	133,3	33,3
2004.	24	114,3	14,3
2005.	27	128,6	28,6

Izvor: Statistika područja «Z», 2006. god.

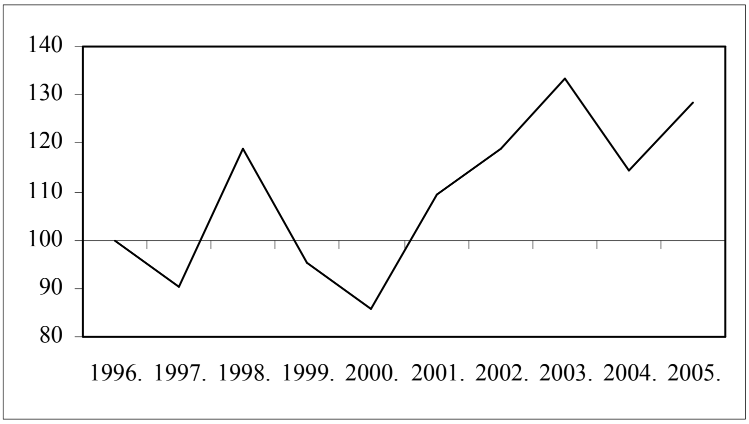
Zadatak je izračunati bazne indekse (1996.=100) i pripadajuće stope promjene te indekse prikazati linijskim grafikonom.

Vrijednosti baznih indeksa navedene su u trećem stupcu tablice, a pripadajuće stope promjene u četvrtom stupcu. Posljednji izračunani bazni indeks iznosi 128,6 i pokazuje da je na svakih 100 t uvezenih banana u 1996. godini dolazilo 128,6 t uvezenih u 2005. godini, ili za 28,6% više.

Grafički prikaz baznih indeksa linijskim grafikonom je sljedeći:

Grafikon 14.

Bazni indeksi uvoza banana na području «Z», 1996.=100



Izvor: Statistika područja «Z», 2006. god.

3.6. Skupni indeksi

Skupni indeksi su relativni pokazatelji dinamike kretanja vrijednosti skupine pojava vremenskog niza koje su na neki način povezane ili su međusobno slične po nekim karakteristikama.

Razdoblje u kojem se iskazuje dinamika skupine pojava naziva se **tekuće ili izvještajno razdoblje (često se označava s 1),** dok se razdoblje s kojim se uspoređuje dinamika naziva **bazno ili početno razdoblje (često se označava s 0).**

Najčešće se računaju:

- **skupni indeks cijena**
- **skupni indeks količina**
- **skupni indeks vrijednosti.**

U praksi postoji problem računanja ovih skupnih indeksa jer nema jedinstvenog izraza kojim bi se oni mogli izračunavati. **U upotrebi su Laspeyresov i Paascheov [9] oblik skupnih indeksa.** Osnovna razlika je u tome da **Laspeyresovi indeksi imaju pondere iz baznog (0) razdoblja, a Paascheovi indeksi imaju pondere iz tekućeg (1) razdoblja.**

Postoje još i neki posebni pojavni oblici ovih indeksa, na primjer **indeksi potrošačkih cijena su posebni oblici indeksa cijena** koje ulaze u potrošačke cijene.

3.6.1. Skupni indeksi cijena

Skupni indeks cijena je relativni pokazatelj dinamike kretanja cijena skupine pojava u tekućem razdoblju u odnosu na bazno razdoblje.

Oblici ovog indeksa su:

a) **Laspeyresov skupni indeks cijena i**

b) **Paascheov skupni indeks cijena.**

[9] Prezimena poznatih statističara.

Laspeyresov skupni indeks cijena je vagana aritmetička sredina individualnih indeksa cijena, gdje su ponderi količine iz baznog, odnosno nultog razdoblja.

Laspeyresov skupni indeks cijena pokazuje za koliko posto su se promijenile cijene skupine pojava zajedno u izvještajnom u odnosu na bazno razdoblje, računajući uz neizmijenjene količine iz baznog razdoblja.

Agregatni oblik ovog indeksa je:

$$P_{01}(q_0) = \frac{\sum_{i=1}^{k} p_{i1} q_{i0}}{\sum_{i=1}^{k} p_{i0} q_{i0}} \cdot 100\,, \qquad (3.6.1)$$

gdje je:

$\sum_{i=1}^{k} p_{i1} q_{i0}$ - suma umnožaka cijena iz izvještajnog i količina iz nultog razdoblja

$\sum_{i=1}^{k} p_{i0} q_{i0}$ - suma umnožaka cijena iz nultog i količina iz nultog razdoblja.

Paascheov skupni indeks cijena je vagana aritmetička sredina individualnih indeksa cijena, gdje su ponderi količine iz izvještajnog, odnosno tekućeg razdoblja.

Paascheov skupni indeks cijena pokazuje za koliko posto su se promijenile cijene skupine pojava zajedno u izvještajnom u odnosu na bazno razdoblje, računajući uz neizmijenjene količine iz izvještajnog razdoblja.

Agregatni oblik ovog indeksa je:

$$P_{01}(q_1) = \frac{\sum_{i=1}^{k} p_{i1} q_{i1}}{\sum_{i=1}^{k} p_{i0} q_{i1}} \cdot 100\,, \qquad (3.6.2)$$

gdje je:

$\sum_{i=1}^{k} p_{i1} q_{i1}$ - suma umnožaka cijena iz izvještajnog i količina iz izvještajnog razdoblja

$\sum_{i=1}^{k} p_{i0} q_{i1}$ - suma umnožaka cijena iz nultog i količina iz izvještajnog razdoblja.

Određene ekonomske analize često u praksi zahtijevaju posebne oblike skupnog indeksa cijena. Od posebne je važnosti **indeks potrošačkih cijena**[10]

---

[10] Do kraja 2003. godine u Republici Hrvatskoj je u službenim statistikama u istoj funkciji bio **indeks troškova života**. **Indeks potrošačkih cijena** uveden je od 2004.

koji odražava promjene cijena dobara i usluga koje koristi referentno stanovništvo radi finalne potrošnje.

Indeks potrošačkih cijena služi za mjerenje inflacije (odnosno porasta cijena) **u privredi, za očuvanje vrijednosti kod ugovora s indeksnim klauzulama kao osnova za deflacioniranje** (uklanjanje utjecaja inflacije, tj. porasta cijena) **određenih vrijednosnih pokazatelja i slično.**

Državni zavod za statistiku Republike Hrvatske ovaj indeks računa na osnovi reprezentativne košarice koju čini oko 540 proizvoda. Svaki mjesec prikuplja se više od 25 000 cijena na unaprijed definiranom uzorku prodajnih mjesta na devet lokacija u zemlji (Zagreb, Slavonski Brod, Osijek, Sisak, Rijeka, Pula, Split, Dubrovnik i Varaždin), odabranih prema kriteriju broja stanovnika i reprezentativnosti za pojedinu statističku regiju.

Pomoću ovog indeksa mjeri se utjecaj potrošačkih cijena na nominalne plaće, odnosno **računaju se realne plaće**.

$$\text{Realne plaće} = \frac{\text{Nominalne pl.}}{\text{Indeksi potr. cijena}} \cdot 100\,, \qquad (3.6.3)$$

gdje su **indeksi u nazivniku izraza bazni po nekom odabranom razdoblju**.

Vrijedi da je:

$$Ind.real.\,pl. = \frac{Ind.no\min.\,pl.}{Ind.\,potr.\ cijena} \cdot 100\,, \qquad (3.6.4)$$

gdje su svi **indeksi bazni po nekom (istom) odabranom razdoblju** radi usporedivosti podataka.

1.6.2. Skupni indeksi količina

Skupni indeks količina je relativni pokazatelj dinamike kretanja količina skupine pojava u tekućem razdoblju u odnosu na bazno razdoblje.

Oblici ovog indeksa su:

a) **Laspeyresov skupni indeks količina**

b) **Paascheov skupni indeks količina.**

---

godine da bi se službena statistička izvješća u Hrvatskoj uskladila sa svjetskim obračunima i pokazateljima.

Laspeyresov skupni indeks količina je vagana aritmetička sredina individualnih indeksa količina, gdje su ponderi cijene iz baznog, odnosno nultog razdoblja.

Laspeyresov skupni indeks količina pokazuje za koliko posto su se promijenile količine skupine pojava zajedno u izvještajnom u odnosu na bazno razdoblje, računajući uz neizmijenjene cijene iz baznog razdoblja.

Agregatni oblik ovog indeksa je:

$$Q_{01}(p_0) = \frac{\sum_{i=1}^{k} q_{i1} p_{i0}}{\sum_{i=1}^{k} q_{i0} p_{i0}} \cdot 100 \,, \qquad (3.6.5)$$

gdje je:

$\sum_{i=1}^{k} q_{i1} p_{i0}$ - suma umnožaka količina iz izvještajnog i cijena iz nultog razdoblja

$\sum_{i=1}^{k} q_{i0} p_{i0}$ - suma umnožaka količina iz nultog i cijena iz nultog razdoblja.

Paascheov skupni indeks količina je vagana aritmetička sredina individualnih indeksa količina, gdje su ponderi cijene iz izvještajnog, odnosno tekućeg razdoblja.

Paascheov skupni indeks količina pokazuje za koliko posto su se promijenile količine skupine pojava zajedno u izvještajnom u odnosu na bazno razdoblje, računajući uz neizmijenjene cijene iz izvještajnog razdoblja.

Agregatni oblik ovog indeksa je:

$$Q_{01}(p_1) = \frac{\sum_{i=1}^{k} q_{i1} p_{i1}}{\sum_{i=1}^{k} q_{i0} p_{i1}} \cdot 100 \,, \qquad (3.6.6)$$

gdje je:

$\sum_{i=1}^{k} q_{i1} p_{i1}$ - suma umnožaka količina iz izvještajnog i cijena iz izvještajnog razdoblja

$\sum_{i=1}^{k} p_{i0} q_{i1}$ - suma umnožaka cijena iz nultog i količina iz izvještajnog razdoblja.

Posebni oblik skupnog indeksa količina je **skupni indeks fizičkog obujma**.

Najprije se pomoću skupnog indeksa cijena vrši postupak deflacioniranja vrijednosnih pokazatelja. Na taj način se iz vrijednosnih pokazatelja odstranjuje utjecaj cijena (najčešće inflacije, tj. porasta cijena).

$$\text{Vrijed. u stalnim cijenama} = \frac{\text{Vrijed. u tek. cijenama}}{\text{Indeksi cijena}} \cdot 100, \qquad (3.6.7)$$

gdje su **indeksi u nazivniku izraza bazni po nekom odabranom razdoblju**.

Nakon toga se može sagledati kretanje vrijednosnih pokazatelja u fizičkom obujmu:

$$Ind.fiz.obujma = \frac{Ind.vrijed.u\,tek.cij.}{Indeks\ cijena} \cdot 100, \qquad (3.6.8)$$

gdje su svi **indeksi bazni po nekom (istom) odabranom razdoblju**, radi usporedivosti podataka.

3.6.3. Skupni indeksi vrijednosti

Skupni indeks vrijednosti je relativni pokazatelj dinamike kretanja vrijednosti skupine pojava u tekućem razdoblju u odnosu na bazno razdoblje.

Vrijednost je umnožak količine i cijene nekog proizvoda (i):

$$V_i = q_i p_i \qquad (3.6.9)$$

Skupni indeks vrijednosti je omjer vrijednosti skupine pojava u izvještajnom i vrijednosti skupine pojava u baznom razdoblju.

Skupni indeks vrijednosti pokazuje za koliko posto su se promijenile vrijednosti skupine pojava zajedno u izvještajnom u odnosu na bazno razdoblje.

Računa se:

$$V_{01} = \frac{\sum_{i=1}^{k} q_{i1} p_{i1}}{\sum_{i=1}^{k} q_{i0} p_{i0}} \cdot 100, \qquad (3.6.10)$$

gdje je:

$\sum_{i=1}^{k} q_{i1} p_{i1}$ - suma umnožaka količina i cijena iz izvještajnog razdoblja

$\sum_{i=1}^{k} q_{i0} p_{i0}$ - suma umnožaka količina i cijena iz nultog razdoblja.

Isti rezultat može se dobiti i množenjem Laspeyresovog skupnog indeksa cijena i Paascheovog skupnog indeksa količina:

$$V_{01} = \frac{P_{01}(q_0) \cdot Q_{01}(p_1)}{100} = \frac{\dfrac{\sum_{i=1}^{k} p_{i1} q_{i0}}{\sum_{i=1}^{k} p_{i0} q_{i0}} \cdot 100 \cdot \dfrac{\sum_{i=1}^{k} q_{i1} p_{i1}}{\sum_{i=1}^{k} q_{i0} p_{i1}} \cdot 100}{100} = \frac{\sum_{i=1}^{k} q_{i1} p_{i1}}{\sum_{i=1}^{k} q_{i0} p_{i0}} \cdot 100 \qquad (3.6.11)$$

ili množenjem Paascheovog skupnog indeksa cijena i Laspeyresovog skupnog indeksa količina:

$$V_{01} = \frac{P_{01}(q_1) \cdot Q_{01}(p_0)}{100} = \frac{\dfrac{\sum_{i=1}^{k} p_{i1} q_{i1}}{\sum_{i=1}^{k} p_{i0} q_{i1}} \cdot 100 \cdot \dfrac{\sum_{i=1}^{k} q_{i1} p_{i0}}{\sum_{i=1}^{k} q_{i0} p_{i0}} \cdot 100}{100} = \frac{\sum_{i=1}^{k} q_{i1} p_{i1}}{\sum_{i=1}^{k} q_{i0} p_{i0}} \cdot 100 \qquad (3.6.12)$$

Primjer 3.6.1.

Treba izračunati indekse količina i cijena po Laspeyresovom i Paascheovom obrascu za jedinicu koja proizvodi tri različita proizvoda. Podaci o proizvodnji tri proizvoda u 2007. i 2008. godini su sljedeći:

Tablica 3.4.

Proizvod	Mjerna jedinica	Količine 2007. q_0	Količine 2008. q_1	Cijene 2007. p_0	Cijene 2008. p_1
A	komad	10	12	25	30
B	litra	30	40	40	40
C	m^2	20	15	60	72

Izvor: Podaci su simulirani

Zadatak je izračunati:
a) za koliko % su se u prosjeku promijenile cijene u 2008. godini u odnosu na 2007. godinu?
b) za koliko % su se u prosjeku promijenile proizvedene količine u 2008. godini u odnosu na 2007. godinu?
c) za koliko % se u prosjeku promijenila vrijednost proizvodnje u 2008. godini u odnosu na 2007. godinu?

Za izračun navedenih indeksa potrebni su sljedeći izračuni:

Proizvod	Mjerna jedinica	q_1p_1	$q_0\,p_0$	q_0p_1	q_1p_0
A	komad	360	250	300	300
B	litra	1600	1200	1200	1600
C	m^2	1080	1200	1440	900
Ukupno		3040	2650	2940	2800

a) Laspeyresovom skupni indeks količina glasi:

$$Q_{01}(p_0)=\frac{\sum_{i=1}^{k} q_{i1}p_{i0}}{\sum_{i=1}^{k} q_{i0}p_{i0}}\cdot 100=\frac{2800}{2650}\cdot 100=105{,}7$$

Paascheovom skupni indeks količina iznosi:

$$Q_{01}(p_1)=\frac{\sum_{i=1}^{k} q_{i1}p_{i1}}{\sum_{i=1}^{k} q_{i0}p_{i1}}\cdot 100=\frac{3040}{2940}\cdot 100=103{,}4$$

Indeks količina izračunan po Laspeyresovom obrascu pokazuje da su se proizvedene količine tri proizvoda 2008. godine u odnosu na 2007. godinu u prosjeku povećale za 5.7%, dok su upotrebom Paascheovog obrasca povećane u prosjeku za 3.4%.

b) Laspeyresovom skupni indeks cijena glasi:

$$P_{01}(q_0)=\frac{\sum_{i=1}^{k} p_{i1}q_{i0}}{\sum_{i=1}^{k} p_{i0}q_{i0}}\cdot 100=\frac{2940}{2650}\cdot 100=110{,}9$$

Paascheovom skupni indeks cijena iznosi:

$$P_{01}(q_1)=\frac{\sum_{i=1}^{k} p_{i1}q_{i1}}{\sum_{i=1}^{k} p_{i0}q_{i1}}\cdot 100=\frac{3040}{2800}\cdot 100=108{,}6$$

Indeks cijena izračunan po Laspeyresovom obrascu pokazuje da su se cijene tri proizvoda 2008. godine u odnosu na 2007. godinu u prosjeku povećale za 10.9%. Upotrebom Paascheovog obrasca indeks cijena pokazuje porast od 8.6% u prosjeku.

c) Skupni indeks vrijednosti proizvodnje iznosi:

$$V_{01} = \frac{\sum_{i=1}^{k} q_{i1} p_{i1}}{\sum_{i=1}^{k} q_{i0} p_{i0}} \cdot 100 = \frac{3040}{2650} \cdot 100 = 114{,}7$$

Indeks vrijednosti proizvodnje pokazuje da se vrijednost proizvodnje tri proizvoda 2008. godine u odnosu na 2007. godinu u prosjeku povećala za 14,7%.

3.7. Modeli trendova

Najčešće su u upotrebi **trend-modeli**:

1. Trend polinomi k-tog stupnja
2. Eksponencijalni trend modeli
3. Hiperbolički trend modeli
4. Asimptotski trend modeli.

3.7.1. Trend polinomi k-tog stupnja

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 + ... + \hat{\beta}_k X^k \qquad (3.7.1)$$

Ocjena parametara najčešće se dobije metodom najmanjih kvadrata zbog njenih optimalnih svojstava. Kaže se da se tom metodom dobiju najbolje linearne nepristrane ocjene parametara (BLUE), naravno uz uvjet da su ispunjene sve pretpostavke koje ova metoda zahtijeva.

- Ako je stupanj polinoma 1, tj. $k = 1 \Rightarrow$ ocjenjuje se model linearnog trenda (odnosno kada se vrijednost pojave u svakoj vremenskoj jedinici mijenja za približno isti apsolutni iznos $\Delta\hat{Y}_t = \hat{Y}_t - \hat{Y}_{t-1} = \hat{\beta}_1$)

- Ako je stupanj polinoma $k \Rightarrow$ ocjenjuje se model trend polinoma k-tog stupnja, (odnosno ako su k-te diferencije vrijednosti vremenskog niza približno konstantne ocjenjuje se (k+1) parametar trend polinoma):

  $$\Delta^k y_t \approx const. \tag{3.7.2}$$

3.7.1.1. Model linearnog trenda

Model linearnog trenda objašnjava linearno kretanje (pozitivno ili negativno) vrijednosti promatranog vremenskog niza kroz vrijeme. Osim prikaza linearnog kretanja pojave vremenskog niza na temelju ocijenjenog modela može se vršiti predviđanje vrijednosti pojave za neka buduća razdoblja.

Model linearnog trenda općenito je oblika:

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + e_t , \tag{3.7.3}$$

gdje je:

Y- ovisna varijabla, tj. vrijednosti vremenskog niza

X - neovisna varijabla, tj. vrijeme (treba napomenuti da se kod trend modela mora izabrati ishodišno razdoblje kojemu se dodjeljuje vrijednost 0. Ako nula nije prva u nizu onda razdoblja prije nultog unatrag imaju vrijednosti: -1,-2,-3,…, a prema naprijed su vrijednosti: 1,2,3,…)

e - slučajna komponenta.

Model ima slučajnu komponentu e, koja upućuje da veze između vrijednosti pojave vremenskog niza i vremena u praksi nisu funkcionalne, nego su statističke ili stohastičke, odnosno oko linije konkretnog linearnog trend modela postoje pozitivna i/ili negativna odstupanja originalnih vrijednosti.

Ocijenjeni linearni trend model je:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X . \tag{3.7.4}$$

Parametri $\hat{\beta}_0$ i $\hat{\beta}_1$ ocjenjuju se tako da pravac $\hat{Y}$ prolazi između stvarnih točaka vremenskog niza i da najbolje tumači vezu između njih, odnosno pravac mora biti takav da odstupanja e_t budu najmanja.

Postoji više različitih metoda za ocjenu ovih parametara, a najčešće korištena metoda je **metoda najmanjih kvadrata,** koja upravo procjenjuje parametre $\hat{\beta}_0$ i $\hat{\beta}_1$ tako da odstupanja e_t budu najmanja. Ona kao i kod regresije daje najbolje linearne nepristrane ocjene i vrlo je često primjenjivana metoda za ocjenu parametara.

$$Y_t = \hat{Y}_t + e_t \tag{3.7.5}$$

Odstupanja originalnih vrijednosti od ocijenjenih vrijednosti e_t mogu biti pozitivna i negativna. Da se međusobno ne bi poništile pozitivne i negativne vrijednosti, **metoda minimalizira sumu kvadrata od** e_t **.**

$$\min \sum_{t=1}^{n} e_t^{\,2} = \min \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2 = \min \sum_{t=1}^{n} \left[Y_t - (\hat{\beta}_0 + \hat{\beta}_1 X_t) \right]^2 \tag{3.7.6}$$

Nakon primjene matematičkog postupka traženja minimuma kao i kod regresije dobiju se dvije jednadžbe s dvije nepoznanice. Tako dobiveni sustav uvijek ima rješenja i **vrijedi da je:**

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t Y_t - n\overline{XY}}{\sum_{t=1}^{n} X_t^{\,2} - n\overline{X}^2} \quad i \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \, , \tag{3.7.7}$$

gdje su:

$$\overline{X} = \frac{\sum_{t=1}^{n} X_t}{n} \quad i \quad \overline{Y} = \frac{\sum_{t=1}^{n} Y_t}{n} \, , \tag{3.7.8}$$

jednostavne aritmetičke sredine varijabli X i Y.

Konačno je ocijenjeni model:

$$\begin{aligned} &\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X, \\ &jedinica \;\; za \;\; X = ... \\ &jedinica \;\; za \;\; Y = ... \\ &X = 0, \;\; ...... \end{aligned} , \tag{3.7.9}$$

gdje je $\hat{\beta}_0$ **konstantni član, tj. očekivana vrijednost vremenskog niza u ishodišnom razdoblju** (kada je nezavisna varijabla jednaka nuli tj. $\hat{\beta}_0 = \overline{Y}$ kada je X=0).

Regresijski koeficijent $\hat{\beta}_1$ **pokazuje prosječnu promjenu zavisne varijable kada nezavisna varijabla poraste za jedinicu vremena** (to u ovisnosti o vremenskoj jedinici može biti prosječna godišnja promjena vrijednosti vremenskog niza, prosječna mjesečna promjena … i tako dalje**).**

Nakon ocijene parametara trend modela postavlja se pitanje **reprezentativnosti**, odnosno sposobnosti **modela da objasni kretanje ovisne varijable vremenskog niza Y kroz vrijeme X.**

U tu svrhu koriste se neki **apsolutni i relativni pokazatelji.** Ovi se pokazatelji, kao i kod regresijske analize, temelje na raspodjeli odstupanja vrijednosti ovisne varijable Y_t u trend modelu od njene aritmetičke sredine $\overline{Y}$ i njenih očekivanih vrijednosti $\hat{Y}_t$.

Vrijedi da je:

$$SP = \sum_{t=1}^{n} \left(\hat{Y}_t - \overline{Y}\right)^2 = \hat{\beta}_0 \cdot \sum_{t=1}^{n} Y_t + \hat{\beta}_1 \sum_{t=1}^{n} X_t Y_t - n \cdot \overline{Y}^2 \qquad (3.7.10)$$

Dakle **SP je suma kvadrata protumačenog dijela,** odstupanja vrijednosti varijable vremenskog niza Y od aritmetičke sredine, **odnosno suma kvadrata odstupanja ocijenjenih vrijednosti varijable Y od aritmetičke sredine**.

$$SR = \sum_{t=1}^{n} \left(Y_t - \hat{Y}_t\right)^2 = \sum_{t=1}^{n} Y_t^2 - \hat{\beta}_0 \cdot \sum_{t=1}^{n} Y_t - \hat{\beta}_1 \sum_{t=1}^{n} X_t Y_t \qquad (3.7.11)$$

SR je suma kvadrata neprotumačenog dijela, odstupanja vrijednosti varijable Y od aritmetičke sredine, **odnosno suma kvadrata odstupanja originalnih ili empirijskih vrijednosti varijable Y od ocijenjenih vrijednosti**. Ova odstupanja su u stvari slučajne pogrješke e_t .

$$ST = \sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2 = \sum_{t=1}^{n} Y_t^2 - n \cdot \overline{Y}^2 \qquad (3.7.12)$$

ST je suma kvadrata ukupnih odstupanja vrijednosti varijable vremenskog niza Y od aritmetičke sredine.

Vrijedi da je:

$$SP + SR = ST \text{ ,} \qquad (3.7.13)$$

Ovaj izraz zove se jednadžba analize varijance i predstavlja temelj analize reprezentativnosti trend modela.

Standardna pogrješka trend modela je apsolutni pokazatelj reprezentativnosti trend modela, a pokazuje prosječni stupanj varijacije stvarnih vrijednosti ovisne varijable u odnosu na očekivane trend vrijednosti.

$$\hat{\sigma}_{\hat{Y}} = \sqrt{\frac{SR}{n-2}} \qquad (3.7.14)$$

Izraz (3.7.14) je standardna pogrješka linearnog trend modela. Ovaj pokazatelj izražen je u originalnim jedinicama mjere varijable vremenskog niza Y. Stoga je na temelju standardne pogrješke trend modela teško uspoređivati reprezentativnost modela s različitim mjernim jedinicama.

Taj problem eliminira **relativni pokazatelj - koeficijent varijacije trend modela, koji predstavlja postotak standardne pogrješke trenda od aritmetičke sredine varijable Y.**

$$\hat{V}_{\hat{Y}} = \frac{\hat{\sigma}_{\hat{Y}}}{\overline{Y}} \cdot 100 \tag{3.7.15}$$

Najmanja vrijednost koeficijenta varijacije je 0%, a najveća nije definirana. **Što je koeficijent varijacije trend modela bliži nuli, to je model reprezentativniji.**

Koeficijent determinacije je pokazatelj reprezentativnosti trend modela koji se temelji na analizi varijance. On se definira kao **omjer sume kvadrata odstupanja protumačenih trend modelom i sume kvadrata ukupnih odstupanja.**

$$r^2 = \frac{SP}{ST} \tag{3.7.16}$$

Koeficijent determinacije pokazuje koliko % je sume kvadrata odstupanja vrijednosti varijable Y od aritmetičke sredine protumačeno trend modelom.

Vrijedi da je:

$$r^2 = 1 - \frac{SR}{ST} \tag{3.7.17}$$

Vrijednost koeficijenta determinacije kreće se u intervalu $0 \le r^2 \le 1$. Trend model je reprezentativniji ako je ovaj pokazatelj bliži 1. Teorijska granica reprezentativnosti modela je 0,9. U praksi je ponekad vrlo teško pronaći varijablu koja dobro objašnjava kretanje vremenskog niza u vremenu, stoga se ta granica reprezentativnosti spušta i niže.

Primjer 3.7.1.

Tablica 3.5.

Zaposlenost u Hrvatskoj u razdoblju od 1995. do 2004. god., godišnji prosjek u tisućama

Godina	Broj zaposlenih (y_t)	Varijabla *vrijeme* (x_t)	$x_t y_t$	x_t^2	y_t^2	Trend vrijednosti ($\hat{y}_t$)
1995.	1196	0	0	0	1430416	1183,1
1996.	1195	1	1195	1	1428025	1200,6
1997.	1187	2	2374	4	1408969	1218,1
1998.	1272	3	3816	9	1617984	1235,5
1999.	1263	4	5052	16	1595169	1253,0
2000.	1258	5	6290	25	1582564	1270,4
2001.	1272	6	7632	36	1617984	1287,9
2002.	1289	7	9023	49	1661521	1305,3
2003.	1330	8	10640	64	1768900	1322,8
2004.	1355	9	12195	81	1836025	1340,3
Ukupno	12617	45	58217	285	15947557	12617

Izvor: Statistički ljetopis Republike Hrvatske 2005. god., str. 132.

Zadatak je ocijeniti model linearnog trenda s ishodištem na početku vremenskog razdoblja i njegovu reprezentativnost te izračunati trend vrijednosti.

Ocijenjeni model linearnog trenda glasi:

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$: $\quad \hat{Y} = 1183{,}1 + 17{,}46X$

X=0, 1995. godine
Jedinica za X je jedna godina
Jedinica za Y je tisuću zaposlenih

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t Y_t - n\overline{X}\overline{Y}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2} = \frac{58217 - 10 \cdot 4{,}5 \cdot 1261{,}7}{285 - 10 \cdot 4{,}5^2} = 17{,}46$$

$$\overline{X} = \frac{\sum_{t=1}^{n} X_t}{n} = \frac{45}{10} = 4{,}5$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 1261{,}7 - 17{,}46 \cdot 4{,}5 = 1183{,}1$$

$$\overline{Y} = \frac{\sum_{t=1}^{n} Y_t}{n} = \frac{12617}{10} = 1261{,}7$$

Koeficijent $\widehat{\beta}_1$ pokazuje da se broj zaposlenih u Hrvatskoj, u analiziranom razdoblju, povećavao prosječno godišnje za 17.460 zaposlenika.

Reprezentativnost ocijenjenog modela linearnog trenda izvodi se iz rezidualnih odstupanja. Varijanca trenda je:

$$\sigma^2_{\widehat{Y}} = \frac{\sum_{t=1}^{n} Y_t^2 - \widehat{\beta}_0 \sum_{t=1}^{n} Y_t - \widehat{\beta}_1 \sum_{t=1}^{n} X_t Y_t}{n} = \frac{15947557 - 1183{,}1 \cdot 12617 - 17{,}46 \cdot 58217}{10} = 353{,}6$$

Standardna devijacija trenda iznosi:

$$\sigma_{\widehat{Y}} = \sqrt{\sigma^2_{\widehat{Y}}} = \sqrt{353{,}6} = 18{,}8$$

Koeficijent varijacije trenda pokazuje da postotak standardne devijacije trenda od aritmetičke sredine varijable Y iznosi 1,49% (manji je od 10%) i ukazuje na dobru reprezentativnost ocijenjenog linearnog trend modela:

$$V_{\widehat{Y}} = \frac{\sigma_{\widehat{Y}}}{\overline{Y}} \cdot 100 = \frac{18{,}8}{1261{,}7} \cdot 100 = 1{,}49$$

Trend vrijednosti ($\widehat{Y}_t$), izračunane ocijenjenom jednadžbom linearnog trenda, pokazane su u posljednjem stupcu tablice 3.5.

3.7.1.2. Trend polinom drugog stupnja

Ako su 2. diferencije vrijednosti vremenskog niza približno konstantne ocjenjuje se (2+1) parametar trend polinoma:

$$\Delta^2 y_t \approx const., \tag{3.7.18}$$

čiji je općeniti oblik:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X + \hat{\beta}_2 \cdot X^2 \tag{3.7.19}$$

Ocjenom parametara metodom najmanjih kvadrata traži se minimum zbroja kvadrata empirijskih odstupanja u odnosu na trend vrijednosti:

Računa se sustav od 3 jednadžbe s 3 nepoznata parametra: $\hat{\beta}_1, \hat{\beta}_2 \ i \ \hat{\beta}_3$.

$$N \cdot \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i \tag{3.7.20}$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3 = \sum_{i=1}^{n} X_i Y_i \tag{3.7.21}$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4 = \sum_{i=1}^{n} X_i^2 Y_i \,. \tag{3.7.22}$$

Rješenje ovog sustava po parametrima $\hat{\beta}_j$ daje rješenje ocjena parametara trend polinoma 2. stupnja.

Jednadžbe analize varijance za model trend polinoma 2. stupnja, tj. za *k=2*:

$$SP = \sum_{t=1}^{n} (\hat{Y}_t - \overline{Y})^2 = \hat{\beta}_0 \cdot \sum_{t=1}^{n} Y_t + \hat{\beta}_1 \cdot \sum_{t=1}^{n} X_t \cdot Y_t + \hat{\beta}_2 \cdot \sum_{t=1}^{n} X_t^2 \cdot Y_t - n\overline{Y}^2 \tag{3.7.23}$$

$$SR = \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{n} Y_t^2 - \hat{\beta}_0 \cdot \sum_{t=1}^{n} Y_t - \hat{\beta}_1 \cdot \sum_{t=1}^{n} X_t \cdot Y_t - \hat{\beta}_2 \cdot \sum_{t=1}^{n} X_t^2 \cdot Y_t \tag{3.7.24}$$

$$ST = \sum_{t=1}^{n} (Y_t - \overline{Y})^2 = \sum_{t=1}^{n} Y_t^2 - n\overline{Y}^2 \tag{3.7.25}$$

Vrijedi da je:

$$SP + SR = ST\,, \tag{3.7.26}$$

što predstavlja jednadžbu analize varijance i predstavlja temelj analize reprezentativnosti trend modela.

Standardna pogrješka trenda je apsolutni pokazatelj reprezentativnosti trend modela, a pokazuje prosječni stupanj varijacije stvarnih vrijednosti ovisne varijable u odnosu na očekivane trend vrijednosti.

$$\hat{\sigma}_{\hat{Y}} = \sqrt{\frac{SR}{n-k-1}} \tag{3.7.27}$$

Ovaj izraz je standardna pogrješka trenda. Izražen je u originalnim jedinicama mjere ovisne varijable Y.

Taj problem eliminira **relativni pokazatelj - koeficijent varijacije trenda, koji predstavlja postotak standardne pogrješke trenda od aritmetičke sredine vremenskog niza Y.**

$$\hat{V}_{\hat{Y}} = \frac{\hat{\sigma}_{\hat{Y}}}{\overline{Y}} \cdot 100 \tag{3.7.28}$$

Najmanja vrijednost koeficijenta varijacije je 0%, a najveća nije definirana. **Što je koeficijent varijacije modela bliži nuli to je model reprezentativniji.** Često se uzima dogovorena granica reprezentativnosti od 10%. Dakle ako je koeficijent varijacije manji od 10% kaže se da je model dobar.

Koeficijent determinacije za model regresijskog polinoma 2. stupnja:

$$r^2 = \frac{SP}{ST} \tag{3.7.29}$$

Korigirani koeficijent determinacije:

$$\bar{r}^2 = 1 - \frac{n-1}{n-(k+1)} \cdot (1-r^2)\,, \tag{3.7.30}$$

je asimptotski nepristrana ocjena koeficijenta determinacije.

Primjer 3.7.2.

Tablica 3.6.

Prijevoz robe u željezničkom prometu na području «Z» u razdoblju od 1998. do 2008. god.

Godina	Prevezeno robe u tisućama *t* (y_t)	Varijabla *vrijeme* (x_t)	$x_t y_t$	x_t^2	x_t^4	$x_t^2 y_t$	$\hat{y}_t$	$(y_t - \hat{y}_t)^2$
1998.	336	-5	-1680	25	625	8400	289,6	2155,3
1999.	247	-4	-988	16	256	3952	251,2	17,6
2000.	185	-3	-555	9	81	1665	231,9	2199,5
2001.	181	-2	-362	4	16	724	231,7	2568,9
2002.	243	-1	-243	1	1	243	250,6	57,0
2003.	311	0	0	0	0	0	288,5	506,3
2004.	385	1	385	1	1	385	345,5	1557,8
2005.	450	2	900	4	16	1800	421,6	804,1
2006.	523	3	1569	9	81	4707	516,8	38,0
2007.	612	4	2448	16	256	9792	631,1	365,4
2008.	750	5	3750	25	625	18750	764,5	209,5
Ukupno	4223	0	5224	110	1958	50418	4223,0	10479,3

Izvor: Statistika područja «Z», 2009. god.

Zadatak je ocijeniti model trend polinoma drugog stupnja i koeficijent varijacije trenda.

Ocijenjeni model trend polinoma drugog stupnja s ishodištem u sredini vremenskog razdoblja glasi:

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2$: $\quad \hat{Y} = 288{,}5 + 47{,}49 \cdot X + 9{,}54 \cdot X^2$

X=0, 2003. godine
Jedinica za *X* je jedna godina
Jedinica za *Y* je 1000 *t*

$$\widehat{\beta}_0 = \frac{\sum_{t=1}^{n} Y_t \sum_{t=1}^{n} X_t^4 - \sum_{t=1}^{n} X_t^2 \sum_{t=1}^{n} Y_t X_t^2}{n\sum_{t=1}^{n} X_t^4 - \sum_{t=1}^{n} X_t^2 \sum_{t=1}^{n} X_t^2} = \frac{4223 \cdot 1958 - 110 \cdot 50418}{11 \cdot 1958 - 110 \cdot 110} = 288{,}5$$

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{t=1}^{n} X_t^2} = \frac{5224}{110} = 47{,}49$$

$$\widehat{\beta}_2 = \frac{n\sum_{t=1}^{n} Y_t X_t^2 - \sum_{t=1}^{n} X_t^2 \sum_{t=1}^{n} Y_t}{n\sum_{t=1}^{n} X_t^4 - \sum_{t=1}^{n} X_t^2 \sum_{t=1}^{n} X_t^2} = \frac{11 \cdot 50418 - 110 \cdot 4223}{11 \cdot 1958 - 110 \cdot 110} = 9{,}54$$

Reprezentativnost ocijenjenog modela trend polinoma drugog stupnja ispitat će se koeficijentom varijacije trenda:

$$\widehat{V}_{\widehat{Y}} = \frac{\widehat{\sigma}_{\widehat{Y}}}{\overline{Y}} \cdot 100 = \frac{36{,}2}{383{,}9} \cdot 100 = 9{,}4$$

$$\overline{Y} = \frac{\sum_{t=1}^{n} Y_t}{n} = \frac{4223}{11} = 383{,}9$$

$$\widehat{\sigma}_{\widehat{Y}} = \sqrt{\frac{SR}{n-k-1}} = \sqrt{\frac{10479{,}3}{11-2-1}} = 36{,}2$$

$k = 2$ (stupanj polinoma)

$$SR = \sum_{t=1}^{n} (Y_t - \widehat{Y}_t)^2 = 10479{,}3$$

Koeficijent varijacije trend modela pokazuje da postotak standardne pogrješke trenda od aritmetičke sredine varijable Y iznosi 9,4% i potvrđuje reprezentativnost ocijenjenog modela trend polinoma drugog stupnja.

3.7.2. Eksponencijalni trend modeli

Eksponencijalni trend modeli su oblika:

$$\hat{Y} = \hat{\beta}_0 \cdot \hat{\beta}_1^{x} \cdot \hat{\beta}_2^{x^2} \cdot ... \cdot \hat{\beta}_k^{x^k} \tag{3.7.31}$$

ili

$$\hat{Y} = e^{\hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 + ... + \hat{\beta}_k X^k} \tag{3.7.32}$$

Ocjena parametara se najčešće dobije metodom najmanjih kvadrata.

- $k = 1 \Rightarrow \quad \hat{\beta}_1 \Rightarrow \quad \overline{S} = (\hat{\beta}_1 - 1) \cdot 100$ - prosječna stopa promjene vrijednosti vremenskog niza u jedinici vremena u %.

  $\hat{\beta}_0 \Rightarrow$ trend vrijednost u ishodištu.

- $k \quad \Rightarrow$ ako su k-te diferencije logaritama vrijednosti vremenskog niza približno konstantne ocjenjuje se (k+1) parametar eksponencijalnog trenda:

$$\Delta^k (\log y_t) \approx const. \tag{3.7.33}$$

3.7.2.1. Jednostavni eksponencijalni trend

Model jednostavne eksponencijalne regresije općenito je oblika:

$$\hat{Y} = \hat{\beta}_0 \cdot \hat{\beta}_1^{X} \tag{3.7.34}$$

Da bi se za ocjenu parametara upotrijebila **metoda najmanjih kvadrata,** potrebno je početni model logaritamskom transformacijom prevesti u logaritamsko-linearni oblik:

$$\log \hat{Y} = \log \hat{\beta}_0 + \log \hat{\beta}_1 \cdot X \tag{3.7.35}$$

Traži se minimum sume kvadrata neprotumačenih ili rezidualnih odstupanja:

$$\min SR = \min \sum_{t=1}^{n} (\log Y_t - \log \hat{Y}_t)^2 = \min \sum_{t=1}^{n} \left[\log Y_t - (\log \hat{\beta}_0 + \log \hat{\beta}_1 X_t)\right]^2 \tag{3.7.36}$$

Nakon primjene matematičkog postupka traženja minimuma dobije se da je:

$$\log \hat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t \log Y_t - n\overline{X}\,\overline{\log Y}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2} \quad i \quad \log \hat{\beta}_0 = \overline{\log Y} - \log \hat{\beta}_1 \overline{X}\ , \tag{3.7.37}$$

gdje su:

$$\overline{X} = \frac{\sum_{t=1}^{n} X_t}{n} \quad i \quad \overline{\log Y} = \frac{\sum_{t=1}^{n} \log Y_t}{n}\ , \tag{3.7.38}$$

jednostavne aritmetičke sredine varijabli X i $\log Y$.

Transformacijom:

$$\hat{\beta}_0 = 10^{\log \hat{\beta}_0} \quad i \quad \hat{\beta}_1 = 10^{\log \hat{\beta}_1}\ , \tag{3.7.39}$$

dobiju se originalne vrijednosti parametara modela.

Parametar $\hat{\beta}_0$ je konstantni član, tj. trend vrijednost u ishodištu: ($\hat{\beta}_0 = \hat{Y}$ kada je X=0).

Parametar $\hat{\beta}_1$ pokazuje ocjenu tempa ili dinamike kretanja zavisne varijable Y (u %) kada nezavisna varijabla poraste za jedinicu vremena. Ovaj parametar interpretira se kao ocjena prosječne stope promjene varijable Y, u jedinici vremena (X):

$$\overline{S} = (\hat{\beta}_1 - 1) \cdot 100 \Rightarrow (u\ \%) \tag{3.7.40}$$

Jednadžbe analize varijance kod modela jednostavnog eksponencijalnog trenda transformiraju se na sličan način logaritmiranjem:

$$SP = \sum_{t=1}^{n} \left(\log \hat{Y}_t - \overline{\log Y}\right)^2 = \log \hat{\beta}_0 \cdot \sum_{t=1}^{n} \log Y_t + \log \hat{\beta}_1 \sum_{t=1}^{n} X_t \log Y_t - n \cdot \overline{\log Y}^2 \tag{3.7.41}$$

$$SR = \sum_{t=1}^{n} \left(\log Y_t - \log \hat{Y}_t\right)^2 = \sum_{t=1}^{n} \log^2 Y_t - \log \hat{\beta}_0 \cdot \sum_{t=1}^{n} \log Y_t - \log \hat{\beta}_1 \sum_{t=1}^{n} X_t \log Y_t \tag{3.7.42}$$

Ova odstupanja u stvari su slučajne pogrješke.

$$ST = \sum_{t=1}^{n} \left(\log Y_t - \overline{\log Y}\right)^2 = \sum_{t=1}^{n} \log^2 Y_t - n \cdot \overline{\log Y}^2 \tag{3.7.43}$$

Vrijedi da je:

$$SP + SR = ST\ , \tag{3.7.44}$$

Standardna pogrješka trenda i ovdje se računa iz ocjene zbroja kvadrata rezidualnih odstupanja, samo što se elementi analize varijance računaju iz transformiranog logaritamski - linearnog trend modela:

$$\hat{\sigma}_{\log \hat{Y}} = \sqrt{\frac{SR}{n-2}} \tag{3.7.45}$$

Relativni pokazatelj, koeficijent varijacije trenda je:

$$\hat{V}_{\log \hat{Y}} = \frac{\hat{\sigma}_{\log \hat{Y}}}{\overline{\log Y}} \cdot 100 \tag{3.7.46}$$

Ako je koeficijent varijacije manji od 10% kaže se da je model dobar.

Koeficijent determinacije je pokazatelj reprezentativnosti modela koji se temelji na analizi varijance. On se definira kao **omjer sume kvadrata odstupanja protumačenih trendom i sume kvadrata ukupnih odstupanja.**

$$r^2 = \frac{SP}{ST} \tag{3.7.47}$$

Koeficijent determinacije pokazuje koliko % je sume kvadrata odstupanja logaritama vrijednosti niza Y od aritmetičke sredine njenih logaritamskih vrijednosti protumačeno trend modelom.

Vrijedi da je:

$$r^2 = 1 - \frac{SR}{ST} \tag{3.7.48}$$

Vrijednost koeficijenta determinacije kreće se u intervalu $0 \le r^2 \le 1$**. Trend model je reprezentativniji ako je ovaj pokazatelj bliži 1.** Teorijska granica reprezentativnosti modela je 0,9. U praksi je ponekad vrlo teško pronaći varijablu koja dobro objašnjava ovisnu pojavu pa se ta granica reprezentativnosti spušta i do 0,6.

Korigirani koeficijent determinacije:

$$\bar{r}^2 = 1 - \frac{n-1}{n-(k+1)} \cdot (1 - r^2),$$

je asimptotski nepristrana ocjena koeficijenta determinacije.

Primjer 3.7.3.

Tablica 3.7.

Krediti odobreni stanovništvu u banci «Z» u razdoblju od 1999. do 2008. god.

Godina	Krediti u mil. kn (Y_t)	Varijabla *vrijeme* (X_t)	$\log Y_t$	X_t^2	$X_t \log Y_t$	$\log \widehat{Y}_t$
1999.	10,5	0	1,02119	0	0,00000	0,94280
2000.	9,8	1	0,99123	1	0,99123	1,08466
2001.	18,3	2	1,26245	4	2,52490	1,22652
2002.	24,8	3	1,39445	9	4,18336	1,36838
2003.	28,4	4	1,45332	16	5,81327	1,51024
2004.	41,6	5	1,61909	25	8,09547	1,65210
2005.	62,5	6	1,79588	36	10,77528	1,79396
2006.	88,8	7	1,94841	49	13,63889	1,93582
2007.	125,1	8	2,09726	64	16,77806	2,07768
2008.	169,2	9	2,22840	81	20,05560	2,21954
Ukupno	579,0	45	15,81168	285	82,85606	15,81168

Izvor: Podaci banke «Z», 2009. god.

Zadatak je ocijeniti model jednostavnog eksponencijalnog trenda i koeficijent varijacije trenda.

Model jednostavnog eksponencijalnog trenda glasi:

$$\widehat{Y} = \widehat{\beta}_0 \cdot \widehat{\beta}_1^{\,X}$$

Za ocjenu parametara «metodom najmanjih kvadrata» model je potrebno logaritamski linearizirati:

$$\log \widehat{Y} = \log \widehat{\beta}_0 + \widehat{\beta}_1 \cdot X$$

Ocjena parametara je sljedeća:

$$\log \widehat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t \log Y_t - n\overline{X}\,\overline{\log Y}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2} = \frac{82{,}85606 - 10 \cdot 4{,}5 \cdot 1{,}58117}{285 - 10 \cdot 4{,}5^2} = 0{,}14186$$

$$\log \widehat{\beta}_0 = \overline{\log Y} - \log \widehat{\beta}_1 \overline{X} = 1{,}58117 - 0{,}14186 \cdot 4{,}5 = 0{,}94280$$

$$\overline{X} = \frac{\sum_{t=1}^{n} X_t}{n} = \frac{45}{10} = 4{,}5 \qquad \overline{\log Y} = \frac{\sum_{t=1}^{n} \log Y_t}{n} = \frac{15{,}81168}{10} = 1{,}58117$$

Ocijenjeni logaritamski linearizirani model je:

$$\log \widehat{Y} = 0{,}94280 + 0{,}14186 \cdot X$$

Transformacijom:

$\widehat{\beta}_0 = 10^{\log\widehat{\beta}_0} = 10^{\log 0,94280} = 8,766$ $\widehat{\beta}_1 = 10^{\log\widehat{\beta}_1} = 10^{\log 0,14186} = 1,386$

dobiju se vrijednosti parametara modela, stoga ocijenjeni model jednostavnog eksponencijalnog trenda glasi:

$\widehat{Y} = 8,766 \cdot 1,386^X$

X=0, 1999. godine
Jedinica za *X* je jedna godina
Jedinica za *Y* je 1 milijun kn

Parametar $\widehat{\beta}_1$ pokazuje ocjenu prosječne stope promjene varijable *Y* u jedinici vremena X:

$$\overline{S} = (\widehat{\beta}_1 - 1) \cdot 100 = (1,386 - 1) \cdot 100 = 38,6\%$$

znači da su se krediti odobreni stanovništvu u analiziranom razdoblju u prosjeku povećavali po stopi od 38,6% godišnje.

Reprezentativnost ocijenjenog modela ispitat će se koeficijentom varijacije trenda:

$$\widehat{V}_{\log\widehat{Y}} = \frac{\widehat{\sigma}_{\log\widehat{Y}}}{\overline{\log Y}} \cdot 100 = \frac{0,05220}{1,58117} \cdot 100 = 3,3$$

$$\widehat{\sigma}_{\log\widehat{Y}} = \sqrt{\frac{SR}{n-2}} = \sqrt{\frac{0,02180}{10-2}} = 0,05220$$

$$SR = \sum_{t=1}^{n} (\log Y_t - \log\widehat{Y}_t)^2 = 0,02180$$

Potreban je sljedeći izračun:

Godina	$\log Y_t$	$\log\widehat{Y}_t$	$(\log Y_t - \log\widehat{Y}_t)^2$
1999.	1,02119	0,94280	0,00615
2000.	0,99123	1,08466	0,00873
2001.	1,26245	1,22652	0,00129
2002.	1,39445	1,36838	0,00068
2003.	1,45332	1,51024	0,00324
2004.	1,61909	1,65210	0,00109
2005.	1,79588	1,79396	0,00000
2006.	1,94841	1,93582	0,00016
2007.	2,09726	2,07768	0,00038
2008.	2,22840	2,21954	0,00008
Ukupno	15,81168	15,81168	0,02180

Koeficijent varijacije trend modela iznosi 3,3% i potvrđuje dobru reprezentativnost ocijenjenog jednostavnog eksponencijalnog modela trenda.

3.7.3. Hiperbolički trend modeli

Hiperbolički trend modeli su oblika:

$$\hat{Y} = \frac{1}{\hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 + ... + \hat{\beta}_k X^k} \tag{3.7.49}$$

Ocjena parametara najčešće se dobije metodom najmanjih kvadrata (BLUE).

- k $\Rightarrow$ ako su k-te diferencije recipročnih vrijednosti vremenskog niza približno konstantne ocjenjuje se (k+1) parametar hiperboličkog trenda:

$$\Delta^k \left(\frac{1}{Y_t}\right) \approx const. \tag{3.7.50}$$

3.7.3.1. Jednostavni hiperbolički trend

Jednostavni hiperbolički trend model je općenito oblika:

$$\hat{Y} = \frac{1}{\hat{\beta}_0 + \hat{\beta}_1 \cdot X} \tag{3.7.51}$$

Da bi se za ocjenu parametara upotrijebila **metoda najmanjih kvadrata,** potrebno je početni model invertnom transformacijom prevesti u linearni oblik:

$$1/\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X \tag{3.7.52}$$

Nakon primjene matematičkog postupka traženja minimuma sume kvadrata neprotumačenih ili rezidualnih odstupanja dobije se da je:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t (1/Y_t) - n\overline{X}\overline{(1/Y_t)}}{\sum_{t=1}^{n} X_t^{\,2} - n\overline{X}^2} \quad i \quad \hat{\beta}_0 = \overline{(1/Y_t)} - \hat{\beta}_1 \overline{X}, \tag{3.7.53}$$

gdje su:

$$\overline{X} = \frac{\sum_{t=1}^{n} X_t}{n} \quad i \quad \overline{(1/Y)} = \frac{\sum_{t=1}^{n} 1/Y_t}{n}. \tag{3.7.54}$$

Parametar $\hat{\beta}_0$, u skladu s originalnom formom modela, može se komentirati: **u ishodišnom razdoblju (kada je X=0) vrijednost pojave vremenskog niza Y odgovarat će recipročnoj vrijednosti parametra $\hat{\beta}_0$, tj. iznositi će $1/\hat{\beta}_0$ jedinica.**

Parametar $\hat{\beta}_1$ ovdje nema logičnu interpretaciju.

Jednadžbe analize varijance kod ovog modela isto se formiraju odgovarajućom transformacijom.

Primjer 3.7.4.

Tablica 3.8.

Proizvodnja proizvoda «Z» u razdoblju od 1994. do 2008. god.

Godina	Proizvodnja u tisućama *t* (Y_t)	Varijabla *vrijeme* (X_t)	$1/Y_t$	$X_t(1/Y)_t$	X_t^2	$1/\widehat{Y}_t$	$\widehat{Y}_t$
1994.	8,5	0	0,11765	0,00000	0	0,11682	8,6
1995.	8,4	1	0,11905	0,11905	1	0,11367	8,8
1996.	9,2	2	0,10870	0,21739	4	0,11053	9,0
1997.	9,5	3	0,10526	0,31579	9	0,10739	9,3
1998.	10,0	4	0,10000	0,40000	16	0,10424	9,6
1999.	10,3	5	0,09709	0,48544	25	0,10110	9,9
2000.	10,2	6	0,09804	0,58824	36	0,09796	10,2
2001.	10,4	7	0,09615	0,67308	49	0,09481	10,5
2002.	10,5	8	0,09524	0,76190	64	0,09167	10,9
2003.	11,3	9	0,08850	0,79646	81	0,08853	11,3
2004.	11,2	10	0,08929	0,89286	100	0,08538	11,7
2005.	12,5	11	0,08000	0,88000	121	0,08224	12,2
2006.	12,4	12	0,08065	0,96774	144	0,07910	12,6
2007.	13,3	13	0,07519	0,97744	169	0,07595	13,2
2008.	14,0	14	0,07143	1,00000	196	0,07281	13,7
Ukupno	161,7	105	1,42222	9,07539	1015	1,42221	161,6

Izvor: Podaci su simulirani.

Zadatak je ocijeniti model jednostavnog hiperboličkog trenda i trend vrijednosti.

Model jednostavnog hiperboličkog trenda glasi:

$$\widehat{Y} = \frac{1}{\widehat{\beta}_0 + \widehat{\beta}_1 X}$$

Za ocjenu parametara «metodom najmanjih kvadrata» model je potrebno invertnom transformacijom linearizirati:

$$1/\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X$$

Ocjena parametara je sljedeća:

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t(1/Y_t) - n\overline{X}\,\overline{(1/Y_t)}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2} = \frac{9{,}07539 - 15\cdot 7\cdot 0{,}09481}{1015 - 15\cdot 7^2} = -0{,}003143$$

$$\widehat{\beta}_0 = \overline{(1/Y_t)} - \widehat{\beta}_1\overline{X} = 0{,}09481 - (-0{,}003143)\cdot 7 = 0{,}116817$$

$$\overline{X} = \frac{\sum_{t=1}^{n} X_t}{n} = \frac{105}{15} = 7 \qquad \overline{(1/Y_t)} = \frac{\sum_{t=1}^{n} 1/Y_t}{n} = \frac{1{,}42222}{15} = 0.09481$$

Ocijenjeni jednostavni hiperbolički model glasi:

$$\widehat{Y} = \frac{1}{0{,}116817 - 0{,}003143X}$$

X=0, 1994. godine
Jedinica za *X* je jedna godina
Jedinica za *Y* je 1 tisuća *t*

Trend vrijednosti prikazane su u posljednjem stupcu tablice 3.8.

3.7.4. Asimptotski trend modeli

Asimptotski trend modeli spadaju u skupinu **pravih nelinearnih trend modela.** Koriste se za opisivanje razvojne tendencije pojava koje se na dugi rok približavaju nekoj graničnoj vrijednosti - **asimptoti**.

Ocjene parametara ne mogu se dobiti metodom najmanjih kvadrata, jer ove modele nije moguće transformirati da budu linearni u parametrima. Spadaju u funkcionalne modele (nisu stohastički ili statistički modeli).

Tri najčešće korištena **asimptotska trend modela:**

1. **Modificirani eksponencijalni trend**
2. **Logistički trend**
3. **Gompertzov trend**

Još se zovu **globalni modeli** ocjene prisutnosti trend komponente u odgovarajućem vremenskom nizu. (Zasnivaju se na primjeni **svih frekvencija vremenskog niza** koje imaju isto značenje ili ponder pri prognozi, bez obzira da li se nalaze na početku vremenskog niza ili su bliže tekućem razdoblju).

Parametri ovih modela **ocjenjuju se** specifičnim metodama ocjena:

a) iterativne metode
b) aproksimativna ocjena parametara

Iterativne metode rade se putem računala pomoću suvremenih statističkih paketa.

Aproksimativne metode ocjene parametara su:

1. metoda parcijalnih suma

2. metoda triju odabranih točaka.

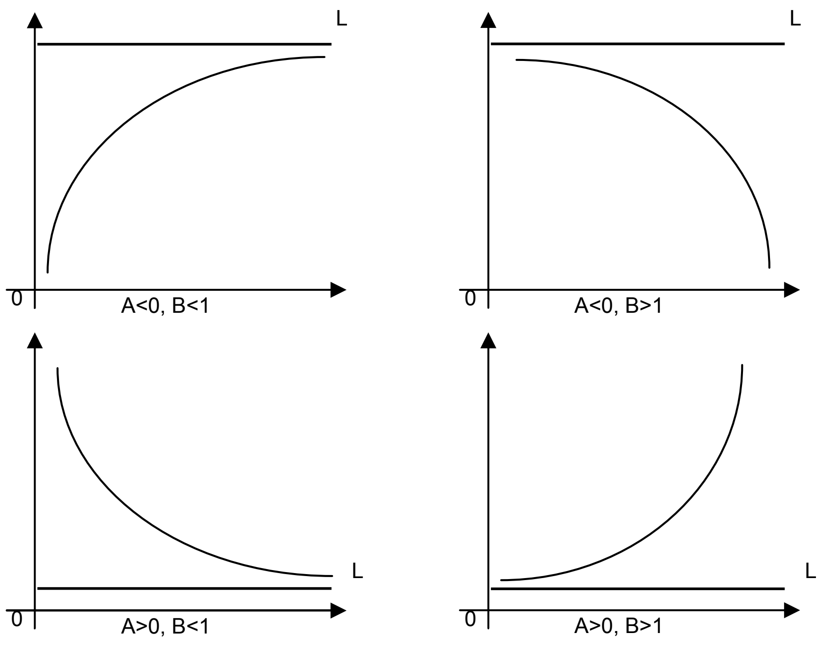
3.7.4.1. Modificirani eksponencijalni trend

Modificirani eksponencijalni trend je oblika:

$$\hat{Y} = L + A \cdot B^x . \tag{3.7.55}$$

Mogući oblici modela modificiranog eksponencijalnog trenda u ovisnosti o parametrima A, B i L prikazani su na slici 3.1.:

Slika 3.1.



Modificirani eksponencijalni trend model upotrebljava se za prikaz trend komponente vremenskog niza kojemu su omjeri prvih diferencija vremenskog niza približno konstantne:

$$\frac{\Delta Y_t}{\Delta Y_{t-1}} \approx const. \tag{3.7.56}$$

Ocjena parametara modificiranog eksponencijalnog trenda:

1. metodom parcijalnih suma

$$B = \sqrt[r]{\frac{S_3 - S_2}{S_2 - S_1}}, \qquad r = \frac{N}{3} \ , \tag{3.7.57}$$

$$A = (S_2 - S_1) \cdot \frac{B - 1}{(B^r - 1)^2} \ , \tag{3.7.58}$$

$$L = \frac{1}{r}\left(\frac{S_1 \cdot S_3 - S_2^2}{S_1 + S_3 - 2S_2} \right), \tag{3.7.59}$$

gdje su:

S_1 - parcijalna suma prve trećine **vrijednost**i vremenskog niza

S_2 - parcijalna suma druge trećine **vrijednosti** vremenskog niza

S_3 - parcijalna suma treće trećine **vrijednosti** vremenskog niza.

2. metodom triju odabranih točaka

$$B = \sqrt[r]{\frac{Y_{III} - Y_{II}}{Y_{II} - Y_I}}, \qquad r = \frac{N - 1}{2}, \tag{3.7.60}$$

$$A = \frac{Y_{II} - Y_I}{B^r - 1} \ , \tag{3.7.61}$$

$$L = Y_I - A \ , \tag{3.7.62}$$

gdje su:

Y_I - **vrijednost** pojave na početku vremenskog niza

Y_{II} - **vrijednost** pojave u sredini vremenskog niza

Y_{III} - **vrijednost** pojave na kraju vremenskog niza.

Primjer 3.7.5.

Tablica 3.9.

Izvoz područja «Z» u razdoblju od 2000. do 2008. god.

Godina	Izvoz u milijunima USD (Y_t)	Varijabla *vrijeme* (X_t)	$\widehat{Y}_t$
2000.	100	0	98
2001.	102	1	103
2002.	109	2	110
2003.	119	3	119
2004.	130	4	130
2005.	145	5	145
2006.	168	6	165
2007.	191	7	191
2008.	222	8	225
Ukupno	1286	36	1286

Izvor: Statistika područja «Z», 2009. god.

Zadatak je ocijeniti model modificiranog eksponencijalnog trenda i trend vrijednosti.

Model modificiranog eksponencijalnog trenda glasi:

$$\widehat{Y} = L + A \cdot B^X$$

Ocjena parametara modificiranog eksponencijalnog trenda metodom parcijalnih suma je sljedeća:

$$B = \sqrt[r]{\frac{S_3 - S_2}{S_2 - S_1}} = \sqrt[3]{\frac{581 - 394}{394 - 311}} = 1{,}31 \qquad r = \frac{n}{3} = \frac{9}{3} = 3$$

$$A = (S_2 - S_1) \cdot \frac{B-1}{(B^r - 1)^2} = (394 - 311)\frac{1{,}31 - 1}{(1{,}31^3 - 1)^2} = 16{,}44$$

$$L = \frac{1}{r}\left(\frac{S_1 \cdot S_3 - S_2^2}{S_1 + S_3 - 2S_2}\right) = \frac{1}{3}\left(\frac{311 \cdot 581 - 394^2}{311 + 581 - 2 \cdot 394}\right) = 81{,}59$$

$$S_1 = 100 + 102 + 109 = 311$$

$$S_2 = 119 + 130 + 145 = 394$$

$$S_3 = 168 + 191 + 222 = 581$$

Ocijenjeni model modificiranog eksponencijalnog trenda glasi:

$\widehat{Y} = 81{,}59 + 16{,}44 \cdot 1{,}31^X$

X=0, 2000. godine
Jedinica za *X* je jedna godina
Jedinica za *Y* je 1 milijun USD-a

Trend vrijednosti prikazane su u posljednjem stupcu tablice 3.9.

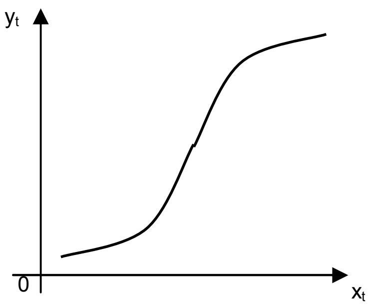
3.7.4.2. Logistički trend

Logistički trend je oblika:

$$\hat{Y} = \frac{1}{L + A \cdot B^x} \tag{3.7.63}$$

Oblik modela logističkog trenda prikazan je na slici 3.2.:

Slika 3.2.



U ekonomiji se krivulja logističkog trenda upotrebljava kod opisivanja životnog vijeka nekog proizvoda (faza uhodavanja, faza ekspanzije i faza stagnantnog razvoja).

Logistički trend model upotrebljava se za prikaz trend komponente vremenskog niza kojemu su omjeri prvih diferencija recipročnih vrijednosti vremenskog niza približno konstantne:

$$\frac{\Delta(1/Y_t)}{\Delta(1/Y_{t-1})} \approx const. \tag{3.7.64}$$

Ocjena parametara logističkog trenda:

1. metodom parcijalnih suma

$$B=\sqrt[r]{\frac{S_3-S_2}{S_2-S_1}},\qquad r=\frac{N}{3} \tag{3.7.65}$$

$$A=(S_2-S_1)\cdot\frac{B-1}{(B^r-1)^2} \tag{3.7.66}$$

$$L=\frac{1}{r}\left(\frac{S_1\cdot S_3-S_2^2}{S_1+S_3-2S_2}\right) \tag{3.7.67}$$

gdje su:

S_1 - parcijalna suma prve trećine **recipročnih vrijednosti** vremenskog niza

S_2 - parcijalna suma druge trećine **recipročnih vrijednosti** vremenskog niza S_3 - parcijalna suma treće trećine **recipročnih vrijednosti** vremenskog niza.

2. metodom triju odabranih točaka

$$B=\sqrt[r]{\frac{(1/Y_{III})-(1/Y_{II})}{(1/Y_{II})-(1/Y_I)}},\qquad r=\frac{N-1}{2} \tag{3.7.68}$$

$$A=\frac{(1/Y_{II})-(1/Y_I)}{B^r-1} \tag{3.7.69}$$

$$L=(1/Y_I)-A \tag{3.7.70}$$

gdje su:

$(1/Y_I)$ - **recipročna vrijednost** pojave na početku vremenskog niza

$(1/Y_{II})$ - **recipročna vrijednost** pojave u sredini vremenskog niza

$(1/Y_{III})$ - **recipročna vrijednost** pojave na kraju vremenskog niza.

Primjer 3.7.6.

Tablica 3.10.

Prodaja proizvoda «Z» u razdoblju od 2000. do 2008. god.

Godina	Prodano tisuća *t* (Y_t)	Varijabla *vrijeme* (X_t)	$1/Y_t$	$\widehat{Y}_t$
2000.	56	0	0,017857	45,5
2001.	65	1	0,015384	74,1
2002.	88	2	0,011364	109,5
2003.	133	3	0,007519	145,1
2004.	186	4	0,005376	174,0
2005.	204	5	0,004902	193,9
2006.	210	6	0,004762	206,0
2007.	211	7	0,004739	212,8
2008.	214	8	0,004673	216,5
Ukupno	1367	36	0,076576	1377,4

Izvor: Podaci su simulirani.

Zadatak je ocijeniti model logističkog trenda i trend vrijednosti.

Model logističkog trenda glasi:

$$\widehat{Y} = \frac{1}{L + AB^X}$$

Ocjena parametara logističkog trenda metodom parcijalnih suma je sljedeća:

$$B = \sqrt[r]{\frac{S_3 - S_2}{S_2 - S_1}} = \sqrt[3]{\frac{0{,}014174 - 0{,}017797}{0{,}017797 - 0{,}044605}} = 0{,}513174 \qquad r = \frac{n}{3} = \frac{9}{3} = 3$$

$$A = (S_2 - S_1) \cdot \frac{B-1}{(B^r - 1)^2} = (0{,}017797 - 0{,}044605)\frac{0{,}513174 - 1}{(0{,}513174^3 - 1)^2} = 0{,}017448$$

$$L = \frac{1}{r}\left(\frac{S_1 \cdot S_3 - S_2^2}{S_1 + S_3 - 2S_2}\right) = \frac{1}{3}\left(\frac{0{,}044605 \cdot 0{,}014174 - 0{,}017797^2}{0{,}044605 + 0{,}014174 - 2 \cdot 0{,}017797}\right) = 0{,}004536$$

$$S_1 = 0{,}017857 + 0{,}015384 + 0{,}011364 = 0{,}044605$$

$$S_2 = 0{,}007519 + 0{,}005376 + 0{,}004902 = 0{,}017797$$

$$S_3 = 0{,}004762 + 0{,}004739 + 0{,}004673 = 0{,}014174$$

Ocijenjeni model logističkog trenda glasi:

$$\widehat{Y} = \frac{1}{0{,}004536 + 0{,}017448 \cdot 0{,}513174^{X}}$$

$X=0$, 2000. godine
Jedinica za X je jedna godina
Jedinica za Y je 1 tisuća t

Trend vrijednosti prikazane su u posljednjem stupcu tablice 3.10.

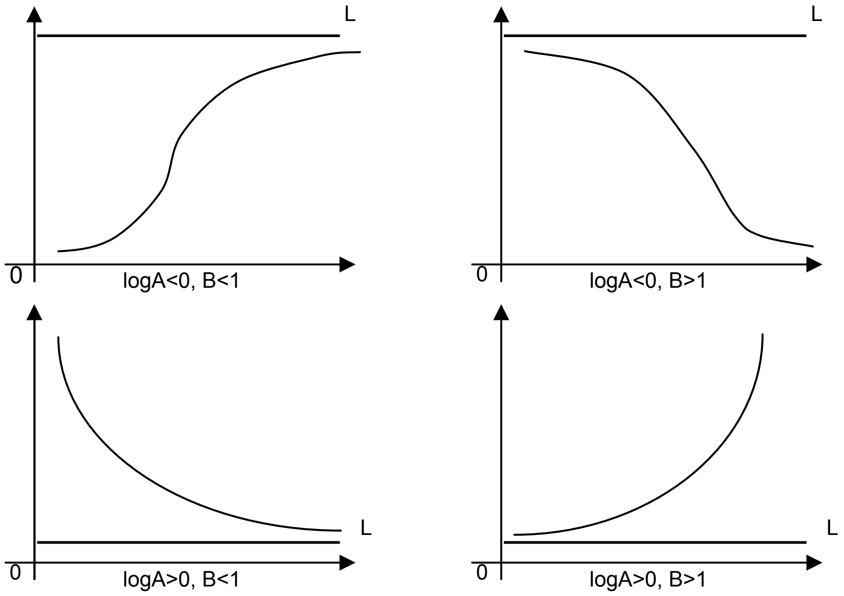
3.7.4.3. Gompertzov trend

Gompertzov trend je oblika:

$$\hat{Y} = L \cdot A^{B^{X}} \tag{3.7.71}$$

Mogući oblici modela Gompertzovog trenda u ovisnosti o parametrima A, B i L prikazani su na slici 3.3.:

Slika 3.3.



U ekonomiji se krivulja Gompertzovog trenda upotrebljava kod opisivanja ekonomskih pojava (faza uhodavanja, faza ekspanzije, faza regresivnog rasta i faza stagnantnog razvoja)

Gompertzov trend model upotrebljava se za prikaz trend komponente vremenskog niza kojemu su omjeri prvih diferencija logaritama vrijednosti vremenskog niza približno konstantne:

$$\frac{\Delta \log Y_t}{\Delta \log Y_{t-1}} \approx const. \tag{3.7.72}$$

Ocjena parametara Gompertzovog trenda:

1. metodom parcijalnih suma

$$B = \sqrt[r]{\frac{S_3 - S_2}{S_2 - S_1}}, \qquad r = \frac{N}{3} \ , \tag{3.7.73}$$

$$\log A = (S_2 - S_1) \cdot \frac{B-1}{(B^r - 1)^2} \ , \tag{3.7.74}$$

$$\log L = \frac{1}{r}\left(\frac{S_1 \cdot S_3 - S_2^2}{S_1 + S_3 - 2S_2}\right), \tag{3.7.75}$$

gdje su:

S_1 - parcijalna suma prve trećine **logaritamskih vrijednosti** vremenskog niza

S_2 - parcijalna suma druge trećine **logaritamskih vrijednosti** vremenskog niza

S_3 - parcijalna suma treće trećine **logaritamskih vrijednosti** vremenskog niza.

2. metodom triju odabranih točaka

$$B = \sqrt[r]{\frac{\log Y_{III} - \log Y_{II}}{\log Y_{II} - \log Y_I}}, \qquad r = \frac{N-1}{2}, \tag{3.7.76}$$

$$\log A = \frac{\log Y_{II} - \log Y_I}{B^r - 1}, \tag{3.7.77}$$

$$\log L = \log Y_I - A \ , \tag{3.7.78}$$

gdje su:

$\log Y_I$ - **logaritam vrijednosti** pojave na početku vremenskog niza

$\log Y_{II}$ - **logaritam vrijednosti** pojave u sredini vremenskog niza

$\log Y_{III}$ - **logaritam vrijednosti** pojave na kraju vremenskog niza.

Primjer 3.7.7.

Tablica 3.11.

Pretplatnici Interneta na području «Z» u razdoblju od 2000. do 2008. god.

Godina	Broj pretplatnika Interneta (Y_t)	Varijabla *vrijeme* (X_t)	$\log Y_t$
2000.	4	0	0,852060
2001.	7	1	0,987955
2002.	8	2	1,028090
2003.	12	3	1,162515
2004.	20	4	1,351030
2005.	32	5	1,536400
2006.	68	6	1,847215
2007.	151	7	2,185599
2008.	393	8	2,596937
Ukupno	695	36	13,547801

Izvor: Statistika područja «Z», 2009. god.

Zadatak je ocijeniti Gompertzov trend.

Gompertzov trend ima oblik:

$$\widehat{Y} = L \cdot A^{B^X}$$

Ocjena parametara Gompertzovog trenda metodom parcijalnih suma je sljedeća:

$$B = \sqrt[r]{\frac{S_3 - S_2}{S_2 - S_1}} = \sqrt[3]{\frac{6{,}629751 - 4{,}040045}{4{,}049945 - 2{,}868105}} = 1{,}297208 \qquad r = \frac{n}{3} = \frac{9}{3} = 3$$

$$\log A = (S_2 - S_1) \cdot \frac{B - 1}{(B^r - 1)^2} = (4{,}049945 - 2{,}868105)\frac{1{,}297208 - 1}{(1{,}297208^3 - 1)^2} = 0{,}251039$$

$$\log L = \frac{1}{r}\left(\frac{S_1 \cdot S_3 - S_2^2}{S_1 + S_3 - 2S_2}\right) = \frac{1}{3}\left(\frac{2{,}868105 \cdot 6{,}629751 - 4{,}049945^2}{2{,}868105 + 6{,}629751 - 2 \cdot 4{,}049945}\right) = 0{,}622993$$

$$S_1 = 0{,}852060 + 0{,}987955 + 1{,}028090 = 2{,}868105$$

$$S_2 = 1{,}162515 + 1{,}351030 + 1{,}536400 = 4{,}049945$$

$$S_3 = 1{,}847215 + 2{,}185599 + 2{,}596937 = 6{,}629751$$

Transformacijom:

$$A = 10^{\log A} = 10^{\log 0{,}251039} = 1{,}782540 \qquad L = 10^{\log L} = 10^{\log 0{,}622993} = 4{,}197526$$

dobiju se vrijednosti parametara modela, stoga ocijenjeni model Gompertzovog trenda glasi:

$$\widehat{Y} = 4{,}197526 \cdot 1{,}782540^{1{,}297208^{X}}$$

X=0, 2000. godine
Jedinica za *X* je jedna godina
Jedinica za *Y* je 1 pretplatnik Internata

3.8. Procjene parametara

Intervalna procjena parametra β_j trend modela ocijenjenog metodom najmanjih kvadrata uz odgovarajući nivo pouzdanosti procjene $(1-\alpha)$ je:

$$\Pr\left\{\hat{\beta}_j - t \cdot Se(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + t \cdot Se(\hat{\beta}_j)\right\} = 1-\alpha \tag{3.8.1}$$

gdje je:

t - odgovarajuća vrijednost iz tablica Studentove distribucije

$1-\alpha$ - odgovarajući nivo pouzdanosti procjene (najčešće se uzima da je 95%),

$Se(\hat{\beta}_j)$ - standardna pogrješka ocijenjenog parametra koja se izračunava:

$$Se(\hat{\beta}_j) = \hat{\sigma}_{\hat{Y}} \cdot \sqrt{s_{jj}}\ , \tag{3.8.2}$$

gdje s_{jj} predstavlja vrijednost odgovarajućeg dijagonalnog elementa u matrici $(X^T X)^{-1}$.

3.9. Pomični prosjeci

"Lokalni" trend modeli ili metode izglađivanja vremenskih nizova između ostalih su pomični prosjeci. Spadaju u **neparametrijske metode za utvrđivanje trend komponente.**

Pomični prosjeci su aritmetičke sredine M uzastopnih frekvencija vremenskog niza (pri čemu je M<N). Mogu biti jednostavni i vagani pomični prosjeci u ovisnosti o tome da li se pri njihovom izračunavanju rabi jednostavna ili vagana aritmetička sredina M uzastopnih frekvencija vremenskog niza.

Neka su:

$$Y_1, Y_2, ..., Y_t, ..., Y_n \tag{3.9.1}$$

frekvencije vremenskog niza.

Jednostavni pomični prosjeci za M=2m+1 (**neparan broj članova pomičnog prosjeka**):

$$\hat{Y}_t^* = \frac{1}{M} \sum_{s=-m}^{m} Y_{t+s}, \quad t = m+1, m+2, ..., (n-m), \tag{3.9.2}$$

npr.: M=3: $\hat{Y}_t^* = \frac{1}{3} \sum_{s=-1}^{1} Y_{t+s}$.

U ovom slučaju kod neparnog broja članova pomičnog prosjeka, vrijednost izračunatog prosjeka se pridružuje središnjem članu od njih M.

Jednostavni pomični prosjeci za M=2m (**paran broj članova pomičnog prosjeka**):

$$\hat{Y}_t^* = \frac{1}{M}\left[\frac{1}{2}Y_{t-m} + \sum_{s=-(m-1)}^{m-1)} Y_{t+s} + \frac{1}{2}Y_{t+m}\right], \; t = m+1, m+2, ..., (n-m), \tag{3.9.3}$$

npr.: M=4: $\hat{Y}_t^* = \frac{1}{4}\left[\frac{1}{2}Y_{t-2} + \sum_{s=-1}^{1} Y_{t+s} + \frac{1}{2}Y_{t+2}\right]$.

Gornji izraz kod parnog broja članova pomičnog prosjeka omogućuje tzv. Centriranje jer bi se u suprotnom pomični prosjeci našli između dvije susjedne vremenske točke.

Jednostavni pomični prosjeci najčešće se primjenjuju kod odstranjivanja sezonskih (do jedne godine) i cikličkih (više od jedne godine) komponenti iz strukture vremenskog niza. Primjenjuje se i u slučaju kada se pojava razvija približno po linearnom trendu. U suprotnom pomični će prosjeci (izglađivanjem) sistematski precjenjivati ili podcjenjivati prisutni trend u vremenskom nizu. U takvoj situaciji se umjesto jednostavnih mogu koristiti vagani pomični prosjeci.

Vagani pomični prosjeci (ponderi W_s obično su unaprijed poznati i tabelirani):

$$\hat{Y}_t^* = \sum_{s=-m}^{m} W_s Y_{t+s}, \quad t = m+1, m+2, ..., (n-m) . \tag{3.9.4}$$

Pri izračunavanju vaganih pomičnih prosjeka uzima se neparan broj frekvencija vremenskog niza (M=2m+1), ponderi su simetrični u odnosu na središnji, a njihov zbroj je jednak nuli.

Primjer 3.9.1.

Tablica 3.12.

Instalirana računala u poslovnici «Z» u razdoblju od 1997. do 2005. god.

Godina	Broj instaliranih računala (Y_t)	Pomični prosjek *M=3*	Pomični prosjek *M=4*
1996.	45	*	*
1997.	64	49	*
1998.	37	43	47
1999.	29	45	49
2000.	69	52	48
2001.	58	53	53
2002.	33	54	56
2003.	72	52	52
2004.	50	56	*
2005.	46	*	*

Izvor: Podaci poslovnice «Z», 2006. god.

Zadatak je izračunati trogodišnje i četverogodišnje pomične prosjeke i grafički ih prikazati.

Izraz za jednostavne pomične prosjeke glasi:

$$\widehat{Y}_t^* = \frac{1}{M}\sum_{s=-m}^{m} Y_{t+s} \qquad t=m+1,\ m+2,\ldots,(n-m)$$

Trogodišnji pomični prosjeci (*M=3, m=1*) iznose:

$$\widehat{Y}_t^* = \frac{1}{3}\sum_{s=-1}^{1} Y_{t+s} \qquad t=2,3,\ldots 9$$

pa su vrijednosti sljedeće:

$$\widehat{Y}_2^* = \frac{1}{3}\left[Y_1 + Y_2 + Y_3\right] = \frac{1}{3}\left[45 + 64 + 37\right] = 49 \qquad \ldots$$

$$\widehat{Y}_9^* = \frac{1}{3}\left[Y_8 + Y_9 + Y_{10}\right] = \frac{1}{3}\left[72 + 50 + 46\right] = 56$$

Četverogodišnji centrirani pomični prosjeci (*M=4, m=2*) iznose:

$$\hat{Y}_t^* = \frac{1}{4}\left[\frac{1}{2}Y_{t-2} + \sum_{s=-1}^{1} Y_{t+s} + \frac{1}{2}Y_{t+2}\right] \qquad t=3,4,\ldots 8$$

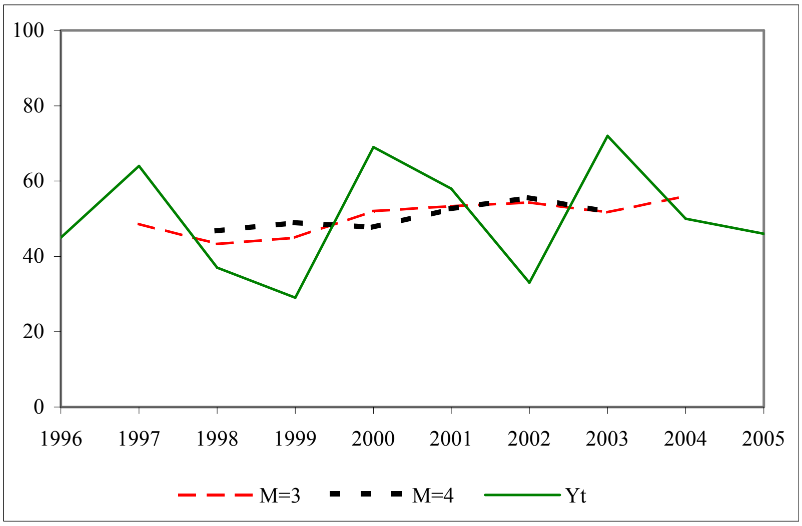
vrijednosti su sljedeće:

$$\hat{Y}_3^* = \frac{1}{4}\left[\frac{1}{2}Y_1 + Y_2 + Y_3 + Y_4 + \frac{1}{2}Y_5\right] = \frac{1}{4}\left[\frac{1}{2}\cdot 45 + 64 + 37 + 29 + \frac{1}{2}\cdot 69\right] = 47 \quad \ldots$$

$$\hat{Y}_8^* = \frac{1}{4}\left[\frac{1}{2}Y_6 + Y7_9 + Y_8 + Y_9 + \frac{1}{2}\cdot Y_{10}\right] = \frac{1}{4}\left[\frac{1}{2}\cdot 58 + 33 + 72 + 50 + \frac{1}{2}46\right] = 52$$

Grafikon 15.

Instalirana računala u poslovnici «Z», trogodišnji i četverogodišnji pomični prosjeci



Izvor: Tablica 3.12.

3.10. Standardna dekompozicija vremenske serije

Svaka **pojava** čije je kretanje predstavljeno u obliku **vremenskog niza sastoji se od više komponenata**, koje su ponekad vidljive na prvi pogled (na primjer iz grafičkog prikaza), a ponekad je potrebna složenija statistička analiza da bi se te komponente utvrdile i analizirale. **Najčešća dekompozicija vremenskog niza je**:

a) **trend komponenta**

b) **sezonska komponenta**

c) **ciklička komponenta**

d) **slučajna komponenta.**

Vrijedi da je **vremenski niz**:

$$Y_t = T + S + C + e, \tag{3.10.1}$$

gdje je: T -trend komponenta, S -sezonska komponenta, C -ciklička komponenta i e -slučajna komponenta. Ovdje se pretpostavlja da je djelovanje svih komponenti na vremenski niz aditivno, tj. njihovo djelovanje na pojavu u apsolutnom iznosu je isto bez obzira na tijek vremena.

Svaki vremenski niz ne mora sadržavati sve ove komponente. S obzirom na složenost vremenskog niza nije uvijek moguće strogo razlučiti pojedine komponente vremenskog niza jer se **može dogoditi da jedna komponenta prigušuje drugu**.

Trend komponenta pretpostavlja da vrijednosti promatranog vremenskog niza imaju dugoročnu tendenciju rasta ili pada kroz vrijeme. Takvo kretanje se u statističkoj obradi vremenskih nizova može izraziti trend-modelima koji mogu biti rastući ili padajući i linearni ili nelinearni, ovisno o tendenciji kretanja vremenskog niza.

Sezonska komponenta pretpostavlja sistematsko kretanje vrijednosti pojave unutar jedne godine, koje se onda ponavlja u svakoj narednoj godini. Takvo kretanje je u praksi najčešće uzrokovano neekonomskim faktorima, kao što su, na primjer klimatske prilike i državni praznici. Sezonska komponenta može biti prisutna kod vremenskih nizova čije su frekvencije, odnosno vrijednosti pojave, mjerene mjesečno ili kvartalno. Na primjer ako je vremenski niz broj turista na nekom području, on će biti veći u ljetnim i zimskim mjesecima zbog praznika i godišnjih odmora. Ta kretanja imaju sezonski karakter i ponavljaju se u svakoj godini.

Ciklička komponenta pretpostavlja period obnavljanja sistematskog kretanja vrijednosti pojave koji je duži od jedne godine. U praksi je ponekad teško pratiti cikličku komponentu jer se često ne raspolaže s dovoljno dugim vremenskim serijama. Ciklička kretanja su u praksi izražena u raznim granama privrede, kao što su građevinarstvo, stočarstvo i slično, a uzroci takvih kretanja su cikličke gospodarske aktivnosti. Ponekad ciklusi kretanja pojave nisu jasno izraženi, već su prigušeni trend komponentom odnosno dugoročnom tendencijom kretanja. Po nekim autorima trend komponenta je ciklička komponenta vremenske serije s veoma dugim periodom obnavljanja.

Slučajna (ili nesistematska) komponenta predstavlja sve ostale utjecaje na vrijednosti promatrane pojave čiji učinak nije sistematski, već se ti utjecaji javljaju s nepredvidivim djelovanjem u nekim vremenskim razdobljima. Na primjer mogu se javiti zbog nepredvidivosti prirode poslovnih i gospodarskih pojava, zbog iznenadne vremenske nepogode i slično.

Primjer 3.10.1.

Tablica 3.13.

Prevezeni putnici pomorskim prometom na području «Z», 2001.-2005. god.

Godina	Prevezeni putnici			
	I. kvartal	II. kvartal	III. kvartal	IV. kvartal
2001.	642	2369	3856	913
2002.	688	2174	4211	1052
2003.	579	2422	4196	863
2004.	727	2519	4327	986
2005.	825	2631	4528	1094

Izvor: Statistika područja «Z», ožujak 2006. god.

Zadatak je izračunati četverogodišnje pomične prosjeke, sezonske faktore, rezidualne faktore (trend i cikličku komponentu) i seriju desezonirati te grafički prikazati zadani niz i desezoniranu pojavu.

Rješenje:

Model glasi: $Y_t = T_t I_{St} \varepsilon_t$ pri čemu je:

T_t trend komponenta, odnosno pomični prosjeci
I_{St} sezonska komponenta
ε_t rezidualna komponenta

Godina, kvartal	Broj putnika (Y_t)	Pomični prosjeci M=4	Prve procjene sezonskih faktora	Sezonski faktori	Desezonirana serija	Rezidualni faktori
2001.: I.	642	*	*	0,3387364	1895,28	*
II.	2369	*	*	1,1570488	2047,45	*
III.	3856	1950,75	1,976676	2,0382841	1891,79	0,969774
IV.	913	1932,13	0,472537	0,4659307	1959,52	1,014178
2002.: I.	688	1952,13	0,352436	0,3387364	2031,08	1,040445
II.	2174	2013,88	1,079511	1,1570488	1878,92	0,932987
III.	4211	2017,63	2,087107	2,0382841	2065,95	1,023953
IV.	1052	2035,00	0,516953	0,4659307	2257,85	1,109507
2003.: I.	579	2064,13	0,280506	0,3387364	1709,29	0,828096
II.	2422	2038,63	1,188056	1,1570488	2093,26	1,026798
III.	4196	2033,50	2,063437	2,0382841	2058,59	1,012340
IV.	863	2064,13	0,418095	0,4659307	1852,21	0,897333
2004.: I.	727	2092,63	0,347411	0,3387364	2146,21	1,025607
II.	2519	2124,38	1,185761	1,1570488	2177,09	1,024815
III.	4327	2152,00	2,010688	2,0382841	2122,86	0,986461
IV.	986	2178,25	0,452657	0,4659307	2116,19	0,971511
2005.: I.	825	2217,38	0,372062	0,3387364	2435,52	1,098381
II.	2631	2256,00	1,166223	1,1570488	2273,89	1,007929
III.	4528	*	*	2,0382841	2221,48	*
IV.	1094	*	*	0,4659307	2347,99	*

Četverogodišnji centrirani pomični prosjeci (*M=4, m=2*) iznose:

$$\widehat{Y}_t^* = \frac{1}{4}\left[\frac{1}{2}Y_{t-2} + \sum_{s=-1}^{1} Y_{t+s} + \frac{1}{2}Y_{t+2}\right] \qquad t=3,4,\ldots 18$$

i prikazani su u trećem stupcu tablice.

U četvrtom stupcu tablice dane su prve procjene sezonskih faktora kao omjer odgovarajućih vrijednosti vremenskog niza i pomičnih prosjeka. S obzirom da procjene sezonskih faktora za iste kvartale variraju treba izračunati prosjek za iste kvartale, a to je učinjeno u sljedećoj tablici:

Godina	Prve procjene sezonskih faktora			
	I. kvartal	II. kvartal	III. kvartal	IV. kvartal
2001.	*	*	1,9766756	0,4725367
2002.	0,3524364	1,0795109	2,0871074	0,5169533
2003.	0,2805063	1,1880557	2,0634374	0,4180948
2004.	0,3474105	1,1857605	2,0106877	0,4526569
2005.	0,3720616	1,1662234	*	*
Prosjek	0,3381037	1,1548876	2,0344770	0,4650605
Sezonski faktori	0,3387364	1,1570488	2,0382841	0,4659307
Sezonski indeksi	33,873640	115,70488	203,82841	46,593072

Prosjek je računan kao jednostavna aritmetička sredina prvih procjena sezonskih faktora istih kvartala. S obzirom da zbroj sezonskih faktora mora biti jednak sezonskom periodu (podaci su kvartalni, pa je to 4), a zbroj prosjeka iznosi: 3,9925288=0,3381037+1,1548876+2,0344770+0,4650605, stoga prosjeke treba korigirati tako da njihov zbroj bude 4. U tu svrhu rabi se korektivni faktor: 1,001871=4/3,9925288 kojim se množi svaki izračunati prosjek te pomnoži sa 100. Na taj su se način dobili sezonski indeksi po kvartalima koji su prikazani u posljednjem retku gornje tablice. Tako npr. sezonski indeks 203,82841 pokazuje da je razina pojave u III. kvartalu svake godine u prosjeku za 103,82841% viša zbog sezonskih utjecaja.

Desezonirana serija, odnosno vrijednosti očišćene od sezonskih utjecaja dobivene su dijeljenjem vrijednosti vremenskog niza sa sezonskim faktorima.

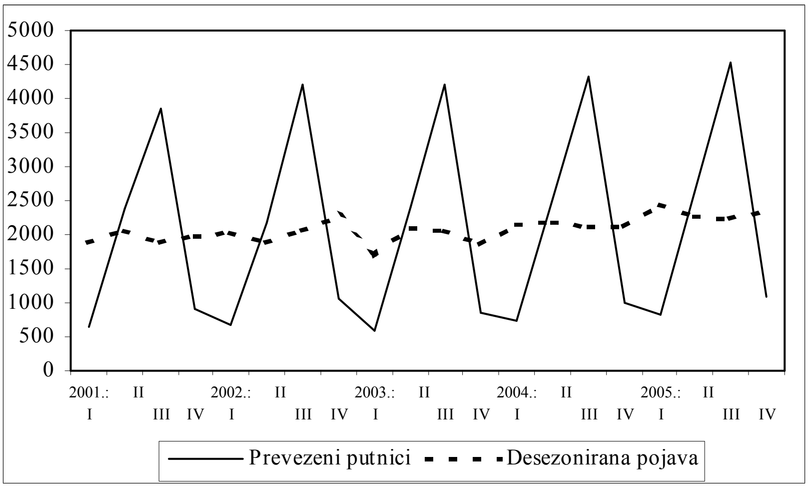
Rezidualni faktori dobiju se dijeljenjem desezoniranih vrijednosti s pomičnim prosjecima kao procjenama trenda. Indeks rezidualnih odstupanja jednak je umnošku rezidualnog faktora i 100, npr. indeks rezidualnih odstupanja u III. kvartalu 2001. godine iznosi 96,9774 a to znači da je razina pojave u tom kvartalu bila niža za 3,026% zbog rezidualnih utjecaja.

Analizirani vremenski niz može se raščlaniti na komponente, tako npr. vrijednost frekvencije vremenskog niza u III. kvartalu 2001. godine iznosi 3856 (treća u nizu), a rastavljena na faktore glasi:

$$Y_3 = T_3 I_3 \varepsilon_3 = 1951 \cdot 2{,}0382841 \cdot 0{,}969774 = 3856$$

Grafikon 16.

Prevezeni putnici i desezonirana pojava



Izvor: Tablica 3.13.

Literatura

1. Aczel, A. D., Complete Business Statistics, 4th edt., Irving McGraw-Hill, New York, 1999.
2. Allen, M. V. and Myddelton, D. R., Essential Management Accounting, Prentice Hall, London, 1992.
3. Anderson, D. R., et. al., Statistics for Business and Economics., 2nd edt., West, St. Paul, 1981.
4. Anderson, D. R. et al. Quantitative Methods for Business. 7th edt. South-Western College Publishing, 1998.
5. Anderson, R. E., Black, W., Hair, J. F. and Tatham, R. L., Multivariate Data Analysis, Prentice Hall, 1998.
6. Anderson, T. W., An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics), Wiley-Interscience, 2003.
7. Arnold, J. A. and Hope, A. J. B., Accounting for Management Decisions, Prentice Hall, London, 1990.
8. Babbie, E., Halley, F., Zaino, J., Adventures in Social Research, Data Analysis Using SPSS^X for Windows 95./98., Pine Forge Press, USA, 2000.
9. Basilevsky, A. T., Statistical Factor Analysis and Related Methods: Theory and Applications, John Wiley and Sons Inc., 1994.
10. Bešter, M. i Bregar, L., Ekonomska statistika, Ekonomski fakultet Ljubljana, Ljubljana, 1986.
11. Blejec, M., Statističke metode za ekonomiste, Ekonomski fakultet Ljubljana, Ljubljana, 1976.
12. Box, G.E.P. and Pierce, D.A., Distributions of Residual Autocorrelations in Autoregressive Integrated Moving Average Models, Journal of the American Statistical Association, 65, 1970., pp. 1509.-1526.
13. Brooks, C. GARCH, Modelling in Finance: A Review of the Software Options, Economics Journal, 107 (443), 1997., pp. 1271.-1276.
14. Brooks, C., Forecasting Stock Return Volatility: Does Volume Help?, Journal of Forecasting, 17, 1998., pp. 59.-80.
15. Brooks, C., Introductory econometrics for finance, University Press, Cambridge, 2002.
16. Canavos, G. C. and Miller, D. M., An Introduction to Modern Business Statistics, Belmont, Wadsworth, 1993.

17. Charniak, E., Statistical Language Learning, A Bradford Book, The MIT Press, Cambridge and Massachusetts, 1993.

18. Chiang, A. C., Osnovne metode matematičke ekonomije. Mate, Zagreb, 1994.

19. Cochran, W. G., Sampling Techniques, Wiley, New York, 1977.

20. Dicky, D.A. and Fuller, W.A., Distribution of Estimators for Time Series Regressions with a Unit Root, Journal of the American Statistical Association, 74, 1979., pp. 427.-431.

21. Dillon, W. R. and Goldstein, M., Multivariate Analysis, Methods and Applications, Wiley, New York, 1984.

22. Dodge, M., Kinata, C. i Stison, C., Kako koristiti Microsoft Excel 97, Znak, Zagreb, 1997.

23. Enders, W., Applied Econometric Time Series, 2nd edt., Wiley, Alabama, 2004.

24. Engle, R.F., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica, 50 (4), 1982., pp. 987.-1007.

25. Feller, W., An Introduction to Probability Theory and its Applications, Vol. I. and Vol. II., Wiley, New York, 1971.

26. Fisher, R. A., Statistical Methods for Research Workers, 11th edt., Oliver and Boyd, Edinburgh, 1950.

27. Fisher, R. A., Statistical Methods, Experimental Design and Scientific Inference, Oxford University Press, Oxford, 1993.

28. Fruk, M., Sezonalnost prinosa dionica na Zagrebačkoj burzi, Financijska teorija i praksa, 28 (4), 2004. pp. 435.-444.

29. Frye, C., Microsoft Excel 2002. Korak po korak, Algoritam, Zagreb, 2003.

30. Fulton, J., Vodič kroz Excel 97., Znak, Zagreb, 1997.

31. GaYnor, P. and Kirkpatrick, R. C., Introduction to Time-Series Modelling and Forecasting in Business and Modelling, McGraw-Hill, Inc., New York, 1994.

32. Gujarati, D. N., Basic Econometrics, McGraw-Hill, Inc., New York, 1995.

33. Hadživuković, S., Tehnika metoda uzorka, Naučna knjiga, Beograd, 1975.

34. Harrell, F. E. Jr., Regression Modelling Strategies, With Applications to Linear Models, Logistic Regression, and Survival Analysis, Springer, New York, 2001.

35. Heij, C., et. al. Econometrics Methods with Applications in Business and Economics, Oxford University Press, New York, 2004.

36. Horvatić, K., Linearna algebra, Golden marketing-Tehnička knjiga, Zagreb, 2004.

37. Johnson, R. A. and Wichern, D. W., Applied Multivariate Statistical Analysis, Prentice Hall, London, 2002.

38. Jolliffe, I.T., Principal Component Analysis, Springer, Berlin, New York: 2002.

39. Jovičić, M., Ekonometrijski metodi, Savremena administracija, Beograd, 1981.

40. Jurun, E. and Pivac, S., Macroeconomic Modelling Relaxing Theoretic Assumptions in the Croatian Financial Sphere, Proceedings of the 8th International Symposium on Operational Research, SOR'05, Nova Gorica, Slovenia, 2005., pp. 285.-290.

41. Kaplan, R. S., The Evaluation of Management Accounting, The Accounting Review, 59 (3), 1984., pp. 95.-101.

42. Karatzas, I. and Shreve, S E., Methods of Mathematical Finance, Springer, Berlin, New York, 2001

43. Kazmier, L. J., Business Statistics, Schaum's, Outline Series, McGraw-Hill, 1996.

44. Kendall, M. and Stuart, A., The Advanced Theory of Statistics, Vol I., Griffin, London, 1973.

45. Kendall, M. and Stuart, A., The Advanced Theory of Statistics, Vol II., Griffin, London, 1974.

46. Kendall, M. and Stuart, A., The Advanced Theory of Statistics, Vol III., Griffin, London.1976.

47. Kish, L., Survey Sampling, Wiley, New York, 1965.

48. Kmenta, J., Počela ekonometrije, MATE d.o.o., Zagreb, 1997.

49. Kolesarić, V. i Petz, B., Statistički rječnik, Tumač statističkih pojmova, Naklada Slap, Zagreb, 2003.

50. De Levie, R., Advanced Excel for Scientific Data Analysis, Oxford University Press, Oxford, 2004.

51. Ljung, G.M., and Box G.E.P., On a Measure of Lack of Fit in Time Series Models, Biometrica, 65 (2), 1978., pp. 297.-303.

52. Lucey, T., Management Information Systems, DP Publication LTD, London, 1989.

53. Mason, D. R. and Lind, D. A., Statistical Techniques in Business and Economics, Boston, IRVIN Publishing, Massachusetts, 1993.

54. Metre, J. G. and Gilbreath, G. H., Statistics for Business and Economics, Plano, Business Publications, Texas, 1983.

55. Mills, T. C., The Econometric Modelling of Financial Time Series, 2nd edt., University Press, Cambridge, 1999.

56. Milton, J. S., et al., Introduction to Statistics, D. C. Heath, Lexington, 1986.

57. Momirović, K., Uvod u analizu nominalnih varijabli, Metodološke sveske 2, JUS, Ljubljana, 1998.

58. Mood, A. M. and Graybill, F. P., Introduction to the Theory of Statistics, McGraw-Hill, New York, 1963.

59. Mott, G. Management Accounting for Decision Makers, Pitman Publishing, London, 1991.

60. Neter, J., et al., Applied Linear Regression Models, IRWIN, Boston, 1989.

61. Neter, J., et al., Applied Statistics, 4th edt., Allyn and Bacon, Boston, 1993.

62. Newbold, P., Statistics for Business and Economics, Prentice-Hall, Englewood Cliffs, 1991.

63. Patterson, K., An Introduction to Applied Econometrics, a time series approach, Palgrave, Hampshire, 2000.

64. Pauše, Ž., Vjerojatnost, informacija, stohastički procesi, Školska knjiga, Zagreb, 2003.

65. Pavlić, I., Statistička teorija i primjena, Tehnička knjiga, Zagreb, 1971.

66. Petz, B., Osnovne statističke metode za nematematičare, Naklada Slap, Zagreb, 2004.

67. Pivac, S. and Jurun, E., Parameter Estimation in Excel, Proceedings of the 28th International Convention MIPRO 2005. Computers in Education, Opatija, 2005., pp. 168.-173.

68. Pivac, S., Jurun, E. and Jujnović, I., Ocjena parametara nelinearnih funkcija u programskom paketu Statistica na primjeru nezaposlenosti, Proceedings of the 29th International Convention MIPRO 2006. Computers in Education, Opatija, 2006., pp. 241.-245.

69. Pivac, S. Jurun, E. and Jerak. N., Rješavanje simultanih jednadžbi kao ekonometrijskog modela pomoću programskog paketa Excel, Proceedings of the 29th International Convention MIPRO 2006. Computers in Education, Opatija, 2006., pp. 246-251.

70. Pivac, S. i Šego, B., Statistika, udžbenik sa zbirkom zadataka za IV. razred srednje ekonomske škole, Alkascript, Zagreb, 2005.

71. Rozga, A., Statistika za ekonomiste, Sveučilište u Splitu, Ekonomski fakultet Split, 2003.

72. Rozga, A. i Grčić, B., Poslovna statistika, Ekonomski fakultet Split, 2003.

73. Seddighi, H., R., Lawler, K. A. and Katos, A. V., Econometrics, A practical approach, Routledge, London and New York, 2006.

74. Seplaki, L., Atorneys' Dictionary and Handbook of Economics and Statistics, Professional Horizons Press, New York, 1991.

75. Serdar, V. i Šošić, I., Uvod u statistiku, Školska knjiga, Zagreb, 1994.

76. Siegel, A. F., Practical Business Statistics, IRVIN Publishing, Boston, Massachusetts, 1994.

77. StatSoft, Statistica System Reference, StatSoft Inc., Tulsa, 2001

78. Studenmund, A. H., Using Econometrics, A Practical Guide, Pearson International Edition, Boston, New York, 2006.

79. Šošić, I., Metode statističke analize, Sveučilišna naklada Liber, Zagreb, 1983.

80. Šošić, I., Zbirka zadataka iz osnova statistike, Sveučilišna naklada Liber, Zagreb, 1985.

81. Šošić, I., Primijenjena statistika, Školska knjiga, Zagreb, 2004.

82. Taylor, S.J., Modelling Financial Time Series, John Wiley & Sons, New York, 1986.

83. Tenjović, L., Statistika u psihologiji, Centar za primenjenu psihologiju, Beograd, 2002.

84. The Central Bureau of Statistics. Republic of Croatia, http:// www.dzs.hr, 2006.

85. Vujković, T., Ekonometrijske metode i tehnike, Informator, Zagreb, 1976.

86. Žarković, S. S., Sampling Methods and Censuses, FAO, Rim, 1965.

87. Weisberg, H.F., Krosnick, J.A. and Bowen, B.D., Survey research, Polling and Data Analysis, 3rd edt., Ohio State University, Assessment Systems Corporation, USA, 2005.

88. Whigham, D., Quantitative Business Methods Using Excel, Oxford University Press, Oxford, 1998.

89. Wonnacott, T. H. and Wonnacott, R. J., Introductory Statistics for Business and Economics, 4th edt., Wiley, New York, 1990.

PRIVITAK UPORABA PROGRAMSKOG PAKETA *STATISTICA*

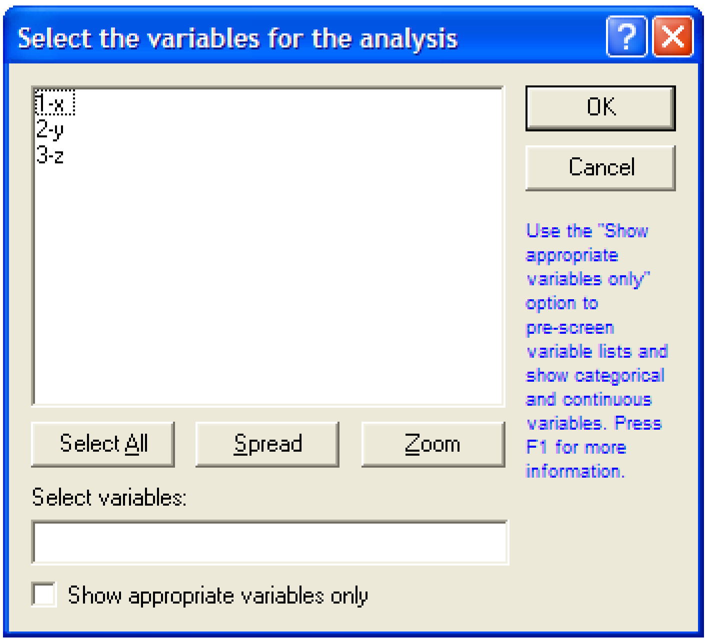
1. Uvod

S razvojem računala, statistički programski paketi postali su osnovni alat statističke analize. Neki od programa koji su namijenjeni statističkoj obradi su *Statistica*, *SPSS*, *SAS* i razni drugi programski paketi. U nastavku je objašnjena uporaba programskog sustava *Statistica*.

Većina zadataka u *Statistici* može se riješiti na više načina, a naveden je u pravilu najjednostavniji način. U zadacima u kojima rješenje ovisi o razini značajnosti ostavljena je razina signifikantnosti 0,05, koja je postavljena kao pretpostavka u *Statistici*. Vrijednost je moguće promijeniti.

U *Statistici* se zadaci rješavaju otvaranjem odgovarajućih prozora u kojima se izabiru potrebne varijable i izračuni. Kada se u prozoru označi sve potrebno izabere se ponuđena tipka, u pravilu **SUMMARY** ili **OK** nakon čega se dobiju rezultati ili se otvara novi prozor.

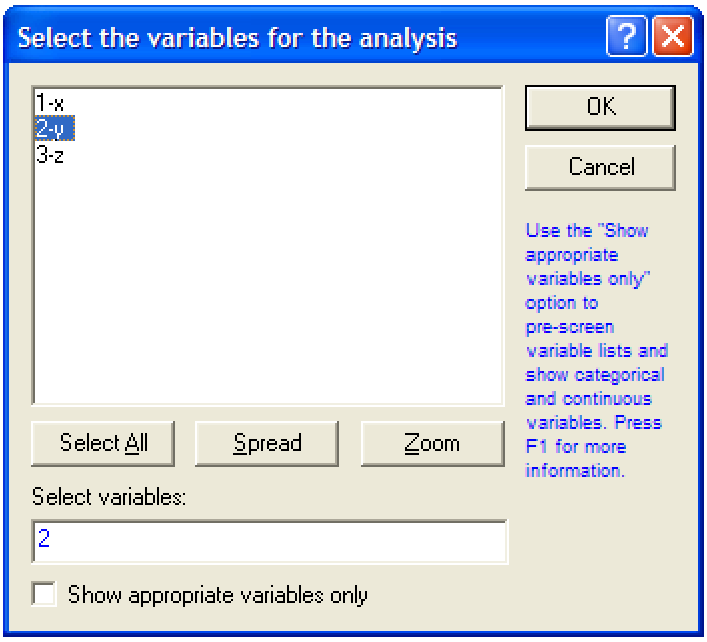
Kada se biraju varijable otvara se prozor s ponuđenim varijablama



Za svaku varijablu piše redni broj i kratki naziv varijable. Varijabla se izabire klikom miša, a za varijablu se upiše redni broj varijable u polju:

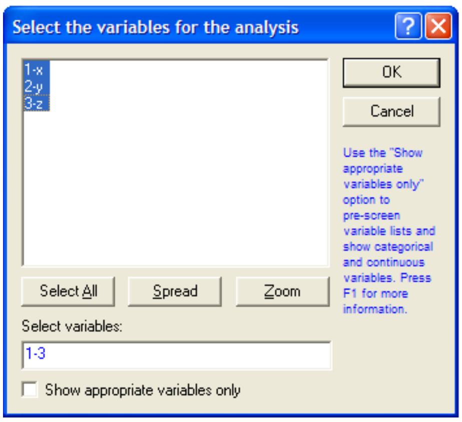
Select variables:

Primjer: izabrana je varijabla y, pod rednim brojem 2.

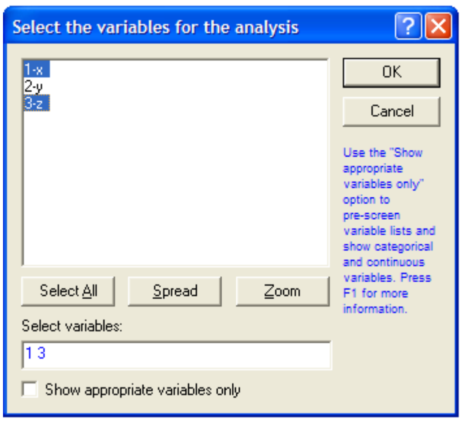


Ako se želi izabrati više varijabli koje su poredane jedna ispod druge, tada se označi prva varijabla, pritisne se tipka **SHIFT** , a nakon toga označi se posljednja varijabla. Ukoliko se žele izabrati varijable koje nisu jedna iza druge, tada se prije izbora sljedeće varijable izabere tipka **CTRL**.

Primjer: izbor varijabli od 1 do 3



Primjer: izbor varijabli 1 i 3



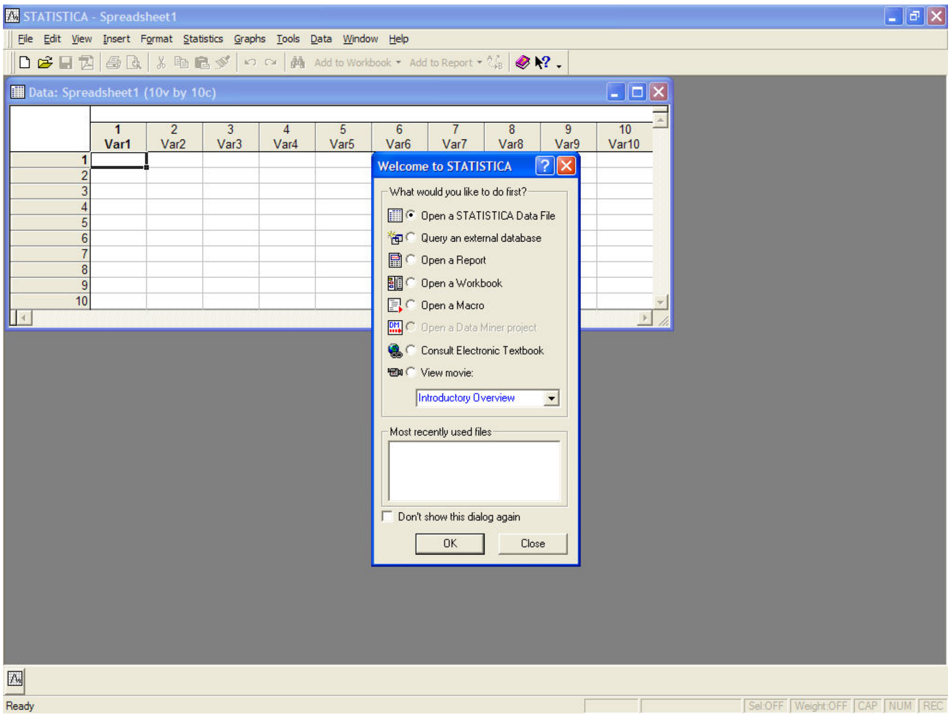
1.1. Pokretanje programa Statistica

Program se pokreće klikom na tipku Start te izborom odgovarajućih menija kojim se dolazi do programa *Statistica*. Najčešće se nalazi u meniju ***Programs*** i pod meniju ***Statistica** 7*.

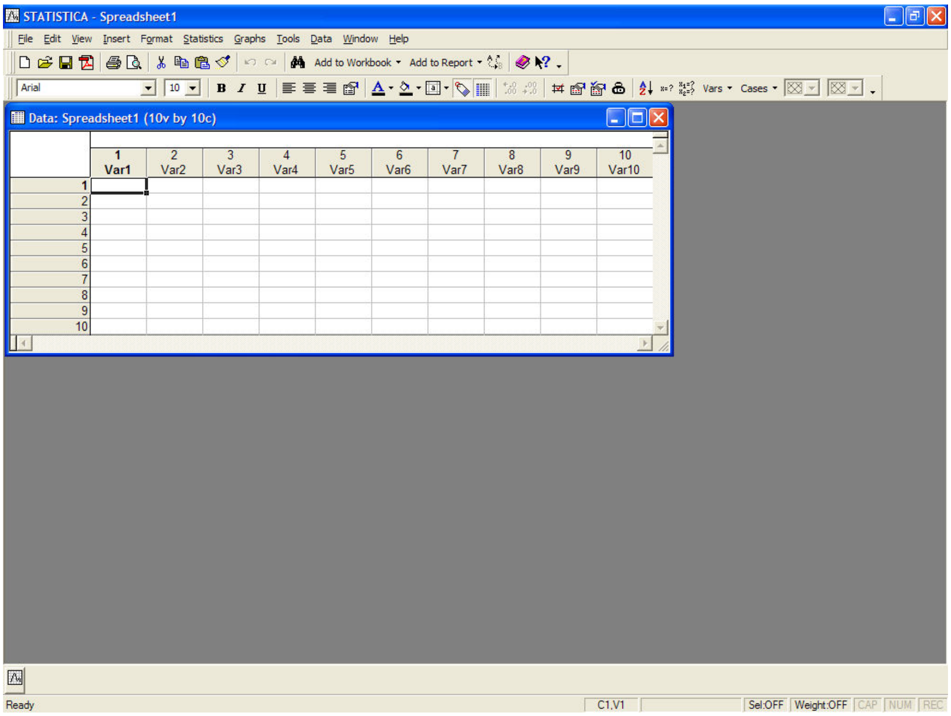
- **START / PROGRAMS / STATISTICA 7 / STATISTICA**

Drugi način pokretanja programa je pomoću ikone: koja se može nalaziti na radnoj površini (desktopu) ili radnoj traci (task baru).

Kada se *Statistica* pokrene pokazuju se dva prozora:



Manji prozor nudi izbor koji se dio sustava *Statistica* želi otvoriti. Izborom predložene opcije ***Open a STATISTICA Data File***, otvara se posljednji korišteni dokument.



Radni list (Spreadsheet) namijenjen je unosu, uređivanju, prikazu i pohrani podataka, a ima oblik tablice. Osnovni element radnog lista je **polje** (cell) u koje se unosi pojedini podatak. Polje se može odrediti kao presjek **stupca** koji predstavlja varijablu (variable) i **retka** koji predstavlja pojedini slučaj (case). Polje se selektira klikom na lijevu tipku miša ili navigacijom pomoću tastature. Polje koje je selektirano je aktivno polje, a koordinate tog polja prikazane su u statusnoj traci (status baru). Adresa polja označava se s C i rednim brojem retka te s V i rednim brojem stupca, tako npr. adresa (C3, V8) označava polje u osmoj varijabli i trećem retku. Varijabla odnosno redak selektira se klikom na **zaglavlje varijable** odnosno na odgovarajuće polje u **pred stupcu**.

U naslovnoj traci nalazi se naziv dokumenta (radnog lista), a u zagradi je naveden broj varijabli (v) i redaka (c).

Prilikom unosa podataka iz tablice naslov se zapisuje u **informativno polje** (info box) i u **zaglavlje** (header). Klikom na informativno polje selektira se cijeli radni list, a dvostrukim klikom selektira se polje i omogućuje unos i promjenu podatka. Zaglavlje je moguće selektirati jednim klikom miša, a podaci se, također, mogu unositi nakon dvostrukog klika lijevom tipkom miša.

U pojedino polje moguće je unijeti niz znakova; dokumente sa zvučnim ili video zapisima, grafovima, animacijama ili bilo koje dokumente kojima je ActiveX kompatibilan.

1.2. Stvaranje novog dokumenta

Primjer 1.1.

Starost i spol izabranih studenata upisanih na prvu godinu studija, akademske godine 2008./2009., na Ekonomski fakultet u Rijeci

Student/ica	*Godine*	*Spol*
A	18	Ženski
B	19	Muški
C	18	Ženski
D	25	Ženski
E	20	Muški

Izvor: Informacijski sustav visokih učilišta (simulirani podaci)

Treba kreirati novi dokument za unos podataka iz gornje tablice.

- **FILE / NEW…**

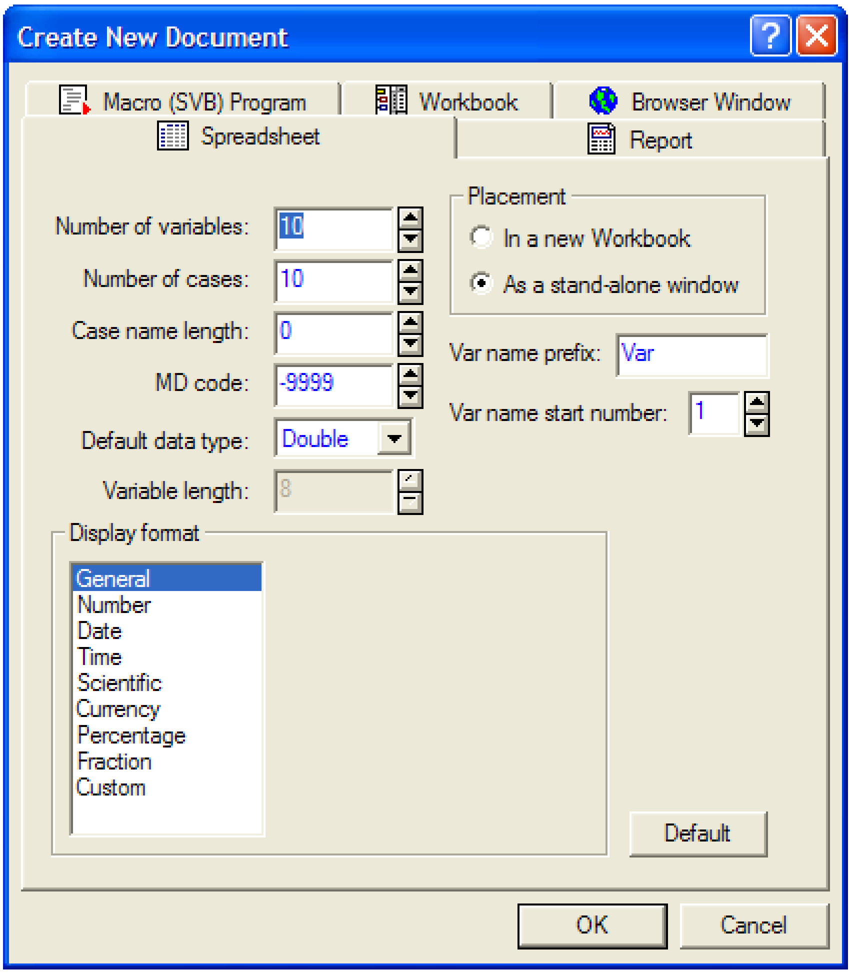
ili

- **<Crtl> + N**

ili

- **Ikona**

Nakon toga otvara se sljedeći prozor:



Ponuđeno je staranje novog **radnog lista** (Spreadsheet), **izvješća** (Report), **makroa** (Macro (SVB) Program), **Radnih zapisa** (Workbook) i **otvaranje internetske stranice** (Browser Window).

Radni list kreira se izborom opcije:

- **Spreadsheet**

- ***Number of variables*** – **Broj varijabli**. Unosi se broj stupaca, može se predvidjeti veći broj stupaca u kojima će biti rezultati izračuna.
- ***Number of cases*** – **Broj redaka**. Upisuje se broj redaka u tablici.

Može se ostaviti ponuđeni broj stupaca i redaka (10×10), a broj varijabli i broj redaka može se na jednostavan način promijeniti tijekom rada.

1.3. Unos podataka

Podatke iz primjera 1.1. treba unijeti u dokument.

- U **zaglavlje** dokumenta upisuje se naziv tablice:

Starost i spol izabranih studenata upisanih na prvu godinu studija, akademske godine 2008./2009., na Ekonomski fakultet u Rijeci

- U **informativno polje** upisuje se naziv pred stupca:

Student/ica

Prije početka unosa podataka potrebno je definirati varijable. Definiranje varijabli moguće je izvesti na više načina. U *Statistici* se većina zadataka može izraditi na više načina, a svejedno je koji se način koristi.

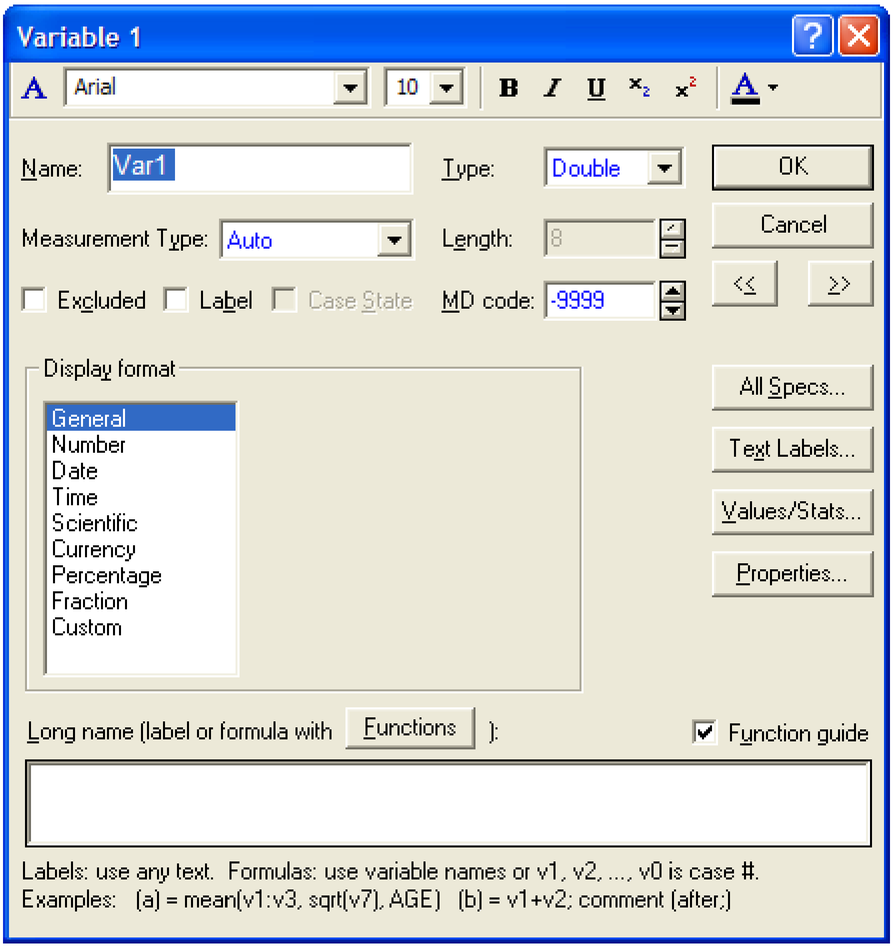
Definiraju se dvije varijable: godine i spol izabranih studenata. Nakon pozicioniranja na polje u određenoj varijabli može se izabrati:

- **DATA / VARIABLE SPECS...**

ili

- **2 puta kliknuti na zaglavlje varijable**

čime se otvara prozor za specifikaciju varijable:



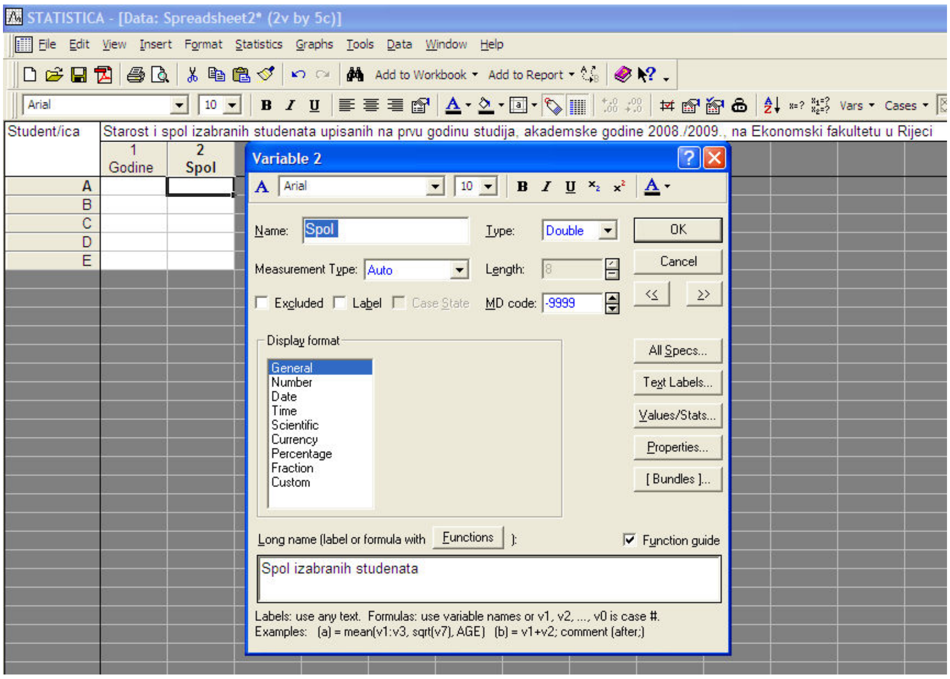
- ***Name*** – **Kratki naziv varijable** – Praktično je ako je kratko ime. Treba izbjegavati „**-**„ i **razmak** te rabiti samo slova engleske abecede. Premda program prihvaća i druge znakove, neke operacije ne mogu se izvoditi, ako se upotrebljavaju ti znakovi.
  - 1. varijabla: upisati ***Godine***
  - 2. varijabla: upisati ***Spol***

- ***Long name (label or formula with*** $\boxed{\textit{Functions}}$ ***)*** **–Dugi naziv varijable (ili formula)** – Unosi se dugi naziv, nema ograničenja dužine, stoga treba dati što bolji opis
  - 1. varijabla: upisati ***Godine starosti izabranih studenata***
  - 2. varijabla: upisati ***Spol izabranih studenata***

U ovu rubriku može se upisati i formula kojom se izračunavaju vrijednosti varijable. Formula započinje znakom jednakosti (=). Dugi naziv varijable u tom slučaju dolazi nakon formule, a između se stavlja znak (;) točka sa zarezom.

- ***Type*** **– Tip varijable**

  Predviđeni tip je *Double*. To je jedini tip varijable koji podržava decimale. Moguće je unositi brojeve, slova i ostale znakove. Najbolje je ostaviti ovaj tip varijable.



Umjesto otvaranja prozora za specifikaciju svake varijable posebno, moguće je izabrati opciju za specificiranje svih varijabli na sljedeće načine:

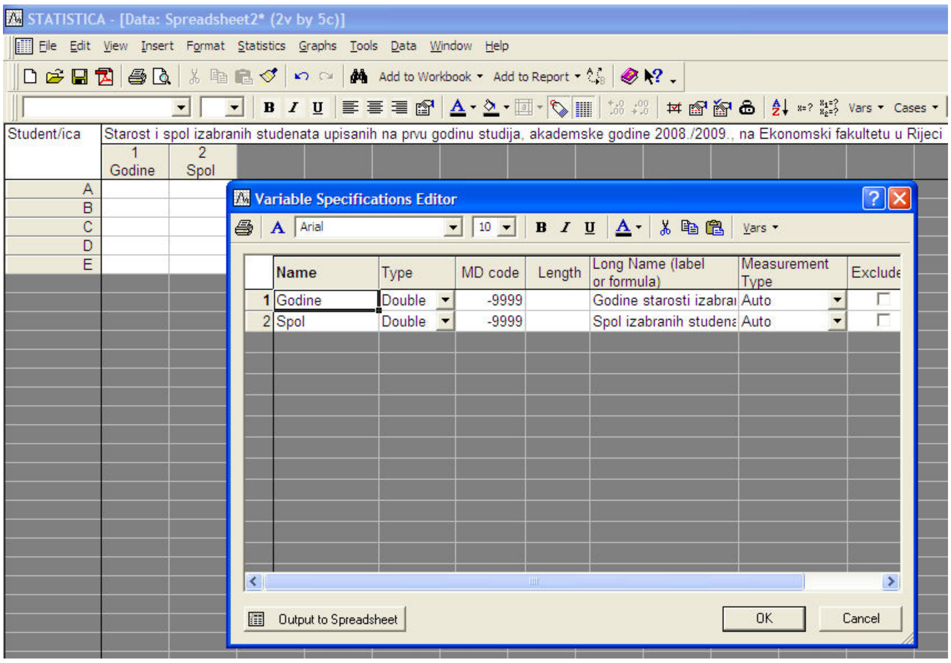
Na već otvorenom prozoru za specifikaciju varijable izabere se:

➢ ALL SPECS...

ili se preko sustava menija izabere

➢ **DATA / ALL VARIABLE SPECS…**

Nakon čega se otvara prozor za specifikaciju svih varijabli.



Ukoliko varijabla može poprimiti neke određene vrijednosti modaliteta, tada se mogu kodirati vrijednosti. Varijabla Spol može poprimiti dvije vrijednosti: *Ženski* i *Muški*.

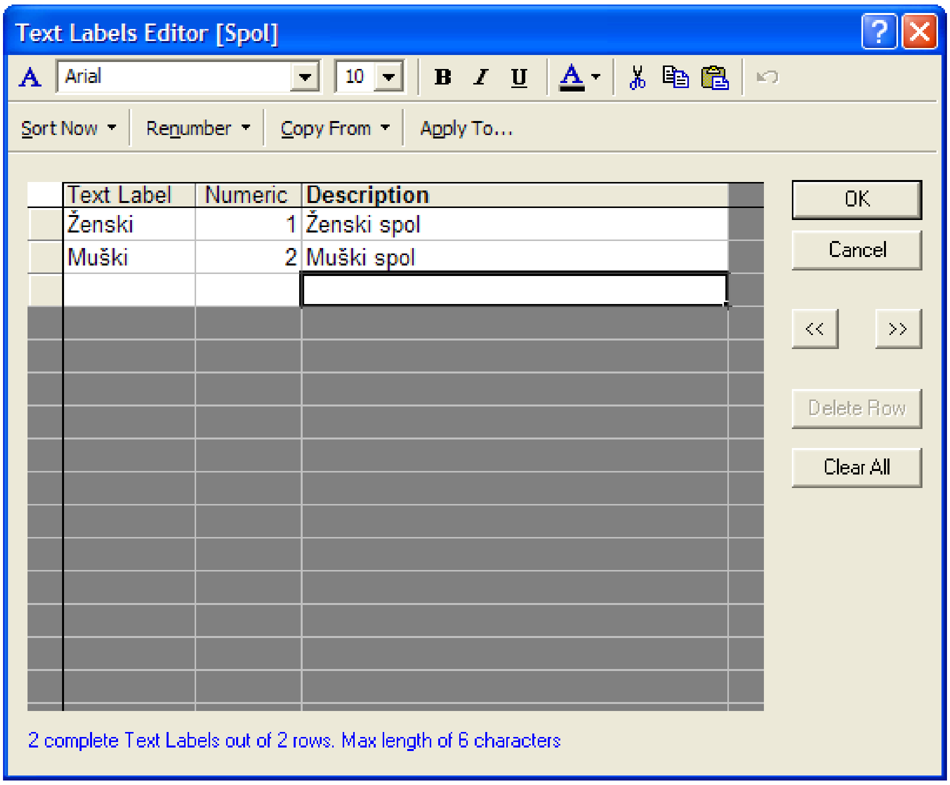
➢ Pozicionirati se na drugoj varijabli i otvoriti prozor za specifikaciju varijable ukoliko već nije otvoren

➢ **DATA / VARIABLE SPECS...**

➢ Izabrati tipku

➢ **Text Labels…**

te se otvara prozor za kodiranje varijabli.

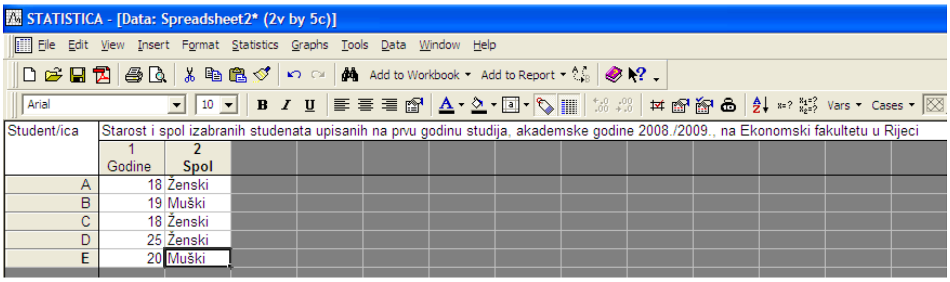


- ***Text label*** – **Modalitet varijable** – unosi se vrijednost koju varijabla može poprimiti

- ***Numeric*** – **Brojčana šifra** – izabire se broj, brojčana šifra, koja predstavlja vrijednost modaliteta

- ***Description*** – **Opis modaliteta** – ovdje se može staviti dugi naziv ili opis kodiranog modaliteta, nije obavezan unos

Nakon definiranja varijabli unose se vrijednosti. Kod unosa pred stupca i varijable godine prepisuju se vrijednosti zadane u tablici, a kod unosa spola unose se šifre: *1* i *2*, a vidi se *Ženski* i *Muški*.



1.4. Spremanje dokumenta

Kod prvog spremanja

- **FILE / SAVE ili SAVE AS…**

ili

- **<Crtl> + S**

ili

- **Ikona** 

Ako se dokument snima prvi put, otvara prozor ***Save As*** u koji se unosi naziv dokumenta i izabere direktorij u koji će biti spremljen. Radni list ima nastavak *.sta*, izvješće ima nastavak *.str*, nastavak makroa je *.svb*, radni zapisi označavaju se s *.stw*.

Ukoliko se u rubrici ***Save as type*** izabere neki drugi nastavak, tada se dokument snima u nekom drugom formatu, npr. Excelu, SPSS-u i sl.

Kad je dokument jednom snimljen, za spremanje izmjena više se ne koristi ***Save As***, nego samo ***Save***. Promjene se tada snimaju u već postojeću datoteku.

1.5. Otvaranje dokumenta

Postojeći dokument otvara se:

- **FILE / OPEN…**

ili

- **<Crtl> + O**

ili

- **Ikona** 

Prikazuje se prozor u kojem se izabere direktorij i dokument koji ima neki od ponuđenih nastavaka. U rubrici ***Files of Type*** može se izabrati neki drugi tip dokumenta.

1.6. Prikazivanje rezultata obrade

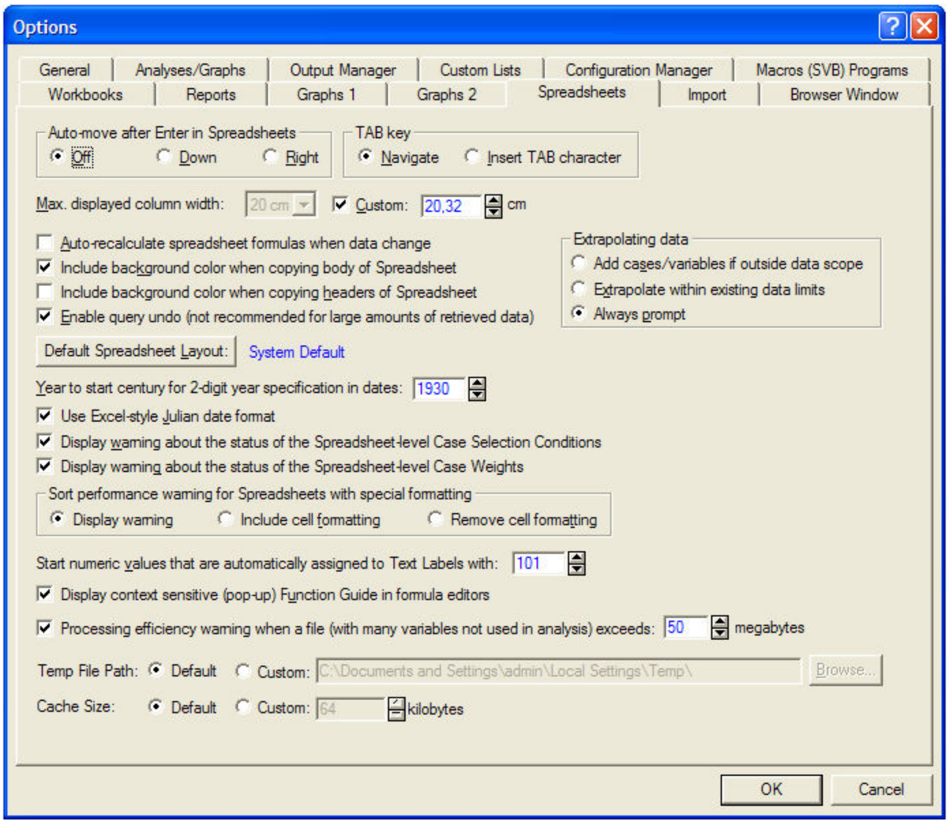
Rezultati statističkih izračuna prikazuju se u prozoru radnih zapisa (Workbook), a moguće ih je dobiti i u obliku izvješća.

Da bi se pokrenulo zapisivanje rezultata u izvješće, potrebno je putem menija izabrati:

- **TOOLS / OPTIONS...**

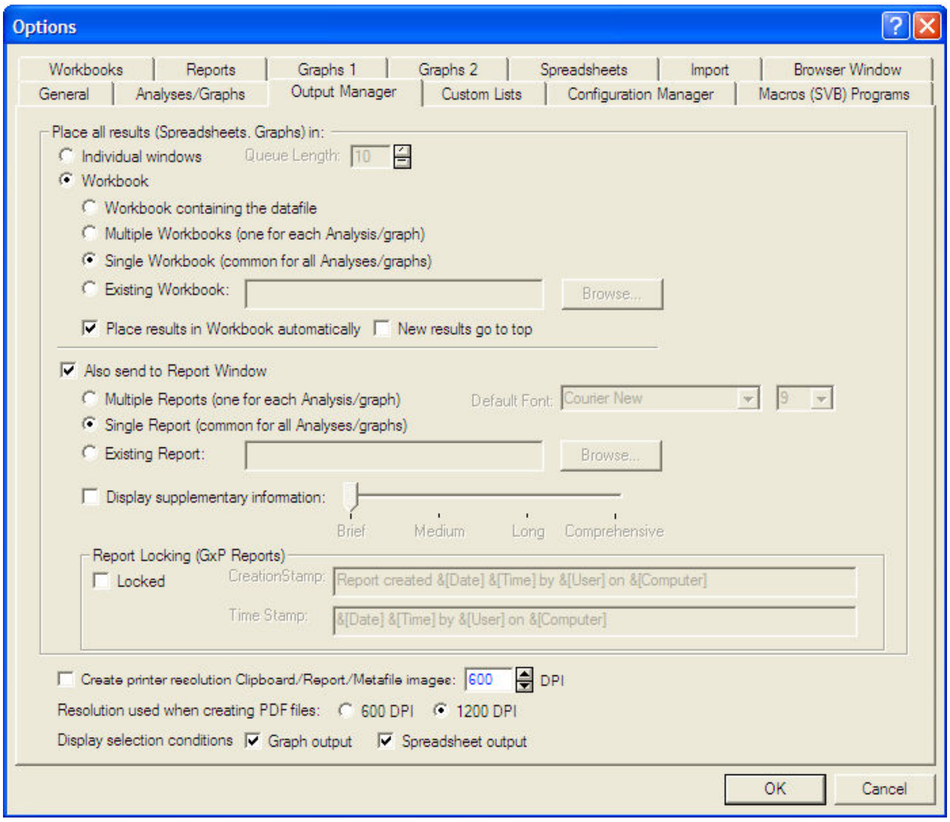
Nakon otvaranja prozora izabere se rubrika:

- **OUTPUT MANAGER**



i označi se:

- ➢ **ALSO SEND TO REPORT WINDOW**



Zapisivanje rezultata u izvješće može se izvršiti na više načina:

- ***Multiple Reports (one for each Analysis/graph)*** – stvaranje više prozora, po jedan prozor za svaku analizu ili grafički prikaz

- ***Single Report (common for all Analysis/graphs)*** – kreira se jedno izvješće, zajedničko za sve analize i grafikone

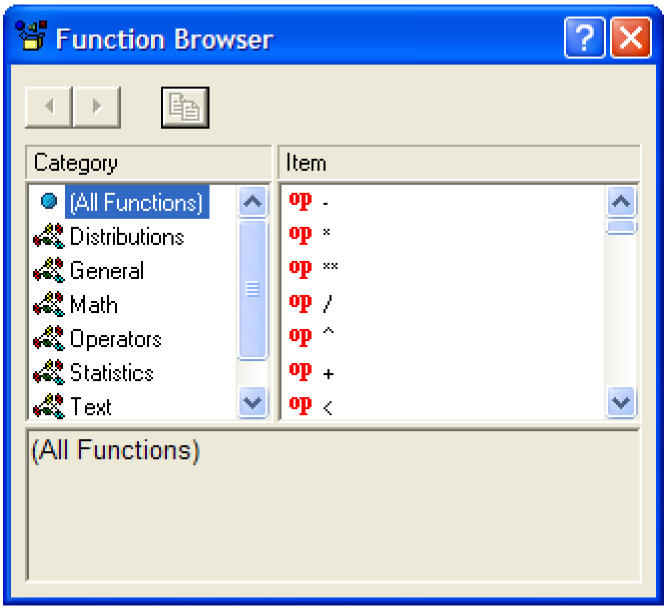
- ***Existing Report*** – rezultati se unose u neko već postojeće izvješće

Uz to moguće je izabrati i:

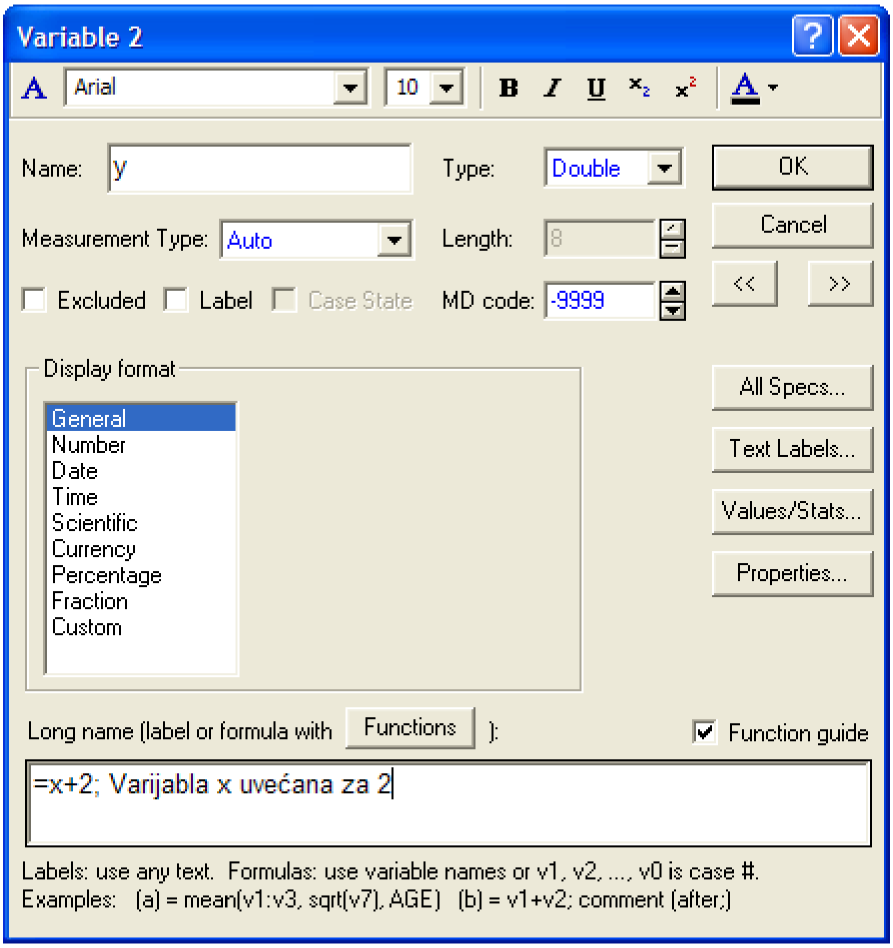
- ***Display supplementary information:*** – prikazuju se dodatne informacije
  - ***Brief*** – kratke
  - ***Medium*** – srednje
  - ***Long*** – duge
  - ***Comprehensive*** – iscrpne

1.7. Definiranje varijable formulom

Varijabla se može definirati formulom. Tada se u rubrici dugi naziv varijable, upisuje znak $\boxed{=}$, a nakon toga slijedi formula. Moguće je koristiti standardne matematičke operacije ili neku od ponuđenih funkcija koja se dobije pritiskom na tipku **Functions**.



Ako se nakon formule upiše **;** može se u nastavku upisati dugi naziv varijable.

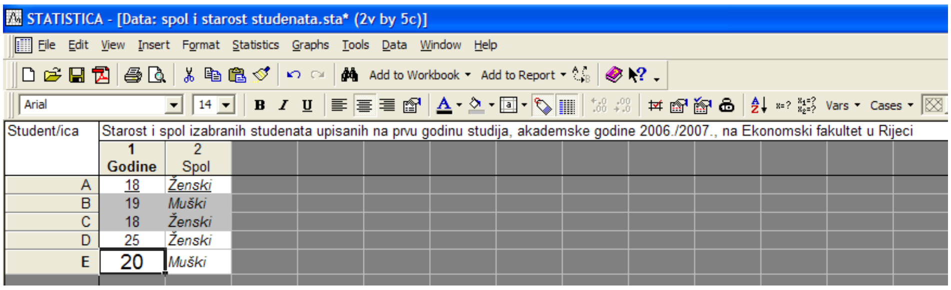


1.8. Uređivanje

Dio dokumenta koji se želi uređivati mora se prvo selektirati. Polje se selektira mišem ili navigacijom pomoću tastature. Varijabla se selektira klikom miša na zaglavlje varijable, redak se selektira klikom miša na naziv retka u pred stupcu. Blok podataka se izabere klikom na gornji lijevi kut bloka podataka i pritiskom na lijevu tipku miša dok se miš povlači do donjeg desnog kuta.

Nakon selektiranja izabranog dijela može se pristupiti uređivanju.

Na slici je prikazan dokument u kojem su vrijednosti varijable "Godine" centrirane, vrijednosti varijable "Spol" napisane su kurzivom (Italic), podaci u retku A su podcrtani, blok podataka od C2V1 do C3V2 je osjenčan, a polje C5V1 ima veću veličinu slova.

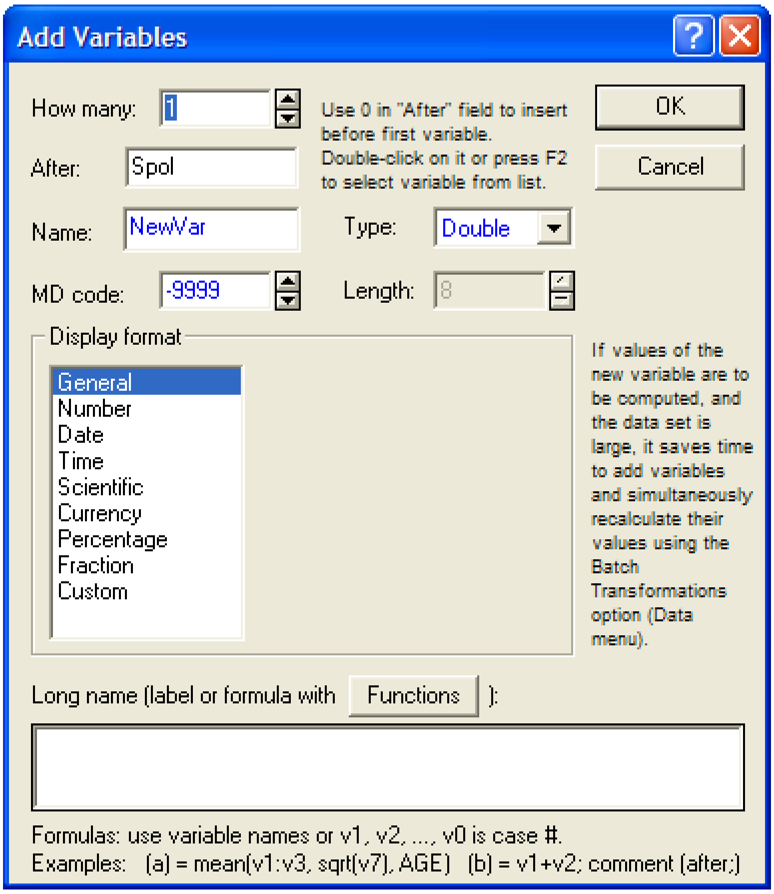


Osim mijenjanja izgleda, moguće je i dodavanje ili brisanje pojedinog stupca ili retka.

Dodavanje novih stupaca:

- **INSERT / ADD VARIABLES…**

Otvara se prozor za dodavanje varijabli:

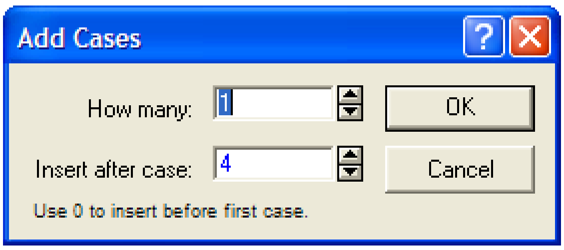


- ***How many*** – broj novih stupaca
- ***After*** – nakon koje varijable se unose nove

Dodavanje novih redaka:

- **INSERT / ADD CASES…**

Otvara se prozor za dodavanje redaka:

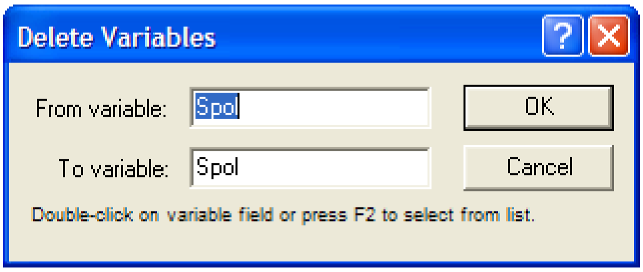


- ***How many*** – broj novih redaka
- ***Insert after case*** – nakon kojeg retka se unose novi

Brisanje stupaca:

- **EDIT / DELETE / VARIABLES…**

Otvara se prozor za brisanje varijabli:



- ***From variable*** – od koje varijable
- ***To variable*** – do koje varijable

Brisanje redaka:

- **EDIT / DELETE / CASES…**

Otvara se prozor za brisanje redaka:

- ***From cases*** – od kojeg retka
- ***To cases*** – do kojeg retka

2. Grafičko prikazivanje nominalnih (atributivnih) nizova

U *Statistici* je moguće napraviti razne vrste grafikona kao i njihove kombinacije. Prikazani su jednostavni stupci, dvostruki stupci i strukturni krug.

Primjer 2.1.

Stanovništvo prema spolu u jadranskim županijama, popis 2001. god.

Županija	*Žene*	*Muškarci*
Istarska	106.375	99.969
Primorsko – goranska	158.290	147.215
Ličko – senjska	27.182	26.495
Zadarska	82.404	79.641
Šibensko – kninska	58.225	54.666
Splitsko – dalmatinska	237.545	226.131
Dubrovačko – neretvanska	63.492	59.378

Izvor: Popis stanovništva 2001, DZS, str. 86. – 97.

a) Prikažite grafički jednostavne stupce za žene
b) Prikažite grafički dvostruke stupce
c) Strukturnim krugom prikažite raspodjelu muškaraca po županijama

Rješavanje zadatka započinje kreiranjem novog dokumenta.

- U zaglavlje dokumenta upisati:

Stanovništvo prema spolu u jadranskim županijama, popis 2001. god.

Definirati pred stupac:

- U informativno polje upisati ***Županija***

    - Upisati vrijednosti u pred stupac:

Županija
Istarska
Primorsko – goranska
Ličko – senjska
Zadarska
Šibensko – kninska
Splitsko – dalmatinska
Dubrovačko - neretvanska

- Definirati varijable:
    - Kratki naziv: ***Zene***, dugi naziv: ***Žene***
        - Unijeti vrijednosti (unositi brojke bez točke i bez razmaka):

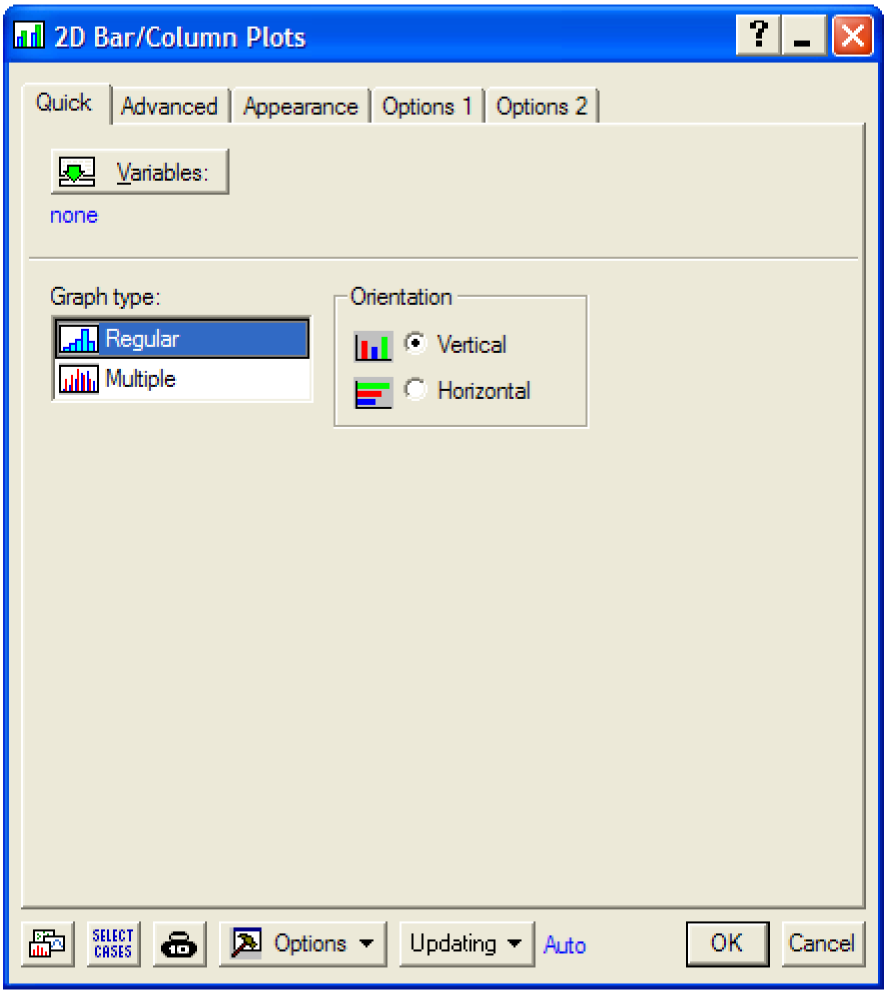
Žene
106375
158290
27182
82404
58225
237545
63492

- Kratki naziv: ***Muskarci***, dugi naziv: ***Muškarci***
    - Unijeti vrijednosti:

Muškarci
99969
147215
26495
79641
54666
226131
59378

a) Prikaz jednostavnih stupaca za žene

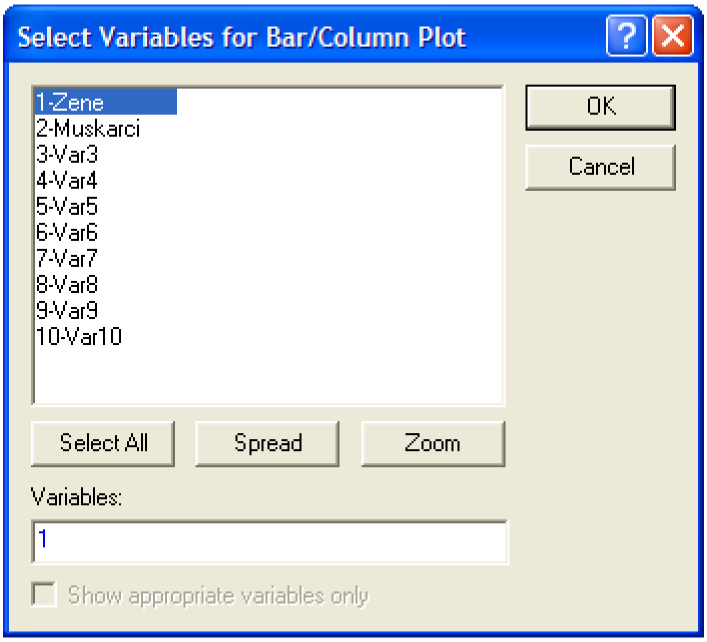
- **GRAPHS / 2D GRAPHS / BAR/COLUMN PLOTS**
  - **QUICK**
    - **Graph type:** ***Regular***
    - **Orientation:** ***Vertical***



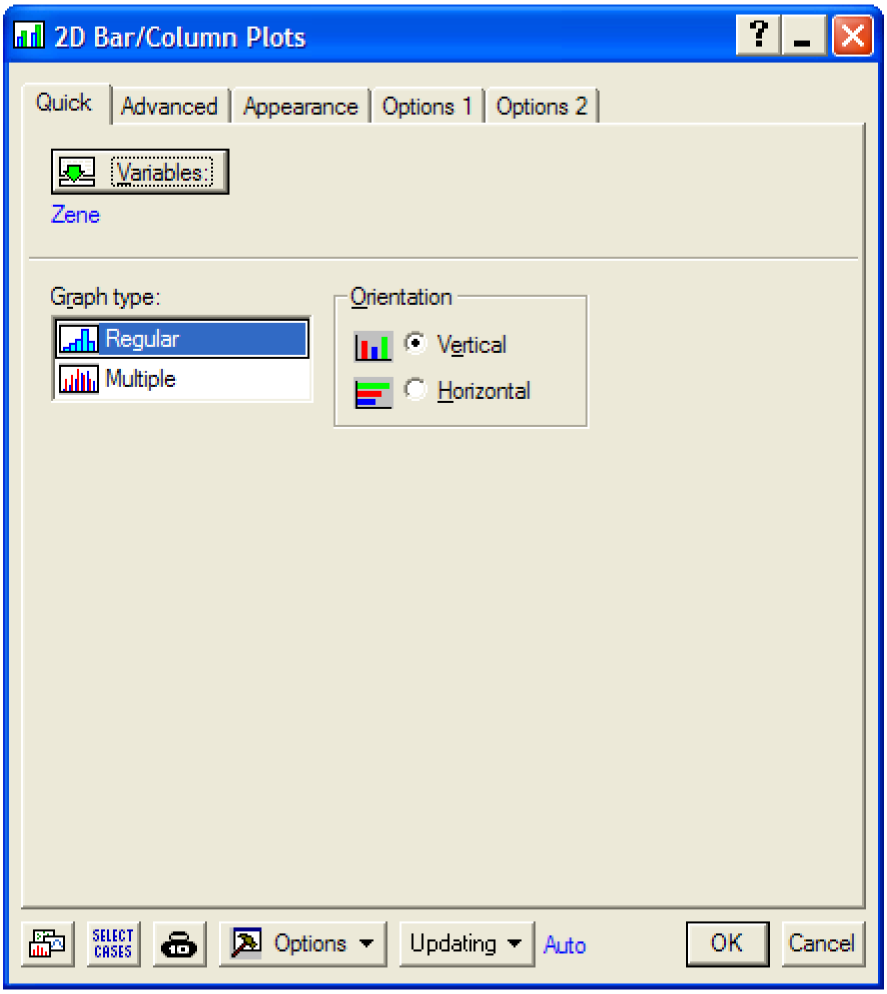
Varijablu koja se prikazuje grafički potrebno je označiti:

- Izabere se tipka: ***Variables***

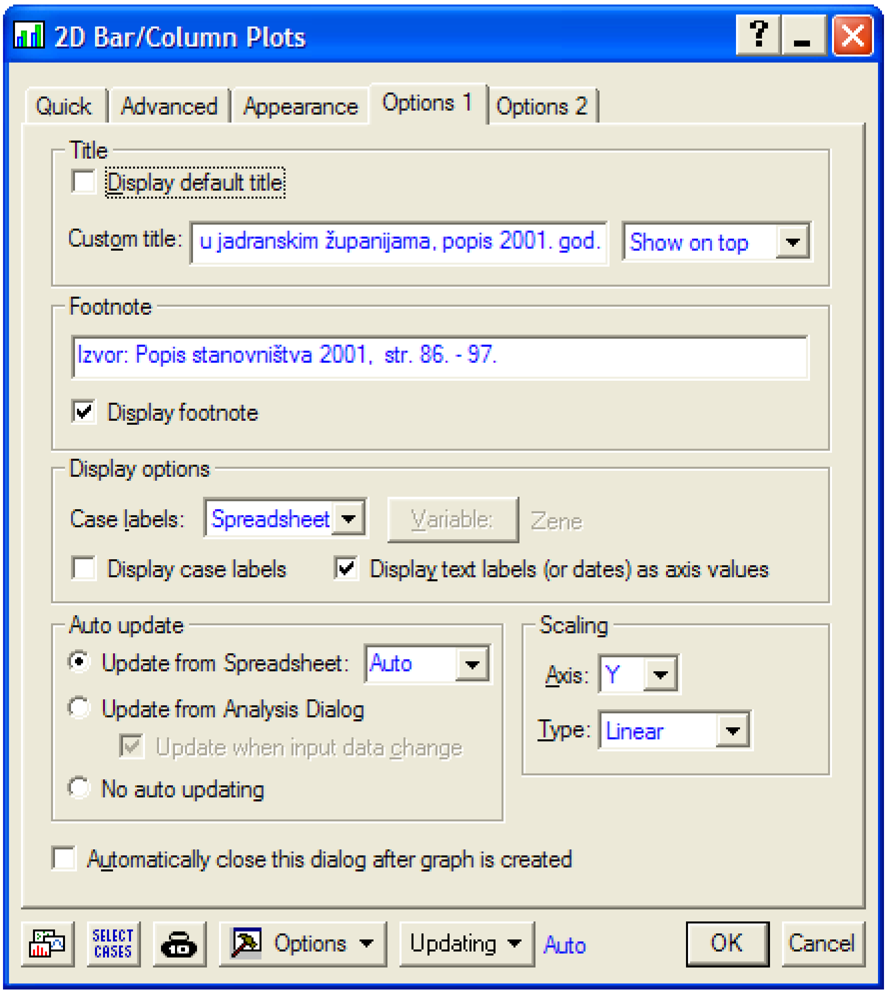
Otvara se prozor za izbor varijable te se izabere varijabla Zene.



- **Variables:** ***1 – Zene***


2D Bar/Column Plots
Quick
Advanced
Appearance
Options 1
Options 2
Variables:
Zene
Graph type:
Regular
Multiple
Orientation
Vertical
Horizontal
SELECT CASES
Options
Updating
Auto
OK
Cancel

- **OPTIONS 1**

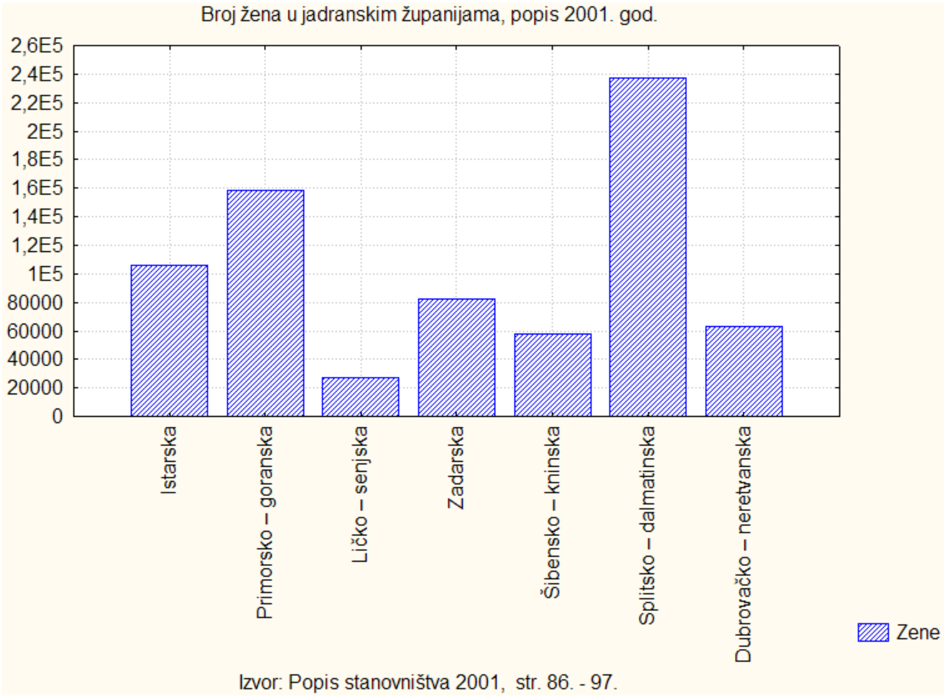


- **Case labels:** ***Case names***
- **Display Default Title – ukloniti oznaku**
- **Custom Title:**

  Broj žena u jadranskim županijama, popis 2001. god.

  - **Show on Top**
- **Footnote:** ***Izvor: Popis stanovništva 2001, str. 86. – 97.***

Dobije se grafikon:



Ukoliko se želi promijeniti nešto na grafikonu, potrebno je prvo selektirati željeni dio klikom miša. Nakon toga se izabere:

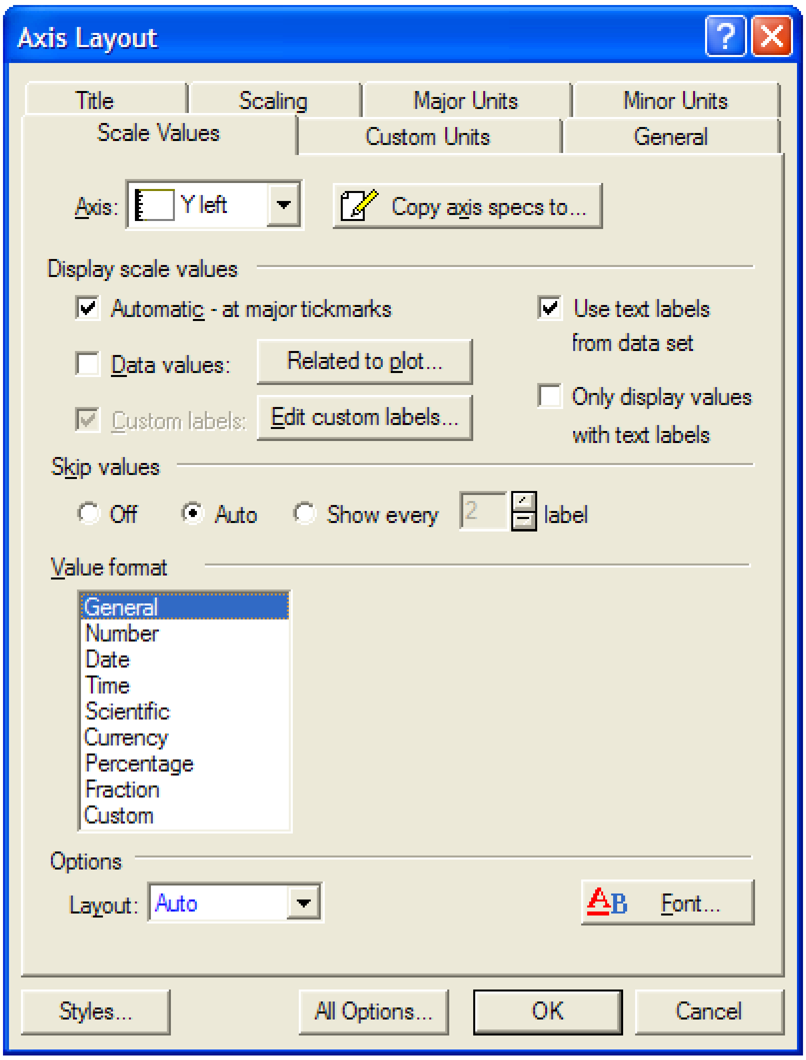
- **FORMAT / SELECTION**

ili

- **2 puta kliknuti**

Promjena formata ispisa vrijednosti na ordinati:

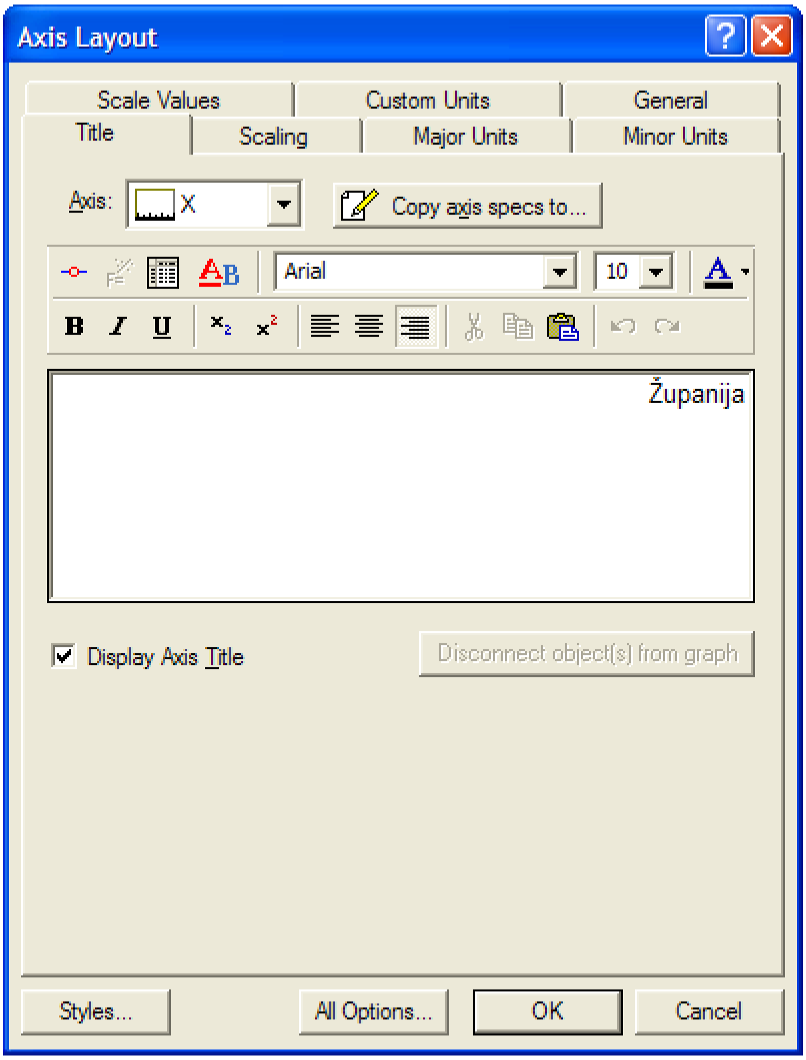
- **Selektirati os y**
- **FORMAT / SELECTION ili 2 puta kliknuti**¸



- **SCALE VALUES**
  - **Automatic – at major tickmarks**
  - **Value Format:** ***General***

Na apscisu staviti naziv *Županija* :

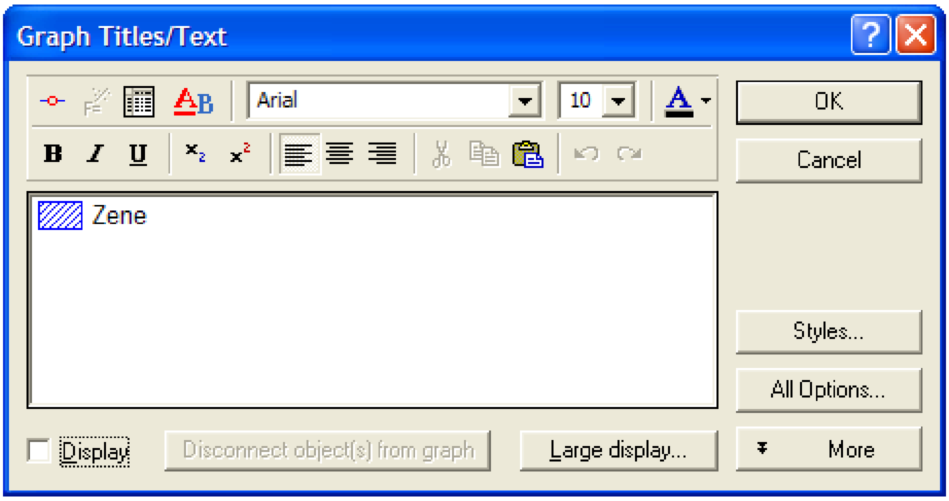
- **Selektirati os x**
- **FORMAT / SELECTION ili 2 puta kliknuti**
  - **TITLE:**



- ***Županija***
- **Poravnati na desno**

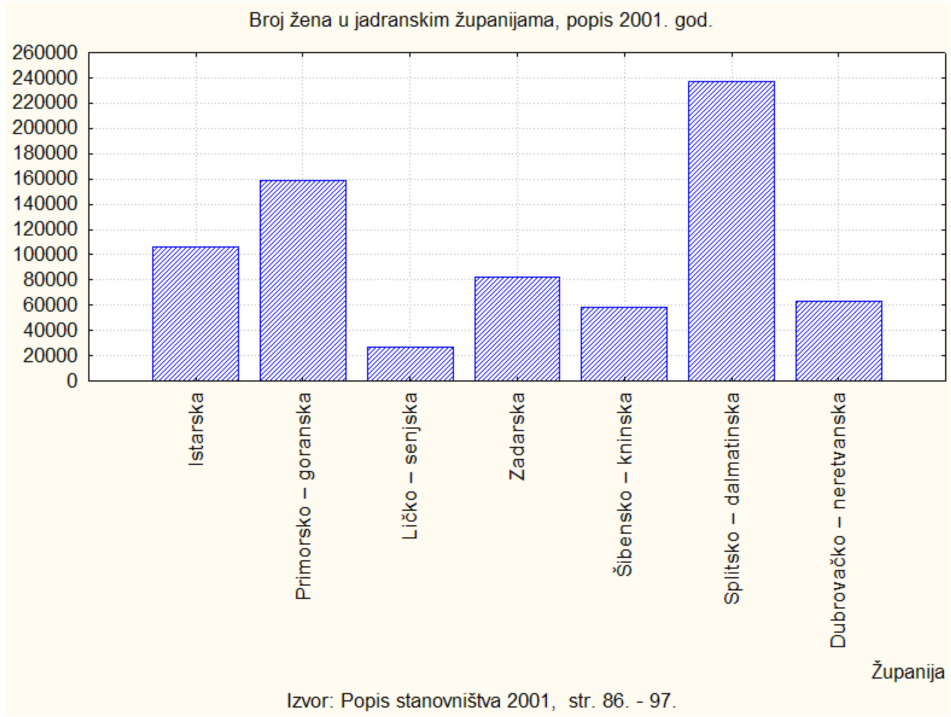
Ukloniti legendu:

- **Selektirati legendu**
- **FORMAT / SELECTION ili 2 puta kliknuti**



- **DISPLAY – ukloniti oznaku**

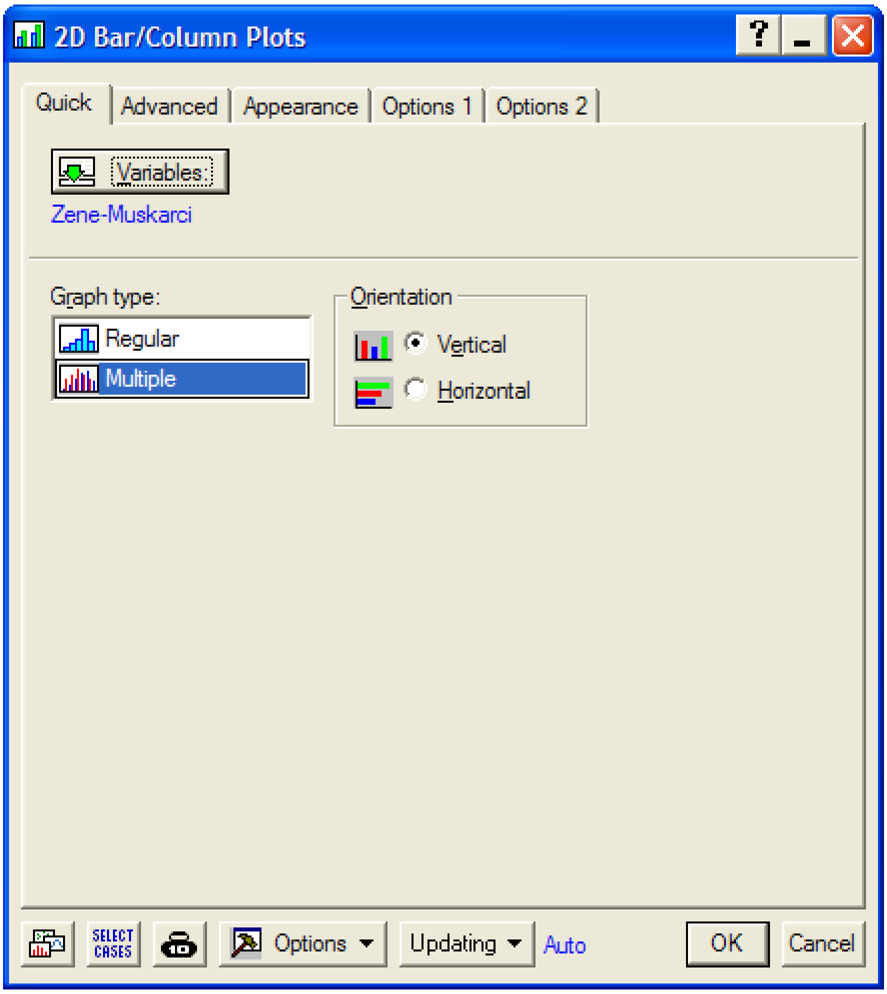
Graf sada izgleda:


Broj žena u jadranskim županijama, popis 2001. god.
260000
240000
220000
200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
Istarska
Primorsko – goranska
Ličko – senjska
Zadarska
Šibensko – kninska
Splitsko – dalmatinska
Dubrovačko – neretvanska
Županija
Izvor: Popis stanovništva 2001, str. 86. - 97.

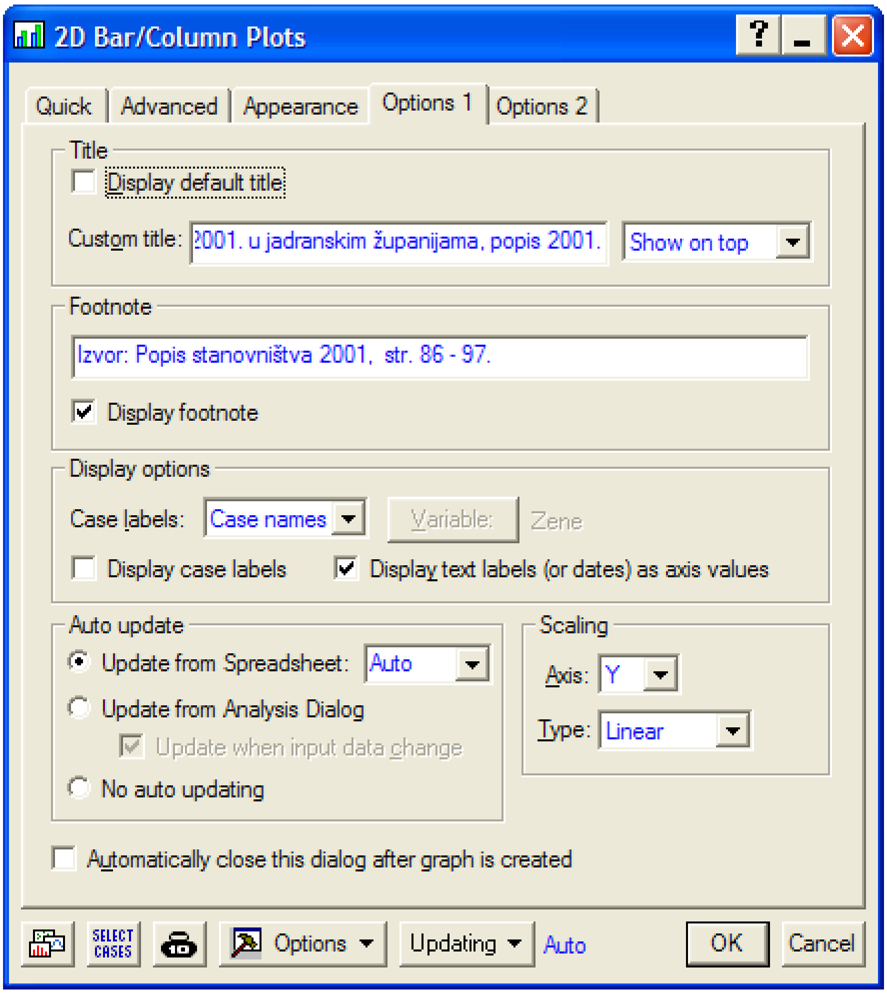
b) Prikaz dvostrukih stupaca

- **GRAPHS / 2D GRAPHS / BAR/COLUMN PLOTS**
  - **QUICK**
    - **Graph type:** ***Multiple***
    - **Orientation:** ***Vertical***
    - **Variables**

1-Zene
2-Muskarci



➢ **OPTIONS 1**

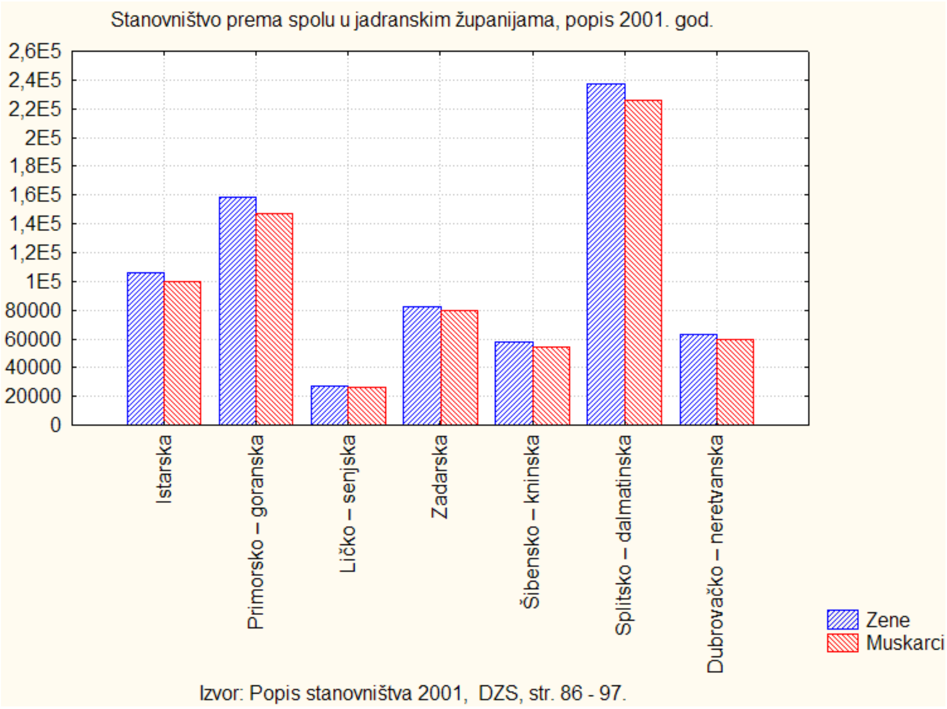


- **Case labels:** ***Case Names***
- **Display Default Title – ukloniti oznaku**
- **Custom Title:**

  Stanovništvo prema spolu u jadranskim županijama, popis 2001. god.

  - **Show on Top**
- **Footnote:** ***Izvor: Popis stanovništva 2001, DZS, str. 86. – 97.***

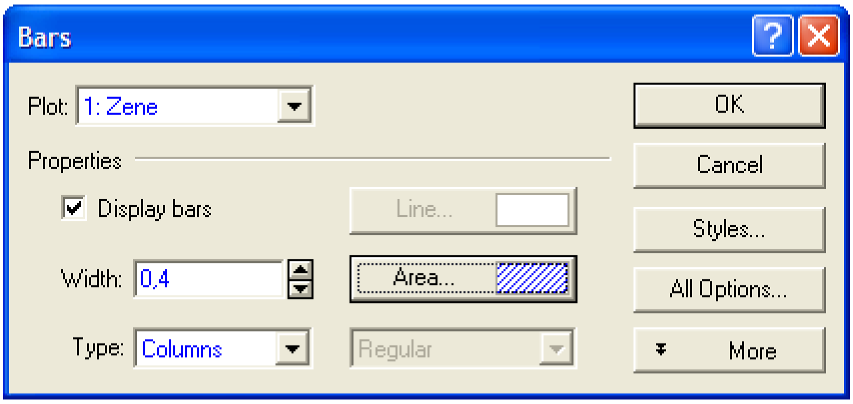
Dobije se grafikon:



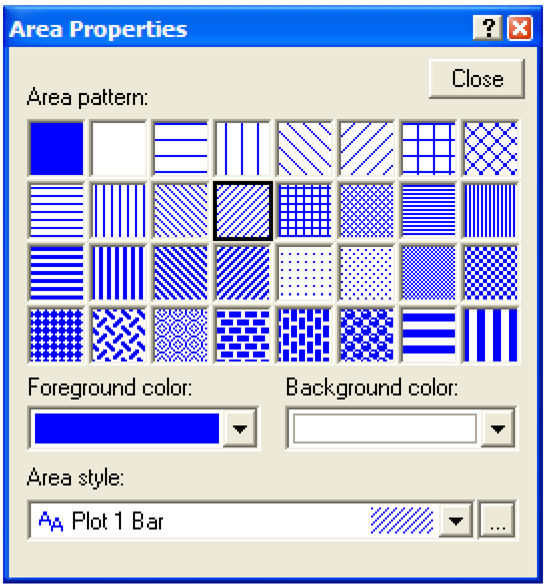
Može se po želji promijeniti izgled nekog stupca.

- **Selektirati "žene" (kliknuti na bilo koji stupac koji predstavlja "žene" – selektirat će se stupci za "žene" u svim županijama)**
- **FORMAT / SELECTION ili 2 puta kliknuti**

Promijeniti po želji:



- **Width**
- **Area**

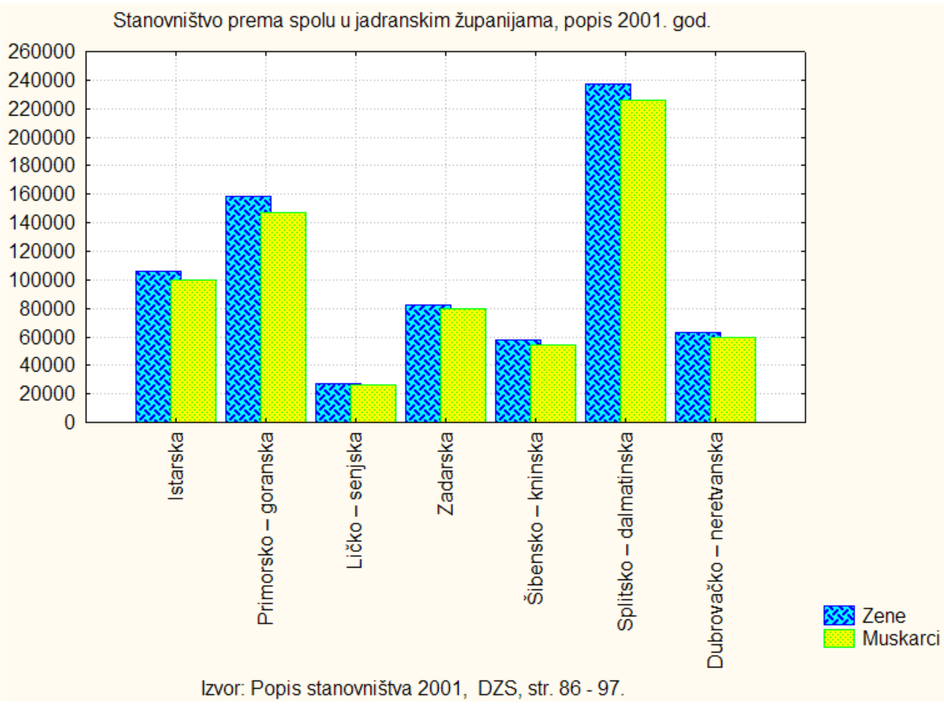


- **Area pattern**
- **Foreground color**
- **Background color**
- **Area style**

- **Type**

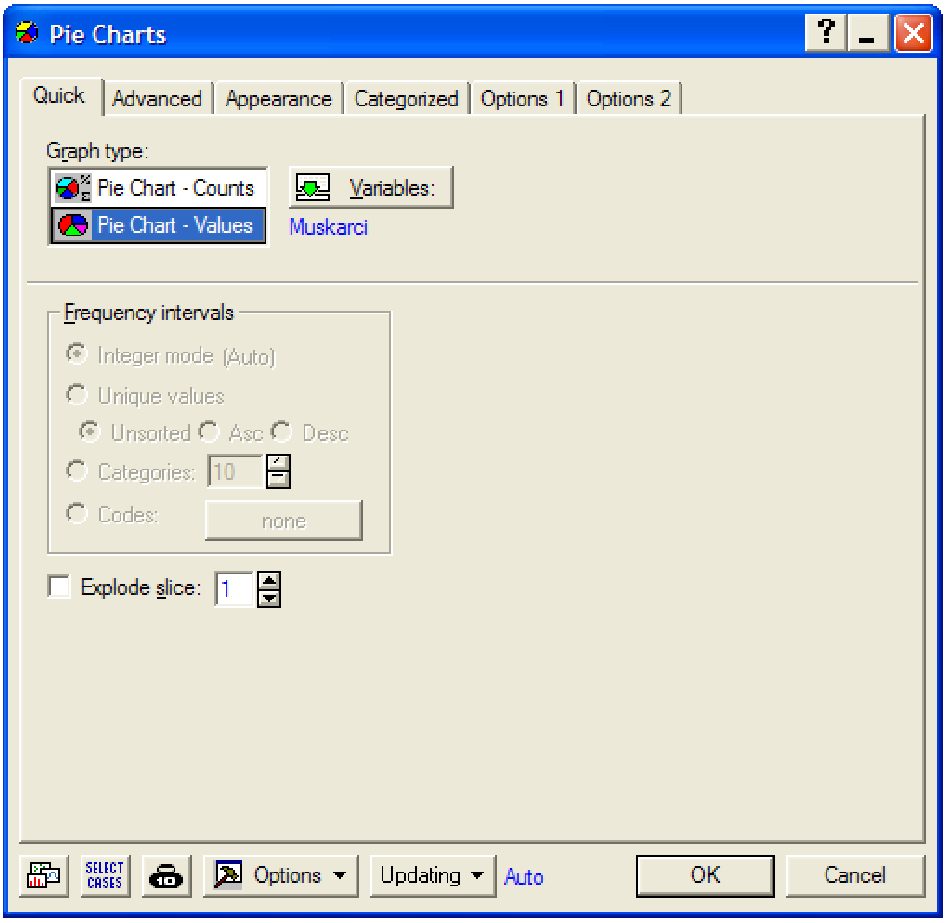
Automatski se promijeni legenda.

Isto se može napraviti za “muškarce”. Sljedeći grafikon je primjer s promijenjenim izgledom stupaca i promijenjenim formatom ispisa vrijednosti na ordinati.



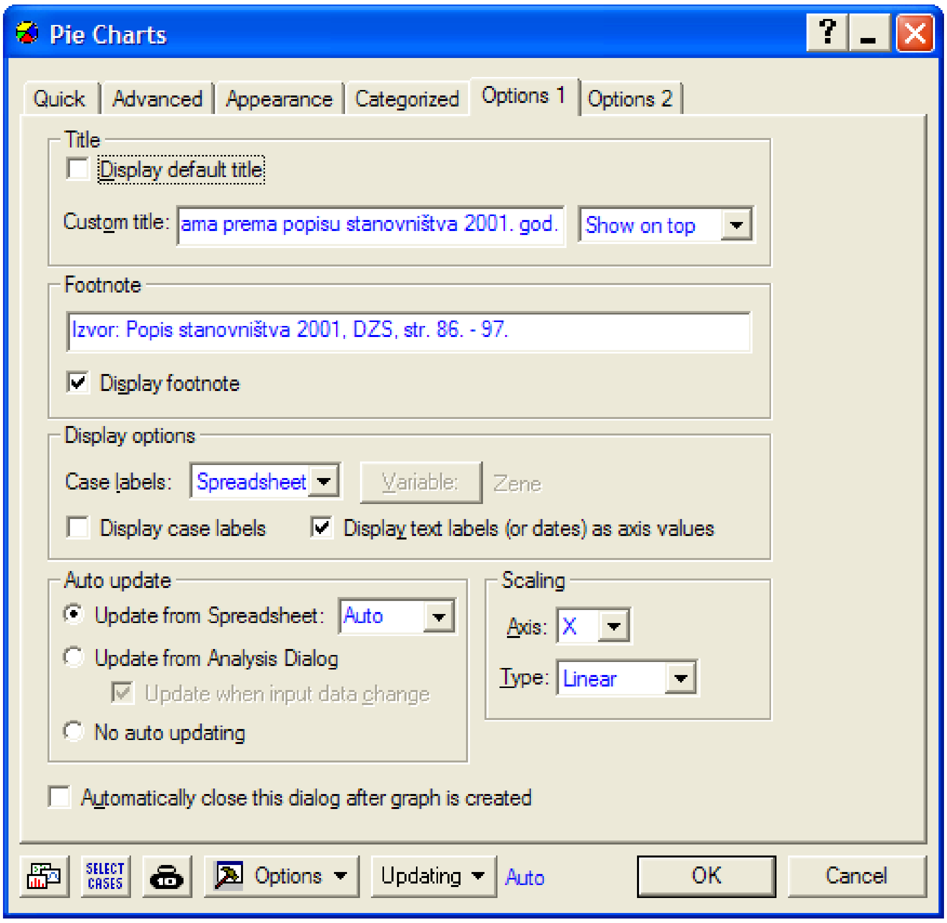
c) Strukturnim krugom treba prikazati raspodjelu muškaraca po županijama

- **GRAPHS / 2D GRAPHS / PIE CHARTS**
  - **QUICK**



  - **Graph type:** ***Pie Chart – Values***
  - **Variables:** ***2 – Muskarci***

- **OPTIONS 1**

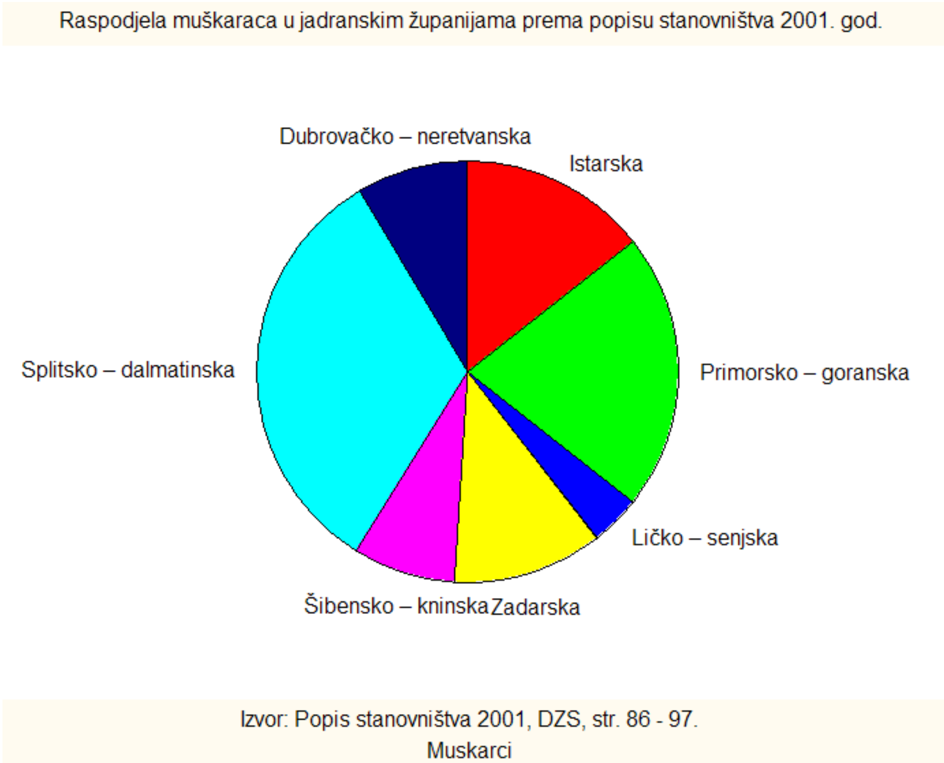


- **Case labels: *Case Names***
- **Display Default Title – ukloniti oznaku**
- **Custom Title:**

  Raspodjela muškaraca u jadranskim županijama prema popisu stanovništva 2001. god.

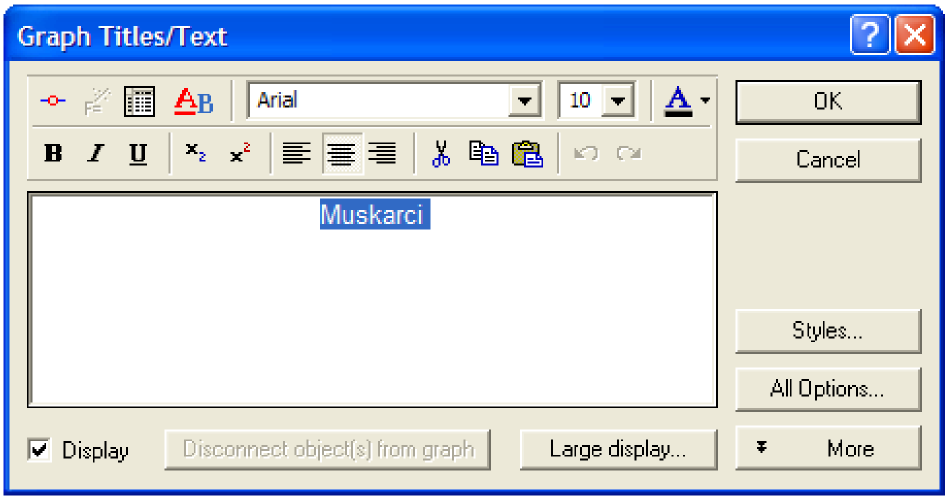
  - **Show on Top**
- **Footnote: *Izvor: Popis stanovništva 2001, DZS, str. 86. – 97.***

Dobije se grafikon:



Ukloniti natpis “Muskarci” na grafikonu:

- **Selektirati natpis “Muskarci”**
- **FORMAT / SELECTION ili 2 puta kliknuti¸**



- **DISPLAY – ukloniti oznaku**

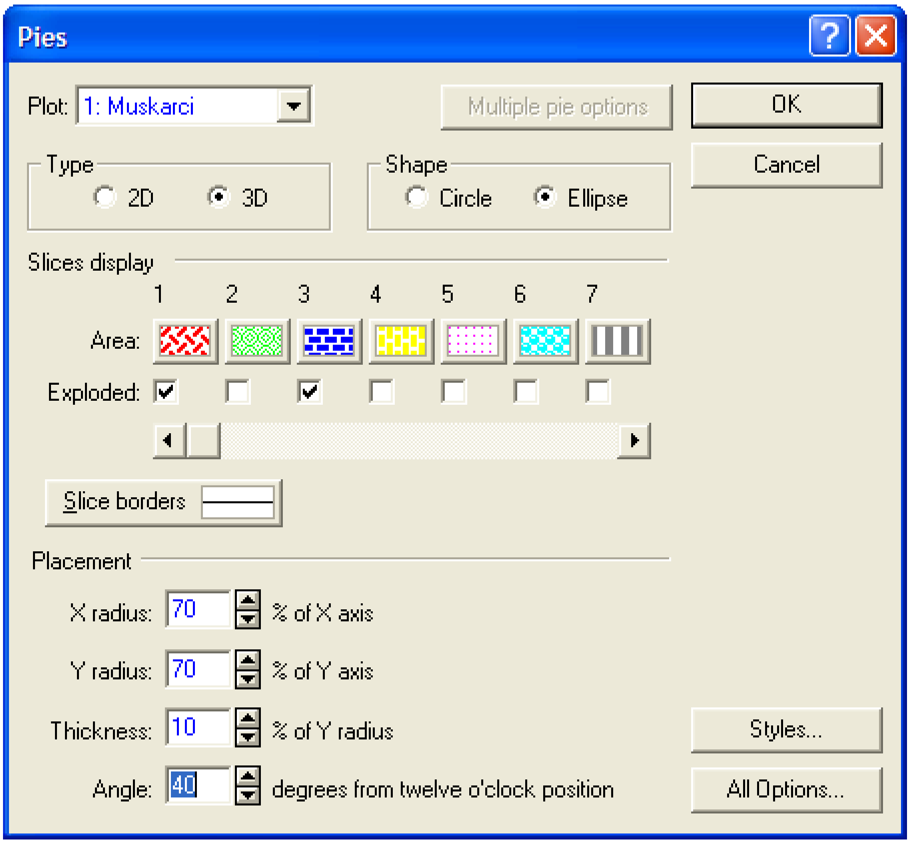
Za svaki kružni isječak može se napisati naziv županije, broj muškaraca u županiji te udio u ukupnom broju muškaraca jadranskih županija

- **Selektirati nazive županija**

- **FORMAT / SELECTION ili 2 x kliknuti**
  - **PROPERTIES**
    - **Text labels (prikazuje se naziv pojedinog kružnog isječka "županije")**
    - **Counts (vrijednost varijable "broj muškaraca")**
      - **Counts format**
        - **Value Format: General**
    - **Percentage (udio u ukupnom zbroju varijable)**
      - **Percentage format**
        - **Decimal Places: 2 (Broj prikazanih decimala)**

Može se mijenjati izgled pojedinog kružnog isječka ili cijelog strukturnog kruga.

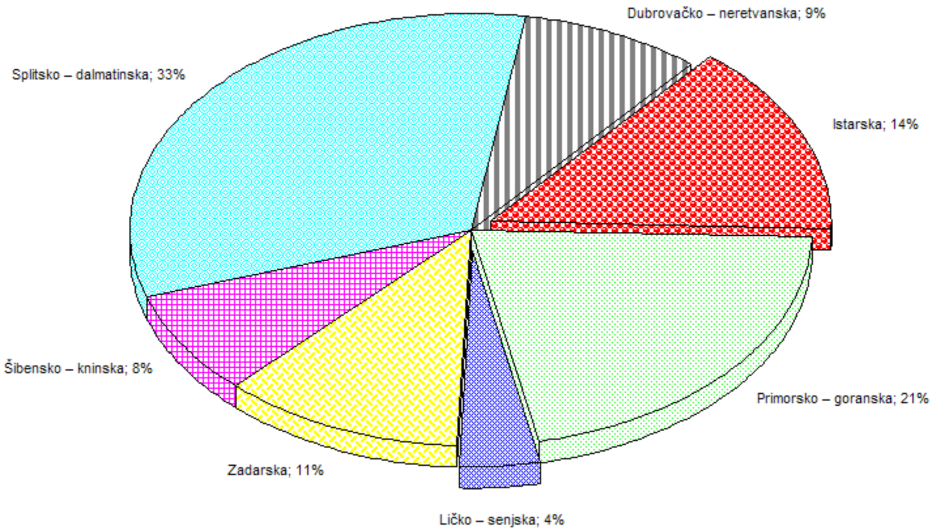
- **Selektirati strukturni krug**



- **FORMAT / SELECTION ili 2 puta kliknuti**
  - **SLICES DISPLAY (izgled pojedinog kružnog isječka)**
    - **Area**
    - **Exploded**
  - **SHAPE (oblik)**
  - **TYPE (dodavanje treće dimenzije)**
  - **PLACEMENT (smještaj grafa)**

Grafikon s promijenjenim izgledom:

Raspodjela muškaraca u jadranskim županijama prema popisu stanovništva 2001. god.



Izvor: Popis stanovništva 2001, DZS, str. 86 - 97.

3. Numerički nizovi

Primjer 3.1.

Zaposleni u ustanovi "A" prema pripadajućem broju dana godišnjeg odmora, na dan 31.12. 2008.

Dani godišnjeg odmora	*Broj zaposlenih*
X_i	f_i
25	7
26	12
27	14
28	18
29	13
30	11
Ukupno	**75**

Izvor: Kadrovska služba (simulirani podaci)

Zadatak je:

a) izračunati srednje vrijednosti

b) odrediti ukupan broj dana godišnjih odmora svih zaposlenih

c) izračunati mjere disperzije

d) odrediti kvartile

e) izračunati najmanju i najveću vrijednost

f) odrediti asimetriju

g) odrediti zaobljenost.

Prisutne su dvije varijable: varijabla "*Dani godišnjeg odmora*" koja predstavlja numeričko obilježje i označava se s $\mathbf{X_i}$ i varijabla "*Broj zaposlenih*" koja predstavlja frekvenciju odnosno ponder što se označava s $\mathbf{f_i}$.

Kreira se novi dokument:

➢ U **zaglavlje** dokumenta upisati:

Zaposleni u ustanovi "A" prema pripadajućem broju dana godišnjeg odmora, na dan 31.12. 2008. godine.

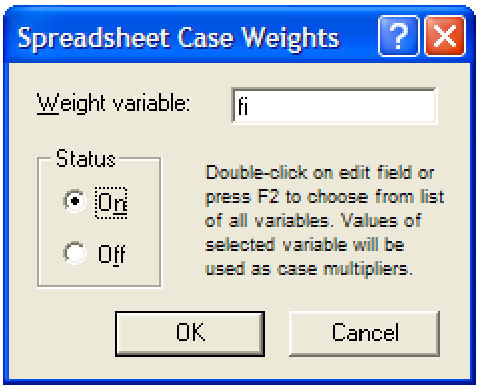
- Definirati varijable:
  - Kratki naziv: ***Xi***, dugi naziv: ***Dani godišnjeg odmora***
  - Kratki naziv: ***fi***, dugi naziv: ***Broj zaposlenih***

Nakon toga unesu se vrijednosti.

Ako postoji stupac s frekvencijama potrebno je prije početka računanja ponderirati. Ponderirati se može na sljedeći način:

- ***TOOLS / WEIGHT…***

Otvara se prozor za ponderiranje:

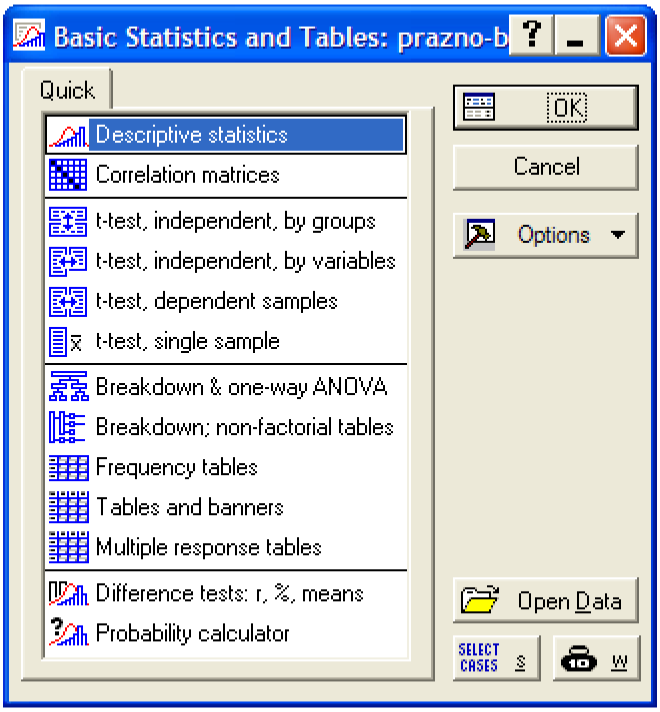


- ***WEIGHT VARIABLE fi***
- ***STATUS On***.

Nakon uključivanja ponderiranja pojavljuje se obavijest da je izvršeno ponderiranje, a u statusnoj traci pojavi se natpis: ***Weight: ON*** što znači da je ponderiranje uključeno. Kada je ponderiranje isključeno natpis je ***Weight OFF***.

Funkcije deskriptivne statistike moguće je pokrenuti putem menija:

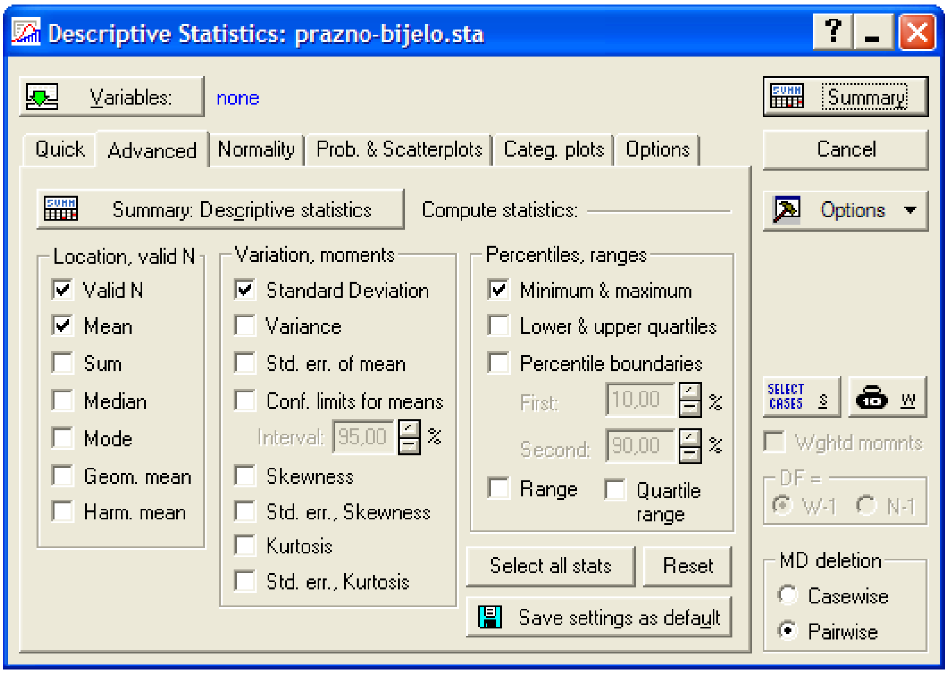
- **STATISTICS / BASIC STATISTIC/TABLES**



U rubrici QUICK zabere se:

- **DESCRIPTIVE STATISTICS**
  - **ADVANCED**

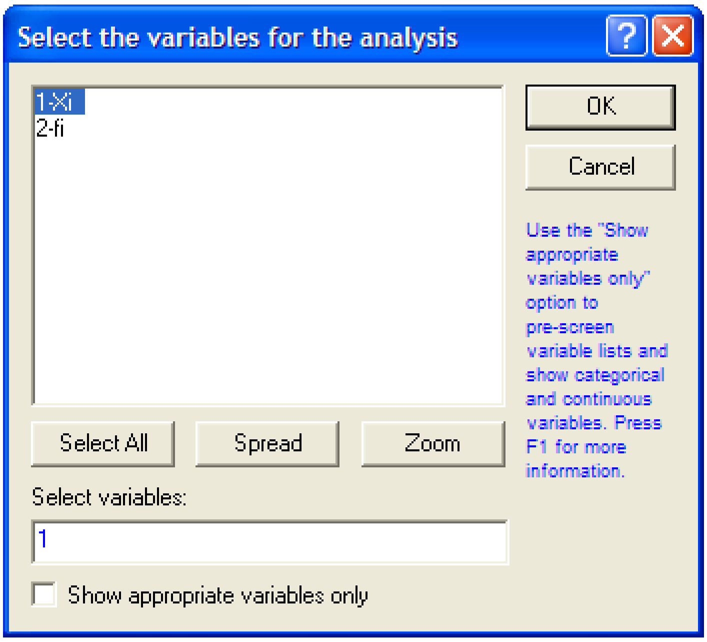
čime se otvara prozor s mjerama deskriptivne statistike.



Prilikom računanja potrebno je označiti na koju se varijablu odnose izračuni:

- Izabere se tipka: ***Variable***

Otvara se prozor za izbor varijable te se izabere varijabla koja predstavlja numeričko obilježje: Xi.



Sada je moguće pristupiti izračunu mjera deskriptivne statistike.

Za kontrolu unosa podataka mogu se rabiti funkcije:

- ***Valid N*** – **zbroj frekvencija**
- ***Sum*** – **zbroj vrijednosti numeričkog obilježja**

Mjere centralne tendencije koje su ponuđene za izračun su:

- ***Mean*** – **aritmetička sredina**
- ***Median*** – **medijan**
- ***Mode*** – **mod**
- ***Geom. mean*** – **geometrijska sredina**
- ***Ham. mean*** – **harmonijska sredina**

Moguće je izračunati i neke potpune mjere raspršenosti:

- ***Standard Deviation*** – **standardna devijacija uzorka** – prilikom izračuna polazi se od pretpostavke da se radi o uzorku, a ne o osnovnom skupu. Vrijednost standardne devijacije i mjera koje iz toga proizlaze, za osnovni skup nešto se razlikuju od vrijednosti za uzorak.
- ***Variance*** – **varijanca uzorka –** varijanca osnovnog skupa razlikuje se od varijance uzorka.

Računati se mogu i neke nepotpune mjere raspršenosti, kao i mjere koje su potrebne za njihovo računanje:

- ***Range*** **– raspon varijacije**
- ***Quartile range*** **– interkvartil**

- ***Minimum & maximum*** **– najmanja i najveća vrijednost**
- ***Lower & upper quartiles*** **– donji i gornji kvartil**

Ponuđeno je određivanje asimetrije i zaobljenosti:

- ***Skewness*** **– koeficijent asimetrije uzorka**
- ***Kurtosis*** **– eksces uzorka** – Gleda se je li mjera manja ili veća od 0, a ne od 3.

Mogu se označiti funkcije koje se traže ili označiti:

- **Select all stats**

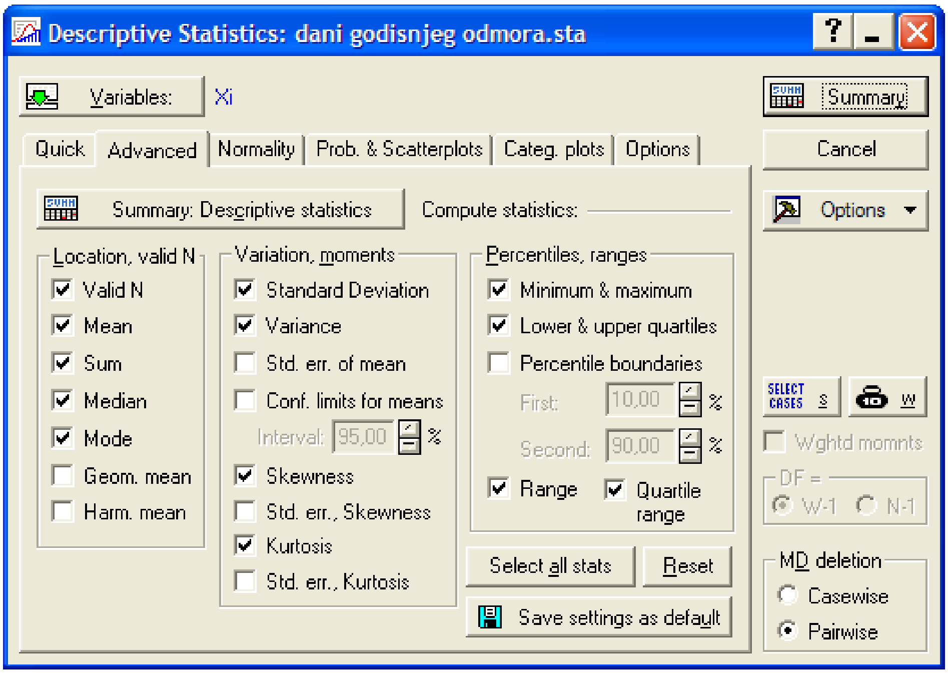
Čime se pokreće izračun svih funkcija.

U primjeru 3.1. za kontrolu unosa može se označiti:

a) Izračunajte srednje vrijednosti	➢ **MEAN, MEDIAN, MODE**
b) Odredite ukupan broj dana godišnjih odmora svih zaposlenih	➢ **SUM**
c) Izračunajte mjere disperzije	➢ **STANDARD DEVIATION, VARIANCE, RANGE, QUARTILE RANGE**
d) Odredite kvartile	➢ **LOWER & UPPER QUARTILES**
e) Izračunajte najmanju i najveću vrijednost	➢ **MINIMUM & MAXIMIM**
f) Odredite asimetriju	➢ **SKEWNESS**
g) Odredite zaobljenost	➢ **KURTOSIS**

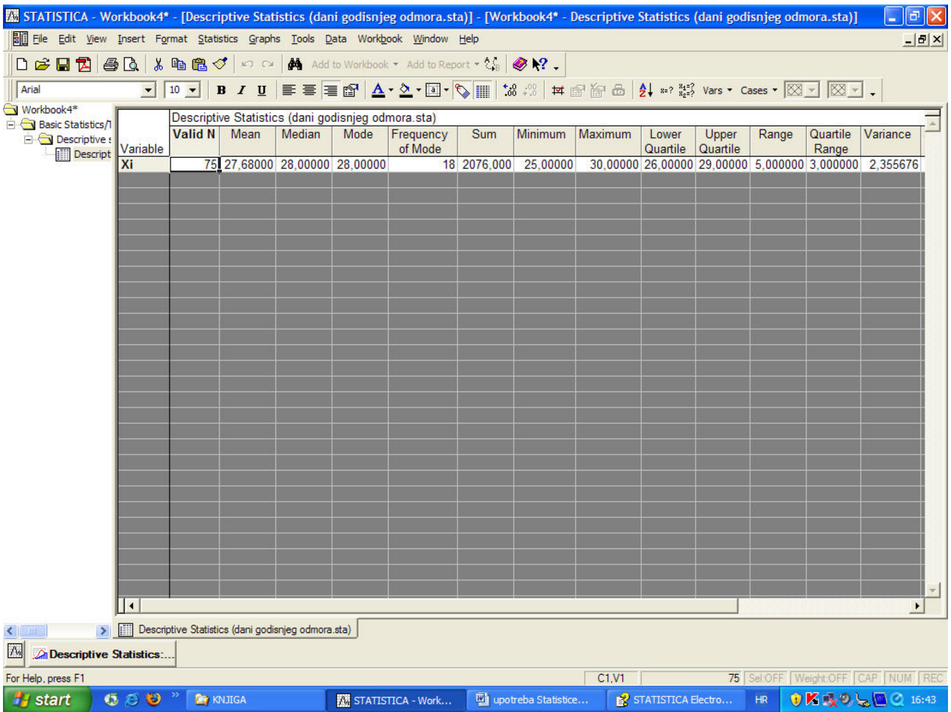
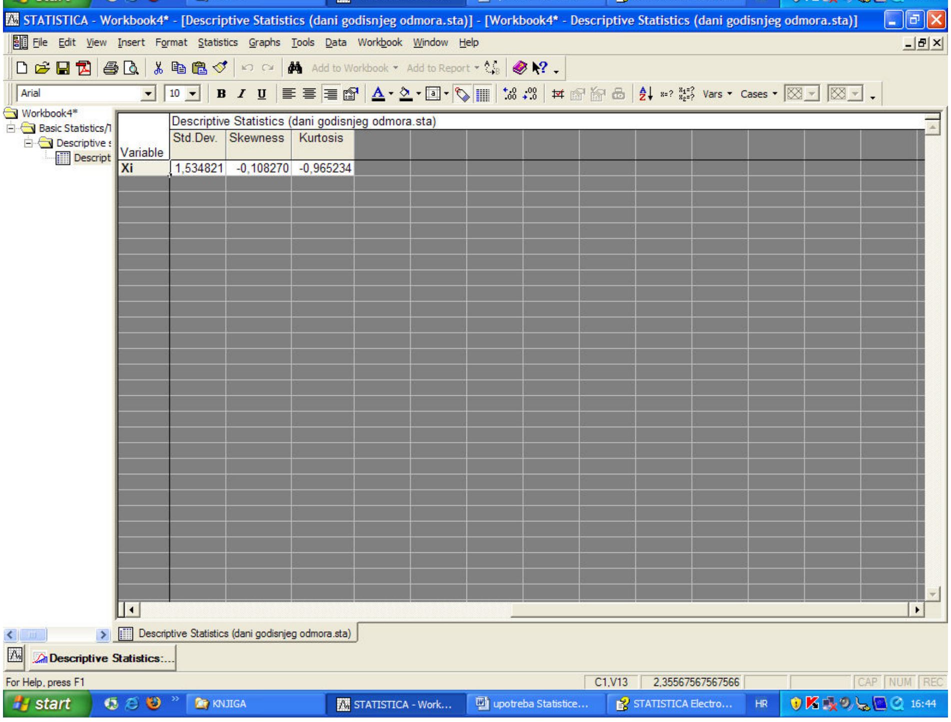
Osim toga za kontrolu unosa može se izabrati :

- **Valid N**



Pritiskom na tipku **SUMMARY** pokreće se računanje.

Otvara se prozor za prikaz rezultata:


STATISTICA - Workbook4* - [Descriptive Statistics (dani godisnjeg odmora.sta)] - [Workbook4* - Descriptive Statistics (dani godisnjeg odmora.sta)]
Descriptive Statistics (dani godisnjeg odmora.sta)
Variable | Valid N | Mean | Median | Mode | Frequency of Mode | Sum | Minimum | Maximum | Lower Quartile | Upper Quartile | Range | Quartile Range | Variance
Xi | 75 | 27,68000 | 28,00000 | 28,00000 | 18 | 2076,000 | 25,00000 | 30,00000 | 26,00000 | 29,00000 | 5,000000 | 3,000000 | 2,355676
For Help, press F1
C1,V1
75

STATISTICA - Workbook4* - [Descriptive Statistics (dani godisnjeg odmora.sta)] - [Workbook4* - Descriptive Statistics (dani godisnjeg odmora.sta)]
Descriptive Statistics (dani godisnjeg odmora.sta)
Variable | Std.Dev. | Skewness | Kurtosis
Xi | 1,534821 | -0,108270 | -0,965234
For Help, press F1
C1,V13
2,35567567567566

Provjera: “ukupan broj zaposlenih” (zbroj frekvencija) je 75 kao što je zadano (Valid N).

a) Izračun srednjih vrijednosti:

***8*= 27,68 dana**
M_e = 28 dana
M_o = 28 dana

Prosječan broj dana godišnjeg odmora zaposlenih u ustanovi A iznosi 27,68 dana.

Među zaposlenima u ustanovi A 50% ima 28 ili manje dana godišnjeg odmora, odnosno 50% zaposlenih ima 28 dana godišnjeg odmora ili više.

Najčešći broj dana godišnjeg odmora je 28, što je slučaj kod 18 radnika.

b) Određivanje ukupnog broja dana godišnjih odmora svih zaposlenih (total):

ΣX = 2076 dana

Ukupan broj dana godišnjeg odmora svih zaposlenih je 2076 dana.

c) Izračun mjera disperzije:

σ = 1,534821 dana*
σ^2 = 2,355676 dana*
R_x = 5 dana
I_Q = 3 dana

**** Pretpostavlja se da se radi o uzorku, a ne o cijelom osnovnom skupu. Standardna devijacija i mjere koje iz toga proizlaze nešto se razlikuju od stvarnih.***

Razlika između najdužeg i najkraćeg godišnjeg odmora je 5 dana. Raspon varijacija središnjih 50% jedinica je 3 dana. Prosječno odstupanje od prosjeka je 1,534821 dana (ovo je približna vrijednost).

d) Određivanje kvartila:

Q_1 = 26 dana
Q_3 = 29 dana

U 25% slučajeva zaposleni imaju 26 ili manje dana godišnjeg odmora, a 75% zaposlenih ima 26 ili više dana godišnjeg odmora.

Zaposleni u 75% slučajeva imaju 29 ili manje dana godišnjeg odmora, a u 25% slučajeva odmor traje 29 ili više dana.

e) Izračun najmanje i najveće vrijednosti:

X_{min} = 25 dana
X_{max} = 30 dana

Zaposleni u ustanovi A imaju najmanje 25, a najviše 30 dana godišnjeg odmora.

f) Odredite asimetriju

Skewness = ***α_3 = -0,108270***

Mjera asimetrije α_3 iznosi -0,108270, a to znači da je raspodjela blago negativno ili lijevostrano asimetrična.

g) Određivanje zaobljenosti:

Kurtosis = ***Eksces = κ = -0,965234***

Eksces je -0,9652342, a to znači da je raspodjela zaobljenija od normalne.

Primjer 3.2.

Proizvodnja stakla 2005. godine u tvornici “B” bila je po pogonima sljedeća: 128, 245, 333, 199 i 275 *t*. Zadatak je:

a) izračunati prosječnu proizvodnju po pogonu
b) izračunati ukupnu proizvodnju u tvornici “B”
c) odrediti srednje vrijednosti
d) izračunati mjere odstupanja od prosjeka
e) odrediti najmanju i najveću proizvodnju u pogonu te raspon varijacija
f) odrediti kvartile i interkvartil.

Sve vrijednosti proizvodnje imaju frekvenciju 1, tj. svaka vrijednost pojavljuje se samo jedanput te nije potrebno ponderirati vrijednosti. Iz tog razloga unosi se samo varijabla Xi – proizvodnja stakla u tonama, a nema varijable frekvencija.

S obzirom da je učestalost svih obilježja jednaka (svaka vrijednost pojavljuje se jedanput) ne postoji mod – najčešće obilježje. Također se ne računaju asimetrija i zaobljenost.

Rješenja:

a) izračun prosječne proizvodnje po pogonu:

8 = 236 t

b) izračun ukupne proizvodnje u tvornici “B”:

ΣX = 1180 t

c) određivanje srednjih vrijednosti:

8 = 236 t
$\boldsymbol{M_e = 245\ t}$
$\boldsymbol{M_o}$ ***= NEMA MODA***

d) izračun mjera odstupanja od prosjeka:

$\boldsymbol{\sigma^2 = 6011}$
$\boldsymbol{\sigma = 77{,}53064\ t}$

e) određivanje najmanje i najveće proizvodnje po pogonu i raspon varijacija:

$\boldsymbol{X_{min} = 128\ t}$
$\boldsymbol{X_{max} = 333\ t}$
$\boldsymbol{R_x = 205\ t}$

f) određivanje kvartila i interkvartila:
$\boldsymbol{Q_1 = 199\ t}$
$\boldsymbol{Q_3 = 275\ t}$
$\boldsymbol{I_Q = 76\ t}$

4. Regresijska i korelacijska analiza

4.1. Jednostavna linearna regresija

Primjer 4.1.

Brodarska kompanija "C" vrši isporuke na udaljenosti do 1000 nautičkih milja. Upravu zanima odnos između udaljenosti koje teret mora prijeći i vremena isporuke u danima. U tu svrhu slučajno je izabrano 10 isporuka izvršenih u ožujku 2006. godine. Udaljenosti u nautičkim miljama i vrijeme isporuke u danima prikazani su u sljedećoj tablici:

Isporuka	*Udaljenost (nautičke milje)*	*Vrijeme isporuke (dani)*
1	751	7
2	532	3
3	632	2
4	917	12
5	503	2
6	732	4
7	830	6
8	866	8
9	601	5
10	763	4

Izvor: Podaci brodarskog društva "C"

Zadatak je:

a) odrediti jednadžbu jednostavne linearne regresije
b) prikazati dijagram rasipanja
c) u dijagramu rasipanja prikazati pravac regresije
d) izvršiti interpolaciju
e) procijeniti vrijeme trajanja isporuke za udaljenost od 600 nautičkih milja.

a) Izračun srednjih vrijednosti:

- U **zaglavlje** dokumenta upisati:

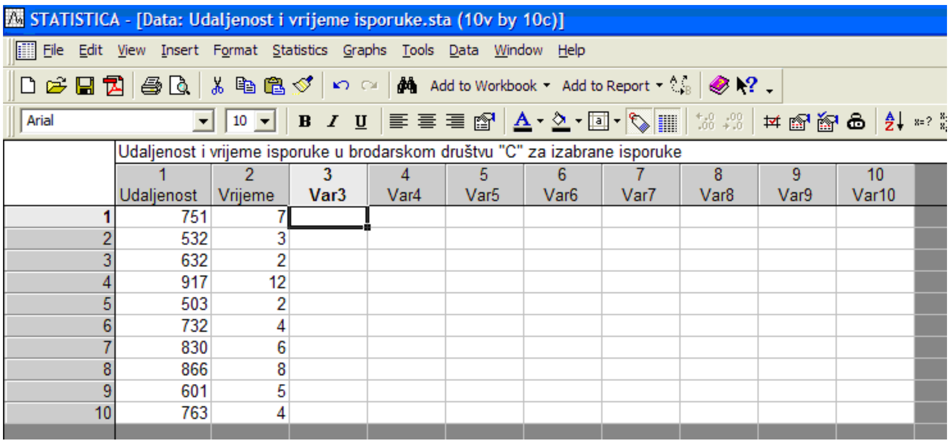
Udaljenosti i vrijeme isporuke u brodarskom društvu "C" za izabrane isporuke

- Definirati varijable:

  - Kratki naziv: $\boxed{\textit{Udaljenost}}$, dugi naziv: $\boxed{\textit{Udaljenost u nautičkim miljama}}$

  - Kratki naziv: $\boxed{\textit{Vrijeme}}$, dugi naziv: $\boxed{\textit{Vrijeme isporuke u danima}}$

- Unijeti vrijednosti



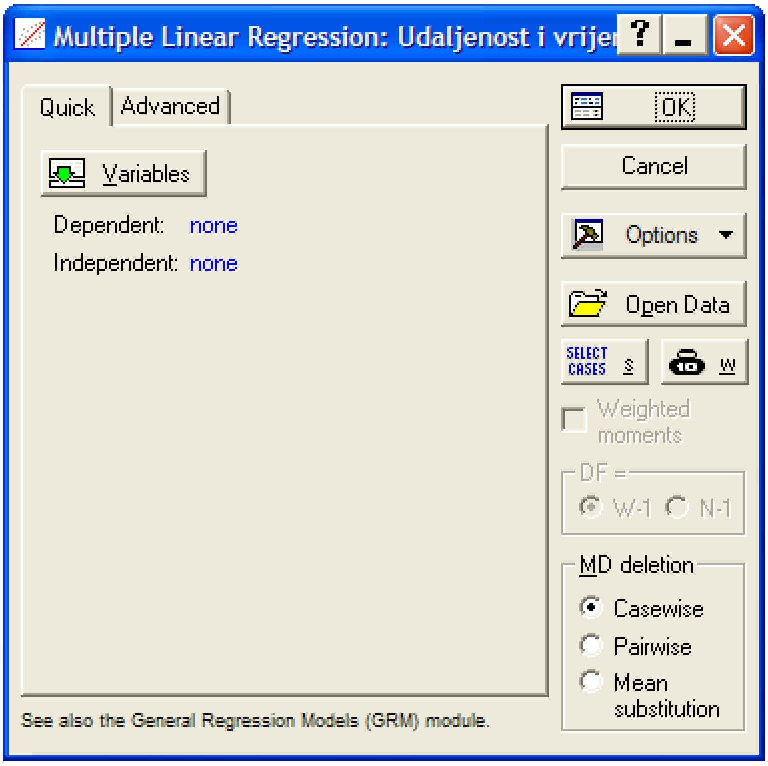
	1 Udaljenost	2 Vrijeme	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1	751	7								
2	532	3								
3	632	2								
4	917	12								
5	503	2								
6	732	4								
7	830	6								
8	866	8								
9	601	5								
10	763	4								

Ocjenjeni model jednostavne linearne regresije glasi:

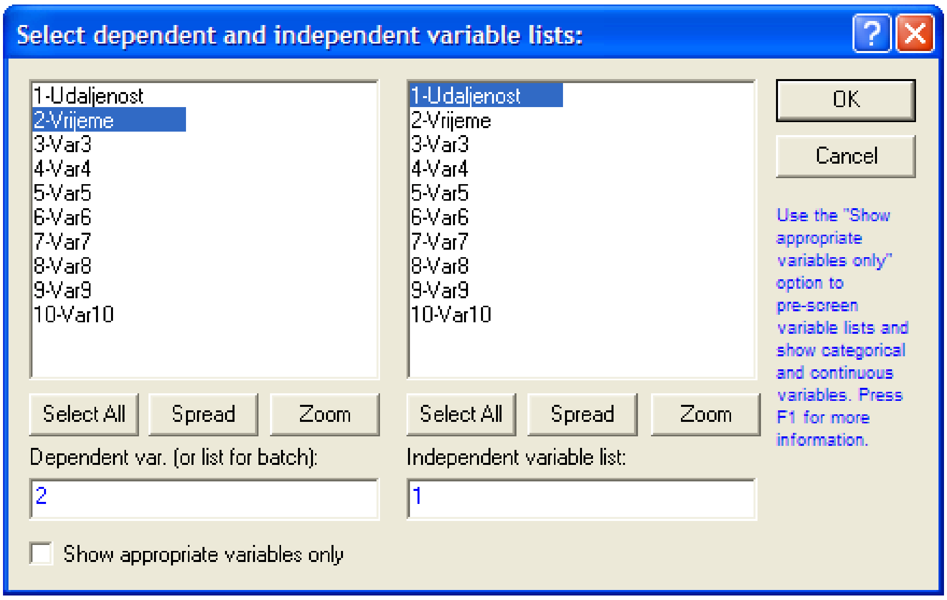
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

Parametri $\hat{\beta}_0$ i $\hat{\beta}_1$ računaju su pomoću procedure *Multiple Regression.* Nezavisna varijabla je "*Udaljenost*", a zavisna varijabla je "*Vrijeme*".

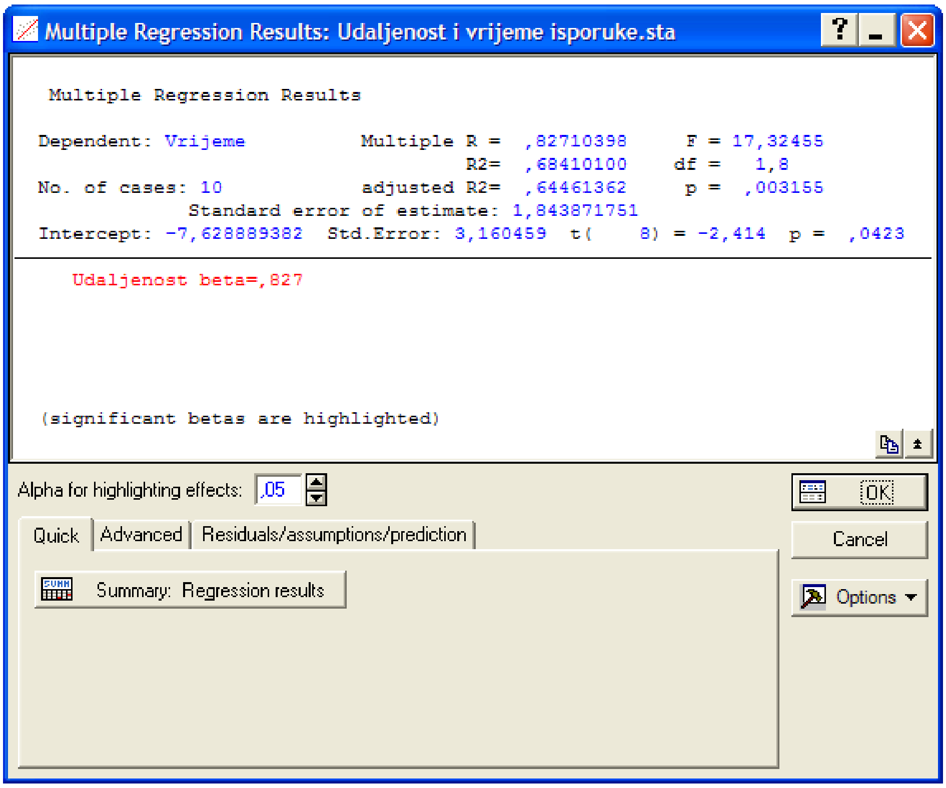
- **STATISTICS / MULTIPLE REGRESSION**
  - **QUICK**



- **Variables**
  - **Dependenet:** $\boxed{\textit{2-Vrijeme}}$ **(zavisna varijabla)**
  - **Independenet:** $\boxed{\textit{1-Udaljenost}}$ **(nezavisna varijabla)**

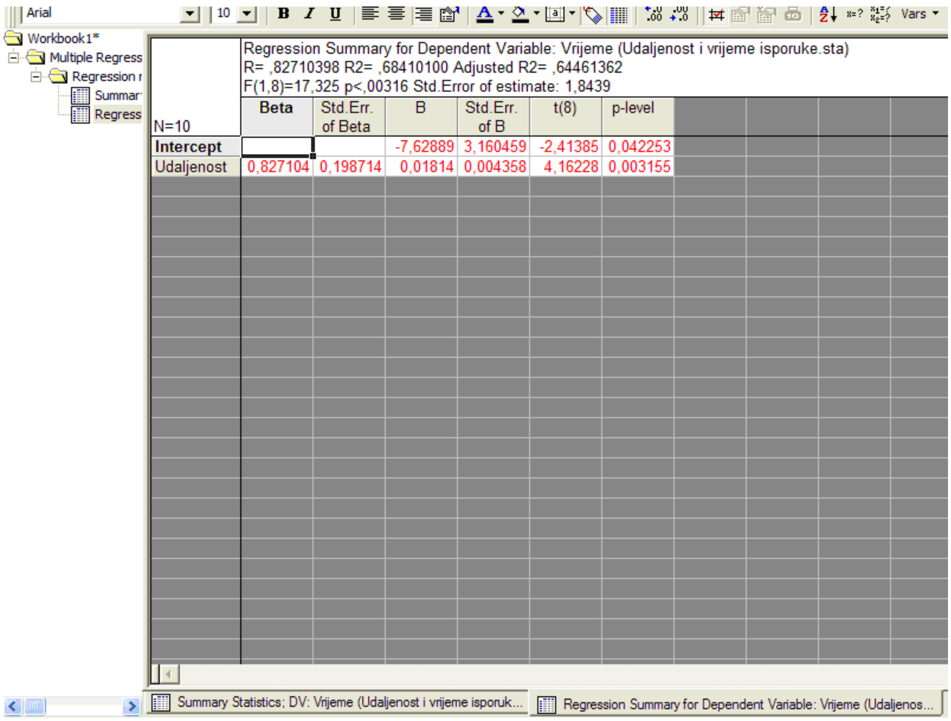


➢ **Pritiskom na tipku OK pokreće se računanje:**



Rezultati se ispisuju u prozor za rezultate, izborom:

- ➢ **QUICK**
  - ➢ **Summary: Regression Results**



Jednadžba regresije postoji samo ukoliko je:

- veza dovoljno jaka ($\bar{R}^2$ = **Adjusted R^2** $> 0{,}5$**)** i
- mala je vjerojatnost da su rezultati dobiveni slučajno **($p < 0{,}05$).**

	Regression Summary for Dependent Variable: Vrijeme (udaljenost R= ,82710398 R2= ,68410100 **Adjusted R2= ,64461362** F(1,8)=17,325 **p<,00316** Std.Error of estimate: 1,8439					
N=10	Beta	Std.Err. of Beta	B	Std.Err. of B	t(8)	p-level
Intercept			-7,62889	3,160459	-2,41385	0,042253
Udaljenost	0,827104	0,198714	0,01814	0,004358	4,16228	0,003155

U zadatku je:

$\bar{R}^2$ = 0,68410100 > **0,5**

P < 0,00316 < **0,05**

Nakon toga provjerava se reprezentativnost svakog parametra. Izračunati parametri uvrštavaju se u jednadžbu samo ako su dovoljno značajni. To je postignuto ako je **p-level < 0,05** za svaki parametar. Te retke *Statistica* dodatno označava crvenom bojom (odbacuje se nul-hipoteza da je parametar = 0).

N=10	Regression Summary for Dependent Variable: Vrijeme (Udaljenost i vrijeme R= ,82710398 R2= ,6841010C Adjusted R2= ,64461362 F(1,8)=17,325 p<,00316 Std.Error of estimate: 1,8439					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(8)	p-level
Intercept			β0 = -7,628889	3,160459	-2,41385	**0,042253<0,05**
Udaljenost	0,827104	0,198714	β1 = 0,01814	0,004358	4,16228	**0,003155<0,05**

Svi parametri su reprezentativni.

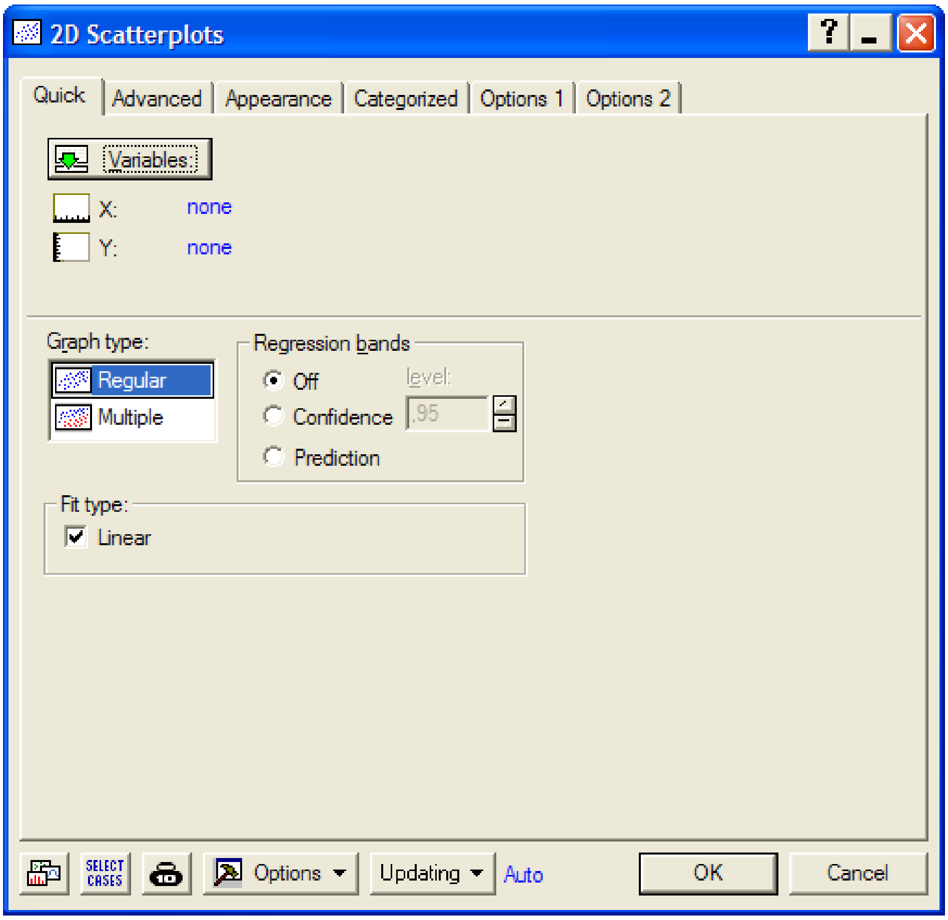
Jednadžba linearne regresije glasi:

$\underline{\hat{Y} = -7{,}62889 + 0{,}01814\ X}$

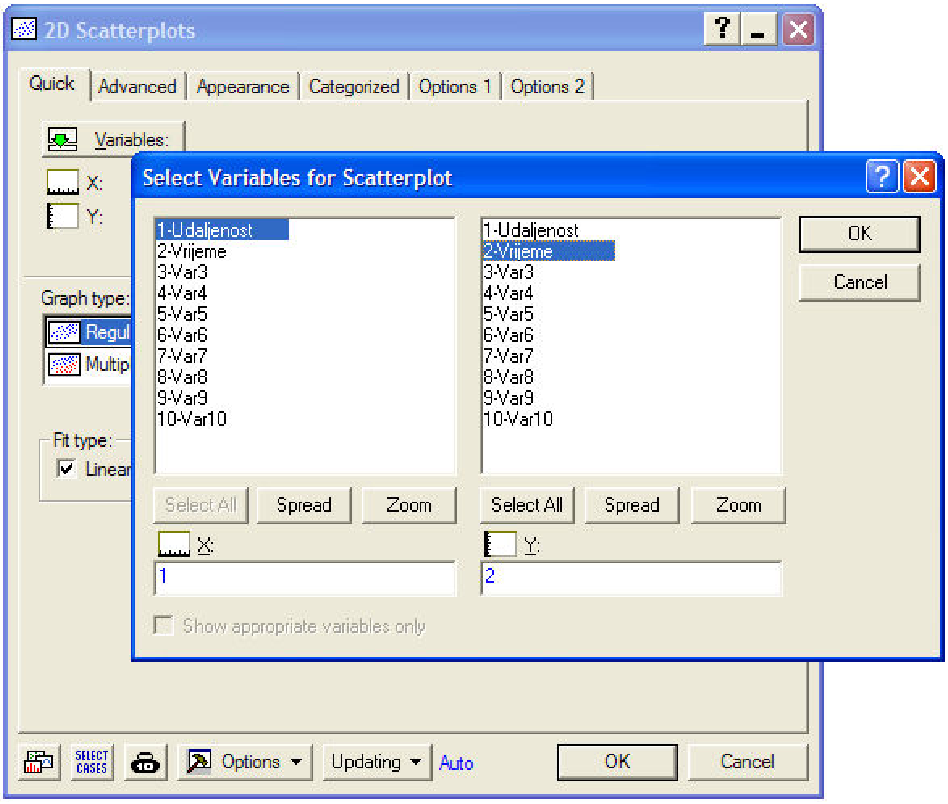
X je udaljenost u nautičkim miljama
Y je vrijeme isporuke u danima

b) Prikaz dijagrama rasipanja:

- **GRAPHS / SCATTERPLOTS**
  - **QUICK**



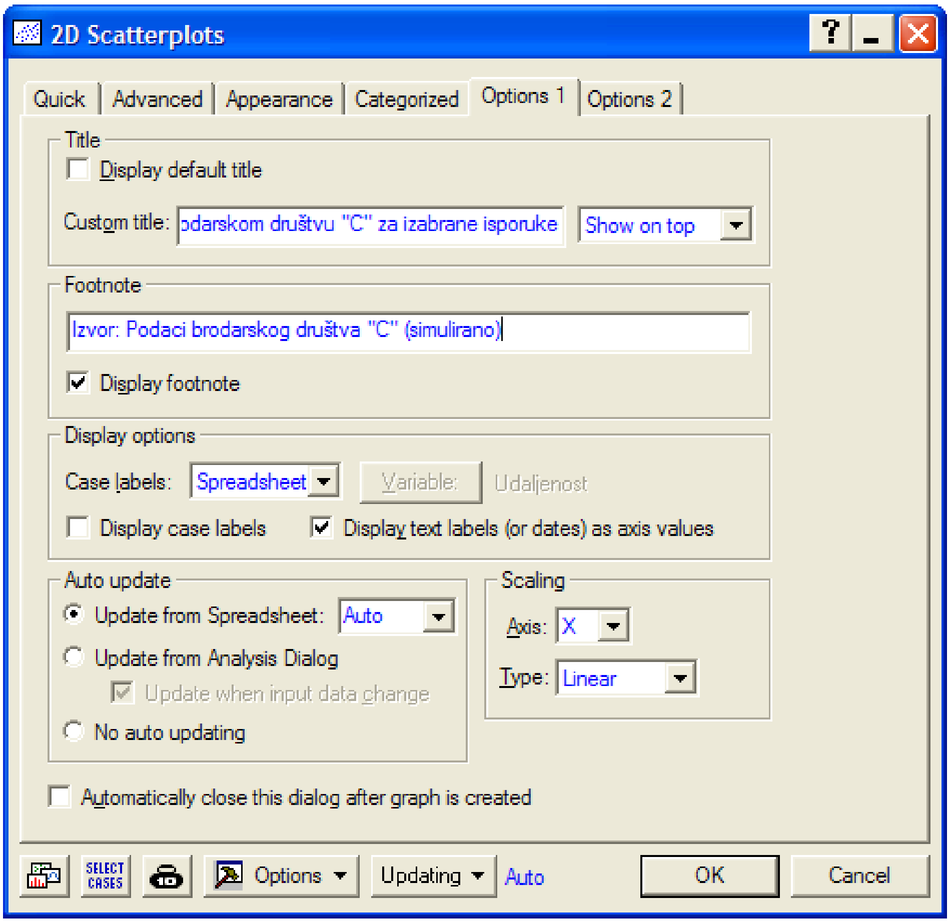
- **Graph type:** ***Regular***
- **Variables:**
  - **X:** ***1 - Udaljenost***
  - **Y:** ***2 – Vrijeme***
- **Fit type:** ***Linear***



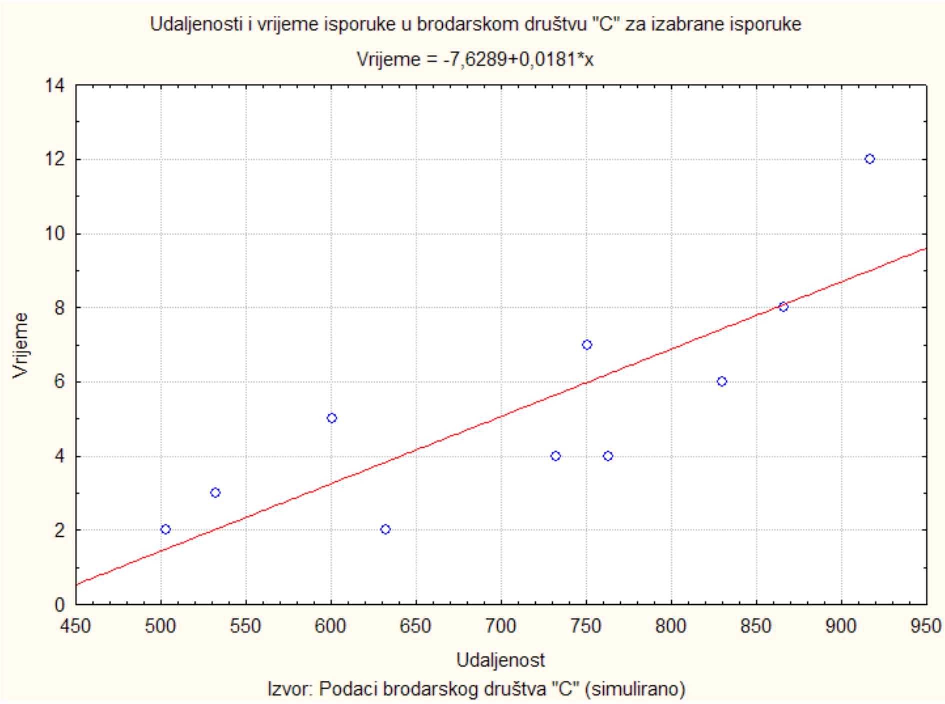
- **OPTIONS1**
  - **Display Default Title – ukloniti oznaku**
  - **Custom Title:**

    Udaljenosti i vrijeme isporuke u brodarskom društvu "C" za izabrane isporuke

    - **Show on Top**
  - **Footnote:** ***Izvor: Podaci brodarskog društva "C" (simulirano)***



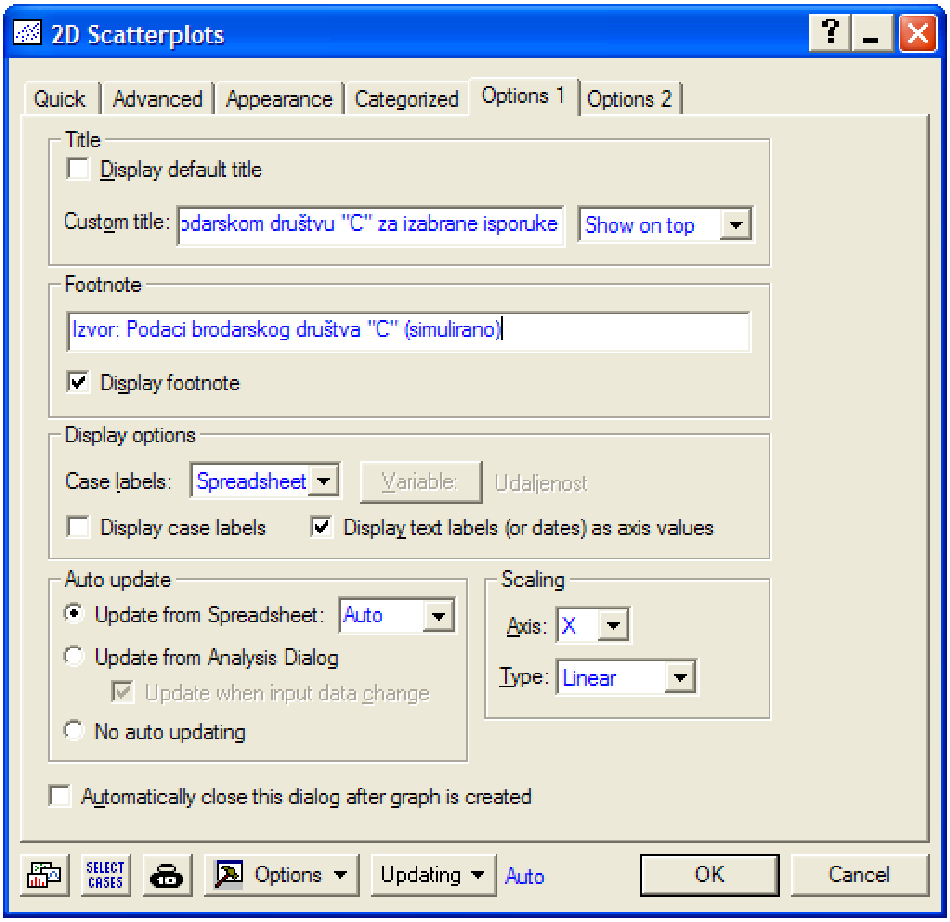
- dobije se grafikon s jednadžbom regresije i ucrtanim pravcem regresije:


Udaljenosti i vrijeme isporuke u brodarskom društvu "C" za izabrane isporuke
Vrijeme = -7,6289+0,0181*x
Vrijeme
Udaljenost
0
2
4
6
8
10
12
14
450
500
550
600
650
700
750
800
850
900
950
Izvor: Podaci brodarskog društva "C" (simulirano)

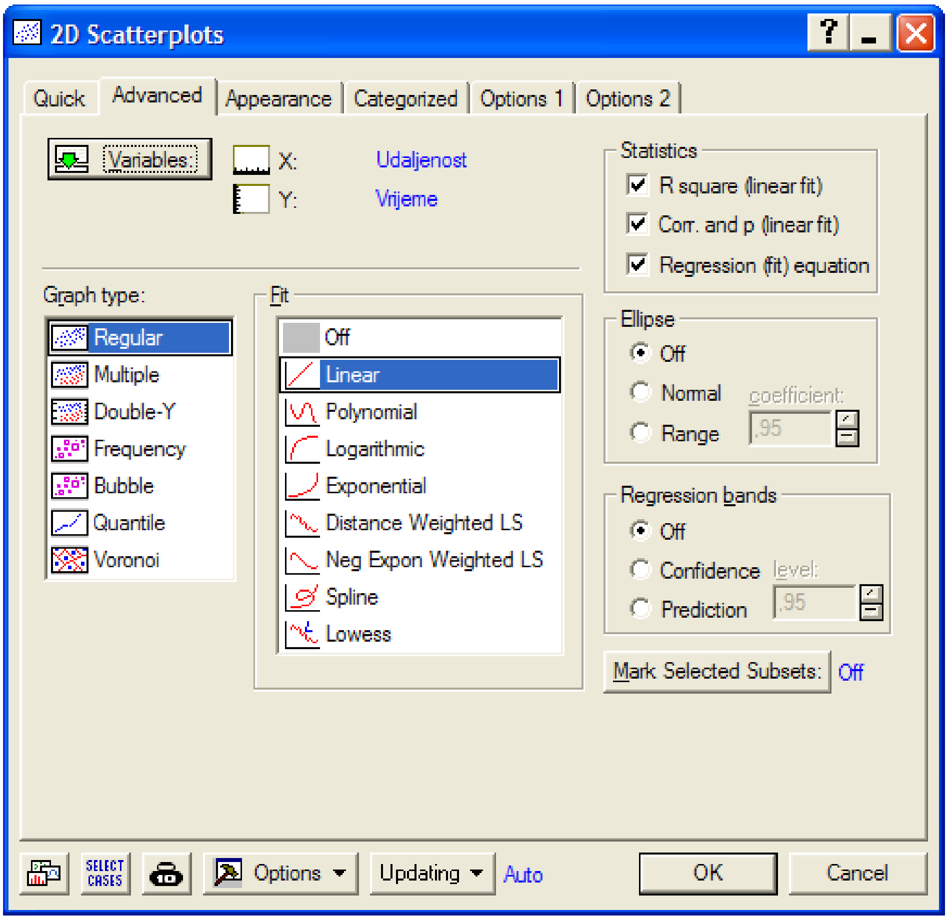
c) Prikaz jednadžbe regresije u dijagramu rasipanja:

➢ **STATISTICS / RESUME...**

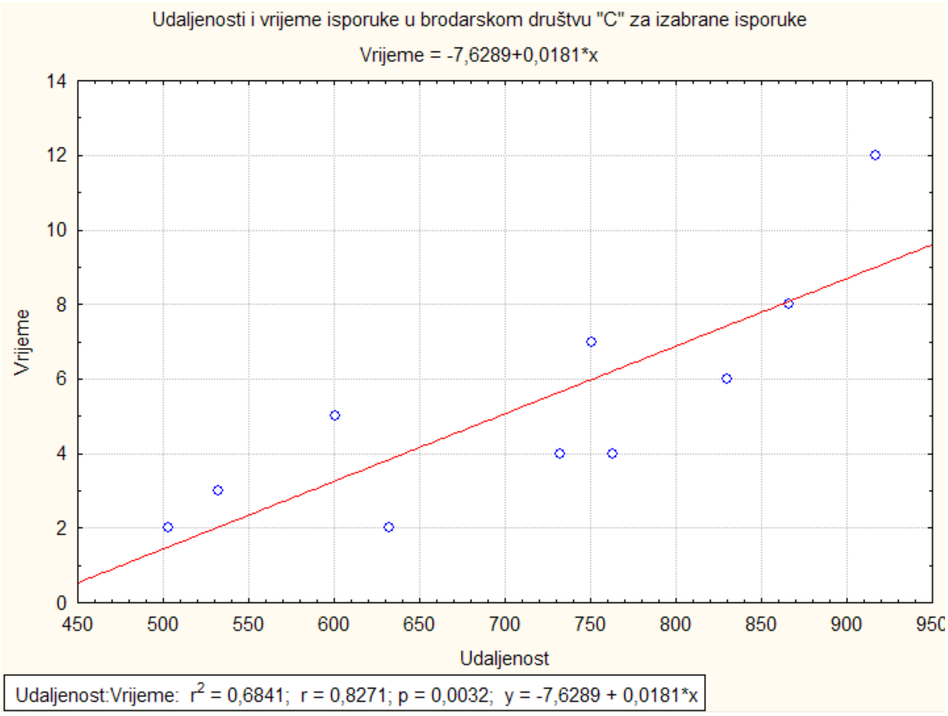
Otvara se zadnji korišteni prozor



- ➢ **Advanced**
  - ➢ **R square (linear fit)** **(R^2)**
  - ➢ **Corr. and p (linear fit)** **(R, p)**
  - ➢ **Regression (fit) equation** **(ocijenjena jednadžba regresije)**



- u grafikonu su dodani i r^2, r, p te jednadžba regresije.

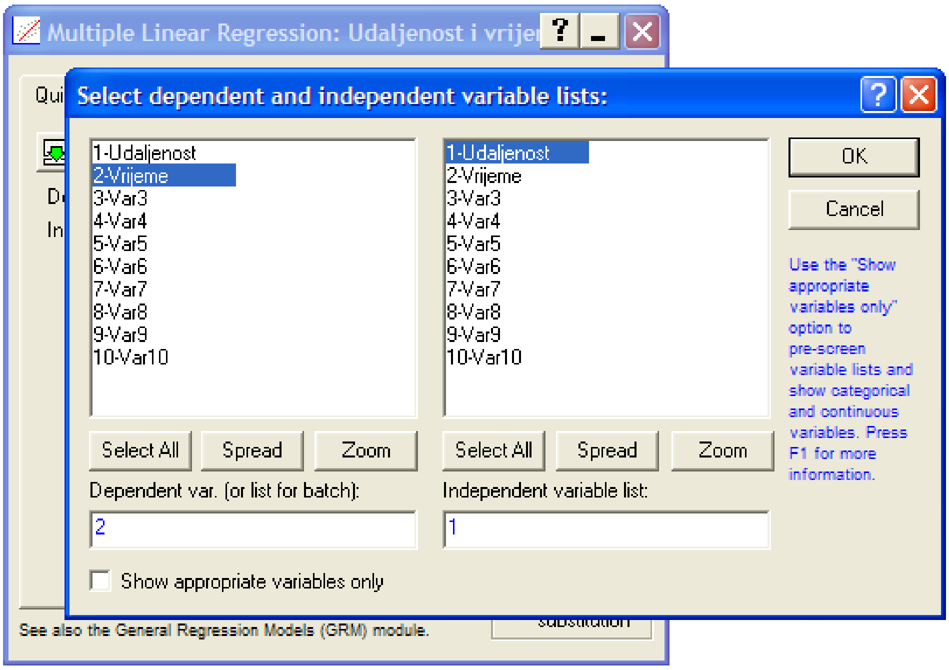


d) Interpolacija:

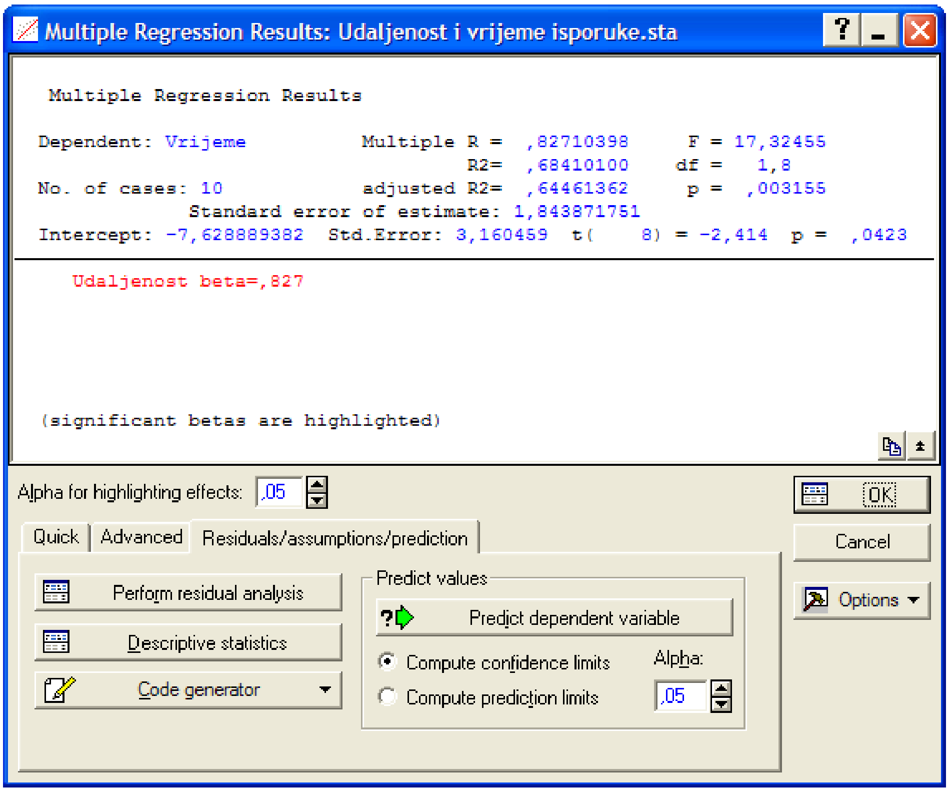
Interpolacija je procjena regresijskih vrijednosti zavisne varijable na temelju zadanih vrijednosti nezavisne varijable. Interpolirati se može ako postoji jednadžba regresije.

Početak računanja je isti kao i kod računanja parametara jednadžbe jednostavne linearne regresije.

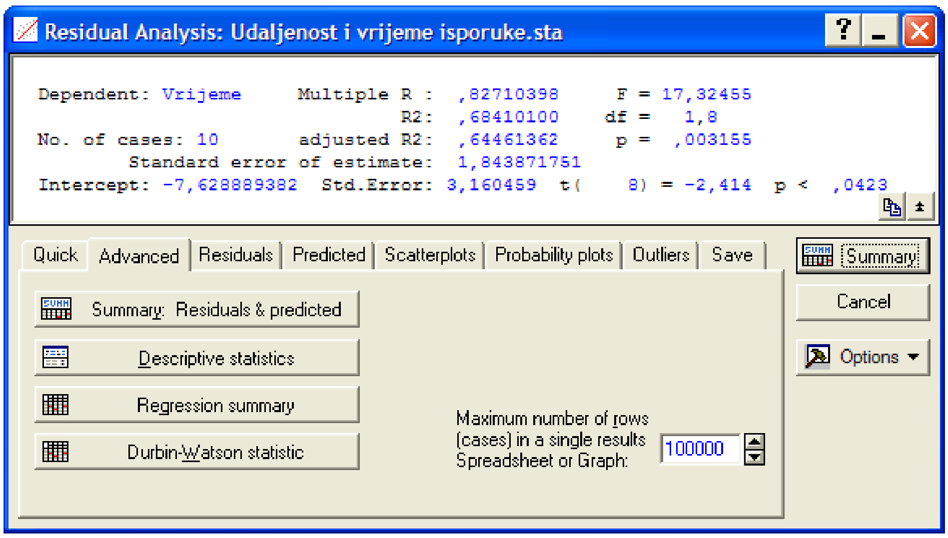
- **STATISTICS / MULTIPLE REGRESSION /**
  - **QUICK**
    - **Variables**
      - **Dependenet:** ***2-Vrijeme***
      - **Independenet:** ***1-Udaljenost***



- **Pritiskom na tipku OK pokreće se računanje:**



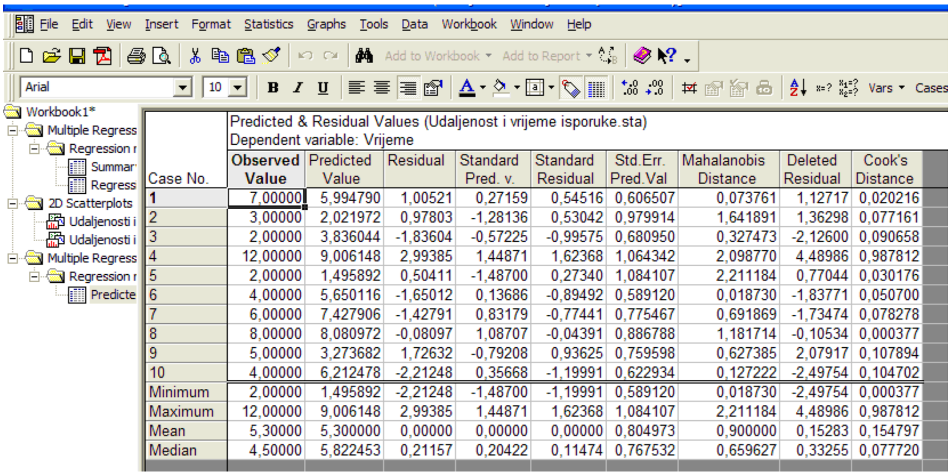
➢ **RESIDUAL/ASSUMPTIONS/PREDICTION**



➢ **PERFORM RESIDUAL ANALISYS**

➢ **ADVANCED**

➢ **Summary: residuals & predicted**



Predicted & Residual Values (Udaljenost i vrijeme isporuke.sta)
Dependent variable: Vrijeme

Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
1	7,00000	5,994790	1,00521	0,27159	0,54516	0,606507	0,073761	1,12717	0,020216
2	3,00000	2,021972	0,97803	-1,28136	0,53042	0,979914	1,641891	1,36298	0,077161
3	2,00000	3,836044	-1,83604	-0,57225	-0,99575	0,680950	0,327473	-2,12600	0,090658
4	12,00000	9,006148	2,99385	1,44871	1,62368	1,064342	2,098770	4,48986	0,987812
5	2,00000	1,495892	0,50411	-1,48700	0,27340	1,084107	2,211184	0,77044	0,030176
6	4,00000	5,650116	-1,65012	0,13686	-0,89492	0,589120	0,018730	-1,83771	0,050700
7	6,00000	7,427906	-1,42791	0,83179	-0,77441	0,775467	0,691869	-1,73474	0,078278
8	8,00000	8,080972	-0,08097	1,08707	-0,04391	0,886788	1,181714	-0,10534	0,000377
9	5,00000	3,273682	1,72632	-0,79208	0,93625	0,759598	0,627385	2,07917	0,107894
10	4,00000	6,212478	-2,21248	0,35668	-1,19991	0,622934	0,127222	-2,49754	0,104702
Minimum	2,00000	1,495892	-2,21248	-1,48700	-1,19991	0,589120	0,018730	-2,49754	0,000377
Maximum	12,00000	9,006148	2,99385	1,44871	1,62368	1,084107	2,211184	4,48986	0,987812
Mean	5,30000	5,300000	0,00000	0,00000	0,00000	0,804973	0,900000	0,15283	0,154797
Median	4,50000	5,822453	0,21157	0,20422	0,11474	0,767532	0,659627	0,33255	0,077720

Rezultati:

Interpolacija je izračunata u varijabli *Predicted Value*

Predicted & Residual Values (Udaljenost i vrijeme isporuke.sta)
Dependent variable: Vrijeme

Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mah D
1	7,00000	**5,994790**	1,00521	0,27159	0,54516	0,606507	
2	3,00000	**2,021972**	0,97803	-1,28136	0,53042	0,979914	
3	2,00000	**3,836044**	-1,83604	-0,57225	-0,99575	0,680950	
4	12,00000	**9,006148**	2,99385	1,44871	1,62368	1,064342	
5	2,00000	**1,495892**	0,50411	-1,48700	0,27340	1,084107	
6	4,00000	**5,650116**	-1,65012	0,13686	-0,89492	0,589120	
7	6,00000	**7,427906**	-1,42791	0,83179	-0,77441	0,775467	
8	8,00000	**8,080972**	-0,08097	1,08707	-0,04391	0,886788	
9	5,00000	**3,273682**	1,72632	-0,79208	0,93625	0,759598	
10	4,00000	**6,212478**	-2,21248	0,35668	-1,19991	0,622934	
Minimum	2,00000	1,495892	-2,21248	-1,48700	-1,19991	0,589120	
Maximum	12,00000	9,006148	2,99385	1,44871	1,62368	1,084107	
Mean	5,30000	5,300000	0,00000	0,00000	0,00000	0,804973	
Median	4,50000	5,822453	0,21157	0,20422	0,11474	0,767532	

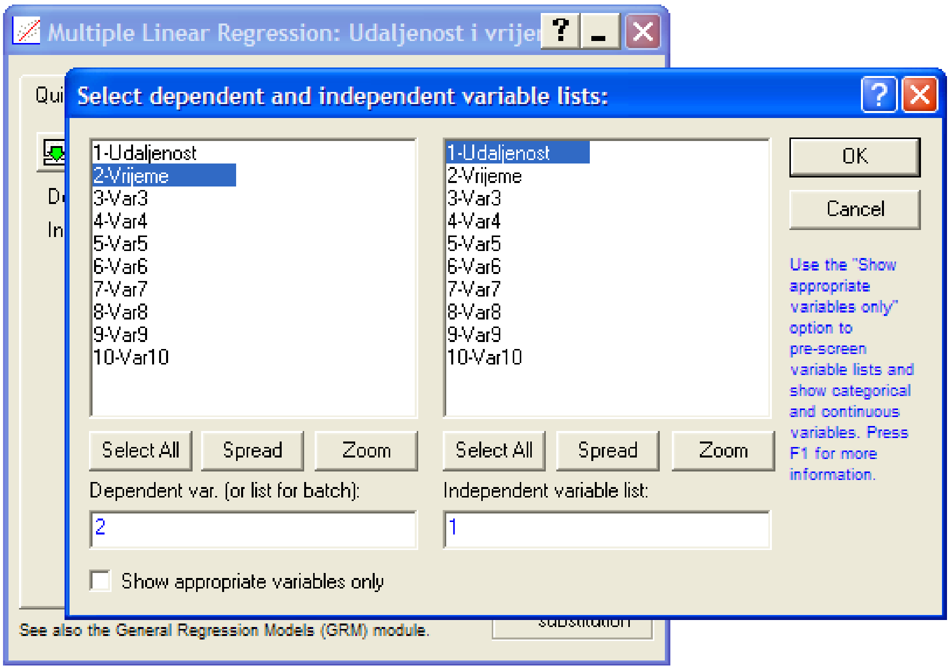
U 1. slučaju, kada je udaljenost iznosila 751 nautičku milju, predviđeno vrijeme isporuke bi bilo 5,99 dana.

e) Procjena vremena trajanja isporuke za udaljenost od 600 nautičkih milja:

Procjena regresijskih vrijednosti zavisne varijable kada nezavisna varijabla poprima vrijednosti koje nisu ranije zadane naziva se ekstrapolacija.

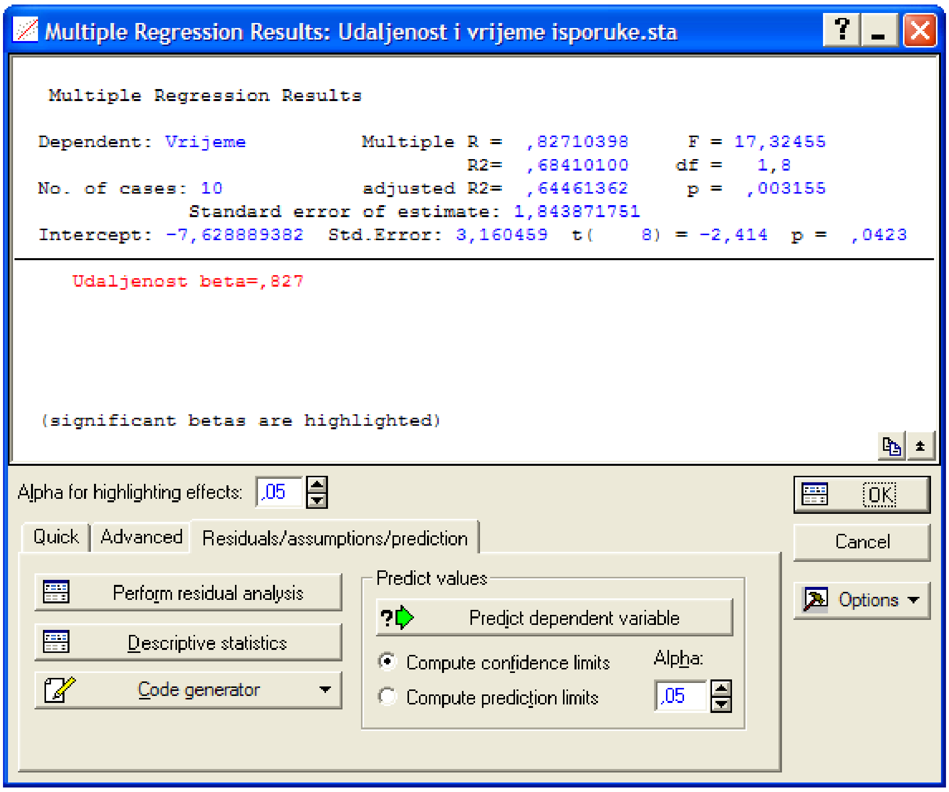
Početak računanja je isti kao i kod računanja parametara jednadžbe jednostavne linearne regresije.

- **STATISTICS / MULTIPLE REGRESSION /**
  - **QUICK**
    - **Variables**

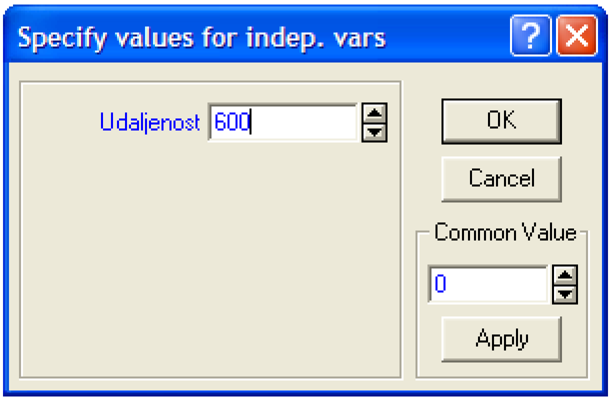


- **Dependenet: *2-Vrijeme***
- **Independenet: *1-Udaljenost***

- **Pritiskom na tipku OK pokreće se računanje:**



➢ **RESIDUAL/ASSUMPTIONS/PREDICTION**



➢ **Predict dependent variable**

➢ **Specifiy values for indep. vars: Udaljenost** $\boxed{600}$

Rezultati:

Variable	Predicting Values for (Udaljenost i vrijeme isporuke.s variable: Vrijeme		
	B-Weight	Value	B-Weight * Value
Udaljenost	0,018141	600,0000	10,88443
Intercept			-7,62889
Predicted			3,25554
-95,0%CL			**1,49745**
+95,0%CL			**5,01364**

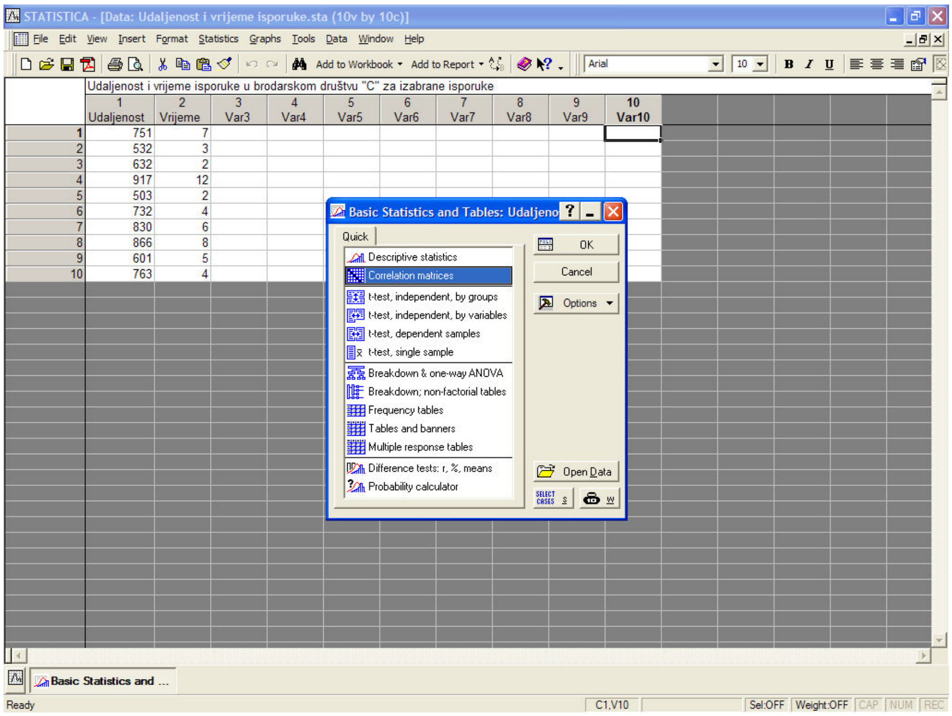
Predviđeno vrijeme trajanja isporuke za udaljenost od 600 nautičkih milja, na razini 95% pouzdanosti, je između 1,49745 i 5,01364 dana.

4.2. Linearna korelacija

Primjer 4.2.

Za podatke iz primjera 4.1. odredite koeficijent linearne korelacije.

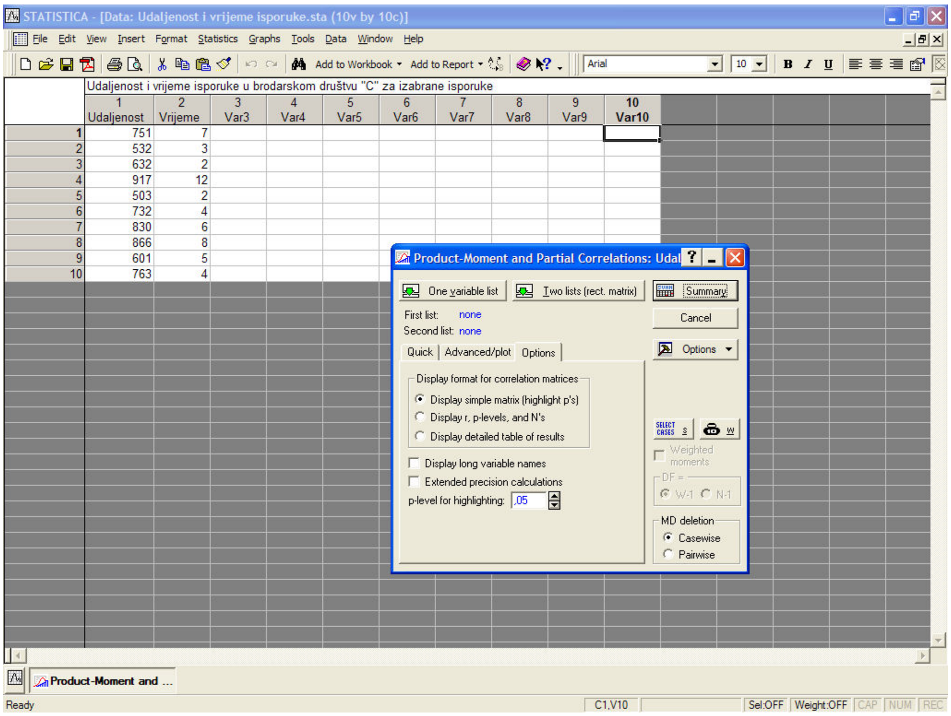
- **STATISTICS / BASIC STATISTIC/TABLES**
    - **QUICK**



Izabire se:

- **CORRELATIONS MATRICES**

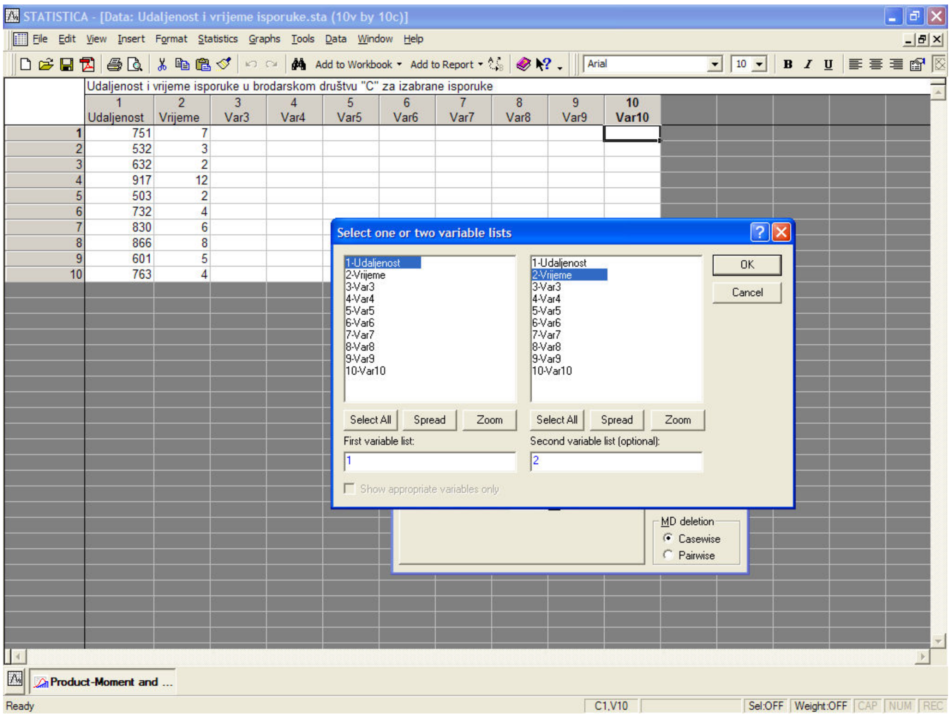
Otvara se prozor za određivanje koeficijenta korelacije:



Varijable je moguće unijeti na dva načina:

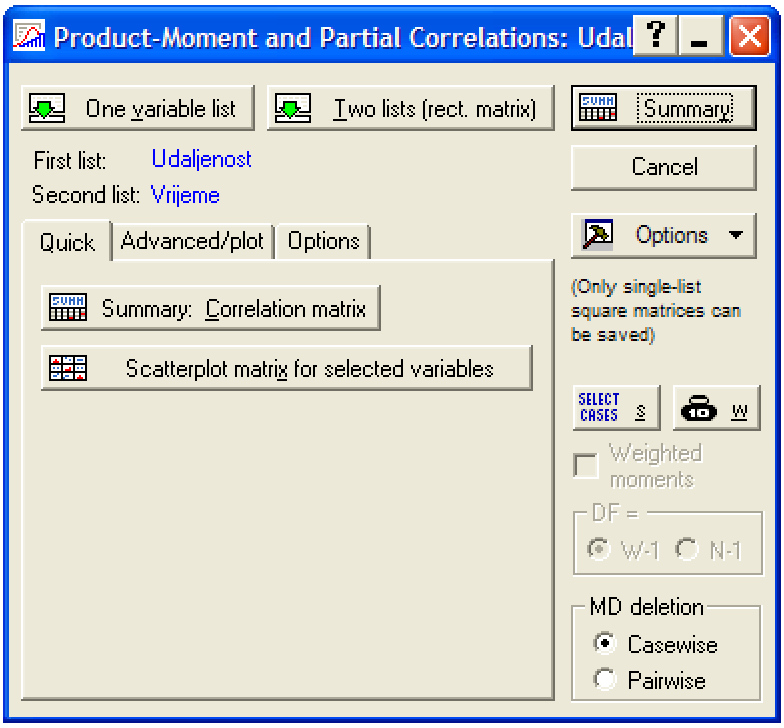
- ***One variable list* – ako se želi dobiti matrica koeficijenata korelacije**
- ***Two variable list (rect. matrix)* – za izračun samo određenih koeficijenata korelacije.**

➢ **Two variable list (rect. matrix)**



➢ **First variable list:** ***1-Udaljenost***

➢ **Second variable list (optional):** ***2-Vrijeme***



- **Quick**
  - **Summary: Correlation matrix**

Rezultat:

	Correlations (Udaljenost i vrijeme isporuke.sta) Marked correlations are significant at p < ,05000 N=10 (Casewise deletion of missing data)
Variable	Vrijeme
Udaljenost	0,83

Da bi se koeficijent korelacije mogao koristiti, mora biti statistički značajan. U prozoru s rezultatima koeficijent korelacije koji je statistički značajan označen je crvenom bojom ($r = 0{,}83$).

Među zadanim varijablama postoji jaka pozitivna korelacija.

4.3. Spearmanov koeficijent korelacije ranga

Bruto domaći proizvod po stanovniku u USD u tekućim cijenama i prema paritetu kupovne moći u zemljama srednje i istočne Europe u 2007. god.

Zemlja	*BDP po stanovniku, tekuće cijene, USD*	*BDP po stanovniku, PPP, USD*
Albanija	3.431	6.319
BIH	3.809	7.081
Bugarska	5.186	11.311
Češka	16.856	24.088
Hrvatska	13.199	17.732
Mađarska	13.752	18.956
Makedonija	3.874	8.561
Poljska	11.143	16.323
Rumunjska	7.850	11.456
Slovačka	13.898	20.275
Slovenija	23.511	27.901
Srbija	5.477	10.019

Izvor: International Monetary Fund, World Economic Outlook Database, April 2009. god.

Zadatak je odrediti korelaciju ranga:

- U zaglavlje dokumenta treba upisati:

Bruto domaći proizvod po stanovniku u USD u tekućim cijenama i prema paritetu kupovne moći u zemljama srednje i istočne Europe u 2007. god.

Stupac zemlja može se unijeti kao pred stupac:

- U informativno polje upisati ***Zemlja***

- Definiranje varijabli:
  - Kratki naziv: ***BDP_pc***, dugi naziv:

    BDP po stanovniku u tekućim cijenama USD

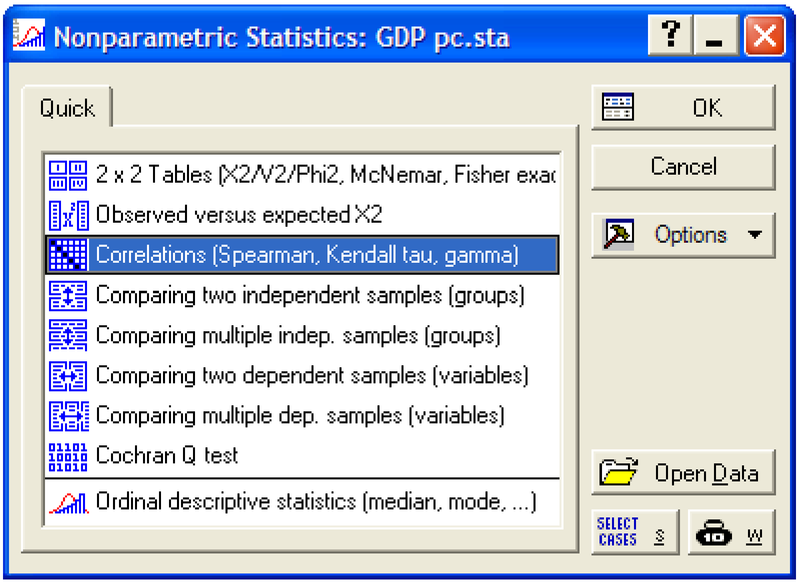
- Kratki naziv: $\boxed{BDP\_pc\_PPP}$, dugi naziv:

  BDP po stanovniku, PPP, USD

- Unijeti vrijednosti.

Korelacija ranga računa se na sljedeći način:

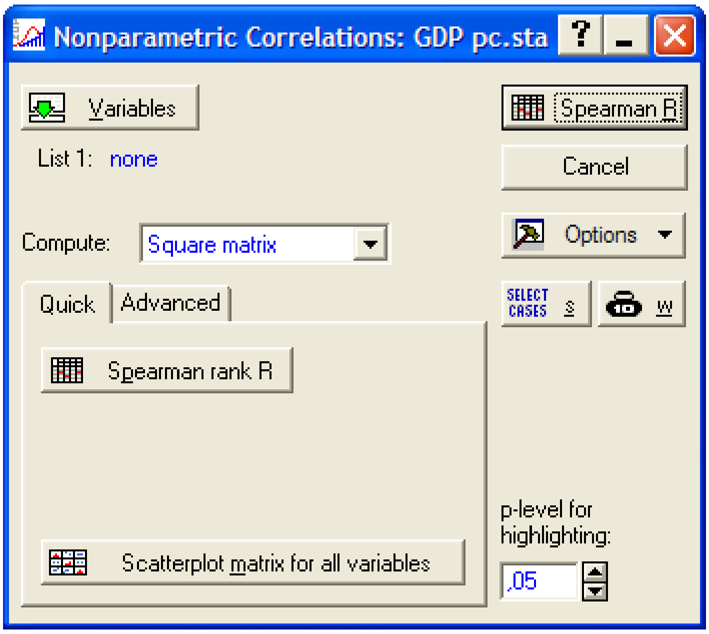
- **STATISTICS / NONPARAMETRICS**
  - **QUICK**



Izabire se:

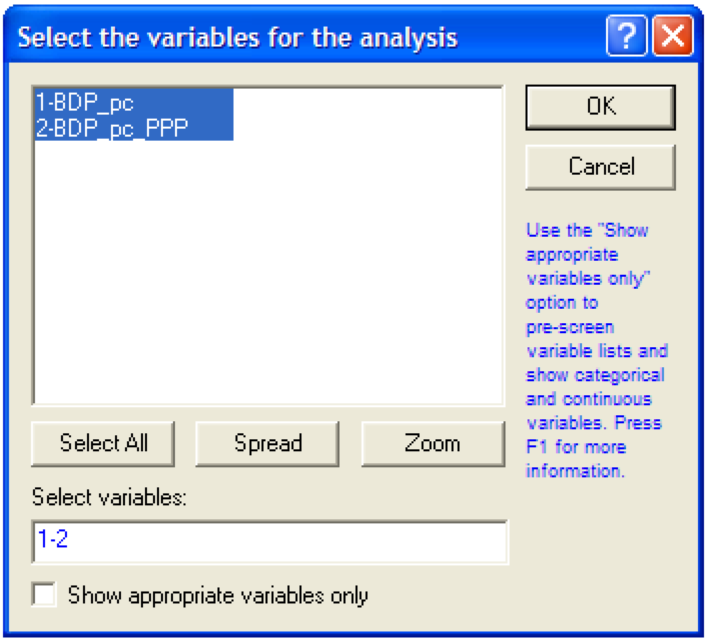
- **CORRELATIONS (SPEARMAN, KENDALL TAU, GAMMA)**

Otvara se prozor za izračun Spearmanove korelacije ranga:



➤ **Variables**

1-BDP_pc
2-BDP_pc_PPP



➤ **Quick**

➤ **Spearman rank R**

Rezultat:

Variable	Spearman Rank Order Correlations (GDP pc.sta) MD pairwise deleted Marked correlations are significant at p <,05000	
	BDP_pc	BDP_pc_PPP
BDP_pc	1,000000	0,993007
BDP_pc_PPP	0,993007	1,000000

Ako je koeficijent statistički značajan na razini 0,05 označen je crvenom bojom. Spearmanov koeficijent korelacije ranga između bruto domaćeg proizvoda po stanovniku u USD u tekućim cijenama i prema paritetu kupovne moći iznosi 0,993. Koeficijent je statistički značajan.

5. Analiza vremenskih nizova

5.1. Bazni i verižni indeksi

Primjer 5.1.

Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999.-2005. god.

Godina	*Proizvodnja u tisućama T*
1999.	1361
2000.	1130
2001.	1078
2002.	1274
2003.	1434
2004.	1288
2005.	1602

Izvor: Podaci tvornice D (simulirani podaci)

Zadatak je:

a) izračunati indekse na bazi proizvodnje u 1999. god.
b) protumačiti dobivene vrijednosti
c) izračunati verižne indekse
d) preračunati indekse 1999.=100 u verižne
e) protumačiti dobivene vrijednosti
f) izračunati indekse prosjek=100
g) protumačiti dobivene vrijednosti.

a) Izračun indeksa na bazi proizvodnje u 1999. god:

- U **zaglavlje** dokumenta upisati:

Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999.-2005. god.

Stupac godine može se unijeti kao pred stupac:

- U informativno polje upisati $\boxed{\textit{Godina}}$
- Definirati varijablu:
  - Kratki naziv: $\boxed{\textit{Yt}}$, dugi naziv:

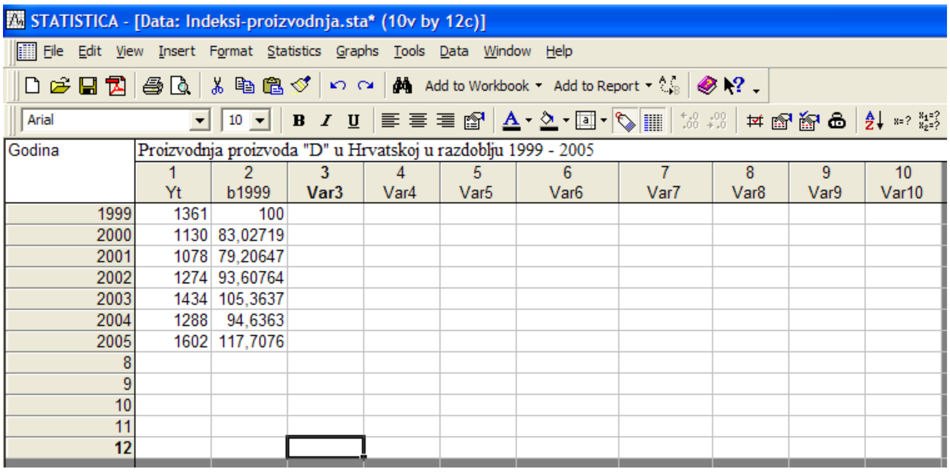
Proizvodnja proizvoda D u tisućama tona

- Unijeti vrijednosti.

Bazni indeksi računaju se po formuli: $I_t = \frac{Y_t}{Y_b} \cdot 100$

Baza je proizvodnja u 1999. god., tj. 1361 tisuća *t*. Otvara se nova varijabla, b1999, u kojoj se bazni indeksi računaju prema formuli: $I_t = \frac{Y_t}{1361} \cdot 100$

- **Definirati novi stupac – *b1999***
- **Long name:** $\boxed{\textit{=Yt/1361*100}}$



Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005

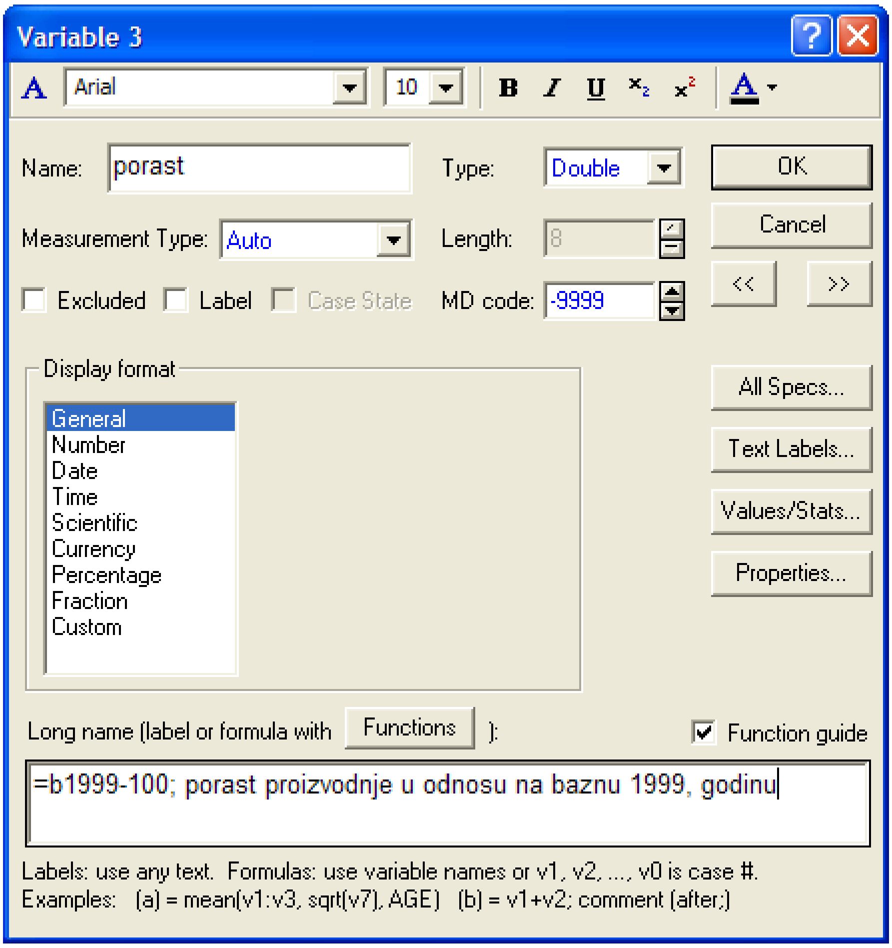
Godina	1 Yt	2 b1999	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1999	1361	100								
2000	1130	83,02719								
2001	1078	79,20647								
2002	1274	93,60764								
2003	1434	105,3637								
2004	1288	94,6363								
2005	1602	117,7076								
8										
9										
10										
11										
12										

b) Objašnjenje dobivenih vrijednosti:

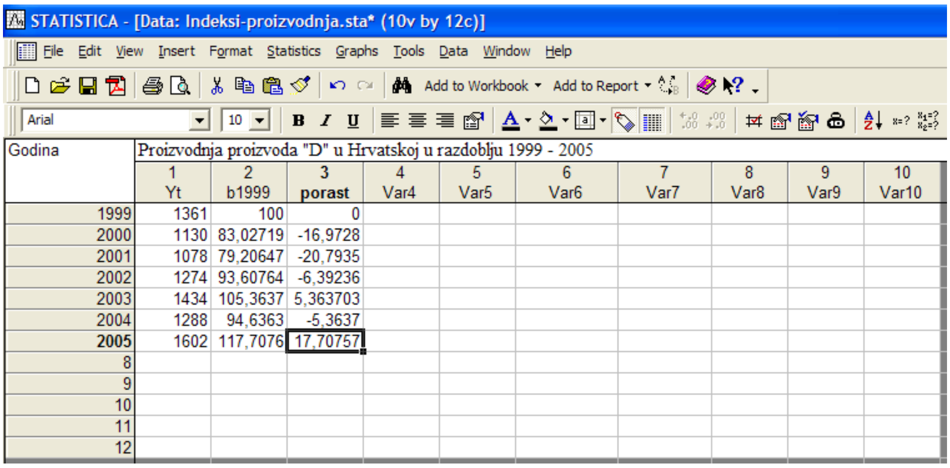
Oduzimanjem broja 100 od vrijednosti indeksa dobije se promjena proizvodnje u odnosu na bazu.

- **Definirati novi stupac –** ***porast***
- **Long name:**

=b1999-100; porast proizvodnje u odnosu na baznu 1999. godinu



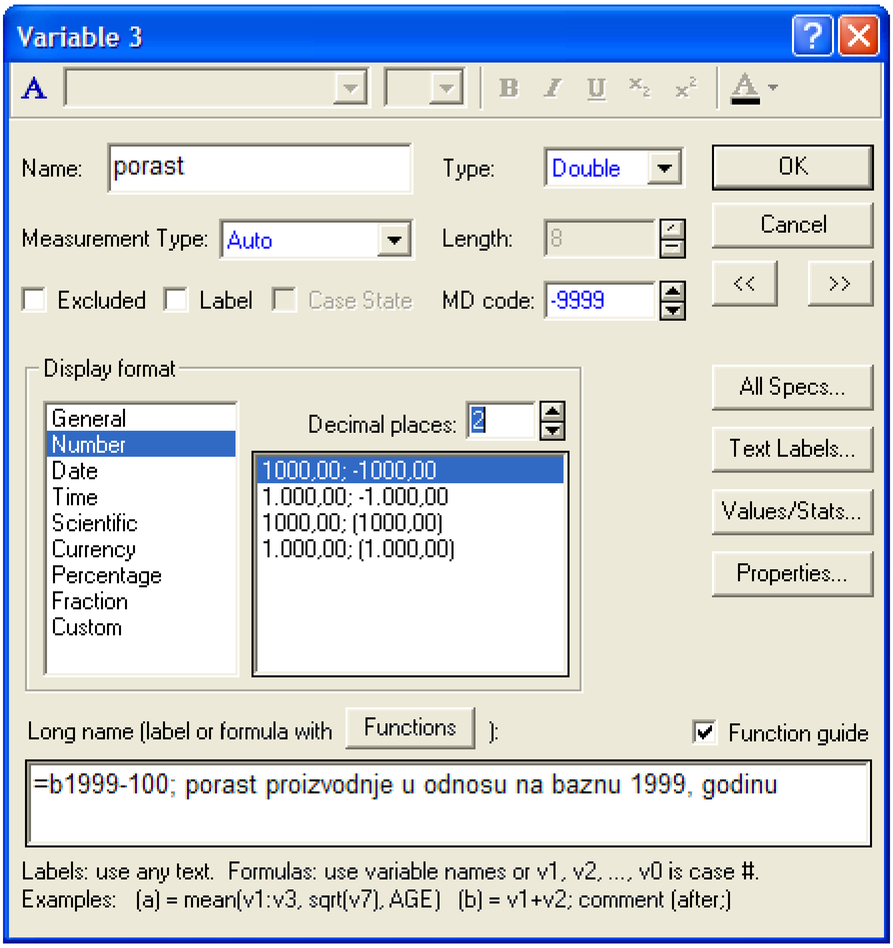
Nakon upozorenja o računanju vrijednosti nove varijable dobiju se tumačenja indeksa:



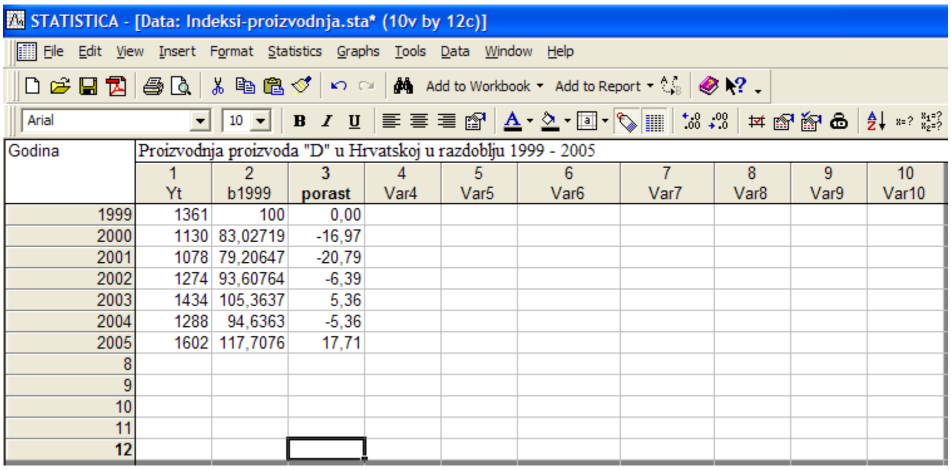
Godina	Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005 1 Yt	2 b1999	3 porast	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1999	1361	100	0							
2000	1130	83,02719	-16,9728							
2001	1078	79,20647	-20,7935							
2002	1274	93,60764	-6,39236							
2003	1434	105,3637	5,363703							
2004	1288	94,6363	-5,3637							
2005	1602	117,7076	17,70757							
8										
9										
10										
11										
12										

Po želji mogu se prikazati vrijednosti varijable zaokružene na dvije decimale:

- **DATA / VARIABLE SPECS...**
  - **Display format** ***Number***
    - **Decimal places** *2*



Varijabla porast s vrijednostima zaokruženima na 2 decimale:



Godina	Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005									
	1 Yt	2 b1999	3 porast	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1999	1361	100	0,00							
2000	1130	83,02719	-16,97							
2001	1078	79,20647	-20,79							
2002	1274	93,60764	-6,39							
2003	1434	105,3637	5,36							
2004	1288	94,6363	-5,36							
2005	1602	117,7076	17,71							
8										
9										
10										
11										
12										

Proizvodnja proizvoda D porasla je 2003. za 5,36% u odnosu na proizvodnju u baznoj 1999. godini.

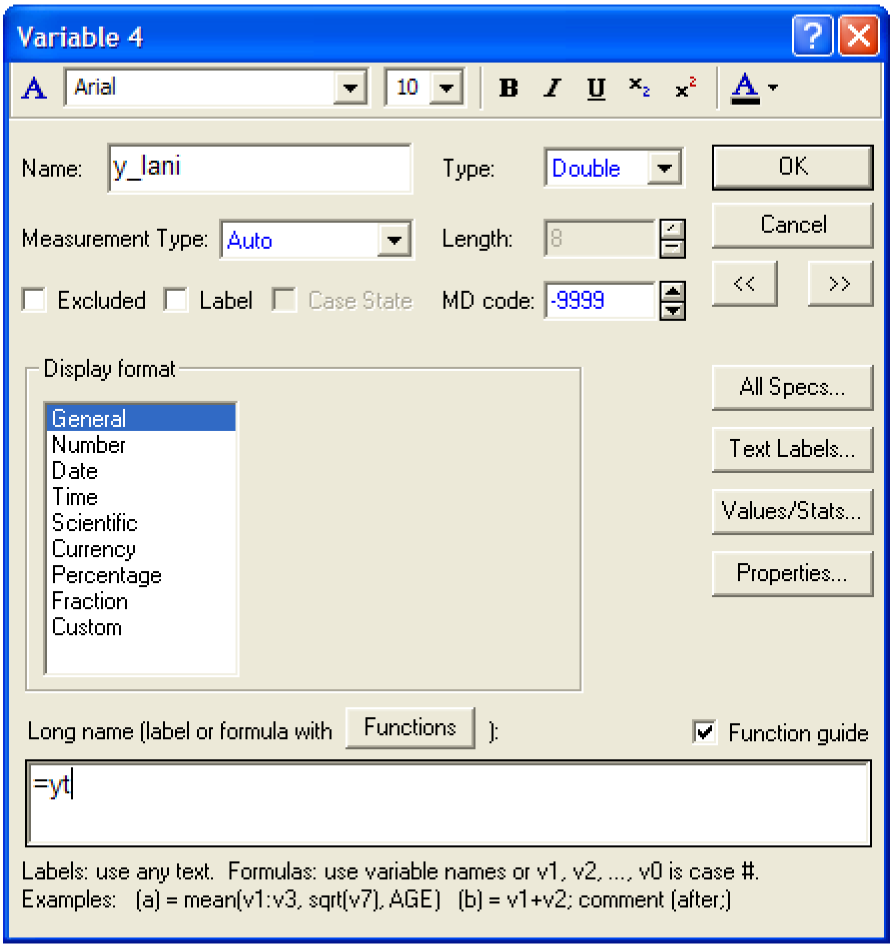
U 2000. god. proizvodnja proizvoda A smanjila se za 16,97% u odnosu na proizvodnju u 1999. god.

c) Izračun verižnih indeksa:

Baza indeksa se mijenja, a u ovom primjeru to je proizvodnja u prethodnoj godini.
1. korak – definirati novu varijablu s vrijednostima prethodne godine. Prvo se kopira varijabla Yt u novu varijablu Y_lani, a zatim se pomakne za 1 godinu unaprijed.

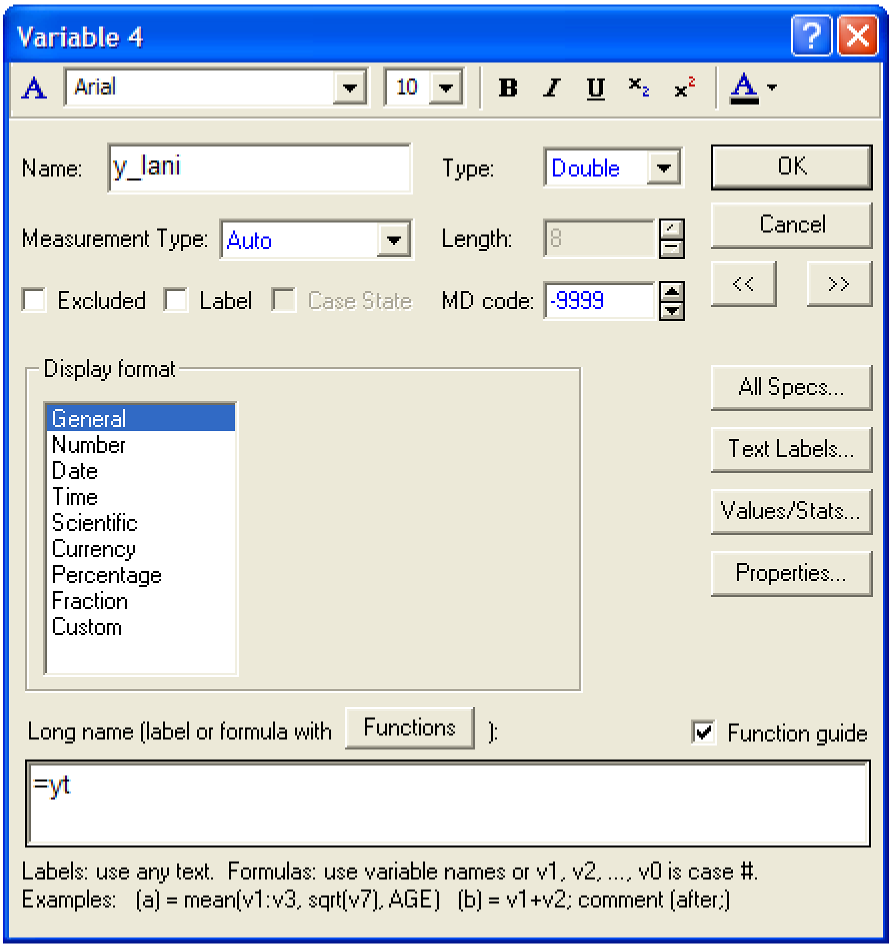
Stvaranje nove varijable s kopiranim vrijednostima varijable Yt:

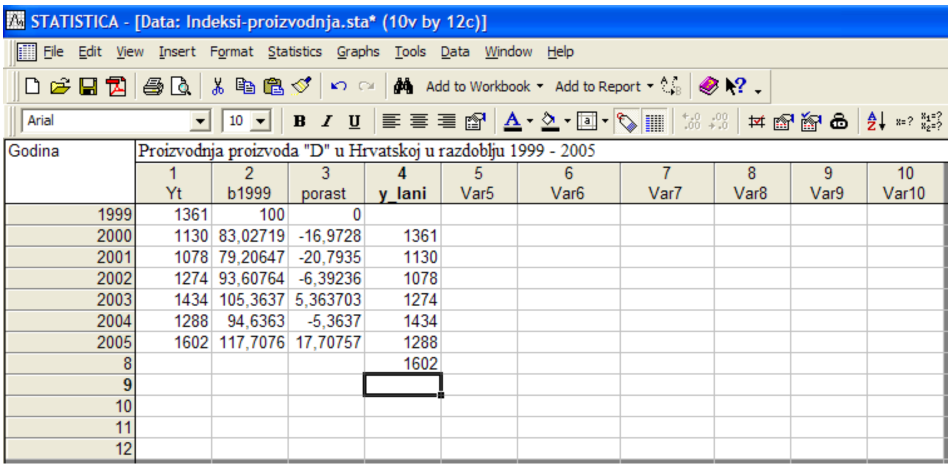
- **Definirati varijablu: Y_lani, dugi naziv** *=Yt*



Pomicanje varijable za 1 godinu Yt:

- **DATA / SHIFT (LAG)**
  - **Variables:** ***4 – Y_lani***
  - **Lag:** ***1***
  - **Direction :** ***Forward***



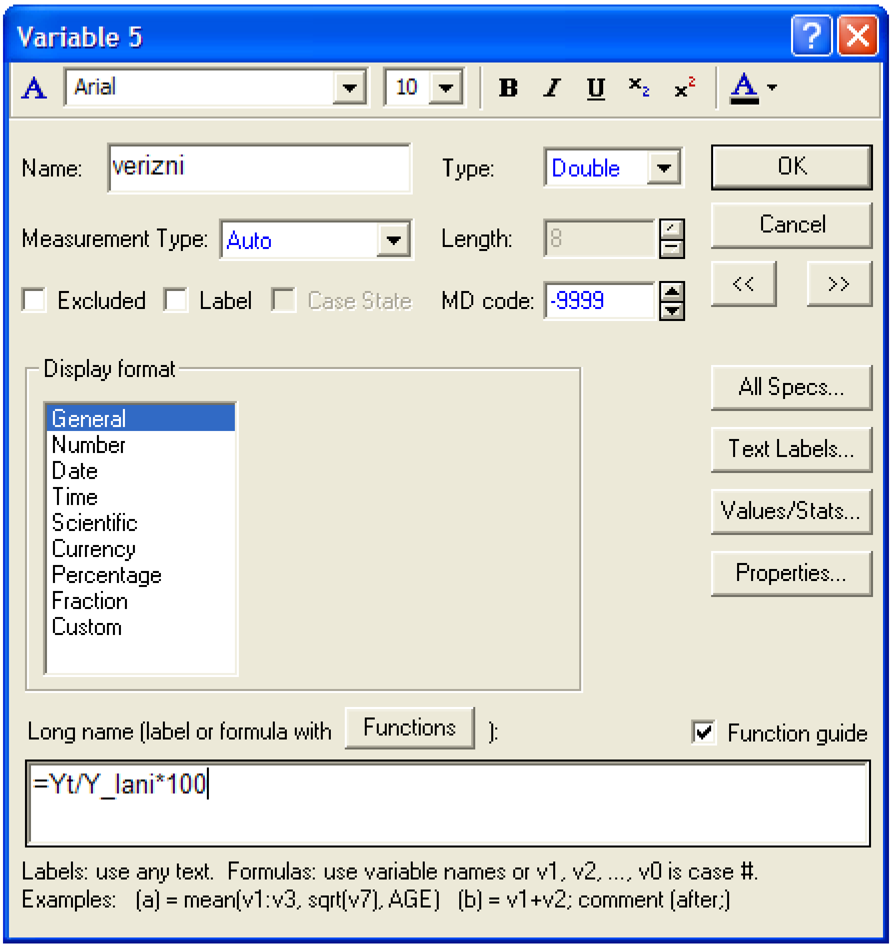


Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005

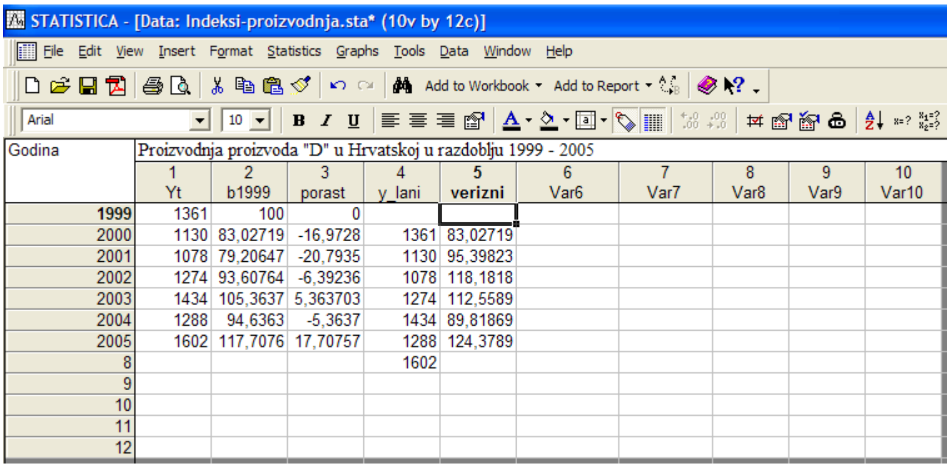
Godina	1 Yt	2 b1999	3 porast	4 y_lani	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1999	1361	100	0							
2000	1130	83,02719	-16,9728	1361						
2001	1078	79,20647	-20,7935	1130						
2002	1274	93,60764	-6,39236	1078						
2003	1434	105,3637	5,363703	1274						
2004	1288	94,6363	-5,3637	1434						
2005	1602	117,7076	17,70757	1288						
8				1602						
9										
10										
11										
12										

2. korak – izračunati verižne indekse:

- **Definirati novi stupac –** ***verizni***
- **Long name:** ***=Yt/Y_lani*100***



Rezultati:



Godina	Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005 1 Yt	2 b1999	3 porast	4 y_lani	5 verizni	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1999	1361	100	0							
2000	1130	83,02719	-16,9728	1361	83,02719					
2001	1078	79,20647	-20,7935	1130	95,39823					
2002	1274	93,60764	-6,39236	1078	118,1818					
2003	1434	105,3637	5,363703	1274	112,5589					
2004	1288	94,6363	-5,3637	1434	89,81869					
2005	1602	117,7076	17,70757	1288	124,3789					
8				1602						
9										
10										
11										
12										

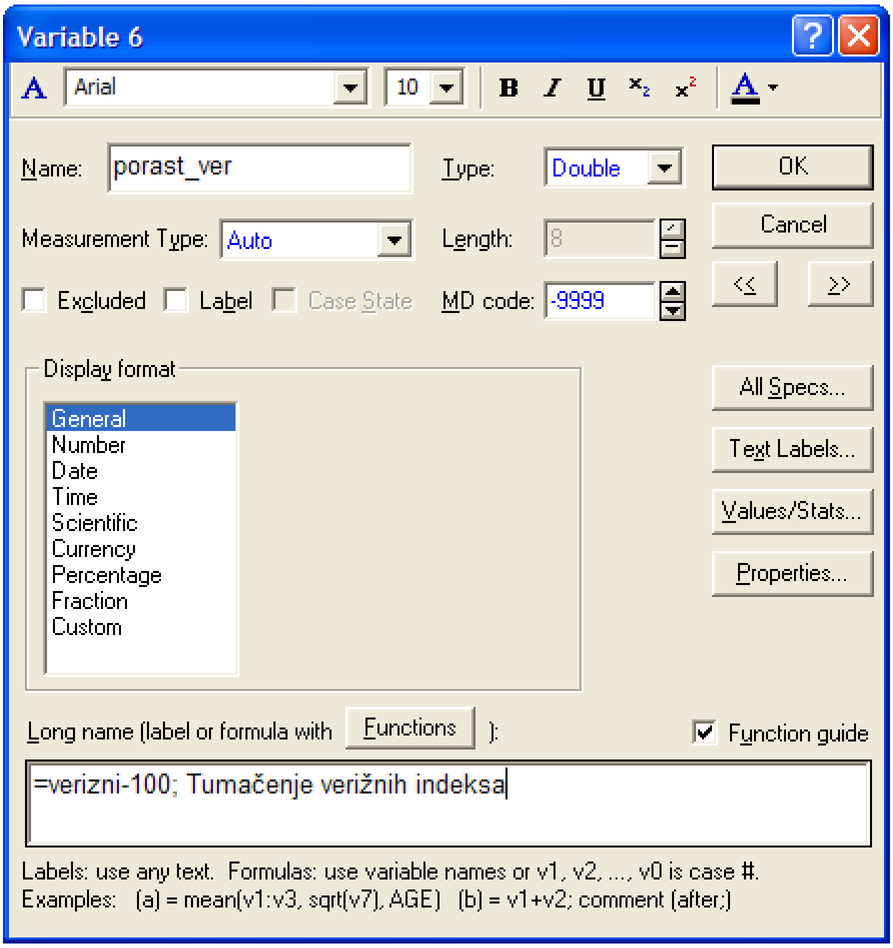
d) Tumačenje dobivenih vrijednosti:

- **Definirati novi stupac – *porast_ver***
- **Long name:**

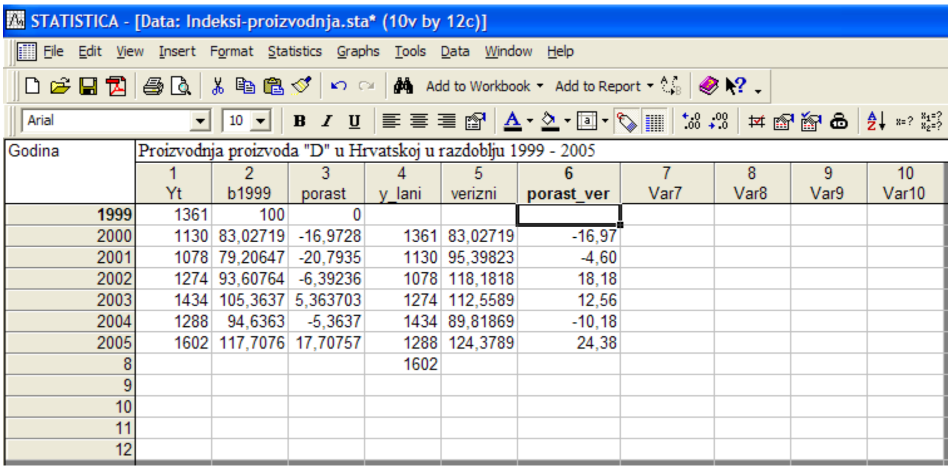
=verizni-100;Tumačenje verižnih indeksa

Varijabla se može zaokružiti na 2 decimale:

- **Display format** $\boxed{\textit{Number}}$
  - **Decimal places** $\boxed{\textit{2}}$


Variable 6
Arial
10
Name: porast_ver
Type: Double
OK
Measurement Type: Auto
Length: 8
Cancel
Excluded
Label
Case State
MD code: -9999
Display format
General
Number
Date
Time
Scientific
Currency
Percentage
Fraction
Custom
All Specs...
Text Labels...
Values/Stats...
Properties...
Long name (label or formula with Functions):
Function guide
=verizni-100; Tumačenje verižnih indeksa
Labels: use any text. Formulas: use variable names or v1, v2, ..., v0 is case #.
Examples: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; comment (after;)

Nakon upozorenja o računanju vrijednosti nove varijable dobiju se tumačenja indeksa:



Godina	1 Yt	2 b1999	3 porast	4 y_lani	5 verizni	6 porast_ver	7 Var7	8 Var8	9 Var9	10 Var10
	Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005									
1999	1361	100	0							
2000	1130	83,02719	-16,9728	1361	83,02719	-16,97				
2001	1078	79,20647	-20,7935	1130	95,39823	-4,60				
2002	1274	93,60764	-6,39236	1078	118,1818	18,18				
2003	1434	105,3637	5,363703	1274	112,5589	12,56				
2004	1288	94,6363	-5,3637	1434	89,81869	-10,18				
2005	1602	117,7076	17,70757	1288	124,3789	24,38				
8				1602						
9										
10										
11										
12										

Proizvodnja **2000.** je za 16,97% manja od proizvodnje u **baznoj 1999.** godini.
Proizvodnja **2001.** je za 4,6% manja od proizvodnje u **baznoj 2000.** godini.
Proizvodnja **2002.** je za 18,18% veća od proizvodnje u **baznoj 2001.** godini.

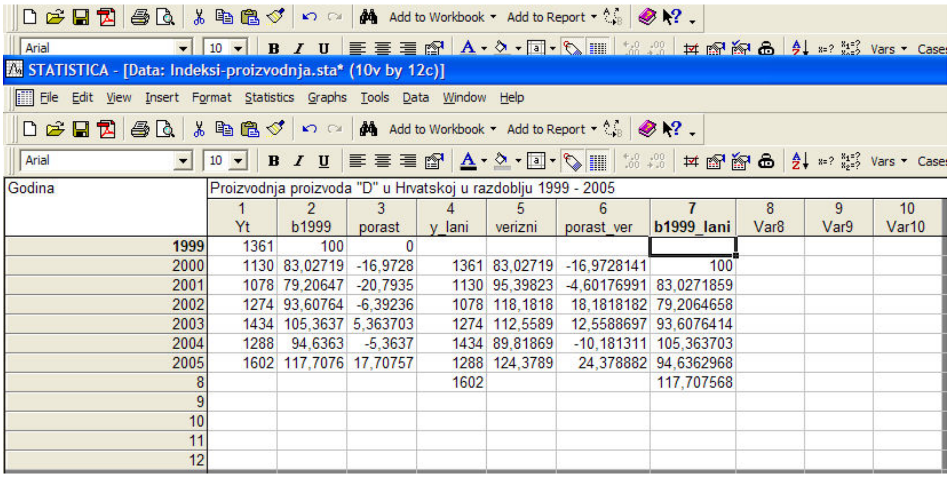
...

e) Pretvaranje indeksa 1999.=100 u verižne:

Bazni indeks se dijeli s prijašnjom vrijednosti.

1. korak – definirati novu varijablu s vrijednostima indeksa na bazi proizvodnje u 1999. god. u prethodnoj godini. Prvo se kreira nova varijabla s kopiranim vrijednostima varijable b1999, a zatim se pomakne za 1 godinu unaprijed:

- **Definirati varijablu: b1999_lani, dugi naziv** $\boxed{\text{=b1999}}$
  - **DATA / SHIFT (LAG)**
    - **Variables:** $\boxed{\textit{7 – b1999\_lani}}$
    - **Lag:** $\boxed{\textit{1}}$
    - **Direction :** $\boxed{\textit{Forward}}$

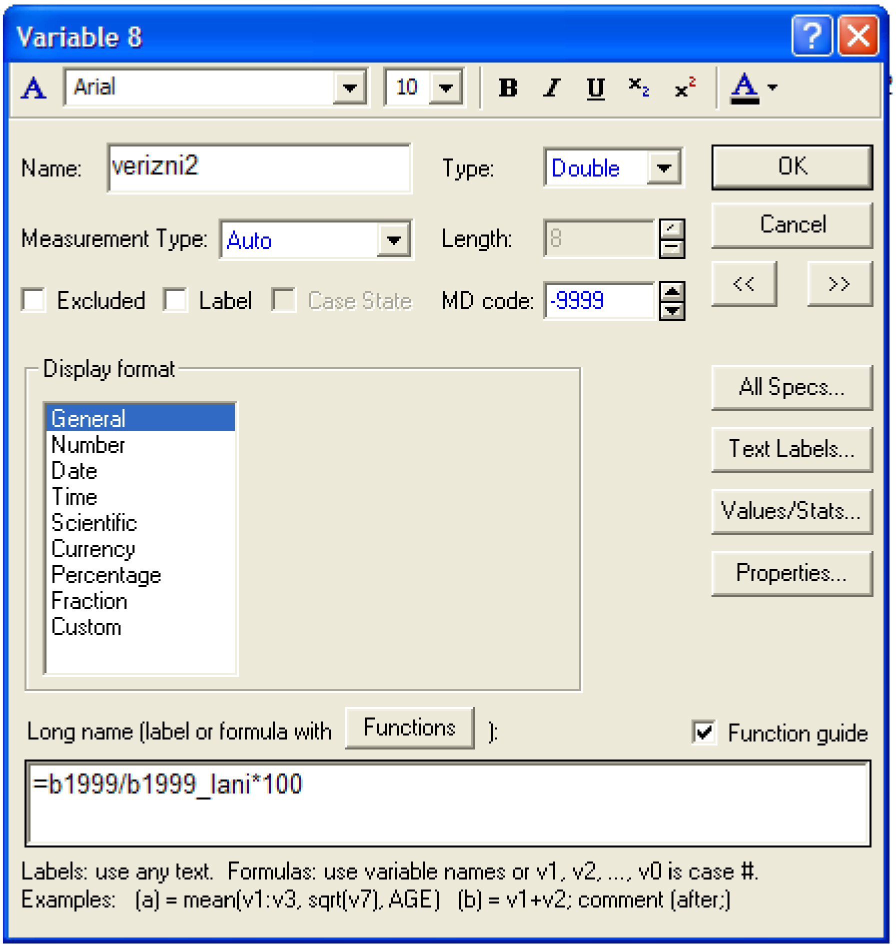


Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005

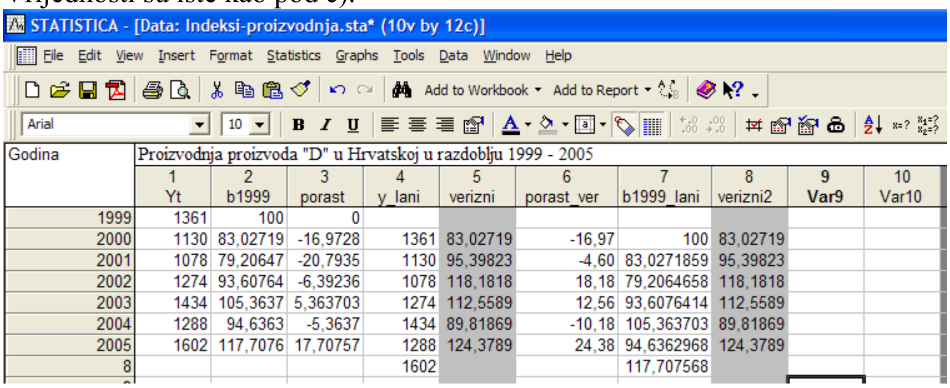
Godina	1 Yt	2 b1999	3 porast	4 y_lani	5 verizni	6 porast_ver	7 b1999_lani	8 Var8	9 Var9	10 Var10
1999	1361	100	0							
2000	1130	83,02719	-16,9728	1361	83,02719	-16,9728141	100			
2001	1078	79,20647	-20,7935	1130	95,39823	-4,60176991	83,0271859			
2002	1274	93,60764	-6,39236	1078	118,1818	18,1818182	79,2064658			
2003	1434	105,3637	5,363703	1274	112,5589	12,5588697	93,6076414			
2004	1288	94,6363	-5,3637	1434	89,81869	-10,181311	105,363703			
2005	1602	117,7076	17,70757	1288	124,3789	24,378882	94,6362968			
8				1602			117,707568			
9										
10										
11										
12										

2. korak – preračunati bazne u verižne indekse:

- **Definirati novi stupac – *verizni2***
- **Long name:** $\boxed{=b1999/b1999\_lani*100}$



Vrijednosti su iste kao pod c).



Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005

Godina	1 Yt	2 b1999	3 porast	4 y_lani	5 verizni	6 porast_ver	7 b1999_lani	8 verizni2	9 Var9	10 Var10
1999	1361	100	0							
2000	1130	83,02719	-16,9728	1361	83,02719	-16,97	100	83,02719		
2001	1078	79,20647	-20,7935	1130	95,39823	-4,60	83,0271859	95,39823		
2002	1274	93,60764	-6,39236	1078	118,1818	18,18	79,2064658	118,1818		
2003	1434	105,3637	5,363703	1274	112,5589	12,56	93,6076414	112,5589		
2004	1288	94,6363	-5,3637	1434	89,81869	-10,18	105,363703	89,81869		
2005	1602	117,7076	17,70757	1288	124,3789	24,38	94,6362968	124,3789		
8				1602			117,707568			

f) Izračun indeksa prosjek=100:

Prosjek=100 znači da je baza indeksa prosjek.
Prvo se izračuna prosjek proizvodnje:

- Označiti varijablu Yt

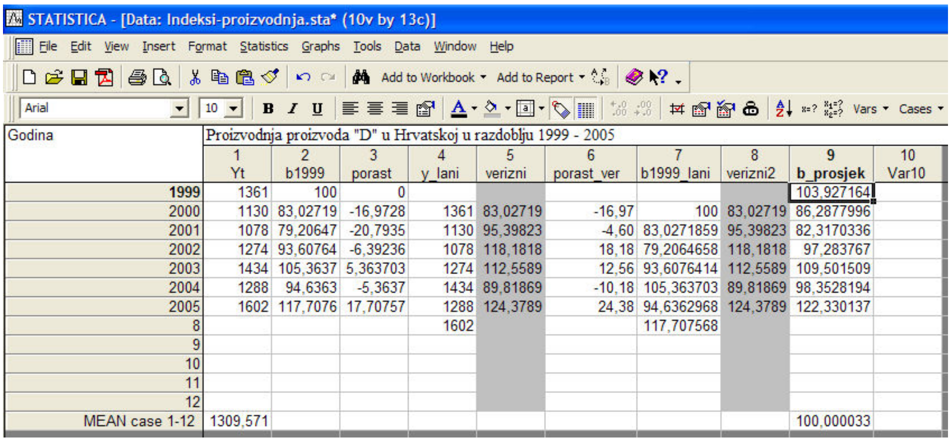
➢ **STATISTICS / STATISTICS OF BLOCK DATA / BLOCK COLUMNS / MEANS**

Godina	Proizvodnj 1 Yt
1999	1361
2000	1130
2001	1078
2002	1274
2003	1434
2004	1288
2005	1602
8	
9	
10	
11	
12	
MEAN case 1-12	1309,571

Rezultat:
$\overline{Y}$ = 1309,571 = baza

Prosjek je zapisan u novome, 13., retku. Kod kreiranja novih varijabli izračunat će se i vrijednosti za taj redak, ali one ne spadaju u rješenje zadatka.

- **Definirati novi stupac – *b_prosjek***
- **Long name:** ***=Yt/1309,571*100***



STATISTICA - [Data: Indeksi-proizvodnja.sta* (10v by 13c)]

Godina	1 Yt	2 b1999	3 porast	4 y_lani	5 verizni	6 porast_ver	7 b1999_lani	8 verizni2	9 b_prosjek	10 Var10
	Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005									
1999	1361	100	0						103,927164	
2000	1130	83,02719	-16,9728	1361	83,02719	-16,97	100	83,02719	86,2877996	
2001	1078	79,20647	-20,7935	1130	95,39823	-4,60	83,0271859	95,39823	82,3170336	
2002	1274	93,60764	-6,39236	1078	118,1818	18,18	79,2064658	118,1818	97,283767	
2003	1434	105,3637	5,363703	1274	112,5589	12,56	93,6076414	112,5589	109,501509	
2004	1288	94,6363	-5,3637	1434	89,81869	-10,18	105,363703	89,81869	98,3528194	
2005	1602	117,7076	17,70757	1288	124,3789	24,38	94,6362968	124,3789	122,330137	
8				1602			117,707568			
9										
10										
11										
12										
MEAN case 1-12	1309,571								100,000033	

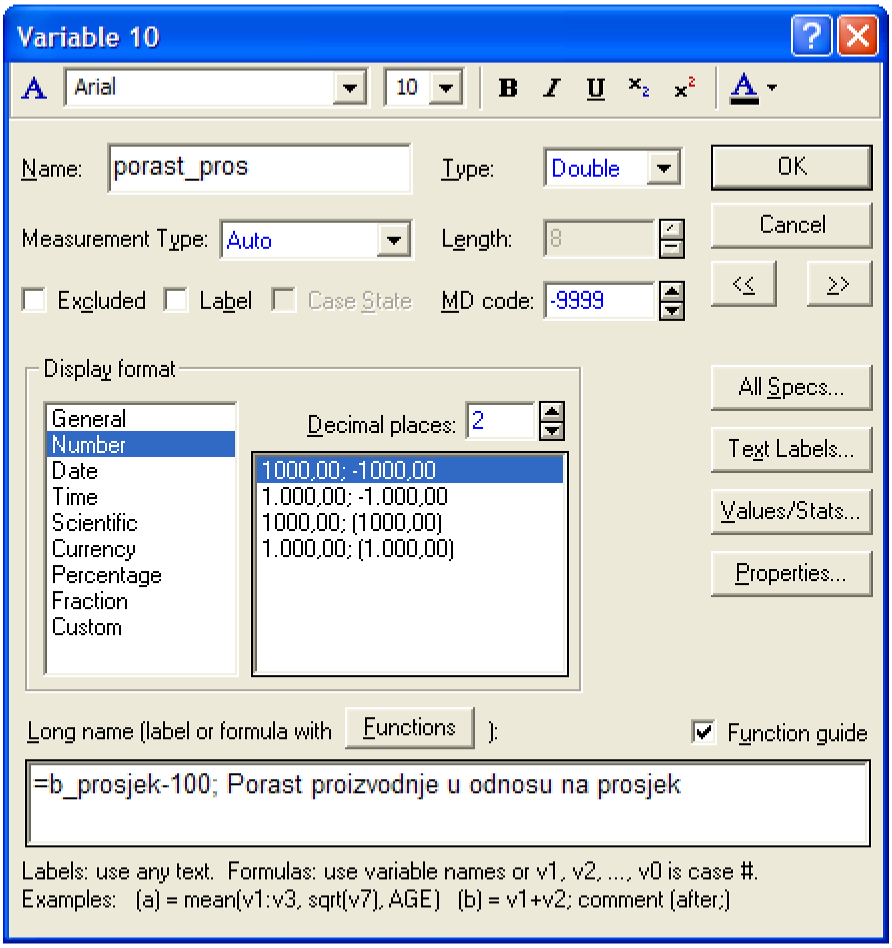
g) Tumačenje dobivenih vrijednosti:

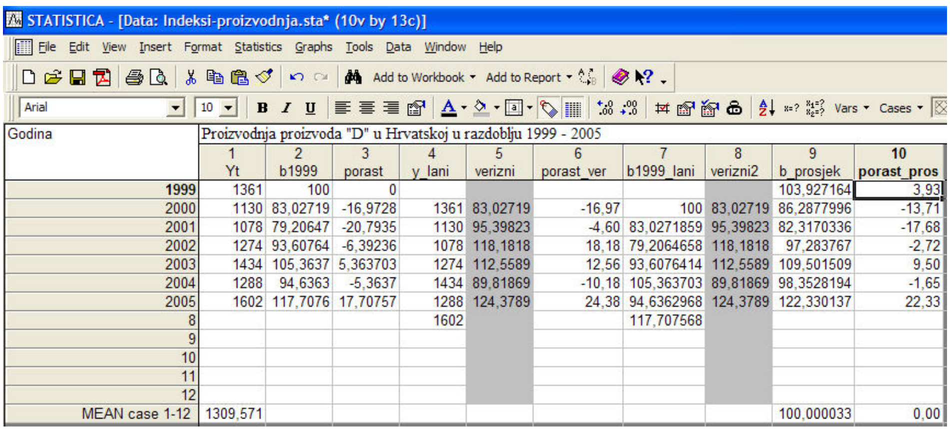
- **Definirati novi stupac – *porast_pros***
- **Long name:**

=b_prosjek-100;Porast proizvodnje u odnosu na prosjek

Varijabla se može zaokružiti na 2 decimale:

- **Display format** ***Number***
  - **Decimal places** ***2***





Proizvodnja proizvoda "D" u Hrvatskoj u razdoblju 1999 - 2005

Godina	1 Yt	2 b1999	3 porast	4 y_lani	5 verizni	6 porast_ver	7 b1999_lani	8 verizni2	9 b_prosjek	10 porast_pros
1999	1361	100	0						103,927164	3,93
2000	1130	83,02719	-16,9728	1361	83,02719	-16,97	100	83,02719	86,2877996	-13,71
2001	1078	79,20647	-20,7935	1130	95,39823	-4,60	83,0271859	95,39823	82,3170336	-17,68
2002	1274	93,60764	-6,39236	1078	118,1818	18,18	79,2064658	118,1818	97,283767	-2,72
2003	1434	105,3637	5,363703	1274	112,5589	12,56	93,6076414	112,5589	109,501509	9,50
2004	1288	94,6363	-5,3637	1434	89,81869	-10,18	105,363703	89,81869	98,3528194	-1,65
2005	1602	117,7076	17,70757	1288	124,3789	24,38	94,6362968	124,3789	122,330137	22,33
8				1602			117,707568			
9										
10										
11										
12										
MEAN case 1-12	1309,571								100,000033	0,00

U 1999. god. proizvodnja je bila za 3,93% veća od prosječne proizvodnje zadanog razdoblja.

Primjer 5.2.

Gubici energije u prijenosu i distribuciji u razdoblju od 1998. do 2005. god. na području "X"

Godina	*Gubici energije u GWh*
1998.	1804
1999.	1711
2000.	2214
2001.	1953
2002.	1974
2003.	2351
2004.	1962
2005.	2432

Izvor: Energetski institut "X" (simulirani podaci)

Zadatak je:

a) izračunati indekse 1998.=100 i odrediti koliko iznosi indeks za 2001. god.
b) utvrditi koliko je povećan gubitak 2001. god. u odnosu na 1998. god.
c) izračunati indekse 2000.=100 i odrediti koliko iznosi indeks za 2004. god.
d) utvrditi koliko je promijenjen gubitak 2004. god. u odnosu na 2000. god.
e) izračunati verižne indekse i odrediti koliko iznosi indeks za 2004. god;
f) odrediti koliko je promijenjen gubitak 2004. god. u odnosu na prethodnu godinu
g) preračunati indekse 1998.=100 u verižne i odrediti koliko iznosi indeks za 2003. god.

Rezultati:

a) 108,26
b) Za 8,26%
c) 88,62
d) Gubitak se smanjio za 11,38%
e) 83,45
f) Smanjen je za 16,55%
g) 119,10.

5.2. Linearni trend

Primjer 5.3.

Izvoz zemlje "B" u razdoblju 2002.-2008. god. u milijunima USD

Godina	*Izvoz u mil. USD*
2002.	6616
2003.	6534
2004.	7204
2005.	7349
2006.	6946
2007.	7856
2008.	8136

Izvor: Zavod za Statistiku zemlje B (simulirani podaci)

Zadatak je:

a) ocijeniti jednadžbu linearnog trenda
b) izvršiti interpolaciju
c) procijeniti koliki će biti izvoz prema trendu u 2011. godini.

a) Ocjena jednadžbe linearnog trenda:

Jednadžba trenda računa se na isti način kao i jednadžba regresije, s time da je nezavisna varijabla vremenska jedinica.

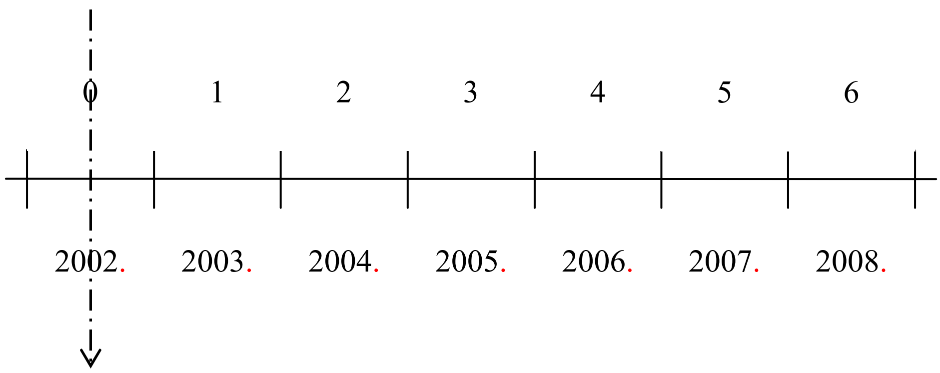
- U **zaglavlje** dokumenta upisati:

Izvoz zemlje "B" u razdoblju od 2002. do 2008. u milijunima USD

Stupac "godine" može se unijeti kao pred stupac:

- U informativno polje upisati ***Godina***
- Definirati varijablu:
  - Kratki naziv: ***Izvoz***, dugi naziv: ***Izvoz u milijunima USD***

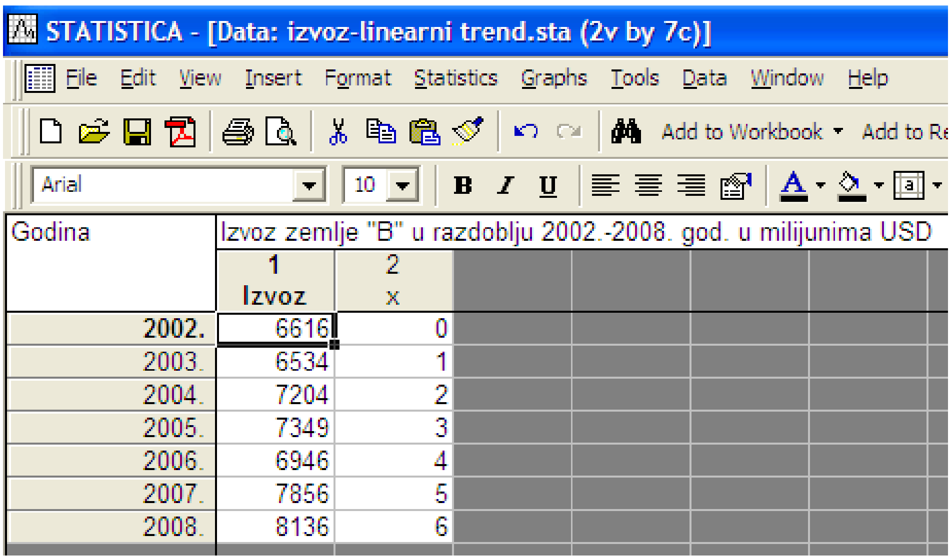
Vremenska varijabla određuje se na sljedeći način:



30.06.2002.

- Definirati varijablu:
  - Kratki naziv: $\boxed{X}$, dugi naziv: $\boxed{\textbf{\textit{Vremenska varijabla}}}$

- Unijeti vrijednosti:



Izvoz zemlje "B" u razdoblju 2002.-2008. god. u milijunima USD

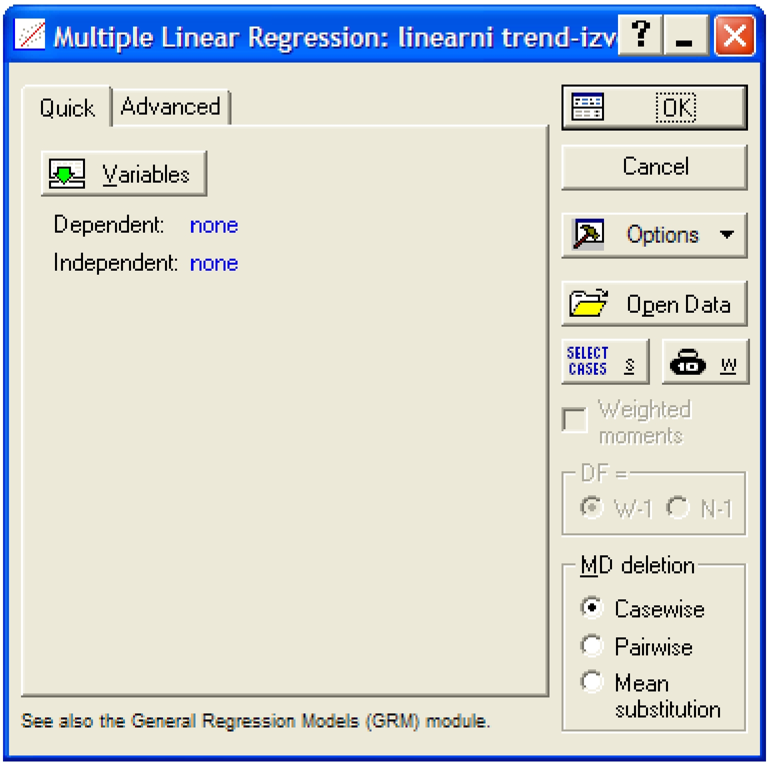
Godina	1 Izvoz	2 x
2002.	6616	0
2003.	6534	1
2004.	7204	2
2005.	7349	3
2006.	6946	4
2007.	7856	5
2008.	8136	6

Ocijenjeni model linearnog trenda glasi:

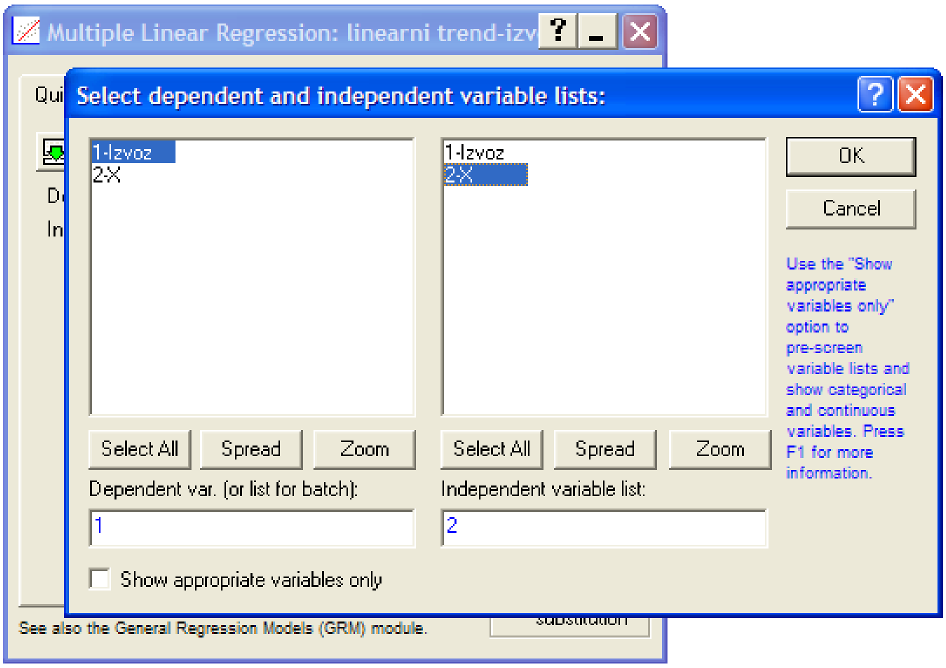
$$
\begin{aligned}
&\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X, \\
&jedinica \ \ za \ \ X = ... \\
&jedinica \ \ za \ \ Y = ... \\
&X = 0, \ \ ......
\end{aligned}
$$

Postupak za rješavanje zadatka:

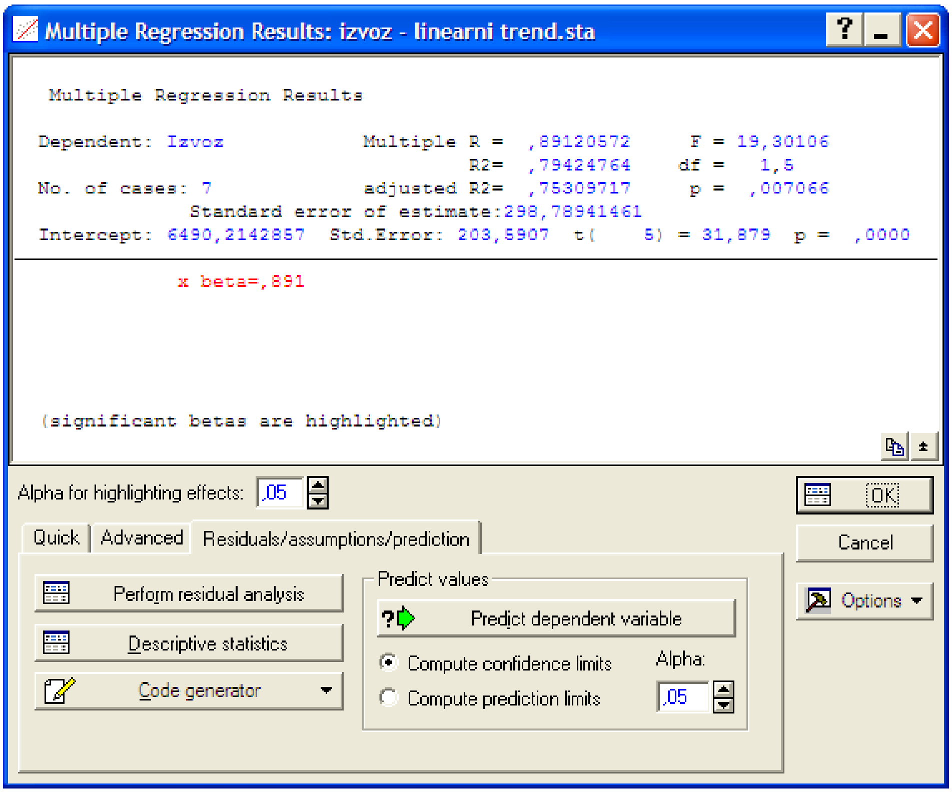
- **STATISTICS / MULTIPLE REGRESSION**
  - **QUICK**



- **Variables**
  - **Dependenet:** ***1-Izvoz*** **(zavisna varijabla)**
  - **Independenet:** ***2-X*** **(nezavisna varijabla)**



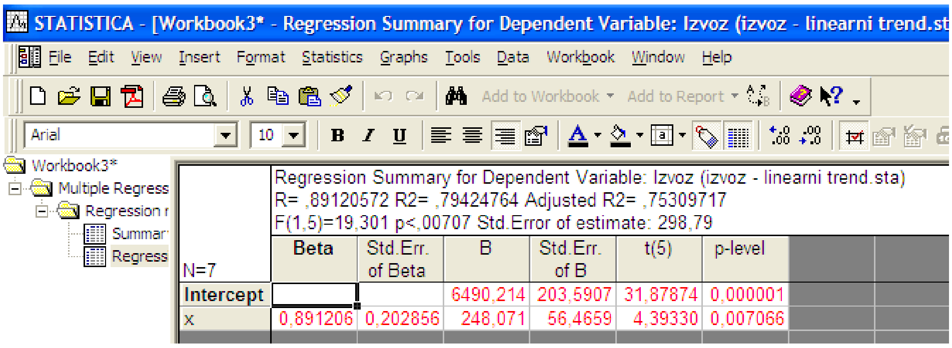
- **OK**



Za prikaz rezultata u prozoru radnih zapisa izabere se:

- **QUICK**
  - **Summary: Regression Results**

Rezultati:



Jednadžba trenda postoji samo ukoliko je:

- veza dovoljno jaka ($\overline{R}^2$ = **Adjusted R^2 > 0,6)** i
- mala je vjerojatnost da su rezultati dobiveni slučajno **(p < 0,05).**

U zadatku je :

$\overline{R}^2$ = 0,75309717 > **0,6**

p < 0,00707 < **0,05**

Može se odrediti jednadžba trenda.

N=7	Beta	Std.Err. of Beta	B	Std.Err. of B	t(5)	p-level
Regression Summary for Dependent Variable: Izvoz (izvoz - linearni trend.sta) R= ,89120572 R2= ,79424764 Adjusted R2= ,75309717 F(1,5)=19,301 p<,00707 Std.Error of estimate: 298,79						
Intercept			**β0=6490,214**	203,5907	31,87874	0,000001
x	0,891206	0,202856	**β1=248,971**	56,4659	4,39330	0,007066

Jednadžba linearnog trenda glasi:

$$\hat{Y} = 6490{,}214 + 248{,}971\,X$$

jedinica za X = 1 godina

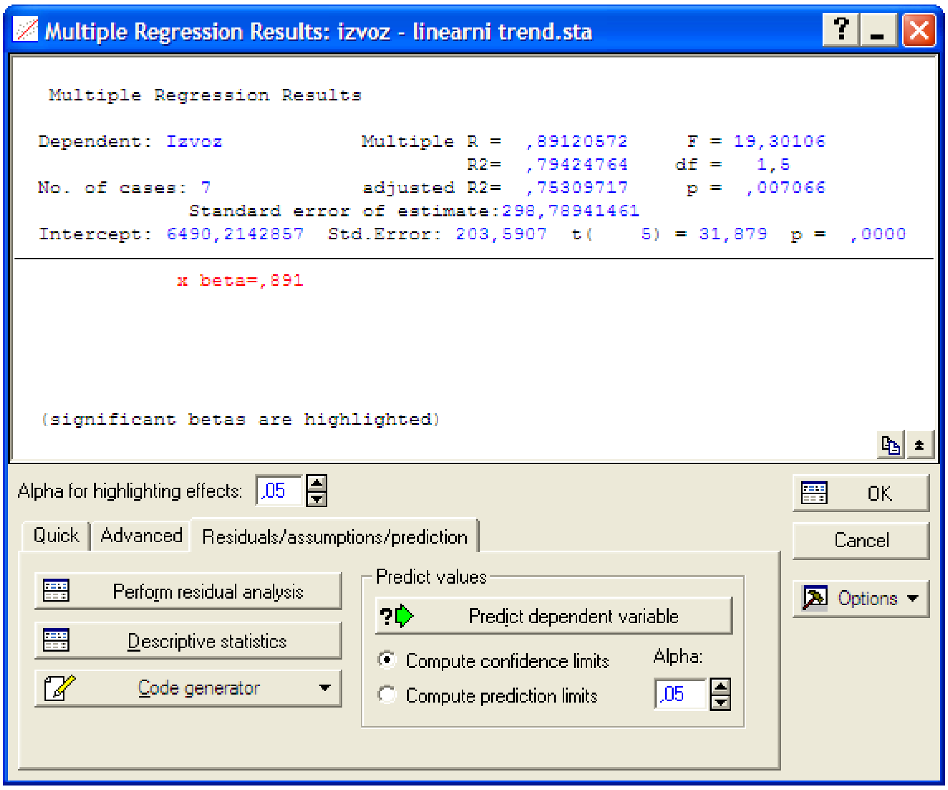
jedinica za Y = izvoz u milijunima USD

X=0, 30.06.2002. godina

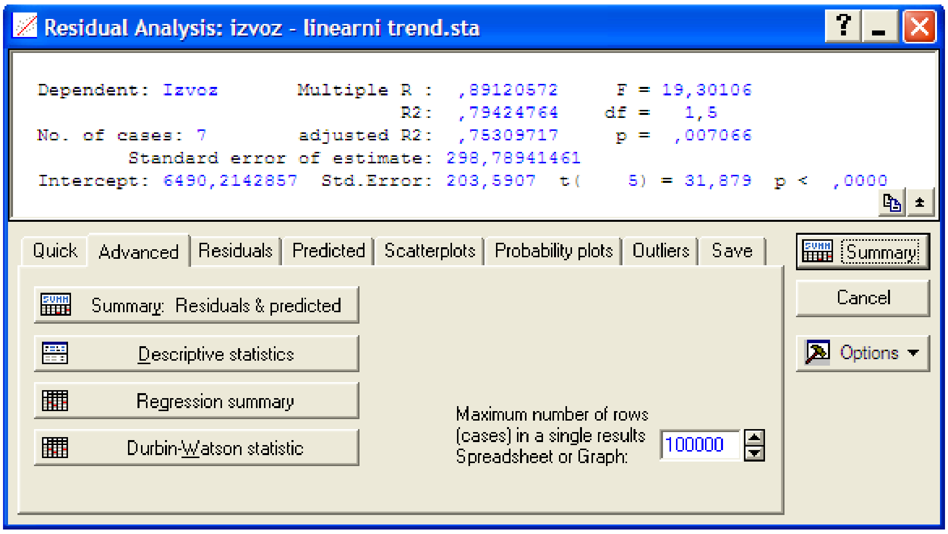
b) Izvršiti interpolaciju

Interpolacija se vrši na isti način kao i kod regresije. Započinje se kao i kod računanja parametara:

- **STATISTICS / MULTIPLE REGRESSION**
  - **QUICK**
    - **Variables**
      - **Dependenet:** *1-Izvoz*
      - **Independenet:** *2-X*
  - **OK**



➢ RESIDUAL/ASSUMPTIONS/PREDICTION



➢ PERFORM RESIDUAL ANALISYS

➢ ADVANCED

➢ Summary: residuals & predicted

Case No.	Predicted & Residual Values (izvoz - linearni trend.sta) Dependent variable: Izvoz Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
2002.	6616,000	**6490,214**	125,786	-1,38873	0,42098	203,5907	1,928571	234,800	0,143358
2003.	6534,000	**6738,286**	-204,286	-0,92582	-0,68371	159,7097	0,857143	-286,000	0,130889
2004.	7204,000	**6986,357**	217,643	-0,46291	0,72842	126,2616	0,214286	264,957	0,070210
2005.	7349,000	**7234,429**	114,571	0,00000	0,38345	112,9318	0,000000	133,667	0,014295
2006.	6946,000	**7482,500**	-536,500	0,46291	-1,79558	126,2616	0,214286	-653,130	0,426630
2007.	7856,000	**7730,571**	125,429	0,92582	0,41979	159,7097	0,857143	175,600	0,049343
2008.	8136,000	**7978,643**	157,357	1,38873	0,52665	203,5907	1,928571	293,733	0,224352
Minimum	6534,000	6490,214	-536,500	-1,38873	-1,79558	112,9318	0,000000	-653,130	0,014295
Maximum	8136,000	7978,643	217,643	1,38873	0,72842	203,5907	1,928571	293,733	0,426630
Mean	7234,429	7234,429	0,000	0,00000	0,00000	156,0079	0,857143	23,375	0,151297
Median	7204,000	7234,429	125,429	0,00000	0,41979	159,7097	0,857143	175,600	0,130889

Rezultati su zapisani u stupcu *Predicted Value*:

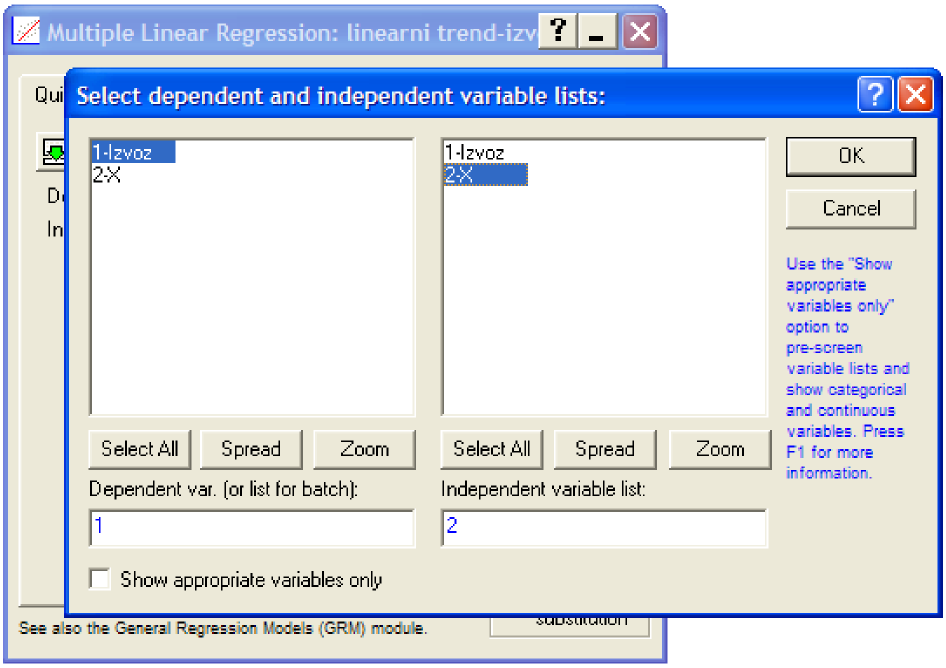
Izvoz bi prema trendu iznosio 6490,2 milijuna USD za 2002. godinu.

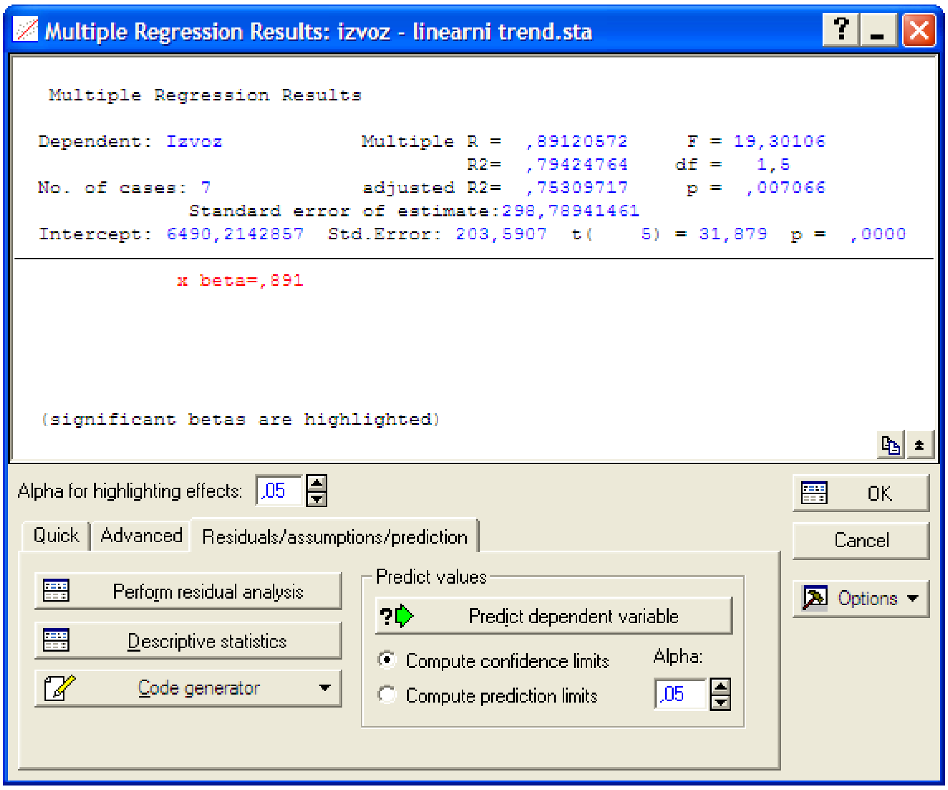
c) Procijeniti koliki će biti izvoz prema trendu u 2011. godini

Ekstrapolacija vrijedi samo ako je trend reprezentativan. Računanje započinje na isti način kao pri računanju parametara jednadžbe trenda.

$X_{2011} = 9$

- **STATISTICS / MULTIPLE REGRESSION /**
  - **QUICK**
    - **Variables**
      - **Dependenet:** ***1-Izvoz***
      - **Independenet:** ***2-X***

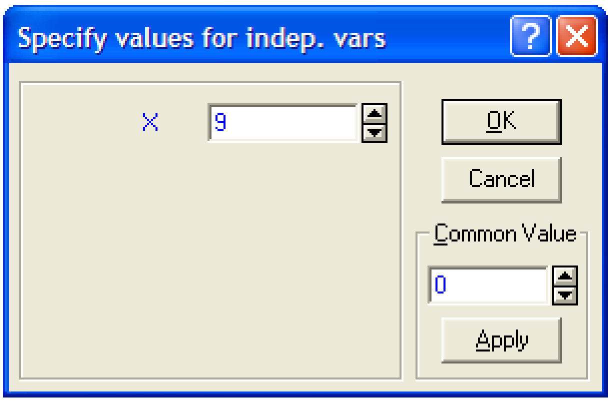




➢ OK

➢ **RESIDUAL/ASSUMPTIONS/PREDICTION**

➢ **Predict dependent variable**



➢ **Specifiy values for indep. vars: x 9**

Rezultati:

Variable	Predicting Values for (izvoz - linearni trend.sta) variable: Izvoz B-Weight	Value	B-Weight * Value
x	248,0714	9,000000	2232,643
Intercept			6490,214
Predicted			8722,857
-95,0%CL			**7804,847**
+95,0%CL			**9640,868**

Na razini 95% pouzdanosti procjenjuje se da će izvoz zemlje "B" u 2011. godini iznositi između 7804,8 i 9640,9 milijuna USD.

Primjer 5.4.

Proizvodnja proizvoda „A" na području "Z" u razdoblju od 1998. do 2008. god.

Godina	Proizvodnja čelika u tis. t
1998.	1741
1999.	1645
2000.	1972
2001.	1921
2002.	2215
2003.	2068
2004.	2297
2005.	2268
2006.	2548
2007.	2415
2008.	2746

Izvor: Zavod za statistiku područja Z (simulirani podaci)

Zadatak je:

a) Odrediti linearni trend
b) Odrediti interpolaciju za 2004. godinu
c) Procijeniti koliki će biti izvoz prema trendu 2014. godine

Rezultati:

a)

$\hat{Y} = 1684{,}682 + 96{,}445\ X$

jedinica za X = 1 godina
jedinica za Y = tisuću T čelika
X=0, 30.06.2002.

b) $\hat{Y}_{2004} = 2263{,}354$ tisuća t

c) $2954{,}364 \leq \hat{Y}_{2014} \leq 3501{,}255$ tis. t, tvrdnja uz 5% rizika

5.3. Trend polinom drugog stupnja

Primjer 5.5.

Uvoz zemlje C u milijunima USD u razdoblju 1997. – 2005. god.

Godina	Uvoz (u milijunima USD)
2000.	6825
2001.	6885
2002.	8671
2003.	9609
2004.	9642
2005.	9762
2006.	10173
2007.	8710
2008.	10120

Izvor: Zavod za statistiku zemlje C (simulirani podaci)

Zadatak je:

a) Izračunati trend polinom drugog stupnja
b) Izvršiti interpolaciju

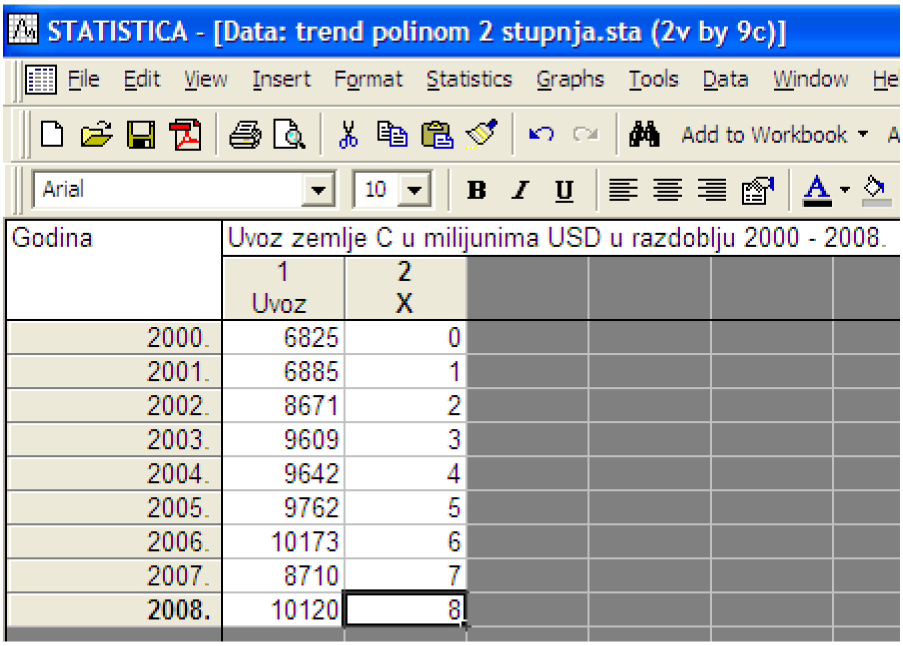
a) Izračun paraboličnog trenda drugog stupnja

- U **zaglavlje** dokumenta upisati:

  Uvoz zemlje C u milijunima USD u razdoblju 2000. – 2008. god.

Stupac godine može se unijeti kao pred stupac:

- U informativno polje upisati ***Godina***
- Definirati varijable:
  - Kratki naziv: ***Uvoz***, dugi naziv: ***Uvoz u milijunima USD***
  - Kratki naziv: ***X***, dugi naziv: ***Vremenska varijabla***



Godina	1 Uvoz	2 X
2000.	6825	0
2001.	6885	1
2002.	8671	2
2003.	9609	3
2004.	9642	4
2005.	9762	5
2006.	10173	6
2007.	8710	7
2008.	10120	8

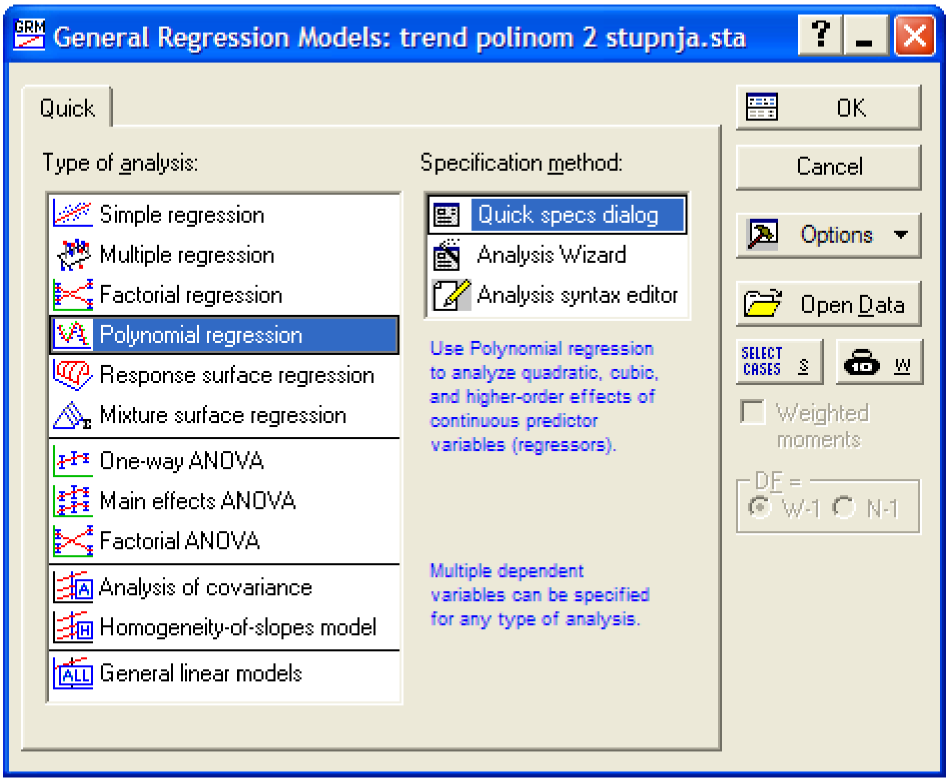
Opći oblik jednadžbe trend polinoma drugog stupnja je:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X + \hat{\beta}_2 \cdot X$$
$$jedinica \;\; za \;\;\; X = ...$$
$$jedinica \;\;\; za \;\;\; Y = ...$$
$$X = 0, \;\; ......$$

Računa se:
- reprezentativnost trenda i
- parametri jednadžbe

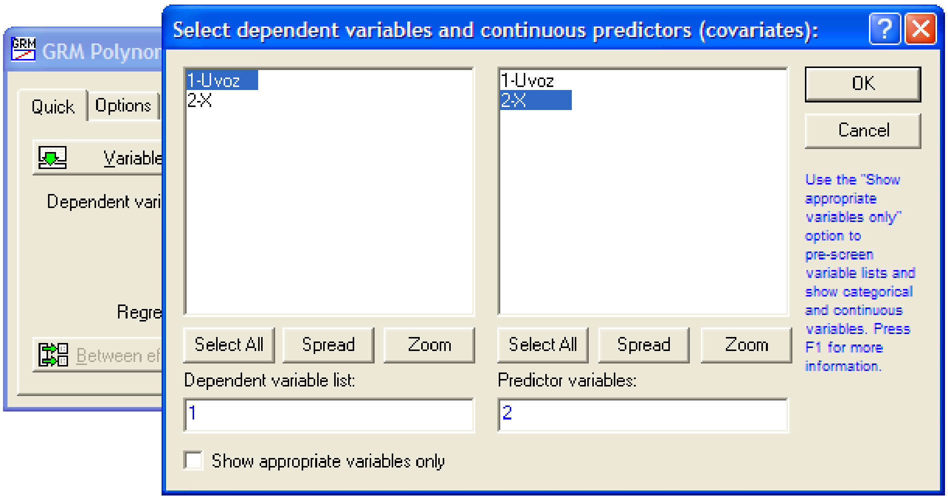
Tek ako je trend reprezentativan može se napisati jednadžba trenda.

- **STATISTICS / ADVANCED LINEAR/NONLINEAR MODELS / GENERAL REGRESSION MODELS**
  - **QUICK**

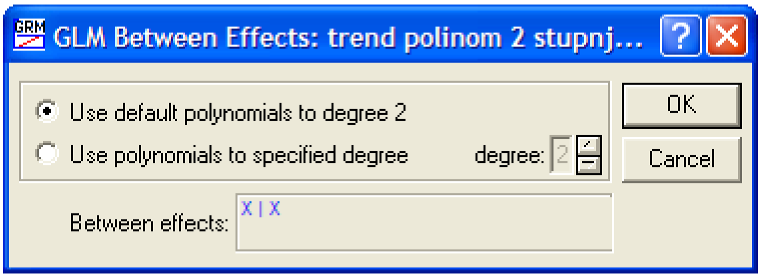


- **POLYNOMIAL REGRESSION**
  - **Quick**

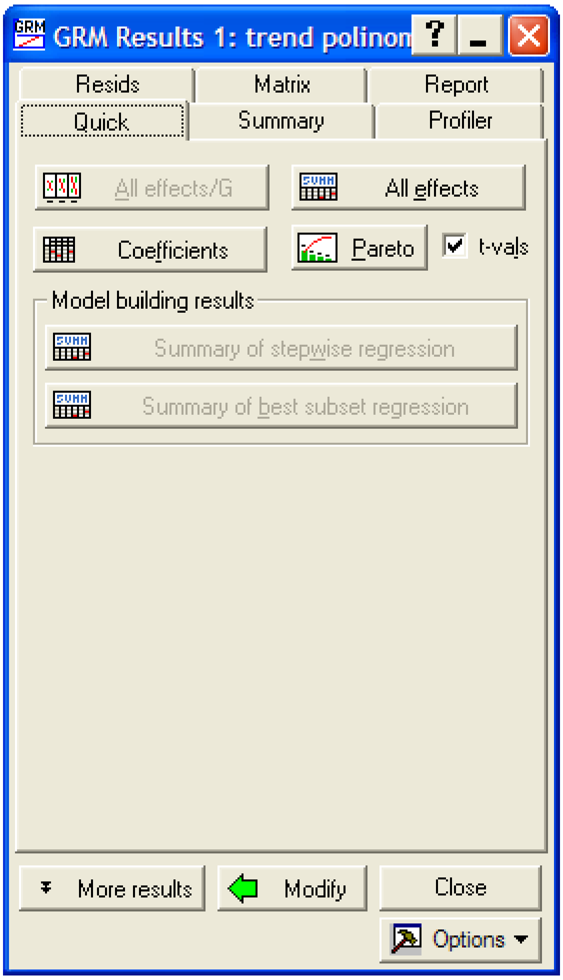
- **Variables**
  - **Dependent:** ***1-Uvoz*** **(zavisna varijabla)**
  - **Predictor variables:** ***2-X*** **(nezavisna varijabla)**



- **Between effects**
  - **Use default polynomials to degree 2**



- ***OK***



- **Quick**
  - **All effects**

Dependnt Variable	Test of SS Whole Model vs. SS Residual (trend polinom 2 stupnja.sta) Multiple R	Multiple R2	Adjusted R2	SS Model	df Model	MS Model	SS Residual
Uvoz	0,899632	0,809339	**0,745785**	10804514	2	5402257	2545294

(Nastavak ispisa)

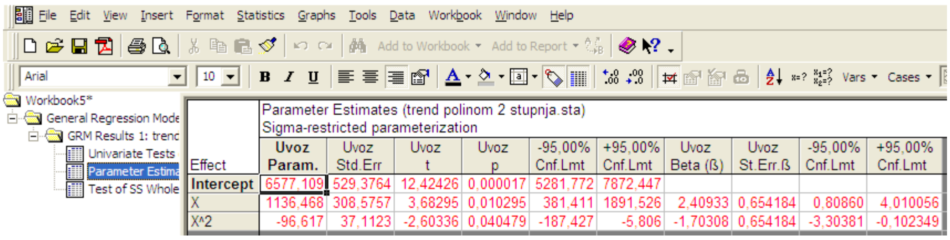
Dependnt Variable	Test of SS Whole Model vs. SS Residual (trend polinom 2 stupnja.sta) df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
Uvoz	2	5402257	2545294	6	424215,6	12,73470	**0,006931**

Trend je reprezentativan:

$\overline{R}^2 = 0{,}745785 > \mathbf{0{,}6}$
P = 0,006931 < **0,05**

Istovremeno su se izračunali i parametri jednadžbe trenda.

Lijevi dio prozora u rezultatima zapisa služi za navigaciju među rezultatima.



Effect	Parameter Estimates (trend polinom 2 stupnja.sta) Sigma-restricted parameterization Uvoz Param.	Uvoz Std.Err	Uvoz t	Uvoz p	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt	Uvoz Beta (ß)	Uvoz St.Err.ß	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
Intercept	6577,109	529,3764	12,42426	0,000017	5281,772	7872,447				
X	1136,468	308,5757	3,68295	0,010295	381,411	1891,526	2,40933	0,654184	0,80860	4,010056
X^2	-96,617	37,1123	-2,60336	0,040479	-187,427	-5,806	-1,70308	0,654184	-3,30381	-0,102349

- Izborom rubrike ***Parameter Estimates*** prelazi se u prozor s izračunatim parametrima modela. (Rubrika *Test of SS Whole Model vs. SS Residual*, sadrži korištene podatke o reprezentativnosti trenda u cjelini).

Effect	Parameter Estimates (trend polinom 2 stupnja.sta) Sigma-restricted parameterization Uvoz Param.	Uvoz Std.Err	Uvoz t	Uvoz p	-95,00% Cnf.Lmt	+95 Cnf
Intercept	**β0=6577,109**	529,3764	12,42426	0,000017	5281,772	787
X	**β1=1136,468**	308,5757	3,68295	0,010295	381,411	189
X^2	**β2=-96,617**	37,1123	-2,60336	0,040479	-187,427	-

Parametri su reprezentativni. Jednadžba trend polinoma drugog stupnja glasi:

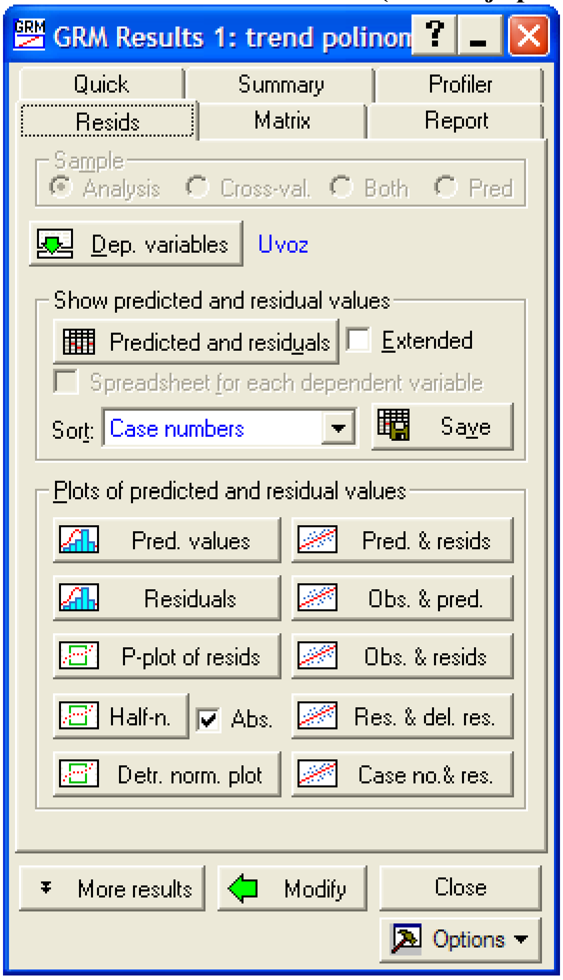
$$\hat{Y} = 6577{,}109 + 1136{,}468\ X - 96{,}617\ X^2$$

jedinica za X = 1 godina
jedinica za Y = uvoz u milijunima USD
X=0, 30.06.2000.

b) Interpolacija

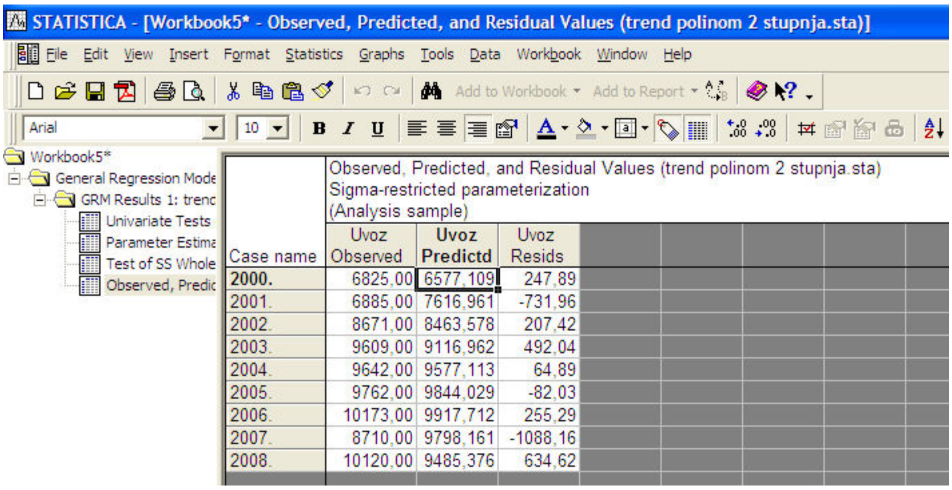
Trend je reprezentativan, pa se može izvršiti interpolacija.
Dovoljno je otvoriti zadnji korišteni prozor

- ....
- **STATISTICS / RESUME (otvaranje posljednjeg rabljenog prozora)**



- **Resids**
    - **Predicted and residuals**

Rezultati:



Case name	Observed, Predicted, and Residual Values (trend polinom 2 stupnja.sta) Sigma-restricted parameterization (Analysis sample) Uvoz Observed	Uvoz Predictd	Uvoz Resids
2000.	6825,00	6577,109	247,89
2001.	6885,00	7616,961	-731,96
2002.	8671,00	8463,578	207,42
2003.	9609,00	9116,962	492,04
2004.	9642,00	9577,113	64,89
2005.	9762,00	9844,029	-82,03
2006.	10173,00	9917,712	255,29
2007.	8710,00	9798,161	-1088,16
2008.	10120,00	9485,376	634,62

Interpolacija je zapisana u rubrici *Predictd*.

Case name	Observed, Predicted, and Residual Values (trend po Sigma-restricted parameterization (Analysis sample) Uvoz Observed	Uvoz Predictd	Uvoz Resids
2000.	6825,00	**6577,109**	247,89
2001.	6885,00	**7616,961**	-731,96
2002.	8671,00	**8463,578**	207,42
2003.	9609,00	**9116,962**	492,04
2004.	9642,00	**9577,113**	64,89
2005.	9762,00	**9844,029**	-82,03
2006.	10173,00	**9917,712**	255,29
2007.	8710,00	**9798,161**	-1088,16
2008.	10120,00	**9485,376**	634,62

Trend vrijednost za 2000. godinu je 6577,109 milijuna USD.
Vrijednost trenda u 2001. godini je 7616,961 milijuna USD itd.
...

Primjer 5.6.

Proizvodnja proizvoda “Y” u razdoblju od 1999. do 2008. god. u milijunima kn

Godina	Proizvodnja (u milijunima kn)
1999.	5297
2000.	5394
2001.	6655
2002.	7135
2003.	7408
2004.	7565
2005.	7822
2006.	8478
2007.	8194
2008.	8772

Izvor: Tvornica Y (simulirano)

Zadatak je:

a) Odrediti trend polinom drugog stupnja
b) Odrediti interpolaciju za 2007. godinu

Rezultati:

a)

$$\hat{Y} = 5197{,}873 + 660{,}177\,X - 31{,}462\,X^2$$

jedinica za X = 1 godina
jedinica za Y = proizvodnja u milijunima kn
X=0, 30.06.1999.

b) $\hat{Y}_{2007} = 8465{,}715$ milijuna kn

5.4. Pomični prosjeci

Primjer 5.7.

Prodaja proizvoda „B“ u tonama u radoblju 2000. do 2008 . god. na području “X”

Godina	**Broj eutanaziranih svinja**
2000.	138
2001.	122
2002.	84
2003.	158
2004.	144
2005.	171
2006.	102
2007.	91
2008.	148

Izvor: Zavod za statistiku područja „X“ (simulirani podaci)

a) Odredite jednostavni krivolinijski trend pomoću pomičnih prosjeka

- U **zaglavlje** dokumenta upisati:

Prodaja proizvoda „B“ u tonama u radoblju 2000. do 2008 . god. na području “X”

Stupac “Godina” može se unijeti kao pred stupac:

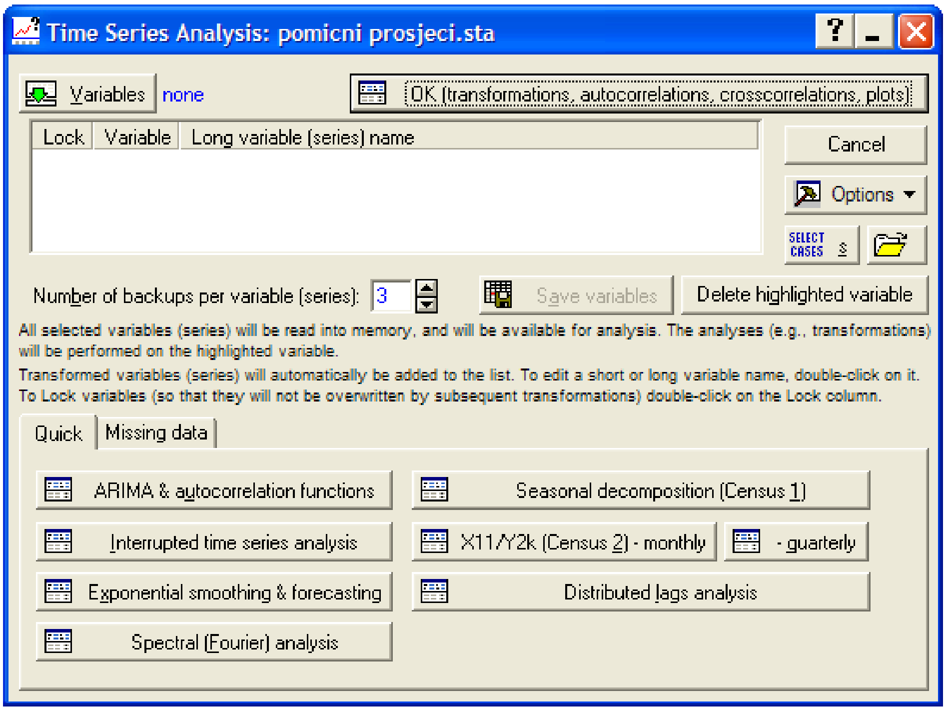
- U informativno polje upisati ***Godina***

- Definirati varijablu:
  - Kratki naziv: ***Prodaja***,

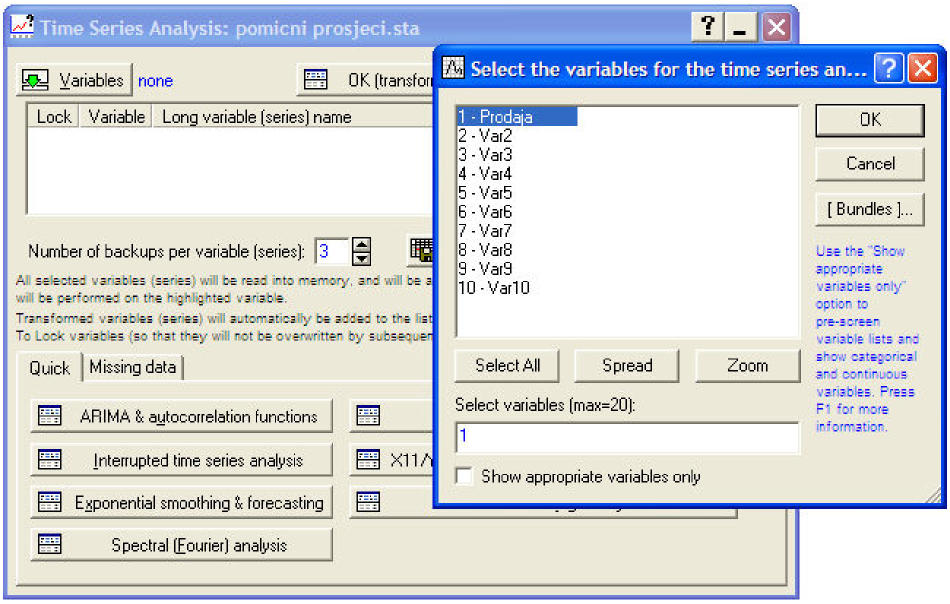
    dugi naziv: ***Prodaja proizvoda „B“ u tonama***

➢ **STATISTICS / ADVANCED LINEAR/NONLINEAR MODELS / TIME SERIES/FORECASTING**

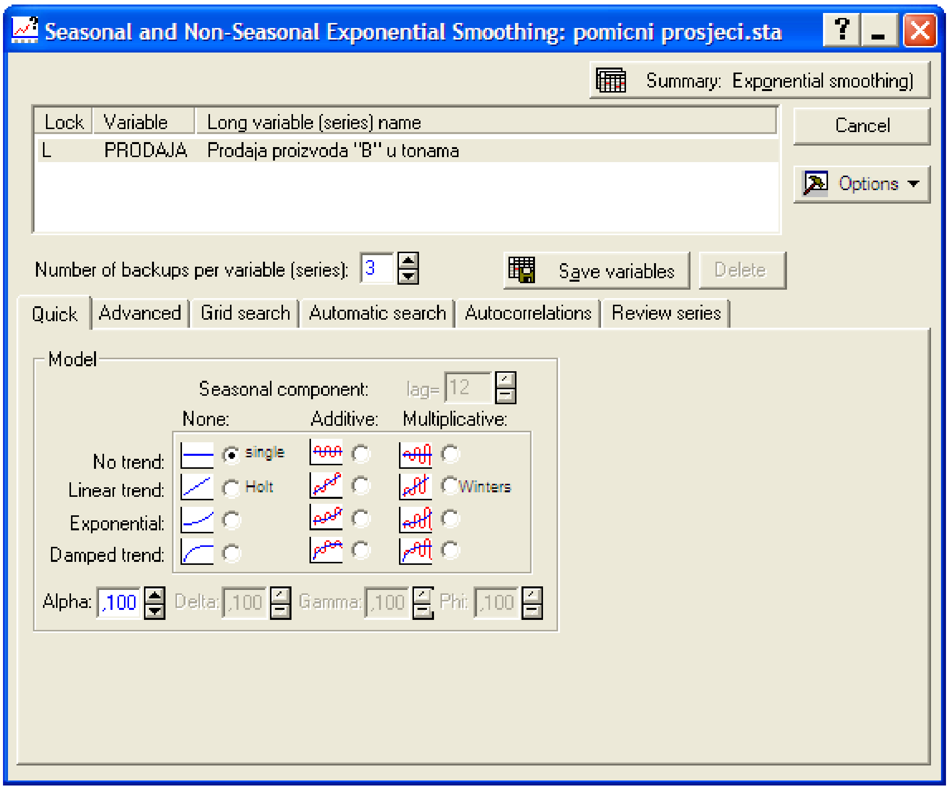
Otvara se prozor za izračun krivolinijskog trenda.



➢ **Variables** ***1- Prodaja***



- **Exponential smoothing & forecasting**
  - **Quick**



U prozoru Quick izabere se model eksponencijalnog izglađivanja niza

- ➢ ***No trend*** **(single) – bez trenda, jednostavno**
- ➢ ***Linear trend*** **(Holt) – s linearnim trendom**
- ➢ ***Exponential*** – s **eksponencijalnim trendom**
- ➢ ***Damped trend*** – **utjecaj trenda se smanjuje s vremenom**

Također je moguće uračunati sezonsku komponentu.

Izabrati jednostavno izglađivanje vremenskog niza, bez trenda:

➢ **No trend (single)**

Rezultati:

Case	Exponential smoothing: S0=128,7 (pomicni prosjeci.sta) No trend,no season; Alpha= ,100 PRODAJA : Prodaja proizvoda "B" u tonama		
	PRODAJA	Smoothed Series	Resids
1	138,0000	**128,6667**	9,3333
2	122,0000	**129,6000**	-7,6000
3	84,0000	**128,8400**	-44,8400
4	158,0000	**124,3560**	33,6440
5	144,0000	**127,7204**	16,2796
6	171,0000	**129,3484**	41,6516
7	102,0000	**133,5135**	-31,5135
8	91,0000	**130,3622**	-39,3622
9	148,0000	**126,4260**	21,5740
10		128,5834	
11		128,5834	
12		128,5834	

Rezultati se nalaze u stupcu: *Smoothed Series*. Brojke koje su zapisane u kasnijim redcima su projekcije za budućnost, ali nisu reprezentativne.

5.5. Sezonska dekompozicija vremenskog niza

Primjer 5.8.

Mjesečni prosjeci srednjeg deviznog tečaja Hrvatske narodne banke za EUR 1997. – 2003. god.

God./mj.	*1*	*2*	*3*	*4*	*5*	*6*
1997.	6,906000	6,940200	6,948000	6,989500	6,941000	6,997200
1998.	6,936200	6,974000	7,035200	7,069200	7,146000	7,203800
1999.	7,387139	7,567448	7,596698	7,591112	7,591861	7,596586
2000.	7,720042	7,730457	7,727280	7,710201	7,683790	7,639108
2001.	7,675722	7,703111	7,680701	7,526926	7,278507	7,320680
2002.	7,568480	7,437848	7,402535	7,395818	7,377827	7,320722
2003.	7,555767	7,620482	7,692318	7,567308	7,546434	7,508844

God./mj.	7	8	9	10	11	12
1997.	100,9408	100,3987	99,7651	99,8072	100,3697	100,2350
1998.	99,8215	99,1672	100,3706	101,1151	100,1417	99,5963
1999.	99,9183	99,9537	100,1552	100,0052	99,8051	100,2180
2000.	100,0170	100,0866	99,5682	99,7673	99,7966	100,1944
2001.	97,0205	102,9684	101,8713	100,3085	100,1084	99,7843
2002.	100,2837	100,2358	99,6603	101,3438	99,8800	99,8567
2003.	99,9275	98,9594	100,1955	99,9126	100,1749	100,1176

Izvor: http://www.hnb.hr/publikac/bilten/statisticki_pregled/h8.xls

Zadatak je:

a) Izvršiti sezonsku dekompoziciju zadanog niza
b) Desezonirati niz

a) Sezonsko dekomponiranje niza

Sezonska dekompozicija je metoda kojom se izdvajaju pojedine komponente vremenskog niza. Model sezonske dekompozicije može biti aditivni ili multiplikativni.

Kod multiplikativnog modela vrijednost zavisne varijable Y_t dobije se kao umnožak sljedećih komponenti:

$$Y_t = T_t \cdot C_t \cdot S_t \cdot I_t$$

gdje je:
t – vremensko razdoblje
Y_t – vrijednost vremenskog niza u trenutku t
T_t – vrijednost trenda u trenutku t
C_t – vrijednost cikličke komponente u trenutku t
S_t – vrijednost sezonske komponente u trenutku t
I_t – vrijednost slučajne komponente u trenutku t.

Trend-komponenta – T_t

Trend pokazuje putanju rasta ili pada vrijednosti kroz duže vremensko razdoblje. Neki faktori koji dugoročno djeluju na trend su: kulturalne snage kao što je npr. veća prihvaćenost žena na poslu, demografske sile (npr. skokovi u veličini populacije, starenje baby boom generacije i posljedično promjene u navikama kupovanja), ekonomske sile (porast raspoloživog prihoda), tehnološke sile (ekspanzija tehnologije mikročipa). U ekonomskim podacima trend se smatra dugoročnim kretanjem koje traje nekoliko godina, općenito 3 do

15 godina. Crta se kao izglađena krivulja koja može imati linearni i krivolinijski oblik.

Sezonska komponenta – S_t

Javlja se u pravilnim sezonskim intervalima. Karakteristična je po ponavljanju iz godine u godinu. Glavni činitelji koji dovode do godišnjeg ponavljanja sheme sezonskih varijacija su vrijeme (klimatski uvjeti) i običaji, time se misli na razne praznike kao i običaje rabljenja godišnjih odmora u određenom razdoblju.

Ciklička komponenta – C_t

Tipičan poslovni ciklus sastoji se od razdoblja prosperiteta, nakon čega slijedi razdoblje recesije, depresije i oporavka. To su značajne fluktuacije koje traju više od godinu dana. U fazi recesije zaposlenost, proizvodnja, Dow Jones Industrial Average te mnogi drugi gospodarski nizovi nalaze se ispod krivulje trenda. Nasuprot tome, u razdoblju rasta ti nizovi su iznad, odnosno veći od trenda.

Slučajna komponenta – I_t

Predstavlja varijaciju u podacima ili nastaje zbog nepredvidljivih događaja kao što su štrajkovi, nevremena, politički nemiri, poplave, potresi i sl. Kratkog su trajanja i ne ponavljaju se.

U ***multiplikativnom modelu*** trend komponenta ima istu jedinicu mjere kao i izvorni niz Y_t. Sezonska, ciklička i slučajna komponenta izražene su relativno te vrijednost komponente veća od 100 znači da utječe da vrijednost niza bude veća od trenda, a ako je komponenta manja od 100 tada niz ima vrijednost manju od trenda.

Ciklička i trend komponenta često se spajaju u jedinstvenu trend-cikličku komponentu, te multiplikativni model tada glasi:

$$Y_t = TC_t \cdot S_t \cdot I_t$$

Prilikom unosa podataka, može se definirati

- zasebno varijabla “godina”, zasebno varijabla “mjesec” (ili “kvartal”) ili
- jedna vremenska varijabla datumskog tipa u koja će sadržavati mjesec i godinu
- te vremenski niz

Prikazan je način zapisa u odvojenim varijablama za mjesec i godinu

- U **zaglavlje** dokumenta upisati:

Mjesečni prosjeci srednjeg deviznog tečaja Hrvatske narodne banke za EUR 1997. – 2003. god.

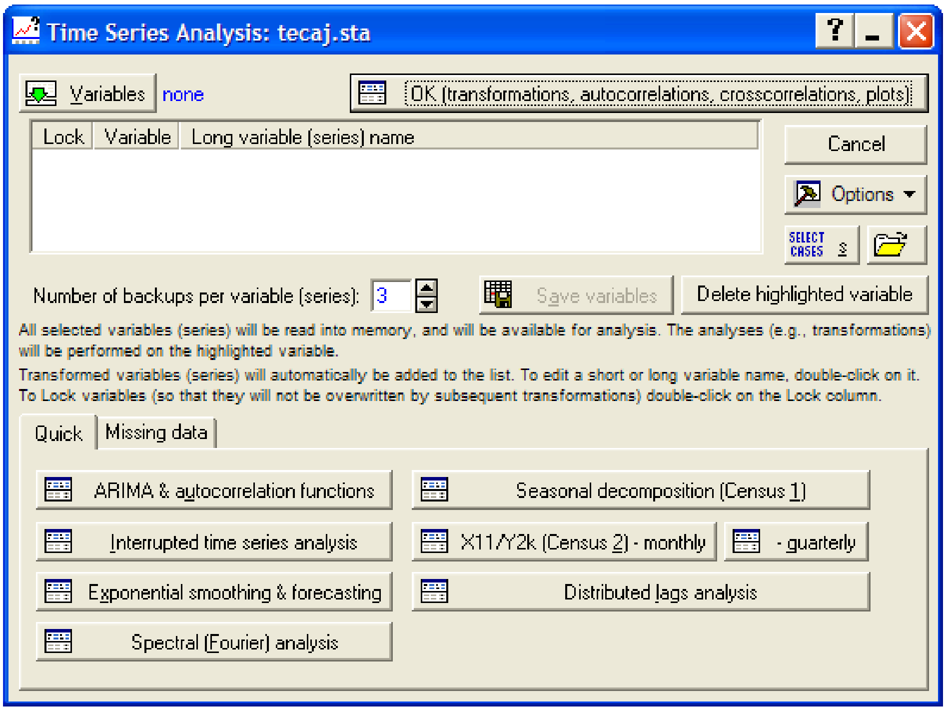
- Definirati varijable:
  - Kratki naziv: ***Godina***
  - Kratki naziv: ***Mjesec***,
  - Kratki naziv: ***Tecaj***, Dugi naziv ***Srednji tečaj EUR-a***

Podaci se unose na način kako je prikazano na slici:

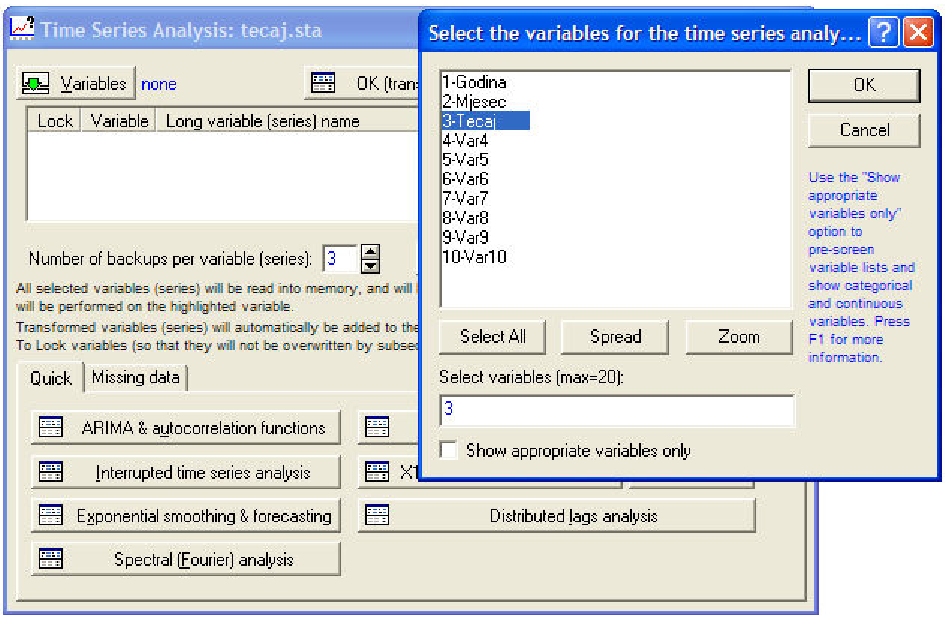
Mjesečni prosjeci srednjeg deviznog tečaja Hrvatske narodne banke za EUR 1997 - 2003

	1 Godina	2 Mjesec	3 Tecaj
1	1997	1	6,906
2	1997	2	6,9402
3	1997	3	6,948
4	1997	4	6,9895
5	1997	5	6,941
6	1997	6	6,9972
7	1997	7	7,0284
8	1997	8	6,9763
9	1997	9	6,9452
10	1997	10	6,9344
11	1997	11	6,9861
12	1997	12	6,9472
13	1998	1	6,9362
14	1998	2	6,974
15	1998	3	7,0352
16	1998	4	7,0692
17	1998	5	7,146
18	1998	6	7,2038
19	1998	7	7,1401
20	1998	8	7,1108
21	1998	9	7,2468
22	1998	10	7,339
23	1998	11	7,339
24	1998	12	7,3291
25	1999	1	7,387139
26	1999	2	7,567448
27	1999	3	7,596698
28	1999	4	7,591112
29	1999	5	7,591861
30	1999	6	7,596586

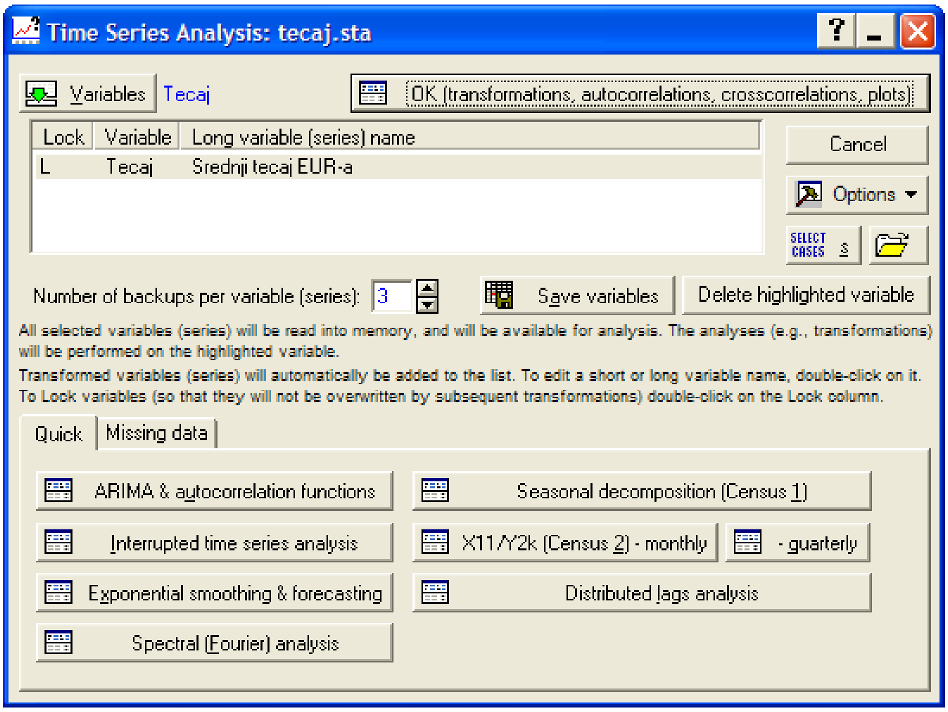
- **STATISTICS / ADVANCED LINEAR/NONLINEAR MODELS / TIME SERIES/FORECASTING**



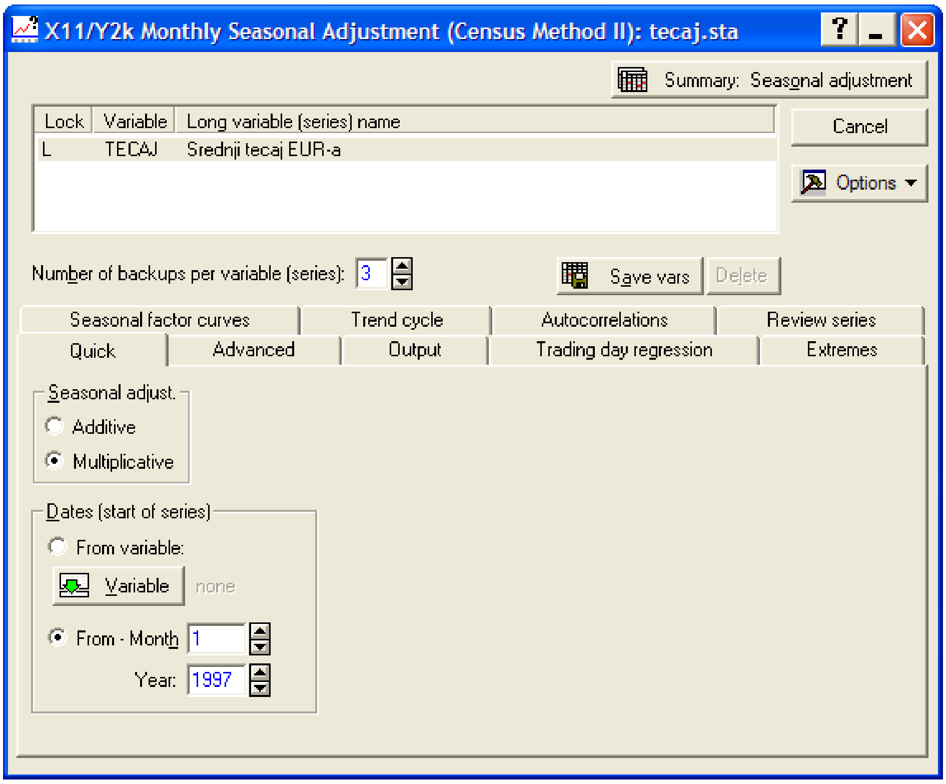
- **Variables: *3-Tecaj***



➢ **X11/Y2k (Census 2) – monthly**



➢ **QUICK**



- **Multiplicative (multiplikativni model)**
- **From – Month *1* (početni mjesec)**
- **Year: *1997* (početna godina)**

U prozoru radnih zapisa izaberu se traženi rezultati:

Trend-ciklička komponenta određuje se pomoću pomičnih prosjeka Hendersonove krivulje koji se mogu očitati u tablici *D 12. Final trend cycle – Henderson Curve*, te su prikazani u tablici 1.

D 12. Final trend cycle - Henderson curve (tecaj.sta)
13-Term moving average selected. I/C ratio is 1,01873
TECAJ : Srednji tecaj EUR-a

Year	January	February	March	April	May	June	July	August	Septembr	October	November	December	Totl
1997	6,897048	6,903940	6,916058	6,931576	6,950955	6,970323	6,986653	6,996059	6,995845	6,986710	6,971384	6,956950	83,46350
1998	6,953470	6,966505	6,995082	7,035307	7,082862	7,132382	7,178830	7,220203	7,258778	7,298327	7,340955	7,385775	85,84847
1999	7,430353	7,473803	7,514766	7,551478	7,581908	7,606281	7,626534	7,645823	7,663472	7,675776	7,681615	7,681898	91,13371
2000	7,679319	7,676235	7,674533	7,672087	7,665148	7,651183	7,631044	7,609037	7,592704	7,588890	7,596752	7,607854	91,64479
2001	7,609174	7,591123	7,553868	7,507168	7,466164	7,442296	7,437638	7,444746	7,452471	7,449054	7,429987	7,401268	89,78495
2002	7,372953	7,354646	7,351119	7,360637	7,376838	7,392964	7,404748	7,411603	7,415787	7,424489	7,442586	7,468465	88,77684
2003	7,498724	7,528298	7,550339	7,563157	7,569153	7,572412	7,578459	7,590710	7,608663	7,627115	7,644669	7,659762	90,99146
Average	7,348720	7,356364	7,365109	7,374487	7,384719	7,395405	7,406273	7,416883	7,426817	7,435765	7,443993	7,451710	

Tablica 1. Hendersonovi pomični prosjeci srednjeg tečaja EUR-a

God./mj.	***1***	***2***	***3***	***4***	***5***	***6***
1997.	6,897048	6,903940	6,916058	6,931576	6,950955	6,970323
1998.	6,953470	6,966505	6,995082	7,035307	7,082862	7,132382
1999.	7,430353	7,473803	7,514766	7,551478	7,581908	7,606281
2000.	7,679319	7,676235	7,674533	7,672087	7,665148	7,651183
2001.	7,609174	7,591123	7,553868	7,507168	7,466164	7,442296
2002.	7,372953	7,354646	7,351119	7,360637	7,376838	7,392964
2003.	7,498724	7,528298	7,550339	7,563157	7,569153	7,572412
Prosjek	*7,348720*	*7,356364*	*7,365109*	*7,374487*	*7,384719*	*7,395405*

God./mj.	***7***	***8***	***9***	***10***	***11***	***12***	***Ukupno***
1997.	6,986653	6,996059	6,995845	6,986710	6,971384	6,956950	83,46350
1998.	7,178830	7,220203	7,258778	7,298327	7,340955	7,385775	85,84847
1999.	7,626534	7,645823	7,663472	7,675776	7,681615	7,681898	91,13371
2000.	7,631044	7,609037	7,592704	7,588890	7,596752	7,607854	91,64479
2001.	7,437638	7,444746	7,452471	7,449054	7,429987	7,401268	89,78495
2002.	7,404748	7,411603	7,415787	7,424489	7,442586	7,468465	88,77684
2003.	7,578459	7,590710	7,608663	7,627115	7,644669	7,659762	90,99146
Prosjek	*7,406273*	*7,416883*	*7,426817*	*7,435765*	*7,443993*	*7,451710*	

Sezonska komponenta izračunata je za svaki mjesec tijekom analiziranog razdoblja, a uz to izračunata je projekcija sezonske komponente za 2004. godinu po mjesecima.

Sezonska komponenta za razdoblje 1997. do 2003. godine vidi se u tablici *D 10. Final seasonal factors*, a prikazana je u tablici 2.

D 10. Final seasonal factors (tecaj.sta)
Table total: 8399,33 Mean: 99,9920 Std.Dev.: ,5934192
TECAJ : Srednji tecaj EUR-a

Year	January	February	March	April	May	June	July	August	Septembr	October	November	December	Avge
1997	99,8816	100,6415	100,6917	100,6038	100,3107	100,1971	99,65994	99,32159	99,50977	99,44300	99,84194	99,62573	99,9774
1998	100,0364	100,7476	100,7917	100,6084	100,2887	100,0792	99,63833	99,31188	99,46633	99,44834	99,83188	99,63490	99,9903
1999	100,2062	100,8722	100,8804	100,5542	100,2398	99,8421	99,61977	99,30889	99,44227	99,53920	99,89001	99,74496	100,0117
2000	100,3233	100,9699	100,9799	100,4290	100,0948	99,5393	99,58641	99,26405	99,35699	99,52702	99,82101	99,68111	99,9644
2001	100,6216	101,1705	101,1606	100,4005	100,0127	99,2934	99,58288	99,28130	99,34277	99,62489	99,91938	99,79317	100,0170
2002	100,6815	101,1812	101,2242	100,2990	99,8648	99,0979	99,58380	99,29314	99,33360	99,65667	99,96665	99,79259	99,9979
2003	100,7521	101,2051	101,2668	100,2686	99,8201	98,9971	99,58473	99,28344	99,31569	99,65274	99,96227	99,71490	99,9853

Tablica 2. Sezonski faktori srednjeg tečaja EUR-a 1997. – 2003. god.

God./mj.	*1*	*2*	*3*	*4*	*5*	*6*
1997.	99,8816	100,6415	100,6917	100,6038	100,3107	100,1971
1998.	100,0364	100,7476	100,7917	100,6084	100,2887	100,0792
1999.	100,2062	100,8722	100,8804	100,5542	100,2398	99,8421
2000.	100,3233	100,9699	100,9799	100,4290	100,0948	99,5393
2001.	100,6216	101,1705	101,1606	100,4005	100,0127	99,2934
2002.	100,6815	101,1812	101,2242	100,2990	99,8648	99,0979
2003.	100,7521	101,2051	101,2668	100,2686	99,8201	98,9971

God./mj.	*7*	*8*	*9*	*10*	*11*	*12*	*Prosjek*
1997.	99,65994	99,32159	99,50977	99,44300	99,84194	99,62573	99,9774
1998.	99,63833	99,31188	99,46633	99,44834	99,83188	99,63490	99,9903
1999.	99,61977	99,30889	99,44227	99,53920	99,89001	99,74496	100,0117
2000.	99,58641	99,26405	99,35699	99,52702	99,82101	99,68111	99,9644
2001.	99,58288	99,28130	99,34277	99,62489	99,91938	99,79317	100,0170
2002.	99,58380	99,29314	99,33360	99,65667	99,96665	99,79259	99,9979
2003.	99,58473	99,28344	99,31569	99,65274	99,96227	99,71490	99,9853

Opća je ocjena da smanjeni priljev eura u početnim mjesecima svake godine uzrokuje povećanje deviznog tečaja, dok sezonski priljev temeljem pritjecanja deviznih sredstava od naplaćenih prihoda za turističke usluge dovodi do relativnog obilja eura na tržištu i posljedičnog smanjenja cijene eura.

S obzirom da se glavnina prihoda ostvaruje na tržištima adoptiranog eura, a sezona svake godine sve više produžuje u netradicionalne turističke mjesece, to je i sezonski činitelj sve ranije opadajući, što je prikazano u tablici 3. vrijednostima manjim od 100 već od 5. mjeseca 2002. i 2003. godine. Slična se kretanja predmnijevaju i za 2004. godinu (tablica 3), što sugerira isplativost konverzije kune za euro u 6. mjesecu i njegovu prodaju u 4. mjesecu uz zaradu na promjeni deviznih tečajeva od otprilike 2,2% u navedenom razdoblju. Drugim riječima, vidljivo je da je sezonski faktor jači u prvoj polovici godine, dok u drugoj polovici godine opada. Međutim zadnjih godina počeo je opadati već u lipnju i polako u svibnju. Taj trend pretpostavljen je i za 2004. godinu, pa se pretpostavlja da će sezonski faktor biti ispod 100 već u svibnju.

Projekcija sezonske komponente za 2004. god. nalazi se u tablici *D 10a. Seasonal factors, one year ahead*. Pretočena je u tablicu 3.

	D 10a. Seasonal factors, one year ahead (tecaj.sta) TECAJ : Srednji tecaj EUR-a												
Year	January	February	March	April	May	June	July	August	Septembr	October	November	December	Avge
2004	100,7873	101,2171	101,2881	100,2533	99,79771	98,94678	99,58520	99,27859	99,30673	99,65077	99,96008	99,67606	99,97898

Tablica 3. Projekcija sezonskih faktora srednjeg tečaja EURu 2004. god.

God./mj.	***1***	***2***	***3***	***4***	***5***	***6***
2003.	100,7873	101,2171	101,2881	100,2533	99,79771	98,94678

God./mj.	***7***	***8***	***9***	***10***	***11***	***12***	***Prosjek***
2003.	99,58520	99,27859	99,30673	99,65077	99,96008	99,67606	99,97898

Slučajna komponenta zapisana je u tablici *D 13. Final irregular series*, koja je prenijeta u tablicu 4. Ukoliko su neke vrijednosti prepoznate kao ekstremi, tada su zamijenjene s vrijednošću 100 čime su dobivene vrijednosti slučajne komponente modificirane za ekstreme. Modificirane slučajne vrijednosti prikazane u tablici 5, dobivene su u iz tablice *E 3. Modified irregular series*.

	D 13. Final irregular series (tecaj.sta) Table total: 8403,39 Mean: 100,040 Std.Dev.: ,7292857 TECAJ : Srednji tecaj EUR-a												
Year	January	February	March	April	May	June	July	August	Septembr	October	November	December	S.D.
1997	100,2485	99,8845	99,7718	100,2305	99,5475	100,1882	100,9408	100,3987	99,7651	99,8072	100,3697	100,2350	0,381740
1998	99,7153	99,3648	99,7835	99,8741	100,6010	100,9213	99,8215	99,1672	100,3706	101,1151	100,1417	99,5963	0,580715
1999	99,2138	100,3775	100,2081	99,9708	99,8917	100,0305	99,9183	99,9537	100,1552	100,0052	99,8051	100,2180	0,279267
2000	100,2064	99,7390	99,7103	100,0675	100,1483	100,3043	100,0170	100,0866	99,5682	99,7673	99,7966	100,1944	0,231174
2001	100,2514	100,3012	100,5125	99,8632	97,4742	99,0659	97,0205	102,9684	101,8713	100,3085	100,1084	99,7843	1,555406
2002	101,9571	99,9506	99,4816	100,1784	100,1488	99,9243	100,2837	100,2358	99,6603	101,3438	99,8800	99,8567	0,721885
2003	100,0086	100,0191	100,6060	99,7869	99,8796	100,1650	99,9275	98,9594	100,1955	99,9126	100,1749	100,1176	0,368911
Std.Dev.	0,8192	0,3210	0,4008	0,1560	1,0011	0,5187	1,1885	1,2428	0,7614	0,6807	0,2045	0,2343	

Tablica 4. Slučajna komponenta srednje vrijednosti EUR 1997. – 2003. god.

God./mj.	*1*	*2*	*3*	*4*	*5*	*6*
1997.	100,2485	99,8845	99,7718	100,2305	99,5475	100,1882
1998.	99,7153	99,3648	99,7835	99,8741	100,6010	100,9213
1999.	99,2138	100,3775	100,2081	99,9708	99,8917	100,0305
2000.	100,2064	99,7390	99,7103	100,0675	100,1483	100,3043
2001.	100,2514	100,3012	100,5125	99,8632	97,4742	99,0659
2002.	101,9571	99,9506	99,4816	100,1784	100,1488	99,9243
2003.	100,0086	100,0191	100,6060	99,7869	99,8796	100,1650
St. dev.	*0,8192*	*0,3210*	*0,4008*	*0,1560*	*1,0011*	*0,5187*

God./mj.	*7*	*8*	*9*	*10*	*11*	*12*	*St. dev.*
1997.	100,9408	100,3987	99,7651	99,8072	100,3697	100,2350	0,381740
1998.	99,8215	99,1672	100,3706	101,1151	100,1417	99,5963	0,580715
1999.	99,9183	99,9537	100,1552	100,0052	99,8051	100,2180	0,279267
2000.	100,0170	100,0866	99,5682	99,7673	99,7966	100,1944	0,231174
2001.	97,0205	102,9684	101,8713	100,3085	100,1084	99,7843	1,555406
2002.	100,2837	100,2358	99,6603	101,3438	99,8800	99,8567	0,721885
2003.	99,9275	98,9594	100,1955	99,9126	100,1749	100,1176	0,368911
St. dev.	*1,1885*	*1,2428*	*0,7614*	*0,6807*	*0,2045*	*0,2343*	

Year	E 3. Modified irregular series (tecaj.sta) Table total: 8400,76 Mean: 100,009 Std.Dev.: ,3696678 TECAJ : Srednji tecaj EUR-a												
	January	February	March	April	May	June	July	August	Septembr	October	November	December	S.D.
1997	100,2485	99,8845	99,7718	100,2305	99,5475	100,1882	100,9408	100,3987	99,7651	99,8072	100,3697	100,2350	0,381740
1998	99,7153	99,3648	99,7835	99,8741	100,6010	100,9213	99,8215	99,1672	100,3706	101,1151	100,1417	99,5963	0,580715
1999	99,2138	100,3775	100,2081	99,9708	99,8917	100,0305	99,9183	99,9537	100,1552	100,0052	99,8051	100,2180	0,279267
2000	100,2064	99,7390	99,7103	100,0675	100,1483	100,3043	100,0170	100,0866	99,5682	99,7673	99,7966	100,1944	0,231174
2001	100,2514	100,3012	100,5125	99,8632	100,0000	99,0659	100,0000	100,0000	100,0000	100,3085	100,1084	99,7843	0,348959
2002	100,0000	99,9506	99,4816	100,1784	100,1488	99,9243	100,2837	100,2358	99,6603	100,0000	99,8800	99,8567	0,226817
2003	100,0086	100,0191	100,6060	99,7869	99,8796	100,1650	99,9275	98,9594	100,1955	99,9126	100,1749	100,1176	0,368911
Std.Dev.	0,3519	0,3210	0,4008	0,1560	0,3015	0,5187	0,3798	0,5346	0,2820	0,4532	0,2045	0,2343	

Tablica 5. Modificirana slučajna komponenta srednje vrijednosti EUR 1997. – 2003. god.

God./mj.	*1*	*2*	*3*	*4*	*5*	*6*
1997.	100,2485	99,8845	99,7718	100,2305	99,5475	100,1882
1998.	99,7153	99,3648	99,7835	99,8741	100,6010	100,9213
1999.	99,2138	100,3775	100,2081	99,9708	99,8917	100,0305
2000.	100,2064	99,7390	99,7103	100,0675	100,1483	100,3043
2001.	100,2514	100,3012	100,5125	99,8632	*100,0000*	99,0659
2002.	*100,0000*	99,9506	99,4816	100,1784	100,1488	99,9243
2003.	100,0086	100,0191	100,6060	99,7869	99,8796	100,1650
St. dev.	*0,3519*	*0,3210*	*0,4008*	*0,1560*	*0,3015*	*0,5187*

God./mj.	*7*	*8*	*9*	*10*	*11*	*12*	*St. dev.*
1997.	100,9408	100,3987	99,7651	99,8072	100,3697	100,2350	0,381740
1998.	99,8215	99,1672	100,3706	101,1151	100,1417	99,5963	0,580715
1999.	99,9183	99,9537	100,1552	100,0052	99,8051	100,2180	0,279267
2000.	100,0170	100,0866	99,5682	99,7673	99,7966	100,1944	0,231174
2001.	*100,0000*	*100,0000*	*100,0000*	100,3085	100,1084	99,7843	0,348959
2002.	100,2837	100,2358	99,6603	*100,0000*	99,8800	99,8567	0,226817
2003.	99,9275	98,9594	100,1955	99,9126	100,1749	100,1176	0,368911
St. dev.	*0,3798*	*0,5346*	*0,2820*	*0,4532*	*0,2045*	*0,2343*	

*Modificirane vrijednosti prikazane su kurzivom (italic)

b) Desezoniranje vremenskog niza

Kada su prikazane sve komponente vremenskog niza moguće je izračunati desezonirani vremenski niz. U slučaju ekstremnih vrijednosti zamjenjuju se s trend-cikličkim vrijednostima. Desezonirani niz prikazan je u tablici 6., koja je dobivena iz tablice *D 11. Final seasonally adjusted series*. Modificirani desezonirani niz zapisan je u tablici 7 koja je dobivena iz tablice *E 2. Modified seasonally adjusted series*.

	D 11. Final seasonally adjusted series (tecaj.sta) Table total: 621,890 Mean: 7,40346 Std.Dev.: ,2519019 TECAJ : Srednji tecaj EUR-a												
Year	January	February	March	April	May	June	July	August	Septembr	October	November	December	Totl
1997	6,914186	6,895964	6,900274	6,947551	6,919502	6,983438	7,052382	7,023951	6,979415	6,973241	6,997160	6,973299	83,56036
1998	6,933673	6,922251	6,979937	7,026450	7,125429	7,198096	7,166018	7,160069	7,285681	7,379711	7,351359	7,355957	85,88463
1999	7,371936	7,502019	7,530401	7,549273	7,573696	7,608603	7,620306	7,642280	7,675363	7,676177	7,666644	7,698643	91,11534
2000	7,695167	7,656197	7,652298	7,677264	7,676512	7,674463	7,632339	7,615630	7,559922	7,571232	7,581297	7,622642	91,61497
2001	7,628303	7,613990	7,592581	7,496902	7,277583	7,372777	7,216033	7,665733	7,591928	7,472034	7,438039	7,385305	89,75121
2002	7,517251	7,351016	7,313012	7,373770	7,387817	7,387366	7,425759	7,429081	7,390594	7,524256	7,433657	7,457760	88,99133
2003	7,499368	7,529737	7,596091	7,547040	7,560037	7,584910	7,572961	7,511722	7,623539	7,620449	7,658037	7,668773	90,97266
Average	7,365697	7,353025	7,366371	7,374036	7,360082	7,401379	7,383685	7,435496	7,443778	7,459586	7,446599	7,451768	

Tablica 6. Desezonirane vrijednosti srednjeg tečaja EUR-a 1997. – 2003. god.

God./mj.	*1*	*2*	*3*	*4*	*5*	*6*
1997.	6,914186	6,895964	6,900274	6,947551	6,919502	6,983438
1998.	6,933673	6,922251	6,979937	7,026450	7,125429	7,198096
1999.	7,371936	7,502019	7,530401	7,549273	7,573696	7,608603
2000.	7,695167	7,656197	7,652298	7,677264	7,676512	7,674463
2001.	7,628303	7,613990	7,592581	7,496902	7,277583	7,372777
2002.	7,517251	7,351016	7,313012	7,373770	7,387817	7,387366
2003.	7,499368	7,529737	7,596091	7,547040	7,560037	7,584910
Prosjek	7,365697	7,353025	7,366371	7,374036	7,360082	7,401379

God./mj.	*7*	*8*	*9*	*10*	*11*	*12*	*Ukupno*
1997.	7,052382	7,023951	6,979415	6,973241	6,997160	6,973299	83,56036
1998.	7,166018	7,160069	7,285681	7,379711	7,351359	7,355957	85,88463
1999.	7,620306	7,642280	7,675363	7,676177	7,666644	7,698643	91,11534
2000.	7,632339	7,615630	7,559922	7,571232	7,581297	7,622642	91,61497
2001.	7,216033	7,665733	7,591928	7,472034	7,438039	7,385305	89,75121
2002.	7,425759	7,429081	7,390594	7,524256	7,433657	7,457760	88,99133
2003.	7,572961	7,511722	7,623539	7,620449	7,658037	7,668773	90,97266
Prosjek	7,383685	7,435496	7,443778	7,459586	7,446599	7,451768	

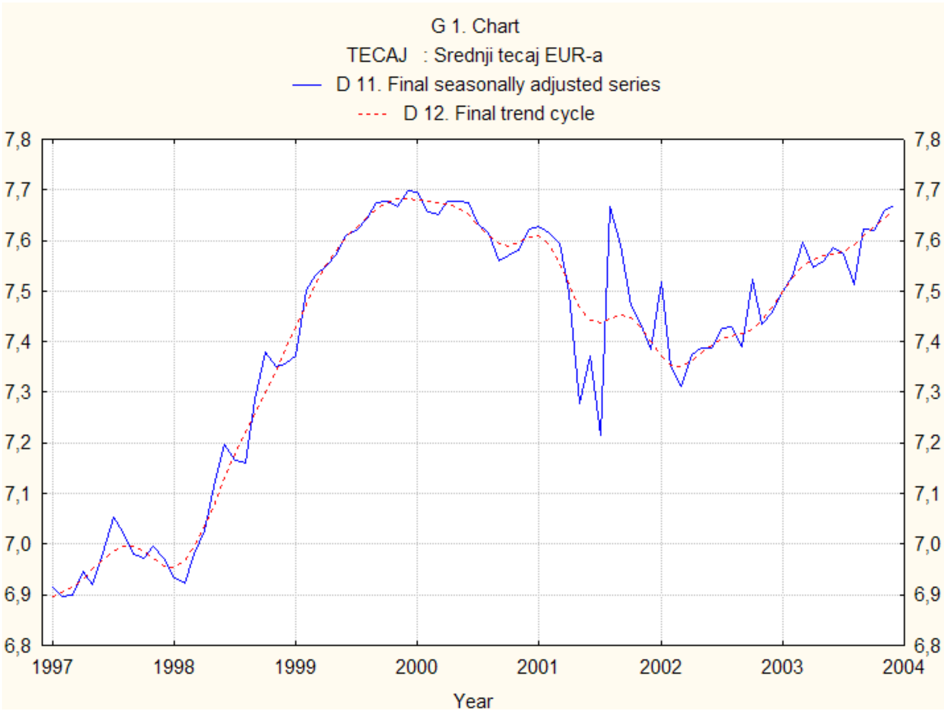
Year	E 2. Modified seasonally adjusted series (tecaj.sta) Table total: 621,696 Mean: 7,40114 Std.Dev.: ,2477892 TECAJ : Srednji tecaj EUR-a												
	January	February	March	April	May	June	July	August	Septembr	October	November	December	Totl
1997	6,914186	6,895964	6,900274	6,947551	6,919502	6,983438	7,052382	7,023951	6,979415	6,973241	6,997160	6,973299	83,56036
1998	6,933673	6,922251	6,979937	7,026450	7,125429	7,198096	7,166018	7,160069	7,285681	7,379711	7,351359	7,355957	85,88463
1999	7,371936	7,502019	7,530401	7,549273	7,573696	7,608603	7,620306	7,642280	7,675363	7,676177	7,666644	7,698643	91,11534
2000	7,695167	7,656197	7,652298	7,677264	7,676512	7,674463	7,632339	7,615630	7,559922	7,571232	7,581297	7,622642	91,61497
2001	7,628303	7,613990	7,592581	7,496902	7,466164	7,372777	7,437638	7,444746	7,452471	7,472034	7,438039	7,385305	89,80095
2002	7,372953	7,351016	7,313012	7,373770	7,387817	7,387366	7,425759	7,429081	7,390594	7,424489	7,433657	7,457760	88,74727
2003	7,499368	7,529737	7,596091	7,547040	7,560037	7,584910	7,572961	7,511722	7,623539	7,620449	7,658037	7,668773	90,97266
Average	7,345084	7,353025	7,366371	7,374036	7,387022	7,401379	7,415343	7,403926	7,423855	7,445333	7,446599	7,451768	

Tablica 7. Modificirane desezonirane vrijednosti srednjeg tečaja EUR 1997. – 2003. god.

God./mj.	***1***	***2***	***3***	***4***	***5***	***6***
1997.	6,914186	6,895964	6,900274	6,947551	6,919502	6,983438
1998.	6,933673	6,922251	6,979937	7,026450	7,125429	7,198096
1999.	7,371936	7,502019	7,530401	7,549273	7,573696	7,608603
2000.	7,695167	7,656197	7,652298	7,677264	7,676512	7,674463
2001.	7,628303	7,613990	7,592581	7,496902	*7,466164*	7,372777
2002.	*7,372953*	7,351016	7,313012	7,373770	7,387817	7,387366
2003.	7,499368	7,529737	7,596091	7,547040	7,560037	7,584910
Prosjek	7,345084	7,353025	7,366371	7,374036	7,387022	7,401379

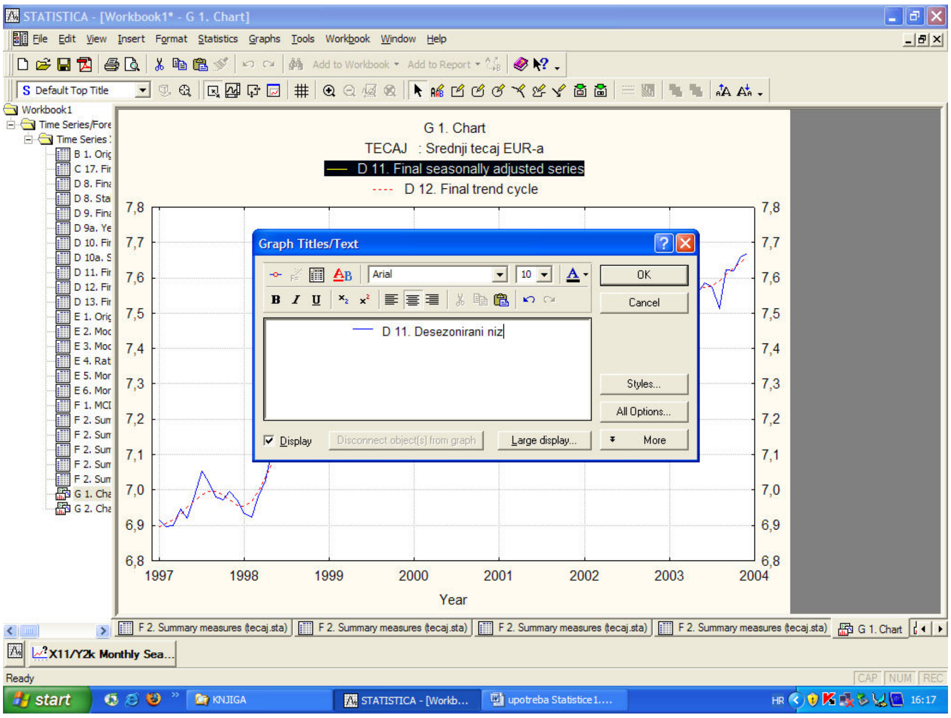
God./mj.	***7***	***8***	***9***	***10***	***11***	***12***	***Ukupno***
1997.	7,052382	7,023951	6,979415	6,973241	6,997160	6,973299	83,56036
1998.	7,166018	7,160069	7,285681	7,379711	7,351359	7,355957	85,88463
1999.	7,620306	7,642280	7,675363	7,676177	7,666644	7,698643	91,11534
2000.	7,632339	7,615630	7,559922	7,571232	7,581297	7,622642	91,61497
2001.	*7,437638*	*7,444746*	*7,452471*	7,472034	7,438039	7,385305	89,80095
2002.	7,425759	7,429081	7,390594	*7,424489*	7,433657	7,457760	88,74727
2003.	7,572961	7,511722	7,623539	7,620449	7,658037	7,668773	90,97266
Prosjek	7,415343	7,403926	7,423855	7,445333	7,446599	7,451768	

Da bi dobiveni rezultati bili jasniji prikazan je grafikon vrijednosti desezoniranog niza i trend-cikličke komponente koji je prikazan na grafikonu 1, a dobiven je iz grafikona *G 1*.

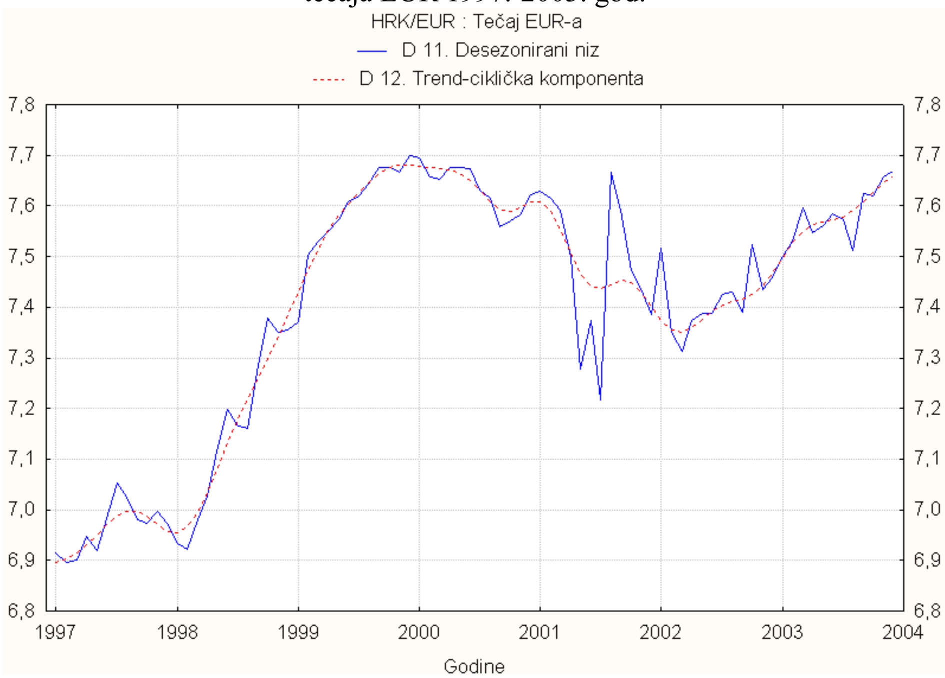


Grafikon se može promijeniti ako se dva puta klikne na dio koji se želi promijeniti i tada se naprave izmjene.

Promijenjen je naslov, a legenda je prevedena na hrvatski jezik:



Grafikon 1. Vrijednosti desezoniranog niza i trend-cikličkih vrijednosti srednjeg tečaja EUR 1997.-2003. god.



Prof. dr. sc. Maja Biljan-August
Prof. dr. sc. Snježana Pivac
Doc. dr. sc. Ana Štambuk

UPORABA STATISTIKE U EKONOMIJI
2. IZDANJE

Izdavač:
Ekonomski fakultet Sveučilišta u Rijeci

Recenzentice:
Prof. dr. sc. Jasna Horvat

Doc. dr. sc. Suzana Marković

Doc. dr. sc. Alemka Šegota

Lektorica:
Kerol Musul-Perić, prof.

Autor naslovnice:
Luka Mičetić, dipl. oec.

Pri izradi naslovnice korišteni su materijali objavljeni na:
www.free-stockphotos.com, www.sxc.hu, www.hnb.hr.

Objavljivanje ovog sveučilišnog udžbenika odobrilo je Povjerenstvo za izdavačku djelatnost Sveučilišta u Rijeci Odlukom – klasa: 602-09/09-01/29, ur. broj: 2170-57-05-09-3 od 25. rujna 2009.

Objavljeno na URL: http://www.efri.hr/prikaz.asp?txt_id=6326
i http://oliver.efri.hr/~statist/biljan-pivac-stambuk-uporaba2.pdf.

ISBN: 978-953-6148-86-8

Rijeka, rujan, 2009.