REPUBLIKA SLOVENIJA
MINISTRSTVO ZA ŠOLSTVO IN ŠPORT

Naložba v vašo prihodnost
OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA
Evropski socialni sklad

POSLOVNA INFORMATIKA S STATISTIKO

ANTON GAMS
MARKO GAMS

Višješolski strokovni program: Gostinstvo in turizem
Učbenik: Poslovna informatika s statistiko
Gradivo za 1. letnik

Avtorja:

mag. Anton Gams, univ. dipl. inž.
ŠOLSKI CENTER VELENJE
Višja strokovna šola

Marko Gams, univ. dipl. ekon.
ŠOLSKI CENTER VELENJE
Višja strokovna šola

Strokovni recenzentki: Nevenka Rozman
Marika Šadl, univ. dipl. ekon.
Lektorica: Lidija Šuster, prof. slov. jez.

Ljubljana, 2009

Gradivo je sofinancirano iz sredstev projekta Impletum Uvajanje novih izobraževalnih programov na področju višjega strokovnega izobraževanja v obdobju od 2008 do 2011.

Projekt oziroma operacijo delno financira Evropska unija iz Evropskega socialnega sklada in Ministrstvo RS za šolstvo in šport. Operacija se izvaja v Operativnem programu razvoja človeških virov za obdobje od 2007 do 2013, razvojne prioritete Razvoj človeških virov in vseživljenjskega učenja ter prednostne usmeritve Izboljšanje kakovosti in učinkovitosti sistemov izobraževanja in usposabljanja.

KAZALO VSEBINE

PREDGOVOR

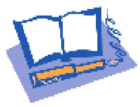
Spoštovani,
veseli naju, da se srečujemo ob tem učbeniku, ki je namenjen podpori na poti pridobivanja znanja na področju Poslovne informatike s statistiko v programu višjega strokovnega izobraževanja Gostinstvo in turizem. Knjiga daje pregled nad znanjem in kompetencami navedenega predmetnega področja, njene vsebine so izbrane skladno z izobraževalnimi cilji, določenimi v katalogu znanja.

Naša dosedanja sodelovanja pri tem predmetu kažejo, da nekateri študenti že imate določeno popotnico srednješolskega znanja s področja statistike, drugi pa prihajate iz šol, kjer vam programi niso nudili tovrstnega znanja, zato začenjate na tem področju na novo. Oboji dovolj dobro poznate in se zavedate pomena zbiranja in objavljanja podatkov ter njihove interpretacije. Praksa potrjuje, da uporaba kvalitetnih podatkov, pravilnih statističnih metod ter znanja statističnega proučevanja dajejo pravilnim informacijam pomen pri odločanju. Pri tem je dandanes nujna podpora informacijsko komunikacijske tehnologije (IKT), ki se kaže v hitrem in stroškovno optimalnem zagotavljanju podatkov, njihovem prenosu, obdelavi, shranjevanju, prezentaciji … Dobro znanje poslovne informatike in statistike vam bo potrebno že pri izpolnjevanju vaših študijskih obveznosti, osvojene kompetence s tega področja pa boste lahko v celoti izkazali v praksi pri pripravi poročil, analiz ter sprejemanju dobrih poslovnih odločitev.

Gradivo je razdeljeno na osem vsebinskih sklopov. V vsakem od njih uvodni predstavitvi sledi uvodna naloga, ki jo rešujemo ob usvajanju novih vsebin v nadaljevanju, v osrednjem delu posameznega poglavja. Opis vsebine dopolnjujejo praktični primeri in naloge, za katere želiva, da vas usmerjajo h kreativnemu razmišljanju in aktivnemu ukvarjanju z vsebino. Poglavja zaključujejo povzetki, ki nudijo hiter pregled obravnavane snovi, tem pa sledijo vprašanja in naloge za samostojno preverjanje znanja. Večina nalog se nanaša na obravnavano področje gostinstva in turizma in so izbrane tako, da pri reševanju krepijo vaše sposobnosti za komunikacijo z IKT za uporabo in obdelavo podatkov, za dostop do podatkovnih virov ...

V učbeniku uporabljava oznake, ki posameznim delom besedila dajejo naslednji pomen:

 Uvod poglavja

 Osrednji del poglavja

 Uvodna naloga

 Povzetek poglavja

 Rešitev uvodne naloge

 Predlagana Excelova funkcija

Strokovni pregled učbenika sta opravili gospa Nevenka Rozman in gospa Marika Šadl, zato se obema za vse pripombe in predloge za izboljšavo najlepše zahvaljujeva. Prav tako se zahvaljujeva gospe Lidiji Šuster, prof. slov. jez., ki je popravila najine jezikovne napake. Hvala tudi sodelavkam in sodelavcem za vzpodbude pri pisanju tega gradiva.

Spoštovane študentke, spoštovani študenti. Želiva vam uspešen študij ter koristno uporabo pridobljenega znanja.
Avtorja

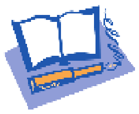
1 UVOD V STATISTIKO

V uvodnem poglavju učbenika bomo spoznali današnji pomen in vlogo statistike. Usvojili in povezali bomo osnovne pojme, na katerih temeljijo statistična proučevanja: enota, populacija, spremenljivka, parameter ter zanje poiskali mesto v priloženih primerih in vajah. Ob spoznavanju organiziranosti in izvajanja nalog državne, lokalne in ostale statistike se bomo naučili, kako lahko iz različnih statističnih virov samostojno pridobimo kakovostne ter časovno in krajevno primerljive podatke, potrebne za naše delo. Spoznali bomo, kako si lahko predvsem z uporabo IKT ob minimalnih stroških zagotovimo pravilne in pravočasne podatke. S praktično uporabo pridobljenega znanja bomo tudi sprotno rešili naslednjo uvodno nalogo.

Uvodna naloga

Na najhitrejši možni način želimo ugotoviti, kolikšno je uradno število tujih turistov iz Nemčije leta 2007 in koliko prenočitev so navedenega leta ti turisti ustvarili v vseh nastanitvenih objektih na območju občine našega stalnega bivališča.

Nalogo bomo rešili v nadaljevanju poglavja, ko bomo med ostalim spoznali tudi načine in poti ažurnega pridobivanja statističnih podatkov z uporabo IKT.



Za današnjo družbo je značilno izobilje podatkov, ki so pomembni pri odločanju na večini področij našega dela in življenja. Podatki bodo lahko koristni, če jih bomo pravilno zbrali, obdelali in interpretirali. Da bomo za to kompetentni, potrebujemo določeno mero znanja, ki ga daje poznavanje statistike. Zato začnimo s prvimi koraki na poti k temu cilju.

1.1 POJEM STATISTIKE IN NJEN POMEN

S statistiko preučujemo množične pojave in povezave med njimi na osnovi številčnih in drugih podatkov. Šadl (2004) opredeljuje, da beseda statistika vključuje:

- sistematično zbrane številske in druge podatke o najrazličnejših pojavih;
- dejavnost, ki se ukvarja z opazovanjem pojavov, zbiranjem podatkov o njih, obdelavo, analizo ter objavljanjem podatkov, in organizacije, ki se s to dejavnostjo ukvarjajo;
- posebno vedo, ki proučuje množične pojave in povezave med njimi. Razvoj statistike je povezan z razvijanjem teorije in metod o zbiranju, obdelavi in analiziranju podatkov;
- vrsto evidence – sistematični zapis, obdelava s statističnimi metodami, analiza, prikaz.

1.2 STATISTIKA V SLOVENIJI

1.2.1 Državna statistika

Statistično dejavnost, ki jo na območju kake države izvajajo državni organi, imenujemo državna statistika. Ta ima nalogo stalnega ali občasnega zbiranja, obdelave in prikazovanja ustreznih statističnih podatkov, ki so predvideni z nacionalnim programom statističnih raziskovanj ter zagotavljanja kakovostnih, pravočasnih, časovno in krajevno ter mednarodno primerljivih podatkov o stanjih in gibanjih na ekonomskem, demografskem, socialnem in okoljskem področju, kot je navedeno na spletni strani http://www.stat.si/ našega najpomembnejšega izvajalca statističnih raziskovanj ter povezovalca dela na področju državne statistike, **Statističnega urada Republike Slovenije** (SURS). Poleg povezovanja in usklajevanja statističnega sistema sodijo med njegove najpomembnejše naloge še mednarodno sodelovanje, predvidevanje potreb uporabnikov, zbiranje, obdelava in izkazovanje podatkov, skrb za njihovo zaupnost …, kot lahko povzamemo s spletne strani SURS (2008).

Poleg Statističnega urada RS izvajajo v tem srednjeročnem obdobju dejavnost državne statistike za določena področja statistik še:

- Banka Slovenije (statistika plačilne bilance, finančne in denarne statistike), http://www.bsi.si/
- Ministrstvo za finance (statistika državnega primanjkljaja in dolga, statistika javnih financ), http://www.mf.gov.si/slov/index.htm
- Agencija Republike Slovenije za javnopravne evidence in storitve (delno poslovne statistike), http://www.ajpes.si/
- Inštitut za varovanje zdravja (statistika zdravstva), http://www.ivz.si/
- Zavod za pokojninsko in invalidsko zavarovanje (statistika pokojnin), http://www.zpiz.si/
- Zavod RS za zaposlovanje (delno statistika zaposlenosti in brezposelnosti), http://www.ess.gov.si/

1.2.2 Viri statističnih podatkov

Pooblaščene institucije objavljajo statistične podatke v spletnih objavah, publikacijah, podatkovnih bazah ter različnih datotekah. Statistični urad RS vse podatke v celoti objavlja na spletnih straneh, kjer so uporabnikom brezplačno na voljo. Podatki na spletni strani SURS so urejeni v 4 velike tematske sklope, ki združujejo 31 statističnih vsebin. Podatki o **turizmu** so uvrščeni v sklop **Ekonomsko področje** (http://www.stat.si/tema_ekonomsko_turizem.asp).

Slika 1: Spletna stran Statističnega urada R Slovenija
Vir: SURS, http://www.stat.si/ (12. 2. 2009)

Kadar želimo z enega mesta dostopati do statističnih podatkov iz različnih virov (tudi na evropskih ravni), je dobrodošla uporaba SI-Stat podatkovnega portala, ki je dosegljiv preko spletne podstrani Statističnega urada RS, http://www.stat.si/pxweb/dialog/statfile2.asp. Tu so uporabnikom na razpolago tudi podatki evropskega statističnega urada (EUROSTAT).

Več podatkov s področja **turizma** je uporabnikom na voljo tudi na spletnih straneh **Slovenske turistične organizacije (STO)**: http://www.slovenia.info/.

Slika 2: Spletna stran Slovenske turistične organizacije
Vir: http://www.slovenia.info/si/Statistični-podatki.htm (12. 2. 2009)

Statistični urad RS izdaja tudi publikacije, ki se nanašajo na območje države, statističnih regij in občin; nekatere publikacije vsebujejo tudi mednarodne statistične preglede. Vse publikacije, ki izidejo v tiskani obliki, so objavljene tudi na spletnih straneh SURS. Med osnovnimi publikacijami (http://www.stat.si/publikacije/pub.asp) izstopa **Statistični letopis**, ki je osrednja serijska publikacija (http://www.stat.si/letopis/). Izhaja konec leta in ponuja najširši nabor statističnih podatkov za preteklo leto ter številne daljše časovne nize podatkov. Tudi v tem učbeniku bo ta publikacija naš temeljni podatkovni vir. Publikacija **Pomembnejši statistični podatki o Sloveniji** (http://www.stat.si/publikacije/pub_psp.asp) izhaja mesečno in vsebuje izbor statističnih podatkov, objavljenih v tekočem mesecu. Uporabnikom so na razpolago tudi rezultati raziskovanj na različnih področjih ter druge zbirke publikacij. Pomembno mesto imajo publikacije s podatki, zbranimi, analiziranimi ter publiciranimi ob popisih prebivalstva, gospodinjstev in stanovanj. Izvedeni so na področju celotne države približno vsakih 10 let, nazadnje je bil leta 2002 (http://www.stat.si/popis2002/si/).

Vstopite na spletno stran Statističnega urada Republike Slovenije in si oglejte vsebine po posameznih področjih. Ugotovite, kateri značilni podatki so uporabnikom na voljo na posameznih področjih in vsebinskih sklopih.

Med množičnimi viri statističnih podatkov so tudi publikacije v elektronski in tiskani obliki, ki jih kreirajo in distribuirajo v lokalnih okoljih. Na področju turizma so njihovi nosilci predvsem lokalne turistične organizacije, občine, turistična podjetja oz. hoteli, moteli, gostišča, kampi …

Razmislite o vsebinah lokalnih publikacij, ki so na področju turizma aktualne na vašem področju.

1.3 TEMELJNI POJMI V STATISTIČNI ANALIZI

Spoznajmo najprej osnovne pojme, na katerih temelji statistična analiza.

1.3.1 Enota in populacija

Osnovni element, ki ga proučujemo, je **enota**. Pri statističnem proučevanju se zavestno odločimo, kaj je enota našega proučevanja (npr.: prebivalec, rojstvo, stalno bivališče; oz. na našem ožjem področju proučevanja: gost, hotel, turistična organizacija, država ipd.).

Enote razvrščamo z ozirom na časovno opredelitev v tri skupine:

- realne enote, ki dejansko obstajajo ter jih opazujemo v danem trenutku (dijak, učitelj, šola, prebivalec, turist, hotel …);
- dogodki, ki se v danem trenutku zgodijo in jih opazujemo v časovnem obdobju (rojstva v letu 2008, nočitve v izbranem mesecu …);
- dogajanja, ki jih opazujemo v danem trenutku ali v časovnem razmiku (gradnja cest, prodaja pohištva …).

Populacija je skupnost enot, ki jih opredelimo in proučujemo. Pojav, ki ga želimo proučevati, moramo razmejiti od drugih pojavov; to storimo tako, da populacijo opredelimo po treh vidikih: vsebinskem, krajevnem in časovnem, kot navaja Košmelj (1996).

Proučujemo obiske gostov z Nizozemske v kampih v Republiki Sloveniji v letu 2008. Pojem Republika Slovenija opredeljuje populacijo krajevno, leto 2008 časovno, vsebina pa so gosti z Nizozemske v kampih.

1.3.2 Vzorec

Množične pojave analiziramo tako, da proučujemo lastnosti enot ustrezno opredeljene populacije. Kadar je populacija obsežna zaradi različnih razlogov (običajno ekonomskih), pogosto ni mogoče oz. ni smotrno pregledati vseh enot. Zato v praksi v takšnih primerih proučujemo primerno izbrano končno podmnožico celotne populacije, ki jo imenujemo **vzorec**. Vzorec lahko izbiramo na več načinov. Najpogosteje izbiramo enote slučajno in govorimo o slučajnem vzorcu. Možni pa so tudi vsaj deloma sistematični načini vzorčenja, kar prikazuje naslednji primer.

Če za vzorec naključno izberemo npr. 15 moških in 15 žensk, je vzorec deloma slučajen, deloma pa določen načrtno, saj smo se odločili za 15 moških in 15 žensk.

Pomembna lastnost vzorca je njegova velikost, to je število vseh enot v vzorcu.

Iz populacije vseh hotelov v R Sloveniji 31. 12. 2008 deloma slučajno izberemo 60 hotelov, in sicer 10 iz Ljubljane, 15 iz zdraviliških krajev, 15 iz obmorskih krajev in 20 iz gorskih krajev. Znotraj vsake skupine smo izbrali hotele slučajno. Pravimo, da je vzorec stratificiran (razdeljen v stratume). V prvem stratumu (Ljubljana) je 10 enot, v drugem 15 itd. V našem primeru vzorec šteje 60 enot (Ostan, 2008).

1.3.3 Spremenljivke in njihove vrste

Enote imajo več lastnosti. Ponavadi ne opazujemo vseh lastnosti, ampak le nekatere, na primer mesečne stroške družine. Med družinami se le-ti spreminjajo, zato pravimo, da so stroški **statistična spremenljivka.** Opazovana lastnost (strošek) enote je torej spremenljivka, ki pri vsaki enoti zavzame svojo vrednost.

Lastnost enote ali opazovana spremenljivka je lahko tako motiv za odločitev za letovanje v izbranem okolju, kot navaja Ostan (2008). Odločimo se lahko npr. najprej, da so njene vrednosti: cene, naravne lepote, obiski ... V drugi raziskavi pa bi se lahko odločili za druge vrednosti, npr. zdravljenje, cene ... Izbor spremenljivk in njihovih vrednosti je torej odvisen od tega, kar želimo proučiti. Nekatere spremenljivke in njihove vrednosti so se v statistiki že uveljavile in jih uporablja veliko število raziskovalcev. Na primer: vrsta gostov (domači, tuji), turistični objekti (hoteli, kampi, zasebne sobe), spol (moški, ženski), število obiskov posameznega kraja, število nočitev itd.

Spremenljivka ima pri vsaki opazovani enoti neko vrednost. Glede na način izražanja njene vrednosti so spremenljivke:

- **številske,** ki so izražene s števili in so lahko:
  - **zvezne**, poljubna vrednost v določenem območju (npr.: vrednost prodaje turističnih aranžmanov v razmiku od 1.520,00 do 1.860,00 evrov …),
  - **diskretne** (število članov v gospodinjstvu, npr.: 1, 2 ali 5 članov …);
- **opisne,** ki imajo najmanj dve vrednosti (spol) ali pa več (kraj prebivališča, izobrazba, registrska številka avtomobila …).

Razmislite o tem, ali je potrebno vrednosti telefonske številke, bančnega računa, telefonske številke ... uvrstiti med številske ali med opisne spremenljivke in utemeljite svojo odločitev.

Spremenljivka "način potovanja", ki ima npr. vrednosti "z letalom", "z avtomobilom", "s partnerjem (ko)", "z družino", "s prijatelji" ... je opisna. "Število otrok v družini" pa je številska diskretna spremenljivka (Ostan, 2008).

V statistični literaturi pogosto srečamo naslednje pojme: univariantna statistična analiza, ko proučujemo eno samo spremenljivko, bivariantna, ko proučujemo istočasno dve spremenljivki, ter multivariantna, ko je le-teh več.

Če proučujemo spremenljivko "število obiskov posameznega turista v kraju A", govorimo o univariantni analizi. Tudi če proučujemo "način rezervacije obiska", uporabljamo univariantno analizo. Ko pa proučujemo tudi povezavo (odvisnost, vpliv) med obema spremenljivkama, govorimo o bivariantni analizi (Ostan, 2008).

1.3.4 Statistični parametri

Posameznim enotam opazujemo določene lastnosti zato, da bi dobili podatke, na podlagi katerih lahko ugotovimo lastnosti celotne populacije. Mero, s katero izražamo lastnost populacije, imenujemo **parameter.** V statistični analizi torej izračunavamo vrednosti parametrov, s katerimi so merjene in izražene lastnosti proučevanega pojava.

Parametre dobimo:

- s preštevanjem enot (npr.: število vseh turistov iz Nemčije leta 2008 v R Sloveniji);
- s seštevanjem vrednosti spremenljivk (npr.: vrednost prihodka od turizma v Sloveniji v preteklem koledarskem letu);
- z razvrščanjem enot v skupine glede na vrednost spremenljivke (npr.: število turistov po krajih obiska);
- z izračunavanjem ustreznih kazalcev (npr.: povprečna starost turistov).

Po prvih treh poteh dobimo enostavne parametre, z izračunavanjem ustreznih kazalcev (npr.: srednjih vrednosti, relativnih števil, mer variabilnosti …) pa izvedene parametre.

Naslednji dve tabeli vsebujeta del podatkov popisa prebivalstva, gospodinjstev in stanovanj v R Sloveniji leta 2002. Opazujmo podatke ter opredelimo obravnavane osnovne pojme statistične analize.

Tabela 1: Podatki popisa prebivalcev leta 2002

Starostne skupine (leta)	Moški	Ženske	Skupaj
0–9	94.928	89.429	184.357
10–19	125.779	120.060	245.839
20–29	150.646	142.018	292.664
30–39	148.672	145.458	294.130
40–49	159.148	151.605	310.753
50–59	122.578	121.125	243.703
60–69	91.476	107.683	199.159
70–79	51.862	90.999	142.861
80 in več	13.487	37.083	50.570
Skupaj	**958.576**	**1.005.460**	**1.964.036**

Vir: SURS, Popis 2002,
http://www.stat.si/popis2002/si/rezultati_slovenija_prebivalstvo_dz.htm (12. 2. 2009)

- Statistična populacija se nanaša na vse prebivalce v R Sloveniji ter je opredeljena krajevno (območje R Slovenije), časovno (stanje na dan 30. 6. 2002 ob 24. uri) ter vsebinsko (podatki se nanašajo na stalne prebivalce).
- Enota v obravnavani populaciji je posamezni prebivalec.
- Pri vsaki enoti so opazovali spol (opisna spremenljivka) ter starost (zvezna številska spremenljivka).
- Parameter je število prebivalcev v opazovanem primeru, njegovo vrednost so ugotovili s preštevanjem enot. Z razvrščanjem prebivalcev po spolu dobimo opisno statistično vrsto, razvrstitev v desetletne starostne skupine pa daje številsko statistično vrsto; obe štejemo za parameter, saj se nanašata na celotno populacijo. Podrobnejšemu spoznavanju statističnih vrst pa bomo našo pozornost namenili v naslednjem poglavju.

Tabela 2: Podatki popisa prebivalstva in gospodinjstev leta 2002

Leto popisa	**Število prebivalcev** (na km^2)	**Povprečno število članov v gospodinjstvu**	**Delež prebivalstva v kmetijstvu** (v %)
1921	64,4	-	-
1931	69,0	4,9	58,8
1948	71,1	3,8	48,9
1953	74,3	3,7	41,1
1961	78,6	3,5	31,6
1971	85,3	3,4	20,4
1981	93,4	3,2	9,2
1991	94,5	3,0	7,6
2002	96,9	2,8	4,0

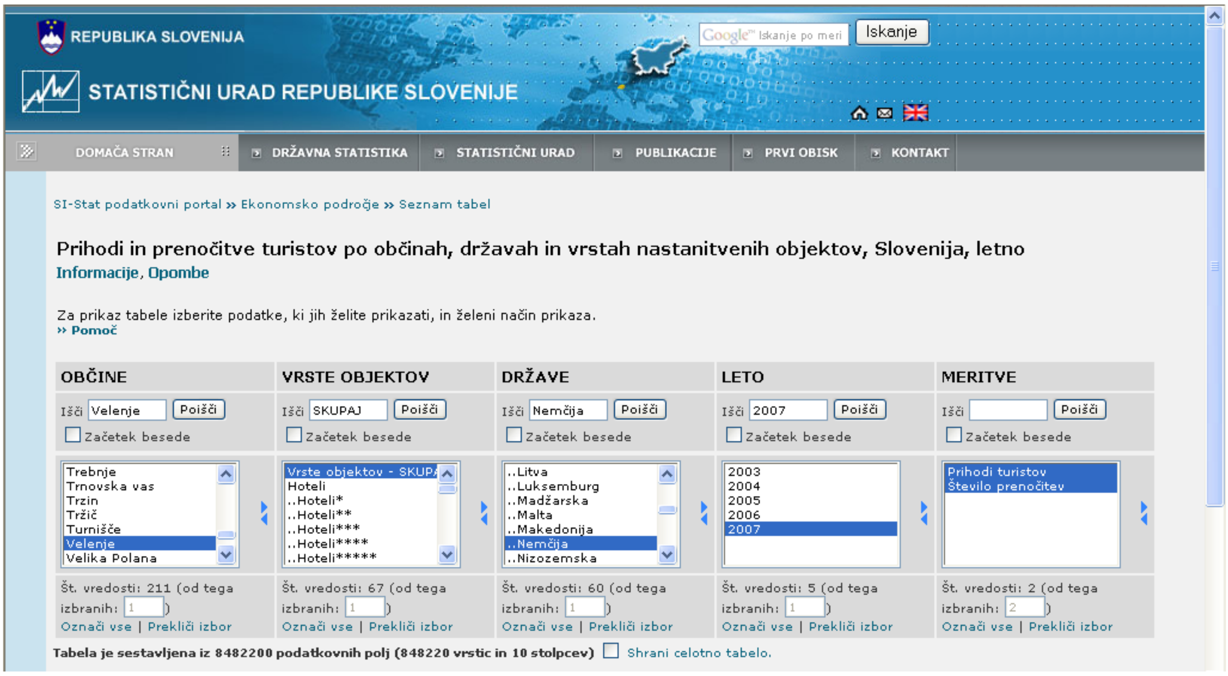
Vir: Tabela 1 ter SURS, Popis 2002, http://www.stat.si/popis2002/si/default (12. 2. 2009)

- V tabeli je predstavljenih devet populacij, ki so jih opazovali ob popisih prebivalcev v navedenih letih; opazovana gospodinjstva predstavljajo osem populacij.
- Enota je posamezni prebivalec oz. gospodinjstvo na dan popisa.
- Pri prebivalcih so ugotavljali tudi podatke o poklicu oz. o dejavnosti (npr.: kmetijstvo, podjetništvo …), kar predstavlja opisno spremenljivko.
- Prikazani kazalci so izračunani na podlagi medsebojne primerjave s popisom dobljenih podatkov. Pri prvem kazalcu gre za razmerje med številom prebivalcev in znano površino naše države (20.273 km^2), za preostala dva je rezultat dobljen z ustrezno medsebojno primerjavo podatkov popisa.

Rešitev uvodne naloge

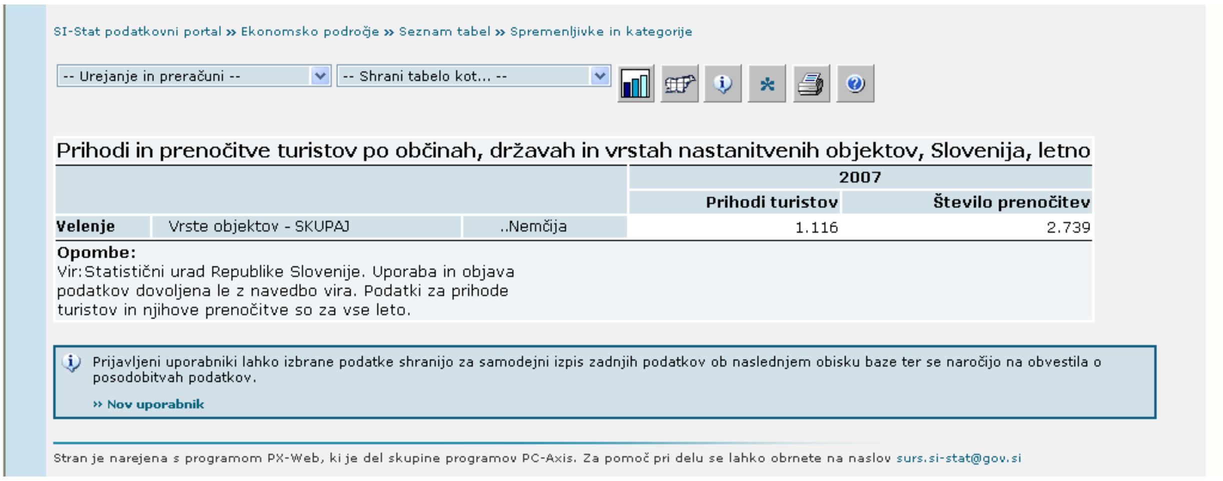
V uvodu postavljena naloga zahteva, da (z uporabo IKT) ugotovimo, kolikšno je uradno število tujih turistov iz Nemčije in število njihovih prenočitev leta 2007 v vseh nastanitvenih objektih na območju (v tem primeru izbrane) Mestne občine Velenje.

Spoznali smo, da državna statistika skrbi tudi za zbiranje, analizo in predstavitev statističnih podatkov po regijah in občinah. Pregled možnosti pridobitve zahtevanih uradnih podatkov, kjer želimo dobiti sumarno število prenočitev nemških turistov v vseh uradnih prenočitvenih objektih (hoteli, moteli, penzioni, gostišča, prenočišča, apartmaji, kampi, turistične kmetije, planinske koče in domovi …) pokaže, da lahko to najučinkoviteje postorimo z uporabo SI-Stat podatkovne baze Statističnega urada R Slovenije. Vstopimo na spletno stran portala http://www.stat.si/pxweb/dialog/statfile2.asp ter izberemo tematsko področje ekonomije in v njem vsebine turizma. Po izbiri možnosti **"Nastanitvena statistika po občinah"** nam portal ponudi tudi opcijo **"Prihodi in prenočitve turistov po občinah, državah in vrstah nastanitvenih objektov, Slovenija, letno"**. V našem primeru populacijo vsebinsko, krajevno in časovno preprosto opredelimo z ustrezno določitvijo pogojev iskanja in pregledovanja podatkov v predstavljenih izbirnih poljih. V izbrani opciji podatkovna baza omogoča poizvedovanje po številu prenočitev in prihodu turistov iz 60-ih držav, in sicer za posamezna leta v obdobju od 2003 do 2007; kriterij iskanja podatkov pa lahko izbiramo (in kombiniramo) po 67-ih vrstah nastanitvenih objektov v 211-ih občinah. Nastavimo lahko tudi način izpisa podatkov, kjer je možno izbirati med vrstičnim ali stolpčnim prikazom zaslonske tabele ter zapisom v Excelovo datoteko.



Slika 3: Določanje pogojev iskanja podatkov na spletni strani portala SI-Stat
Vir: SURS, http://www.stat.si/pxweb/Dialog/varval.asp?ma=2118102S&ti (12. 2. 2009)

Rezultat iskanja nam podatkovni portal SI-Stat vrne na svoji spletni strani, kjer lahko v zaslonski tabeli preberemo rezultat našega poizvedovanja. Ta je v našem primeru obisk 1.116 turistov in njihovih skupaj 2.739 nočitev v letu 2007 v Mestni občini Velenje.



Slika 4: Prikaz rezultatov iskanja podatkov na podatkovnem portalu SI-Stat
Vir: SURS, SI-Stat podatkovni portal, http://www.stat.si/pxweb/Dialog/ (12. 2. 2009)

V poglavju smo spoznali osnovne pojme v statistiki, kot so: enota, populacija, spremenljivka ter parameter kot temeljno mero izražanja lastnosti proučevanega pojava. Na priloženih zgledih smo pokazali njihovo uporabo in pomen pri reševanju praktičnih primerov. Spoznali smo izvajalce statistične dejavnosti na državnem nivoju ter pomembnejše publikacije in ostale značilne vire statističnih podatkov, ki jih objavljajo pooblaščene institucije. Ob reševanju uvodne naloge smo ugotovili uporabnost dobrega poznavanja teh ustanov in njihovih razpoložljivih podatkovnih virov, tudi dostopa do podatkovnih baz. Za našo usposobitev, kako iz množice izbrati koristne podatke ter jih sistematično in ažurno spremeniti v koristno informacijo, smo izvedli postopek sodobnega *pridobivanja želenih podatkov z internetno pomočjo preko prosto dostopnega podatkovnega portala SI-Stat Statističnega urada Republike Slovenije. Tako smo tudi izpostavili in dali ustrezen pomen vlogi informacijske in komunikacijske tehnologije že na prvem koraku statističnega preučevanja določenega pojava v praksi, to je zbiranju kvalitetnih podatkov. S tem želimo dati poudarek, da je tako pridobljena izkušnja vedno bolj potrebna tudi zaradi aktualnih zahtev stroškovno obvladljivega in ažurnega načina pridobivanja želenih podatkov in našega poslovnega komuniciranja v celoti. Dobra usposobljenost komunikacije z IKT za zbiranje, uporabo, obdelavo, prenos ... podatkov, za dostop do informacijskih virov ter obdelavo informacij se vam bo obrestovala pri uspešnem študijskem delu in kasnejšem reševanju poslovnih problemov.*

Naloge

1. Vstopite na spletno stran Statističnega urada R Slovenije, oglejte si vsebine objavljenih osnovnih in promocijskih publikacij. Primerjajte jih ter opišite njihove značilnosti (pomoč: http://www.stat.si/publikacije/pub.asp).

2. Na spletni strani Statističnega urada R Slovenije odprite Statistični letopis 2008, poglavje 25 – Turizem. Razmislite, kako bi za podatke v posameznih tabelah opredelili osnovne statistične pojme: enoto, populacijo, spremenljivke, parametre.

3. Poiščite spletno stran Slovenske turistične organizacije, oglejte si njene vsebine ter primerjajte publikacije s statističnimi podatki, ki so pri tej organizaciji na voljo uporabnikom v elektronski obliki, ter zanje določite osnovne statistične pojme (pomoč: http://www.slovenia.info/). Ugotovite podobnosti in razlike s podatki, objavljenimi na ustreznih spletni straneh Statističnega urada R Slovenije.

4. V knjižnici poiščite publikacije z objavljenimi statističnimi podatki na področju gostinstva in turizma ter najdite primere, kjer so vrednosti spremenljivke določene zvezno oziroma diskretno.

5. Vstopite na spletno stran Državnega izpitnega centra ter odprite najnovejše letno poročilo o poklicni maturi. Razmislite, kako bi za podatke o doseženem uspehu kandidatov v vašem srednješolskem izobraževalnem programu opredelili osnovne statistične pojme (pomoč: http://www.ric.si/).

6. Razmislite o publikacijah, ki ste jih spoznali na praksi ali vašem študentskem delu. Opišite spremenljivke enot, katerih vrednosti so bile opisne oziroma številske.

7. Spomnite se svojega izpolnjevanja prijavnega obrazca ob vašem vpisu na višjo strokovno šolo. Opredelite vrsto vprašanj ter pripadajočim spremenljivkam izražene vrednosti.

2 ZBIRANJE, UREJANJE IN PRIKAZOVANJE PODATKOV

S statističnimi metodami na določen način analiziramo lastnosti družbenih in ekonomskih pojavov. O njih moramo v ta namen imeti na razpolago zbrane ter ustrezno obdelane podatke. Celoten postopek, s katerim zberemo podatke o pojavu in analiziramo njegove lastnosti, pogosto imenujemo statistično raziskovanje. To obsežno delo, kot navaja B. Košmelj (1996), obsega več opravil, od zbiranja podatkov, prek njihove obdelave oziroma urejanja do ustrezne in pregledne predstavitve. Pri tem si moramo prizadevati za natančnost in odgovornost ter krepiti čut za estetiko in obliko prikazanih podatkov in izračunanih parametrov. Pokaže se, da je pri tem zopet nujna podpora IKT z orodji za učinkovito izdelavo ter predstavitev podatkov v tabelah in grafih, s katerimi v praksi povečamo preglednost statističnih analiz in poročil.

Uvodna naloga

Zberite podatke o izbranem pojavu, jih smiselno uredite ter zanje kreirajte tabelo in grafikon. Prizadevajte si, da bo vaš prikaz podatkov pravilen in pregleden, zato pri oblikovanju tabele in grafa uporabite postopke, ki so v aplikaciji Excel potrebni za učinkovito urejanje podatkov in oblikovanje tabel ter kreiranje in dopolnjevanje različnih vrst grafov.

Delo zbiranja, urejanja in prikazovanja podatkov obsega aktivnosti, ki si sledijo v določenem zaporedju, medsebojno povezano in usklajeno. Ta opravila so:

- načrtovanje statističnega raziskovanja,
- statistično opazovanje oz. zbiranje podatkov,
- urejanje in obdelava podatkov,
- prikazovanje podatkov in
- statistična analiza opazovanega pojava.

2.1 NAČRTOVANJE STATISTIČNEGA RAZISKOVANJA

Tudi za statistično raziskovanje velja, da je dobro pripravljen načrt pogoj za uspešno delo in kakovosten rezultat, kot navaja B. Košmelj (1996). Z **načrtovanjem raziskovanja** predvidimo celoten postopek raziskovanja ter rešimo vprašanja, povezana z raziskovanjem; ta se nanašajo predvsem na:

- vsebinska vprašanja, kjer izstopata dve – opredelitev populacije in določitev opazovanih spremenljivk;
- analitična vprašanja – izbrani parametri, ki bodo izračunani, morajo izražati lastnosti proučevanega pojava, kar je osnovni cilj raziskovanja;
- organizacijska in tehnična vprašanja – izbira najprimernejšega načina zbiranja, obdelave in prikaza podatkov;
- finančna vprašanja, da se s predvidenimi sredstvi uspešno izvede raziskovanje ter zagotovi ustrezna kakovost podatkov.

2.2 STATISTIČNO OPAZOVANJE

Rezultat **statističnega opazovanja** je množica podatkov, ki so osnova za statistično analizo pojava. Za uspešnost statističnega raziskovanja je bistvena kakovost zbranih podatkov. Ti morajo biti popolni (za vse opazovane enote po predvidenih spremenljivkah) ter točni (ustrezati morajo dejanskemu stanju).

Z ozirom na število opazovanih enot delimo statistična opazovanja na:

- **popolna**, saj z njimi zberemo podatke za vse enote preučevane populacije; tu ločimo:
  - popis realnih enot (registracija gostov v recepciji hotela, registracija motornih vozil, popis prebivalcev, popis stanovanj …) in
  - sprotno spremljanje dogodkov (evidenca rojstev, evidenca prometnih nesreč, podatki o proizvodnji …);
- **delna,** kjer zberemo podatke le za del enot proučevanega pojava.

Delno opazovanje imenujemo tudi **vzorčenje,** ko se iz seznama vseh enot (okvira vzorčenja) zajame del populacije. Parametre izračunamo za vzorec in jih posplošimo na celotno populacijo. Torej ne ugotavljamo dejanskih statističnih parametrov za celotno populacijo, ampak te le ocenimo (Šadl, 2008). Z vzorcem lahko dobimo le oceno parametra. Pomembno za celoten postopek vzorčenja pa je, da lahko kakovost vzorčnih ocen ugotavljamo samo v primeru, če je bila izbira enot v vzorec slučajna. Kot mero za kakovost je namreč mogoče na podlagi vzorčnih podatkov pri slučajnem vzorcu izračunati standardno napako ocene parametra.

Vsekakor je potrebno in smiselno še pred izvajanjem statističnega opazovanja preveriti statistične vire o preučevanem pojavu, če so morda zanj že zbrani, obdelani in objavljeni podatki. Kadar so potrebni podatki na razpolago v evidencah institucij, pooblaščenih za statistična proučevanja (spoznali smo jih že v prejšnjem poglavju), ti predstavljajo **sekundarni vir** statističnih podatkov. Če teh ni in izvajalec opazovanja sam zbere podatke, govorimo o **primarnem viru** – in sicer:

- z neposrednim opazovanjem (zdravnik – podatke o zdravstvenem stanju pacientov, učitelj športne vzgoje – športne dosežke dijakov …);
- s posrednim opazovanjem po osebah, ki enoto poznajo in dajo podatke (šola – podatke o dijakih in njihovem uspehu, starši o otrocih …).

Oseba, ki izvaja opazovanje, vpisuje podatke v za to pripravljene vprašalnike. Na tem dokumentu so v obliki vprašanj spremenljivke, po katerih bomo opazovali enote. »Pri sestavljanju vprašanj moramo upoštevati sposobnosti in pripravljenost tistih, ki ga bodo izpolnjevali. Torej morajo biti vprašanja kratka in razumljiva najširšemu krogu ljudi« (Šadl, 2008, 10). Anketiranje kot zbiranje podatkov se izvaja tudi dopisno, telefonsko ali prek spleta.

Spomnite se vprašalnika – prijavnice, ki ste jo izpolnili pri vpisu v višjo strokovno šolo. Ali so bila po vašem mnenju vprašanja ustrezna? Ste uporabili navodila za izpolnjevanje in ali ste bili seznanjeni z namenom zbiranja podatkov?

2.3 OBDELAVA IN UREJANJE PODATKOV

S statističnim opazovanjem je običajno zbrana obsežna množica podatkov, ti se nanašajo na veliko število opazovanih enot ter več spremenljivk z veliko vrednostmi le-teh. Zato moramo podatke najprej **urediti** in **prikazati** v pregledni obliki, ki omogoča prepoznavanje značilnosti populacije in izračunavanje parametrov.

Kadar ima spremenljivka manjše število vrednosti, vsaki vrednosti priredimo skupino, npr.: dijake razvrstimo po "spolu" v dve skupini, po "učnem uspehu" v pet skupin. Za spremenljivke z veliko vrednostmi, kot so npr.: "stalno prebivališče dijakov", "znesek štipendije" … pa v skupine združujemo sorodne vrednosti. Za spremenljivko "stalno prebivališče dijakov" bi tako opredelili skupino z večjim geografskim območjem (občina, regija), za spremenljivko "znesek štipendije" pa bi lahko izbrali skupine – te pri številskih spremenljivkah navadno imenujemo **razredi** – npr. tako, kot kaže naslednja tabela.

Tabela 3: Opredelitev razredov za znesek štipendije

Znesek štipendije v €
nad 25 do 50
nad 50 do 75
nad 75 do 100
nad 100 do 125
nad 125 do 150

Vir: Lasten

V razred zajamemo niz vrednosti številske spremenljivke, ki so omejene z dvema mejama, v vsakem razredu pa lahko določimo:

- spodnjo mejo razreda: $y_{j,\min}$
- zgornjo mejo razreda: $y_{j,\max}$
- širino razreda: $d_j = y_{j,\max} - y_{j,\min}$
- sredino razreda: $y_j = \dfrac{y_{j,\min} + y_{j,\max}}{2}$

Pri oblikovanju skupin moramo upoštevati zahtevi, da so skupine:

- opredeljene enolično (vsako enoto lahko glede na vrednost razvrstimo le v eno skupino);
- homogene, enovite, morajo združevati sorodne vrednosti in biti opredeljene skladno s cilji raziskovanja.

To velja za vse vrste spremenljivk v skupinah, pri številskih moramo upoštevati še posebnosti zvezne in diskretne spremenljivke.
Z razvrščanjem enot v skupine oz. razrede dobimo **statistične vrste.**

Statistične vrste

V času raziskave nastalo statistično vrsto, ki je najpogosteje neurejena, najprej uredimo.

Prvih 10 kandidatov je pri opravljanju poklicne mature zbralo naslednje število točk:

14	*17*	*13*	*21*	*18*	*12*	*11*	*22*	*20*	*19*

Vrednosti, urejene po velikosti (naraščajoče ali padajoče), dajo spodnjo urejeno statistično vrsto, imenovano ***ranžirna vrsta****. Ranžirna vrsta je primerna za manjše podatkovne nize, za obsežnejše množice pa uporabljamo frekvenčne porazdelitve.*

11	*12*	*13*	*14*	*17*	*18*	*19*	*20*	*21*	*22*

Z razvrščanjem enot po časovni, krajevni ali vsebinski (stvarni) opredelitvi dobimo **časovne, krajevne in stvarne vrste.**

Časovne vrste so lahko:

- trenutne (momentne), če se podatki nanašajo na časovne trenutke ter
- razmične (intervalne), kadar se podatki nanašajo na časovne razmike.

Trenutna časovna vrsta

Kot primer trenutne časovne vrste si poglejmo podatke o številu prebivalcev za območje sedanje R Slovenije po popisih v obdobju od leta 1921 do 2002. Tabela 4 predstavlja podatke, ki se za posamezno leto popisa praviloma nanašajo na stanje 31. 3. ob 24. uri.

Tabela 4: Število prebivalcev na ozemlju Slovenije ob popisih 1921–2002

Leto popisa	Moški	Ženske	Skupaj
1921	622.168	682.632	1.304.800
1931	673.248	724.402	1.397.650
1948	675.353	764.447	1.439.800
1953	712.034	792.393	1.504.427
1961	760.770	830.753	1.591.523
1971	835.998	891.139	1.727.137
1981	918.766	973.098	1.891.864
1991	923.643	989.712	1.913.355
2002	958.576	1.005.460	1.964.036

Vir: SURS, Letopis 2008, http://www.stat.si/letopis/2008/04 (12. 2. 2009)

Razmična časovna vrsta

Tabela 5: Prihodi tujih turistov v R Slovenijo v letu 2007 po mesecih

Mesec	Prihodi tujih turistov
Januar	86.464
Februar	75.439
Marec	97.148
April	138.430
Maj	150.105
Junij	181.520
Julij	240.511
Avgust	277.922
September	195.228
Oktober	128.982
November	92.284
December	87.299

Vir: SURS, Letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)

Krajevna vrsta

Tabela 6: Število turističnih potovanj naših prebivalcev v letu 2007 (v tisoč)

Država	Turistična potovanja	Poslovna potovanja
Hrvaška	1.253	65
Italija	173	48
Avstrija	148	56
Nemčija	140	76
Bosna in Hercegovina	132	24
Srbija	87	33

Vir: SURS, Letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)

Stvarna vrsta

Tabela 7: Študentje VSŠ v študijskem letu 2006/07 po posameznih področjih

Program	Redni	Izredni	Skupaj
Elektroenergetika	0	118	118
Elektronika	98	148	246
Gostinstvo	321	76	397
Gradbeništvo	139	495	634
Hortikultura (vrtnarstvo)	133	65	198
Informatika	352	416	768
Kmetijstvo	64	57	121
Komercialist	592	2.267	2.859
Komunala	123	561	684
Lesarstvo	169	91	260
Mehatronika	579	588	1.167
Multimediji	62	245	307
Poslovni sekretar	519	1.500	2.019
Poštni promet	270	139	409
Promet	207	683	890
Računovodja	387	1.335	1.722
Rudarstvo in geotehnologija	44	22	66
Strojništvo	555	922	1.477
Telekomunikacije	157	51	208
Turizem	425	312	737
Upravljanje podeželja in krajine	338	84	422
Ustni higienik	0	27	27
Živilstvo	293	188	481
Živilstvo in prehrana	164	33	197
Skupaj	**5.991**	**10.423**	**16.414**

Vir: SURS, Letopis 2008, http://www.stat.si/letopis/2008/06 (12. 2. 2009)

2.4 PRIKAZOVANJE STATISTIČNIH PODATKOV

Statistične vrste predstavljamo v **tabelah** in z **grafikoni**. Prednost tabel je, da so lahko poljubno velike, v njih lahko prikažemo natančne podatke, razvrščene tudi po več spremenljivkah hkrati. Grafikoni so po drugi strani zelo pregledni in nazorni; v praksi se običajno oba načina dopolnjujeta. Tako tabele kot grafikone danes oblikujemo in dopolnjujemo s programskimi orodji; najpogosteje se uporablja elektronska preglednica, kjer izstopata Excel in Calc, oba odlikuje preprosta in učinkovita uporaba.

2.4.1 Prikazovanje podatkov v tabelah

Podatke vnašamo v polja, dobljena na presečiščih vrstic in stolpcev. Prvi ali zadnji stolpec je zbirni s seštevkom podatkov vrstic, prva ali zadnja vrstica je zbirna s seštevkom podatkov stolpcev. Nad tabelo napišemo naslov, da je razvidno, kateri podatki so v njej prikazani, pod njo pa vir, iz katerega so bili podatki črpani.

Naslov tabele: → Tabela 8: Priporočen izgled elementov tabele

		GLAVA		
Č	*polje (celica)*		s	Z s
E			t	b t
L		*vrstica*	o	i o
O			l	r l
			p	n p
			e	i e
			c	c
	Z B I R N A	V R S T I C A		

Vir: → Vir: Prirejeno po Šadl (2006)

Tabele delimo na:

- enostavne (enorazsežne): prikazujejo eno statistično vrsto, podatki so razvrščeni po eni spremenljivki;
- sestavljene: prikazujejo več istovrstnih statističnih vrst za več časovnih obdobij;
- kombinirane (dvorazsežne): prikazujejo podatke razvrščene po dveh spremenljivkah.

Na spletni strani Statističnega urada R Slovenije odprite publikacijo Popis 2002 ter si oglejte oblike in obseg objavljenih tabel.

Enostavna tabela

Enostavna tabela prikazuje eno samo statistično vrsto ter ima podatke nanizane v samo enem stolpcu ali eni vrstici.

Tabela 9: Prenočitve turistov v Sloveniji po mesecih v letu 2007 (v tisoč)

Mesec	**Število prenočitev**
Januar	472.257
Februar	497.207
Marec	502.120
April	567.645
Maj	618.581
Junij	811.129
Julij	1.226.242
Avgust	1.325.017
September	776.080
Oktober	572.521
November	440.712
December	451.797
Skupaj	**8.261.308**

Vir: SURS, Statistični letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)

Preizkusite postopek, s katerim lahko v Excelu preoblikujete (konvertirate) podatke v preglednici iz stolpcev v vrstice ter v obratni smeri.

Sestavljena tabela

Ta tabela je sestavljena iz več vrst (v našem primeru časovnih), ki pa se nanašajo na isto spremenljivko.

Tabela 10: Prenočitve turistov v RS po vrstah krajev in po mesecih v letu 2007

Mesec	**Ljubljana**	**Zdraviliški kraji**	**Obmorski kraji**	**Gorski kraji**	**Drugi turist. kraji**	**Ostali kraji**	**Skupaj**
Januar	32.247	174.893	46.847	160.232	55.036	3.002	472.257
Februar	30.172	192.035	62.141	155.752	54.130	2.977	497.207
Marec	44.763	193.195	94.439	103.269	62.491	3.963	502.120
April	56.839	205.039	139.707	90.652	70.355	5.053	567.645
Maj	63.012	204.743	152.500	121.040	71.286	6.000	618.581
Junij	70.573	226.267	242.890	177.266	85.627	8.506	811.129
Julij	90.903	282.111	390.450	351.524	100.542	10.712	1.226.242
Avgust	94.852	324.738	390.007	390.569	115.077	9.774	1.325.017
September	70.756	233.817	206.121	166.992	91.620	6.774	776.080
Oktober	62.192	233.336	115.359	83.564	73.422	4.648	572.521
November	46.830	201.339	87.706	42.215	59.233	3.389	440.712
December	44.162	179.683	64.442	103.568	56.363	3.579	451.797
Skupaj	**707.301**	**2.651.196**	**1.992.609**	**1.946.643**	**895.182**	**68.377**	**8.261.308**

Vir: SURS, Statistični letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)

Kombinirana tabela

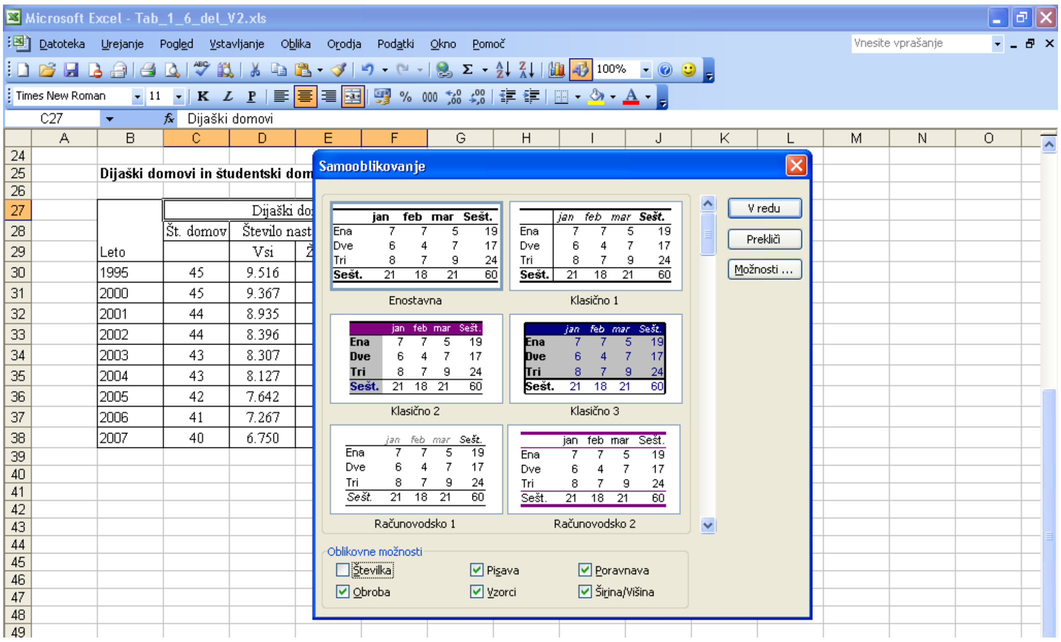
Prikazuje podatke več statističnih vrst, ki se nanašajo na dve ali več različnih spremenljivk.

Tabela 11: Dijaški in študentski domovi v letu 2006

Leto	Dijaški domovi				Študentski domovi		
	Število domov	Število nastanjenih		Vzgojitelji	Število domov	Število nastanjenih	
		Vsi	Ženske			Vsi	Ženske
1995	45	9.516	5.052	371	19	8.876	5.107
2000	45	9.367	4.892	363	19	9.198	5.209
2001	44	8.935	4.834	334	20	9.100	5.311
2002	44	8.396	4.611	321	20	9.017	5.340
2003	43	8.307	4.637	319	21	9.009	5.283
2004	43	8.127	4.537	316	22	9.655	5.258
2005	42	7.642	4.200	303	29	10.010	5.878
2006	41	7.267	3.958	278	32	10.256	6.021
2007	40	6.750	3.762	233	34	10.606	6.314

Vir: SURS, Statistični letopis 2008, http://www.stat.si/letopis/2008/06 (12. 2. 2009)

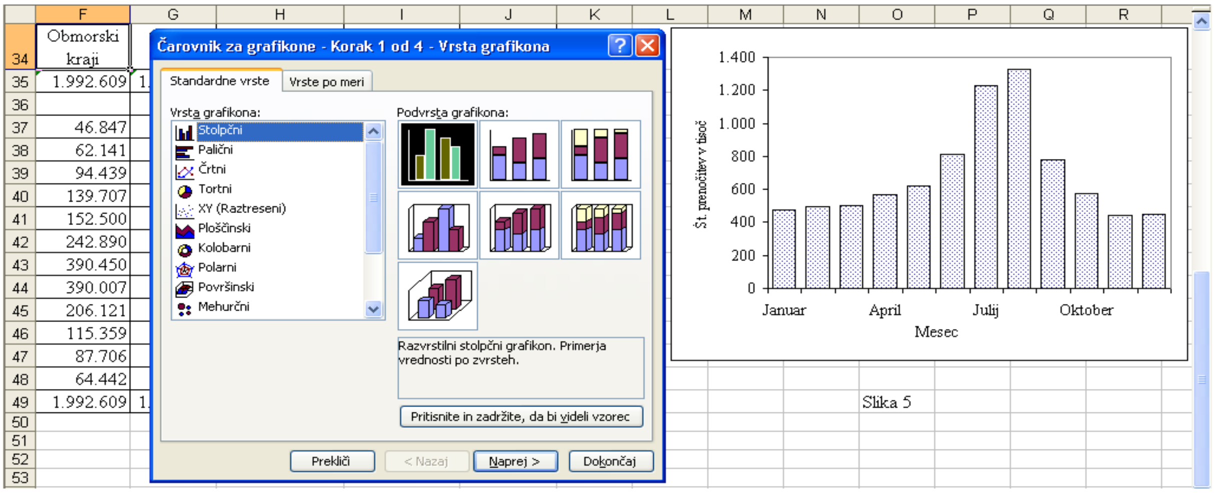
Na spletni strani Statističnega urada R Slovenije odprite Statistični letopis 2008 ter si v poglavju Turizem izberite prikaz s sestavljeno tabelo. Podatke ustrezno prenesite v svojo elektronsko preglednico Excel ter uporabite pravila oblikovanja tabel. Presodite, ali si pri oblikovanju lahko pomagamo tudi s postopkom za samooblikovanje tabel.



Slika 5: Izbor možnosti samooblikovanja tabel v Excelu
Vir: Microsoft Excel

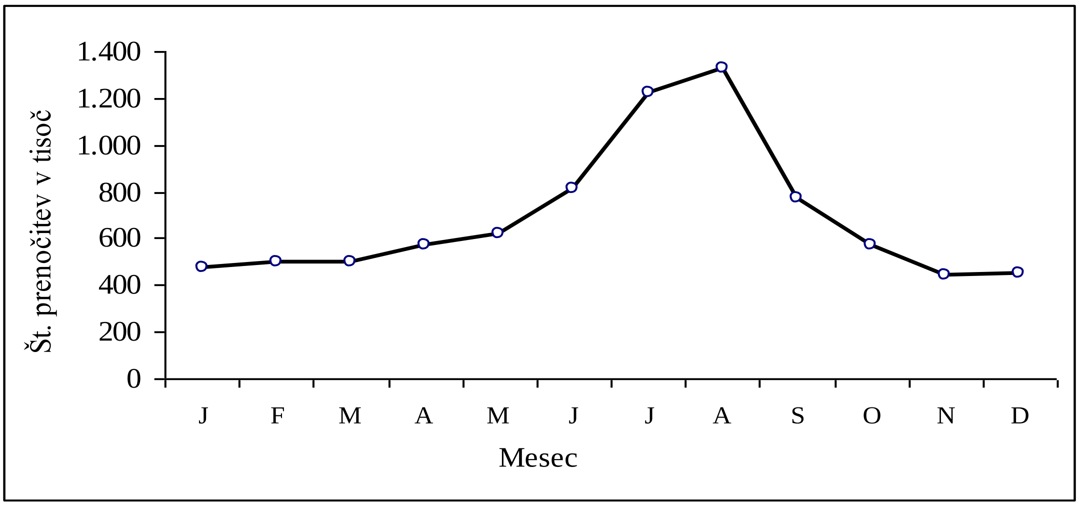
2.4.2 Prikazovanje podatkov z grafi

Uporaba grafikonov najpogosteje pomeni dopolnitev prikazovanja statističnih podatkov v tabelah. Z grafikoni lahko lastnosti opazovanih pojavov bolje poudarimo, kot je to možno storiti v tabelah. Da z njimi dosežemo želeno preglednost, moramo grafikon ustrezno narisati in opremiti. Iz podatkov, nanizanih in oblikovanih npr. v elektronski preglednici Excel, tudi grafe s pomočjo "Čarovnika za grafikone" najprej preprosto kreiramo in nato po svoji želji še dopolnjujemo (Gams, 2007). Grafi morajo biti narisani tako, da realno prikažejo osnovne značilnosti podatkov. Pri tem smo pozorni na pravilni izbor ustrezne vrste grafa, da so v njem zajeti pravilni podatkovni nizi, da je graf primerno opremljen (naslov, legenda, merilo in vrednosti na oseh, obrobe …).



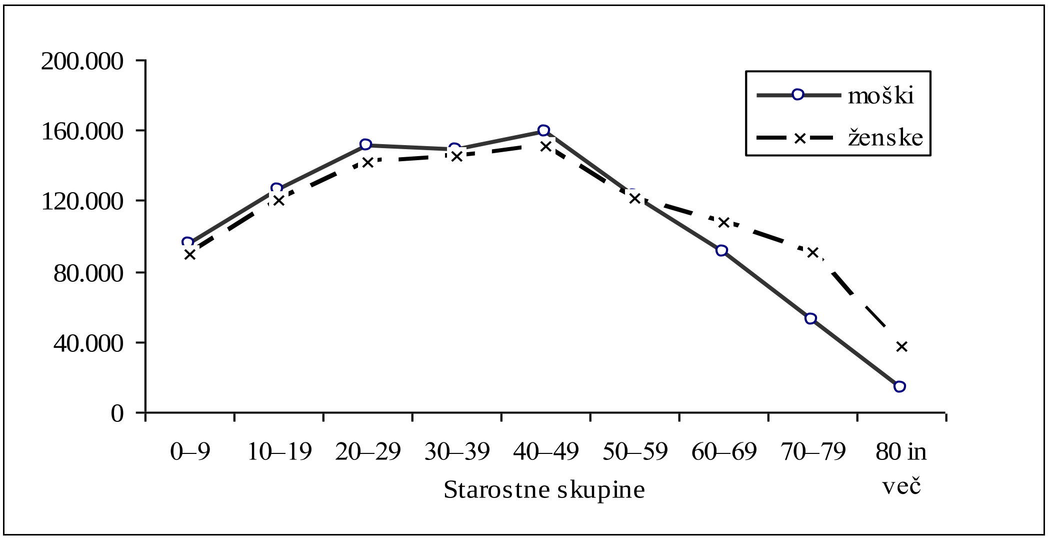
Slika 6: Uvodni korak kreiranja grafikonov v Excelu
Vir: Microsoft Excel

Najpogosteje uporabljamo **linijske (črtne)** in **stolpčne grafikone**. Naslednje slike prikazujejo njihov izgled pri predstavitvah podatkov iz posameznih tabel.



Slika 7: Linijski graf prenočitev turistov v Sloveniji v letu 2007 po mesecih
Vir: Tabela 9

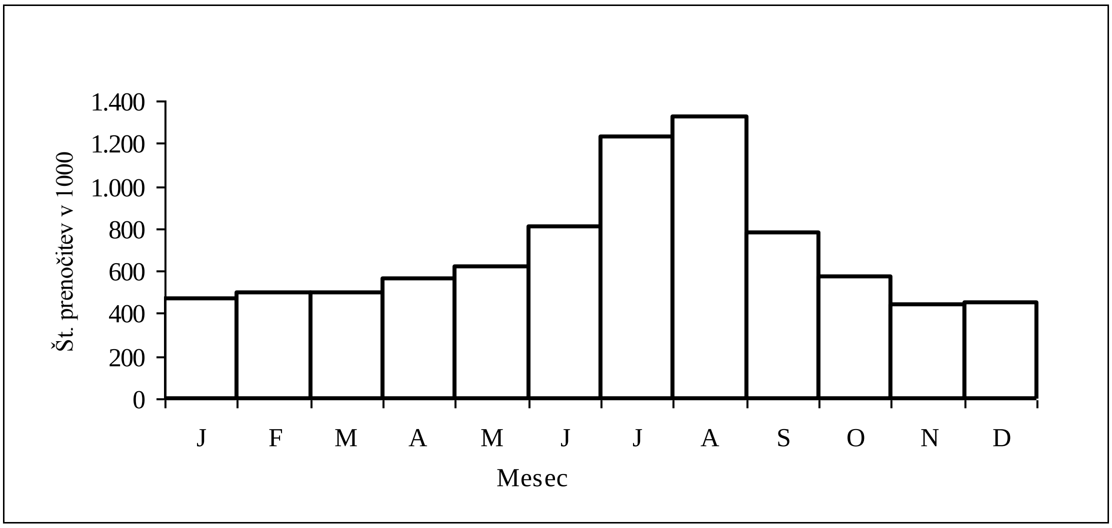
Z linijskimi grafi najpogosteje predstavljamo številske in časovne statistične vrste. Paziti moramo na ustrezno izbiro merila na obeh oseh. Tako navaja Gregorc (2009) ter izpostavlja njihovo primernost tudi za predstavitev trenda časovnih vrst, kar bomo spoznali v 8. poglavju.



Slika 8: Linijski graf predstavitve prebivalcev v R Sloveniji ob popisu 2002 po starostnih skupinah
Vir: Tabela 1

V obeh črtnih grafikonih so točke podanih vrednosti na sredinah intervalov, nanizanih na abscisi, saj je prikazana razmična vrsta. V primerih, kadar je v grafu prikazana trenutna vrsta, kjer pa se točke nanašajo točno nad trenutki, na katere se podatki nanašajo.

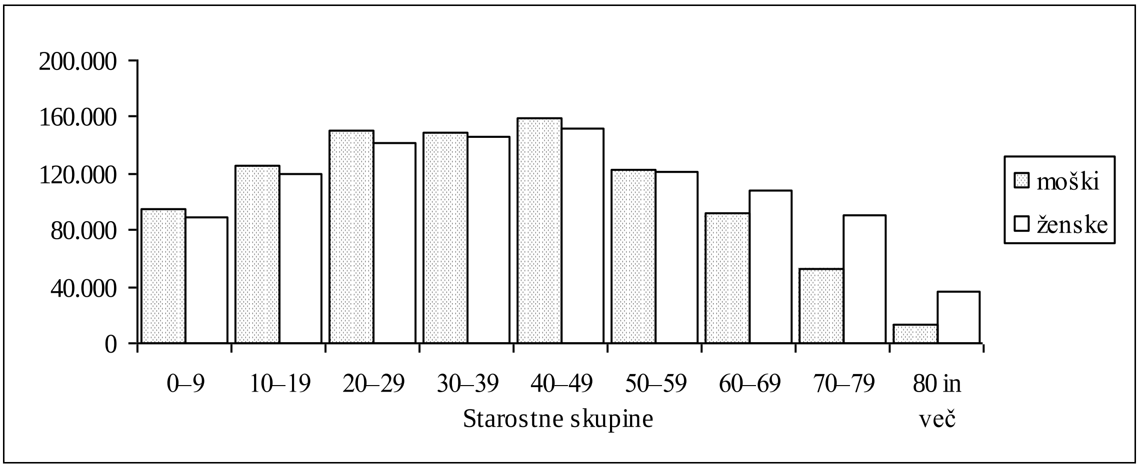
Časovne vrste predstavljamo tudi s stolpčnimi grafi, predvsem, kadar želimo izpostaviti primerjavo sosednjih vrednosti, ki so izražene z višino posameznega stolpca. Sicer pa se s stolpčnimi grafi predstavljajo predvsem opisne statistične vrste, v nadaljevanju pa bomo spoznali, da so zelo uporabni tudi pri prikazovanju struktur pojavov in njihovih primerjav.



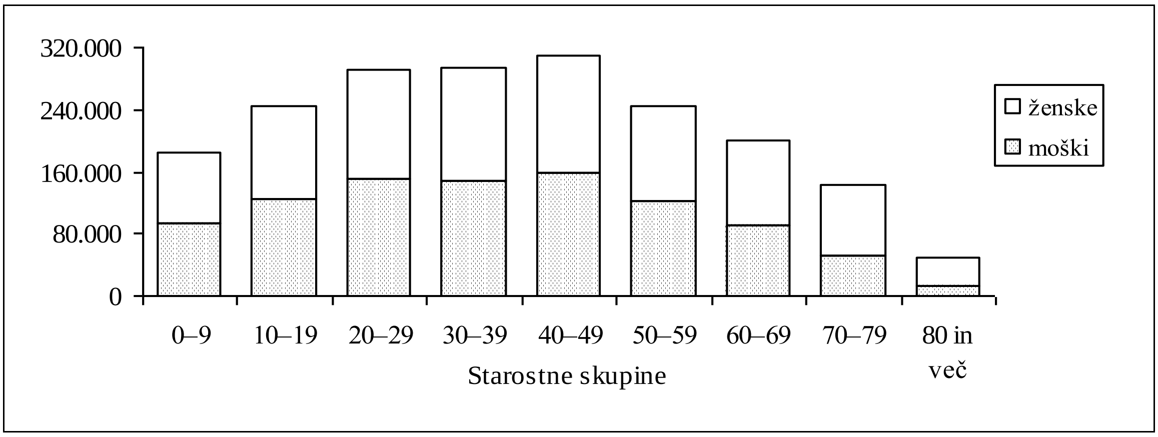
Slika 9: Stolpčni graf prenočitev turistov v Sloveniji v letu 2007 po mesecih
Vir: Tabela 9

Kadar v stolpčnem diagramu predstavljamo podatke dveh ali več statističnih vrst, imamo na razpolago varianto s porazdeljenimi ali s sestavljenimi stolpci. V Excelu lahko preprosto spreminjamo tudi širino stolpcev in njihov medsebojni razmik, s čimer lahko vizualno poudarimo značilnost preučevanega pojava (Gregorc, 2009).

Slika 10 prikazuje število prebivalcev v R Slovenije ob popisu 2002 ločenih po spolu in prikazanih s porazdeljenimi stolpci, slika 11 pa z naloženimi stolpci.



Slika 10: Prebivalci v R Sloveniji ob popisu 2002
Vir: Tabela 1



Slika 11: Prebivalci v R Sloveniji ob popisu 2002
Vir: Tabela 1

Pogled v odprto okno Excelovega "čarovnika" za grafikone na sliki 9 kaže na možnosti kreiranja dodatnih vrst grafikonov in njihovih podvrst. Poglejmo še nekaj nadaljnjih značilnih oblik grafične predstavitve podatkov.

Krožne grafe bomo uporabili pri prikazu strukture opazovanih pojavov. Poln krog (360°) predstavlja celoto ene statistične vrste, njegovi segmenti pa so sorazmerni njenim posameznim delom. Za primerjavo med statističnimi vrstami se uporabljajo različno veliki krogi, postavljeni v soseščini ali pa se uporabi **kolobarni graf**, ki ga tvorijo koncentrični krogi. Tudi tega bomo uporabili pri prikazu struktur. Krajevne statistične vrste so pogosto vrisane v zemljevide v obliki **kartogramov** (npr. regionalna porazdelitev države …) ali **diagramskih kart** (npr. stolpčni prikaz količine padavin …) (Gregorc, 2009).

Na spletni strani Slovenske turistične organizacije si oglejte grafične predstavitve statističnih podatkov. Ob opazovanju jih primerjajte in utemeljite razlike.

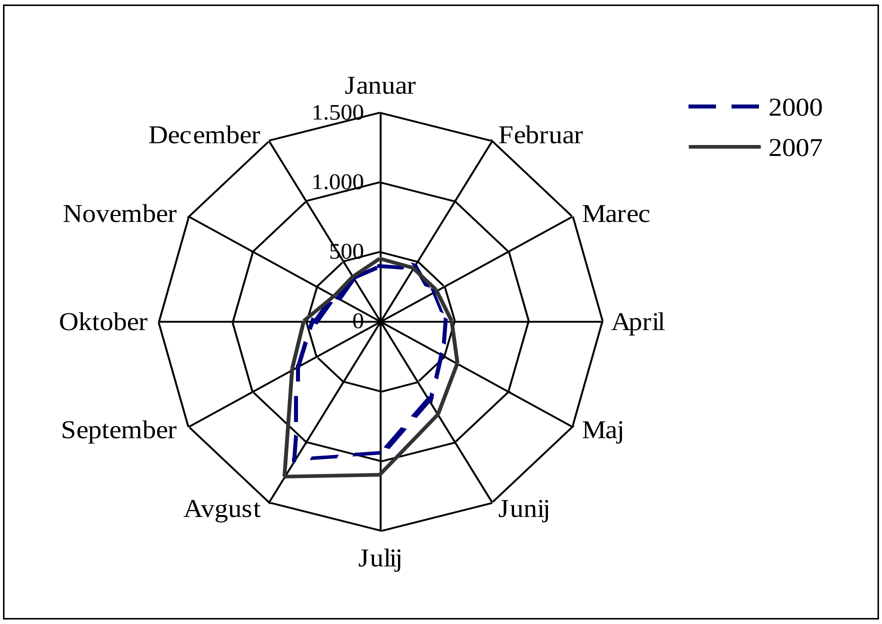
Iz statističnega letopisa prenesite podatke s krajevno statistično vrsto ter zasnujte ustrezno tabelo in graf. Obe predstavitvi dopolnite in izboljšajte s postopki, ki so v Excelu potrebni za oblikovanje tabel ter kreiranje in dopolnjevanje različnih vrst grafov.

Za prikaz časovnih vrst z izrazitejšimi sezonskimi gibanji, kar je na področju turizma pogosta značilnost, uporabljamo polarne grafikone. Tabela 12 ter slika 12 prikazujeta mesečna gibanja prenočitev tujih turistov za leti 2000 in 2007.

Tabela 12: Prenočitve turistov v Sloveniji po mesecih v letih 2000 in 2007

	Število prenočitev (v 1000)	
Mesec	**2000**	**2007**
Januar	389,1	442,5
Februar	429,6	452,1
Marec	402,6	432,8
April	446,8	484,7
Maj	488,3	604,3
Junij	676,1	777,5
Julij	952,8	1.103,9
Avgust	1.163,4	1.284,7
September	644,2	689,1
Oktober	448,8	507,3
November	330,2	355,6
December	347,1	368,2

Vir: SURS, Statistični letopis 2001, http://www.stat.si/letopis/2001/27 (12. 2. 2009), Statistični letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)



Slika 12: Prenočitve turistov v Sloveniji v letih 2000 in 2007 po mesecih
Vir: Tabela 12

Statistično delo se začne s statističnim raziskovanjem oziroma z zbiranjem, urejanjem in prikazovanjem podatkov o preučevanem pojavu. Verodostojnost statističnih analiz in rezultatov je pogojena s kakovostjo podatkov, zato je potrebno tej uvodni fazi statističnega procesa vedno nameniti dovolj časa in naporov: delo najprej skrbno načrtovati, proučiti razpoložljive sekundarne statistične vire, s pomočjo sistematskega opazovanja odgovorno zbrati potrebne podatke za naš primarni vir, podatke urediti in prikazati v pregledni obliki, ki omogoča prepoznavanje značilnosti populacije in izračunavanje parametrov oz. izvajanje nadaljnjih faz statističnega procesa. Za učinkovito analizo podatkov je v vseh fazah proučevanja pomembna njihova pravilnost in preglednost. Utrdili in nadgradili smo znanja sodobne računalniške tehnike pri učinkoviti izdelavi preglednih tabel in grafikonov. Tako smo v obeh dosedanjih poglavjih ugotovili, da že v postopkih zbiranja, urejanja in prikazovanja podatkov obstaja veliko možnosti, da se tovrstno delo in opravila učinkovito podprejo z uporabo sodobne IKT.

Naloge

1. Na spletni strani Slovenske turistične organizacije poiščite v poglavju Raziskave in razvoj objavljene letne, mesečne in ostale statistične podatke (pomoč: http://www.slovenia.info/si/Statistični-podatki.htm?ppg_statisticni_podatki2005=0&lng=1).
   - Primerjajte objavljene podatke, ugotovite bistvene značilnosti ter načine njihovega zbiranja in vire.
   - Določite vrste in oblike statističnih vrst, v katere so podatki urejeni.
   - Presodite značilnosti njihove predstavitve v tabelah in grafih.
   - Poiščite postopek, kako bi lahko izbrane podatke brez napak prenesli v svojo aplikacijo. Razmislite, kaj morate za to pravico postoriti.

2. Na spletni strani Statističnega urada Slovenije odprite poglavje Turizem v Statističnem letopisu 2008.
   - Ugotovite lastnosti statističnih vrst ter oblike in vrste uporabljenih tabel pri naslednjih objavljenih podatkih:
     a) Sobe in ležišča po vrstah nastanitvenih objektov in po vrstah krajev
     b) Prihodi turistov po državni pripadnosti
     c) Prenočitve turistov po državni pripadnosti
     d) Prihodi in prenočitve turistov po vrstah krajev
     e) Prenočitve turistov po vrstah nastanitvenih objektov
     f) Prenočitve turistov po vrstah krajev in po mesecih, 2007
     g) Število prenočitev na turističnih potovanjih domačega prebivalstva
     h) Povprečni dnevni izdatki v evrih na turista na turističnih potovanjih domačega prebivalstva
     i) Število turističnih potovanj domačega prebivalstva, starega 15 let ali več, po najbolj obiskanih državah.
   - Opišite detajle, ki jih izpostavljajo grafikoni in kartogrami v poglavju Turizem.

3. Razmislite in razčlenite postopke zbiranja, urejanja in prikazovanja podatkov, ki ste ga spoznali v času prakse v vašem podjetju.

3 RELATIVNA ŠTEVILA

Podatki, ki so zbrani, urejeni in predstavljeni v tabelah, so navadno izraženi v absolutnih številih ter niso najbolj primerni za medsebojne primerjave. V statistični analizi opazovane pojave pogosto medsebojno primerjamo, npr.: posamezni pojav sorazmernostno primerjamo z drugim pojavom ali istovrstnim pojavom v drugem času ali kraju. Pogosto je potrebna tudi primerjava posameznih delov pojava s celoto in posameznih delov med seboj. Zagato glede primerjanja podatkov rešimo, če za zbrana absolutna števila izračunamo medsebojna razmerja – pretvorimo jih v relativna, npr. deleže v odstotkih. V tem poglavju bomo spoznali enostavne tovrstne statistične kazalce – strukture, statistične koeficiente in indekse – ter ugotovili postopke njihovega izračunavanja in uporabe. Na priloženih primerih bomo relativna števila z uporabo računalnika izračunali ter jih prikazali v tabeli in ustreznem grafu. V naslednjih poglavjih bomo nabor parametrov dopolnili še z zahtevnejšimi merami za izražanje lastnosti opazovanega pojava.

Uvodna naloga

Ocenjujemo obiske tujih turistov v Sloveniji in primerjamo deleže, ki pripadajo gostom iz posameznih držav. Poiščimo postopek, s katerim bomo ugotovili razliko (v odstotnih točkah) ter relativno razliko (v odstotkih) med številom prenočitev tujih turistov iz Italije in Avstrije leta 2007 v naši državi. Primerjajte oba izračunana rezultata ter ju interpretirajte.

Računsko so relativna števila količniki iz dveh absolutnih števil oziroma podatkov (lahko so to istovrstni ali raznovrstni podatki), katerih primerjava je smiselna. Glede na podatke, iz katerih jih računamo, ločimo:

- strukture,
- statistične koeficiente in
- indekse.

3.1 STRUKTURE

Strukture izračunavamo kot razmerje med dvema istovrstnima podatkoma ter jih izražamo v obliki strukturnih deležev, odstotkov ali odtisočkov. Če označimo podatek, ki se nanaša na del pojava z Y_j ter z Y celoto, na katero primerjamo, jih izračunamo, kot navaja Šadl (2004):

- strukturni delež:

$$P_j = \frac{Y_j}{Y} \qquad \text{velja:} \quad \sum P_j = 1$$

- strukturni odstotek (procent):

$$P_j\% = \frac{Y_j}{Y} \cdot 100 \qquad \text{velja:} \quad \sum P_j\% = 100$$

- strukturni odtisoček (promil):

$$P_j\text{‰} = \frac{Y_j}{Y} \cdot 1000 \qquad \text{velja:} \quad \sum P_j\text{‰} = 1000$$

3.1.1 Enorazsežne strukture

O **enorazsežni strukturi** govorimo takrat, kadar je opazovani pojav razčlenjen le po vrednosti ene spremenljivke. Značilen primer izračuna enorazsežne strukture prikazuje naslednja tabela.

Razmislite, kako za spodnje podatke izračunamo strukturni odstotek, npr. za prenočitve turistov iz Italije ter ostalih držav?

Tabela 13: Prenočitve tujih turistov v Sloveniji v letu 2007

Država	**Št. prenočitev** (v 1000)	**Struktura** (v %)
Avstrija	668,9	15,2
Hrvaška	295,6	6,7
Italija	911,2	20,7
Madžarska	134,3	3,1
Nizozemska	199,2	4,5
Nemčija	641,2	14,6
Združeno kraljestvo	304,7	6,9
Druge države	1.244,2	28,3
Skupaj	**4.399,2**	**100,0**

Vir: SURS, Letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)

Strukturni odstotek npr. za prenočitve turistov iz Italije izračunamo:

$$P_j\,\%_{\,Italija} = \frac{Y_j}{Y} \cdot 100 = 20{,}7\ \%$$

kar pomeni, da so turisti iz Italije leta 2007 opravili 20,7 % prenočitev od skupnega števila vseh prenočitev tujih turistov v navedenem letu.

3.1.2 Dvorazsežne strukture

Dvorazsežno strukturo je možno izračunati, če pojav razčlenjujemo po vrednostih dveh spremenljivk. V tem primeru lahko izračunamo tri strukture, in sicer dve, ki izražata razčlenitev pojava z vidika posamezne spremenljivke, in strukturo z vidika obeh spremenljivk hkrati.

Tabela 14: Zaposlene osebe v zdravstvu in socialnem varstvu v R Sloveniji 31. 12. 2007 po spolu in strokovni usposobljenosti

Spol	**Stopnja strokovne usposobljenosti**				**Skupaj**
	podiplomska	**visoka**	**višja**	**srednja**	
moški	338	3.125	606	2.962	7.031
ženske	333	8.054	4.514	16.476	29.377
Skupaj	**671**	**11.179**	**5.120**	**19.438**	**36.408**

Vir: SURS, Letopis 2008, http://www.stat.si/letopis/2008/12 (12. 2. 2009)

Od izbire, kaj v populaciji izberemo za celoto in kaj za dele, lahko iz prejšnjih podatkov izračunamo tri različne strukture:

- strukturo zaposlenih oseb v zdravstvu in socialnem varstvu po spolu (tabela 15),
- strukturo zaposlenih oseb v zdravstvu in socialnem varstvu po stopnji strokovne usposobljenosti (tabela 16) ter
- strukturo zaposlenih oseb v zdravstvu in socialnem varstvu po spolu in stopnji strokovne usposobljenosti (tabela 17).

Uporaba Excela nam računanje struktur močno olajša in poenostavi. Pozorni pa moramo biti na pravilno uporabo naslavljanja celic (absolutno, relativno, kombinacija) s podatki, ki predstavljajo celoto, ter tistimi, ki predstavljajo primerljive dele populacije!

Tabela 15: Struktura zaposlenih oseb po spolu v zdravstvu in socialnem varstvu v R Sloveniji 31. 12. 2007

Spol	**Stopnja strokovne usposobljenosti**				**Skupaj**
	podiplomska	**visoka**	**višja**	**srednja**	
moški	50,4	28,0	11,8	15,2	19,3
ženske	49,6	72,0	88,2	84,8	80,7
Skupaj	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**

Vir: Tabela 14

Tabela 16: Struktura zaposlenih oseb po stopnji strokovne usposobljenosti v zdravstvu in socialnem varstvu v R Sloveniji 31. 12. 2007

Spol	**Stopnja strokovne usposobljenosti**				**Skupaj**
	podiplomska	**visoka**	**višja**	**srednja**	
moški	4,8	44,4	8,6	42,1	100,0
ženske	1,1	27,4	15,4	56,1	100,0
Skupaj	**1,8**	**30,7**	**14,1**	**53,4**	**100,0**

Vir: Tabela 14

Tabela 17: Struktura zaposlenih oseb v zdravstvu in socialnem varstvu v R Sloveniji 31. 12. 2007 po spolu in stopnji strokovne usposobljenosti

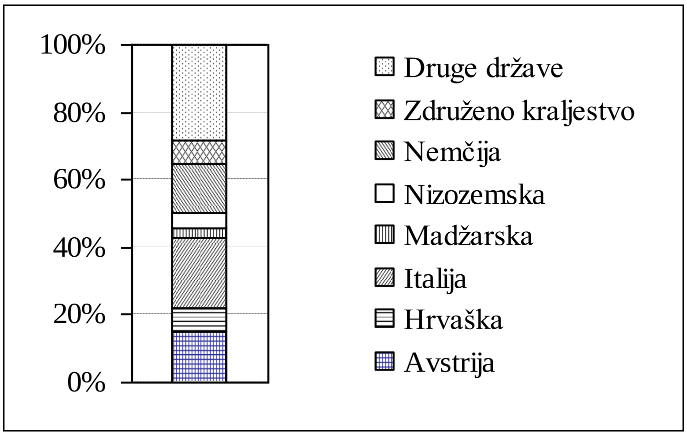
Spol	**Stopnja strokovne usposobljenosti**				**Skupaj**
	podiplomska	**visoka**	**višja**	**srednja**	
moški	0,9	8,6	1,7	8,1	19,3
ženske	0,9	22,1	12,4	45,3	80,7
Skupaj	**1,8**	**30,7**	**14,1**	**53,4**	**100,0**

Vir: Tabela 14

Predstavljeni primeri potrjujejo, da izračunane strukture ena drugo dopolnjujejo, zato v praksi izberemo tiste kazalce, ki najbolje osvetljujejo analizirani pojav z vidika ciljev proučevanja.

3.1.3 Grafično prikazovanje struktur

Prikaz s strukturnim stolpcem



Slika 13: Struktura prenočitev tujih turistov v Sloveniji v letu 2007
Vir: Tabela 13

Tabela 18: Prenočitve tujih turistov v Sloveniji v letih 1995, 2000 in 2007

Država	Število prenočitev v 1000			Struktura (v %)		
	1995	2000	2007	1995	2000	2007
Avstrija	440,7	527,0	668,9	18,1	15,5	15,2
Hrvaška	212,8	251,1	295,6	8,7	7,4	6,7
Italija	387,8	650,6	911,2	15,9	19,1	20,7
Madžarska	58,1	86,2	134,3	2,4	2,5	3,1
Nizozemska	83,4	125,2	199,2	3,4	3,7	4,5
Nemčija	571,6	772,8	641,2	23,5	22,7	14,6
Združeno kraljestvo	65,8	152,5	304,7	2,7	4,5	6,9
Druge države	615,3	838,7	1.244,2	25,3	24,6	28,3
Skupaj	**2.435,5**	**3.404,1**	**4.399,2**	**100,0**	**100,0**	**100,0**

Vir: SURS, Letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)

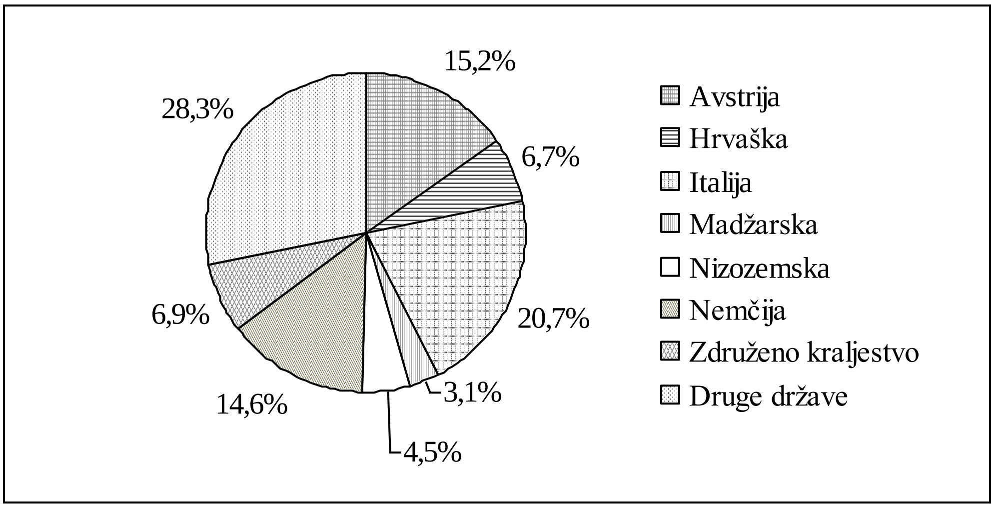
Slika 14: Struktura prenočitev tujih turistov v Sloveniji v letih 1995, 2000 in 2007
Vir: Tabela 13

Prikaz s strukturnim krogom

Pri morebitnem ročnem izrisu prikaza struktur s strukturnim krogom moramo strukturne odstotke – prikazane kot krožne izseke – preračunati v ločne stopinje, kar storimo tako, da jih pomnožimo s 3,6; poln krog 360° namreč predstavlja celoto 100 %.

Torej: $P^{o} = 3{,}6 \cdot P\%$

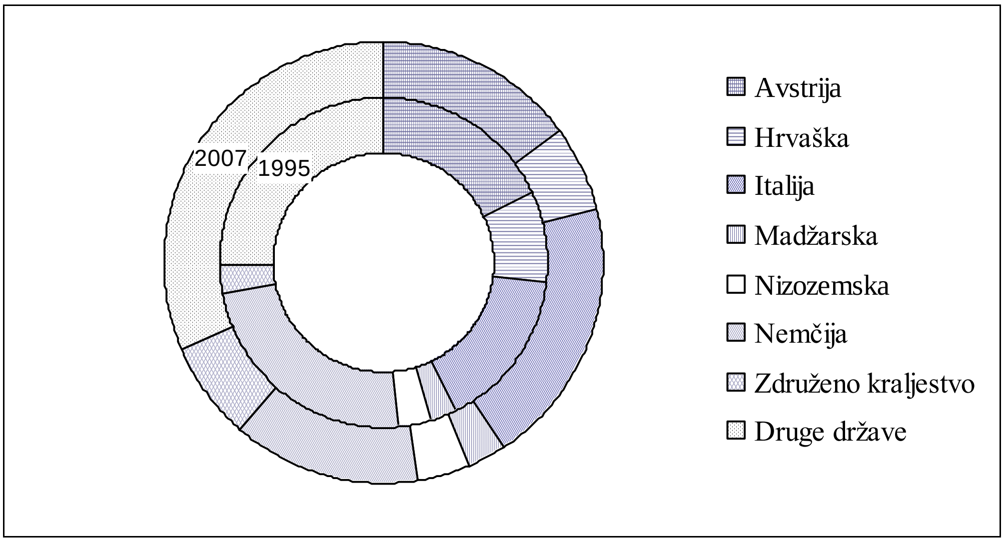
Pri kreiranju tovrstnih prikazov z elektronsko preglednico (npr. z Excelom) nam to uskladitev aplikacija preračuna samodejno.



Slika 15: Struktura prenočitev tujih turistov v Sloveniji v letu 2006
Vir: Tabela 13

Prikaz struktur z dvema krogoma oz. s kolobarjem

Enorazsežne strukture dveh pojavov je možno prikazati tudi s kolobarnim grafom.



Slika 16: Struktura prenočitev tujih turistov v Sloveniji v letih 1995 in 2006
Vir: Tabela 18

3.2 STATISTIČNI KOEFICIENTI

3.2.1 Kaj so statistični koeficienti?

Statistični koeficient je – kot navaja Šadl (2004) – količnik med dvema raznovrstnima podatkoma, ki pa sta v medsebojni vsebinski zvezi, saj morajo biti izpolnjeni naslednji pogoji:

- podatka sta vsebinsko smiselna,
- sta enako krajevno opredeljena,
- nanašata se na isti časovni trenutek ali razmik.

Če z X in Y označimo na raznovrstna pojava nanašajoče se podatke, koeficient izračunamo kot razmerje:

$$K = \frac{Y}{X} \cdot E$$

V imenovalcu je podatek, na osnovo katerega koeficient računamo, E pa je faktor, ki je odvisen od tega, na koliko enot podatka v imenovalcu koeficient računamo (1, 10, 100, 1.000, 1.000.000).

Kadar je smiselna primerjava podatkov, ki ju v števcu in imenovalcu zamenjamo, računamo tudi recipročni (obratni) koeficient K_{rec}.

$$K_{rec} = \frac{X}{Y} \cdot E$$

Iz podatkov v spodnji tabeli bomo izračunali nekatere koeficiente. Pred izračunom pa presodite, katere od navedenih podatkov je v izračunih koeficientov smotrno primerjati.

Tabela 19: Osnovne šole, učenci in učitelji v R Sloveniji v šolskih letih od 2000/01 do 2006/07

Šolsko leto	Št. šol	Št. učencev	Št. učiteljev
2000/01	448	180.874	15.382
2001/02	447	177.755	15.625
2002/03	448	175.743	16.039
2003/04	446	177.083	17.145
2004/05	446	171.358	17.446
2005/06	447	167.616	17.713
2006/07	447	164.991	17.671

Vir: SURS, Statistični letopis 2008, http://www.stat.si/letopis/2008/06 (12. 2. 2009)

Pri izračunavanju statističnih koeficientov primerjamo podatka, ki se nanašata na različni populaciji, vendar pa sta v vsebinski zvezi.

V navedenem primeru smo se odločili za koeficiente, ki nam kažejo razmerja med naslednjimi, v prejšnji tabeli navedenimi podatki:

Tabela 20: Število učencev na šolo in na učitelja ter število učiteljev na šolo v R Sloveniji v šolskih letih od 2000/01 do 2006/07

Šolsko leto	**Štev. učencev na šolo**	**Štev. učencev na učitelja**	**Štev. učiteljev na šolo**
2000/01	403,7	11,8	34,3
2001/02	397,7	11,4	35,0
2002/03	392,3	11,0	35,8
2003/04	397,0	10,3	38,4
2004/05	384,2	9,8	39,1
2005/06	375,0	9,5	39,6
2006/07	369,1	9,3	39,5

Vir: Tabela 19

Za šolsko leto 2001/02 velja, da je število učiteljev na šolo enako $K = \frac{Y}{X} = \frac{15382}{448} = 34{,}3$.

3.2.2 Primerjava podatka, ki se nanaša na časovni interval, s podatkom, ki se nanaša na časovni trenutek

Kadar se eden od primerjanih podatkov nanaša na časovni interval, drugi pa na časovni trenutek, ju v takšni obliki ni možno neposredno primerjati. Podatke, ki se nanašajo na časovne trenutke, moramo najprej prilagoditi tako, da izračunamo povprečje, ki se nanaša na tisti časovni interval, za katerega upoštevamo intervalni podatek. V statističnih publikacijah so tudi pogosto že objavljeni podatki, ki izražajo povprečje, npr. število zaposlenih sredi meseca in podobno.

Obrazec za izračun koeficienta se spremeni v:

$$K = \frac{Y}{\overline{X}} \cdot E$$

kjer je $\overline{X}$ izračunano povprečje podatkov, nanašajočih se na časovne trenutke.

Postopek za izračun povprečja je odvisen od trenutka, v katerem pojav opazujemo, in sicer, kot navaja Trstenjak (2001):

- če pojav opazujemo v sredini obdobij, izračunamo povprečje iz N podatkov:

$$\overline{X} = \frac{1}{N}(X_1 + X_2 + \ldots + X_N)$$

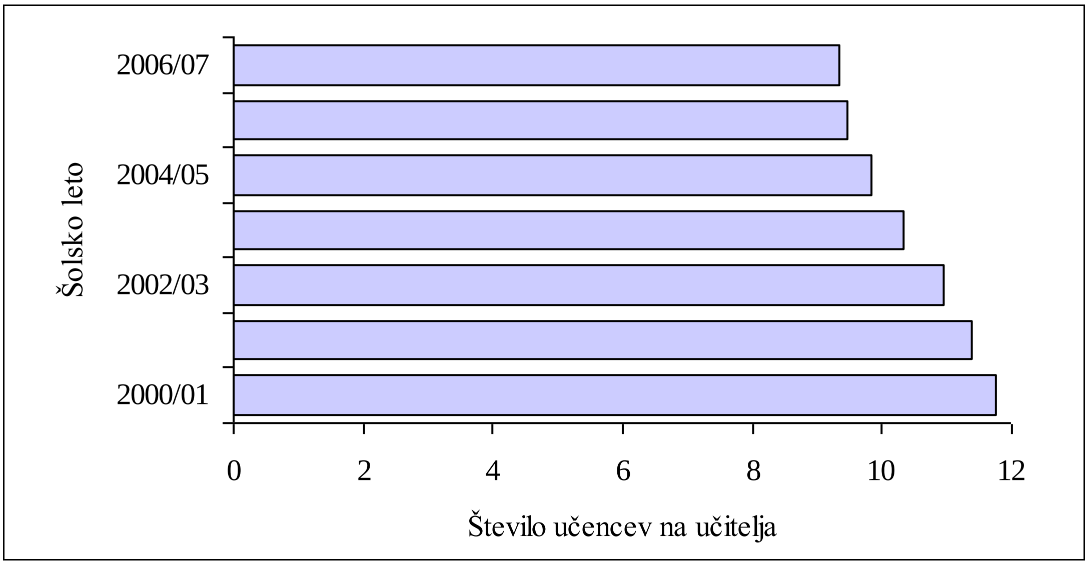
- kadar pojav opazujemo na začetku ali koncu obdobij, najprej izračunamo povprečja za posamezna obdobja, nato pa iz teh povprečij izračunamo povprečje za celotno opazovano obdobje tako:

$$\overline{X} = \frac{1}{N}(\overline{X}_1 + \overline{X}_2 + \ldots + \overline{X}_N) = \frac{1}{N}\left(\frac{X_0 + X_1}{2} + \frac{X_1 + X_2}{2} + \ldots + \frac{X_{N-1} + X_N}{2}\right) =$$

$$= \frac{1}{N}\left(\frac{X_0}{2} + X_1 + X_2 + \ldots + X_{N-1} + \frac{X_N}{2}\right)$$

3.2.3 Grafično prikazovanje statističnih koeficientov

Statistične koeficiente lahko zelo nazorno prikažemo s stolpci v navpični oz. trakovi v vodoravni legi, njihova dolžina je sorazmerna vrednosti koeficienta.



Slika 17: Število učencev na učitelja v šolskih letih od 2000/01 do 2006/07
Vir: Tabela 20

3.3 INDEKSI

Indeksi so relativna števila, pri katerih primerjamo dva istovrstna podatka, ki se nanašata na različen kraj, čas ali stvarno področje. Indeksi torej izražajo razmerje med dvema členoma krajevne, časovne ali stvarne vrste. Za razliko od struktur in koeficientov, pri katerih lahko primerjamo le absolutne podatke, lahko računamo indekse tudi iz drugih relativnih števil. Indeksi so neimenovana števila, izračunana po obrazcu:

$$I_{j/0} = \frac{Y_j}{Y_0} \cdot 100$$

Y_j – podatek, ki ga primerjamo s podatkom v imenovalcu

Y_0 – podatek, ki je osnova (baza) za primerjavo

Če od indeksa odštejemo vrednost 100, dobimo relativno razliko (stopnjo):

$D_{j/0}\% = I_{j/0} - 100$, ki je izražena v odstotkih

3.3.1 Krajevni indeksi

O **krajevnih indeksih** govorimo, če primerjamo dva istovrstna podatka, ki se nanašata na dve različni geografski področji.

Izračunajmo indekse za število internetnih priključkov v letu 2007 ter na osnovi njihovih vrednostih razmislimo o nivoju uporabe IKT v nekaterih evropskih državah.

Tabela 21: Število internetnih priključkov na 100 gospodinjstev v letu 2007 v nekaterih evropskih državah ter izračunani krajevni indeksi

Država	**Štev. int. priključkov** (na 100 gospodinjstev)	**Indeks** (Slovenija = 100)	**Stopnja** (v %)
Avstrija	60	103,4	3,4
Belgija	60	103,4	3,4
Bolgarija	19	32,8	–67,2
Češka republika	35	60,3	–39,7
Danska	78	134,5	34,5
Estonija	53	91,4	–8,6
Francija	49	84,5	–15,5
Grčija	25	43,1	–56,9
Italija	43	74,1	–25,9
Luksemburg	75	129,3	29,3
Madžarska	38	65,5	–34,5
Nemčija	71	122,4	22,4
Nizozemska	83	143,1	43,1
Poljska	41	70,7	–29,3
Portugalska	40	69,0	–31,0
Slovaška	46	79,3	–20,7
Slovenija	**58**	**100,0**	**0,0**
Španija	45	77,6	–22,4
Švedska	79	136,2	36,2
Združeno kraljestvo	67	115,5	15,5

Vir: SURS, Statistični letopis 2008, http://www.stat.si/letopis/2008/33 (12. 2. 2009), http://epp.eurostat.ec.europa.eu (12. 2. 2009)

Vrednosti v prvi vrstici izračunamo: $I_{j/0} = I_{Avstrija/Slovenija} = \frac{60}{58} \cdot 100 = 103{,}4$

$$D_{j/0}\% = I_{Avstrija/Slovenija} - 100 = 103{,}4 - 100 = 3{,}4\,\%$$

3.3.2 Časovni indeksi

Časovne indekse uporabljamo najpogosteje. Računamo jih z razmerjem dveh istovrstnih podatkov, nanašajočih se na različna časovna trenutka ali intervala ter izražajo relativne spremembe pojava v času. Iz časovnih vrst lahko izračunamo:

- **indekse s stalno osnovo** $I_{j/0}$, ko posamezne podatke za dano obdobje v časovni vrsti Y_j primerjamo vedno z istim podatkom Y_0, določenim za osnovo:

$$I_{j/0} = \frac{Y_j}{Y_0} \cdot 100$$

- **verižne indekse** V_j, ko vsak člen časovne vrste v danem obdobju Y_j primerjamo s členom Y_{j-1} prejšnjega obdobja; osnova je torej predhodni člen:

$$I_{j/j-1} = V_j = \frac{Y_j}{Y_{j-1}} \cdot 100$$

Izračunajmo indekse s stalno osnovo in verižne indekse za število prvič registriranih osebnih avtomobilov v R Sloveniji v letih 1998 do 2007. Pojasnite značilnosti obeh izračunanih relativnih številih.

Tabela 22: Število prvič registriranih osebnih avtomobilov v R Sloveniji v letih od 1998 do 2007 in izračunani indeksi s stalno osnovo in verižni indeksi

Leto	**Štev. registr. osebnih avtomobilov**	**Indeksi s stalno osnovo** (1998 = 100)	**Verižni indeksi**
1998	67.283	100,0	-
1999	77.973	115,9	115,9
2000	62.444	92,8	80,1
2001	53.651	79,7	85,9
2002	61.278	91,1	114,2
2003	71.605	106,4	116,9
2004	78.808	117,1	110,1
2005	79.438	118,1	100,8
2006	81.539	121,2	102,6
2007	94.364	140,2	115,7

Vir: SURS, Statistični letopis 2002, http://www.stat.si/letopis/index_vsebina.asp (12. 2. 2009)
Statistični letopis 2008, http://www.stat.si/letopis/2008/21 (12. 2. 2009)

Za leto 2007 velja: $I_{j/0} = I_{2007/1998} = \frac{94364}{67283} \cdot 100 = 140{,}2$

$$V_j = V_{2007} = \frac{Y_j}{Y_{j-1}} \cdot 100 = \frac{94364}{81539} \cdot 100 = 115{,}7$$

Pogled na izračunane vrednosti obeh vrst indeksov v tabeli pokaže dve posebnosti, zato razmislite, zakaj v prvi vrsti (za leto 1998) ne moremo izračunati verižnega indeksa ter kje je razlog, da sta v drugi vrsti (za leto 1999) vrednosti obeh indeksov enaki.

3.3.3 Koeficienti in stopnje rasti

Za analizo sprememb časovnih vrst uporabljamo tudi kazalca:

- **koeficient rasti** K_j, ta izraža razmerje med dvema podatkoma, ki se nanašata na zaporedna časovna trenutka ali intervala; izračunamo ga kot verižni indeks in kaže relativne spremembe členov v časovni vrsti:

$$K_j = \frac{Y_j}{Y_{j-1}}$$

- **stopnjo rasti** S_j, ta kaže neposredno relativno razliko med dvema podatkoma, ki se nanašata na zaporedna časovna trenutka ali intervala; izračunamo jo:

$$S_j = \frac{Y_j - Y_{j-1}}{Y_{j-1}} \cdot 100$$

Proučimo značilne vrednosti kazalcev dinamike pri spreminjanju pojava, in sicer: kadar le-ta upada, narašča oziroma se časovno ne spreminja.

Vrednosti kazalcev dinamike, ko se pojav časovno spreminja, kot navaja Šadl (2004):

Kazalec dinamike	Oznaka	Pojav upada	Pojav narašča	Pojav se ne spreminja
Verižni indeks	V_j	$V_j < 100$	$V_j > 100$	$V_j = 100$
Koeficient rasti	K_j	$K_j < 1$	$K_j > 1$	$K_j = 1$
Relativna razlika	D_j	$D_j < 0$	$D_j > 0$	$D_j = 0$
Stopnja rasti	S_j	$S_j < 0$	$S_j > 0$	$S_j = 0$

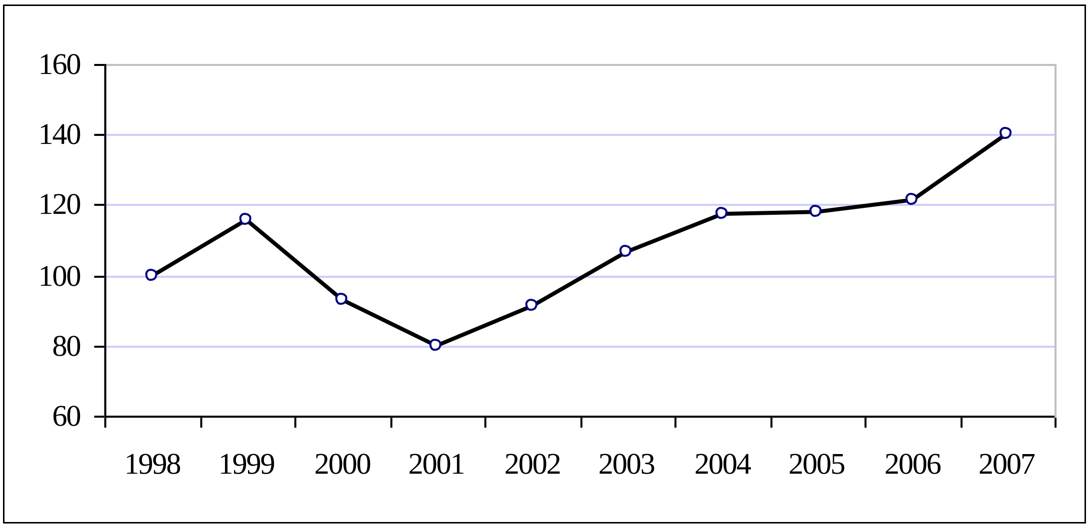
3.3.4 Povezave med kazalci dinamike

Šadl (2004) navaja, da veljajo med obravnavanimi kazalci dinamike naslednje povezave:

- $S_j = \frac{Y_j - Y_{j-1}}{Y_{j-1}} \cdot 100 = \frac{Y_j}{Y_{j-1}} \cdot 100 - \frac{Y_{j-1}}{Y_{j-1}} \cdot 100 = \frac{Y_j}{Y_{j-1}} \cdot 100 - 100 = V_j - 100$
- $S_j = V_j - 100 = 100 \cdot K_j - 100$ in
- $S_j = 100 \cdot (K_j - 1)$, iz česar sledi: $K_j = \frac{S_j}{100} + 1$

3.3.5 Grafično prikazovanje indeksov

Za grafični prikaz indeksov najpogosteje uporabljamo linijski grafikon in stolpce. Prvi se prioritetno uporabljajo za indekse s stalno osnovo, predvsem pri prikazu indeksnih vrst za več pojavov, ko želimo primerjati njihovo dinamiko.



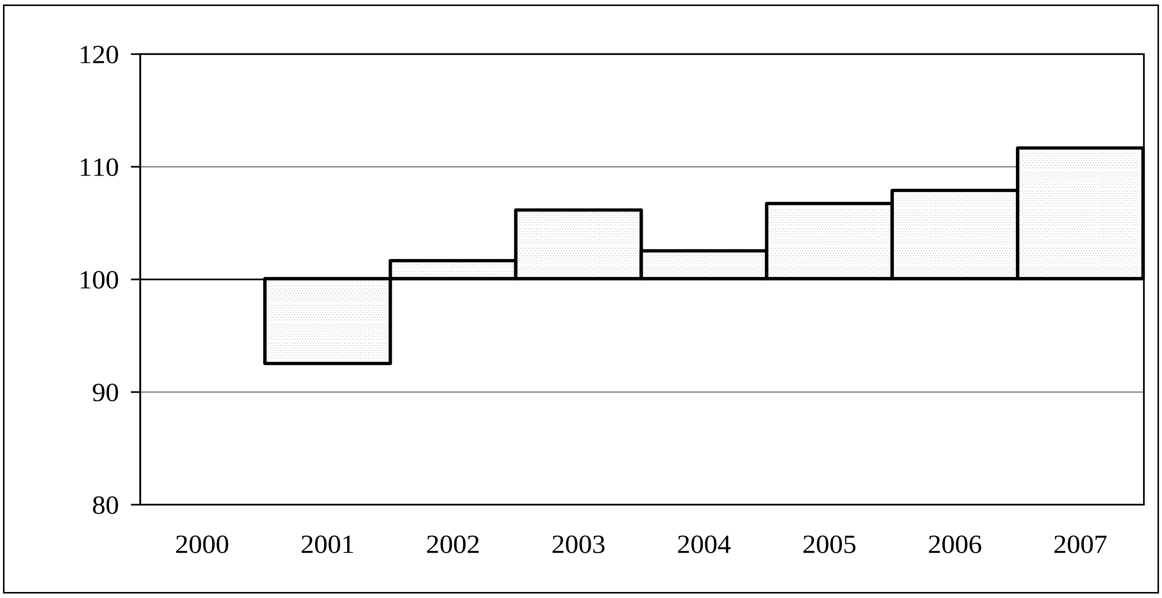
Slika 18: Indeksi s stalno osnovo za število prvič registriranih osebnih avtomobilov v R Sloveniji v letih od 1998 do 2007
Vir: Tabela 22

Verižne indekse prikazujemo s stolpci; ti so za večje vrednosti indeksov – tiste nad osnovo ($V_j = 100$) – obrnjeni navzgor, za manjše indekse pa navzdol. Tako lahko iz grafikona tudi neposredno odčitamo stopnje rasti.

Tabela 23: Število z letali prepeljanih potnikov v R Sloveniji v letih od 2000 do 2007 in izračunani verižni indeksi

Leto	**Štev. potnikov** (v 1000)	**Verižni indeksi**
2000	866	-
2001	801	92,5
2002	814	101,6
2003	864	106,1
2004	885	102,4
2005	944	106,7
2006	1.018	107,8
2007	1.136	111,6

Vir: SURS, Statistični letopis 2003, http://www.stat.si/letopis/2003/22 (12. 2. 2009), Statistični letopis 2008, http://www.stat.si/letopis/2008/21 (12. 2. 2009)



Slika 19: Verižni indeksi za število potnikov, prepeljanih z letali v R Sloveniji v letih od 2000 do 2007
Vir: Tabela 23

3.4 RAZLIKA IN RELATIVNA RAZLIKA ZA RELATIVNA ŠTEVILA

Tudi za relativna števila je smiselno izračunavati **razliko** in **relativno razliko.** Pri teh izračunih moramo upoštevati posebnosti, ki izhajajo iz definicije posameznih relativnih števil, zajetih v dve skupini: prevladujočih neimenovanih (strukture, indeksi, stopnje) ter imenovanih relativnih števil (statistični koeficienti). Skladno z ugotovitvami tega podpoglavja pa bomo lahko tudi rešili našo uvodno nalogo.

3.4.1 Razlika in relativna razlika med imenovanima relativnima številoma

Razlika med dvema imenovanima relativnima številoma je tudi imenovano število. Relativna razlika med dvema imenovanima relativnima številoma pa je neimenovano število.

V tabeli 2 imamo zbrane značilne podatke ob posameznih uradnih popisih prebivalcev. Ugotovite razliko in relativno razliko za povprečni števili članov v gospodinjstvih v letih 2002 ter 1931. Bodite pozorni, kako sta izraženi obe izkazani vrednosti.

Tabela 2 navaja podatka, da je bilo na sedanjem ozemlju Slovenije leta 1931 ob popisu prebivalstva povprečno 4,9 članov na gospodinjstvo, ob popisu prebivalstva 2002 pa le 2,8 članov na gospodinjstvo. Za izračunana statistična koeficienta znašata obe razliki:

- razlika: $D_{2002/1931} = 2{,}8 - 4{,}9 = -2{,}1$ člana na gospodinjstvo
- relativna razlika: $D_{2002/1931}\,\% = 100 \cdot \dfrac{2{,}8 - 4{,}9}{4{,}9} = -42{,}6\ \%$

Izračunana rezultata povesta, da je imelo v Sloveniji gospodinjstvo leta 2002 povprečno 2,1 člana manj kot leta 1931, oz. se je v navedenem obdobju povprečno število članov na gospodinjstvo zmanjšalo za 42,6 %.

3.4.2 Razlika in relativna razlika med neimenovanima relativnima številoma

Med dvema neimenovanima številoma (kot so strukture, indeksi, stopnje) je razlika izražena v odstotnih točkah, relativna razlika pa v % (ali ‰).

Rešitev uvodne naloge

Pridobljeno znanje nam sedaj omogoča, da rešimo nalogo, zastavljeno v uvodu poglavja za izračun razlike ter relativne razlike med številoma prenočitev turistov iz Italije in Avstrije leta 2007 v naši državi. Oba izračunana rezultata še primerjajte ter ju razložite.

V tabeli 13 smo na osnovi statističnih podatkov SURS izračunali, da so leta 2007 v prenočitvah vseh tujih turistov prenočitve turistov iz Italije znašale 20,7 %, prenočitve turistov iz Avstrije pa 15,2 %. Za navedena strukturna odstotka iskani vrednosti znašata:

- razlika: $D_{Italija/Avstrija} = 20{,}7 - 15{,}2 = 5{,}5$ odstotnih točk
- relativna razlika: $D_{Italija/Avstrija}\,\% = 100 \cdot \dfrac{20{,}7 - 15{,}2}{15{,}2} = 36{,}2\ \%$

Pri računanju razlike in relativne razlike med relativnima številoma je važno razlikovanje pomena in načina računanja obeh vrednosti ter tolmačenje rezultatov. Predvsem za relativno razliko, ki je pri vseh relativnih številih izražena v %, ne smemo spregledati potrebe, da interpretacijo rezultata pravilno povežemo s konkretno primerjalno vrednostjo, ki je bila izbrana za opazovani pojav.

V tem poglavju smo spoznali vrsto relativnih števil, ki jih uporabljamo kot najenostavnejše statistične parametre. Naučili smo se izračunati strukture, statistične koeficiente in indekse, na priloženih primerih spoznali pravilno izbiro relativnega števila ter ugotovili njihovo uporabnost pri primerjavi statističnih podatkov. Kadar želimo podatke opazovanega pojava razčleniti oziroma prikazati njegovo sestavo, uporabljamo strukturna števila, ko primerjamo raznovrstne, vendar vsebinsko povezane (časovno in krajevno enako opredeljene) podatke, računamo statistične koeficiente, z najpogosteje uporabljanimi indeksi pa primerjamo istovrstne urejene podatke in analiziramo njihovo dinamiko. V tem poglavju smo se naučili dinamiko pojavov spremljati z navedenimi kazalci dinamike. Naše proučevanje časovnih vrst bomo v 8. poglavju še razširili in poglobili s spoznavanjem dejavnikov spreminjanja pojavov, računanjem parametrov ter iskanjem metod in postopkov, ki ob določenih pogojih omogočajo napoved nadaljnjega razvoja pojava. V primerih, uporabljenih v tem poglavju, smo tudi izpostavili pomen računalniške podpore pri izračunavanju relativnih števil ter grafični predstavitvi njihovih vrednosti.

<u>Naloge</u>

1. Opišite relativna števila, ki jih uporabljate pri ostalih predmetih v modulih Ekonomika in poslovanje, Komunikacija, Trženje ter Poslovanje v gostinstvu in turizmu. Pojasnite, kako so ta števila pri teh predmetih poimenovana in določite, v katero skupino lahko te kazalce uvrstimo.

2. V tabeli 7 zbrani podatki predstavljajo število študentov višjih strokovnih šol, vpisanih v programe posameznih področij v študijskem letu 2007/08. Izračunajte:
   - strukturo študentov po področjih,
   - strukturo študentov po načinu študija (redni oz. izredni) ter
   - strukturo študentov po področjih in načinu študija.

3. Podatke v tabeli 11, v obdobju od leta 2000 do 2007, dopolnite z izračunanimi indeksi s stalno osnovo (2000 = 100) ter z verižnimi indeksi.

4. Tabelo 23 dopolnite z izračunanimi indeksi s stalno osnovo, pri izbranem (2000 = 100) ter jih grafično predstavite.

5. Na spletni strani Statističnega urada Slovenije odprite v Statističnem letopisu 2008 poglavje Turizem. Pred vami so tabele s statističnimi vrstami, ki ste jih proučevali v nalogah prejšnjega poglavja. Iz spodnjega nabora tabel poiščite tiste, ki omogočajo računanje struktur, statističnih koeficientov ter časovnih in krajevnih indeksov. Za vsako od navedenih relativnih števil tudi izračunajte vrednosti v po eni izbrani tabeli ter zanje kreirajte tudi grafično predstavitev.
   - Sobe in ležišča po vrstah nastanitvenih objektov in po vrstah krajev
   - Prihodi turistov po državni pripadnosti
   - Prenočitve turistov po državni pripadnosti
   - Prihodi in prenočitve turistov po vrstah krajev
   - Prenočitve turistov po vrstah nastanitvenih objektov
   - Prenočitve turistov po vrstah krajev in po mesecih, 2007
   - Število prenočitev na turističnih potovanjih domačega prebivalstva
   - Povprečni dnevni izdatki v evrih na turista na turističnih potovanjih domačega prebivalstva
   - Število turističnih potovanj domačega prebivalstva, starega 15 let ali več, po najbolj obiskanih državah.

4 FREKVENČNE PORAZDELITVE

*Ob statističnem raziskovanju imamo običajno zbrano množico podatkov, nanašajočih se na vrednosti spremenljivk niza enot opazovane populacije. Da si olajšamo prepoznavanje značilnosti populacije in izračunavanje njenih parametrov, podatke uredimo in prikažemo v pregledni obliki***.** *Pri manjšem številu vrednosti spremenljivke je to ranžirna vrsta, pri njihovem večjem številu pa statistična vrsta, ki jo imenujemo frekvenčna porazdelitev. V njej prikažemo razdelitev posameznih vrednosti za proučevani pojav – preprosteje povedano: preštejemo, kolikokrat se posamezna vrednost pojavi. V tem poglavju bomo spoznali pravila sestavljanja podatkov v frekvenčne porazdelitve, značilne za posamezne vrste spremenljivk, se naučili frekvenčno porazdelitev opisati ter vrednosti v njej nazorno grafično predstaviti.*

Uvodna naloga

Poiščite uradne podatke o uspehu dijakov na poklicni maturi 2007 (pomoč: spletna stran Državnega izpitnega centra http://www.ric.si/). Podatke pregledno razporedite v ustrezno število razredov, določite in razložite pomen osnovnih kazalcev, porazdelitev opišite, jo grafično predstavite ter interpretirajte značilnosti dobljene razporeditve.

V točki 2.3 smo spoznali, da krajše podatkovne nize uredimo v ranžirne vrste, za obsežnejše množice pa bomo uporabili frekvenčne porazdelitve. »Frekvenčna porazdelitev je statistična vrsta, ki prikazuje sestavo velikosti populacije po skupinah izbrane spremenljivke (najpogosteje po razredih številske spremenljivke). Število enot v posameznem razredu imenujemo frekvenca razreda.« (Korenjak, 2008, 2).

4.1 OBLIKOVANJE FREKVENČNIH PORAZDELITEV

Pri oblikovanju frekvenčnih porazdelitev je potrebno zagotoviti enoličnost opredelitve razredov in preglednost podatkov, upoštevaje značilnosti opisne in številske spremenljivke.

4.1.1 Frekvenčne porazdelitve za opisne spremenljivke

Pri opisnih spremenljivkah z manjšim številom vrednosti izberemo njim enako število skupin (spol, učni uspeh …). Kadar pa je število vrednosti veliko, v izbranih skupinah združujemo sorodne vrednosti, kot npr. v spodnji tabeli.

Tabela 24: Prihodi in prenočitve turistov po vrstah krajev v R Sloveniji v letu 2007

Vrste krajev	**Prihodi turistov** (v 1000)	**Prenočitve turistov** (v 1000)
Ljubljana	372,2	707
Zdraviliški kraji	632,5	2.651
Obmorski kraji	545,5	1.993
Gorski kraji	663,3	1.947
Drugi turistični kraji	434,1	895
Drugi kraji	33,6	68
Skupaj	**2.681,2**	**8.261**

Vir: SURS, Statistični letopis 2008, http://www.stat.si/letopis/2008/25 (12. 2. 2009)

4.1.2 Frekvenčne porazdelitve za številske spremenljivke

Za številske spremenljivke, ki imajo najpogosteje veliko število vrednosti, določimo skupine, ki jih imenujemo razredi in to tako, da vsako enoto lahko uvrstimo le v en razred. Razrede razmejimo tako, da določimo za opazovano številsko spremenljivko Y spodnjo mejo razreda $y_{j,\min}$ ter zgornjo mejo razreda $y_{j,\max}$. Pri določanju mej razredov moramo upoštevati dejstvo, da so številske spremenljivke lahko zvezne ali diskretne ter da so njihove vrednosti lahko zaokrožene ali pa niso zaokrožene.

Bodimo pozorni pri določevanju mej razredov zvezne številske spremenljivke, kjer vrednosti v posamezne razrede razvrščamo glede na cele vrednosti. Ugotovite, kakšna je razlika med opisoma " ***od*** *(vključno)* ***… do pod …****" ter "****nad … do*** *(vključno)* ***…*** *"?*

Tabela 25: Spodnje in zgornje meje, širine in sredine razreda za zvezno spremenljivko

Starostne skupine	**Širina razreda** d_j	**Spodnja meja** $y_{j,\min}$	**Zgornja meja** $y_{j,\max}$	**Sredina razreda** y_j
do pod 10	-	-	10	-
od 10 do pod 20	10	10	20	15
od 20 do pod 30	10	20	30	25
od 30 do pod 40	10	30	40	35
od 40 do pod 50	10	40	50	45
od 50 do pod 60	10	50	60	55
od 60 do pod 70	10	60	70	65
od 70 do pod 80	10	70	80	75
od 80 naprej	-	80	-	-

Vir: Lasten

V zgornji tabeli prvi razred nima spodnje meje, zadnji razred pa nima zgornje meje; takšnim razredom pravimo odprti razredi ter jih opredelimo, kadar so najmanjše oz. največje vrednosti zelo razpršene. Zaradi odprtih razredov pogosto tudi ne moremo izračunati vseh parametrov. Meje razredov so tukaj podane zvezno, saj velja: $y_{j,\min} = y_{j-1,\max}$

Pri diskretnih spremenljivkah zveznost zagotovimo s popravkom za zveznost:

$$\Delta y_j = \frac{y_{j,\min} - y_{j-1,\max}}{2} \qquad \text{za} \quad y = 2, 3, 4 \ldots r$$

ki ga k zgornjim mejam prištevamo, od spodnjih pa odštevamo.

Tabela 26: Določitev obeh mej, širine in sredine razreda za diskretno spremenljivko

Število članov v gospodinjstvu	d_j	$y_{j,\min}$	$y_{j,\max}$	y_j
od 1 do 2	2	0,5	2,5	1,5
od 3 do 4	2	2,5	4,5	3,5
od 5 do 6	2	4,5	6,5	5,5

Vir: Lasten

4.2 SESTAVLJANJE FREKVENČNE PORAZDELITVE

Vemo, da frekvenčno porazdelitev kreiramo z namenom večje preglednosti razporeditve enot, kar olajša ugotavljanje določenih značilnosti pojava, saj se te vidneje pokažejo pri urejenih podatkih.

Koraki sestavljanja frekvenčne porazdelitve za populacijo s številom enot N so naslednji:

- poiščemo $y_{\min}$ in $y_{\max}$
- v razmiku od $y_{\min}$ *do* $y_{\max}$ določimo primerno število razredov r; to je odvisno od velikosti populacije, ni pa pravila za njegovo izbiro; za srednje velike populacije je število razredov najpogosteje med 8 in 16; kadar nimamo omejitev, lahko uporabimo Sturgesovo pravilo (Korenjak, 2008):

  $r = 1 + \log_2 N \cong 1 + 3{,}32 \cdot \log_{10} N$
- določimo širino razredov,
- določimo meje razredov,
- enote razvrstimo v razrede; če jih preštejemo, dobimo frekvence razredov.

Tabela 27: Frekvenčna porazdelitev uspeha dijakov na poklicni maturi 2007

Točke	**Štev. dijakov**	$y_{j,\min}$	$y_{j,\max}$	y_j	d_j
8–9	270	7,5	9,5	8,5	2
10–11	1.046	9,5	11,5	10,5	2
12–13	1.716	11,5	13,5	12,5	2
14–15	1.643	13,5	15,5	14,5	2
16–17	1.364	15,5	17,5	16,5	2
18–19	863	17,5	19,5	18,5	2
20–21	434	19,5	21,5	20,5	2
22–23	117	21,5	23,5	22,5	2

Vir: RIC, Državni izpitni center, Letno maturitetno poročilo o poklicni maturi 2007, http://www.ric.si/ (12. 2. 2009)

Za nazoren prikaz frekvenčne porazdelitve moramo izbrati ustrezno število razredov, ki je odvisno od velikosti populacije. Pri določitvi se ravnamo po smiselnosti glede na predstavljeno vsebino. Za srednje velike populacije je število razredov navadno med 8 in 16. Razmislite, kaj pomeni za natančnost (in na drugi strani za preglednost porazdelitve) povečanje ali zmanjšanje izbranega števila razredov.

4.3 OPIS FREKVENČNE PORAZDELITVE

Pri analizi frekvenčne porazdelitve uporabljamo naslednje pojme (Šadl, 2004):

- **frekvenca** f_j je število enot v posameznem razredu; seštevek frekvenc v vseh razredih je enak številu enot N v opazovani populaciji:

  $$\sum_{j=1}^{r} f_j = N$$

- **relativna frekvenca** $f_j^0 = \frac{f_j}{N}$ izraža delež enot v posameznem razredu; z njo odpravimo vpliv velikosti populacije, zanjo velja:

$$\sum_{j=1}^{r} f_j^0 = 1$$

- **kumulativa frekvenc** F_j predstavlja vsoto frekvenc do določenega razreda, dobimo jo s postopnim seštevanjem frekvenc, in sicer:
  - v prvem razredu je enaka njegovi frekvenci $F_1 = f_1$
  - v naslednjih razredih pa velja: $F_j = F_{j-1} + f_j$

  za dani razred njena izračunana vrednost določa število enot z vrednostjo, ki je manjša ali enaka zgornji meji tega razreda;

- **kumulativa relativnih frekvenc** F_j^0 določa delež enot z vrednostjo, ki je manjša ali enaka zgornji meji tega razreda; njen izračun je podoben prejšnji kumulativi frekvenc:

$$F_j^0 = F_{j-1}^0 + f_j^0$$

Tabela 28: Frekvenčna porazdelitev uspeha dijakov na poklicni maturi 2007 z izračunanimi relativnimi frekvencami ter kumulativo absolutnih in relativnih frekvenc

Točke	**Št. dijakov** f_j	**Delež dijakov** f_j^0	**Kumulativa frekvenc** F_j	**Kumulat. relativnih frekvenc** F_j^0
8–9	270	0,0362	270	0,0362
10–11	1.046	0,1403	1.316	0,1766
12–13	1.716	0,2302	3.032	0,4068
14–15	1.643	0,2204	4.675	0,6273
16–17	1.364	0,1830	6.039	0,8103
18–19	863	0,1158	6.902	0,9261
20–21	434	0,0582	7.336	0,9843
22–23	117	0,0157	7.453	1,0000
Skupaj	**7.453**	**1,0000**		

Vir: Tabela 27

Interpretacija predstavljenih kazalcev npr. v tretjem razredu ($j = 3$):

- $f_3 = 1.716$, kar pomeni, da je 1.716 dijakov opravilo poklicno maturo z oceno 12 ali 13 točk;

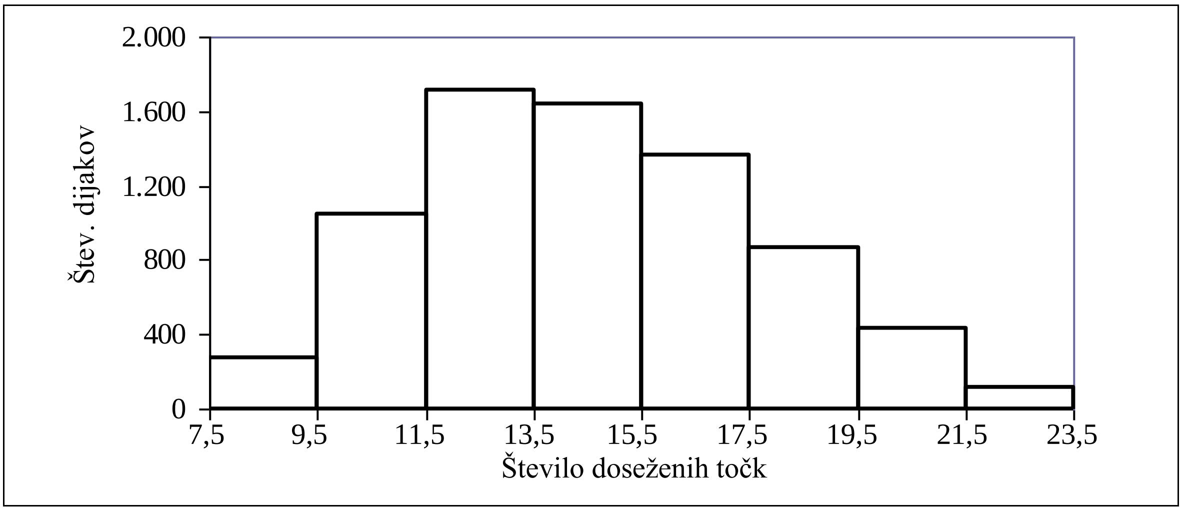
- $f_3^0 = 0,2302$, to pomeni, da je 23,02 % dijakov opravilo poklicno maturo z oceno 12 ali 13 točk;

- $F_3 = 3.032$, kar pomeni, da je 3.032 dijakov opravilo poklicno maturo z oceno največ do 13 točk;

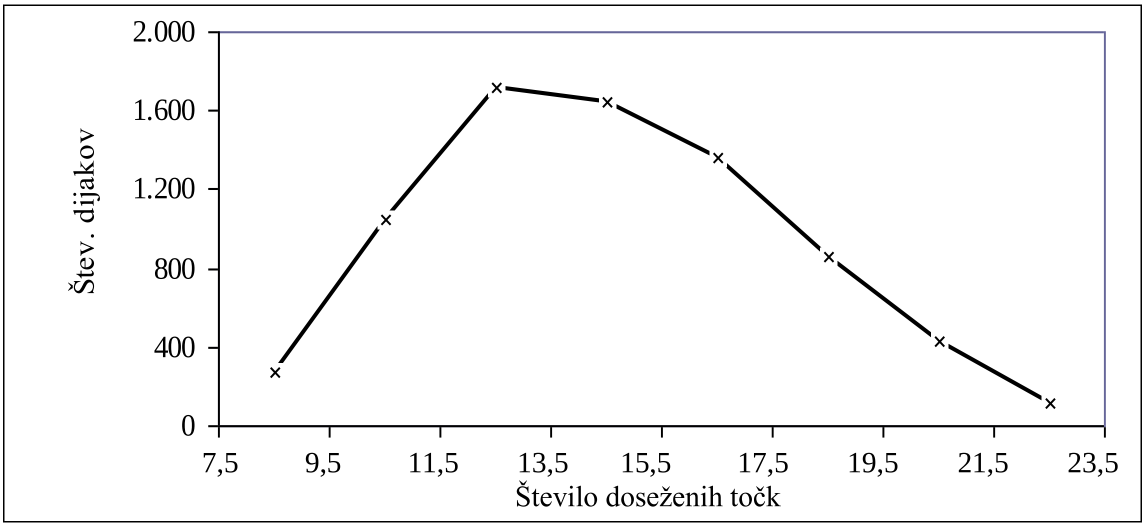
- $F_3^0 = 0,4068$, kar pomeni, da je 40,68 % dijakov opravilo poklicno maturo z oceno največ do 13 točk.

4.4 GRAFIČNO PRIKAZOVANJE FREKVENČNIH PORAZDELITEV

Frekvenčne porazdelitve grafično prikazujemo s stolpčnim grafikonom (histogramom) ali z linijskim grafikonom (poligonom). S slednjim prikazujemo tudi kumulative frekvenc – ogiva.



Slika 20: Histogram frekvenčne porazdelitve uspeha dijakov na poklicni maturi 2007
Vir: Tabela 27



Slika 21: Poligon frekvenčne porazdelitve uspeha dijakov na poklicni maturi 2007
Vir: Tabela 27

4.5 FREKVENČNE PORAZDELITVE Z NEENAKO ŠIROKIMI RAZREDI

Kadar je razmik med najmanjšo in največjo vrednostjo spremenljivke velik ter ima proučevana populacija veliko število enot, oblikujemo frekvenčne porazdelitve z neenako širokimi razredi. Zaradi različnih širin razredov izračunavamo **gostoto frekvenc** g_j ter **gostoto relativnih frekvenc** φ_j, ki kažeta stopnjo zgostitve:

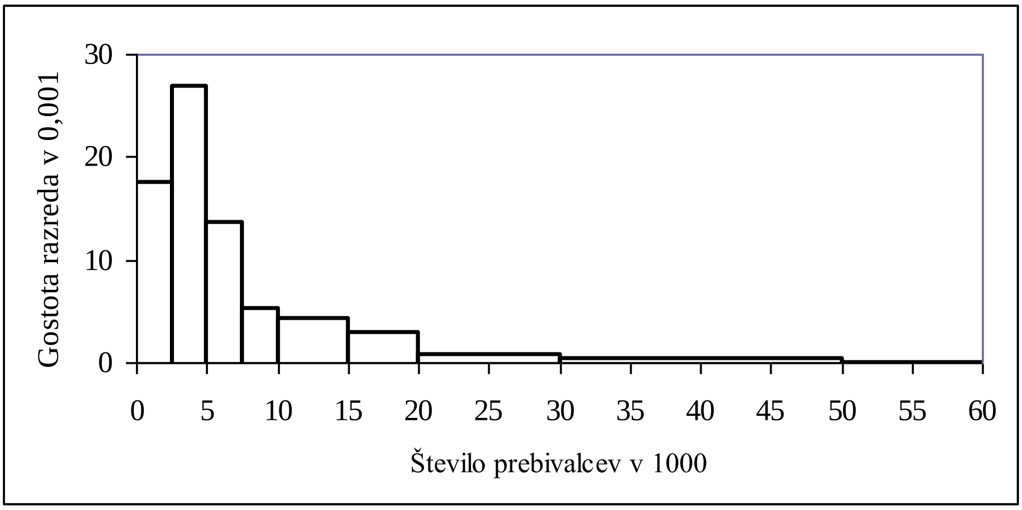
$$g_j = \frac{f_j}{d_j} \qquad \varphi_j = \frac{f_j^0}{d_j}$$

Tabela 29: Občine po številu prebivalcev v R Sloveniji 1. 1. 2008 z izračunanimi gostotami razredov

Število prebivalcev	**Število občin** f_j	**Širina razreda** d_j	**Gostota razreda** g_j v 10^{-3}
do 2.500	44	2.500	17,6
2.501–5.000	66	2.500	26,4
5.001–7.500	34	2.500	13,6
7.501–10.000	13	2.500	5,2
10.001–15.000	21	5.000	4,2
15.001–20.000	15	5.000	3,0
20.001–30.000	8	10.000	0,8
30.001–50.000	6	20.000	0,3
50.001–120.000	2	70.000	0,029
120.001–280.000	1	160.000	0,006

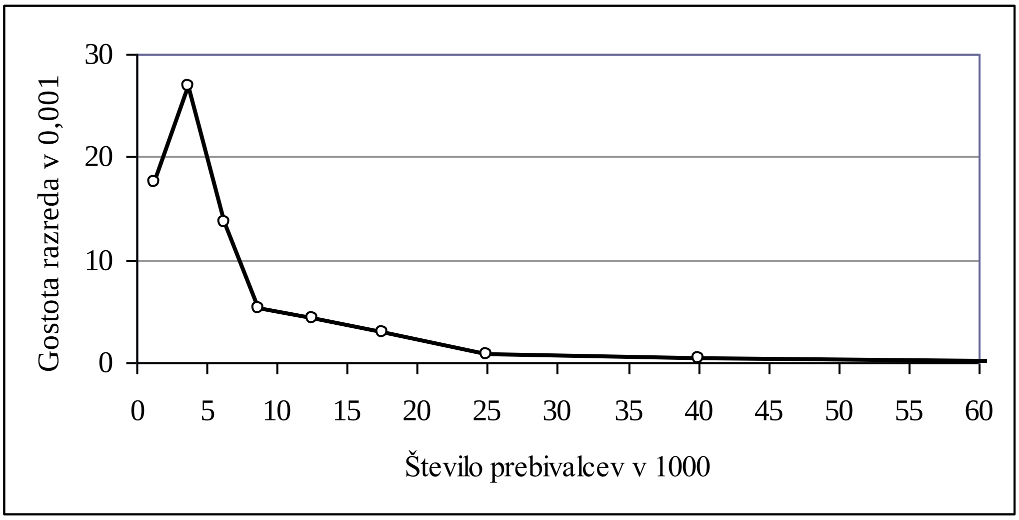
Vir: SURS, Statistični letopis 2008, http://www.stat.si/letopis/2008/02 (12. 2. 2009) ter http://www.stat.si/letopis/2008/31 (12. 2. 2009)

Frekvenčno porazdelitev z neenakomerno širokimi razredi grafično predstavljamo s histogramom ter s poligonom. Na abscisno os nanašamo meje razredov, na ordinatno pa vrednosti gostote frekvenc.



Slika 22: Histogram frekvenčne porazdelitve števila prebivalcev za občine v R Sloveniji
Vir: Tabela 29

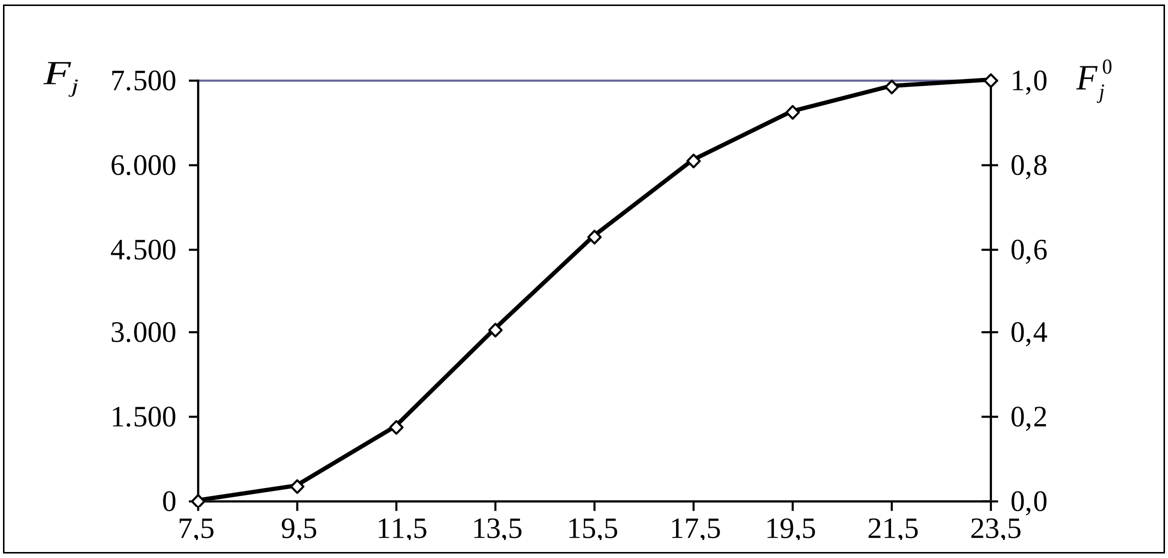
Pojasnite razlog, zakaj so v naslednji predstavitvi frekvenčne porazdelitve s poligonom vrisane točke vnesenih vrednosti za gostote frekvenc na sredini razredov, točke pa so povezane z lomljeno črto oziroma z daljicami.



Slika 23: Poligon frekvenčne porazdelitve števila prebivalcev za občine v R Sloveniji
Vir: Tabela 29

4.6 GRAFIČNI PRIKAZ KUMULATIVE FREKVENC

Kumulative frekvenc najpogosteje prikazujemo z linijskim grafom. Njihovo vrednost, odčitano na ordinatni osi, vnašamo kot točke nad zgornje meje razredov. Da lahko pregledno odčitamo vrednosti kumulative relativnih frekvenc, je skala zanjo izpisana na sekundarni ordinatni osi.



Slika 24: Ogiva – kumulativa frekvenčne porazdelitve uspeha dijakov na poklicni maturi 2007
Vir: Tabela 28

Ogiva je prikaz kumulativ frekvenc z linijskim grafikonom. V grafu kumulative frekvenc nanašamo nad zgornje meje razredov.

S pomočjo grafičnega prikaza kumulativ frekvenc ocenite število kandidatov in njihov %, ki so opravili poklicno maturo z oceno največ do 17 točk.

Odgovor: *Iz zgornje slike lahko odčitamo vrednost 6.000 kandidatov, kar je dobro primerljivo z izračunano vrednostjo v tabeli* 28, *ki znaša 6.039 kandidatov.*

Množico zbranih podatkov je potrebno pregledno predstaviti, zato le-te pri manjšem številu razvrščamo v ranžirne vrste, sicer pa v frekvenčne porazdelitve. Frekvenčna porazdelitev je statistična vrsta, kjer so sorodne vrednosti uvrščene v skupine (razrede), posamezni skupini oz. razredu pa prirejena ustrezna reprezentativna vrednost, ki je nova vrednost spremenljivke, kot navaja Ferligoj (1995). Oblikovanje skupin (oz. razredov pri številčnih podatkih) je potrebno izpeljati skrbno; določene morajo biti enolično, tako da je vsako enoto možno uvrstiti le v eno samo skupino oz. razred. Na obravnavanih primerih smo opredelili skupine in razrede, zanje določili primerno število in širino razredov ter podatke razvrstili. Naučili smo se izračunati frekvenco, relativno frekvenco, kumulativo frekvence ter kumulativo relativne frekvence, s katerimi je frekvenčna porazdelitev opisana. Njihove značilne porazdelitve smo ob podpori Excela tudi grafično prikazali. Ugotovljene lastnosti frekvenčnih porazdelitev pa bomo tudi koristno uporabili pri nadgradnji znanja v 6. poglavju.

Naloge

1. Razmislite in uporabite postopek, s katerim lahko izbrani podatkovni niz v Excelu najpreprosteje preoblikujemo v padajočo ali naraščajočo ranžirno vrsto.

2. Tabela 1 nam predstavlja podatke popisa prebivalcev iz leta 2002 po starostnih skupinah. Za skupno število prebivalcev tabelo dopolnite z izračunanimi relativnimi frekvencami in kumulativo relativnih frekvenc, razložite vse kazalce v četrtem razredu ter frekvenčno porazdelitev grafično prikažite s poligonom.

3. Študente so anketirali o mesečni višini nagrade, prejete za opravljanje prakse. Podatki, izraženi v frekvencah, so naslednji:

Znesek štipendije v €	**Število študentov**
nad 50 do 75	2
nad 75 do 100	4
nad 100 do 125	5
nad 125 do 150	7
nad 150 do 175	11
nad 175 do 200	19
nad 200 do 225	13
nad 225 do 250	8
nad 250 do 275	7
nad 275 do 300	4

Vir: Lasten

Tabelo dopolnite z izračunanimi relativnimi frekvencami in kumulativo relativnih frekvenc ter jih prikažite grafično. Grafično ocenite število študentov, ki so prejeli nagrado od 220 € do 240 €.

5 SREDNJE VREDNOSTI

V dosedanjih korakih spoznavanja statističnega preučevanja smo si prizadevali, da bi množico neurejenih podatkov naredili bolj pregledno. V prejšnjem poglavju smo jo v ta namen pri frekvenčnih porazdelitvah uredili v določeno število razredov, ki jih predstavljajo sredine razredov. Želja po poenostavljanju množice podatkov gre pogosto še dlje in to z namenom iskanja take značilne vrednosti sredin, ki bi lahko dovolj reprezentativno predstavljale celoten niz podatkov. Sredin, ki jih je mogoče izbrati kot tipičnega predstavnika populacije, je v statistiki več. Tako bomo v tem poglavju spoznali vrednosti, ki jih imenujemo s skupnim imenom mere srednjih vrednosti: aritmetično sredino, mediano, modus, geometrijsko sredino ter harmonično sredino. Proučili bomo njihove lastnosti ter razmislili o ustrezni izbiri srednje vrednosti pri analizi vsebine obravnavanih podatkov tako, da bo izbrani parameter najbolje predstavil značilnosti preučevanega pojava.

<u>Uvodna naloga</u>

V restavraciji dva natakarja strežeta gostom. Spremljava števila postreženih gostov v treh urah kaže naslednje število postreženih oseb:

Tabela 30: Število postreženih gostov

Obdobje	**Natakar A**	**Natakar B**
1. ura	11	9
2. ura	13	8
3. ura	15	22

Vir: Lasten

Oba natakarja sta v opazovanem obdobju postregla enako število gostov, tudi aritmetični sredini navedenih podatkov sta enaki, vseeno pa podatki kažejo na njuno neenakomerno obremenjenost. Poiščimo tako srednjo vrednost, da bomo z njo lahko ugotovili, ali je bil pri navedeni strežbi kdo bolj zaposlen oziroma je imel bolj dinamično delo (Knežević, 2006).

5.1 POJEM IN VRSTE SREDNJIH VREDNOSTI

Pri spoznavanju frekvenčnih porazdelitev smo videli, da so vrednosti spremenljivke pri posameznih opazovanih enotah raznolike ter variirajo od enote do enote. Opazili pa smo tudi, da so nekatere vrednosti spremenljivke pogostejše, da se pojavlja gostitev enot ter se vrednosti spremenljivke le nekoliko razlikujejo od kake osrednje vrednosti. Odpre se vprašanje, kdaj je neka srednja vrednost, nanašajoča se na osrednjo težnjo niza podatkov, lahko dober predstavnik vseh opazovanih enot. Srednja vrednost je kot mera centralne tendence tipična vrednost in je pogosto uporabljan statistični parameter.

Mere srednjih vrednosti, ki jih v nadaljevanju preučujemo, so:

- aritmetična sredina,
- mediana,
- modus,
- harmonična sredina in
- geometrijska sredina.

Vsaka od naštetih srednjih vrednosti ima konkreten pomen v praksi ter s svojimi posebnimi lastnostmi in na svoj način odkriva tisto, kar je tipično za opazovane enote preučevanega pojava.

5.2 ARITMETIČNA SREDINA

Aritmetična sredina M je najbolj znana in najpogosteje uporabljana srednja vrednost, imenovana tudi povprečje. Če vsoto vrednosti spremenljivke y označimo z Y, velja:

$$M = \frac{Y}{N} = \bar{y}$$

kjer je N enak številu vseh enot opazovane spremenljivke.

Računamo jo lahko samo za številčne spremenljivke. Pri njenem izračunu moramo upoštevati, ali izračunavamo (podobno velja tudi za ostale parametre) aritmetično sredino iz posameznih vrednosti ali iz podatkov, zbranih v frekvenčni porazdelitvi.

5.2.1 Izračun aritmetične sredine iz posameznih vrednosti

V tem preprostem primeru velja obrazec:

$$M = \frac{Y}{N} = \frac{1}{N}(y_1 + y_2 + \ldots + y_N) = \frac{1}{N}\sum_{i=1}^{N} y_i$$

V elektronski preglednici Excel lahko aritmetično srednjo vrednost preprosto izračunamo z uporabo statistične funkcije **AVERAGE**(number1;number2;...)

Izračunajmo povprečno oceno dosežkov 10-ih kandidatov pri opravljanju poklicne mature, katerih rezultati so bili predstavljeni v točki 2.3:

11	*12*	*13*	*14*	*17*	*18*	*19*	*20*	*21*	*22*

Odgovor: M = 16,7 točke

5.2.2 Izračun aritmetične sredine iz frekvenčne porazdelitve

Pri velikih populacijah so podatki navadno zbrani v frekvenčnih porazdelitvah, tu so vrednosti spremenljivke zajete v razredih; za vsak razred poznamo število enot oz. frekvenco, ne pa posamezne vrednosti – vemo le, da so njihove vrednosti med spodnjo mejo razreda $y_{j,\min}$ in zgornjo mejo razreda $y_{j,\max}$. V tem primeru upoštevamo sredino razreda y_j kot predstavnika vseh vrednosti, vključenih v razred. Vsoto vrednosti Y^* v tem primeru zato ocenimo, in sicer kot seštevek produktov sredine razredov in njim pripadajočih frekvenc, po obrazcu:

$$Y^* = \sum_{j=1}^{k} y_j f_j$$

Ocenjeno vsoto vrednosti Y^* delimo s številom enot in dobimo oceno aritmetične sredine M^*, torej:

$$M^* = \frac{Y^*}{N} = \frac{\sum_{j=1}^{k} y_j f_j}{N} = \frac{\sum_{j=1}^{k} y_j f_j}{\sum_{j=1}^{k} f_j}$$

kjer pomeni:

k	število razredov
y_j	sredino v j razredu
f_j	frekvenco v j razredu, utež

Aritmetično sredino, računano iz frekvenčne porazdelitve, imenujemo **tehtana (ponderirana) aritmetična sredina,** frekvence pa **teže (ponderji).**

Značilnost aritmetične sredine je – kot navaja Korenjak Černe (2009 a) – tudi dejstvo, da je vsota kvadriranih odklonov vrednosti spremenljivke y od vrednosti M najmanjša. To je zelo pomembna lastnost, saj je iz kvadratov odklonov od aritmetične sredine izpeljan parameter za merjenje variabilnosti podatkov – varianca, ki jo bomo spoznali v nadaljevanju.

Iz frekvenčne porazdelitve v tabeli 27 izračunajmo doseženo povprečno točkovno vrednost uspeha dijakov na poklicni maturi leta 2007.

Odgovor: M = 14,6 točke

5.3 MEDIANA

Centralna vrednost ali **mediana** *Me* je srednja vrednost, ki se v množici številčnih podatkov, urejenih po velikosti, nahaja natančno v sredini: polovica enot proučevane populacije ima od nje manjše vrednosti, polovica pa večje.

5.3.1 Izračun mediane iz posameznih vrednosti

Vrednosti številske spremenljivke razvrstimo po velikosti od najmanjše do največje vrednosti. Dobimo urejeno naraščajočo **ranžirno vrsto**, kjer dobi vsaka enota svoje zaporedno mesto oz. **absolutni rang** R. Ta rang ima vrednosti od 1 do N, če je v vrsti N opazovanih enot. Mediana je tista vrednost, ki je točno na sredini ranžirne vrste, njen rang izračunamo takole:

$$R_{Me} = \frac{N+1}{2}$$

Vrednost mediane je odvisna od števila opazovanih enot. Pri njenem izračunu sta dve možnosti:

a) pri lihem številu opazovanih enot je mediana enaka podatku na sredini vrste: $Me = y_R$

b) če je v ranžirni vrsti sodo število enot, izračunan rang po zgornjem obrazcu ni celo število; mediano izračunamo kot povprečje obeh podatkov na sredini vrste:

$$Me = \frac{y_{R-0,5} + y_{R+0,5}}{2}$$

Za podatke v točki 5.2.1 izračunajmo še vrednost mediane.

Tabela 31: Ranžirna vrsta

Točke	11	12	13	14	17	18	19	20	21	22
Rang	1	2	3	4	5	6	7	8	9	10

Vir: Lasten

$$R_{Me} = \frac{N+1}{2} = \frac{10+1}{2} = 5{,}5 \qquad Me = \frac{17+18}{2} = 17{,}5 \text{ točk}$$

Z uporabo statistične funkcije **MEDIAN**(number1;number2;...) lahko mediano izračunamo v aplikaciji Excel.

5.3.2 Izračun mediane iz frekvenčne porazdelitve

Kadar je število opazovanih enot veliko, so ponavadi vrednosti razvrščene v frekvenčni porazdelitvi. Posameznih vrednosti spremenljivke ne poznamo, zato bomo lahko dobili za mediano le približno vrednost oz. oceno. V postopku njenega izračuna ranžirno vrsto nadomestimo s kumulativo frekvenc.

Rang mediane izračunamo po prejšnjem obrazcu. Mediana se nahaja v razredu, ki je prvi z večjo kumulativo od ranga; ta razred imenujemo medialni razred.

$$F_{j-1} < R_{Me} < F_j$$

Oceno mediane izračunamo po prilagojenem obrazcu:

$$Me^* = y_{j,\min} + d_j \frac{R_{Me} - F_{j-1}}{f_j}$$

kjer pomenijo:

- $y_{j,\min}$ spodnjo mejo medialnega razreda,
- d_j širino medialnega razreda,
- f_j frekvenco medialnega razreda,
- F_{j-1} kumulativo frekvenc pred medialnim razredom.

Iz frekvenčne porazdelitve, podane v tabeli 28, izračunajmo mediano za uspeh dijakov na poklicni maturi leta 2007.

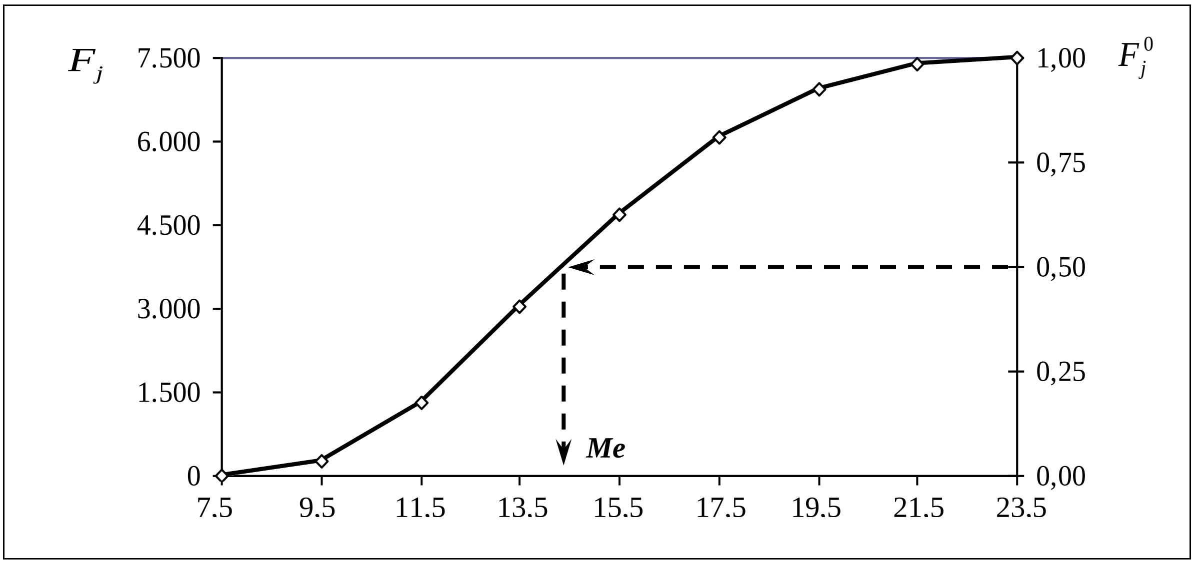
Odgovor: $R_{Me} = \frac{7454}{2} = 3727$; $\quad Me^* = 13{,}5 + 2 \cdot \frac{3727 - 3032}{1643} = \underline{14{,}3 \text{ točke}}$

Korenjak Černe (2009 a) navaja med prednosti mediane, da je preprosto razumljiva mera ter za določanje njene vrednosti ni potrebno poznati vseh vrednosti. To je posebej ugodno, če sta prvi in (ali) zadnji razred odprta: zadošča poznavanje vrednosti tistih enot, ki ležijo okoli sredine v ranžirni vrsti. Njena slabost pa je premajhna občutljivost za spremembe vrednosti, saj se vrednost mediane spremeni le, če spremembe vrednosti spremenljivk povzročijo premik le-teh iz ene polovice ranžirne vrste v drugo.

Presodite, ali je možno izračunati mediano za frekvenčno porazdelitev v primerih, kadar je medialni razred prvi razred. Razložite argumente za svojo odločitev.

5.3.3 Grafično določanje mediane

Oceno za mediano lahko preprosto določimo iz grafičnega prikaza kumulativ frekvenc (iz ogive) tako, da odčitamo vrednost številske spremenljivke, ki ustreza kumulativi relativnih frekvenc $F_j^0 = 0{,}5$. V spodnji sliki vidimo, da se odčitana vrednost mediane (14,5 točke) dobro ujema z izračunano vrednostjo v primeru na prejšnji strani ($Me^* = 14{,}3$ točke).



Slika 25: Grafično določanje mediane iz kumulative frekvenčne porazdelitve uspeha dijakov na poklicni maturi 2007
Vir: Tabela 28

5.4 MODUS

Srednjo vrednost, okoli katere so vrednosti proučevane populacije najgostejše, imenujemo **modus** Mo ali najpogostejša vrednost. Je vrednost, ki dominira v neki množici ter ni odvisna od vrednosti enot, ampak izključno od njihove gostitve. Modus lahko računamo za opisne in številčne spremenljivke ter tudi tiste podatkovne porazdelitve, ki imajo odprt začetni ali končni razred. Posamezne porazdelitve lahko imajo dva modusa (so bimodalne) ali celo več (so polimodalne).

5.4.1 Izračun modusa iz posameznih vrednosti

Iz posameznih vrednosti lahko določimo modus tako, da pogledamo, katera vrednost se največkrat pojavlja med opazovanimi vrednostmi (npr. najpogostejša ocena pri pisnem ocenjevanju določenega predmeta).

V aplikaciji Excel lahko v množici podatkov modus poiščemo z uporabo statistične funkcije **MODE**(number1;number2;...).

Izračunajmo modus naslednjih dosežkov 10-ih kandidatov pri opravljanju poklicne mature:

16	*12*	*13*	*16*	*17*	*18*	*19*	*20*	*16*	*21*

Odgovor: Mo = 16 točk

5.4.2 Izračun modusa iz frekvenčne porazdelitve

Za obsežne populacije, ki so grupirane v frekvenčni porazdelitvi, pri izračunu modusa izhajamo iz modusnega (modalnega) razreda, ki ima največ enot, torej $f_{\max}$. Približno vrednost za modus Mo^* izračunamo z linearno interpolacijo po obrazcu:

$$Mo^* = y_{j,\min} + d_j \frac{f_j - f_{j-1}}{2f_j - f_{j-1} - f_{j+1}}$$

kjer so:

$y_{j,\min}$ spodnja meja modusnega razreda,

d_j širina razreda,

f_j frekvenca modusnega razreda,

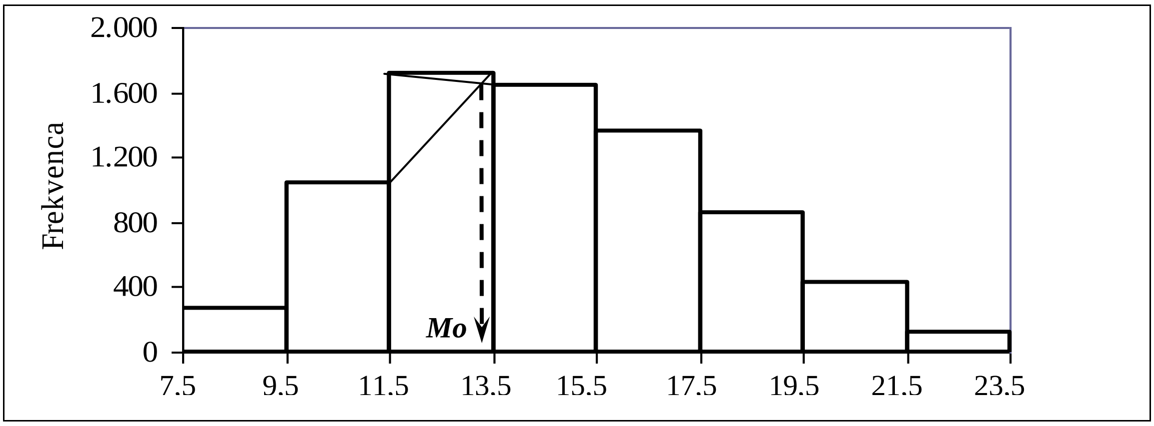
f_{j-1} frekvenca predmodusnega razreda,

f_{j+1} frekvenca pomodusnega razreda.

Iz frekvenčne porazdelitve, podane v tabeli 28, izračunajmo vrednost modusa za uspeh dijakov na poklicni maturi leta 2007.

Odgovor: $Mo^* = 11{,}5 + 2 \cdot \frac{1716 - 1046}{2 \cdot 1716 - 1046 - 1643} = \underline{13{,}3 \text{ točke}}$

Modus lahko tudi odčitamo z grafa frekvenčne porazdelitve v histogramu kot abciso iz točk preseka dveh diagonal, prve med spodnjo mejo modusnega in pomodusnega razreda ter druge med zgornjo mejo predmodusnega in modusnega razreda. V spodnji sliki odčitana vrednost modusa se tudi dobro ujema z zgoraj izračunano vrednostjo ($Mo^* = 13{,}3$ točke).



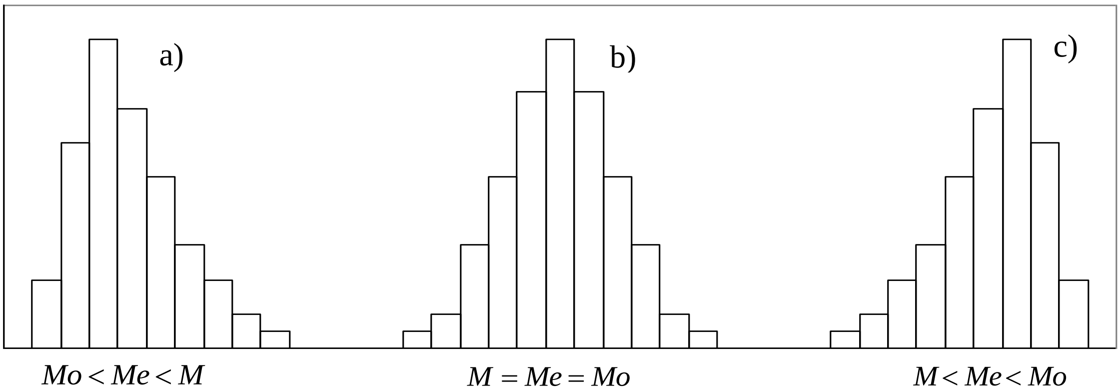
Slika 26: Grafično določanje modusa iz frekvenčne porazdelitve uspeha dijakov na poklicni maturi 2007
Vir: Tabela 27

Razmislite in presodite, če je možno izračunati modus za frekvenčno porazdelitev v primerih, kadar je modusni razred prvi ali zadnji razred. Argumentirajte svojo odločitev.

Med prednosti modusa Korenjak Černe (2009 a) navaja dejstvo, da ga lahko določamo iz osnovnih podatkov preprosto s preštevanjem ter nanj ne vplivajo spremembe vrednosti posameznih enot (vse dokler gostitev na nekem drugem mestu ne preseže stopnje gostitve v sprva ugotovljenem modusu). Ta premajhna občutljivost na spremembe posameznih vrednosti pa se v določenih primerih opazovanja pojavov pokaže tudi kot slabost tega parametra; prav tako pa je ugotavljanje modusa lahko vprašljivo pri bimodalnih in polimodalnih porazdelitvah ter pri zaokroževanju vrednosti podatkov proučevanega pojava**.**

5.5 ODNOSI MED ARITMETIČNO SREDINO, MEDIANO IN MODUSOM

Aritmetična sredina, mediana in modus imajo lahko različne ali enake vrednosti, odvisno od oblike frekvenčne porazdelitve. Odvisno od gostitve vrednosti so enomodusne (unimodalne) porazdelitve lahko simetrične ali asimetrične. Izkaže se, da so srednje vrednosti: aritmetična sredina, mediana in modus, enake za simetrične porazdelitve, ki imajo eno mesto zgostitve. Če pa je porazdelitev asimetrična, bodisi v desno ali v levo, navedene srednje vrednosti niso več medsebojno enake. Naslednja slika izpostavlja lastnosti, značilne za večino unimodalnih porazdelitev.



Slika 27: Aritmetična sredina, mediana in modus za: a) asimetrično porazdelitev v desno, b) simetrično ter za c) asimetrično porazdelitev v levo
Vir: Lasten

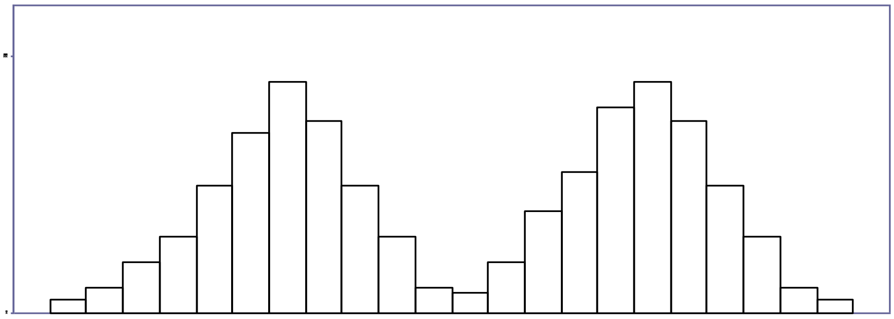
Pri unimodalni porazdelitvi, asimetrični v desno, približno velja relacija (Knežević, 2006):

$$Me - Mo = 2 \cdot (M - Me)$$

Pri unimodalni porazdelitvi, asimetrični v levo, približno velja relacija:

$$Mo - Me = 2 \cdot (Me - M)$$

Izkaže se, da je pri asimetričnih porazdelitvah aritmetična sredina manj primerna za predstavitev vseh opazovanih enot, kot sta to lahko mediana in modus, saj prva v teh primerih ne izraža dovolj značilnosti opazovanega pojava.



Slika 28: Bimodalna frekvenčna porazdelitev
Vir: Lasten

5.6 HARMONIČNA SREDINA

Harmonična sredina H je recipročna vrednost aritmetične sredine, izračunane iz recipročnih vrednosti opazovane spremenljivke (Trstenjak, 2001). Torej je enaka:

$$H = \frac{N}{\frac{1}{y_1} + \frac{1}{y_2} + \frac{1}{y_3} \ldots + \frac{1}{y_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{y_i}}$$

V statistični analizi uporabljamo harmonično sredino, ko analiziramo povprečja. Pogosto jo zato srečamo pri izračunavanju srednjih vrednosti relativnih števil, in sicer pri izračunavanju povprečnih strukturnih deležev statističnih koeficientov, strukturnih odstotkov (oz. odtisočkov) ter posameznih stopenj.

Harmonično sredino lahko v Excelu izračunamo z uporabo statistične funkcije **HARMEAN**(number1;number2;...).

V primerih, ko se posamezne vrednosti pojavljajo z različno pogostnostjo, uporabljamo **tehtano harmonično sredino**, ki je enaka:

$$H = \frac{f_1 + f_2 + f_3 + \ldots + f_k}{\frac{f_1}{y_1} + \frac{f_2}{y_2} + \frac{f_3}{y_3} + \ldots + \frac{f_k}{y_k}} = \frac{\sum_{j=1}^{k} f_j}{\sum_{j=1}^{k} \frac{f_j}{y_j}}$$

Z našim znanjem lahko sedaj tudi dokončamo uvodno nalogo:

<u>Rešitev uvodne naloge</u>

Za število postreženih gostov v opazovanem obdobju izračunajmo vrednosti obeh harmoničnih sredin:

Tabela 32: Harmonični sredini za število postreženih gostov

Obdobje	**Natakar A**	**Natakar B**
1. ura	11	9
2. ura	13	8
3. ura	15	22
H	12,8	10,7

Vir: Tabela 30

Analiza podatkov s pomočjo harmonične sredine pokaže, da je bil pri strežbi gostov prvi natakar bolj zaposlen, saj velja relacija: $H_A > H_B$, kljub temu, da je za oba niza enako skupno število gostov ($N_A = N_B = 39$) ter ugotovljeni isti aritmetični srednji vrednosti ($M_A = M_B = 13$). Preračunano na uro iz znane harmonične sredine tako velja, da je za postrežbo gosta natakar A porabil 4,7 minute, natakar B pa 5,6 minute.

5.7 GEOMETRIJSKA SREDINA

Geometrijska sredina G je za N pozitivnih vrednosti spremenljivke Y enaka N-temu korenu iz produkta teh številskih vrednosti:

$$G = \sqrt[N]{y_1 \cdot y_2 \cdot y_3 \cdot \ldots \cdot y_N} = \sqrt[N]{\prod_{i=1}^{N} y_i}$$

Če izraz logaritmiramo, dobimo povezavo z aritmetično sredino:

$$\log G = \frac{1}{N}(\log y_1 + \log y_2 + \log y_3 + \ldots + \log y_N) = \frac{1}{N}\sum_{i=1}^{N} \log y_i$$

V Excelu lahko za obseg pozitivnih podatkov geometrijsko sredino izračunamo z uporabo statistične funkcije **GEOMEAN**(number1;number2;...).

Tudi geometrijsko sredino uporabljamo pri izračunavanju srednjih vrednosti iz relativnih števil, predvsem pri analizi časovnih vrst ter za izračune povprečne stopnje rasti, povprečnega koeficienta rasti in povprečnega verižnega indeksa, kar bomo spoznali v naslednjem poglavju.

5.8 IZRAČUNAVANJE POVPREČIJ IZ RELATIVNIH ŠTEVIL

Za opazovani pojav imamo zbrane vrednosti relativnih števil za nekaj zaporednih let, npr.:

Leto	**2002**	**2003**	**2004**	**2005**	**2006**
Pojav	Y_0	Y_1	Y_2	Y_3	Y_4
Koeficient rasti		K_1	K_2	K_3	K_4
Verižni indeks		V_1	V_2	V_3	V_4
Stopnja rasti		S_1	S_2	S_3	S_4

Poznamo že obrazce za izračun kazalcev rasti:

$$K_j = \frac{Y_j}{Y_{j-1}} \qquad V_j = 100 \cdot \frac{Y_j}{Y_{j-1}} \qquad S_j = V_j - 100 = 100 \cdot (K_j - 1)$$

Sedaj razmislimo in oblikujmo obrazce izračuna povprečnih vrednosti kazalcev za obdobje več let vnaprej.

5.8.1 Računanje povprečnega koeficienta rasti

Podatek Y_N za pojav v obdobju N izrazimo z začetnim stanjem Y_0, pomnoženim s koeficientom rasti K_j od obdobja do obdobja:

$$Y_N = Y_0 \cdot K_1 \cdot K_2 \cdot \ldots \cdot K_N$$

Enako stalno relativno spremembo iz obdobja v obdobje kaže **povprečni koeficient rasti** $\overline{K}$, torej velja:

$Y_N = Y_0 \cdot \overline{K}^N$ iz česar sledi:

$$\bar{K} = \sqrt[N]{\frac{Y_N}{Y_0}}$$

Povprečni koeficient rasti lahko z geometrijsko sredino izračunamo tudi iz koeficientov rasti za posamezna obdobja:

$$\bar{K} = \sqrt[N]{K_1 \cdot K_2 \cdot \ldots \cdot K_N}$$

Izračunajmo povprečni koeficient rasti za podatke o vpisanih študentih v slovenske višje šole v petletnem obdobju, predstavljenem v naslednji tabeli. Preverite, kako različna osnova v indeksni vrsti vpliva na vrednost povprečnega koeficienta.

Tabela 33: Število študentov, vpisanih v višje šole v študijskih letih 2002/03 do 2006/07, koeficienti rasti in indeksi s stalno osnovo (2002 = 100 in 2004 = 100)

Leto	**Št. študentov**	K_j	$I_{j/2002}$	$I_{j/2004}$
2002/03	8.796	-	100,0	69,7
2003/04	11.099	1,262	126,2	87,9
2004/05	12.621	1,137	143,5	100,0
2005/06	14.246	1,129	162,0	112,9
2006/07	15.831	1,111	180,0	125,4

Vir: SURS, Letopis 2007, http://www.stat.si/letopis/2007/06 (12. 2. 2009)

Povprečni koeficient rasti bomo izračunali po treh različnih poteh. V praksi se za način, po katerem izračunavamo navedeni kazalec, odločimo na osnovi razpoložljivih podatkov.

- Izračun povprečnega koeficienta rasti iz osnovnih podatkov:

$\bar{K} = \sqrt[N]{\frac{Y_N}{Y_0}} = \sqrt[4]{\frac{15831}{8796}} = 1{,}158$ iz česar sledi povprečna stopnja rasti:

$$\bar{S} = (\bar{K} - 1) \cdot 100 = 15{,}8\ \%$$

- Izračun povprečnega koeficienta rasti iz koeficientov rasti:

$\bar{K} = \sqrt[4]{K_1 \cdot K_2 \cdot K_3 \cdot K_4} = \sqrt[4]{1{,}262 \cdot 1{,}137 \cdot 1{,}129 \cdot 1{,}111} = 1{,}158$ ter

$$\bar{S} = (\bar{K} - 1) \cdot 100 = 15{,}8\ \%$$

– Izračun povprečnega koeficienta rasti iz indeksov s stalno osnovo:

$\bar{K} = \sqrt[N]{\frac{Y_{N/2002}}{Y_{1/2002}}} = \sqrt[4]{\frac{180{,}0}{100{,}0}} = 1{,}158$, kar pomeni, da je v opazovanem obdobju število študentov naraščalo povprečno letno za 15,8 %.

Enako lahko izračunamo še iz druge indeksne vrste (2004/05 = 100):

$$\bar{K} = \sqrt[N]{\frac{Y_{N/2004}}{Y_{1/2004}}} = \sqrt[4]{\frac{125{,}0}{69{,}7}} = 1{,}158$$

iz česar izhaja povprečna stopnja rasti:

$\bar{S} = (\bar{K} - 1) \cdot 100 = 15{,}8\,\%$

V prejšnjem izračunu se je potrdila predpostavka, da različna osnova v indeksni vrsti ne vpliva na vrednost povprečnega koeficienta rasti.

5.8.2 Računanje povprečnega verižnega indeksa

Povprečni verižni indeks $\bar{V}$ je možno izračunati na tri načine:

- iz povprečnega koeficienta rasti:

$\bar{V} = 100 \cdot \bar{K}$

- z geometrijsko sredino iz verižnih indeksov:

$\bar{V} = \sqrt[N]{V_1 \cdot V_2 \cdot \ldots \cdot V_N}$ ali pa

- iz podatkov, nanašajočih se na začetno in končno stanje opazovanega pojava:

$$\bar{V} = 100 \cdot \sqrt[N]{\frac{Y_N}{Y_0}}$$

5.8.3 Računanje povprečne stopnje rasti

Povprečne stopnje rasti $\bar{S}$ ne moremo izračunati neposredno iz stopenj rasti za posamezna obdobja S_j, saj so te v posameznih obdobjih lahko enake nič ali so negativne, kar bi posledično prineslo za povprečno stopnjo nesmiseln rezultat.

Tabela 34: Stopnje rasti, verižni indeksi in koeficienti rasti za število prepeljanih potnikov z letali v R Sloveniji v letih od 2002 do 2007

Leto	**Štev. potnikov** (v 1000)	S_j	V_j	K_j
2002	814	-	-	-
2003	864	6,1	106,1	1,061
2004	885	2,4	102,4	1,024
2005	944	6,7	106,7	1,067
2006	1.018	7,8	107,8	1,078
2007	1.136	11,6	111,6	1,116

Vir: Tabela 23

Zaradi prejšnje ugotovitve računamo povprečno stopnjo rasti $\bar{S}$ posredno preko povprečnega verižnega indeksa $\bar{V}$ ali povprečnega koeficienta rasti $\bar{K}$, in sicer:

- stopnje rasti spremenimo v verižne indekse in iz njih najprej, po eni izmed prej nakazanih poti, izračunamo $\bar{V}$:

$\bar{V} = \sqrt[N]{V_1 \cdot V_2 \cdot \ldots \cdot V_N} = \sqrt[5]{106{,}1 \cdot 102{,}4 \cdot 106{,}7 \cdot 107{,}8 \cdot 111{,}6} = 106{,}9$ iz česar sledi:

$\bar{S} = \bar{V} - 100 = 6{,}9\ \%$ ali:

- stopnje rasti spremenimo v koeficiente rasti ter iz njih najprej izračunamo $\bar{K}$:

$\bar{K} = \sqrt[N]{K_1 \cdot K_2 \cdot \ldots \cdot K_N} = \sqrt[5]{1{,}061 \cdot 1{,}024 \cdot 1{,}067 \cdot 1{,}078 \cdot 1{,}116} = 1{,}069$ ter nato:

$\bar{S} = (\bar{K} - 1) \cdot 100 = 6{,}9\ \%$

Po obeh poteh pridemo do enakega rezultata, da je v obdobju od 2002 do 2007 število z letali prepeljanih potnikov povprečno letno naraščalo za 6,9 %.

5.8.4 Ocenjevanje pojava v prihodnosti

Razmislimo, ali lahko na osnovi povprečnega koeficienta rasti, izračunanega za določeno časovno obdobje, **ocenimo pojav v prihodnosti,** seveda ob predpostavki, da bodo tudi v prihodnosti enake razmere.

Na podlagi izračunanih relacij med kazalci ter števila prepeljanih potnikov z letali v R Sloveniji v letih od 2002 do 2007, navedenih v tabeli 34, ocenimo število potnikov, ki se bodo prepeljali z letali v letu 2010.

Izhodišče je število potnikov, prepeljanih z letali v letu 2007, torej $Y_0 = 1.136.000$ in število let $N = 3$ (2008, 2009, 2010).

Uporabimo relacijo:

$$Y_N = Y_0 \cdot \bar{K}^N$$

$$Y_{2010} = Y_{2007} \cdot 1{,}069^3 = 1.388.000 \text{ potnikov.}$$

Ocenjujemo, da bo leta 2010 z letali prepeljanih 1.388.000 potnikov, seveda ob predpostavki, da bodo razmere ostale enake.

Značilnosti proučevanega pojava pogosto ugotavljamo z izračunanimi srednjimi vrednostmi, saj te – vsaka na svoj način – kažejo gostitev oziroma centralno tendenco podatkov. V tem poglavju smo spoznali teorijo in postopke izračunavanja aritmetične sredine, mediane, modusa, geometrijske sredine ter harmonične sredine. Predstavljene značilnosti, njihov pomen in razlike med različnimi povprečji, nam v praksi pomagajo pri odločitvah, kdaj je – z ozirom na podatke in namen analize – primerno uporabiti katero srednjo vrednost. Izberemo tisto oz. računamo tiste mere, s katerimi najbolje opišemo značilnosti opazovanega pojava, kot navaja Šadl (2008). Tudi pri tem delu si lahko dodobra pomagamo z računalniško podporo Excelovih statističnih funkcij. Na dodanih primerih so nam izračunane vrednosti tudi pokazale, kako lahko iz primerjanja različnih rezultatov posameznih srednjih mer, izračunanih za isto porazdelitev podatkov, sklepamo in razložimo njene značilnosti. Poznavanje vsebin tega poglavja nam bo dobrodošlo tako v naslednjem 6. poglavju, ko bomo mediano povezali s kvantili, kot tudi pri ugotavljanju razpršenosti podatkov, s čimer se bomo ukvarjali v 7. poglavju.

Naloge:

1. Tabela 27 podaja frekvenčno porazdelitev doseženega točkovnega uspeha dijakov na poklicni maturi leta 2007. Dobljene vrednosti njegove aritmetične sredine, mediane in modusa medsebojno primerjajte ter določite vrsto in značilnosti navedene asimetrične porazdelitve.

2. Modus števila obiskov turistov iz Italije v izbranem zdravilišču je 4. Kaj to pomeni?

3. Za izbrani hotel so znani naslednji letni indeksi rasti cen, in sicer:

120	110	101	105	112	110	110

   Izračunajte povprečni letni indeks rasti cen v opazovanih sedmih časovnih obdobjih!

4. Prvotna cena opazovanega blaga je bila 100 €. Naprej se je blago podražilo za 14 %, nato za 7 % ter nazadnje za 5 %. Kolikšen je povprečen indeks podražitev in koliko znaša nova cena?

5. Tabela 1 vsebuje podatke popisa prebivalcev iz leta 2002 po starostnih skupinah. Izračunajte mediano in modus za skupno število prebivalcev ter ločeno za moške in ženske. V histogramu za skupno število prebivalcev določite obliko frekvenčne porazdelitve.

6. Izračunajte povprečno dobo bivanja gostov v R Sloveniji leta 2007, če so za posamezne vrste krajev podane vrednosti povprečne dobe bivanja in števila gostov. Potrebne uradne podatke za rešitev naloge si poiščite v Statističnem letopisu 2008.
   (Pomoč: Statistični letopis 2008, poglavje Turizem,
   http://www.stat.si/letopis/LetopisVsebina.aspx?poglavje=25&lang=si)

7. Izračunajte povprečno dobo bivanja gostov v R Sloveniji leta 2007, če so za posamezne vrste krajev podane vrednosti povprečne dobe bivanja in števila prenočitev. Potrebne uradne podatke za rešitev naloge si poiščite v Statističnem letopisu 2008.
   (Pomoč: Statistični letopis 2008, poglavje Turizem,
   http://www.stat.si/letopis/LetopisVsebina.aspx?poglavje=25&lang=si)

6 RANGI IN KVANTILI

V 4. poglavju smo preučevali frekvenčne porazdelitve ter spoznali pravila preglednega razvrščanja podatkov v ranžirne vrste in v frekvenčne porazdelitve. Eden od kazalcev, s katerim smo frekvenčno porazdelitev opisali, je bila tudi kumulativa frekvenc. Pri grafični predstavitvi kumulative frekvenc dane spremenljivke smo ugotovili, da lahko njeno vrednost v grafikonu ocenimo, sedaj pa bomo ugotovili postopek njenega izračuna. Lahko bomo rešili tudi obraten problem, ko na podlagi znanega podatka o deležu vrednosti, manjših oziroma večjih od dane vrednosti, izračunamo njen položaj, kot opredeljuje Šadl (2008). To omogočajo kvantili, ki jim namenjamo pozornost v tem poglavju. V dve izmed možnosti njihove praktične uporabe pa nas vpeljuje v razmišljanje pri naslednji nalogi.

<u>Uvodna naloga</u>

Pisni izpit je opravljalo 12 kandidatov, ki so od možnih 100 točk dosegli naslednji uspeh:

63	*88*	*52*	*27*	*38*	*91*	*42*	*96*	*34*	*48*	*71*	*84*

Izračunajmo:

A) Koliko % kandidatov je izpit opravilo, če je bilo potrebno za to doseči 45 točk?
B) Kolikšna bi morala biti točkovna meja za pozitivno oceno, da bi izpit opravilo 75 % kandidatov?

6.1 ABSOLUTNI IN KVANTILNI RANG

V prejšnjih poglavjih smo že spoznali **absolutni rang** R kot zaporedno številko, prirejeno vsaki enoti v naraščajoči ranžirni vrsti danih vrednosti opazovane spremenljivke. Ta absolutni rang R sicer določa položaj statistične enote v množici glede na vrednost spremenljivke, vendar je njegova slabost ta, da moramo ob njem vedno navajati še število enot – ni namreč vseeno, ali ima nekdo npr. rang 3 med 10 ali 100 športniki, kot navaja Leskošek (2008 b). Absolutni rang R lahko pri N enotah spremenljivke zavzame vsa cela števila od 1 do N in ima lastnost diskretne spremenljivke.

»**Kvantilni rang P** nam pove (v deležu), na katerem delu ranžirnega razmika leži določena enota, ki ji pripada rang R« (Devjak, 2008, 6). Določa, koliko odstotkov enot ima manjše in koliko večje vrednosti od izbrane enote. Ima vse vrednosti na razmiku od 0 do 1 in lastnost zvezne spremenljivke (Šadl, 2008).

Povezava med absolutnim in relativnim rangom je določena z obrazcem:

$$P = \frac{R - 0,5}{N} \quad \text{oziroma}$$

$$R = P \cdot N + 0,5$$

Popravek 0,5 se uporablja, ker je rang R diskretna veličina.

6.2 KVANTILI IN NJIHOVA UPORABA

Kvantili so vrednosti številske spremenljivke y, urejene v ranžirno vrsto, ki pripadajo določenemu kvantilnemu rangu in ranžirno vrsto razdelijo na k enakih delov (Knežević, 2006). Ti statistični izračuni nam pomagajo določiti, kje se posamezna enota nahaja v primerjavi z drugimi enotami in njenimi vrednostmi.

Najpogosteje se uporabljajo naslednji kvantili:

- **Mediana:** $Me(P = 0{,}50)$. Poleg že znanih lastnosti mediane, ugotovljenih v prejšnjem poglavju, lahko mediano iz zornega kota kvantilov opišemo še z naslednjimi dejstvi:
  - $k = 2$: mediana je kvantil, ki razdeli ranžirno vrsto točno na dva dela,
  - pripada kvantilnemu rangu $P = 0{,}50$,
  - je vrednost, od katere ima 1/2 enot manjšo, 1/2 pa večjo vrednost.

- **Kvartili:** $Q_1(P = 0{,}25)$, $Q_2(P = 0{,}50)$, $Q_3(P = 0{,}75)$:
  - $k = 4$: kvartili so kvantili, ki razdelijo ranžirno vrsto na 4 enake dele,
  - npr. 3. kvartil je vrednost, ki pripada kvantilnemu rangu $P = 0{,}75$,
  - npr. 3. kvartil je vrednost, od katere ima 75 % enot manjšo, 25 % pa večjo vrednost.

- **Decili:** $D_1(P = 0{,}10)$, $D_2(P = 0{,}20)$, ..., $D_9(P = 0{,}90)$:
  - $k = 10$: decili so kvantili, ki razdelijo ranžirno vrsto na desetine,
  - npr. 4. decil je vrednost, ki pripada kvantilnemu rangu $P = 0{,}40$,
  - npr. 4. decil je vrednost, od katere ima 40 % enot manjšo, 60 % enot pa večjo vrednost.

- **Centili:** $C_1(P = 0{,}01)$, $C_2(P = 0{,}02)$, ..., $C_{99}(P = 0{,}99)$:
  - $k = 100$: centili so kvantili, ki razdelijo ranžirno vrsto na stotine,
  - npr. 38. centil je vrednost, ki pripada kvantilnemu rangu $P = 0{,}38$,
  - npr. 38. centil je vrednost, od katere ima 38 % enot manjšo, 62 % enot pa večjo vrednost.

Med posameznimi kvantili tako veljajo naslednje relacije:

$$Me = Q_2 = D_5 = C_{50}$$
$$D_1 = C_{10}$$
$$Q_1 = C_{25}$$

V programu Excel lahko s statistično funkcijo **QUARTILE**(array;quart) izračunamo iz množice podatkov enega od štirih kvartilov.

Knežević (2006) navaja, da kvantile uporabljamo pri izračunu nekaterih statistično-analitičnih kazalcev razpršenosti in asimetrije. Nadalje pa so kvantili posebej uporabni pri določanju pozicije statistične enote v množici glede na vrednost spremenljivke ter pri obratnem problemu: določanju vrednosti spremenljivke, da bi se le-ta razporedila na opredeljeno mesto v množici. Omogočajo primerjavo vrednosti (znotraj populacije, dveh enot, ki pripadata različnim populacijam, iste enote v različnih časovnih obdobjih …).

Kazalce variabilnosti in asimetrije bomo spoznali v naslednjem poglavju. Drugo, zgoraj navedeno lastnost kvantilov, pa sedaj uporabimo pri reševanju uvodne naloge. Podobna primera sta s podrobnejšo razlago predstavljena v Šadl (2008) ter v Ferligoj (1995).

Rešitev uvodne naloge

V uvodu tega poglavja je izpostavljena naloga s podatki, da je 12 kandidatov opravljalo izpit ter od možnih 100 točk doseglo naslednji uspeh:

63 88 52 27 38 91 42 96 34 48 71 84

Iščemo odgovora na vprašanji:

A) Koliko % kandidatov je izpit opravilo, če je bilo potrebno za to doseči 45 točk?
B) Kolikšna bi morala biti točkovna meja za pozitivno oceno, da bi izpit opravilo 75 % kandidatov?

Nalogo rešimo v dveh delih, za vsako vprašanje posebej.

A) V prvem vprašanju je znan kvantil y_P = 45 točk, naši podatki izpitnih rezultatov pa so posamični. Postopek reševanja naloge poteka v naslednjih korakih (Devjak, 2008):

- Najprej podatke uredimo v ranžirno vrsto:

Tabela 35: Ranžirna vrsta dosežkov na izpitu

y	27	34	38	42	48	52	63	71	84	88	91	96
R	1	2	3	4	5	6	7	8	9	10	11	12

Vir: Lasten

- Za podani $y_P = 45$ točk poiščemo sosednji vrednosti spremenljivke, tako da velja: $y_0 \le y_P \le y_1$, kar v našem primeru pomeni:
  $y_0 = 42 < y_P = 45 < y_1 = 48$ ter
  $R_0 = 4 < R_P < R_1 = 5$

- Rang R_P za vrednost y_P izračunamo z linearno interpolacijo (podrobneje je postopek opisan v http://www.zavod-irc.si/docs/Skriti_dokumenti/STATISTIKA_ekon_sadl.pdf, stran 62) po obrazcu:

$$R_P = R_0 + \frac{y_P - y_0}{y_1 - y_0} \cdot (R_1 - R_0) = 4{,}5$$

- Sedaj lahko izračunamo pripadajoči kvantilni rang P_P:

$$P_y = \frac{R_P - 0{,}5}{N} = \frac{4}{12} = \frac{1}{3} = 0{,}333$$

Rezultat nam pove, da je na izpitu dosežek do 45 točk zbrala 1/3 kandidatov. To pomeni, da je izpit opravilo 2/3 oz. 66,7 % kandidatov.

V prvem delu naloge smo iz znanega kvantila računali njegov kvantilni rang. Postopek reševanja je bil naslednji:

$$y_P \rightarrow R_P \rightarrow P_y$$

B) V drugem delu naloge poizvedujemo, kolikšna bi bila točkovna meja za pozitivno oceno, če bi moralo izpit opraviti 75 % kandidatov. Tako je znan kvantilni rang $P_y = 0{,}25$; iskana vrednost je 25. centil oziroma 1. kvartil, ki ga izračunamo v naslednjih korakih:

- Iz relacije med obema rangoma računamo rang R_P :
  $R_P = P_y \cdot N + 0{,}5 = 3{,}5$
- Ta izračunani rang ustreza neenačbam:
  $R_0 \leq R_P \leq R_1$, oziroma za naše vrednosti:
  $R_0 = 3 < R_P < R_1 = 4$ ter
  $y_0 \leq y_P \leq y_1$, kar v našem primeru pomeni:
  $y_0 = 38 < y_P < y_1 = 42$
- Želeno vrednost kvantila izračunamo na podlagi obrazca:
  $$y_P = y_0 + \frac{R_P - R_0}{R_1 - R_0} \cdot (y_1 - y_0) = 40 \text{ točk}$$
  Izračunana vrednost pomeni, da bi izpit opravilo 75 % kandidatov, če bi bila točkovna meja za pozitivno oceno 40 točk.

V drugem delu naloge smo iz znanega kvantilnega ranga računali njegov kvantil, in sicer po naslednjem postopku po korakih:

$$P_y \to R_P \to y_P$$

Kvantilne range in kvantile smo doslej v tem poglavju računali le iz ranžirne vrste. Določamo jih lahko tudi iz frekvenčne porazdelitve na način, ki je predstavljen v učbeniku na spletni strani http://www.zavod-irc.si/docs/Skriti_dokumenti/STATISTIKA_ekon_sadl.pdf.

Z dopolnitvijo spoznanj 4. poglavja smo ob znanem absolutnem rangu ugotovili značilnosti kvantilnega ranga. Temu smo dodali spoznanja o kvantilu kot novo izraženi obliki standardnih vrednosti. Ugotovili smo neposredno povezavo med kvantili in mediano, ki smo jo preučevali v 5. poglavju. Pomembnejše pa je spoznanje, da s pravkar osvojenim znanjem, kot navaja Devjak (2008), lahko rešujemo dve vrsti problemov, in sicer: določanje pozicije statistične enote v množici glede na vrednost spremenljivke (z uporabo ranga in kvantilnega ranga) ter določanje vrednosti spremenljivke, da bi se le-ta razporedila na opredeljeno mesto v množici (kvantil). Svoj pomen imajo kvantili tudi pri ugotavljanju variabilnosti populacije, preučevanje te lastnosti pa je osrednja tema naslednjega poglavja.

Naloga

1. Pri izvajanju praktične vaje so pri 14-ih kandidatih merili čas, ki so ga le-ti potrebovali za njeno pravilno izvedbo. Ta je v naslednji tabeli izražen v minutah:

31	34	27	24	19	21	23	35	26	28	22	32	30	25

   Izračunajte:
   - Koliko odstotkov kandidatov je za nalogo potrebovalo 30 in več minut časa?
   - Koliko časa je za nalogo potreboval kandidat, od katerega ima četrtina vrstnikov višjo in tri četrtine nižjo vrednost dosežka?

7 VARIABILNOST

S statističnim opazovanjem zbrana množica podatkov je osnova za izračunavanje statističnih parametrov, od katerih vsak odkriva določene lastnosti pojava. V petem poglavju smo spoznali, da se vrednosti podatkov pogosto zgoščajo okrog določenih osrednjih vrednosti, čeprav so razlike med njimi lahko tudi velike. Govorimo o statistični razpršenosti oz. variabilnosti, ki jo razumemo kot variiranje (odklanjanje) vseh ali dela individualnih vrednosti neke populacije okoli mere sredine. V tem poglavju bomo najprej spoznali v praksi uveljavljene absolutne in relativne mere variabilnosti ter ugotovili njihove prednosti in slabosti. Računsko jih bomo uporabili za medsebojno primerjanje izbrane množice podatkov ter z njihovo pomočjo zmogli ugotovili bistvene značilnosti primerjanih pojavov. K temu bomo še dodali spoznanja o lastnostih normalne porazdelitve, najpogostejšega modela pri opisu, analizi in interpretaciji rezultatov statističnih raziskav.

<u>Uvodna naloga</u>

Naslednja tabela podaja število prijavljenih gostov v posameznih dnevih tedna v dveh hotelih. Ugotovimo podobnosti in razlike v obeh časovnih vrstah ter na podlagi uveljavljenih mer variabilnosti izračunajmo in utemeljimo rezultat, v katerem od obeh hotelov je njihovo število bolj variiralo.

Tabela 36: Število prijavljenih gostov v dveh izbranih hotelih

	Ponedeljek	Torek	Sreda	Četrtek	Petek	Sobota	Nedelja
Hotel A	67	96	98	102	103	144	160
Hotel B	30	84	91	105	110	154	196

Vir: Lasten

Podobnosti v obeh časovnih vrstah so še preprosto ugotovljive, saj podatki kažejo, da je v obeh hotelih enako skupno tedensko število gostov $(N = 770)$, *enaka je tudi aritmetična sredina* $(M = 110)$. *Za rešitev drugega dela naloge moramo uporabiti parametre variabilnosti, zato jih najprej spoznajmo, nalogo pa bomo rešili v nadaljevanju poglavja.*

Variabilnost spremenljivke merimo lahko z merami variabilnosti, ki jih delimo na dve skupini, in sicer kot navaja Šadl (2004):

- absolutne mere variabilnosti (variacijski razmik, varianca, standardni odklon …),
- relativne mere variabilnosti (koeficient variabilnosti …).

7.1 VARIACIJSKI RAZMIK

Variacijski razmik VR je najenostavnejša mera variabilnosti; izračunamo jo kot razliko med največjo in najmanjšo vrednostjo opazovane spremenljivke:

$$VR = y_{\max} - y_{\min}$$

Iz naslednjih podatkov ocen 10-ih kandidatov pri opravljanju poklicne mature:

14 17 13 21 18 12 11 22 20 19

ugotovimo, da znaša VR = <u>11 točk.</u>

S kvocientom rezultata s povprečnim številom točk M = 16,7 ugotovimo, da razpon med najnižjo in najvišjo oceno predstavlja 65,9 % navedene povprečne vrednosti.

7.2 KVARTILNI ODKLON

Kvartilni odklon Q je določen z obrazcem:

$$Q = \frac{Q_3 - Q_1}{2}$$

Kvartilni odklon je enak polovici kvartilnega razmika.

Izračunajte kvartilni odklon za prejšnji primer podanih 10-ih ocen poklicne mature!

<u>*Odgovor:*</u> $Q =$ <u>3,25 točke</u>

Na izračun variacijskega razmika ter kvartilnega odklona vplivajo le posamezne vrednosti, zato ti dve meri dajeta le približni sliki razpršenosti. Boljši so tisti parametri, ki jih izračunamo iz vseh vrednosti spremenljivke. Ena takšnih mer je varianca.

7.3 VARIANCA

Varianca VAR in aritmetična sredina M sta temeljna parametra v statistični analizi. Pri izračunu variance upoštevamo razlike vseh posameznih vrednosti opazovane številske spremenljivke Y od aritmetične sredine M; ker je vsota teh razlik enaka 0, jih kvadriramo. Varianco lahko računamo iz posameznih vrednosti ter iz frekvenčnih porazdelitev.

7.3.1 Izračun variance iz posameznih vrednosti

Varianca je definirana kot povprečje kvadratov odklonov posameznih vrednosti od aritmetične sredine ter jo izračunamo po obrazcu:

$$VAR = \sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - M)^2 \qquad \text{oz. poenostavljeno:} \qquad VAR = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i^2 - M^2)$$

V Excelu lahko varianco, ki temelji na podani populaciji, izračunamo z uporabo statistične funkcije **VARP**(number1;number2;...).

7.3.2 Izračun variance iz frekvenčne porazdelitve

Posamezne vrednosti spremenljivke so v frekvenčni porazdelitvi nadomeščene s sredinami razredov, v izračunu zato upoštevamo razlike (odklone) njihovih vrednosti od aritmetične sredine:

$$VAR = \sigma^2 = \frac{1}{N} \cdot \sum_{j=1}^{N} f_j \cdot (y_j - M)^2$$

Varianca je sicer temeljna mera variabilnosti, ki pa jo najpogosteje uporabljamo posredno tako, da izračunamo kvadratni koren iz njene vrednosti. To mero variabilnosti imenujemo standardni odklon.

7.4 STANDARDNI ODKLON

Standardni odklon σ je najpogosteje uporabljana mera razpršenosti. Predstavlja povprečje odklonov v opazovani množici vrednosti enot in je določen z:

$$\sigma = \sqrt{VAR} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - M)^2}$$

Na celotni podani populaciji lahko standardni odklon izračunamo tudi z Excelovo statistično funkcijo **STDEVP**(number1;number2;...).

Pri podani frekvenčni porazdelitvi, kjer so podatki grupirani v skupine, velja:

$$\sigma = \sqrt{VAR} = \sqrt{\frac{1}{N} \cdot \sum_{j=1}^{N} f_j \cdot (y_j - M)^2}$$

Standardni odklon je izražen v istih merskih enotah kot opazovana številska spremenljivka. Višja vrednost σ pomeni tudi večjo razpršenost opazovanih vrednosti.

Za standardni odklon velja, da je pri simetričnih porazdelitvah njegova vrednost približno enaka 1/6 variacijskega razmika, torej:

$$\sigma \approx \frac{1}{6} VR \quad \text{oziroma} \quad VR \approx 6\sigma$$

Velja tudi, da so pri simetričnih porazdelitvah vrednosti opazovane spremenljivke le redko zunaj intervala $(M - 3\sigma,\ M + 3\sigma)$. Kot bomo spoznali, to še posebej velja za normalne (Gaussove) porazdelitve. V povezavi z njo standardni odklon tudi najlažje razumemo.

S pomočjo formul in statističnih funkcij elektronske preglednice izračunajmo za podatke v naslednji tabeli aritmetično sredino ter varianco in standardni odklon. Zanje tudi določite pravilne merske enote.

Tabela 37: Porazdelitev starosti turistov v določenem hotelu v izbranem letu ter izračun pripadajočih značilnih mer variabilnosti

Leta	f_j	y_j	$f_j \cdot y_j$	$(y_j - M)$	$(y_j - M)^2$	$f_j \cdot (y_j - M)^2$
15–19	85	17,0	1.445,0	–24,5	600,7	51.056
20–24	115	22,0	2.530,0	–19,5	380,6	43.766
25–29	143	27,0	3.861,0	–14,5	210,5	30.100
30–39	330	34,5	11.385,0	–7,0	49,0	16.209
40–49	371	44,5	16.509,5	3,0	9,0	3.320
50–59	288	54,5	15.696,0	13,0	169,0	48.610
60–69	168	64,5	10.836,0	23,0	529,0	88.808
Skupaj	**1.500**					**281.869**

Vir: Lasten

Za navedene podatke izračunamo naslednje vrednosti: aritmetična sredina 41,5 let, varianca 187,9 (let)^2 ter standardni odklon 13,7 let .

7.5 KOEFICIENT KVARTILNEGA ODKLONA

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Knežević (2006) navaja, da lahko koeficient **kvartilnega odklona** V_Q zavzame vrednosti med 0 (ko ni razlik med razpršitvami) in 1 (kamor se približuje s povečevanjem navedenih razlik).

7.6 KOEFICIENT VARIABILNOSTI

Pogosto želimo primerjati variabilnost dveh ali več pojavov (Čibej, 1994), katerih spremenljivke so izražene v različnih merskih enotah. V takšnih primerih uporabljamo **koeficient variabilnosti** $KV\ \%$, ki je najbolj znana relativna mera variabilnosti.

$$KV\ \% = 100 \cdot \frac{\sigma}{M}$$

<u>Rešitev uvodne naloge</u>

Dokončajmo nalogo iz začetka poglavja ter v ta namen za podatke o številu prijavljenih gostov v obeh hotelih izračunajmo variacijski razmik, varianco, standardni odklon in koeficient variabilnosti; njihove vrednosti primerjajmo ter ugotovimo značilnosti.

Tabela 38: Število prijavljenih gostov ter izračun pripadajočih značilnih mer variabilnosti

	Hotel A			Hotel B		
Dan	**Št. gostov**	$(y_j - M)$	$(y_j - M)^2$	**Št. gostov**	$(y_j - M)$	$(y_j - M)^2$
Ponedeljek	67	–43	1.849	30	–80	6.400
Torek	96	–14	196	84	–26	676
Sreda	98	–12	144	91	–19	361
Četrtek	102	–8	64	105	–5	25
Petek	103	–7	49	110	0	0
Sobota	144	34	1.156	154	44	1.936
Nedelja	160	50	2.500	196	86	7.396
Skupaj	**770**	**0**	**5.958**	**770**	**0**	**16.794**

Vir: Tabela 36

<u>Izračunani parametri so naslednji:</u>

	Hotel A	**Hotel B**
$M =$	110	110
$VR =$	93	166
$VAR =$	851,1	2.399,1
$\sigma =$	29,2	49,0
$KV\ \% =$	26,5 %	44,5 %

Skupini rezultatov potrjujeta, da število gostov bolj variira v hotelu B. V obeh hotelih imamo v opazovanem obdobju enega tedna sicer enako skupno število gostov, enaki sta tudi aritmetični sredini *M*, toda izračunana standardna odklona σ ter koeficienta variabilnosti

$KV\ \%$ imata v drugem hotelu znatno višji vrednosti. To pa ravno kaže na večjo variabilnost števila prijavljenih gostov v hotelu B.

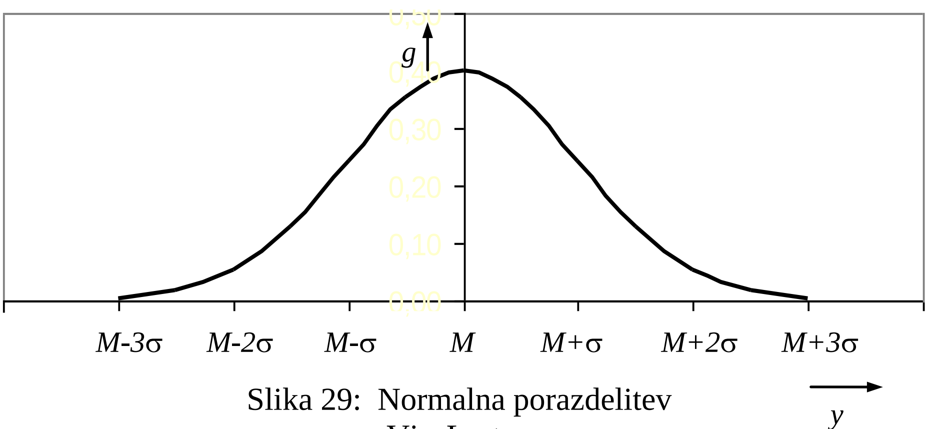
Tudi pri primerjavah lastnosti skupin, katerih podatki imajo različno velike aritmetične sredine in standardne odklone, je za analizo stopnje variiranja merodajna medsebojna primerjava vrednosti pripadajočih koeficientov variabilnosti.

7.7 LASTNOSTI NORMALNE PORAZDELITVE

V statistiki pomeni frekvenčna porazdelitev razpršenost vrednosti spremenljivke ali pogostost pojavljanja rezultata v opazovani množici, kot poudarja Knežević (2006). Naši dosedanji primeri so nam predstavljali stvarne (empirične) porazdelitve. Poznamo pa tudi porazdelitve, ki bazirajo na teoretičnih izhodiščih ter vnaprej znano večino statističnih parametrov (M, Me, Mo, σ). Te teoretične porazdelitve so osnova za vzorčenje, na osnovi katerega so razvite metode ocenjevanja in sklepanja, zato v praksi lahko dobro služijo za modele pri opisu, analizi ter interpretaciji rezultatov naših raziskav. Najpogosteje uporabljana teoretična porazdelitev je normalna (Gaussova) porazdelitev.

Normalna porazdelitev je temeljna teoretična porazdelitev. Pomembna je po svojih teoretičnih lastnostih in praktičnem pomenu, saj je – kot navaja Korenjak Černe (2009 b) – v vsakdanjem življenju niz porazdelitev, ki so po svojih značilnostih podobne normalni porazdelitvi. Je simetrična porazdelitev, pri kateri se enote gostijo v obliki Gaussove krivulje; ta prikazuje gostoto relativnih frekvenc g ter jo skladno s Trstenjakom (2001) zapišemo kot:

$$g^0(y) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(y-M)^2}{2\sigma^2}}$$



Slika 29: Normalna porazdelitev
Vir: Lasten

Za vse normalne porazdelitve na splošno velja, da v:

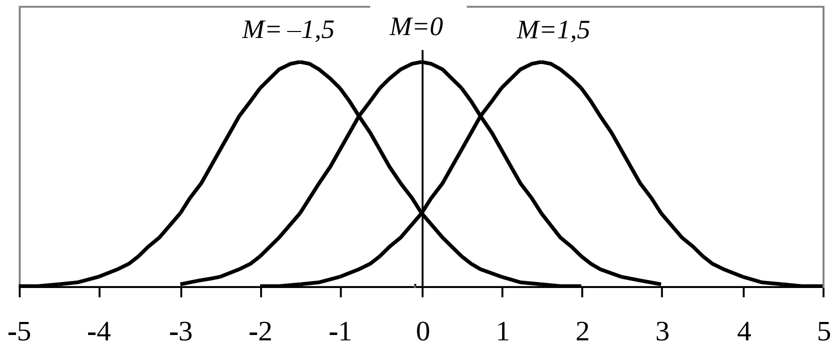
- intervalu $(M - \sigma$ do $M - \sigma)$ leži 68,3 % vseh vrednosti spremenljivke,
- intervalu $(M - 2\sigma$ do $M - 2\sigma)$ leži 95,4 % vseh vrednosti spremenljivke in
- intervalu $(M - 3\sigma$ do $M - 3\sigma)$ leži kar 99,7 % vseh njenih vrednosti.

Značilnost normalnih porazdelitev je, da je krivulja g normalno porazdeljene spremenljivke:

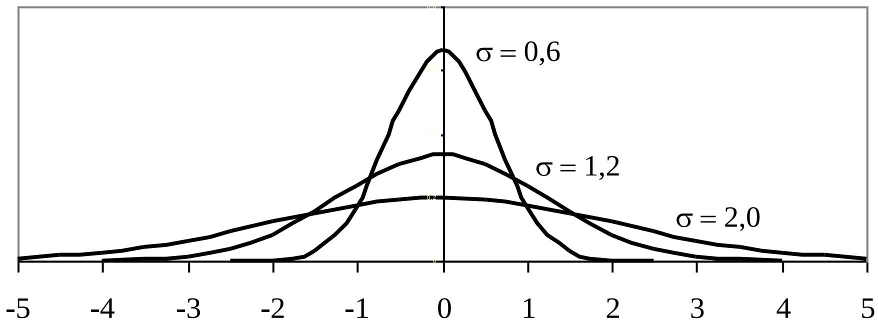
- enomodusna in zvonaste oblike ter
- simetrična, zato velja $M = Me = Mo$.

Normalna porazdelitev je natanko določena z dvema parametroma:

- aritmetično sredino M, ki določa lego krivulje ter s
- standardnim odklonom σ, ki določa njeno obliko.



Slika 30: Odvisnost normalne porazdelitve od parametra *M*
Vir: Lasten



Slika 31: Odvisnost normalne porazdelitve od parametra σ
Vir: Lasten

Razmislimo ter ob reševanju naslednje naloge tvorno uporabimo spoznane značilnosti normalne porazdelitve (prirejeno po Korenjak Černe (2009 b)).

Predpostavimo, da je teža štruc kruha, kupljenih v določeni pekarni, normalno porazdeljena slučajna spremenljivka z aritmetično sredino 1 kg in s standardnim odklonom 2 dag. V tednu dni v pekarni spečejo in prodajo 1.000 štruc. Izračunajmo:

a) *Kolikšna je lahko največja in najmanjša teža pri večini štruc?*
b) *Koliko % in koliko štruc skupaj je težjih od 98 dag in lažjih od 102 dag?*
c) *Koliko štruc je težjih od 102 dag ter lažjih od 98 dag?*
d) *Število štruc, ki so lažje od 96 dag.*
e) *Koliko štruc ustreza kakovosti teže, če je dovoljeno odmikanje 4 dag od aritmetične sredine?*

Rešitve s komentarjem

a) *Najmanjša teža večine štruc je lahko 94 dag, največja pa 106 dag.*
(v intervalu $(M-3\sigma,\ M+3\sigma)$ leži 99,7 % vseh vrednosti spremenljivke)
b) *Na omenjenem intervalu se nahaja 68,3 % izdelkov, to je 683 štruc.*
(v intervalu $(M-\sigma,\ M+\sigma)$ leži 68,3 % vseh vrednosti spremenljivke).
c) *Lažjih od 98 dag ter težjih od 102 dag je (100 % – 68,3 %) = 31,7 % izdelkov oz. 317 štruc; polovica (15,85 %) od teh je lažjih od 98 dag, polovica (15,85 %) pa težja od 102 dag.*
d) *Lažjih od 96 dag je 23 štruc, saj je skupaj težjih od 104 dag ter lažjih od 96 dag: (100 % – 95,4 %) = 4,6 % izdelkov oz. 46 kosov.*
e) *Na intervalu med 96 dag in 104 dag je 95,4 % izdelkov oz. 954 štruc.*

Zbrani podatki o proučevanih pojavih so običajno razpršeni oziroma se bolj ali manj gostijo okrog srednjih vrednosti. Ostin (2008) navaja, da se lahko vrednosti spremenljivk med seboj močno razlikujejo – variirajo ali pa so si zelo podobne. Da lahko te lastnosti spremenljivk medsebojno ustrezno primerjamo, jih moramo meriti. V ta namen smo najprej spoznali postopke izračunavanja ter pomen in razlike med merami variabilnosti, ki se najpogosteje uporabljajo. Te so: variacijski razmik, kvartilni odklon, varianca, standardni odklon, koeficient kvartilnega odklona ter koeficient variabilnosti.

Na priloženih primerih iz prakse smo z računalniško podporo izračunali več mer variabilnosti. Primerjava dobljenih rezultatov nam je dala osnovo za argumentirano sklepanje o značilnostih proučevanega pojava. Poudarek smo dali dvema pomembnima merama variabilnosti: varianci in standardnemu odklonu, saj prva velja skupaj z aritmetično sredino za temeljna parametra v statistični analizi. Praksa potrjuje, da je standardni odklon σ *najpogosteje uporabljana mera razpršenosti, njegova opazna vloga pa je tudi v tesni povezavi z lastnostmi normalne porazdelitve; zanjo smo tudi spoznali karakteristike in zakonitosti. Vloga normalne porazdelitve je vsekakor pomembna pri statističnem proučevanju pojavov, saj so v praksi pojavljajoče se porazdelitve pogosto podobne normalni, ker so te simetrične, unimodalne – velja torej relacija* $M = Me = Mo$ *– ter zvonaste oblike, kot navaja Korenjak Černe (2009 b). Pri normalni porazdelitvi je gostota relativnih frekvenc, ki smo jo že spoznali pri opisovanju frekvenčnih porazdelitev v 4. poglavju, definirana z obrazcem Gaussove krivulje in značilno grafično podobo. Pri tem je normalna porazdelitev natanko določena z dvema parametroma, in sicer: lego krivulje določa aritmetična sredina* M*, njeno obliko pa standardni odklon* σ*.*

Poglobljeno znanje za reševanje vprašanj s tega področja si lahko pridobimo v literaturi, navedeni v 9. poglavju, pa tudi na spletu. Za pomoč navajamo še nekaj povezav:

- http://ablejec.nib.si/Statistika/Vaje-iz-Statistike.pdf
- http://www.fsp.uni-lj.si/Statistika/2006/NormalnaPorazdelitev.pdf
- http://www.visoka-sola.com/pripone/Stud-S-Operacijske%20raziskave%20in%20statistika.ppt
- http://www.student-info.net/sis-mapa/skupina_doc/fu/knjiznica_datoteke/1232009155_zip_vsebina_i_statistika_prosojnice_predavanja_2_devjak_2006_redni_lj.pdf
- http://agri.bfro.uni-lj.si/predmeti/BiomRac/zapiski/05.Razprsenost.pdf

Naloge

1. V določenem časovnem obdobju smo opazovali zamude letal, katerih število v 10-minutnih intervalih je naslednje:

Zamuda (minut)	0–10	11–20	21–30	31–40	41–50	51–60
Število letal	14	18	9	12	13	7

Izračunajte aritmetično sredino, standardni odklon, koeficient variabilnosti, mediano in modus.

2. V izbranem hotelu so proučevali bivanje turistov glede na število nočitev. Rezultate tega proučevanja prikazuje naslednja frekvenčna porazdelitev:

Dnevi	1–5	6–10	11–15	16–20	21–25	26–30
Število gostov	46	31	22	12	13	5

   Ugotovite vrednosti aritmetične sredine, standardnega odklona, koeficienta variabilnosti, mediane in modusa.

3. V naslednji tabeli je podana frekvenčna porazdelitev uspeha dijakov na poklicni maturi. Kolikšne so vrednosti aritmetične sredine, standardnega odklona, koeficienta variabilnosti, mediane in modusa? Rezultate povežite z ugotovitvami o značilnostih frekvenčnih porazdelitev v 4. poglavju ter srednjih vrednosti v 5. poglavju. Na primerjavi vrednosti aritmetične sredine, mediane in modusa utemeljite vrsto dobljene nesimetrične porazdelitve.

Točke	8–9	10–11	12–13	14–15	16–17	18–19	20–21	22–23
Št. dijakov	270	1.046	1.716	1.643	1.364	863	434	117

   Vir: RIC, Državni izpitni center, Letno maturitetno poročilo o poklicni maturi 2007;
   http://www.ric.si/mma_bin.php/$fileI/2008052807112324/$fileN/LP-PM-2007.pdf

4. Za podatke v tabeli 10 izračunajte aritmetično sredino, varianco, standardni odklon in koeficienta variabilnosti. Rezultate preverite in ugotovite, če ste izračunanim vrednostim določili pravilne merske enote.

5. Naslednja naloga se nahaja v Korenjak Černe (2009 b, 25):
   »Predpisana teža zavitkov je 1 kg, standardni odklon pa 1 dag. Predpostavljamo, da delujejo na proizvodnjo le slučajni dejavniki in se teža porazdeljuje normalno.
   a) Prikažite skico porazdelitve za težo zavitkov.
   b) Koliko odstotkov zavitkov bo težkih med 99 in 101 dag?
   c) Koliko odstotkov zavitkov bo lažjih od 99 dag?
   d) Koliko odstotkov zavitkov bo težjih od 99 dag?
   e) Koliko odstotkov zavitkov bo neustrezne kakovosti, če je toleranca teže 100 ± 2 dag?«

8 ANALIZA ČASOVNIH VRST

V dosedanjih razmišljanjih smo spoznali, kako s statistično dejavnostjo opazujemo pojave, o njih zbiramo podatke ter jih obdelujemo in računamo parametre. Pri ugotavljanju značilnosti pojavov v preteklosti pridemo do ugotovitev, da se pri njih pogosto pojavlja medsebojni vpliv med spremenljivkami, zato upravičeno predpostavljamo, da so med vrednostmi le-teh določene soodvisnosti. Zlasti v turizmu in gostinstvu lahko najdemo takšnih primerov kar precej, npr: letna obdobja in število turistov v obmorskih krajih, finančni stroški reklame in tržni delež, število gostov ženskega in moškega spola na določenem turističnem območju, zunanja temperatura in število gostov v obmorskih krajih … Tako se pojavi vprašanje: Ali je možno na podlagi izračuna medsebojne povezanosti spremenljivk ter vrednosti ene od njih napovedati vrednost druge spremenljivke?

Praksa kaže, da se – še zlasti na področju turizma – velik del pojavov časovno spreminja. V tem poglavju bomo spoznali najpomembnejše dejavnike teh sprememb ter metode njihovega določanja in računanja parametrov. Ugotovili bomo, kako lahko ob poznavanju dosedanjih lastnosti določenega pojava tudi predvidevamo njegovo bodoče spreminjanje, podobno kot navaja B. Košmelj (1996). Videli bomo, da statistične metode pod določenimi pogoji nudijo orodja za takšne napovedi trenda in prognoze razvoja pojava v prihodnosti. Izbrane pristope bomo preizkusili in ocenili njihovo ustreznost.

<u>Uvodna naloga</u>

Na osnovi statističnih podatkov o letnem številu z letali prepeljanih potnikov v R Sloveniji v obdobju od 1999 do 2007, napovejmo njihovo predvideno število v letih 2008, 2009 in 2010. Preizkusimo različne možnosti, jih medsebojno primerjajmo in ocenimo ter poiščimo najustreznejši pristop pri napovedi trenda v našem primeru. Ugotovljeno vrednost tudi primerjajmo z napovedjo, dobljeno pri izračunu s povprečnimi koeficienti rasti v poglavju 5.8.4.

Tabela 39: Število z letali prepeljanih potnikov v R Sloveniji v letih od 1999 do 2007

Leto	1999	2000	2001	2002	2003	2004	2005	2006	2007
Št. potnikov	780	866	801	814	864	885	944	1.018	1.136

Vir: SURS, Statistični letopis 2003, http://www.stat.si/letopis/2003/22 (12. 2. 2009) in Statistični letopis 2008, http://www.stat.si/letopis/2008/21 (12. 2. 2009)

V tem poglavju bomo postopno odgovorili na v uvodu postavljeni vprašanji. Ko bomo spoznali dejavnike spreminjanja pojavov in sestavine časovnih vrst, bomo z znanjem temeljnih zakonitosti regresije in koleracije zmogli za izbrani pojav tudi ugotoviti osnovno smer njegovega nadaljnjega gibanja.

8.1 DEJAVNIKI SPREMINJANJA POJAVOV IN SESTAVINE ČASOVNIH VRST

Podatki v časovnih vrstah prikazujejo spremembe pojavov v odvisnosti od časa, zato z njimi proučujemo časovni razvoj pojavov. Poznamo že osnovna orodja za njihovo proučevanje, kot so indeksi, koeficienti rasti in stopnje rasti, ki nam kažejo dinamiko časovne vrste. Za nadaljnje spoznavanje značilnosti pa bomo sedaj spoznali tiste postopke, s katerimi lahko pri razčlenjevanju časovnih vrst njihove komponente ocenimo s statističnimi metodami, kot to opredeljuje Terminološki slovar (2008).

Spremembe časovne vrste povzročajo **slučajni in neslučajni dejavniki.** Košmelj (1996) in Leskošek (2008) navajata, da lahko večino vzorcev časovnih vrst opišemo z naslednjimi komponentami:

- **Trend** T kaže dolgoročno gibanje pojava, podaja smer (premico ali krivuljo) razvoja. Dolgoročne spremembe so pogojene npr. z biološkimi dejavniki, spremembami v okolju …
- **Sezonska komponenta** S se nanaša na ponavljajoče se spremembe, nastale zaradi letnih časov, vremena … Te oscilacije imajo svojo konstantno dolžino, najpogosteje se nanaša na eno leto (turizem, kmetijstvo …).
- **Ciklična komponenta** C se nanaša na neperiodična gibanja okoli trenda. Dolžina cikla ni določena, a je daljša od enega leta. Vzroki: dolgoročne spremembe v gospodarstvu in v okolju, moda …
- **Slučajna komponenta** (neregularni vplivi) I se nanaša na nesistematična in slučajna gibanja, ki jih ne moremo pojasniti s prejšnjimi tremi komponentami T, S in C.

V časovnih vrstah niso vedno prisotne vse navedene sestavine. V njih nikoli ne manjka slučajna komponenta, v večini praktičnih vzorcev časovnih vrst pa najdemo še predvsem komponenti trenda in sezonskosti.

»Teorija časovnih vrst skuša identificirati te komponente, jih razumeti in uporabiti pri napovedovanju. Za vsako časovno vrsto skušamo identificirati njen model. Modelov je veliko vrst. Najsplošnejša sta dva« (Košmelj, 2008, 5):

- aditivni model:

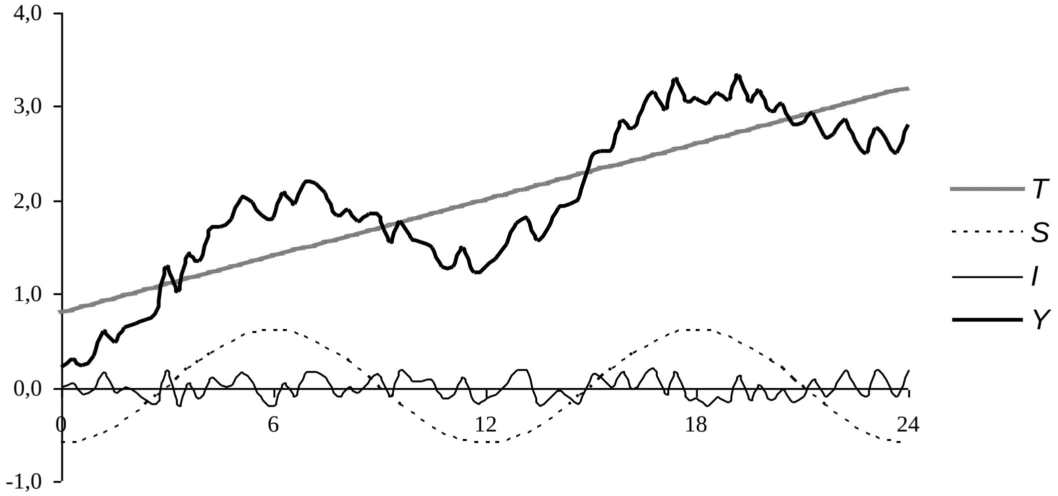
  $$Y = T + S + C + I$$

- multiplikativni model:

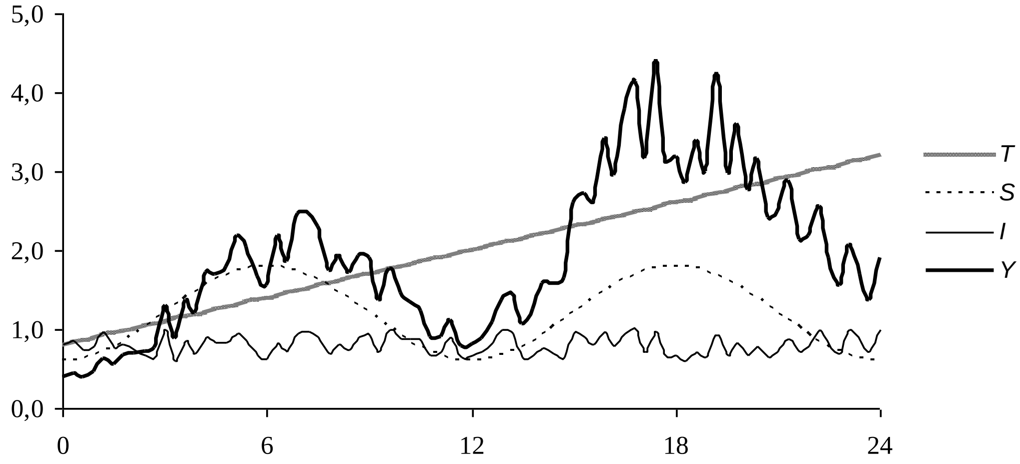
  $$Y = T \cdot S \cdot C \cdot I \quad \text{oziroma:}$$

  $$\log Y = \log T + \log S + \log C + \log I$$

Primerjajmo grafično oba modela:

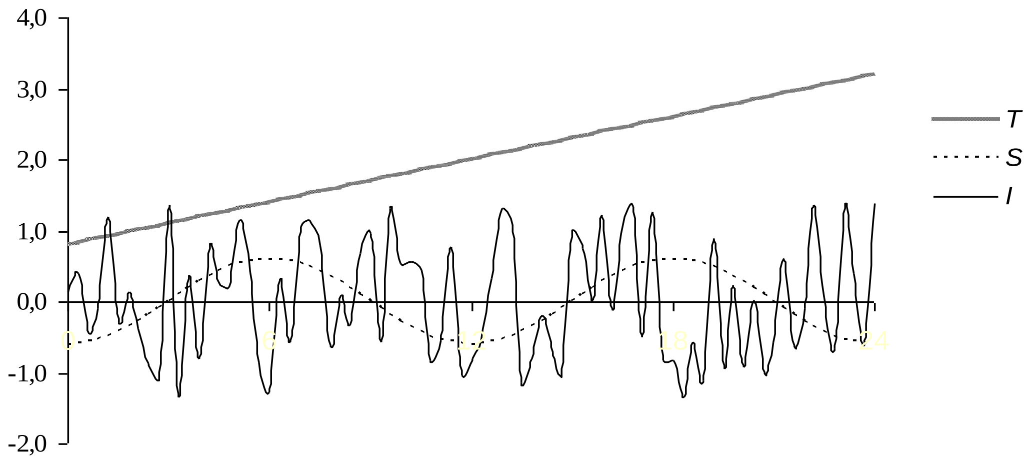


Slika 32: Aditivni model $Y = T + S + I$ za časovno vrsto
Vir: Lasten

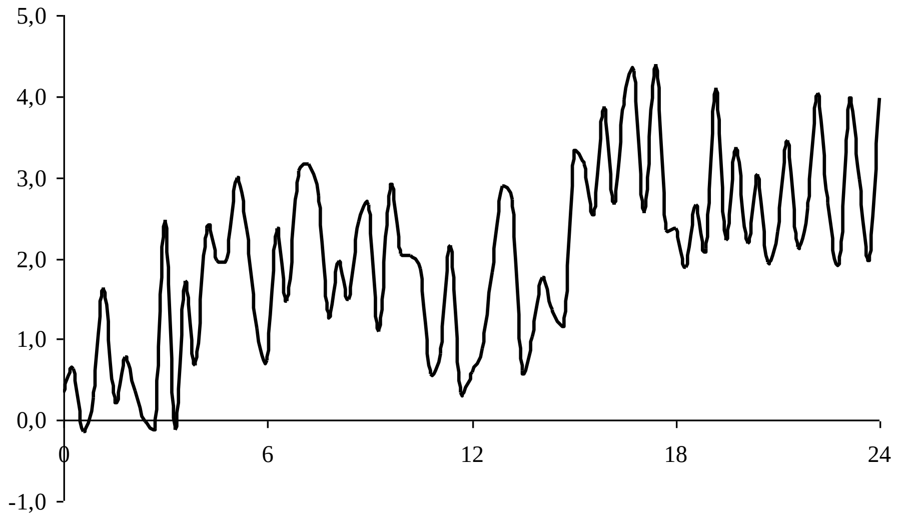


Slika 33: Multiplikativni model $Y = T \cdot S \cdot I$ za časovno vrsto
Vir: Lasten

Kadar je v praksi variabilnost slučajne komponente prekomerna ter izstopajoča, lahko ta močno zabriše vpliv trenda in sezonske komponente v časovni vrsti, kar je razvidno iz naslednjih dveh slik.



Slika 34: Linearni trend ter sezonska in izstopajoče velika slučajna komponenta
Vir: Lasten



Slika 35: Aditivni model za časovno vrsto z izstopajočo slučajno komponento
Vir: Podatki s slike 34

Za ugotovitev trenda, sezonske ali ciklične komponente je zaradi zmanjšanja napake potrebno v časovni vrsti najprej izločiti vpliv slučajne spremenljivke. To naredimo s postopki in tehnikami glajenja, čemur sledi poskus prileganja podatkov predvideni matematični funkciji, s čimer se bomo spoznali v nadaljevanju poglavja.

8.2 DOLOČANJE TRENDA, KORELACIJA IN REGRESIJA

Ljubič (2008) navaja, da trend T podaja osnovno smer gibanja pojava ter ga tolmačimo kot prirastek oziroma zmanjšanje vrednosti pojava v časovni enoti. Obliko trenda lahko določimo:

- prostoročno na grafičnem prikazu ali
- analitično z metodami regresijske analize.

8.2.1 Regresija in koleracija

V življenju so pojavi pogosto medsebojno soodvisni, zato v takšnih primerih zbrani podatki medsebojno kolerirajo. Ugotovili bomo, da statistika ne omogoča le izračuna medsebojne povezanosti spremenljivk, temveč lahko te koleracije uporabimo kot možno napoved iz ene spremenljivke v drugo (Knežević, 2006).

O **regresiji** na splošno govorimo v primerih, ko sta dva ali več pojavov v medsebojni odvisnosti. Kadar nastopata v medsebojni odvisnosti samo dva pojava oz. količini, govorimo o enostavni regresiji (Pelhan, 2008).

Odvisnost je enostranska $X \to Y$, kadar je pojav X vzrok ter pojav Y posledica (na primer: odvisnost med telesno višino staršev in njihovih otrok). V primerih, da ni možno določiti, kaj je vzrok in kaj posledica, pa je odvisnost dvostranska $X \leftrightarrow Y$ (na primer: odvisnost med telesno višino bratov oz. sester).

S pomočjo regresije lahko poiščemo tako funkcijo, ki najbolje podaja medsebojno odvisnost pojavov. Torej skozi območje točk (x_j, y_j) položimo takšno simetralo oz. krivuljo, ki se čim bolje prilega raztresenim točkam obeh spremenljivk. Ta krivulja, ki podaja medsebojno odvisnost spremenljivk, je regresijska funkcija $y = f(x)$, ki jo lahko v splošni obliki zapišemo kot:

$$y = ax^0 + bx^1 + cx^2 + \cdots + nx^n$$

Kadar ima regresijska funkcija le prva dva člena, govorimo o **linearni regresiji** oziroma regresiji 1. reda:

$$y = ax^0 + bx^1 \quad \text{oz.} \quad y = a + bx$$

Pri tej regresiji, ki je najpreprostejša in jo bomo obravnavali, se enačba regresijske funkcije poenostavi v enačbo premice. Nanjo se bomo tudi naslonili pri določanju linearnega trenda. Pri regresijah višjih redov, ki so nelinearne oz. krivuljne, pa ima regresijska funkcija večje število členov.

Kako dobra je matematična povezava med količinama X in Y, ki ju povezuje regresijska premica, ugotavljamo s pomočjo **koleracije.** Knežević (2006) navaja, da je Pearson oblikoval postopek računanja **koeficienta koleracije** r_{xy} linearne povezanosti spremenljivk, in sicer kot razmerje:

$$r_{xy} = \frac{N \cdot \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{\sqrt{\left(\left(N \cdot \sum x_i^2\right) - \left(\sum x_i\right)^2\right) \cdot \left(\left(N \cdot \sum y_i^2\right) - \left(\sum y_i\right)^2\right)}}$$

Kompleksnost napisanega obrazca je nespodbudna za praktično uporabo, zato poiščemo prikladnejše možnosti ugotavljanja vrednosti koeficienta koleracije, uporabljane v praksi.

Pearsonov koeficient korelacije lahko – kot navaja Kovačič (2009) – izračunamo tudi s pomočjo naslednjega izraza:

$$r_{xy} = \frac{C_{xy}}{\sigma_x^2 \cdot \sigma_y^2}$$

pri čemer je kovarianca C_{xy}, ki kaže linearno povezanost med spremenljivkama, opredeljena kot povprečje produktov dvojic odklonov za vsak podatkovni par v dveh množicah podatkov in je določena z izrazom:

$$C_{xy} = \frac{1}{N} \cdot \sum \left(x_i - M_x\right) \cdot \left(y_i - M_y\right)$$

Najlažje Pearsonov koeficient kot mero linearne povezanosti dveh številskih spremenljivk izračunamo v Excelu s statistično funkcijo **CORREL**(array1;array2).

Izračunu kovariance pa je namenjena statistična funkcija **COVAR**(array1;array2).

Koeficient korelacije r_{xy} kaže smer in moč korelacije ter ima vrednost v intervalu [–1, 1]. Njegove značilne vrednosti so predvsem:

- +1; približevanje r_{xy} vrednosti +1 pomeni močno pozitivno povezanost (z večanjem vrednosti ene spremenljivke se veča tudi vrednost druge),
- 0; med spremenljivkama ni linearne povezanosti, obstaja lahko samo slučajna variabilnost,
- –1; približevanje r_{xy} vrednosti –1 pomeni visoko negativno linearno povezavo (z večanjem vrednosti prve spremenljivke se manjša vrednost druge).

Knežević (2004) tudi navaja, da je za natančnejšo predstavitev koeficienta koleracije na voljo tudi **determinacijski koeficent** r^2. Dobimo ga s kvadriranjem koeficienta koleracije r_{xy}; z njegovo vrednostjo se pokaže, kolikšen delež variance si delita obe spremenljivki, ki imata skupno razpršenost – delež pojasnjene variance.

V prodajni službi spremljajo stroške oglaševanja in vrednost povečanja prodaje. Utemeljiti želimo soodvisnost le-teh, če so podatki za opazovane trge naslednji:

Reklama (€)	*1.500*	*820*	*1.200*	*1.200*	*2.200*	*1.900*	*900*	*2.800*	*950*	*1.300*
Prodaja (€)	*1.850*	*1.200*	*1.050*	*1.750*	*3.100*	*2.500*	*1.050*	*3.100*	*1.500*	*1.850*

<u>Odgovor:</u> Izračunani koeficient koleracije r = 0,94 kaže na zelo visoko povezanost obeh časovnih vrst.

Spremljali smo ceno izbranega izdelka ter njegovo prodajo v določenem časovnem obdobju; zanima nas medsebojna odvisnost navedenih spremenljivk.

Cena (€)	*3,10*	*3,30*	*3,50*	*3,80*	*3,90*	*4,10*	*4,20*	*4,30*	*4,40*
Prodaja (kosov)	*130*	*125*	*122*	*118*	*112*	*104*	*106*	*97*	*104*

Odgovor: *Ugotovljeni koeficient koleracije r = –0,96 kaže na zelo močno negativno povezanost obeh spremenljivk.*

S korelacijo izražamo povezanost, z regresijo pa soodvisnost dveh spremenljivk. Pri linearni regresiji, oz. regresiji 1. reda predpostavljamo, da se opazovanemu pojavu najbolj prilega premica z enačbo:

$$y' = a + bx$$

pri tem y' pomeni ocenjeno vrednost y pri danem x.

Kako izračunati oba parametra premice, da bo ta skozi "korelacijski oblak" položena tako, da se čimmanj odklanja od točk vrednosti spremenljivke? Leskošek (2008, 2009) opredeljuje, da je običajen kriterij najmanjših kvadratov, kjer bo vsota kvadriranih napak najmanjša:

$$\sum (y - y')^2 = \min$$

Ob tem kriteriju sta ocenjeni vrednosti parametrov regresijske premice (Knežević, 2004):

$$b = \frac{C_{xy}}{\sigma_x^2} = \frac{N \cdot \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{N \cdot \sum (x_i)^2 - (\sum x_i)^2}$$

$$a = M_y - b \cdot M_x = \frac{\sum y_i - b \cdot \sum x_i}{N}$$

8.2.2 Ugotavljanje trenda

V tem poglavju obravnavamo dinamiko pojavov, pri katerih je množica kronološko urejenih podatkov za neko časovno obdobje nanizanih v ustrezno časovno vrsto, zato pri uporabi predhodnih obrazcev dejansko nastopata dve spremenljivki: opazovani pojav Y in čas t.

Če podatek v časovni vrsti označimo z Y_t, z enačbo y' izračunano vrednost trenda T_t in $t = 1, 2, \ldots, N$, velja zahteva metode najmanjših kvadratov zapisana kot:

$$\sum_{t=1}^{N} (Y_t - T_t)^2 = \min$$

Povežimo predstavljena dejstva ter iz pravkar ugotovljenega najdimo postopek za določanje trenda časovne vrste.

Kot navajajo Devjak (2009), B. Košmelj (1996) ter K. Košmelj (2008) lahko zgornje računanje vrednosti parametrov regresijskega modela ter s tem računanje parametrov **linearnega trenda** močno poenostavimo s transformiranjem časovne skale tako, da ordinatno os premaknemo v sredino časovnega intervala.

V ta namen je v praksi uveden t. i. tehnični čas x_t, za katerega velja pogoj:

$$\sum_{t=1}^{N} x_t = 0$$

Način določanja intervalov tehničnega časa za liho in sodo število letnih obdobij:

	Leto	*2000*	*2001*	*2002*	*2003*	*2004*	*2005*	*2006*	*2007*	$\sum x_t$
N=2k +1	x_t		*–3*	*–2*	*–1*	*0*	*1*	*2*	*3*	*0*
N=2k	x_t	*–3,5*	*–2,5*	*–1,5*	*–0,5*	*0,5*	*1,5*	*2,5*	*3,5*	*0*

S tem se sedaj za funkcijo linearnega trenda:

$$T_t = a + bx_t$$

izračun obeh parametrov poenostavi v:

$$a = \frac{\sum_{t=1}^{N} Y_t}{N} = \overline{Y} = M_Y \qquad \text{ter} \qquad b = \frac{\sum_{t=1}^{N} Y_t \cdot x_t}{\sum_{t=1}^{N} x_t^2}$$

Excelova statistična funkcija **TREND**(known_y's;known_x's;new_x's;const) vrne vrednost vzdolž linearnega trenda.

<u>Rešitev uvodne naloge:</u>

Z osvojitvijo znanja predstavljenih postopkov lahko začnemo v nalogi iz začetka poglavja izračunavati parametre in pripadajoče vrednosti trenda. Pri primerjanju in vrednotenju dobljenih vmesnih rezultatov ter iskanju optimalne rešitve si lahko zelo pomagamo tudi z Excelovo grafično podporo.

Tabela 40: Število z letali prepeljanih potnikov v R Sloveniji v letih od 1999 do 2007 ter izračun trenda

Leto	**Število potnikov** (v 1000)	x_t	$Y_t \cdot x_t$	x_t^2	T_t
1999	780	–4	–3.120	16	752
2000	866	–3	–2.598	9	789
2001	801	–2	–1.602	4	826
2002	814	–1	–814	1	864
2003	864	0	0	0	901
2004	885	1	885	1	938
2005	944	2	1.888	4	975
2006	1.018	3	3.054	9	1.013
2007	1.136	4	4.544	16	1.050

Vir: Tabela 39

Izračunana parametra linearnega regresijskega modela sta:

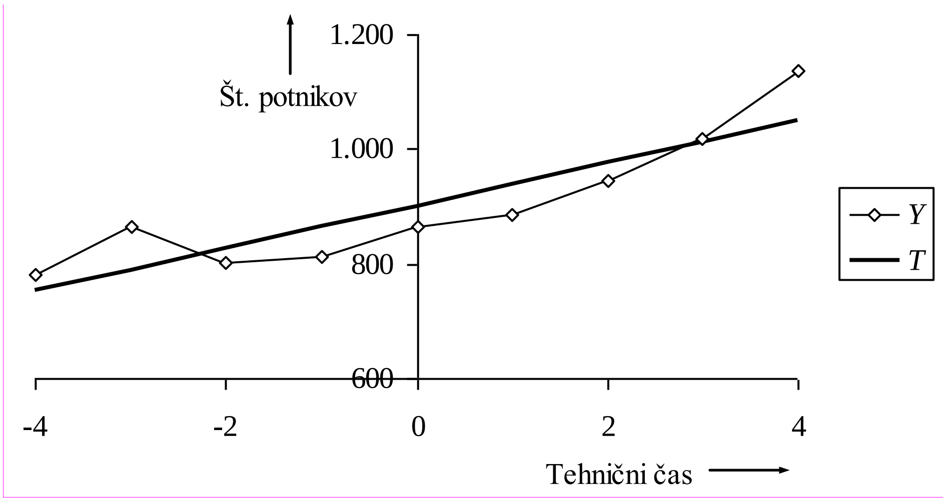
$$a = 901 \quad \text{in} \quad b = 37{,}28$$

Premica linearnega trenda je tako določena s funkcijo:

$$T_t = 901 + 37{,}28 \cdot x_t$$

Napovedi za naslednja izbrana 3 leta znašajo:

$$T(2008) = T(t=5) = 1.087.000$$
$$T(2009) = T(t=6) = 1.125.000$$
$$T(2010) = T(t=7) = 1.162.000$$



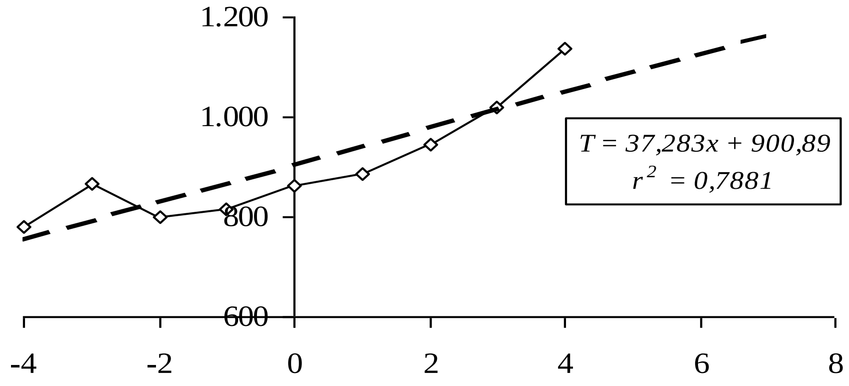
Slika 36: Število z letali prepeljanih potnikov v R Sloveniji v letih od 1999 do 2007 ter vrednost trenda
Vir: Tabela 40

Učinkovito in nazorno podporo pri določanju trenda najdemo tudi v Excelu. V pogovornem oknu urejanja in dopolnjevanja grafov se lahko pri izbiri vrste dodane trendne črte odločamo med linearno, potenčno, logaritmično, eksponentno in polinomsko (od 2. do 6. stopnje) funkcijsko odvisnostjo ter glajenjem časovne vrste s pomočjo drsečega povprečja.

Slika 37: Dodajanje trendne črte v Excelu
Vir: Microsoft Excel 2003

Vendar se Excelova podpora pri določanja trenda ne konča samo z vrisom trendne premice ali krivulje. Imamo še dodatne možnosti: za izbrano vrsto regresije lahko uporabimo praktično poljuben časovni interval (vnaprej in nazaj) napovedovanja ter določimo, da aplikacija poleg grafične predstavitve trendne črte prikaže še njeno enačbo ter izračunano vrednost determinacijskega koeficienta. S tem je dana uporabniku prijazna možnost preizkušanja in primerjanja in ocenjevanja na poti k izbiri optimalnega rezultata.

Napovejmo sedaj z Excelom linearno trendno vrednost našega opazovanega števila potnikov za izbrana tri naslednja leta.



Slika 38: Napovedovanje v Excelu z linearnim trendom
Vir: Tabela 40 in Microsoft Excel 2003

Odčitana trendna vrednost leta 2010 (pri $t = 7$) se ujema s prej izračunanim številom potnikov, ki znaša 1.162.000 oseb. Po zaokrožitvi decimalk se tudi oba parametra premice v celoti ujemata s prejšnjim izračunom. Pogled na graf še bolj kot na prejšnji sliki izpostavi slabo ujemanje poteka linearne trendne črte z vrednostmi časovne vrste, kar potrjuje tudi precejšen odmik determinacijskega koeficienta r_{xy}^2 od maksimalne možne linearne povezave z vrednostjo 1,00.

Kvaliteto trenda ocenjujemo po navedbah Devjaka (2009) z dvema merama: s standardnim odklonom σ_T vrednosti pojava od vrednosti funkcije trenda ter s koeficientom variacije trenda KV_T:

$$\sigma_T = \sqrt{\frac{\sum (Y_t - T_t)^2}{N}}$$

$$KV_T\,\% = \frac{\sigma_T}{M_Y} \cdot 100$$

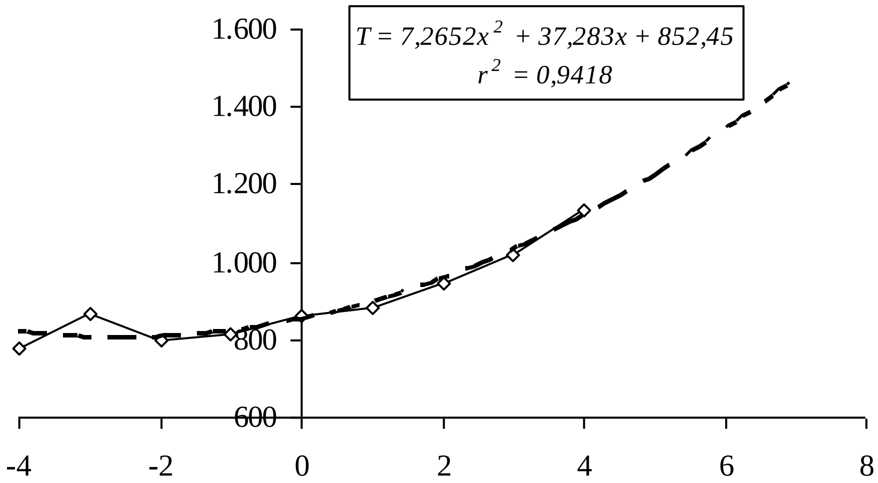
Manjša vrednost obeh mer odraža ustreznejšo funkcijo trenda.
V našem primeru sta izračunani vrednosti za uporabljen linearen trend:

$\sigma_T = 149.733$ potnikov
$KV_T\,\% = 16{,}6\ \%$

Upravičeno lahko dvomimo, da so rezultati, dobljeni na podlagi linearnega trenda, v našem primeru optimalni. Raziščimo, ali lahko najdemo takšno trendno črto, ki bi imela pri reševanju našega primera boljše karakteristike.

Nelinearna regresija

Podatkom v časovni vrsti se marsikdaj bolje kot premica prilega kakšna krivulja. Slika 37 kaže, da imamo lahko pri preizkušanju možnosti ter iskanju nadaljnjih, optimalnejših rešitev učinkovito podporo Excela. Izkoristimo to ter poglejmo, kakšne rezultate dobimo, če se pri iskanju boljših možnosti prileganja dodane trendne črte v Excelu odločimo za polinomsko funkcijo 2. reda, torej izberemo parabolično odvisnost.



Slika 39: Napovedovanje v Excelu s paraboličnim trendom
Vir: Tabela 40 in Microsoft Excel 2003

Za izbrani **parabolični trend** nam Excel vrne gornjo grafično odvisnost s pripisanimi parametri parabole in precej ugodnejšo vrednostjo determinacijskega koeficienta, kot jo je ta imel pri prej proučevanem linearnem trendu. Tudi izračun obeh mer kvalitete trenda sedaj kaže na njuno nižjo in s tem boljšo vrednost:

$$\sigma_T = 78.517 \text{ potnikov}$$
$$KV_T\% = 8{,}7\ \%$$

S pomočjo paraboličnega trenda dobljene napovedi za izbrana 3 nadaljnja leta kažejo višje vrednosti od tistih, dobljenih na podlagi linearnega trenda in znašajo:

$$T(2008) = T(t=5) = 1.220.000$$
$$T(2009) = T(t=6) = 1.337.000$$
$$T(2010) = T(t=7) = 1.469.000$$

Postopek ocenjevanja pojava v prihodnosti smo si zamislili že v 5. poglavju o relativnih številih. Na osnovi povprečnega koeficienta rasti, izračunanega za določeno časovno obdobje, smo takrat za podatke iste časovne vrste ocenili, da bo npr. leta 2010 z letali prepeljanih 1.388.000 potnikov. Primerjava pokaže, da je ta vrednost višja od tiste, izračunane s pomočjo linearnega trenda, ter nižja od pravkar ugotovljene s paraboličnim trendom, saj velja:

$$T_{lin(2010)} = 1.162.000 < 1.388.000 < T_{par(2010)} = 1.469.000$$

Dobili smo tri različne ocene napovedanih bodočih vrednosti. Še kakšno več bi dala izbira naslednje trendne črte z drugačnim prileganjem opazovanim podatkom. Kako se odločiti, kadar imamo na isto vprašanje več odgovorov? Uporaba IKT omogoča hitro in cenovno ugodno pridobivanje podatkov, ki so podlaga za poslovne in druge odločitve; te pa sprejemajo odgovorne osebe na podlagi svojih kompetenc in odgovornosti ter poslovnega rizika.

Dodatno znanje in pogled v analitične podrobnosti o paraboličnem trendu si lahko pridobimo v učbeniku: http://www.zavod-irc.si/docs/Skriti_dokumenti/STATISTIKA_ekon_sadl.pdf.

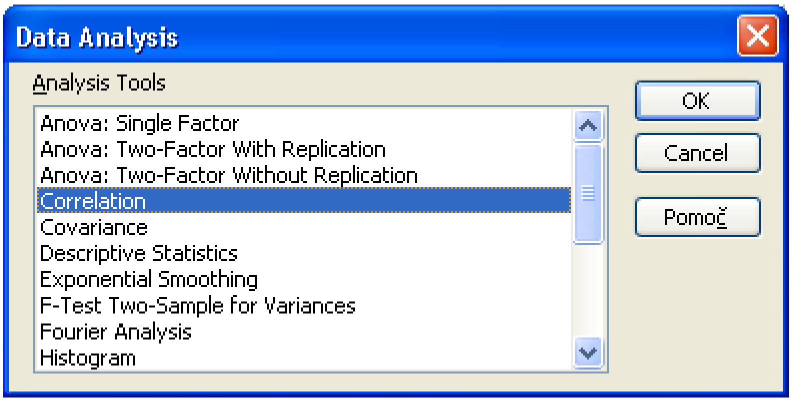
V Excelu je možno trendno črto dodati kateremukoli nizu podatkov v stolpčnem, paličnem, črtnem, ploščinskem in raztresenem (XY) grafikonu. Trendne črte ni mogoče dodati 3D-grafikonu, polarnemu, krožnemu, kolobarnemu in mehurčnemu grafikonu.

Metoda drsečih sredin

Med postopki in tehnikami glajenja, s katerimi se poskuša v časovni vrsti eliminirati vpliv slučajne komponente, s ciljem odkrivanja trenda, sezonske in ciklične komponente, se poleg regresijskih funkcij uporablja metoda drsečih sredin. Pri tem se originalni časovni vrsti priredi časovna vrsta drsečih sredin, katere členi so zaporedja povprečij, izračunanih iz delov niza podatkov opazovane časovne vrste. V grafu na novo kreirane časovne vrste se najbolje vidi, kako se zgladijo nihanja v podatkih in jasneje pokaže trend.

Več o metodi drsečih sredin na: http://les.bf.uni-lj.si/uploads/media/casovne_vrste.pdf

*V meniju **Orodja, Analiza podatkov** nudi Excel tudi zbirko podatkovnih orodij, imenovano "Orodja za statistično analizo", uporabno predvsem pri zahtevnejših statističnih analizah. Za uporabo posameznega orodja moramo zagotoviti potrebne podatke in parametre. Orodje uporabi primerne statistične funkcije z makri za obdelavo in prikaže rezultate v izhodni tabeli. Nekatera orodja izdelajo poleg izhodnih tabel tudi grafikone.*



Slika 40: Vstopno okno orodij za statistično analizo
Vir: Microsoft Excel 2003

8.3 ANALIZA PERIODIČNIH NIHANJ

Kako analizirati pojave, ki so periodični, se ponavljajo na krajša, a enako dolga časovna obdobja (po urah dneva, dnevih v tednu, mesecih v letu), npr.: cestni ali letalski promet, poraba energije, promet v trgovinah …? Kakšne metode lahko glede na dolžino periode takšnih pojavov uporabljamo pri njihovem proučevanju?

Pri ponavljajočih se periodičnih nihanjih pojavov je najdaljša perioda eno leto, v tem primeru govorimo o sezonskih nihanjih. Za proučevanje periodičnega nihanja pojavov potrebujemo podatke po obdobjih znotraj posamezne periode za več period. Pri tem želimo iz podatkov v časovnih vrstah izločiti vse ostale sestavine razen periodične, ki jo izrazimo s periodičnim indeksom. To je težje narediti pri sezonskih pojavih, kjer se poleg periodičnih in slučajnih nihanj navadno kaže še vpliv trenda ali cikličnih nihanj.

B. Košmelj (1996) navaja, da je za analizo periodične komponente najpreprostejša metoda vsot. Ta v časovni vrsti predpostavlja le prisotnost periodičnih in slučajnih nihanj, zato je ta metoda uporabna predvsem pri pojavih s kratkoročnimi periodami do 1 meseca, pri analizi sezonskih nihanj pa jo je možno uporabiti kot približek.

Periodični indeks I_p za posamezno obdobje $p = 1, 2, 3 \ldots P$ znotraj periode izračunamo po obrazcu:

$$I_p = \frac{S_p}{\overline{S}} \cdot 100$$

potrebni vmesni vrednosti sta določeni z:

$$S_p = \sum_{t=1}^{N} Y_{tp} \qquad \text{in} \qquad \overline{S} = \frac{\sum_{p=1}^{P} S_p}{P}$$

kjer so:

Y_{tp} – opazovane vrednosti pojava znotraj periode,

S_p – njihove vsote po posameznih obdobjih ter

$\overline{S}$ – povprečje vsot.

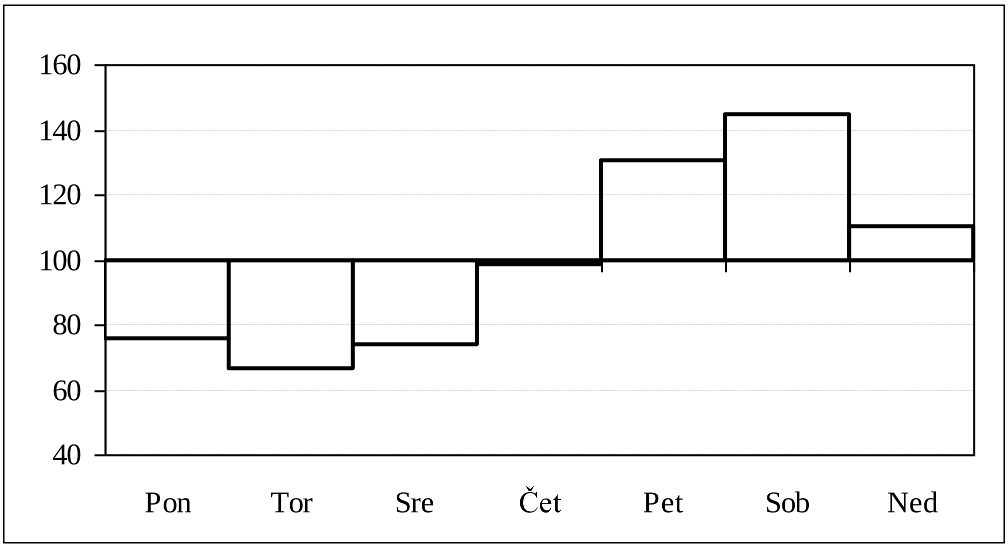
Za podatke v naslednji tabeli analizirajmo periodično nihanje števila nočitev gostov v zdravilišču po dnevih v tednu, ki smo jih opazovali šest tednov ter grafično predstavimo periodičnost opazovanega pojava.

Tabela 41: Število nočitev po dnevih v tednu ter izračunani pripadajoči periodični indeksi

	Ponedeljek	Torek	Sreda	Četrtek	Petek	Sobota	Nedelja
1. teden	73	60	78	106	125	144	110
2. teden	72	63	70	96	121	141	97
3. teden	69	61	72	84	122	139	106
4. teden	71	64	76	89	123	128	112
5. teden	73	65	63	92	128	132	98
6. teden	74	66	62	94	126	142	105
I_p	75,8	66,5	73,8	98,4	130,6	144,8	110,1

Vir: Lasten

Postopek izračuna periodičnih indeksov I_p po zgornjih obrazcih je podrobneje predstavljen v učbeniku: http://www.zavod-irc.si/docs/Skriti_dokumenti/STATISTIKA_ekon_sadl.pdf.



Slika 41: Periodični indeksi števila nočitev po dnevih v tednu
Vir: Tabela 41

Šadl (2008) navaja primer uporabe metode vsot tudi pri analizi sezonskih pojavov (za katere perioda traja leto dni), prav tako značilnih za področje turizma. Podobno kot smo v našem prejšnjem primeru periodične indekse izračunali po dnevih v tednu, se sezonski indeksi ugotavljajo npr. po četrtletjih, seveda iz večletnih zbranih podatkov preučevanega pojava. Z njihovo pomočjo se tako lahko, recimo, iz znane ocene predvidenega skupnega letnega števila turistov izračuna ocena njihovega števila po posameznih četrtletjih istega leta.

Dodatno znanje iz analize periodičnih nihanj si lahko pridobimo v učbeniku: http://www.zavod-irc.si/docs/Skriti_dokumenti/STATISTIKA_ekon_sadl.pdf.

Naša proučevanja pojavov so pokazala, da se jih pomemben del s časom spreminja. Že v 3. poglavju smo se naučili dinamiko pojavov spremljati s kazalci dinamike: indeksi, koeficienti in stopnjami rasti. V tem poglavju smo proučevanje časovnih vrst poglobili s spoznavanjem dejavnikov spreminjanja pojavov, računanjem parametrov ter iskanjem metod in postopkov, ki ob določenih pogojih omogočajo napoved nadaljnjega razvoja pojava.

Oba predstavljena modela opisa časovnih vrst (aditivni in multiplikativni model) izpostavljata štiri njihove značilne komponente, in sicer: trend ter sezonsko, ciklično in slučajno komponento časovne vrste. Za določanje prvih treh komponent, od katerih je v praksi vsekakor najpogosteje iskan trend, je potrebno v časovni vrsti najprej izločiti vpliv slučajne spremenljivke, nato pa poiskati tako črto oziroma matematično funkcijo, ki se bo najbolje prilegala podatkom proučevane časovne vrste. Uporabili smo lastnost, da lahko s korelacijo izražamo povezanost, z regresijo pa soodvisnost dveh spremenljivk. Pri linearni regresiji predpostavljamo, da se opazovanemu pojavu najbolj prilega premica, zato je v tem primeru njena enačba osnova za izračun parametrov trenda; njihov izračun pa močno poenostavimo s transformacijo časovne skale. Premica sicer v mnogo primerih ustrezno opiše smer razvoja pojava, vendar se podatkom v časovni vrsti marsikdaj bolje kot premica prilega kakšna krivulja. Uporabili in preizkusili smo možnosti iskanja nadaljnjih, optimalnejših rešitev ter pri tem ugotovili uporabniško prijazno podporo Excela. Z njim lahko pri izbiri vrste dodane trendne črte učinkovito in nazorno odločamo med več ponujenimi oblikami črt oziroma vrstami matematičnih funkcij.

S parametroma, kot sta koeficient korelacije ter determinacijski koeficient, smo ugotavljali ustreznost matematične povezave; standardni odklon vrednosti pojava od vrednosti funkcije trenda ter koeficient variacije trenda pa sta bili uporabljeni meri za ugotovitev kvalitete trenda. Poglavje smo vsebinsko zaključili s predstavitvijo pristopa k analizi periodičnih nihanj.

Naloge

Pri reševanju navedenih nalog v čim večji meri uporabljajte funkcije in orodja elektronske preglednice Excel. Dobljene rešitve interpretirajte in kritično ovrednotite.

1. Naslednja tabela prikazuje hitrost vožnje v km/h za različne izbrane hitrosti istega avtomobila in pripadajočo porabo goriva v litrih. Ugotovite soodvisnost spremenljivk z izračunom koeficienta korelacije.

Hitrost (km/h)	50	60	70	80	90	100	110	120	130
Gorivo (l)	5,3	5,9	6,5	6,7	7,0	7,1	7,6	8,2	9,0

2. Izračunajte koeficient koleracije, ki podaja odvisnost med višinama očetov in odraslih sinov, za podatke v spodnji tabeli:

Višina očetov (cm)	190	165	172	187	196	171	184	192	176
Višina sinov (cm)	177	171	169	191	196	167	190	181	175

3. Ugotovite stopnjo povezanosti med številom ženskih in moških gostov v izbranem hotelu za naslednje podatke v 10-dnevnem opazovanem obdobju:

Število žensk	61	58	26	36	38	46	53	59	44	41	26
Število moških	50	45	14	29	37	41	48	51	46	39	38

4. Naslednja tabela podaja statistične podatke o letnem številu prenočitev tujih turistov v obdobju od 1997 do 2006. Izračunani trend grafično predstavite skupaj z dejanskimi vrednostmi ter napovejte število prenočitev v letu 2007 in ga primerjajte z doseženim!

Leto	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Prenočitve	3.078	3.062	2.741	3.404	3.813	4.021	4.175	4.363	4.399	4.489

Vir: SURS, Statistični letopis 2008; http://www.stat.si/letopis/2008/25_08/25-02-08.htm

5. Tabela 22 v poglavju 3 podaja statistične podatke o številu prvič registriranih osebnih avtomobilov v RS v obdobju od 1997 do 2006. Izračunajte linearni trend ter napovejte število registriranih avtomobilov v letu 2007 in ga primerjajte z dejansko doseženim. Izračunajte standardni odklon in koeficient variacije trenda. (Pomoč: http://www.stat.si/).

6. V tabeli 33 v poglavju 5 so predstavljeni statistični podatki o številu študentov, vpisanih v višje šole v študijskih letih od 2002/03 do 2006/07. Izračunajte trend ter napovejte število študentov v preteklem študijskem letu ter dobljeno vrednost primerjajte z dejansko doseženo. Poiščite, primerjajte in preizkusite z Excelom najprimernejšo obliko trendne črte. (Pomoč: http://www.stat.si/).

9 LITERATURA

1. Čibej, J. A. *MATEMATIKA. Kombinatorika, verjetnostni račun, statistika.* Ljubljana: DZS, ISBN: 86-341-xxxx-0, 1994.

2. Devjak, S. *STATISTIKA. Analiza časovnih vrst, trend* (online). 2009. (citirano 4. 1. 2009). Dostopno na naslovu: http://www.fu.uni-lj.si/personal/zigaa/Bolonja-visokosolski/STAT_TREND_SPLET.pdf.

3. Devjak, S. *STATISTIKA. Kvantili* (online). 2008. (citirano 28. 12. 2008). Dostopno na naslovu: http://www.fu.uni-lj.si/personal/zigaa/Bolonja-visokosolski/STAT_Kvant_SPLET.pdf.

4. Eurostat. *Evropski statistični podatki* (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu: http://epp.eurostat.ec.europa.eu.

5. Ferligoj, A. *Osnove statistike na prosojnicah* (online). Ljubljana, 1995. (citirano 18. 12. 2008.). Dostopno na naslovu: http://shrani.si/f/I/hJ/3yOve8FC/osaf.pdf.

6. Gams, A. *Delo z elektronsko preglednico Excel.* Velenje: ŠC Velenje, 2006.

7. Gregorc, F. *Učbenik: Poslovna informatika s statistiko.* Višješolski strokovni program: Gostinstvo in turizem. Ministrstvo za šolstvo in šport (online). 2009. (citirano 1. 2. 2009). Dostopno na naslovu: http://www.zavod-irc.si/docs/Skriti_dokumenti/Gregorc_Poslovna_informatika_s_statistiko.pdf.

8. Knežević, M. *Statistika z uporabo računalnika.* Portorož: Turistica, ISBN: 961-6496-18-5, 2006.

9. Korenjak Černe, S. *Statistika 1 – VPŠ, 3. del: Urejanje podatkov, frekvenčne porazdelitve, ranžirna vrsta, kvantili* (online). 2008. (citirano 18. 12. 2008). Dostopno na naslovu: http://miha.ef.uni-lj.si/_dokumenti3plus2/190109/3-UrejanjeKvantili-VPS07.pdf.

10. Korenjak Černe, S. *Statistika 1 – VPŠ: Srednje vrednosti* (online). 2009 a. (citirano 19. 1. 2009). Dostopno na naslovu: http://miha.ef.uni-lj.si/_dokumenti3plus2/190109/4-Srednjevrednosti-VPS07-1.pdf.

11. Korenjak Černe, S. *Statistika 1 – VPŠ: Mere variabilnosti*, *normalna porazdelitev.* (online). 2009 b. (citirano 21. 1. 2009). Dostopno na naslovu: http://miha.ef.uni-lj.si/_dokumenti3plus2/190109/7-MereVariabNP-VPS07-3.pdf.

12. Košmelj, B. *Statistika.* Ljubljana: DZS, ISBN: 86-341-1470-8, 1996.

13. Košmelj, K. *Časovne vrste* (online). 2008. (citirano 26. 12. 2008). Dostopno na naslovu: http://les.bf.uni-lj.si/uploads/media/casovne_vrste.pdf.

14. Kovačič, M. On-line slovarček statističnih pojmov. OSNOVNA STATISTIČNA ANALIZA (online). 2009. (citirano 2. 1. 2009). Dostopno na naslovu: http://www.ljudmila.org/matej/statistika/mva.html.

15. Leskošek, B. *Statistika: Osnovni pojmi, Urejanje in grafično prikazovanje podatkov, Mere srednjih vrednosti in razpršenosti* (online). 2008 a. (citirano 18. 12. 2008). Dostopno na naslovu: http://www.fsp.uni-lj.si/AKIS/Statistika/OsnovniPojmi.pdf.

16. Leskošek, B. *Rangi in kvantili* (online). 2008 b. (citirano 22. 12. 2008). Dostopno na naslovu: http://www.fsp.uni-lj.si/bojan/Statistika/2007/RangiKvantili.pdf.

17. Leskošek, B. *Regresija* (online). 2009. (citirano 3. 1. 2009). Dostopno na naslovu: http://www.fsp.uni-lj.si/Statistika/2007/Regresija.pdf.

18. Leskošek, B. *Členitev časovne vrste* (online). 2008. (citirano 26. 12. 2008). Dostopno na naslovu: http://www.fsp.uni-lj.si/Statistika/2006/CasovneVrste.pdf.

19. Ljubič, T. *Planiranje in vodenje proizvodnje: metode, modeli, tehnike. Napovedovanje - stohastično planiranje* (online). 2008. (citirano 27. 12. 2008). Dostopno na naslovu: http://www1.fov.uni-mb.si/ljubic/images/Mtp08_Stohasticno_planiranje.pdf.

20. Microsoft Corporation. *Pomoč in nasveti za Microsoft Office Excel 2003.* (online). 2008. (citirano 24. 12. 2008). Dostopno na naslovu: http://office.microsoft.com/sl-si/excel/FX100646961060.aspx?CTT=96&Origin=CL100570551060.

21. Ostan, I. *PISTA: Praktična Interaktivna STAtistika za turistične delavce* (online). 2008. (citirano 24. 12. 2008). Dostopno na naslovu: http://www3.turistica.si/kribel/articles/pista/.

22. Pelhan, N. *Statistika in verjetnost* (online). 2008. (citirano 28. 12. 2008). Dostopno na naslovu: http://fov.pelhan.net/datoteke/1_Letnik/Statistika_in_verjetnost/STVAJEN.doc.

23. RIC, Državni izpitni center. *Letno maturitetno poročilo o poklicni maturi 2007* (online). 2008. (citirano 4. 12. 2008). Dostopno na naslovu: http://www.ric.si/mma_bin.php/$fileI/2008052807112324/$fileN/LP-PM-2007.pdf.

24. STO, Slovenska turistična organizacija. *Statistični podatki* (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu: http://www.slovenia.info/si/Statistični-podatki.htm?ppg_statisticni_podatki2005=0&lng=1.

25. SURS, Statistični urad Republike Slovenije. *Državna statistika* (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu: http://www.stat.si/.
Popis 2002 (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu:
http://www.stat.si/popis2002/si/,
http://www.stat.si/popis2002/si/rezultati_slovenija_prebivalstvo_dz.htm,
http://www.stat.si/popis2002/si/default.htm,
http://www.stat.si/popis2002/si/rezultati/rezultati_red.asp?ter=SLO&st=103.
Statistični letopis 2001 (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu:
http://www.stat.si/letopis/2001/27_01/27-17-01.xls.
Statistični letopis 2002 (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu:
http://www.stat.si/letopis/index_vsebina.asp?poglavje=22&leto=2002&jezik=si.
Statistični letopis 2003 (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu:
http://www.stat.si/letopis/2003/22_03/22-18-03.xls.
Statistični letopis 2007 (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu:
http://www.stat.si/letopis/2007/06_07/06-14-07.xls.

Statistični letopis 2008 (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu:
http://www.stat.si/letopis/,
http://www.stat.si/letopis/2008/04_08/04-02-08.htm,
http://www.stat.si/letopis/2008/25_08/25-09-08.htm
http://www.stat.si/letopis/2008/06_08/06-14-08.htm,
http://www.stat.si/letopis/2008/25_08/25-17-08.htm,
http://www.stat.si/letopis/2008/25_08/25-10-08.htm,
http://www.stat.si/letopis/2008/06_08/06-23-08.htm,
http://www.stat.si/letopis/2008/25_08/25-05-08.htm,
http://www.stat.si/letopis/2008/12_08/12-05-08.htm,
http://www.stat.si/letopis/2008/06_08/06-05-08.htm,
http://www.stat.si/letopis/2008/33_08/33-22-08.htm,
http://www.stat.si/letopis/2008/21_08/21-23-08.htm,
http://www.stat.si/letopis/2008/21_08/21-18-08.htm,
http://www.stat.si/letopis/2008/25_08/25-06-08.htm,
http://www.stat.si/letopis/2008/02_08/02-03-08.htm,
http://www.stat.si/letopis/2008/31_08/31-01-08.htm.

SI-STAT. Podatkovni portal. (online). 2009. (citirano 12. 2. 2009). Dostopno na naslovu:
http://www.stat.si/pxweb/dialog/statfile2.asp,
http://www.stat.si/pxweb/Dialog/Saveshow.asp,
http://www.stat.si/pxweb/Dialog/varval.asp?ma=2118102S&ti=Prihodi+in+preno%E8itve+turistov+po+ob%E8inah%2C+dr%9Eavah+in+vrstah+nastanitvenih+objektov%2C+Slovenija%2C+letno&path=../Database/Ekonomsko/21_gostinstvo_turizem/01_21181_nastanitev_obcine/&lang=2.

26. Šadl, M. *Statistika.* Celovec – Ljubljana – Dunaj: Mohorjeva, ISBN: 3-7086-0084-3, 2004.

27. Šadl, M. *Učbenik: Statistika.* Višješolski strokovni program: Ekonomist. Ministrstvo za šolstvo in šport (online). 2008. (citirano 21. 12. 2008). Dostopno na naslovu: http://www.zavod-irc.si/docs/Skriti_dokumenti/STATISTIKA_ekon_sadl.pdf.

28. Terminološki slovar. *Členitev časovne vrste* (online). 2008.
(citirano 26. 12. 2008). Dostopno na naslovu:
http://www.najdi.si/trident/quickpreview.jsp?q=%22%C4%8Dasovne+vrste%22&hpage=web&qpts=12&rn=66183649.

29. Trstenjak, I. *Statistika. Delovni zvezek z osnovami teorije.* Ljubljana: Modrijan, ISBN 961-6357-50-6, 2001.

30. Uporabljene ikone in cliparti (online). 2009. (citirano 20. 2. 2009). Dostopno na naslovu:
http://www.iconsxtras.com/iconsxtras-free-icons-details.asp?prodid=29968,
http://www.themexp.org/preview.php?mid=41805&type=icon&view=views&page=&cat=&n_ame=xToon.zip&proddesc=Icons&catdesc=Designer&namedesc=xToon,
http://www.wpclipart.com/education/.

PRILOGA

V učbeniku so bile za izračune predstavljenih primerov in nalog uporabljene naslednje Excelove statistične funkcije:

AVERAGE(number1;number2;...)	Vrne aritmetično srednjo vrednost argumentov.
CORREL(array1;array2)	Vrne korelacijski koeficient med dvema naboroma podatkov.
COVAR(array1;array2)	Vrne kovarianco, povprečje produktov dvojic odklonov za vsak podatkovni par v dveh množicah podatkov.
FORECAST(x;known_y's;known_x's)	Z uporabo linearne regresije izračuna ali napove bodočo vrednost spremenljivke.
GEOMEAN(number1;number2;...)	Vrne geometrično srednjo vrednost za matriko ali obseg pozitivnih podatkov.
HARMEAN(number1;number2;...)	Vrne harmonično srednjo vrednost za množico podatkov.
LINEST(known_y's;known_x's;const;stats)	Izračuna parametre linearnega trenda z uporabo metode najmanjših kvadratov.
MAX(number1;number2;...)	Vrne največjo vrednost v množici vrednosti.
MEDIAN(number1;number2;...)	Iz množice števil vrne mediano.
MIN(number1;number2;...)	Vrne najmanjše število v množici vrednosti.
MODE(number1;number2;...)	Iz množice števil vrne modus.
NORMDIST(x;mean;standard_dev;cumulative)	Vrne normalno porazdelitev za podano srednjo vrednost in standardni odklon.
PEARSON(array1;array2)	Vrne Pearsonov korelacijski koeficient.
QUARTILE(array;quart)	Iz množice podatkov vrne kvartil.
RANK(number;ref;order)	Vrne rang števila številske spremenljivke.
SLOPE(known_y's;known_x's)	Vrne naklon regresijske premice.
STDEVP(number1;number2;...)	Izračuna standardni odklon na temelju celotne populacije.
TREND(known_y's;known_x's;new_x's;const)	Vrne vrednost vzdolž linearnega trenda.
VARP(number1;number2;...)	Izračuna varianco, ki temelji na celotni populaciji.

Več in podrobno na Microsoft Office Online
http://office.microsoft.com/sl-si/excel/CH062528311060.aspx

Projekt **Impletum**
Uvajanje novih izobraževalnih programov na področju višjega strokovnega izobraževanja v obdobju 2008–11

Konzorcijski partnerji:

Operacijo delno financira Evropska unija iz Evropskega socialnega sklada ter Ministrstvo RS za šolstvo in šport. Operacija se izvaja v okviru Operativnega programa razvoja človeških virov za obdobje 2007–2013, razvojne prioritete Razvoj človeških virov in vseživljenjskega učenja in prednostne usmeritve Izboljšanje kakovosti in učinkovitosti sistemov izobraževanja in usposabljanja.