

Damijana Kastelec in Katarina Košmelj

STATISTIČNA ANALIZA PODATKOV
S PROGRAMOMA EXCEL 2003 IN R

Oktober 2009

Recenzenta: prof. dr. Janez Stare in prof. dr. Andrej Blejec

Lektorica: Jasmina Antonijević

Oblikovanje besedila in slik: dr. Damijana Kastelec

Založnik: Biotehniška fakulteta, Univerza v Ljubljani

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

004.42:311.1(075.8)

KASTELEC, Damijana

Statistična analiza podatkov s programoma Excel 2003 in R
[Elektronski vir] / Damijana Kastelec in Katarina Košmelj. - El.
knjiga. - Ljubljana : Biotehniška fakulteta, 2009

Način dostopa (URL): <http://www.bf.uni-lj.si/agronomija/o-oddelku/katedre-in-druge-org-enote/za-genetiko-biotehnologijo-statistiko-in-zlahtnjenje-rastlin/statistika/sap-2003.html>

ISBN 978-961-6379-14-4

1. Košmelj, Katarina

247854592

Copyright ©Damijana Kastelec in Katarina Košmelj, 2009

PREDGOVOR

Knjiga je zasnovana kot učbenik, vezan na pouk statistike na različnih študijskih programih. Delo je nastajalo več let, se nadgrajevalo in dopolnjevalo. Namen učbenika je pridobitev osnovnih računalniških znanj, ki so vezana na pouk statistike in na delo z enostavnimi podatkovnimi zbirkami. Statistična vsebina obsega poglavja iz osnovne statistike: opisna statistika, osnove statističnega sklepanja, verjetnostne porazdelitve, intervali zaupanja ter preverjanje statističnih domnev o povprečju in Bernoullijevi verjetnosti, enostavno linearno regresijsko analizo, korelacijo ter nekatere hi-kvadrat preizkuse. Poseben poudarek je na ustreznih grafičnih prikazih podatkov in rezultatov.

Računalniška analiza temelji na uporabi programa MS Excel 2003 ter programa R. Prvega sva izbrali zaradi splošne uporabnosti in dostopnosti študentom, drugega pa zaradi dejstva, da je brezplačno dostopen na Internetu in omogoča resno statistično analizo. Ker sodiva, da je uporaba programa R v osnovni obliki prezahtevna za naše študente, za analizo uporabljamo grafični vmesnik R Commander v slovenskem jeziku, ki omogoča podoben način dela kot program Excel, poleg tega pa uporabniku omogoča relativno enostaven preskok na pisanje ukazov v ukazne vrstice programa R.

Knjiga ni namenjena branju, temveč aktivnemu reševanju nalog z uporabo računalnika. Besedilo vsebuje opise programov in njihovih funkcij ter orodij, nato sledijo naloge za utrjevanje snovi. Naloge si sledijo po principu »od lažjega k težjemu«, vendar zahtevajo aktivno delo »bralca«, saj bo le tako knjiga dosegla svoj namen. V večini primerov so predstavljene tudi rešitve nalog.

Avtorici

KAZALO

1	Grafični prikazi in računanje.....	1
1.1	Uvod.....	1
1.2	Datoteka v Excelu - zvezek.....	3
1.3	Preglednica in celica v preglednici.....	4
1.3.1	Vnos podatkov v celico.....	4
1.3.2	Določanje višine vrstic in širine stolpcev.....	4
1.3.3	Oblika vsebine celice.....	5
1.3.4	Spreminjanje vsebine v celicah.....	7
1.4	Označene celice.....	7
1.4.1	Označevanje.....	7
1.4.2	Kopiranje.....	8
1.4.3	Brisanje.....	8
1.4.4	Vstavljanje.....	9
1.5	Grafično prikazovanje podatkov.....	9
1.5.1	Čarovnik za grafikone.....	11
1.5.2	Spreminjanje grafikona.....	12
1.6	Formule.....	17
1.6.1	Kopiranje formul.....	18
1.6.2	Vstavljanje funkcij.....	19
1.6.3	Kopiranje na poseben način.....	21
2	Relativna števila in grafični prikazi.....	26
2.1	Oblikovanje osi grafikona.....	26
2.2	Oblikovanje niza vrednosti odvisne spremenljivke.....	32
2.3	Časovne vrste, indeksi, stopnje rasti in gostote ter njihovi grafični prikazi.....	35
2.4	Koeficienti in njihovi grafični prikazi, povprečni koeficient.....	46
3	Opisne statistike, slučajna izbira in frekvenčna porazdelitev.....	51
3.1	Uporaba vgrajenih funkcij v Excelu.....	51
3.1.1	Geometrijska sredina.....	52
3.2	Napake pri uporabi formul.....	53
3.3	Opisne statistike.....	54

3.3.1	Uporaba funkcij	54
3.3.2	Uporaba orodja Descriptive Statistics	56
3.4	Slučajna izbira	60
3.5	Frekvenčna porazdelitev in kvantili	63
4	Podatkovna zbirka in vrtilne tabele.....	73
4.1	Podatkovna zbirka	73
4.2	Urejanje zapisov	75
4.3	Izbira zapisov	76
4.3.1	Samodejni filter	76
4.3.2	Napredni filter	79
4.4	Kontrola podatkov	81
4.5	Vrtilne tabele	81
4.5.1	Čarovnik za vrtilne tabele.....	83
4.5.2	Dodatno oblikovanje vrtilnih tabel.....	87
4.5.3	Orodna vrstica Vrtilna tabela	89
4.5.4	Seznam polj vrtilne tabele	97
4.5.5	Nastavitve polj vrtilne tabele.....	98
4.6	Vrtilni grafikon.....	99
5	Opisne statistike in grafični prikazi s programom R.....	112
5.1	Program R in grafični vmesnik R Commander	112
5.1.1	Kako naložiti program R?	112
5.1.2	Kako naložiti R Commander?	116
5.1.3	Vnos podatkov.....	120
5.2	Opisne statistike in grafični prikazi.....	125
5.2.1	Številski povzetki	125
5.2.2	Okvir z ročaji.....	126
5.2.3	Histogram	129
5.2.4	Frekvenčna porazdelitev in stolpični diagram.....	130
5.3	Slučajno vzorčenje v R.....	133
5.4	Slučajna izbira enot v načrtovanju poskusa	135
6	Normalna in binomska porazdelitev z Excelom	139
6.1	Normalna porazdelitev, funkciji NORMDIST in NORMINV	139

6.1.1	Gostota verjetnosti.....	140
6.1.2	Izračun verjetnosti	142
6.1.3	Izračun kvantilov	143
6.2	Generator slučajnih števil v Excelu.....	146
6.2.1	Generiranje vrednosti, porazdeljenih po normalni porazdelitvi	146
6.3	Binomska porazdelitev, funkcija BINOMDIST	148
6.4	Aproksimacija binomske porazdelitve z normalno porazdelitvijo.....	151
7	Verjetnostne porazdelitve s programom R.....	154
7.1	Porazdelitev vzorčnih aritmetičnih sredin.....	154
7.2	Porazdelitev vzorčnih varianc	162
7.3	Studentova porazdelitev (t-porazdelitev)	166
7.4	Binomska porazdelitev	170
8	Intervali zaupanja	173
8.1	Interval zaupanja za povprečje	173
8.1.1	σ je znana.....	173
8.1.2	σ ni znana.....	175
8.2	Velikost vzorca za povprečje pri normalni porazdelitvi	183
8.2.1	Standardni odklon je znan	183
8.2.2	Standardni odklon ni znan – iterativni postopek	184
8.3	Interval zaupanja in velikost vzorca za p pri binomski porazdelitvi $b(n, p)$...	186
9	Preverjanje statističnih domnev za en vzorec.....	189
9.1	Povprečje	189
9.2	Bernoullijeva verjetnost	197
10	Preverjanje statističnih domnev za dva vzorca in interval zaupanja za razliko	200
10.1	Razlika povprečij za neodvisna vzorca	200
10.2	Razlika povprečij za odvisna vzorca	212
10.3	Razlika Bernoullijevih verjetnosti za neodvisna vzorca	219
10.4	Razlika Bernoullijevih verjetnosti za odvisna vzorca	223
11	Linearna regresija in korelacija.....	226
11.1	Linearna regresija	226
11.1.1	Excelovo orodje za linearno regresijo	229

11.1.2	Linearna regresija s programom R	231
11.2	Korelacija	240
11.2.1	Pearsonov in Spearmanov koeficient korelacije v Excelu	240
11.2.2	Pearsonov in Spearmanov koeficient korelacije v programu R	245
12	Hi-kvadrat preizkusi	251
12.1	Hi-kvadrat preizkus za preverjanje domneve o porazdelitvi	251
12.2	Hi-kvadrat za preverjanje povezanosti opisnih spremenljivk	261
12.3	Hi-kvadrat za preverjanje homogenosti struktur	265
13	DODATEK.....	269
13.1	Uporaba kalkulatorja za izračun povprečja in standardnega odklona	269
13.2	Uporaba kalkulatorja za linearno regresijo in korelacijo	270

1 Grafični prikazi in računanje

Novi pojmi:

- datoteka v Excelu - zvezek
- vnos podatkov
- grafično prikazovanje podatkov
- formule - osnove

Datoteke s podatki:

v tem poglavju jih ni

Ustvarjene datoteke:

PREHODNOST.XLS

OKRASNE RASTLINE.XLS

VPIS.XLS

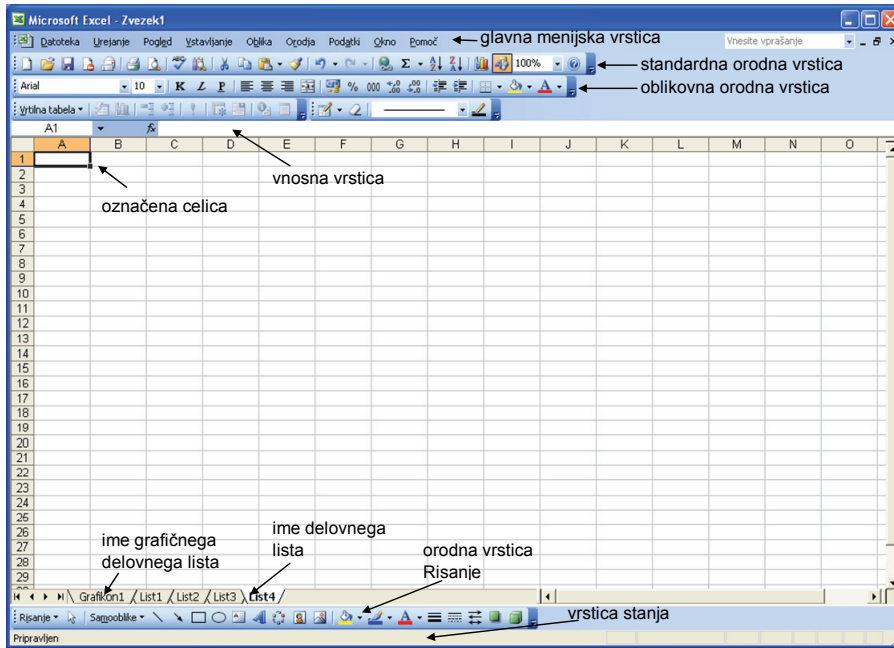
1.1 Uvod

Program Microsoft Excel je namenjen vnosu, grafičnemu prikazovanju in analizam različnih vrst podatkov. Excel omogoča enostavno delo tudi pri zelo kompleksnih analizah podatkov. Tabele s podatki in slike lahko brez težav prenesemo v datoteke urejevalnika besedil Word in obratno.

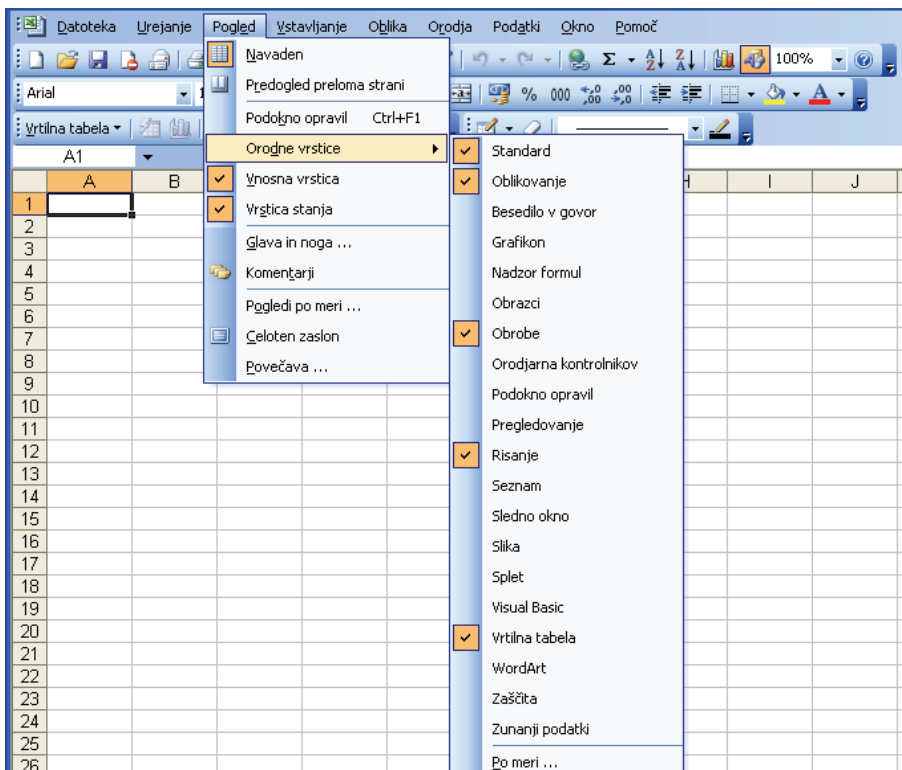
Okno Microsoft Excel pri osnovni nastavitvi vsebuje:

- glavno menijsko vrstico z meniji *Datoteka*, *Urejanje*, *Pogled*, *Vstavljanje*, *Oblika*, *Orodja*, *Podatki*, *Okno* in *Pomoč*;
- standardno orodno vrstico z ikonami za delo z datotekami in označenimi deli besedila ter orodja za uporabo funkcij, risanje grafov itd.;
- oblikovno orodno vrstico z ikonami za oblikovanje besedila;
- vnosno vrstico, ki je pod zadnjo orodno vrstico;
- vrstico stanja, ki je zadnja vrstica v oknu Excela in podaja informacijo o operaciji, ki se trenutno izvaja (npr. Pripravljen).

1 Grafični prikazi in računanje



Če katere od zelenih orodnih vrstic ni na zaslonu, uporabimo ukaz **Pogled/Orodne vrstice**.




1.2 Datoteka v Excelu - zvezek

Delo z datotekami poteka enako kot v Wordu (meni **Datoteka**, ukazi **Nova**, **Odpri**, **Zapri**, **Shrani**, **Shrani kot**). Datoteke, napisane v Excelu, imajo končnico .XLS.

PREHODNOST 1

Shranite datoteko *Zvezek 1* z imenom PREHODNOST.XLS. ♦

Datoteka v Excelu je organizirana kot zvezek, ki ima na poseben način oblikovane delovne liste. Z enega delovnega lista na drugega se premikamo s klikom jezička delovnega lista (*List1*, *List2*, ...) v predzadnji vrstici na spodnjem robu okna¹. Delovne liste lahko poljubno dodajamo, jih preimenujemo, premikamo, kopiramo ali brišemo z ukazi v meniju, ki se pokaže z desnim klikom jezička delovnega lista. Delovni list ima lahko v osnovi dve obliki: **preglednico** (vnaprej določena imena so *List1*, *List2*, ...) ali **graf** (vnaprej določena imena *Grafikon1*, *Grafikon2*, ...).

Excel omogoča preklic zadnjih nekaj zaporednih ukazov z ukazom **Urejanje/Razveljavi** ali s klikom ikone **Razveljavi** . Nekaterih ukazov ne moremo razveljaviti.

Excel omogoča delo z več datotekami (okni) hkrati. Seznam odprtih datotek je v meniju **Okno** ali v opravilni vrstici okolja Okna na dnu zaslona.

Ko se z miško premikamo po preglednici, ima lahko kazalec miške tri različne oblike:

- bel križec nadomešča kazalec, ko se z miško premikamo po preglednici ali označujemo posamezne dele preglednice;
- črni križec se pojavi, ko se z miško postavimo v desni spodnji vogal označenih celic v preglednici;
- črni križec s puščicami na koncih se pojavi, ko se z miško postavimo na rob označenih celic v preglednici.

¹ Če te vrstice ni, izberemo v ukazu **Orodja/Možnosti** na kartici *Pogled* v polju **Možnosti okna** možnost **Jezički listov**.

1.3 Preglednica in celica v preglednici

Preglednica je tabela, sestavljena iz velikega števila celic, ki jih določajo stolpci in vrstice. Stolpci so označeni s črkami, vrstice pa s števkami. Oznake stolpcev in vrstic stojijo v **naslovni vrstici** in v **naslovnem stolpcu**. Vsaka celica v preglednici ima svoj naslov: celica A1 stoji v prvem stolpcu in v prvi vrstici. Celice so namenjene vnosu podatkov. Med celicami se premikamo s tipkami na tipkovnici (←, ↑, →, ↓, Home, End, PgUp, PgDn) ali pa z miško. **Celica je označena**, kadar so njeni robovi odebeljeni. Označimo jo s klikom. Če pa v celici hkrati vidimo tudi utripajoči kazalec, je **celica aktivna**. Celico aktiviramo z dvoklikom.

1.3.1 Vnos podatkov v celico

Podatek vnesemo v celico tako, da jo označimo, natipkamo njeno vsebino in vnos potrdimo s pritiskom na tipko ENTER ali pa celico kakorkoli zapustimo. Če želimo vnos že natipkanega podatka preprečiti, pritisnemo na tipko ESC. V celice preglednice lahko vnašamo konstante ali formule. Konstante so: števila, datumi, časi, denarne vrednosti, odstotki, besedilo. O formulah bomo več povedali v naslednjem poglavju. Način vnosa števil, datumov, časa in denarnih vrednosti je odvisen od **Področnih nastavitev** v okolju *Okna*.²

1.3.2 Določanje višine vrstic in širine stolpcev

Če je vsebina v celici daljša od širine celice, je v celoti vidna samo, če je desna sosednja celica prazna. Drugače vidimo le del vsebine, ki ga določa širina celice. Če je celica označena ali aktivirana, vidimo celotno vsebino celice v vrstici za vnos.

Če v celico vnesemo številsko vrednost in se po kliku gumba ENTER vsebina celice spremeni v #####, to pomeni, da je celica preozka za izpis in jo je potrebno razširiti.

- Višino vrstic in širino stolpcev spreminjamo tako, da se s križcem miške postavimo na mejo med dvema vrsticama oz. stolpcema v naslovni vrstici oz. stolpcu. Ko se

² Program Nadzorna plošča (angl. Control Panel), podprogram Področne in jezikovne nastavitve (angl. Regional and Language Options), v pogovornem oknu kliknemo gumb Nastavi (angl. Customize).

križec spremeni v dvosmerno puščico, s potegom miške prestavimo mejo vrstice/stolpca.

- Kadar je v celici že neka vsebina, potrebno širino stolpca določimo z dvoklikom meje med stolpcem in njegovim desnim sosedom v naslovni vrstici.
- Višino vrstice oz. širino stolpca lahko spremenimo še z ukazom **Oblika/Vrstica**, **Višina** ali **Oblika/Stolpec**, **Širina**. Pred tem moramo označiti tiste vrstice ali stolpce, ki jim želimo prirediti spremembe.

Pri vnosu podatkov se številske vrednosti v osnovnem formatu poravnajo desno, besedilo pa se poravna levo. Tak način poravnave omogoča hitro kontrolo vrste vnesenih podatkov (število, besedilo). Če odtipkamo napačen znak za decimalno ločilo, Excel vneseni podatek razume kot besedilo in ne kot število³.

PREHODNOST 2

V datoteko PREHODNOST.XLS vnesite podatke iz tabele 1.1. Širino stolpcev nastavite tako, da bo vidna celotna vsebina celic. Delovni list poimenujte »Podatki«. ♦

Tabela 1.1: Prehodnost generacije študentov 1. letnika v šolskem letu 2002/2003 v 2. letnik na univerzitetnih študijskih programih Biotehniške fakultete (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003)

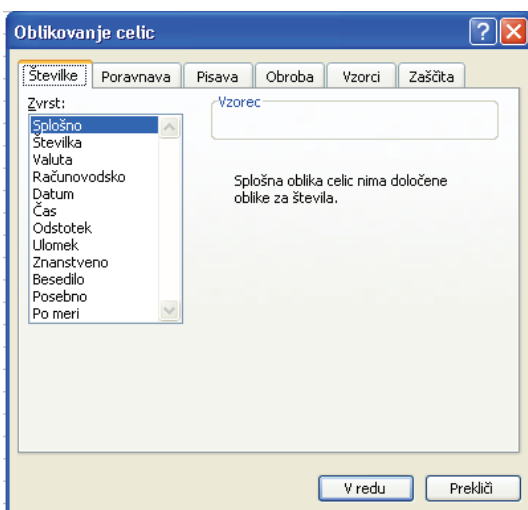
Študijski program	Napredovali v 2. letnik	Ponovni vpis v 1. letnik	Neznano
Kmetijstvo-agronomija	46	21	18
Biologija	52	13	9
Krajinska arhitektura	27	2	3
Lesarstvo	14	16	38
Mikrobiologija	44	20	2
Kmetijstvo-zootehnika	22	21	14
Živilska tehnologija	53	11	4

1.3.3 Oblika vsebine celice

Vsebini vsake celice določimo oblikovne parametre z možnostmi, ki jih nudi ukaz **Oblika/Celice**. Če hočemo obliko določiti več celicam naenkrat, jih označimo, lahko pa uporabimo tudi ukaza **Oblika/Vrstica** ali **Oblika/Stolpec**, ki dodelita oblikovne parametre celim vrsticam oz. stolpcem.

³ Kot decimalno ločilo po slovenskem pravopisu uporabljamo vejico. Drugačne nastavitve za ločila, oblike datumov idr. naredimo v okolju *Okna, Start/Nadzorna plošča, Področne nastavitve*.

Z ukazom **Oblika/Celice** se odpre pogovorno okno, ki vsebuje šest kartic:



- **Številke** - na tej kartici v polju **Zvrst** določimo željeno zvrst podatkov: **Številka**, **Datum**, **Čas**, **Odstotek**, **Ulomek**, **Znanstveno**, **Valuta**, **Besedilo** idr. Poleg že določenih oblik lahko uporabimo tudi take, ki jih določimo sami ob izbiri zvrsti **Po meri**.

Če želimo v celico vpisati število v obliki besedila, npr. letnico, moramo na začetek števila napisati apostrof (') ali pa celici pred vnosom podatka dodeliti zvrst **Besedilo**;

- **Poravnava** - poravnava besedila v celici;
- **Pisava** - oblika, velikost, tip znakov;
- **Obroba** - risanje obrob in črt v tabelah;
- **Vzorci** - senčenje, vzorci in barve v celicah;
- **Zaščita** - zaščita vsebine celice.

PREHODNOST 3

Uporabite ukaz **Oblika/Celice** in preglejte vsebino kartic. Katere oblikovne parametre lahko določimo na posamezni kartici?

Besedilo v prvi vrstici tabele 1.1 naj bo izpisano krepko, ozadje celic naj bo sivo. Določite, naj se besedilo v celicah prve vrstice izpiše v več vrstic, če je stolpec preozek.

V tabeli naj bodo vse celice obrobljene s črto. ♦

1.3.4 Spreminjanje vsebine v celicah

Vsebino celic lahko popravljamo na tri načine:

- ponoven vnos vsebine celice. Ta način uporabimo, če je vsebina celice kratka. Označimo celico, ponovno napišemo vsebino in potrdimo vnos s pritiskom na tipko ENTER. Pri tem nova vsebina celice nadomesti staro;
- spreminjanje v vrstici za vnos. Ta način uporabimo, če želimo popraviti samo del vsebine. Označimo celico in postavimo utripajoči kazalec v vrstico za vnos podatkov. V tej vrstici lahko popravljamo kot v urejevalniku besedila (del besedila lahko označimo in ga kopiramo ali premeščamo, posamezne znake pobrišemo ali pripišemo, spreminjamo oblikovne parametre, itd.). Vnos zaključimo s tipko ENTER;
- spreminjanje v aktivni celici. Celico aktiviramo z dvoklikom. Pri tem se v celici pokaže celotna vsebina celice in utripajoči kazalec. Besedilo v celici spremenimo.

1.4 Označene celice

1.4.1 Označevanje

Eno celico ali skupino celic lahko označimo s potegom miške, pri tem se označene celice pobarvajo svetlo modro, samo celica, s katero smo začeli označevanje, pa ostane bela. V Excelu so označene skupine celic poimenovane na osnovi leve zgornje in desne spodnje celice. Primer: označen del preglednice, ki vsebuje celice A1, A2, A3, B1, B2, B3, C1, C2, C3, se imenuje A1:C3.

PREHODNOST 4

Označite tabelo 1.1, ki ste jo vpisali v nalogi PREHODNOST 2. ♦

Celo vrstico oz. stolpec označimo s klikom na polje naslova vrstice oz. stolpca. Naenkrat lahko označimo tudi celice, ki se ne stikajo. Najprej označimo prvo skupino celic, pritisnemo tipko CTRL in jo držimo, medtem ko označujemo še preostale skupine celic.

1.4.2 Kopiranje

Z ukazi **Urejanje/Izreži**, **Urejanje/Kopiraj**, **Urejanje/Prilepi** ali pa z uporabo ikon v standardni orodni vrstici označene celice kopiramo ali predstavljamo po preglednici, med delovnimi listi zvezka in med različnimi zvezki. Ko uporabimo ukaza **Urejanje/Izreži** ali **Urejanje/Kopiraj**, se okoli označenih celic pojavi črtkana utripajoča črta. Ta črta je prisotna, dokler je vsebina označenih celic v odlagališču. Mesto v preglednici, kamor želimo kopirati ali prestaviti označene celice, določimo tako, da označimo celico, ki predstavlja položaj levega zgornjega vogala prenesene vsebine odlagališča. Z ukazom **Urejanje/Prilepi** jo postavimo na izbrano mesto v preglednici.

Pri kopiranju ali prestavljanju posamične celice moramo paziti, da to celico le označimo in ne tudi aktiviramo.

PREHODNOST 5

Označite prvi in tretji stolpec v tabeli 1.1, uporabite ukaz **Urejanje/Kopiraj** in izbrane celice z ukazom **Urejanje/Prilepi** kopirajte pod tabelo 1.1 na isti delovni list. Oglejte si položaj kopiranih celic, nato pa uporabite ukaz **Razveljavi**. ♦

PREHODNOST 6

Kopirajte tabelo 1.1 na naslednji delovni list v datoteki PREHODNOST.XLS. Prestavite zadnja dva stolpca v tabeli 1.1 za en stolpec v desno in nato uporabite ukaz **Urejanje/Razveljavi**. ♦

1.4.3 Brisanje

Z ukazom **Urejanje/Počisti** v celicah pobrišemo različne dele vsebine celic: **Vse** - pobriše vse; **Oblikovanje** - pobriše oblikovne parametre v celici (meje, vzorci, osenčitve, oblika črk, ...); **Vsebina** - pobriše podatke v celici, oblikovni parametri za nove podatke pa ostanejo. Tipka DELETE deluje enako kot ukaz **Urejanje/Počisti**, **Vsebina**.

Ukaz **Urejanje/Izbriši** odstrani celice. Ko uporabimo ta ukaz, se odpre pogovorno okno z naslednjimi možnostmi: **Pomakni celice v levo** - celice desno od izbranih se pomaknejo v levo; **Pomakni celice navzgor** - celice pod izbranimi se pomaknejo navzgor; **Celotno vrstico** - odstrani označene vrstice; **Celoten stolpec** - odstrani označene stolpce.

PREHODNOST 7

Z ukazom **Urejanje/Počisti**, Vse pobrišite vsebino celic, ki vsebujejo podatke za smer študija Biologija v tabeli 1.1, nato uporabite ukaz **Urejanje/Razveljavi**. Postopek ponovite z ukazom **Urejanje/Počisti**, Vsebina in nato še z ukazom **Urejanje/Počisti**, **Oblikovanje**. Na koncu naj bo tabela nespremenjena. ♦

PREHODNOST 8

Iste celice pobrišite z ukazom **Urejanje/Izbriši**, **Premakni celice navzgor**. Nato uporabite ukaz **Urejanje/Razveljavi**, da bo na koncu ostala tabela nespremenjena. ♦

1.4.4 Vstavljanje

V preglednico lahko vstavljamo posamezne prazne celice, vrstice, stolpce ali skupine celic z ukazi v meniju **Vstavljanje**. Prazno vrstico vstavimo tako, da se v preglednici postavimo v vrstico, nad katero želimo vstaviti prazno vrstico in uporabimo ukaz **Vstavljanje/Vstavi vrstico**. S tem se vsebina celic v označeni in vseh naslednjih vrsticah premakne za eno vrstico navzdol. Hkrati lahko vrinemo tudi več vrstic, tako da označimo več vrstic, nato pa uporabimo ukaz **Vstavljanje/Vstavi vrstico**.

Stolpce, posamezne celice in skupine celic vrivamo po podobnem postopku.


PREHODNOST 9

Vrinite v tabelo 1.1 vrstico med podatke za smer študija Biologija in Krajinska arhitektura in vnesite v vrinjeno vrstico podatke za smer študija Gozdarstvo (20 vpisanih v 2. letnik, 10 ponovno vpisanih v 1. letnik in 11 neznan). ♦

PREHODNOST 10

Med prva dva stolpca v tabeli 1.1 vrinite stolpec "Kratice" in vanj vpišite kratice za smeri študija (AG, BI, GO, KA, LE, MI, ZOO, ŽT). ♦

1.5 Grafično prikazovanje podatkov

V Excelu lahko podatke grafično prikažemo na veliko različnih načinov. Pri tem nam pomaga **Čarovnik za grafikone** .

Podatke lahko prikazujemo v dveh ali treh dimenzijah. Grafikoni, ki jih prikazujemo v dvorazsežnem koordinatnem sistemu, vsebujejo dve osi: vodoravno os X in navpično os

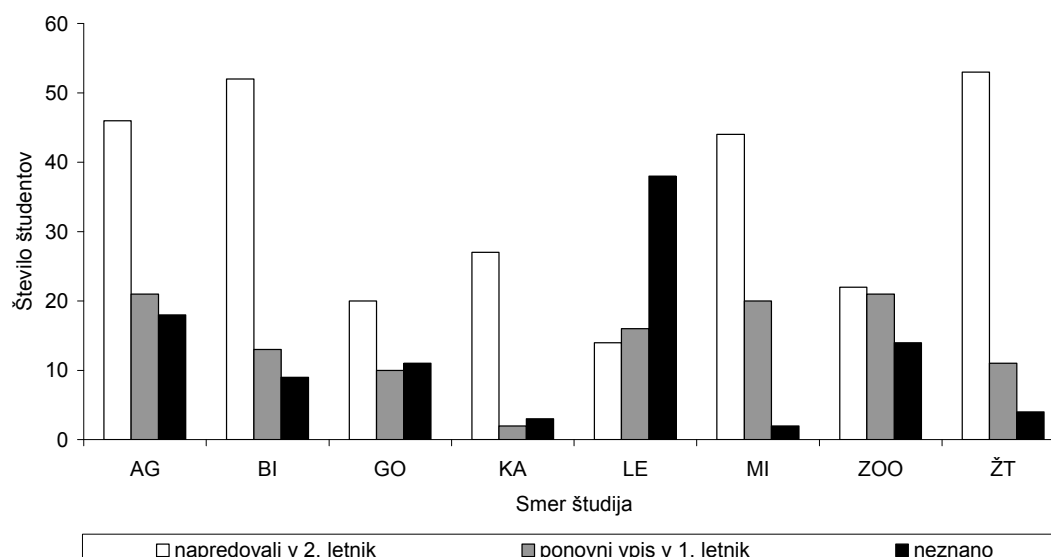
Y. Na osi X prikazujemo neodvisno spremenljivko, na osi Y pa odvisno. Spremenljivka na osi X je lahko številska ali opisna (kategorična). V primeru, da na istem grafikonu prikazujemo dve različni odvisni spremenljivki pri isti neodvisni spremenljivki, lahko uporabimo dve osi Y: primarno na levi strani in sekundarno na desni strani grafikona.

Podatki za grafikon so: niz vrednosti spremenljivke na osi X, nizi vrednosti odvisnih spremenljivk na oseh Y, oznake odvisnih spremenljivk za izdelavo legende. Pri izdelovanju grafikonov pogosto uporabljamo deljene skupine označenih celic.

Čarovnik za grafikone v Excelu ima štiri korake. Iz koraka v korak se premikamo z gumbom *Naprej*, če pa želimo popraviti določila v prejšnjih korakih, se vrnemo nazaj z gumbom *Nazaj*. Z gumbom *Končaj* delo čarovnika za grafikone zaključimo. Gumb *Končaj* lahko uporabimo v vsakem koraku. Tedaj čarovnik za grafikone zaključi svoje delo in grafikon se prikaže v preglednici.


S klikom področja grafikona se v vogalih in na sredini stranic pojavijo črni kvadratici. Če se z miško postavimo na te kvadratke, se križec miške spremeni v puščico in s potegom velikost grafikona poljubno spreminjamo.

V nadaljevanju bomo oblikovali grafikon na sliki 1.1.




Slika 1.1: Prehodnost generacije študentov 1. letnika v šolskem letu 2002/2003 v 2. letnik na univerzitetnih študijskih programih Biotehniške fakultete, **grafični prikaz s stolpci** (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003)

1.5.1 Čarovnik za grafikone

Ob kliku ikone **Čarovnik za grafikone**  v standardni orodni vrstici se na zaslonu odpre pogovorno okno čarovnika za grafikone.

PREHODNOST 11

Označite zadnje štiri stolpce v tabeli 1.1. Prva vrstica označenih celic naj vsebuje podatke za legendo, prvi stolpec pa vrednosti opisne spremenljivke na osi X. V preglednici s podatki kliknite ikono . ♦

1. korak - Čarovnik za grafikone, Korak 1 od 4 – Vrsta grafikona

V prvem koraku na kartici **Standardne vrste** na seznamu **Vrsta grafikona** s klikom imena izberemo vrsto grafikona (stolpčni, palični, črtni,...). V polju **Podvrsta grafikona** se pokažejo ikone s shematičnim prikazom podvrst grafikonov. S klikom na ustrezno ikono izberemo podvrsto grafikona.

PREHODNOST 12

Izberite vrsto grafikona *Stolpčni* in prvo podvrsto grafikona. ♦

2. korak - Čarovnik za grafikone, Korak 2 od 4 – Izvor podatkov za grafikon

V tem koraku na kartici **Obseg podatkov** določimo celice s podatki za grafikon. Če smo te celice predhodno že označili, se v polju **Obseg podatkov** pojavi ime skupine označenih celic. V polju **Nizi v** določimo, ali so podatki v vrsticah (**Vrsticah**) ali v stolpcih (**Stolpcih**).

Na drugi kartici **Niz** lahko posamezne nize spremenimo, odstranimo ali dodamo na grafikon. V seznamu **Nizi** so navedena imena nizov na grafikonu. Obseg podatkov za posamezen niz spremenimo tako, da kliknemo ime niza v seznamu in ponovno določimo polji **Ime** in **Vrednosti**. Vsebina celice, ki jo določimo v polju **Ime**, je v legendi na grafikonu. V polju **Vrednosti** označimo naslov celic, v katerem so vrednosti za izbrani niz. Posamezen niz odstranimo z grafikona tako, da označimo njegovo ime v seznamu **Nizi** in kliknemo gumb **Odstrani**. Nov niz dodamo na grafikon s klikom na

gumb **Dodaj** in določitevijo polj **Ime** in **Vrednosti**. V polje **Oznake osi kategorije (X)** napišemo naslov bloka z oznakami kategorij oz. vrednostmi na osi X.

PREHODNOST 13

V drugem koraku čarovnika za grafikone preverite, ali ste označili prave celice. ♦

3. korak – Čarovnik za grafikone, Korak 3 od 4 – Možnosti grafikonov

V tem koraku čarovnika za grafikone dodatno oblikujemo grafikon z izbiro možnosti na šestih karticah: **Naslovi**, **Osi**, **Mrežne črte**, **Legenda**, **Oznake podatkov**, **Podatkovna tabela**. Ko spreminjamo določila na karticah, se izgled grafikona sproti spreminja.

PREHODNOST 14

Preizkusite različne možnosti v tretjem koraku čarovnika za grafikone in opazujte, kako se grafikon spreminja. Končna izbira možnosti naj bo taka, da bo grafikon čim bolj podoben sliki 1.1. Opremite grafikon z naslovi. ♦

4. korak - Čarovnik za grafikone, Korak 4 od 4 – Mesto grafikona

V četrtem koraku čarovnika za grafikone določimo mesto grafikona v zvezku. Možnost **Kot nov list** izberemo, če želimo, da se grafikon pojavi na novem grafičnem listu v zvezku. Ime **Grafikon 1** lahko spremenimo po lastni izbiri. Možnost **Kot predmet v** izberemo, če želimo, da se grafikon pojavi kot objekt na enem izmed že obstoječih delovnih listov v zvezku.

Priporočamo, da so grafični prikazi na grafičnih delovnih listih in ne kot objekt v preglednici. Grafične liste smiselno vsebinsko poimenujemo.

PREHODNOST 15

V četrtem koraku čarovnika postavite grafikon na nov grafični list z imenom »Slika1.1«. ♦

1.5.2 Spreminjanje grafikona

Potem ko čarovnik za grafikone opravi svoje delo, lahko grafikon še dodatno oblikujemo in spreminjamo. Z desnim klikom na posamezni del grafikona se odpre meni z ukazi, ki pomagajo oblikovati izbrani del grafikona. Del grafikona, ki je trenutno izbran, je označen s črnimi kvadrati. Če izberemo os X ali Y, se črni kvadrati pojavijo

na začetku in na koncu osi. Če izberemo niz vrednosti odvisne spremenljivke, se črni kvadratki pojavijo na izbrani krivulji (stolpcu, ...).

Deli grafikona so:

- **risalna površina** je med osema X in Y. To področje oblikujemo z ukazoma **Oblikuj risalno površino** in **Oblikuj mrežne črte**. Mrežne črte sestavljajo mrežo v ozadju slike;
- **področje grafikona** oblikujemo z ukazom **Oblikuj področje grafikona**. Tu lahko določimo obliko vseh znakov v grafikonu naenkrat (naslovi, oznake na oseh); določimo lahko, ali naj bo grafikon v okvirčku ali ne, kakšna bo barva ozadja itd.;
- **področje nizov vrednosti spremenljivk** je odvisno od vrste grafikona: npr. točke, krivulja, stolpci, krožni izseki itd. Prikaze nizov podatkov oblikujemo z ukazom **Oblikuj nize podatkov**;
- **področje osi X** tvorijo: črta, ki ponazarja os X, črtice, ki določajo skalo na osi ali pa ločijo posamezne kategorije (*Glavna črtica, Pomožna črtica*), oznake teh črtic (*Oznake črtic*). Področje osi X oblikujemo z ukazom **Oblikuj os**;
- **področje naslova osi X** je pod osjo X. Vanj napišemo naslov osi X z ukazom **Možnosti grafikona** na kartici *Naslovi*;
- **področje osi Y** tvorijo: črta, ki ponazarja os Y, črtice, ki določajo skalo na osi ali pa ločijo posamezne kategorije (*Glavna črtica, Pomožna črtica*) in oznake teh črtic (*Oznake črtic*). Področje osi X oblikujemo z ukazom **Oblikuj os**;
- **področje naslova osi Y** je levo od osi. Vanj napišemo naslov osi Y z ukazom **Možnosti grafikona** na kartici *Naslovi*;
- **področje legende** sestavljajo oznake in naslovi oznak v legendi. Legendo oblikujemo in ji določimo položaj na grafikonu z ukazom **Oblikuj legendo**;
- **področje naslova grafa** je nad grafom. Vanj napišemo naslov grafikona z ukazom **Možnosti grafikona** na kartici *Naslovi*;

PREHODNOST 16

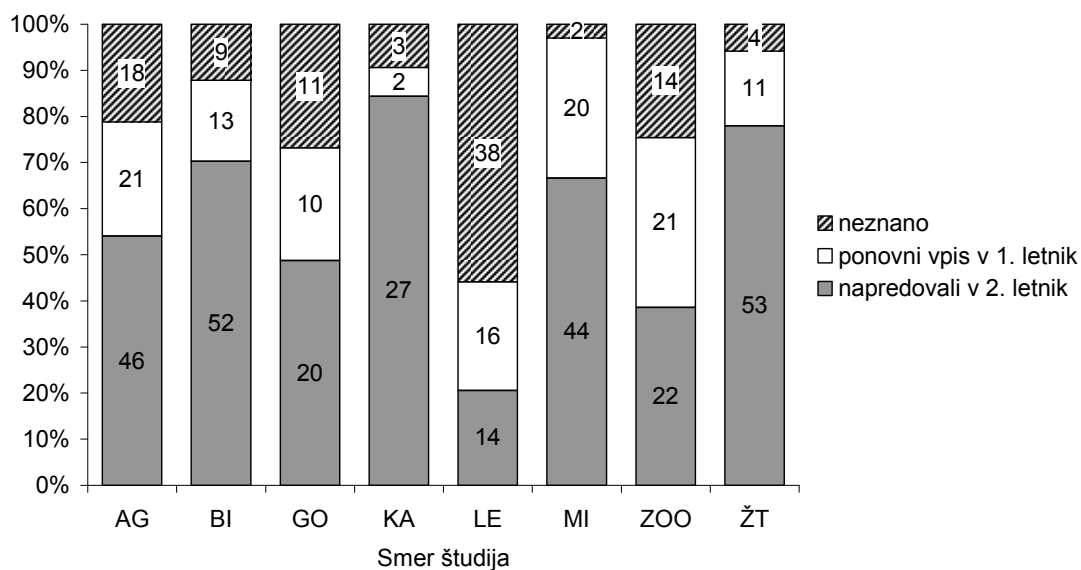
Z desnim klikom na področje grafikona odprite meni z ukazi za oblikovanje grafikona in v meniju izberite ukaz **Oblikuj področje grafikona**. Na kartici *Pisava* izberite velikost črk 16, odebeljeni tisk (***Krepko***) in pisavo Times New Roman. ♦

PREHODNOST 17

Z desnim klikom na področje slike odprite meni in izberite ukaz **Oblikuj risalno površino**. V polju **Obroba** izberite možnost **Brez** in enako v polju **Področje**. ♦

PREHODNOST 18

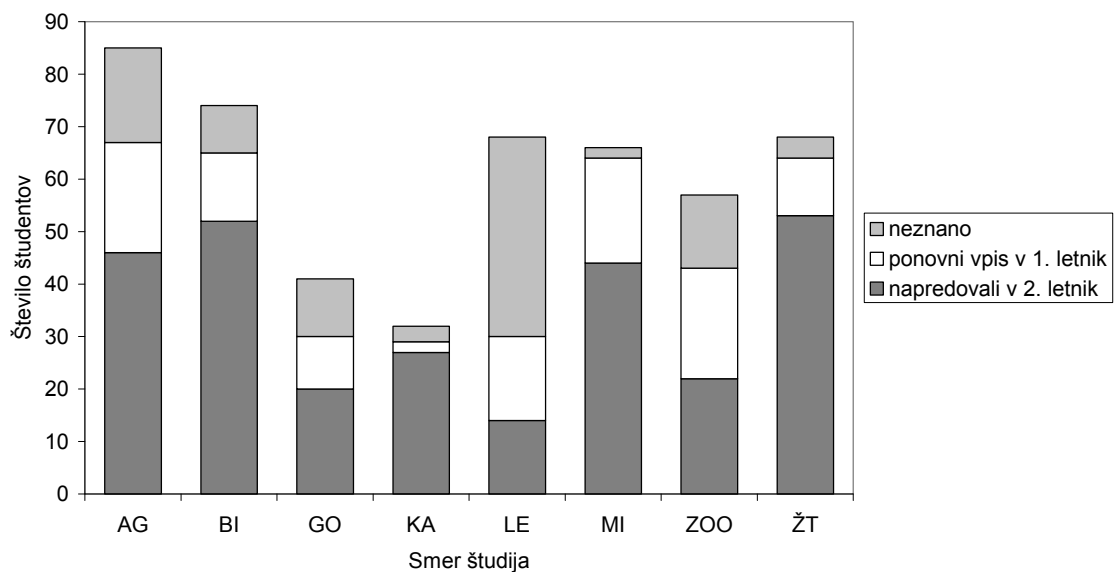
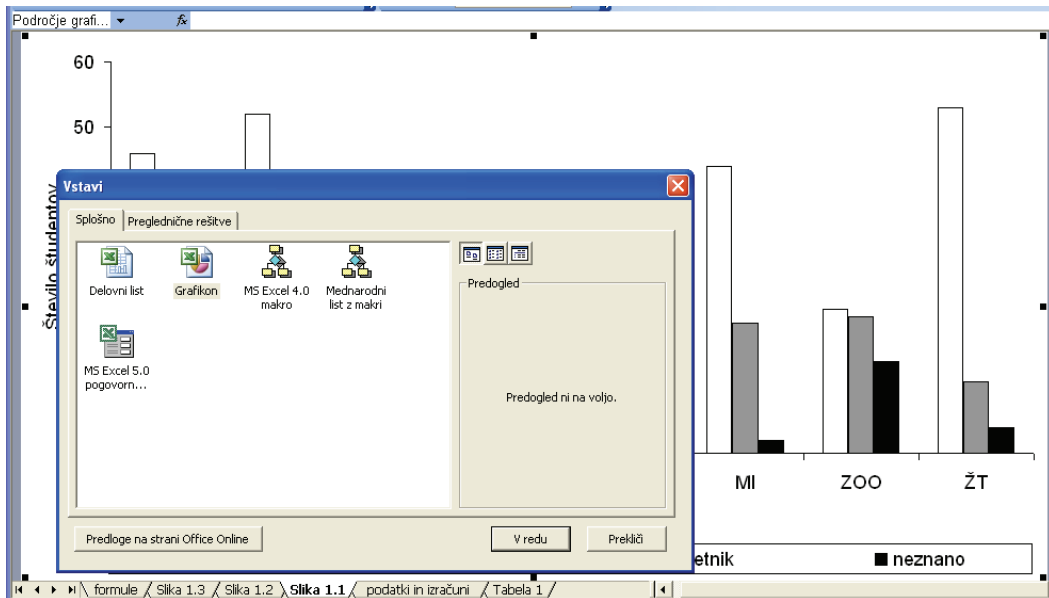
Na podlagi podatkov v tabeli 1.1 oblikujte še en grafikon, prikazan na sliki 1.2. Grafikon naj bo na svojem grafičnem listu z imenom »Slika 1.2«. V tem primeru je za vsako študijsko smer s strukturnimi stolpci prikazana struktura prehodnosti študentov, vpisanih v 1. letnik v šolskem letu 2002/03. ♦



Slika 1.2: Struktura prehodnosti študentov 1. letnika v šol. l. 2002/03 na Biotehniški fakulteti v Ljubljani, po smereh študija (%); **grafični prikaz s strukturnimi stolpci** (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003). Na strukturnih stolpcih je prikazano absolutno število študentov.

Kadar oblikujemo različne vrste grafikonov na podlagi istih podatkov, lahko uporabimo naslednjo pot: z desnim klikom na jeziček z imenom že oblikovanega grafičnega lista se odpre pomožni meni, na katerem kliknemo na ukaz **Vstavi**. Odpre se pogovorno okno, v

katerem kliknemo na ikono **Grafikon**. S tem ponovno aktiviramo Čarovnika za grafikone na podlagi istih podatkov.

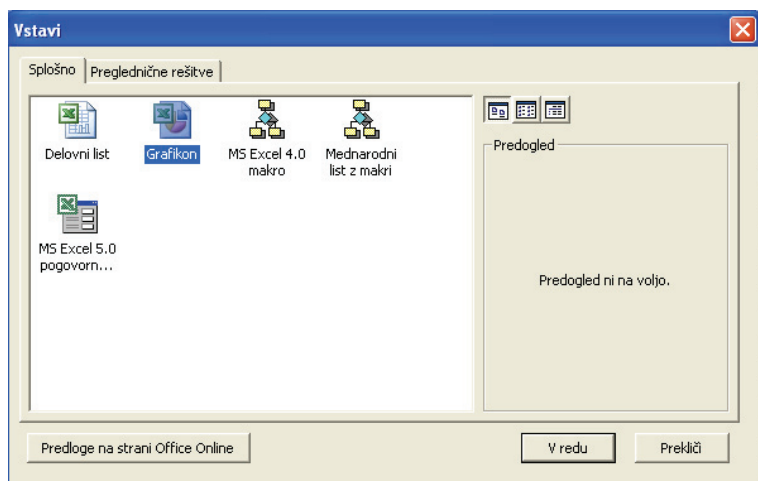


Slika 1.3: Prehodnost generacije študentov 1. letnika v šolskem letu 2002/2003 v 2. letnik na univerzitetnih študijskih programih Biotehniške fakultete, **grafični prikaz z naloženimi stolpci** (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003)

PREHODNOST 19

Na podlagi podatkov v tabeli 1.1 oblikujte še tretji grafikon, prikazan na sliki 1.3. Grafikon naj bo na svojem grafičnem listu z imenom »Slika 1.3«. V tem primeru so podatki za posamezno smer študija prikazani z naloženimi stolpci. Katero vrednost predstavlja višina stolpca na sliki 1.3? ♦

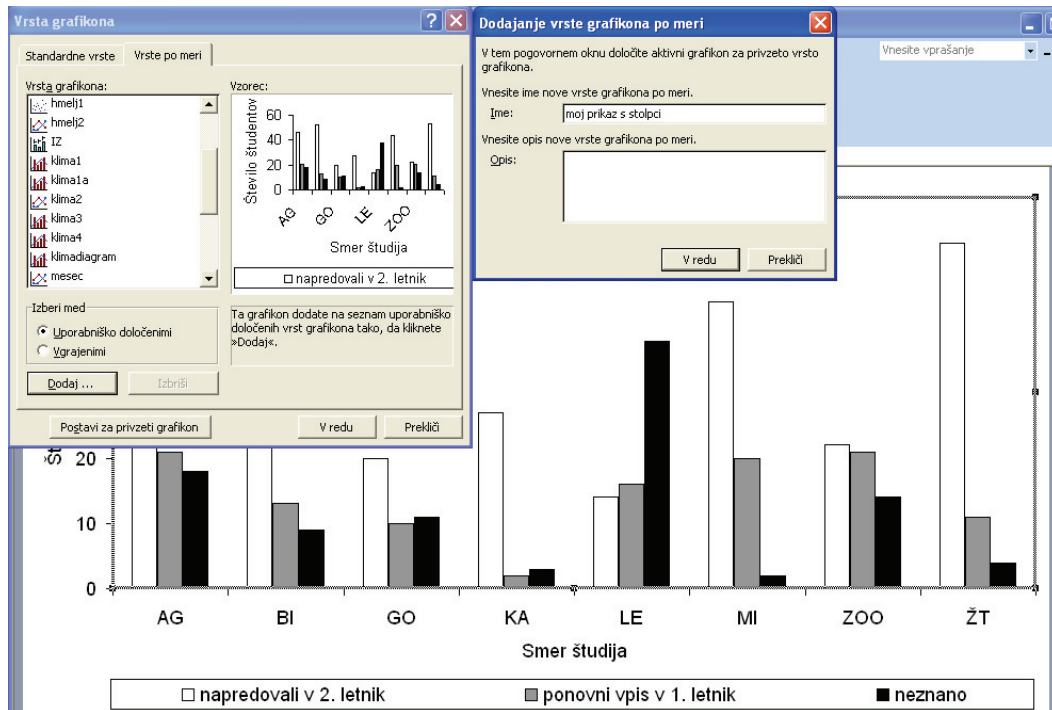
Če potrebujemo **prazen grafični list**, na katerega želimo kopirati sicer že oblikovani grafikon ali pa želimo grafikon oblikovati z orodjem **Risanje**, kliknemo na enega izmed jezičkov obstoječih delovnih listov v zvezku, nato z desnim klikom odpremo meni in izberemo ukaz **Vstavi**:



V pogovornem oknu **Vstavi** izberemo ikono **Grafikon** in kliknemo na gumb **V redu**. S tem aktiviramo Čarovnika za grafikone, v katerem takoj kliknemo na gumb **Dokončaj**. S tem se odpre nov delovni list z imenom »Grafikon #« (# je številka, ki je odvisna od tega, koliko grafičnih delovnih listov v zvezku že obstaja). Če smo v zvezku predhodno že oblikovali grafikone, ponavadi tako vstavljen grafični delovni list prikazuje podatke, s katerimi smo se nazadnje ukvarjali. Da se grafikona znebimo, z desnim klikom izberemo **Področje grafikona** in uporabimo ukaz **Počisti**; s tem dobimo prazen grafični list.

Večkrat se zgodi, da z dodatnim oblikovanjem že pripravljenih vrst grafikonov oblikujemo novo vrsto grafikona, ki bi jo radi uporabili tudi na drugih podatkih. V takem primeru definiramo t. i. **Uporabniško določeni grafikon**. Desni klik miške na že oblikovani grafikon odpre meni, na katerem izberemo ukaz **Vrsta grafikona**, v pogovornem oknu ukaza na kartici **Vrste po meri** izberemo možnost **Uporabniško določeni** in s klikom na gumb **Dodaj** v seznam dodamo novo vrsto grafikona, kateri moramo določiti tudi ime. S tem se ime nove vrste

grafikona pojavi na seznamu in jo lahko uporabimo pri kasnejšem oblikovanju podobnih grafikonov.



1.6 Formule

V celici Excelove preglednice uporabimo formulo, kadar želimo kaj izračunati. Formule so lahko preproste - seštevanje, odštevanje, deljenje, množenje, primerjave, lahko pa so tudi zapletene matematične funkcije. Formulo, ki jo napišemo v celico, si lahko predstavljamo kot enačbo, katere rezultat je napisan v celici, enačba pa v vrstici za vnos. Formule so v osnovi sestavljene iz vrednosti in operatorjev. V Excelu formule vedno začnemo pisati z znakom =, vnos formule v celico pa potrdimo s pritiskom na tipko **ENTER**. Ob tem se v celico izpiše rezultat formule.

Aritmetični operatorji so: + (seštevanje), - (odštevanje), / (deljenje) in * (množenje), ^ (potenca). Logični operatorji so = (enako), < (manjše), > (večje), >= (večje ali enako), <= (manjše ali enako), <> (neenako). Vrstni red računanja v bolj zapletenih formulah določamo z oklepaji.

FORMULE 1

V datoteki PREHODNOST.XLS se postavite na prazni delovni list in poimenujte to stran »Formule«. V celico A1 napišite formulo $= (24*60-5)*5$ in v celico B1 formulo $=24*60-5*5$. V čem je razlika med formulama? Bodite pozorni na uporabo oklepajev. ♦

Namesto vrednosti lahko v formulo vpišemo tudi naslov celice (referenco), v kateri je podatek, ki ga želimo uporabiti v formuli⁴. V tem primeru se rezultat formule samodejno popravi, če spremenimo vsebino referenčne celice.

FORMULE 2

V celico A3 vnesite število 5 in v celico B3 napišite formulo $=A3*10$. Nato spremenite podatek v celici A3 na število 4. Poglejte, kaj se zgodi z vrednostjo v celici B3.

Navodilo: Drug način pisanja reference celice v formulo je s klikom na to celico: npr. najprej natipkamo znak =, nato z miško kliknemo na celico A3, ob tem se referenca celice izpiše v formulo, nadaljujemo s tipkanjem znaka * in številke 10. ♦

1.6.1 Kopiranje formul

Formule lahko kopiramo in predstavljamo po preglednici, med stranmi preglednice in med datotekami. Pri tem pa moramo biti posebno pazljivi, če so v formulah reference. Excel kopira formule tako, da uporablja t.i. **relativno naslavljanje celic**: upošteva relativni položaj referenčne celice glede na celico, v katero smo vnesli formulo. Bistvo relativnega naslavljanja bomo spoznali na zgledih.

FORMULE 3

Kopirajte formulo v celici B3 v celico B4 in nato še v skupino celic C3:F3. Nato si oglejte rezultate in formule v celici B4 in v skupini celic C3:F3. ♦

FORMULE 4

V celico A4 vpišite formulo $=A3+1$. Kaj se zgodi z vrednostjo v celici B4? ♦

FORMULE 5

Skupino celic A4:B4 kopirajte v skupino celic A5:B10. Oglejte si rezultate in formule v skupini celic A5:B10. ♦

⁴ Ni pomembno ali napišemo veliko ali majhno črko (a3 ali A3), po pritisku na tipko **ENTER** Excel samodejno izpiše reference celic z velikimi črkami.

Relativno naslavljanje celic pa ni vedno ustrezno, zato Excel pozna tudi **absolutno in mešano naslavljanje celic**. Absolutno naslovimo celico tako, da pred ime stolpca in vrstice v imenu celice postavimo znak \$ (primer: \$A\$1). Dva znaka \$ v referenco celice najhitreje dodamo s pritiskom tipke **F4**. Na primer za formulo = \$A\$3*10 najprej natipkamo znak =, nato natipkamo ali kliknemo na celico A3 in takoj za tem pritisnemo tipko **F4** ter dodamo *10. Če je formula že napisana z relativno naslovljeno celico A3, jo spremenimo v absolutno naslovljeno celico tako, da utripajoči kazalec v vnosni vrstici postavimo na enega izmed treh možnih položajev: pred ali za črko A, ali za številko 3, nato kliknemo tipko **F4**.

Če kopiramo formulo, v kateri smo uporabili absolutno naslavljanje celice, bo ne glede na to, kam formulo kopiramo, v formuli ostala ista referenčna celica.

FORMULE 6

Popravite formulo v celici B3 na =\$A\$3*10 in jo kopirajte v skupino celic B4:B10 in v skupino celic C3:F3. Kakšna je razlika v rezultatih kopiranih formul z relativnim in absolutnim naslavljanjem celic? ♦

Mešano naslavljanje celic uporabimo takrat, kadar absolutno naslovimo samo stolpec ali pa samo vrstico reference (\$A3 ali A\$3).



FORMULE 7


Popravite formulo v celici B3 na =\$A3*10 in jo kopirajte v skupino celic B4:B10 in v skupino celic C3:F3. V katerih celicah se rezultati spremenijo? ♦

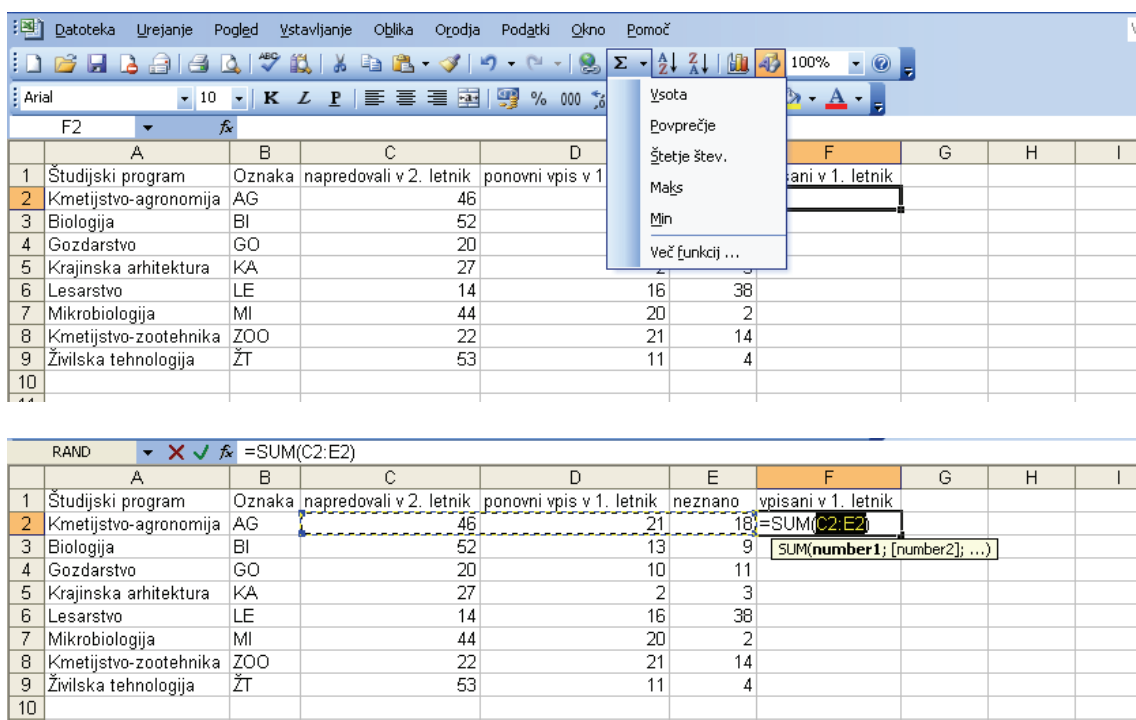
FORMULE 8

Popravite formulo v celici B3 na =A\$3*10 in jo kopirajte v skupino celic B4:B10 in v skupino celic C3:F3. V katerih celicah se rezultati spremenijo? ♦

1.6.2 Vstavljanje funkcij

Excel ima veliko vgrajenih funkcij. V standardni orodni vrstici stoji ikona  (Samodejna vsota), ki omogoča hitro seštevanje po stolpcih ali vrsticah. Klik na puščico na ikoni  poleg računanja vsote omogoča tudi hitro računanje povprečja izbranih vrednosti (**Povprečje**), prešteje koliko vrednosti smo izbrali (**Štetje šte.**) ter

poišče minimalno in maksimalno vrednost (**Min**, **Maks**). Druge funkcije najdemo na seznamu, ki se odpre s klikom na ukaz **Več funkcij**. Za vstavljanje funkcij lahko uporabimo ukaz **Vstavljanje/Funkcija**, hitrejša pot pa je s klikom na ikono  v vrstici za vnos. Funkcijo lahko vnesemo tudi z vpisom v celico ali pa z vpisom v vrstico za vnos (celica, v kateri želimo izračun, mora biti izbrana). Na ta način lahko sestavljamo poljubne formule.



	A	B	C	D	E	F	G	H	I
1	Študijski program	Oznaka	napredovali v 2. letnik	ponovni vpis v 1. letnik	neznano vpisani v 1. letnik				
2	Kmetijstvo-agronomija	AG	46	21	18	=SUM(C2:E2)			
3	Biologija	BI	52	13	9	SUM(number1; [number2]; ...)			
4	Gozdarstvo	GO	20	10	11				
5	Krajinska arhitektura	KA	27	2	3				
6	Lesarstvo	LE	14	16	38				
7	Mikrobiologija	MI	44	20	2				
8	Kmetijstvo-zootehnika	ZOO	22	21	14				
9	Živilska tehnologija	ŽT	53	11	4				
10									

Natančneje se bomo z načini uporabe zahtevnejših formul seznanili v prihodnjih vajah.

PREHODNOST 20

Za podatke v tabeli 1.1 v datoteki PREHODNOST.XLS (skupina celic A1:E9) izračunajte:

- v skupino celic F2:F9 izračunajte število vpisanih študentov v 1. letnik v šol. l. 2002/03 za vsako smer študija;
- v celico F10 izračunajte skupno število vpisanih v 1. letnik na BF v Ljubljani;
- v skupino celic G2:I9 izračunajte strukturo prehodnosti iz 1. letnika (%) za vsako smer študija;
- v skupino celic J2:J9 izračunajte strukturo vpisanih v 1. letnik v šol. l. 2002/03 na BF po smereh (%).

Navodilo: Formule vpišite samo za prvo smer študija, v druge celice jih kopirajte (poteg z miško⁵). Pri kopiranju pazite na relativno in absolutno naslavljanje celic. ♦

1.6.3 Kopiranje na poseben način

Poleg običajnega načina kopiranja z ukazoma **Urejanje/Kopiraj** in **Urejanje/Prilepi**, ki ju že poznamo, Excel omogoča tudi posebno kopiranje z ukazoma **Urejanje/Kopiraj** in **Urejanje/Posebno lepljenje**. V pogovornem oknu *Posebno lepljenje* imamo dve skupini možnosti, *Prilepi* in *Operacija*. V skupini *Prilepi* so naslednje možnosti:

	A	B	C	D	E	F	G	H	I	J
1	Študijski program	Oznake napredovali v 2. letnik ponovni vpis v 1. letnik	neznano vpisani v 1. letnik	% 2. letnik	% ponovno	% neznano	% BF			
2	Kmetijstvo-agronomija	AG				85	54,1	24,7	21,2	17,3
3	Biologija	BI				74	70,3	17,6	12,2	15,1
4	Gozdarstvo	GO				41	48,8	24,4	26,8	8,4
5	Krajinska arhitektura	KA				32	84,4	6,3	9,4	6,5
6	Lesarstvo	LE				68	20,6	23,5	55,9	13,8
7	Mikrobiologija	MI				66	66,7	30,3	3,0	13,4
8	Kmetijstvo-zootehnika	ZOO				57	38,6	36,8	24,6	11,6
9	Živilska tehnologija	ŽT				68	77,9	16,2	5,9	13,8
10						491				

Prilepi	
<input checked="" type="radio"/> Vse	<input type="radio"/> Preverjanje veljavnosti
<input type="radio"/> Formule	<input type="radio"/> Vse razen obrob
<input type="radio"/> Vrednosti	<input type="radio"/> Širine stolpcev
<input type="radio"/> Oblike	<input type="radio"/> Formule in številске oblike
<input type="radio"/> Komentarje	<input type="radio"/> Vrednosti in številске oblike
Operacija	
<input checked="" type="radio"/> Brez	<input type="radio"/> Pomnoži
<input type="radio"/> Dodaj	<input type="radio"/> Deli
<input type="radio"/> Odvzemi	
<input type="checkbox"/> Preskoči prazne	<input type="checkbox"/> Transponiraj
Prilepi s povezavo V redu Prekliči	

- možnost *Vse* je izbrana pri osnovni nastavitvi in kopira vsa določila vsebine odlagališča;

⁵ Z miško se postavite v desni spodnji vogal celice s formulo in ko se oblika in barva križca spremenita, potegnite z miško čez celice, v katere želite kopirati formulo.

- možnost **Formule** kopira formule brez oblikovnih določil;
- možnost **Vrednosti** uporabimo, če želimo kopirati vrednosti, izračunane na osnovi formul. Če bi uporabili navadno kopiranje, bi zaradi relativnega naslavljanja celic v formulah dobili drugačne vrednosti;
- možnost **Oblike** kopira oblikovna določila vsebine odlagališča.

Možnosti v skupini **Operacija** določajo operacije med vsebino celic, ki jih želimo kopirati in vsebino celic, kamor želimo kopirati:

- možnost **Brez** določi, naj vsebina, ki jo želimo kopirati, zamenja vsebino izbrane celice. Ta možnost je izbrana pri osnovni nastavitvi;
- možnosti **Dodaj** in **Odvzemi** določita, naj se vsebina, ki jo želimo kopirati, prišteje oziroma odšteje od vsebine v celici, kamor želimo kopirati;
- možnosti **Pomnoži** in **Deli** določita, da se vsebina celice, kamor želimo kopirati, pomnoži oziroma deli z vsebino celice, ki jo kopiramo.

Možnost **Preskoči prazne** izberemo, kadar želimo, da se prazne celice ne kopirajo.

Možnost **Transponiraj** kopira vsebino odlagališča tako, da vrstice postavi v stolpce: prva vrstica postane prvi stolpec.

PREHODNOST 21

Celice F1:J9 kopirajte na nov delovni list. Zakaj se vrednosti v celicah na novem delovnem listu spremenijo? ♦

PREHODNOST 22

Celice F1:J9 kopirajte na nov delovni list z ukazoma **Urejanje/Kopiraj** in **Urejanje/Posebno lepljenje, Vrednosti**. Ali celice na novem delovnem listu še vsebujejo formule? ♦

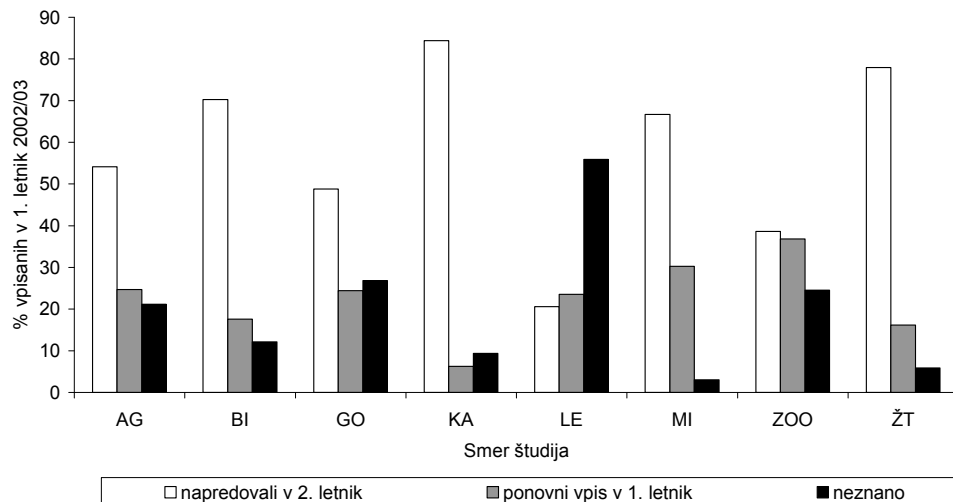
PREHODNOST 23

Tabelo A1:J9 kopirajte na nov delovni list z ukazoma **Urejanje/Kopiraj** in **Urejanje/Posebno lepljenje, Transponiraj**. ♦

PREHODNOST 24

Podatke iz tabele 1.1 prikažite še z razrezanimi strukturnimi stolpci (slika 1.4). Delovni list z grafikonom v preglednici poimenujte z imenom »Slika 1.4«.

Razmislite: ali bi sliko 1.4 lahko naredili tudi samo na podlagi podatkov v tabeli 1.1 (brez dodatnih izračunov)? Zakaj pri sliki 1.2 dodatni izračuni niso bili potrebni? ♦

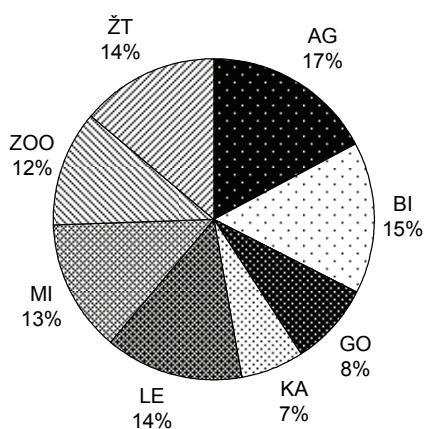


Slika 1.4: Prehodnost generacije študentov 1. letnika v šolskem letu 2002/2003 v 2. letnik na univerzitetnih študijskih programih Biotehniške fakultete, grafični prikaz z **razrezanimi strukturnimi stolpci** (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003)

PREHODNOST 25

Grafično prikažite strukturo po smereh študija za vse vpisane v 1. letnik na BF, tako kot prikazuje slika 1.5. Delovni list z grafikonom v preglednici poimenujte z imenom »Slika 1.5«. (ukaz **Vrsta grafikona**, kartica *Vrste po meri*, **ČB-tortni**).

Navodilo: grafikon lahko oblikujete tako na podlagi absolutnih podatkov, (števila študentov po smereh) kot tudi na podlagi relativnih vrednosti - izračunane strukture. Naredite oba primera: enkrat za vhodne podatke grafikona izberite kratice smeri študija in število vpisanih v 1. letnik v šol. l. 2002/03 za posamezno smer študija, drugič pa za vhodne podatke grafikona izberite kratice smeri študija in izračunano strukturo vpisanih v 1. letnik v šol. l. 2002/03 na BF po smereh študija. ♦



Slika 1.5: Struktura po smereh študija (%) za vpisane v 1. letnik v šol. l. 2002/03 na Biotehniški fakulteti v Ljubljani, grafični prikaz s **strukturnim krogom**.

Domača naloga 1.1: Okrasne rastline

V tabeli 1.2 so podatki o številu vzgojenih okrasnih rastlin po vrstah okrasnih rastlin za leti 2003 in 2006 v Sloveniji.

Tabela 1.2: Število vzgojenih rastlin (1000) glede na vrsto v Sloveniji za leti 2003 in 2006 (Vir: SURS, 2006: Popis vrtnarstva v Republiki Sloveniji v letu 2006)

Vrsta rastlin	Vzgojene rastline (1000)	
	Leto 2003	Leto 2006
Sobne rastline	1981	925
Balkonske rastline	6046	6711
Enoletnice	3687	2843
Dvoletnice	5059	4302
Večletne in lesnate rastline	1949	1329

OKRASNE RASTLINE 1

Podatke iz tabele 1.2 prepisite v excelovo datoteko z imenom OKRASNE RASTLINE.XLS in jih grafično prikažite. Razmislite, katera vrsta grafikona je najustreznejša. ♦

OKRASNE RASTLINE 2

S strukturnima stolpcema na istem grafikonu prikažite strukturo vzgojenih rastlin po njihovi vrsti za leti 2003 in 2006. ♦

Domača naloga 1.2: Vpis na BF

Tabela 1.3: Število redno vpisanih študentov na univerzitetne programe študija v šol. l. 2006/07 po smereh študija na Biotehniški fakulteti v Ljubljani (Vir: Univerza v Ljubljani BF, Poročilo za leto 2007)

Študijski program	1. letnik	2. letnik	3. letnik	4. letnik	Absolventi
Biologija	87	80	89	85	88
Biotehnologija	55	48	35	29	
Gozdarstvo	51	63	31	19	46
Kmetijstvo-Agronomija	83	69	68	78	70
Kmetijstvo-Zootehnika	85	43	41	38	63
Krajinska arhitektura	33	35	33	31	50
Lesarstvo	71	23	23	14	37
Mikrobiologija	64	46	45	48	64
Živilska tehnologija	77	69	60	60	78

VPIS 1

Podatke iz tabele 1.3 prepisite v datoteko VPIS.XLS. ♦

VPIS 2

Na enem grafikonu prikažite podatke o vpisu na BF v šolskem letu 2006/2007 (tabela 1.3). ♦

VPIS 3

Za vsako smer študija izračunajte strukturo študentov po letnikih študija (brez absolventov).
Strukture za vse smeri grafično prikažite s strukturnimi stolpci na istem grafikonu. ♦

VPIS 4

Za vse vpisane na BF v študijskem letu 2006/2007 z razrezanimi strukturnimi stolpci grafično prikažite in izračunajte strukturo študentov po smereh študija. ♦

VPIS 5

Za vsako smer študija posebej izračunajte odstotek absolventov in izračune grafično prikažite. ♦

2 Relativna števila in grafični prikazi

Novi pojmi:

- oblikovanje grafikonov v Excelu
- grafični prikaz časovnih vrst
- izračun in grafični prikaz relativnih števil

Datoteke s podatki:

PREBIVALSTVO.XLS

GOSTOTA PREBIVALSTVA.XLS

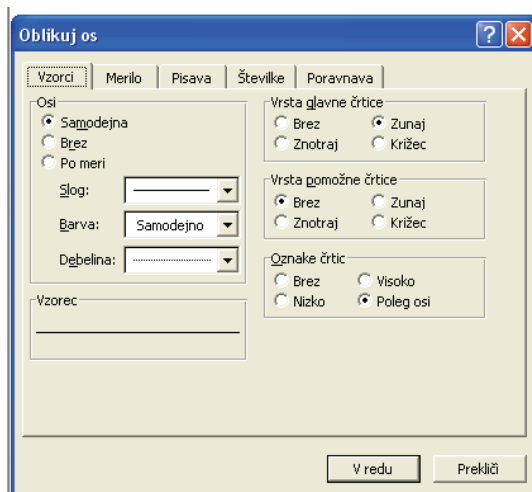
Datoteke, ustvarjene v prejšnjih poglavjih:

VPIS.XLS

OKRASNE RASTLINE.XLS

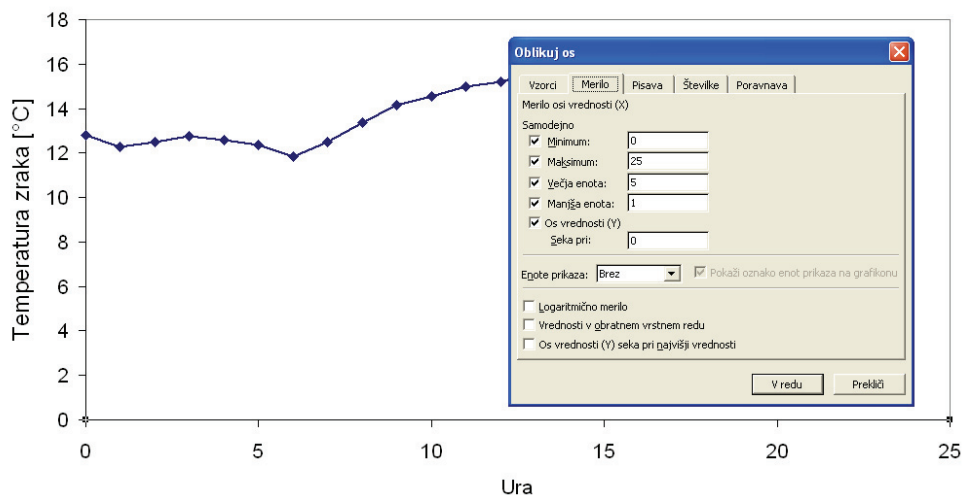
2.1 Oblikovanje osi grafikona

Meni z ukazi za oblikovanje osi grafikona aktiviramo z desnim klikom na os grafikona. Prvi ukaz v meniju je **Oblikuj os**, drugi pa **Počisti**. Ukaz **Počisti** uporabimo, kadar želimo os v grafikonu odstraniti. Os lahko ponovno narišemo v grafikon v tretjem koraku čarovnika - ***Možnosti grafikona***. S klikom na ukaz **Oblikuj os** se odpre pogovorno okno:

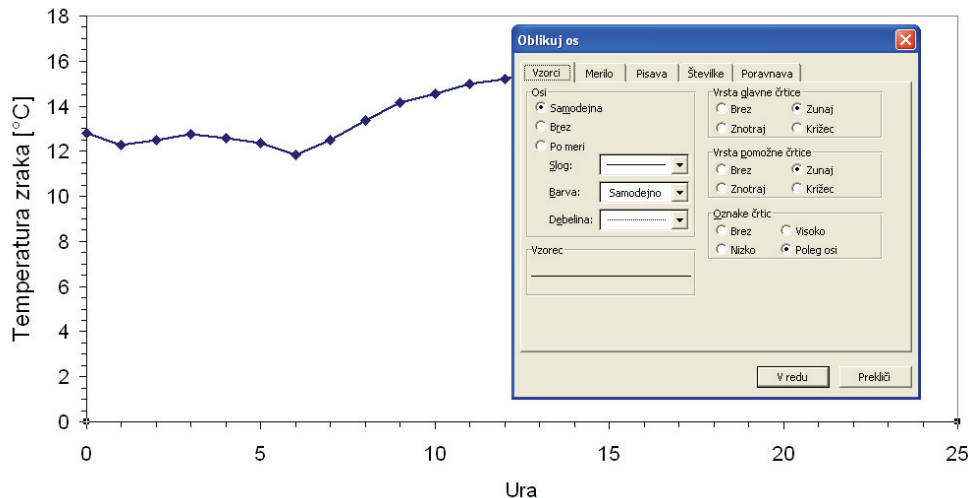


V pogovornem oknu ***Oblikuj os*** je več kartic:

- **Vzorci** - na tej kartici v polju **Osi** določimo obliko črte, ki predstavlja os; v polju **Vrsta glavne črtice** izberemo položaj črtic na osi; le-te ločujejo posamezne kategorije ali pa označujejo položaj posameznih vrednosti na osi; v polju **Vrsta pomožne črtice** izberemo položaj pomožnih (krajših) črtic, ki so med glavnimi črticami na osi - ob njih se ne izpisujejo oznake; v polju **Oznake črtic** določimo položaj oznak kategorij ali vrednosti ob glavnih črticah na osi;
- **Merilo**- vsebina te kartice ni vedno enaka, odvisna je od vrste grafikona in od vrste osi (številsko ali opisna/kategorična). Za številsko os (X ali Y) so možnosti, ki jih lahko določimo na tej kartici, naslednje:



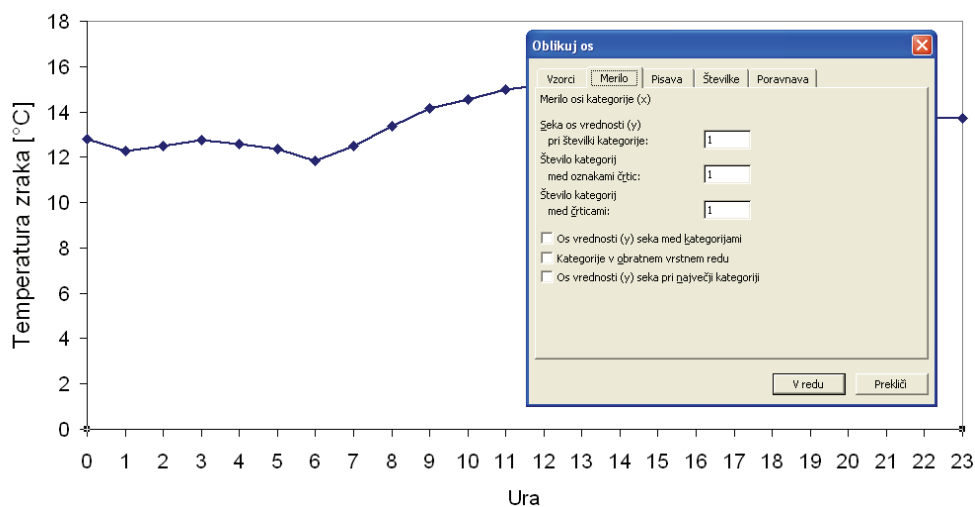
Na zgornji sliki je izbrana os X, vrsta grafikona je XY (Raztreseni). Če na kartici **Vzorci** določimo za **Vrsto pomožne črtice** možnost **Zunaj**, os X izgleda takole:



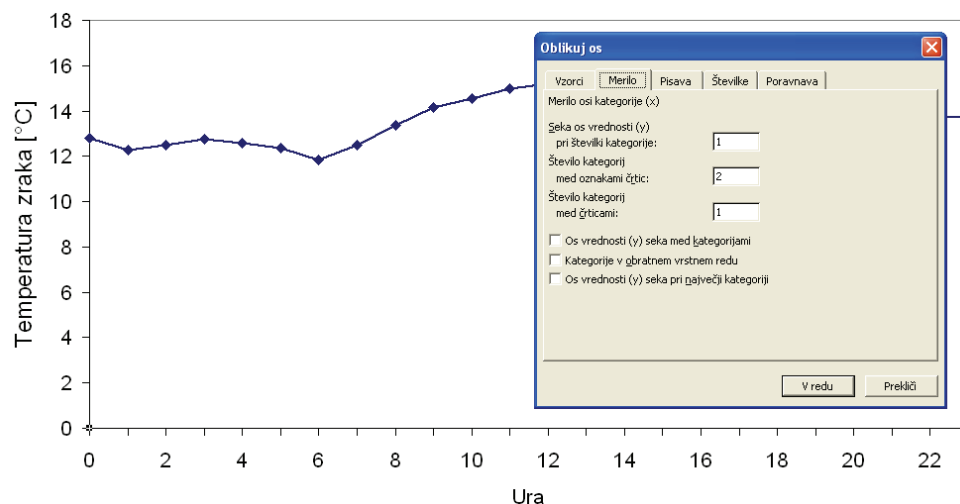
- **Minimum:** vrednost na spodnjem koncu osi;
- **Maksimum:** vrednost na zgornjem koncu osi;
- **Večja enota:** razmak med glavnimi črticami na osi;
- **Manjša enota:** razmak med pomožnimi črticami na osi;
- **Os kategorije(X) seka pri.:** če imamo izbrano os Y, ali pa **Os kategorije(Y) seka pri.:** če imamo izbrano os X: v polje napišemo vrednost na osi Y oz. X, pri kateri jo seka os X oz. Y;
- **Enote prikaza:** v tem polju lahko izberemo, v kakšnih enotah bodo prikazane številke ob osi (Brez, Stotice, Tisočice, Milijoni, Bilijoni, Triljoni);
- **Logaritmično merilo:** logaritemska skala na osi Y (Excel vrednost spremenljivke na osi Y logaritmiraj);
- **Vrednosti v obratnem vrstnem redu:** največja vrednost na osi Y je prikazana na spodnjem koncu osi in obratno;
- **Os kategorije (X) seka pri najvišji vrednosti.:** če imamo izbrano os Y ali pa **Os kategorije (Y) seka pri najvišji vrednosti.:** če imamo izbrano os X: ta možnost določa, da je os X na zgornjem robu grafikona oz. os Y na desnem robu grafikona.

Navedena določila v ukazu **Oblikuj os** so ista za os Y in os X, če je spremenljivka na osi X številska. Če pa je spremenljivka na osi X opisna, so na kartici *Merilo* za os X drugačne možnosti:

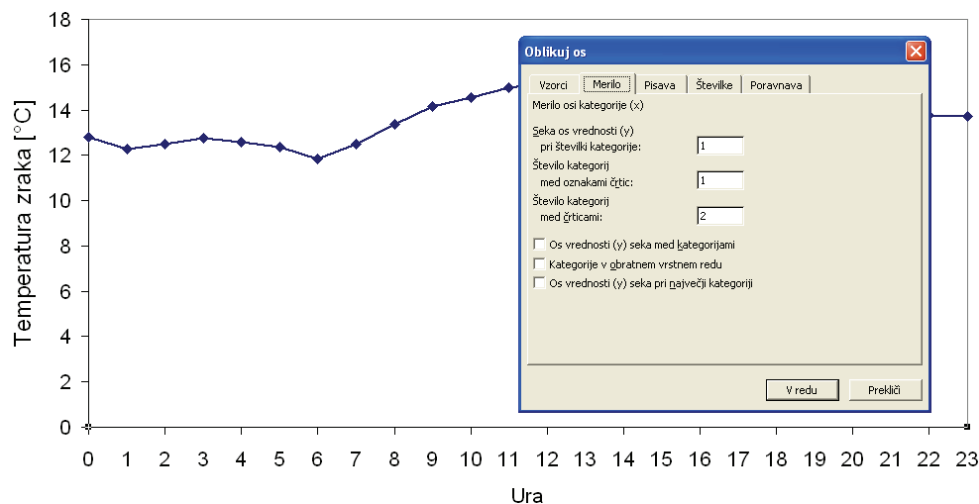
- **Seka os vrednosti (Y) pri številki kategorije:** os Y seka os X pri izbrani zaporedni številki kategorije, ki jo Excel določi glede na vrstni red podatkov v preglednici;
- **Število kategorij med oznakami črtic:** številski razmak za oznake glavnih črtic; k vmesnim glavnim črticam se oznake kategorij ne izpišejo;
- **Število kategorij med črticami:** številski razmak za glavne črtice; vmesne glavne črtice se na osi ne narišejo;
- možnost **Os vrednosti (Y) seka med kategorijami:** če je ta možnost izbrana, se vrednost spremenljivke na osi Y nariše na sredini med dvema glavnima črticama na osi X, v nasprotnem primeru pa se vrednost spremenljivke na osi Y nariše nad glavno črtico na osi X;
- možnost **Kategorije v obratnem vrstnem redu:** kategorije na osi X se pokažejo v obratnem vrstnem redu, kot so navedene v preglednici;
- možnost **Os kategorije (Y) seka pri najvišji kategoriji:** os Y je na desnem robu grafikona.



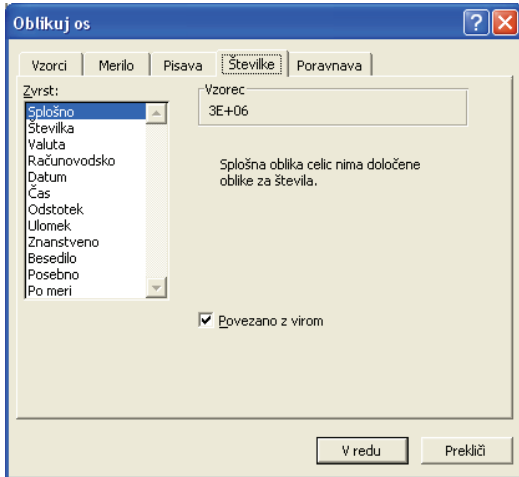
Na zgornji sliki je izbrana os X, vrsta grafikona je **Črtni**, zato je v tem primeru os X opisna (kategorična). Na kartici **Merilo** možnost **Os vrednosti (y) seka med kategorijami** ni izbrana (po osnovnih nastavitvah je izbrana). Če na kartici **Merilo** določimo za **Število kategorij med oznakama črtic: 2**, os X izgleda takole:



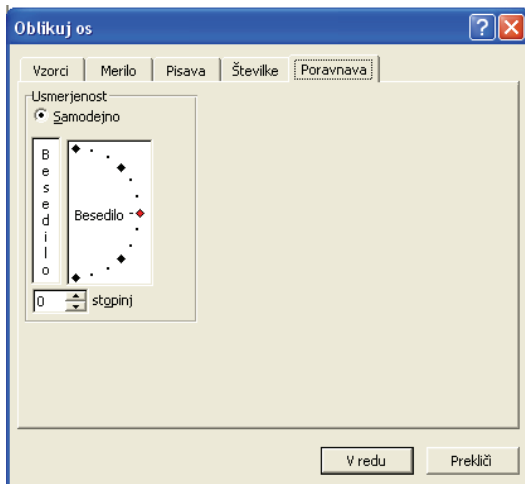
Če na kartici **Merilo** določimo za **Število kategorij med črticami: 2**, os X izgleda takole:



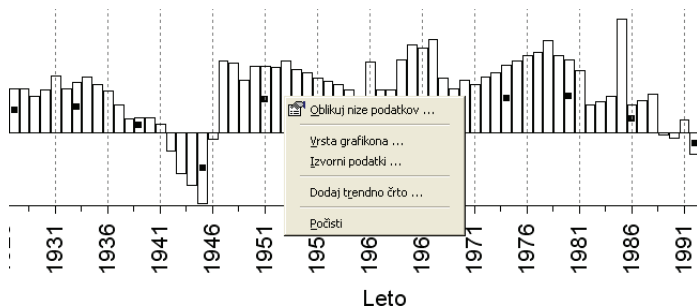
- **Pisava** - določimo obliko oznak na osi;
- **Številke** - določimo obliko vrednosti števil na osi;



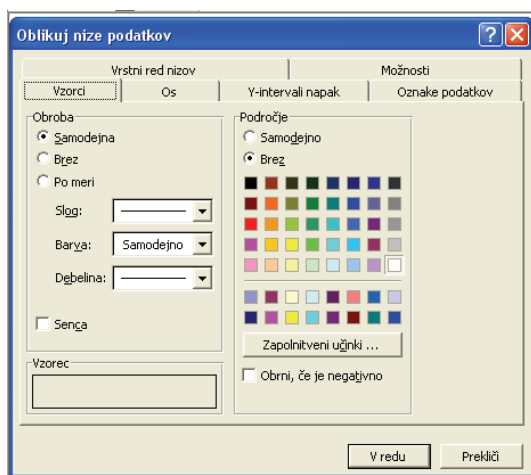
- **Poravnava** - določimo položaj oznak na osi (vodoraven, navpičen izpis, ...).



2.2 Oblikovanje niza vrednosti odvisne spremenljivke

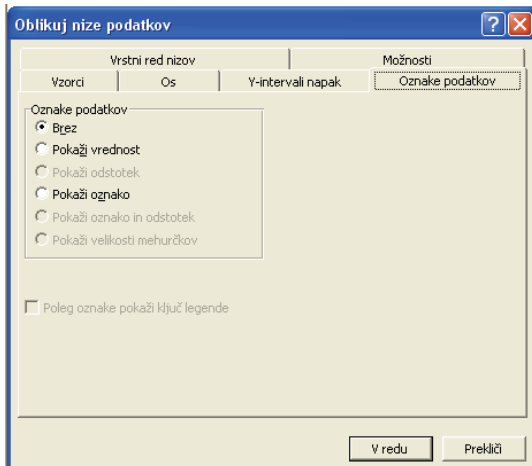


Z desnim klikom na predhodno izbrani objekt v grafikonu (npr. črta, stolpec) se pokaže meni z ukazi za oblikovanje tega objekta. Z ukazom **Počisti** pobrišemo prikaz izbranega objekta. Z ukazom **Oblikuj nize podatkov...** se odpre pogovorno okno *Oblikuj nize podatkov*, ki vsebuje več kartic:

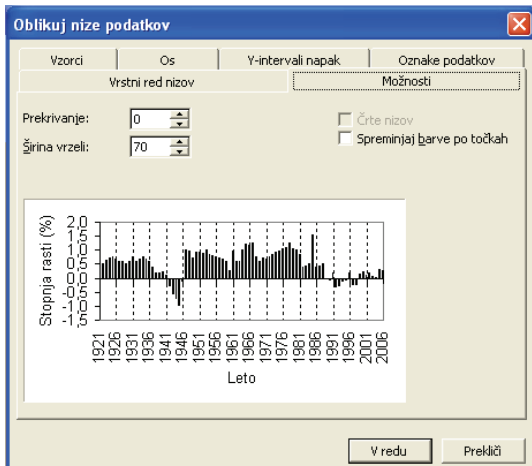


- **Vzorci** - na tej kartici določimo vzorec objekta, ki prikazuje niz podatkov (npr. neprekinjena, prekinjena, odebeljena črta, različne oblike točk, šrafura in barva stolpcev);
- **Os** - možnost **Primarna os** izberemo, če so vrednosti za izbrani niz podatkov prikazane na levi osi Y, možnost **Sekundarna os** izberemo za prikaz na desni osi Y. Excel pri osnovni nastavitvi vse nize podatkov prikaže na levi osi Y;

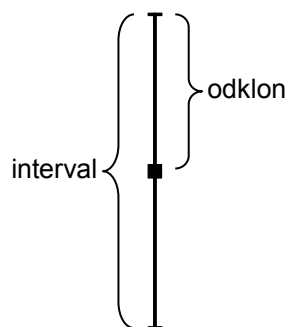
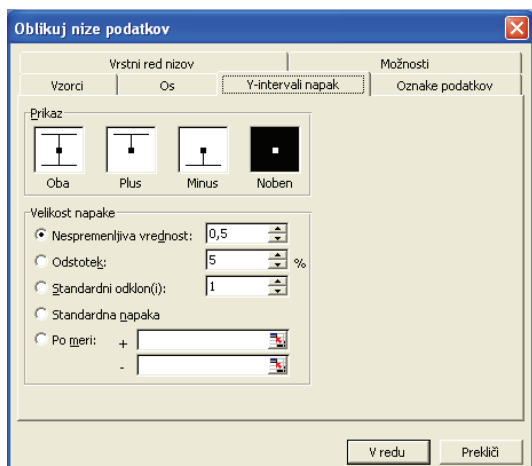
- **Oznake podatkov** - z izbiro možnosti **Pokaži vrednost** se ob vsakem prikazanem podatku izpiše vrednost Y. Z izbiro možnosti **Pokaži oznako** se ob vsakem prikazanem podatku izpiše vrednost X;



- **Možnosti** - na stolpčnem grafikonu nastavimo širino vrzeli med stolpci. V osnovni nastavitvi je **Širina vrzeli** 150, če želimo, da se stolpci na grafikonu držijo skupaj, pa mora biti v polju **Širina vrzeli** vrednost 0. V polju **Prekrivanje** določimo širino, za katero se zaporedni stolpci prekrivajo;



- **Y-intervali napak** - ob predstavitelju niza podatkov (točki ali stolpcu) zna Excel narisati tudi interval, katerega obliko določimo v razdelku *Prikaz* (**Oba**, **Plus**, **Minus**, **Noben**). V razdelku *Velikost napake* določimo velikost odklonov intervala. Če izberemo možnost **Nespremenljiva vrednost**, bodo vsi odkloni enako široki, njihovo širino pa določimo z izbiro vrednosti (0,5 po osnovni nastavitvi). Možnost **Odstotek** določi odklon intervala kot odstotek vrednosti, ki jo prikazuje predstavitelj (v tem primeru so intervali pri različnih vrednostih različno široki). Če uporabimo možnost **Standardni odklon(i)**, Excel izračuna povprečje (**AVERAGE**) in standardni odklon (**STDEV**) vseh vrednosti, ki jih prikazuje izbrani niz. Intervale nariše tako, da njihovo sredino postavi na vrednost povprečja, odklon intervalov pa je določen z izbranim številom standardnih odklonov. Podobno deluje možnost **Standardna napaka**, le da je odklon intervala namesto s standardnim odklonom določen s standardno napako za povprečje⁶. Peta možnost **Po meri** omogoča, da sami izberemo velikost odklona intervala, v tem primeru lahko narišemo tudi nesimetrični interval. To možnost uporabimo za grafični prikaz intervalov zaupanja (8. poglavje).



⁶ Standardna napaka za povprečje je standardni odklon, deljen s korenem števila podatkov izbranega niza.

2.3 Časovne vrste, indeksi, stopnje rasti in gostote ter njihovi grafični prikazi

Zgled 2.1: Naravno gibanje prebivalstva v Sloveniji

V datoteki PREBIVALSTVO.XLS so podatki o številu stalnega prebivalstva, živorojenih in umrlih v Sloveniji po letih v obdobju 1921-2006 (Vir: spletna stran SURS). Število prebivalstva je ocenjeno na dan 30.6., število živorojenih in umrlih pa se nanaša na celotno leto.

PREBIVALSTVO 1

Datoteko PREBIVALSTVO.XLS shranite v vašo mapo. Postavite se na delovni list »Podatki in izračuni«. Za boljši pregled podatkov in izračunov zamrznite podokna tako, da bodo na zaslonu vedno vidne prve štiri vrstice in prvi stolpec tabele s podatki.

Navodilo: Označite celico B5 in uporabite ukaz **Okno/Zamrzni podokna**. Prestavite se do spodnjega desnega roba preglednice s podatki in na podlagi videza zaslona ugotovite, kaj ste z ukazom **Okno/Zamrzni podokna** dosegli.⁷ ♦

PREBIVALSTVO 2

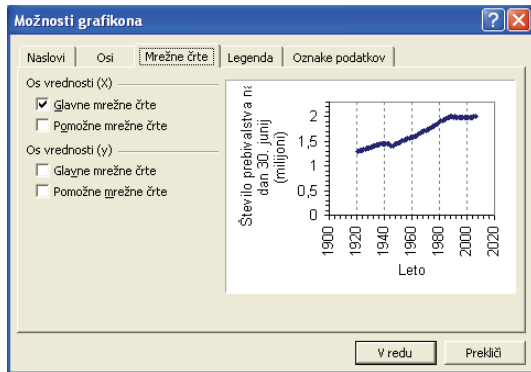
Grafično predstavite število prebivalstva v Sloveniji v obdobju 1921-2006 (slika 2.1).

Navodilo: Časovne vrste običajno prikazujemo z vrsto grafikona XY (Raztreseni).

Na delovnem listu najprej označimo podatke za leta in za število prebivalstva na dan 30. 6., (celice A4:B90) nato pa v Čarovniku za grafikone izberemo vrsto grafikona XY (Raztreseni). Excel na tem grafikonu samodejno priredi osi X podatke iz levega stolpca (Leta), podatki v desnem stolpcu (Število) pa se narišejo kot niz z vrednostmi na osi Y.

V okviru Čarovnika za grafikone poleg ostalih že poznanih nastavitev ustrezno določite položaj mrežnih črt:

⁷ Z enkratno uporabo omenjenega ukaza na izbranem delovnem listu v datoteki se v meniju **Okno** pojavi ukaz **Odmrzni podokna**.



Ko zaključite delo v čarovniku za grafikone, dodatno oblikujte os Y tako, da bodo oznake črtic na osi izpisane v milijonih in da bo maksimalna prikazana vrednost 2.100.000. Poleg tega naj bodo glavne črtice na osi 500.000 vsak sebi, vidne pa morajo biti tudi pomožne črtice 100.000 vsak sebi:

Na osi X naj se letnice izpišejo vsako peto leto, v obdobju 1921-2010. Ko določate vrednosti na kartici **Merilo**, možnosti za samodejno nastavitvev ne smejo biti izbrane:

Oblikuj os

Vzorci: Merilo | Pisava | Številke | Poravnava

Merilo osi vrednosti (X)

Samodejno

Minimum: 1921

Maksimum: 2010

Večja enota: 5

Manjša enota: 1

Os vrednosti (Y)

Šeka pri: 1921

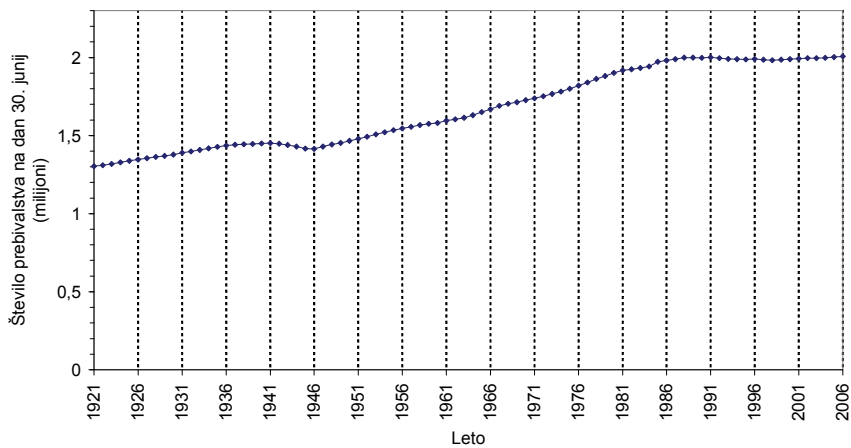
Epote prikaza: Brez Pokaži oznako enot prikaza na grafikonu

Logaritmčno merilo

Vrednosti v obratnem vrstnem redu

Os vrednosti (Y) šeka pri najvišji vrednosti

V redu Preklič



Slika 2.1: Število prebivalstva v Sloveniji na dan 30. junija v obdobju 1921-2006

PREBIVALSTVO 3

Izračunajte verižne indekse in stopnje rasti za število prebivalstva v Sloveniji v obdobju 1921-2006. ♦

PREBIVALSTVO 4

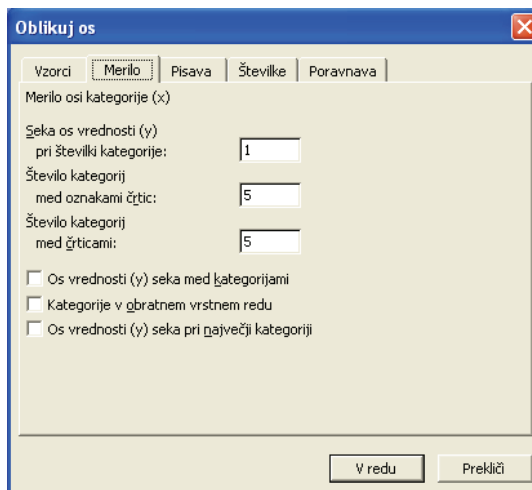
Letno stopnjo rasti za število prebivalstva v Sloveniji grafično prikažite (slika 2.2).

Navodilo: Za grafični prikaz stopnje rasti je primeren prikaz s stolpci (vrsta grafikona Stolpčni).

Označimo celice s podatki o stopnji rasti (D4:D90, celica D5 mora biti prazna) in izberemo vrsto grafikona Stolpčni. V drugem koraku čarovnika za grafikone na kartici *Niz* določimo, naj bodo oznake na osi X letnice (A5:A90). Excel samodejno nariše grafikon tako, da os X seka os Y pri vrednosti 0.

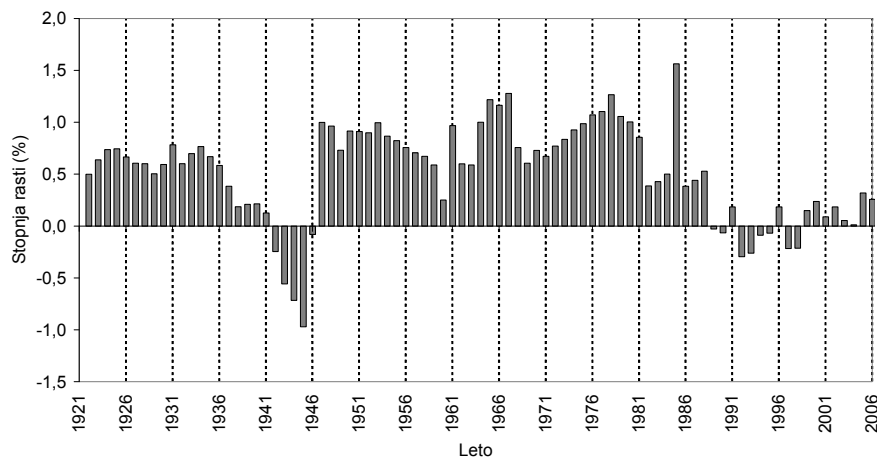
Po oblikovanju grafikona s čarovnikom postavite oznake črtic osi X (letnice) na spodnji rob grafikona z ukazom **Oblikuj os**, kartica *Vzorci*, razdelek *Oznake črtic: Nizko*.

Pokončne mrežne črte nastavite na 5 let narazen, prav tako naj bodo 5 let narazen oznake črtic na osi X. Možnost ***Os vrednosti (y) seka med kategorijami*** naj ne bo izbrana. To dosežete z dodatnim oblikovanjem osi X:



Kako se oblika osi X spremeni, če je možnost ***Os vrednosti (y) seka med kategorijami*** izbrana?

Širina vrzeli med stolpci naj bo 70 (**Oblikuj niz podatkov**, kartica *Možnosti*, polje ***Širina vrzeli***). ♦



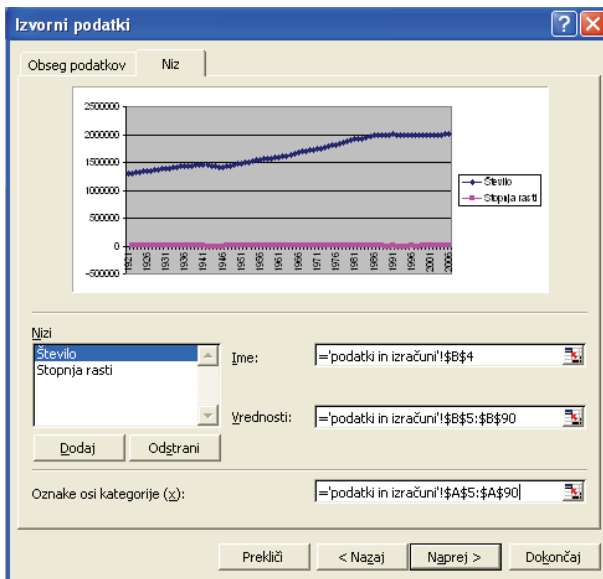
Slika 2.2: Letna stopnja rasti za število prebivalstva na dan 30. junij v Sloveniji v obdobju 1921-2006

PREBIVALSTVO 5

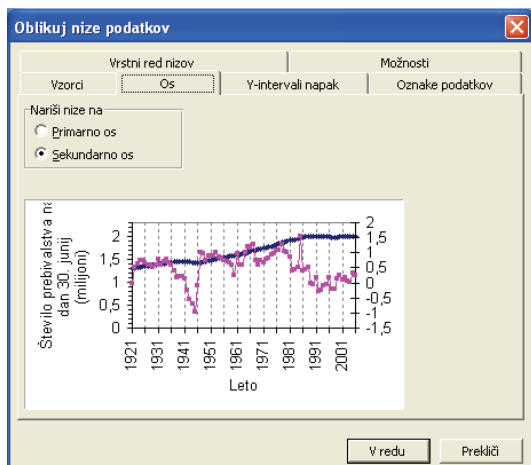
Grafično predstavite število prebivalstva in njegovo letno stopnjo rasti v Sloveniji v obdobju 1921-2006 na istem grafikonu (slika 2.3).

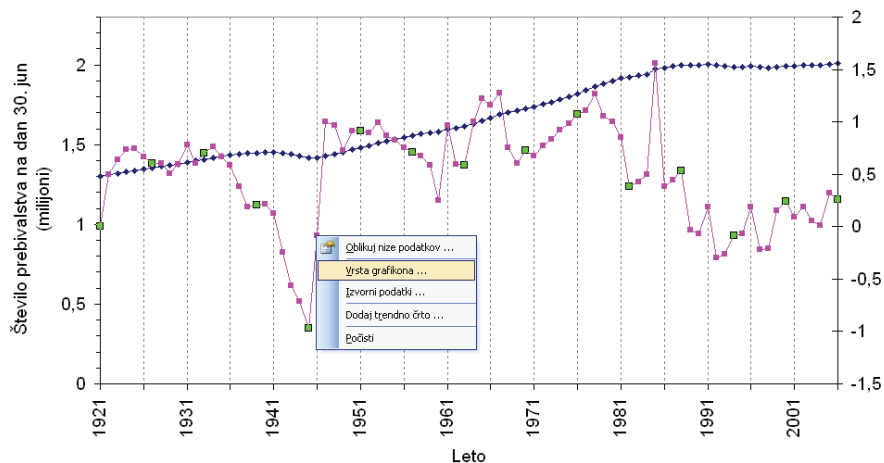
Navodilo: označimo celice s podatki o srednjem številu prebivalstva in o stopnji rasti (B4:B90; D4:D90, celica D5 mora biti prazna) in izberemo črtni grafikon⁸, ki omogoča, da kasneje za prikaz stopnje rasti izberemo stolpce. V drugem koraku čarovnika za grafikone na kartici *Niz* določimo, naj bodo oznake X osi letnice (A5:A90).

⁸ Sicer za grafični prikaz časovne vrste v splošnem uporabljamo grafikon vrste XY (Raztreseni), tu pa zaradi zahteve, da mora biti ena vrsta prikazana s stolpci, uporabimo črtni grafikon.



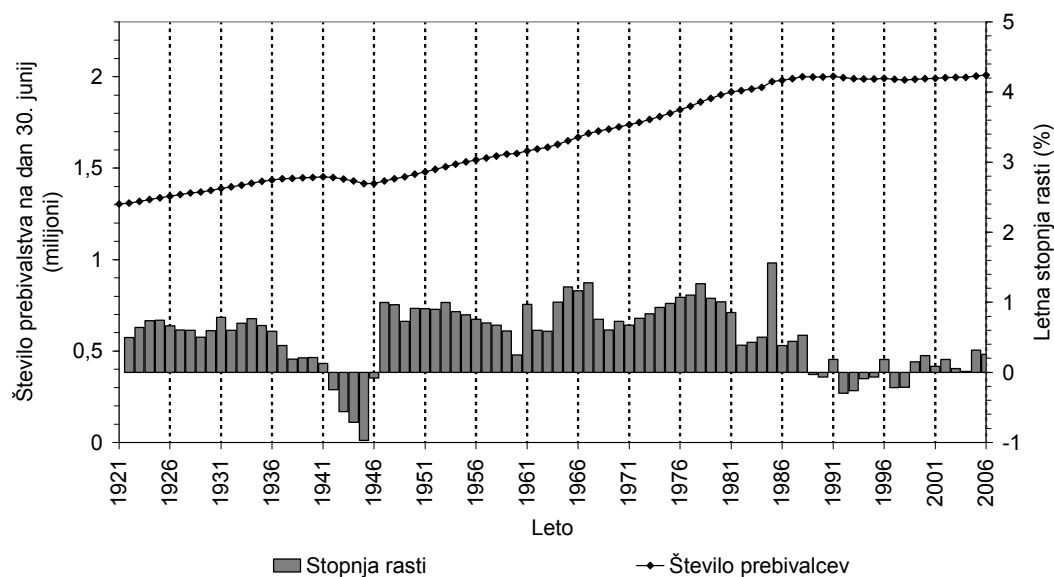
Da se vrednosti za stopnjo rasti pokažejo na sekundarni osi Y, s klikom izberemo drugi niz podatkov, uporabimo ukaz **Oblikuj niz podatkov** in na kartici *Os* izberemo možnost **Sekundarna os**. Črto, ki prikazuje stopnje rasti, spremenimo v stolpec tako, da ta niz najprej izberemo in uporabimo ukaz **Vrsta grafikona** na pomožnem meniju, ki se pokaže ob desnem kliku miške, ter izberemo vrsto **Stolpčni**.





Os X in primarno os Y oblikujete tako, kot kaže slika 2.3.

Na koncu nastavite merilo na sekundarni Y osi na interval od -1 do 5, dodajte ime sekundarne Y osi in zmanjšajte širino vrzeli med stolpci na 20. ♦



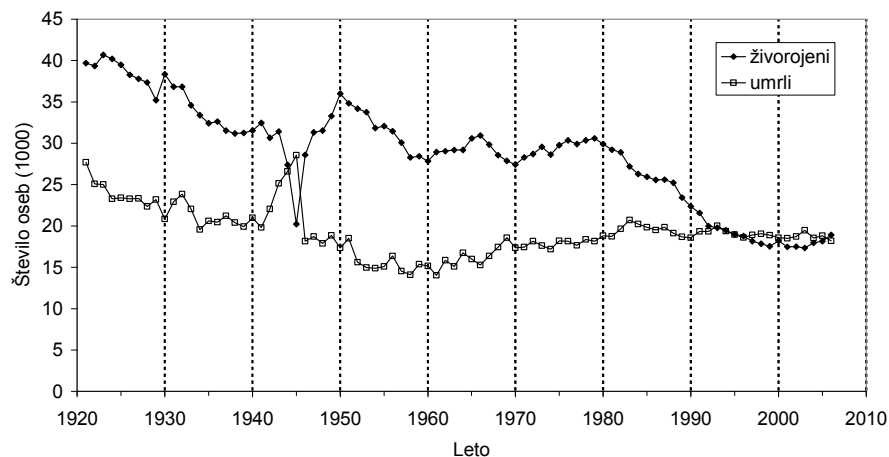
Slika 2.3: Število prebivalstva v Sloveniji na dan 30. junija v obdobju 1921-2006 in pripadajoča stopnja rasti.

PREBIVALSTVO 6

Na istem grafikonu prikažite podatke o številu živorojenih in umrlih (grafični prikaz dveh časovnih vrst).

Navodilo: Če na delovnem listu izberemo tri stolpce podatkov: A4:A90, E4:F90, grafikon vrste XY (Raztreseni) podatke A5:A90 (leta) prikaže na osi X, podatke E5:E90 kot prvi niz in podatke F5:F90 kot drugi niz. Podatek v celicah A4 predstavlja naslov osi X, podatka v celicah E4 in E5 pa predstavljata imeni nizov, ki se pojavita v legendi.

Grafikon oblikujte tako, da bo čimbolj podoben sliki 2.4, list v datoteki pa poimenujte »Slika 2.4«. Bodite pozorni na oblikovanje osi X in Y. ♦



Slika 2.4: Število živorajenih in umrlih v Sloveniji v obdobju 1921-2006.

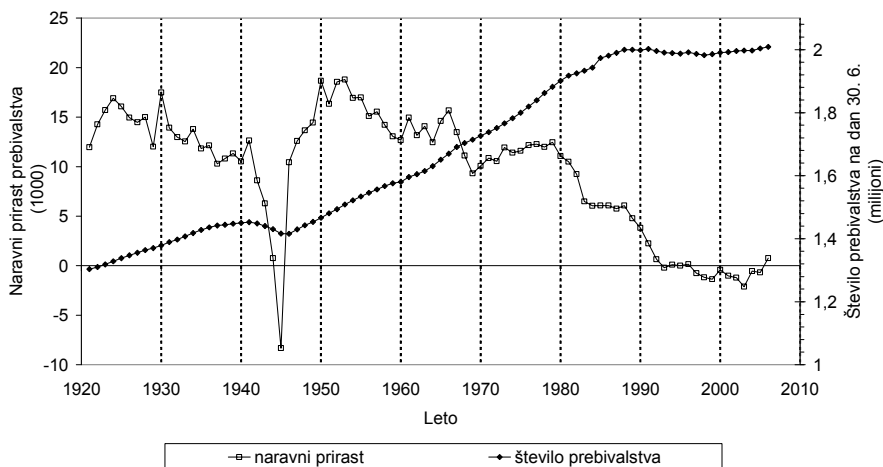
PREBIVALSTVO 7

Izračunajte naravni prirast v posameznih letih. Naravni prirast je razlika med številom živorajenih in številom umrlih. Vrednosti za leti 1945 in 2006 obrazložite. ♦

PREBIVALSTVO 8

Na istem grafikonu predstavite naravni prirast (primarna os Y) in število prebivalstva (sekundarna os Y) v Sloveniji v obdobju 1921-2006 (slika 2.5).

Navodilo: Označimo celice A4:B90 in G4:G90 ter izberimo XY (Raztreseni) grafikon. Vrednosti za število prebivalcev postavimo na sekundarno os Y, po končanem delu s čarovnikom za grafikone z ukazom **Oblikuj nize podatkov** na kartici *Os*, v razdelku *Nariši nize* na izberemo možnost **Sekundarna os**. ♦



Slika 2.5: Naravni prirast in število prebivalcev na dan 30. 6. v Sloveniji po letih (1921-2006)

PREBIVALSTVO 9

Izračunajte verižne indekse in stopnjo rasti za število živorojenih in za število umrlih za obdobje 1921-2006.

Izračunane stopnje rasti vsebinsko smiselno grafično prikažite (dva grafikona na delovnih listih z imeni »Slika 2.2a« in »Slika 2.2b«). ♦

PREBIVALSTVO 10

Za vsako leto izračunajte koeficiente, ki izražajo število rojenih na 1000 prebivalcev (stopnja rodnosti) in število umrlih na 1000 prebivalcev (stopnja umrljivosti).

Stopnjo rodnosti in stopnjo umrljivosti grafično prikažite na istem grafikonu (ime lista »Slika 2.4a«). Sliko primerjajte s sliko 2.4. Obrazložite razlike. ♦

PREBIVALSTVO 11

Izračunajte indekse, ki ponazarjajo, kako se je število prebivalstva spreminjalo glede na leto 1961. Izračune grafično predstavite in obrazložite vrednosti za leti 1945 in 2006 (ime lista z grafikonom naj bo »Slika 2.6«).

PREBIVALSTVO 12

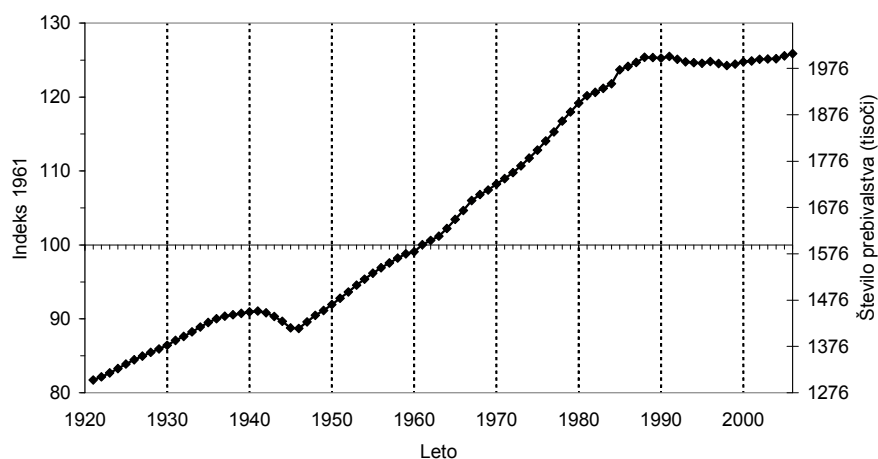
Na sliko, ki prikazuje indeks z osnovo 1961 za število prebivalstva v Sloveniji, dodajte še sekundarno os, ki predstavlja originalne vrednosti - število prebivalstva na dan 30. 6. v tisočih (slika 2.6).

Navodilo: Sekundarne osi Y ne moremo dobiti, če na grafikonu prikazujemo samo en niz podatkov. Zato najprej na grafikon dodajte še en niz - število prebivalstva (Y_2). Skala na osi Y

se samodejno poveča na območje od 0 do 2500000, vidimo niz za število prebivalstva, niz za indekse je blizu ničle. Za število prebivalstva izberemo sekundarno os. Končno moramo uskladiti merili za primarno in sekundarno os tako, da se krivulji na grafikonu prekrijeta. V ta namen najprej določimo minimum in maksimum za primarno os – Y_1 (npr. 80 in 130), nato izračunamo minimum in maksimum za Y_2 , ki ustrežata vrednostima na primarni osi (število prebivalstva v letu 1961 je bilo 1595450):

	Minimum	Maksimum
Primarna os Y_1	80	130
Sekundarna os Y_2	$1595450 * 0,8 = 1276360$	$1595450 * 1,3 = 2074085$
Zaokrožene vrednosti na tisočice	1276000	2074000

Na kartici *Merilo* sekundarne osi Y lahko dobljene vrednosti zaokrožimo na tisočice natančno, da dobimo za oznake črtic cela števila, črti obeh nizov pa sta pri dani grafični natančnosti prekriti. ♦



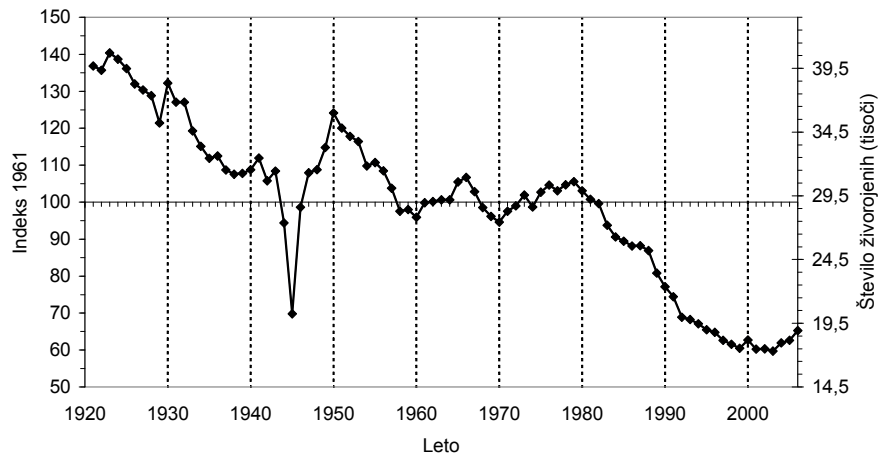
Slika 2.6: Indeksi z osnovo 1961 za število prebivalstva in število prebivalstva na dan 30. 6. v Sloveniji v obdobju 1921-2006

PREBIVALSTVO 13

Izračunajte indekse z osnovo 1961 za število živorojenih in za število umrlih v Sloveniji v danem obdobju. Izračune za leti 1945 in 2006 obrazložite. ♦

PREBIVALSTVO 14

Izračunane indekse z osnovo 1961 za število živorojenih grafično prikažite skupaj z originalnimi podatki, kot kaže slika 2.7. Grafikon postavite na grafični delovni list z imenom »Slika 2.7«. ♦



Slika 2.7: Indeksi z osnovo 1961 za število živorojenih in število živorojenih v Sloveniji v obdobju 1921-2006

PREBIVALSTVO 15

Izračunane indekse z osnovo 1961 za število umrlih grafično prikažite skupaj z originalnimi podatki. Grafikon postavite na grafični delovni list z imenom »Slika 2.7a«. ♦

2.4 Koeficienti in njihovi grafični prikazi, povprečni koeficient

Zgled 2.2: Vpisani študenti in diplomanti na BF

Tabela 2.1: Število redno vpisanih študentov na univerzitetne programe študija v šol. l. 2006/07 in število diplomantov v letu 2007 po smereh študija na Biotehniški fakulteti v Ljubljani (Vir: Univerza v Ljubljani BF, Poročilo za leto 2007)

Študijski program	Število vpisanih 2006/07	Število diplomantov 2007
Biologija	429	78
Gozdarstvo	210	16
Kmetijstvo-Agronomija	368	46
Kmetijstvo-Zootehnika	270	27
Krajinska arhitektura	182	25
Lesarstvo	168	21
Mikrobiologija	267	47
Živilska tehnologija	344	32

VPIS 6

Podatke iz tabele 2.1 dodajte k podatkom iz tabele 1.3 v datoteko VPIS.XLS. Za vsako smer študija izračunajte razmerje število vpisanih v šol. l. 2006/07 na število diplomantov v letu 2007 na BF. Kako imenujemo izračunana števila?

Navodilo: $k_i = \frac{\text{Število vpisanih}_i}{\text{Število diplomantov}_i}$, $i = 1, \dots, 8$ ♦

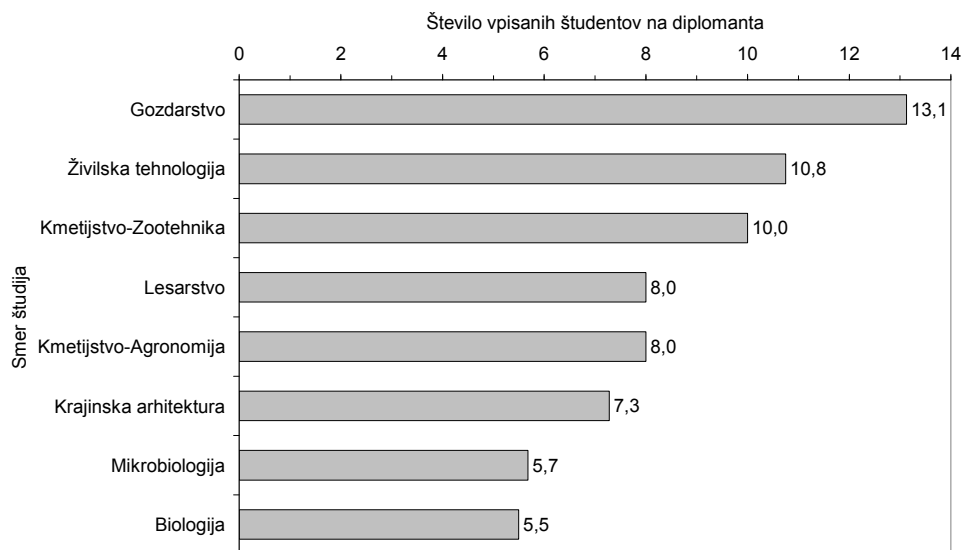
VPIS 7

Grafično prikažite število vpisanih v šol. l. 2006/07 na diplomanta v letu 2007 po smereh študija na BF (slika 2.8).

Navodilo: Najprej tabelo s podatki in izračuni uredite tako, da bo vrstni red smeri študija na BF določen z velikostjo koeficienta v naraščajočem vrstnem redu (ukaz **Podatki/Razvrsti**).

Za grafični prikaz izberite palični grafikon. Os X (v tem primeru grafikona je to vertikalna os) oblikujte tako, da bo izbrana možnost **Os vrednosti (y) seka pri največji kategoriji**.

Ob »palice« naj se izpišejo tudi vrednosti koeficientov (ukaz **Oblikuj nize podatkov**, kartica **Oznake podatkov**, možnost **Vrednosti**). ♦



Slika 2.8: Grafični prikaz koeficientov: število vpisanih v šol. l. 2006/07 na diplomanta v letu 2007

VPIS 8

Izračunajte povprečno število vpisanih študentov v šol. l. 2006/07 na diplomanta v letu 2007 na BF (\bar{k}).

Navodilo:
$$\bar{k} = \frac{\sum_{i=1}^K \check{\text{Število vpisanih}}_i}{\sum_{i=1}^K \check{\text{Število diplomantov}}_i} \quad K = 8 \text{ je število smeri študija na BF} \blacklozenge$$

VPIS 9

Ali bi to povprečje lahko izračunali, če bi imeli na voljo samo podatke o številu vpisanih študentov na diplomanta (samo koeficiente, k_i) za posamezne smeri študija? \blacklozenge

VPIS 10

Na nov delovni list kopirajte podatke za študijsko smer, število vpisanih in koeficiente. Izračunajte povprečni koeficient za BF. S katero sredino ste izračunali to povprečje?

Navodilo:
$$\bar{k} = \frac{\sum_{i=1}^K \check{\text{Število vpisanih}}_i}{\sum_{i=1}^K \frac{\check{\text{Število vpisanih}}_i}{k_i}}$$

to je **tehtana harmonična sredina**. \blacklozenge

VPIS 11

Na nov delovni list kopirajte podatke za študijsko smer, število diplomantov in koeficiente. Izračunajte povprečni koeficient za BF. S katero sredino ste izračunali to povprečje?

Navodilo:
$$\bar{k} = \frac{\sum_{i=1}^K k_i \cdot \check{\text{Število diplomantov}}_i}{\sum_{i=1}^K \check{\text{Število diplomantov}}_i},$$

to je **tehtana aritmetična sredina**. ♦

Domača naloga 2.1: Okrasne rastline (nadaljevanje domače naloge 1.1)

V tabeli 2.2 so podatki o številu vzgojenih okrasnih rastlin in o številu pridelovalcev po vrstah okrasnih rastlin za leti 2003 in 2006 (popis vrtnarstva SURS) v Sloveniji.

Tabela 2.2: Pridelava okrasnih rastlin, Slovenija, 2003 in 2006 (Vir: SURS)

Vrsta rastlin	Vzgojene rastline (1000)		Število pridelovalcev	
	2003	2006	2003	2006
Sobne rastline	1981	925	134	93
Balkonske rastline	6046	6711	306	258
Enoletnice	3687	2843	296	247
Dvoletnice	5059	4302	287	229
Večletne in lesnate rastline	1949	1329	123	92

OKRASNE RASTLINE 3

Podatke o številu pridelovalcev dodajte v datoteko OKRASNE RASTLINE.XLS. Za vsako od vrst okrasnih rastlin izračunajte indekse, na podlagi katerih primerjajte število vzgojenih okrasnih rastlin v letu 2006 glede na leto 2003. Obrazložite izračunane vrednosti. Indekse grafično prikažite. ♦

OKRASNE RASTLINE 4

Za vsako od vrst okrasnih rastlin izračunajte indekse, na podlagi katerih primerjajte število pridelovalcev okrasnih rastlin v letu 2006 glede na leto 2003. Obrazložite izračunane vrednosti. Indekse grafično prikažite. ♦

OKRASNE RASTLINE 5

Za vsako od vrst okrasnih rastlin izračunajte število vzgojenih rastlin (1000) na pridelovalca v letih 2003 in 2006. Obrazložite vrednost za sobne rastline. Kako imenujemo izračunana relativna števila? ♦

Domača naloga 2.2: Gostota prebivalstva v Sloveniji

V datoteki GOSTOTA PREBIVALSTVA.XLS so podatki o številu gospodinjstev in o številu moških in žensk po popisih od leta 1921 dalje na ozemlju Slovenije (tabela 2.3).

GOSTOTA PREBIVALSTVA 1

Izračunajte število prebivalcev za posamezni popis. ♦

GOSTOTA PREBIVALSTVA 2

Izračunajte gostoto prebivalstva za posamezni popis, ob tem pa upoštevajte površino sedanje Slovenije (20273 km²). Izračune grafično prikažite. ♦

Tabela 2.3: Gospodinjstva in prebivalstvo v Sloveniji, popisi 1921 – 2002.

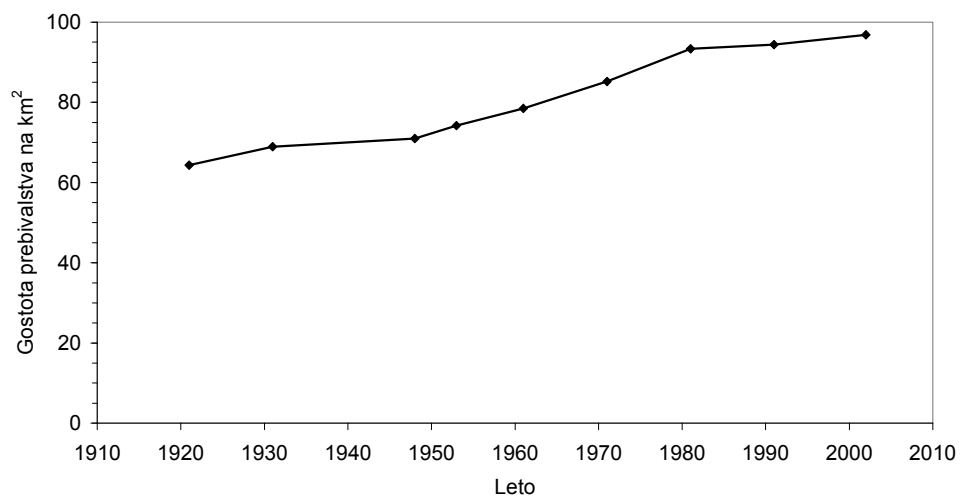
Leto popisa	Število gospodinjstev	Moški	Ženske
1921	...	622168	682632
1931	287228	673248	724402
1948	380950	675353	764447
1953	410976	712034	792393
1961	458853	760770	830753
1971	515531	835998	891139
1981	594571	918766	973098
1991	632278	923643	989712
2002	684847	958576	1005460

GOSTOTA PREBIVALSTVA 3

Izračunajte povprečno število članov gospodinjstva za posamezni popis. ♦

GOSTOTA PREBIVALSTVA 4

Izračunajte število žensk na enega moškega za posamezni popis. ♦



Slika 2.9: Gostota prebivalstva v Sloveniji ob popisih. Upoštevana je današnja površina Slovenije.

3 Opisne statistike, slučajna izbira in frekvenčna porazdelitev

Novi pojmi:

- ukaz **Vstavi/Funkcija**
- orodje **Analiza podatkov/Descriptive statistics**
- slučajna izbira, funkcija **RANDBETWEEN**

Datoteke s podatki:

PREBIVALSTVO.XLS

PADAVINE.XLS

LJUBLJANSKI MARATON.XLS

DEBELINA DEBEL.XLS



KATJA.XLS

Ustvarjene datoteke:

SPLET.XLS

ROJSTNO IME.XLS

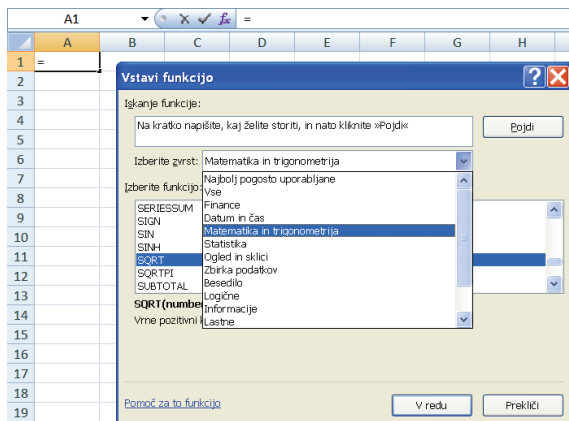
3.1 Uporaba vgrajenih funkcij v Excelu

V Excelu lahko uporabimo različne že vnaprej vgrajene funkcije, ki so razporejene v večje skupine (zvrsti). Ob uporabi take funkcije najprej označimo celico preglednice, v katero se bo izpisal rezultat, nato uporabimo ukaz **Vstavljanje/Funkcija** ali klik na ikono  v vrstici za vnos. Bljižnica za nekaj funkcij je klik na puščico na ikoni , ki odpre meni, v katerem izberemo eno izmed funkcij *Vsota*, *Povprečje*, *Štetje števil*, *Min*, *Max* ali ukaz **Več funkcij**. V pogovornem oknu *Vstavi funkcijo* določimo funkcijo v dveh korakih.

1. korak

V pogovornem oknu sta dva seznama: *Izberite zvrst funkcije* - seznam skupin funkcij in *Izberite funkcijo* - seznam imen funkcij za izbrano zvrst funkcij. Pod njima je napisana oblika za uporabo funkcije in kratko pojasnilo, kaj funkcija izračuna. Najprej s klikom izberemo skupino funkcij in nato še ime zelene funkcije (npr. kvadratni koren najdemo

v skupini funkcij *Matematika in trigonometrija*, ime funkcije je SQRT). S klikom na gumb **V redu** se odpre naslednje pogovorno okno.



2. korak

Vsebina tega okna je odvisna od izbrane funkcije, vedno pa je tu eno ali več polj za vnos argumentov funkcije. Argumenti funkcij so lahko števila, zaporedje števil, naslovi celic, zaporedje naslovov celic, naslovi skupin celic ali pa funkcije. Argumente vpisujemo v polja tako, da postavimo utripajoči kazalec v polje in jih napišemo ali pa označimo z miško ustrezne celice v preglednici.

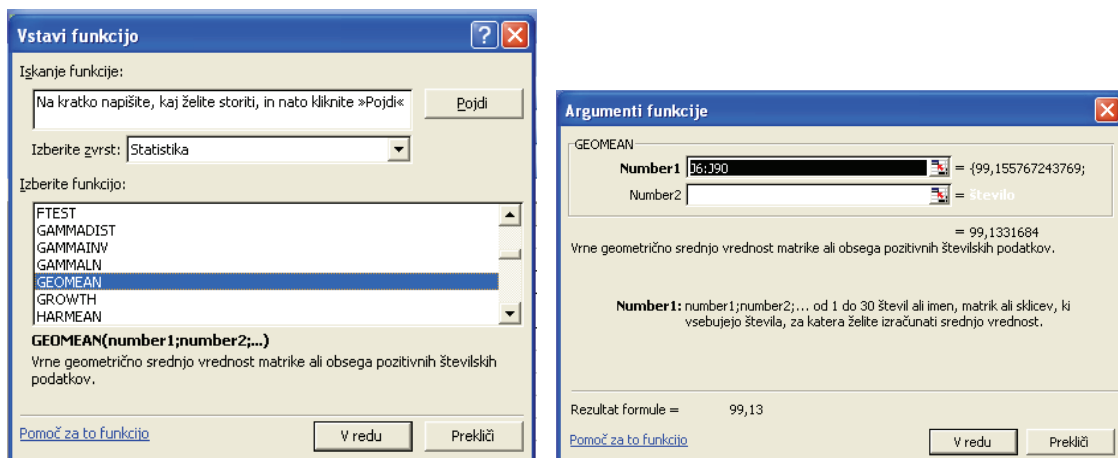
V polju *Rezultat formule* se po vnosu argumentov izpiše rezultat. Z gumbom **V redu** se rezultat pokaže v označeni celici.

3.1.1 Geometrijska sredina

Povprečni verižni indeks (\bar{I}) izračunamo kot geometrijsko sredino verižnih indeksov ($I_{i/i-1}$):

$$\bar{I} = \sqrt[T]{I_{2/1} \cdot I_{3/2} \cdot \dots \cdot I_{T/T-1}}$$

Poglejmo si primer uporabe funkcije, ki v izbrano celico vpiše geometrijsko sredino izbranih števil. Ta funkcija se imenuje GEOMEAN.



87	2003	8,7	9,7	99,0	104,0	-1,0	4,0
88	2004	9,0	9,3	103,7	95,2	3,7	-4,8
89	2005	9,1	9,4	101,1	101,6	1,1	1,6
90	2006	9,4	9,1	104,3	96,6	4,3	-3,4
91				Povprečni verižni indeks:		Povprečna stopnja rasti:	
92				99,13	99,51	-0,87	-0,49
93							

PREBIVALSTVO 16

V datoteki PREBIVALSTVO.XLS izračunajte povprečni letni verižni indeks in povprečno letno stopnjo rasti za število živorojenih in število umrlih v Sloveniji v obdobju 1921-2006.

Navodilo: Povprečno stopnjo rasti (\bar{S}) izračunamo tako, da od povprečnega verižnega indeksa odštejemo 100:

$$\bar{S} = \bar{I} - 100 \quad \blacklozenge$$

3.2 Napake pri uporabi formul

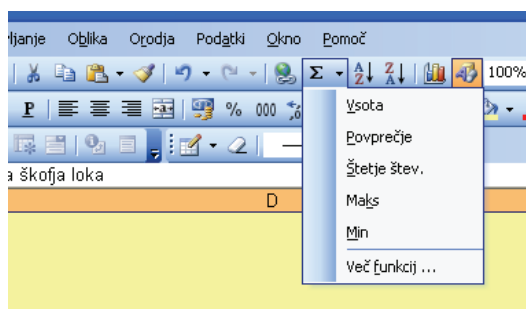
Če pri računanju s formulami naredimo napako, Excel v celico namesto rezultata napiše oznako za vrsto napake npr.:

- #DEL/0! v formuli so argumenti napisani tako, da pride do deljenja z 0;
- #ŠTEV! v formuli smo narobe določili argument (npr. besedilo namesto števila);
- #SKLIC! referenčna celica je neveljavna (npr. pobrisana vsebina referenčne celice);
- #N/V vrednost, ki je navedena kot argument formule, ni na voljo formuli ali funkciji;
- ##### rezultat je predolg, da bi se lahko izpisal v celico. To ni prava napaka. Če ustrezno razširimo stolpec, se bo rezultat izpisal.

3.3 Opisne statistike

3.3.1 Uporaba funkcij

Opisne statistike (mere sredine in mere variabilnosti) Excel izračuna z že vgrajenimi funkcijami, ki jih v pogovornem oknu ukaza **Vstavi funkcijo** najdemo v skupini **Statistika**. Nekatere med njimi najdemo tudi v meniju ob uporabi ikone Σ :



Možnost **Povprečje** izvede funkcijo **AVERAGE** (aritmetična sredina), **Štetje šte.** funkcijo **COUNT** (število podatkov), **Maks** izvede funkcijo **MAX** (maksimum) in **Min** izvede funkcijo **MIN** (minimum). Spodaj je navedenih še nekaj drugih funkcij v skupini **Statistika**⁹, ki jih uporabljamo pri računanju opisnih statistik:

MEDIAN	mediana;	MODE	modus;
STDEV	vzorčni standardni odklon;	STDEVP	populacijski standardni odklon;
VAR	vzorčna varianca;	VARP	populacijska varianca;
HARMEAN	harmonična sredina;		
QUARTILE	minimalna vrednost, prvi kvartil, mediana, tretji kvartil, maksimalna vrednost;		
PERCENTILE	k -ti percentil, k je vrednost iz intervala od 0 do 1.		

⁹ Izračun kvantilov (QUARTILE in PERCENTILE) je v Excelu malo drugačen, kot je predstavljeno v knjigi Uporabna statistika (Košmelj, 2007). Razlike so pri izračunavanju kvantilnega (relativnega) ranga: $R = P \cdot n + 0,5$ (Košmelj), $R = P \cdot (n - 1) + 1$ (Excel). V splošnem so razlike minimalne.

Zgled 3.1: Anketa - število ur deskanja na svetovnem spletu

Spodaj so navedeni podatki o številu ur deskanja po svetovnem spletu na teden (*DSS*) za 14 študentov 2. letnika Biotehnologije in Mikrobiologije v šol. l. 2006/07.

10 7 14 20 2 2 5 2 7 4 0 7 5 15

SPLET 1

Podatke o *DSS* prepišite v novo datoteko z imenom SPLET.XLS. ♦

SPLET 2

Izračunajte opisne statistike za *DSS*: minimum, maksimum, modus, mediano in povprečje. Izračuni naj imajo smiselno število decimalnih mest (tabela 3.1). ♦

SPLET 3

Izračunajte še prvi in tretji kvartil za *DSS* (tabela 3.1). Obrazložite izračune.

Navodilo: uporabite funkcijo QUARTILE(array; quart), ki ima dva argumenta: »array« je obseg podatkov in »quaer« je številka kvartila. Obseg podatkov je naslov celic s podatki, za katere računamo kvartile (npr. A2:A15), za številke kvartilov pa velja: 0 - minimum, 1 - prvi kvartil, 2 - mediana, 3 - tretji kvartil in 4 - maksimum. ♦

SPLET 4

Ali v podatkih o *DSS* obstajajo osamelci?

Navodilo: spodnjo (l_1) in zgornjo (l_2) mejo za osamelce izračunamo po formulah:

$$l_1 = Q_1 - \frac{3}{2}Q \qquad l_2 = Q_3 + \frac{3}{2}Q \qquad Q = Q_3 - Q_1 \text{ je kvartilni razmik (tabela 3.1).}$$

SPLET 5

Izračunajte centila, s katerima se ranžirna vrsta za podatke razdeli na tri enake dele. Obrazložite rezultate (tabela 3.1).

Navodilo: uporabite funkcijo PERCENTILE(array; k), »array« je obseg podatkov, ki ga določimo enako kot pri funkciji QUARTILE, »k« pa je relativni rang ustreznega centila. ♦

Tabela 3.1: Izračuni opisnih statistik za spremenljivko DSS; izpis formul¹⁰ (levo); izpis rezultatov (desno)

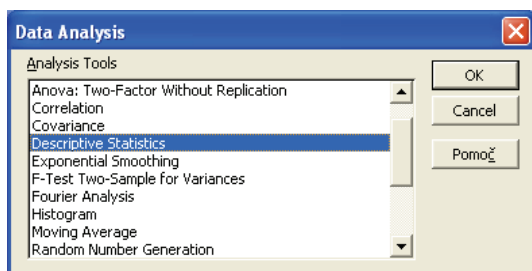
	A	B	C	D		A	B	C	D
1	DSS		Minimum	=MIN(A2:A15)	1	DSS		Minimum	0
2	10		Maksimum	=MAX(A2:A15)	2	10		Maksimum	20
3	7		Modus	=MODE(A2:A15)	3	7		Modus	7
4	14		Mediana, Q ₂ = Me	=MEDIAN(A2:A15)	4	14		Mediana, Q ₂ = Me	7
5	20		Povprečje	=AVERAGE(A2:A15)	5	20		Povprečje	7,5
6			Q ₁	=QUARTILE(A2:A15;1)	6			Q ₁	4
7	2		Q ₂ = Me	=QUARTILE(A2:A15;2)	7	2		Q ₂ = Me	7
8	5		Q ₃	=QUARTILE(A2:A15;3)	8	5		Q ₃	10
9	2		Q	=D8-D6	9	2		Q	6
10	7		Meja za sp. osamelce	=D6-1,5*D9	10	7		Meja za sp. osamelce	-5,00
11	4		Meja za zg. osamelce	=D8+1,5*D9	11	4		Meja za zg. osamelce	19,00
12	0		C _{33,3}	=PERCENTILE(A2:A15;1/3)	12	0		C _{33,3}	5,0
13	7		C _{66,7}	=PERCENTILE(A2:A15;2/3)	13	7		C _{66,7}	7,0
14	5		Varianca	=VAR(A2:A15)	14	5		Varianca	33,6
15	15		Standardni odklon	=STDEV(A2:A15)	15	15		Standardni odklon	5,8
16			KV%	=D15/D5*100	16			KV%	76,9

SPLET 6

Za DSS izračunajte še varianco (s^2), standardni odklon (s) in koeficient variacije ($KV\%$) (tabela 3.1). ♦

3.3.2 Uporaba orodja Descriptive Statistics

Ogledali si bomo uporabo orodja **Descriptive Statistics**, ki združuje več funkcij. V splošnem orodju, ki jih uporabljamo pri statistični analizi podatkov, najdemo v meniju **Orodja** pod ukazom **Analiza podatkov**¹¹.

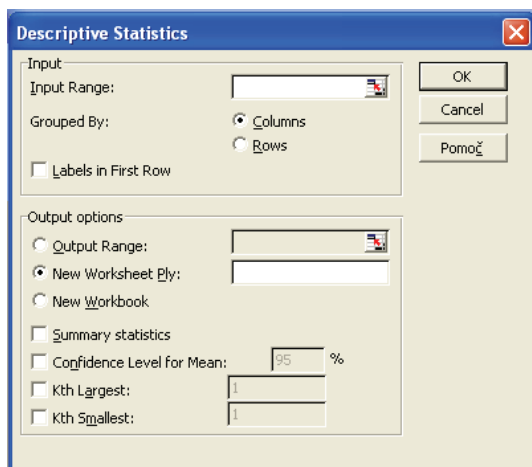


V pogovornem oknu **Descriptive Statistics**¹² določimo položaj in obliko vhodnih podatkov (vrednosti izbranih spremenljivk) in excelovih izračunov v zvezku:

¹⁰ Izpis formul dosežemo s tem, da v meniju **Orodja** uporabimo ukaz **Možnosti** in na kartici *Pogled* izberemo možnost **Formule**.

¹¹ Če ukaza *Analiza podatkov* na meniju *Orodja* ni, uporabite ukaz *Orodja/Dodatki* in izberite možnost *Orodja za analizo*. Ukaz je v meniju naslednjič, ko odprete Excel.

¹² Ukazi v pogovornem oknu niso prevedeni v slovenščino.



- v polje ***Input Range (Vhodni obseg)*** vpišemo skupino celic, v kateri so vrednosti spremenljivk, za katere računamo opisne statistike. Skupina celic se v to polje izpiše, če podatke označimo z miško;
- v polju ***Grouped By (Združeno po)*** določimo, ali so spremenljivke v stolpcih (***Columns***) ali v vrsticah (***Rows***).
- v polju ***Labels in First Row (Oznake v prvi vrstici)*** določimo, ali so v prvi vrstici skupine celic, določene v polju ***Input Range (Vhodni obseg)***, napisane oznake za spremenljivke.

Mesto izpisa opisnih statistik določimo z izbiro ene izmed treh možnosti v razdelku ***Output Options (Izhodne možnosti)***:

- možnost ***Output Range (Izhodni obseg)*** izberemo, če želimo rezultate izpisati na eno izmed že obstoječih strani v preglednici. V polje napišemo naslov celice, ki predstavlja levi zgornji vogal izpisa. To celico lahko označimo z miško in njen naslov se napiše v polje;
- možnost ***New Worksheet Ply (Nov delovni list)*** izberemo, če želimo rezultate izpisati na nov delovni list. V polje napišemo ime novega delovnega lista, saj sicer dobi nov list naslov Listxx. Pri tem xx pomeni številko, ki je odvisna od tega, koliko delovnih listov imamo že v zvezku (datoteki);

- možnost *New Workbook (Nov delovni zvezek)* izberemo, če želimo rezultate izpisati v novo datoteko. V polje napišemo ime datoteke.



Katere opisne statistike naj se izpišejo, določimo v razdelku *Output options (Izhodne možnosti)*:

- možnost *Summary statistics (Statistični povzetek)* izberemo, če želimo izpis vseh opisnih statistik. Brez te možnosti Excel izračuna samo odklon intervala zaupanja za povprečje.
- v polju *Confidence Level for Mean (Stopnja zaupanja za povprečje)* določimo stopnjo zaupanja za interval zaupanja za aritmetično sredino. Po osnovni nastavitvi je to 95 %;
- v poljih *Kth Largest (K-ti največji)* in *Kth smallest (K-ti najmanjši)* je po osnovni nastavitvi napisana številka 1, kar pomeni, da se bosta med opisnimi statistikami izpisali maksimalna in minimalna vrednost. Če izberemo številko 2, se bosta izpisali druga najmanjša in druga največja vrednost.

SPLET 7

Izračunajte opisne statistike za *DSS* z uporabo orodja **Descriptive statistics**.

Navodilo: uporabite ukaz **Orodja/Analiza podatkov/Descriptive statistics**. Postavite kazalec miške v polje *Input Range* in v preglednici označite stolpec s podatki. Izberite možnost **Grouped By Columns. Labels in First Row** izberite glede na to, kako ste določili skupino celic v polju *Input Range*. Izberite možnost **Summary Statistics**. Določite, naj se rezultati izpišejo na nov delovni list z imenom »Opisne statistike«. ♦

V preglednico se na novo stran izpišejo rezultati, ki jih vidimo v prvih dveh stolpcih tabele 3.2. V tretji stolpec preglednice 3.2 smo dodali slovenska imena opisnih statistik. Rezultati se po osnovni nastavitvi izpišejo na toliko decimalnih mest, kot so široke celice. To ponavadi ni smiselno. Število decimalnih mest določimo z ukazom **Oblika/Celice, Številke** oz. z uporabo ikone  (zmanjšanje) ali  (povečanje).

SPLET 8

Določite, naj se vsi rezultati, razen variacijskega razmika, minimuma, maksimuma, vsote in števila enot v vzorcu, izpišejo na eno decimalno mesto. ♦

Tabela 3.2: Opisne statistike za število ur deskanja po svetovnem spletu za študente 2. letnika BT in MB v šol. l. 2006/07 (DSS)

DSS		Razlaga izračunov
Mean	7,1	aritmetična sredina
Standard Error	1,5	standardna napaka povprečja
Median	6	mediana
Mode	7	modus
Standard Deviation	5,8	vzorčni standardni odklon
Sample Variance	33,2	vzorčna varianca
Kurtosis	0,4	sploščenost porazdelitve
Skewness	1,0	asimetričnost porazdelitve
Range	20	variacijski razmik
Minimum	0	minimalna vrednost
Maximum	20	maksimalna vrednost
Sum	100	vsota vrednosti
Count	14	število vrednosti
Confidence Level(95,0%)	3,3	odklon intervala zaupanja

Domača naloga 3.1: Število dni s padavinami

Za meteorološko postajo Ljubljana Bežigrad imamo podatke o številu dni s padavinami v obdobju 1951-2005 v datoteki PADAVINE.XLS.

PADAVINE 1

Izračunajte opisne statistike za število dni s padavinami v Ljubljani v obdobju 1951-2005. ♦

PADAVINE 2

Določite, naj se vsi rezultati, razen variacijskega razmika, minimuma, maksimuma, vsote in števila enot v vzorcu, izpišejo na eno decimalno mesto. ♦

PADAVINE 3

Izračunajte še prvi in tretji kvartil za število dni s padavinami v obdobju 1951-2005. Obrazložite izračune. ♦

PADAVINE 4

Ali v podatkih o številu dni s padavinami v Ljubljani obstajajo osamelci? ♦

PADAVINE 5

Izračunajte centila, s katerima se ranžirna vrsta za podatke razdeli na tri enake dele. Obrazložite rezultate. ♦

PADAVINE 6

Izračunajte opisne statistike še za podatke o številu dni s padavinami za prvo in zadnje tridesetletno obdobje, 1951-1980 in 1976-2005. Primerjajte rezultate. ♦

3.4 Slučajna izbira

Zgled 3.2: Maraton

Datoteka LJUBLJANSKI MARATON.XLS vsebuje podatke za 403 moških udeležencev, ki so bili uvrščeni na Ljubljanskem maratonu 2005.

MARATON 1

Za vsakega maratonca na podlagi podatka o letu rojstva (*LR*) izračunajte *starost* tekačev v letu 2005. Za starost tekačev:

- izračunajte dve meri sredine in ju vsebinsko obrazložite;
- izračunajte dve meri variabilnosti in ju vsebinsko obrazložite.

Ali v podatkih obstajajo osamelci? Izračunajte mejo za spodnje/zgornje osamelce. ♦

MARATON 2

Za vsakega udeleženca preračunajte njegov čas teka (*Netto*) v minute (sekunde zanemarite), novo spremenljivko pa poimenujte *čas teka*. Tako kot spremenljivko *starost* analizirajte tudi *čas teka*.

Navodilo: Za izračun spremenljivke *čas teka* uporabite funkciji, ki ju najdete v skupini funkcij *Datum in čas*, HOUR in MINUTE: HOUR(G2)*60+MINUTE(G2). V primeru, da se rezultat formule izpiše v obliki 00:00:00, uporabite ukaz **Oblika/Celice** in na kartici *Številke* izberite *Splošno* zvrst.

MARATON 3

Primerjajte variabilnost *starosti* in *časa teka*.

Navodilo: Za vsako spremenljivko posebej izračunajte koeficient variabilnosti. ♦

MARATON 4

Grafično prikažite odvisnost časa teka od starosti.

Navodilo: izberite XY (Raztreseni) grafik, na osi Y je čas teka in na osi X je starost. ♦

MARATON 5

Analiziramo vzorec velikosti 15. Koliko vzorcev velikosti 15 lahko naredimo iz populacije velikosti 403 (brez ponavljanja in če vrstni red ni pomemben)? S katero funkcijo s področja kombinatorike se to izračuna?

Navodilo: Poiščite ustrezno funkcijo v zvrsti *Matematika in trigonometrija* in jo uporabite. ♦

MARATON 6

S sistematično izbiro izberite vzorec velikosti 15.

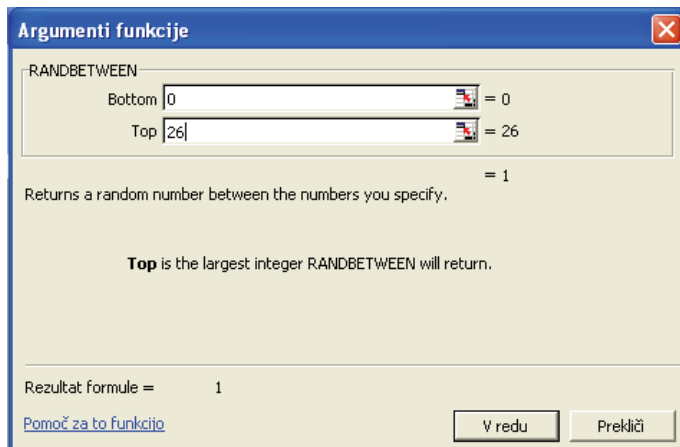
Navodilo: Pri sistematičnem vzorčenju najprej izračunamo korak vzorčenja in vrednost ustrezno zaokrožimo:

$$\text{Korak} = \text{ROUND}(403/15;0) = 27$$

Začetek vzorčenja med 0 in 26 določimo po principu slučajnosti. V Excelu si vzorčenje olajšamo tako, da najprej izračunamo vrednosti nove spremenljivke z imenom *izbor*, ki predstavlja ostanek deljenja zaporedne številke tekača (*Zap.št.*) s korakom vzorčenja 27. Uporabimo funkcijo MOD, ki vrne ostanek po deljenju s 27 (primer: $\text{MOD}(A2; 27) = 1$, $\text{MOD}(A28;27) = 0$):

	A	B	C	D	E	F	G	H	I	J	K
1	Zap št.	Štarba št.	Priimek	Ime	LR	Država	Netto	Bruto	starost	čas teka	izbor
2	1	5	NJOROGE	SAMUEL	1977	KEN	02:15:45	02:15:47	28	135	1
3	2	7	KIPCHUMBA	JOSHUA	1968	KEN	02:17:47	02:17:49	37	137	2
4	3	9	RUTTO	JOSEPH	1977	KEN	02:18:24	02:18:26	28	138	3
5	4	4	KIPKIRUI	DAVID	1977	KEN	02:18:54	02:18:56	28	138	4
6	5	8	BANDAWE	PHILLIP	1973	ZIM	02:21:28	02:21:30	32	141	5
7	6	2	KEŽAR	ROMAN	1966	SLO	02:22:08	02:22:08	39	142	6
8	7	6	NGENY	DAVID	1981	KEN	02:23:49	02:23:51	24	143	7
9	8	1	NSHIMIRIMANA	JOACHIM	1973	BDI	02:24:12	02:24:15	32	144	8
10	9	355	VENCELJ	TONI	1972	SLO	02:38:40	02:38:42	33	158	9
11	10	199	ČAUŠEVIČ	SADET	1971	BIH	02:44:22	02:44:26	34	164	10
12	11	438	ILAR	STANISLAV	1963	SLO	02:44:40	02:44:43	42	164	11
13	12	380	SAJEVEC	PETER	1976	SLO	02:47:37	02:47:38	29	167	12
14	13	279	ZUPANC	MILAN	1963	SLO	02:48:22	02:48:25	42	168	13
15	14	755	GOLOB	IVAN	1954	SLO	02:49:37	02:49:41	51	169	14
16	15	763	KLOBUČARIČ	DAMIR	1981	CRO	02:49:54	02:50:10	24	169	15
17	16	744	MARAŽ	KLAVDIJ	1962	SLO	02:50:20	02:50:26	43	170	16
18	17	156	KURMANŠEK	SEBASTJAN	1975	SLO	02:51:33	02:51:40	30	171	17
19	18	695	ERJAVEC	ALEŠ	1975	SLO	02:53:24	02:53:27	30	173	18
20	19	742	KOBE	PRIMOŽ	1981	SLO	02:53:17	02:53:32	24	173	19
21	20	565	ROBIČ	ANDREJ	1976	SLO	02:53:55	02:54:00	29	173	20
22	21	405	REBOLJ	JERNEJ	1977	SLO	02:54:33	02:54:38	28	174	21
23	22	277	WEBER	THOMAS	1969	CRO	02:54:45	02:54:52	36	174	22
24	23	354	ERJAVEC	ANTON	1958	SLO	02:55:06	02:55:07	47	175	23
25	24	737	MARINIČ	MILOŠ	1953	SLO	02:55:30	02:55:34	52	175	24
26	25	115	FERFILA	STANE	1951	SLO	02:55:29	02:55:39	54	175	25
27	26	254	KLENOVŠEK	TOMAŽ	1968	SLO	02:55:44	02:55:49	37	175	26
28	27	701	GAVEZ	DAVORIN	1966	SLO	02:56:47	02:56:51	39	176	0
29	28	462	PANČUR	ANTON	1963	SLO	02:57:11	02:57:17	42	177	1
30	29	323	GREET	MARK	1967	GBR	02:57:45	02:57:56	38	177	2
31	30	439	KRAŠOVEC	DUŠAN	1967	SLO	02:58:39	02:58:54	38	178	3

Začetek vzorčenja določa slučajno izbrano število na intervalu od 0 do 26. Slučajno izbiro števila naredimo z uporabo funkcije `RANDBETWEEN(bottom; top)`, ki jo najdemo na seznamu, če za zvrst funkcije izberemo skupino *Vse*:



Nato uporabimo samodejni filter, da izberemo zapise z vrednostjo spremenljivke *izbor* enako vrednosti, ki jo je vrnila funkcija `RANDBETWEEN`. Dobljene podatke kopirajte na nov delovni list »vzorec«. ♦

MARATON 7

Na izbranem vzorcu tekačev ponovite analizo *starosti* in *časa teka*. ♦

MARATON 8

Za *starost* in *čas teka* primerjajte povprečje za populacijo in za vzorec. Izračunajte absolutno in relativno napako vzorčne ocene.

Navodilo: absolutno napako izračunamo z izrazom $|\bar{x} - \mu|$ in relativno napako izračunamo z izrazom $|\bar{x} - \mu|/\mu$. ♦

MARATON 9

Primerjajte variabilnost *starosti* in *časa teka* v vzorcu. ♦

MARATON 10

Grafično prikažite odvisnost časa teka od starosti za podatke iz vzorca. ♦

3.5 Frekvenčna porazdelitev in kvantili

Zgled 3.3: Debelina debel dreves

V tabeli 3.3 in v datoteki DEBELINA DEBEL.XLS so podatki o številu dreves v posameznem razredu glede na debelino (premer) debla.

DEBELINA DEBEL 1

Razrede v tabeli 3.3 dopolnite z njihovimi karakteristikami (spodnja (x_{\min}), zgornja meja (x_{\max}), sredina (x_i) in širina (d_i)).

Za vsak razred izračunajte relativno frekvenco ($f_i\%$), kumulativno frekvenco (F_i) in kumulativno relativnih frekvenc ($F_i\%$). Izračune oblikujte v preglednico, kot kaže tabela 3.3. Obrazložite vrednosti za razred številka 5.

Tabela 3.3: Število dreves glede na debelino debla

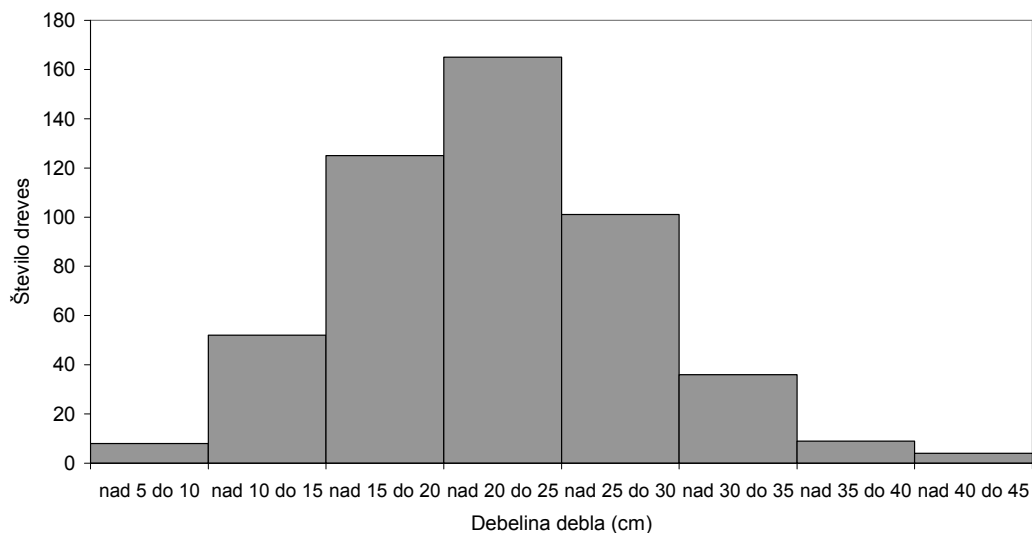
Razred	Debelina debla (cm)	Število dreves	Spodnja meja	Zgornja meja	Sredina razreda	Širina razreda	Relativna frekvenca	Kumulativna frekvenc	Kumulativna rel. frekvenc
i		f_i	x_{\min}	x_{\max}	x_i	d_i	$f_i\%$	F_i	$F_i\%$
1	nad 5 do 10	8							
2	nad 10 do 15	52							
3	nad 15 do 20	125							
4	nad 20 do 25	165							
5	nad 25 do 30	101							
6	nad 30 do 35	36							
7	nad 35 do 40	9							
8	nad 40 do 45	4							

DEBELINA DEBEL 2

Ali je potrebno za vsak razred izračunati gostoto frekvence? Utemeljite odgovor. ♦

DEBELINA DEBEL 3

Grafično prikažite podatke iz tabele 3.3 s histogramom (slika 3.1). ♦



Slika 3.1: Histogram za debelino debla za 500 dreves

DEBELINA DEBEL 4

Grafično prikažite podatke iz tabele 3.3 s poligonom (slika 3.2).

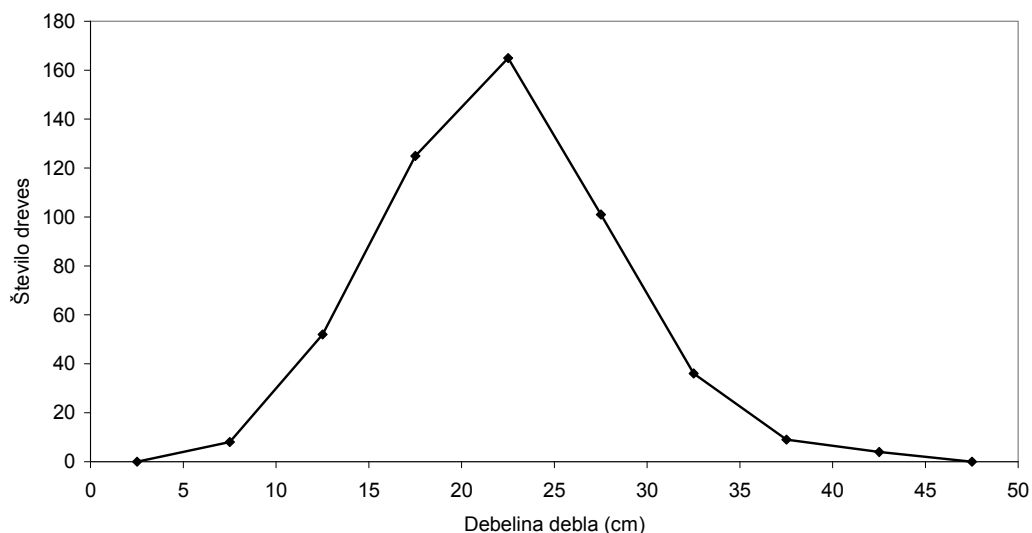
Navodilo: Za grafični prikaz poligona moramo v tabelo dodati razred z zaporedno številko 0 in razred z zaporedno številko 9:

Razred	Debelina debla (cm)	Število dreves	Spodnja meja	Zgornja meja	Sredina razreda	Širina razreda	Relativna frekvenca	Kumulativna frekvenc	Kumulativna rel. frekvenc
i		f_i	x_{\min}	x_{\max}	x_i	d_i	$f_i\%$	F_i	$F_i\%$
0	nad 0 do 5	0	0	5	2,5	5	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
9	nad 45 do 50	0	45	50	47,5	5	0	500	100

Za risanje poligona uporabimo grafikon vrste XY (Raztreseni). Na osi X so sredine razredov, na osi Y so frekvence.

Preden začnemo delo s čarovnikom za grafikone označimo samo stolpec s frekvencami (podatki za os Y), nato v drugem koraku čarovnika za grafikone na kartici *Niz*, v polju **Vrednosti X** izberemo sredine razredov (vrednosti za os X). Tak postopek je potreben, ker XY (Raztreseni)

grafikon vzame podatke iz levega označenega stolpca za os X in iz desnega označenega stolpca za os Y. Vrstni red stolpcev v tabeli 3.3 tako ne ustreza excelovim zahtevam za risanje poligona. ♦



Slika 3.2: Poligon za debelino debla za 500 dreves





DEBELINA DEBEL 5

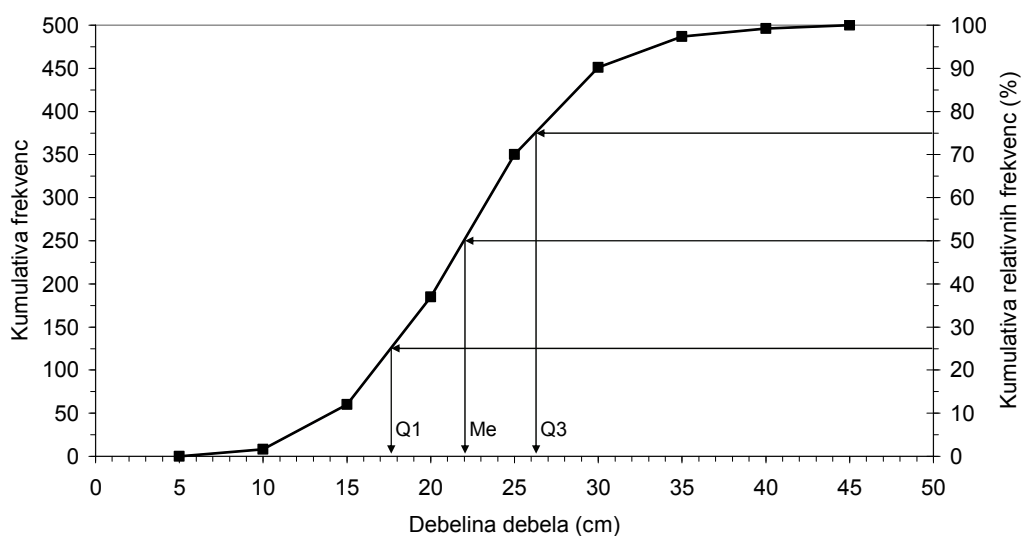
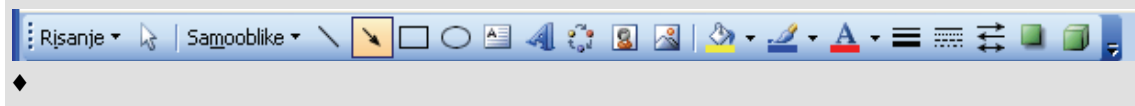
Grafično predstavite kvartile za debelino dreves (slika 3.3).

Navodilo: Najprej moramo grafično prikazati kumulativno relativnih frekvenc (grafični prikaz se imenuje ogiva). Ogivo narišemo tako, da bo na primarni osi Y skala za kumulativno frekvenc, na sekundarni osi Y pa skala za kumulativno relativnih frekvenc (slika 3.3).

Preden kliknemo na ikono čarovnika za grafikone, označimo podatke za zgornje meje razredov ($x_{i\max}$), ki določajo vrednosti na osi X ter dva niza podatkov: kumulativno frekvenc (F_i) in kumulativno relativnih frekvenc ($F_i\%$). Izberemo vrsto grafikona XY (*Raztreseni*). Dobimo dve lomljeni črti, eno za F_i in drugo za $F_i\%$ (ta je blizu osi X, saj prikazuje vrednosti, ki so v absolutnem pomenu ≤ 100).

Dokončno oblikovanje grafikona naredimo izven čarovnika (ukazi, ki jih dobimo z desnim klikom miške na posamezen del grafikona): nizu $F_i\%$ določimo skalo na sekundarni osi Y in ustrezno popravimo merila obeh osi Y (500 dreves predstavlja 100 %). S tem dosežemo, da se črti, ki prikazujeta omenjena dva niza podatkov, prekrivata.

Za grafični prikaz kvartilov uporabimo ikoni *Puščica*  in *Polje z besedilom*  v orodni vrstici Risanje¹³; klik na ikono  in poteg z miško na grafikonu nariše puščico kot samostojen objekt; klik na ikono  in poteg z miško na področju grafikona nanj postavi prazno polje za besedilo, v katerega napišemo oznako posameznega kvartila, za kar potrebujemo malo ročne spretnosti. Za večjo natančnost pred tem primerno oblikujemo merilo na oseh X in Y (slika 3.3).




Slika 3.3: Kumulativna frekvenc (F) in kumulativna relativnih frekvenc ($F\%$) za debelino debela za 500 dreves v tabeli 3.3. Grafični prikaz kvartilov (Q_1 , $Q_2 = Me$, Q_3)

DEBELINA DEBEL 6

Izračunajte približke za vse tri kvartile debeline debela. Vsebinsko obrazložite dobljene vrednosti.

Navodilo:
$$Q_i = x_{0\min} + \frac{F(Q_i) - F(x_{0\min})}{f_0} d_0,$$

¹³ Če orodne vrstice Risanje ni na zaslonu, kliknemo na ikono  v standardni orodni vrstici poleg ikone čarovnika za grafikone.

Q_i je izbrani kvantil (Q_1 , ali Q_2 ali Q_3), $x_{0\min}$ je spodnja meja razreda, v katerem je izbrani kvantil (»ničelni razred«), $F(Q_i)$ je rang izbranega kvantila, $F(x_{0\min})$ je kumulativa frekvenca pri $x_{0\min}$ (razreda pred »ničelnim« razredom), f_0 je frekvenca in d_0 širina »ničelnega razreda«.

Izračune naredite najprej z računalom. Nato izračune ponovite v excelovi datoteki v pripravljeno tabeli. ♦

DEBELINA DEBEL 7

Na delovni list z imenom »Statistike« kopirajte frekvence (f_i) in sredine razredov (x_i) iz tabele 3.3. Izračunajte približke za povprečno debelino debla (\bar{x}) za varianco debeline debla (s^2) ter za standardni odklon debeline debla (s).

Navodilo:

$$\bar{x} = \frac{1}{\sum_{i=1}^K f_i} \sum_{i=1}^K f_i x_i$$

K je število razredov, x_i je sredina in f_i frekvenca i -tega razreda

$$s = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^K f_i (x_i - \bar{x})^2 \right)}$$

Potenco 2 (npr. $(x_i - \bar{x})^2$) v Excelu v formuli napišemo tako, da istočasno pritisnemo tipki **AltGr** in tipko s številko **3**. Tako izračunamo za prvi razred izraz $f_i \cdot (x_i - \bar{x})^2$ v celici F3 s formulo =C3*(D3-\$C\$13)^2.

Kvadratni koren v Excelu izračunamo s funkcijo SQRT: v celico C15 napišemo formulo =SQRT(C14).

Tabela 3.4: Izračun približka za povprečje, varianco in standardni odklon za podatke iz frekvenčne tabele; izpis formul (zgoraj); izpis rezultatov (spodaj)

	A	B	C	D	E	F
1	Razred	Debelina debla (cm)	Število dreves	Sredina razreda	Pomožni izračuni	
2			f_i	x_i	$f_i * x_i$	$f_i * (x_i - x_p)^2$
3	1	nad 5 do 10	8	7,5	=C3*D3	=C3*(D3-\$C\$13)^2
4	2	nad 10 do 15	52	12,5	=C4*D4	=C4*(D4-\$C\$13)^2
5	3	nad 15 do 20	125	17,5	=C5*D5	=C5*(D5-\$C\$13)^2
6	4	nad 20 do 25	165	22,5	=C6*D6	=C6*(D6-\$C\$13)^2
7	5	nad 25 do 30	101	27,5	=C7*D7	=C7*(D7-\$C\$13)^2
8	6	nad 30 do 35	36	32,5	=C8*D8	=C8*(D8-\$C\$13)^2
9	7	nad 35 do 40	9	37,5	=C9*D9	=C9*(D9-\$C\$13)^2
10	8	nad 40 do 45	4	42,5	=C10*D10	=C10*(D10-\$C\$13)^2
11		Vsota	=SUM(C3:C10)		=SUM(E3:E10)	=SUM(F3:F10)
12						
13		približek za povprečje x_{povp}	=E11/C11			
14		približek za varianco s^2	=F11/(C11-1)			
15		približek za standardni odklon s	=SQRT(C14)			

	A	B	C	D	E	F
1	Razred	Debelina debla (cm)	Število dreves	Sredina razreda	Pomožni izračuni	
2			f_i	x_i	$f_i * x_i$	$f_i * (x_i - x_p)^2$
3	1	nad 5 do 10	8	7,5	60	1712,30
4	2	nad 10 do 15	52	12,5	650	4822,32
5	3	nad 15 do 20	125	17,5	2188	2679,61
6	4	nad 20 do 25	165	22,5	3713	22,59
7	5	nad 25 do 30	101	27,5	2778	2912,53
8	6	nad 30 do 35	36	32,5	1170	3871,33
9	7	nad 35 do 40	9	37,5	337,5	2126,13
10	8	nad 40 do 45	4	42,5	170	1659,75
11		Vsota	500		11065	19806,55
12						
13		približek za povprečje x_{povp}	22,13			
14		približek za varianco s^2	39,69			
15		približek za standardni odklon s	6,30			

Zgled 3.4: Rojstno ime KATJA

S spletne strani <http://www.stat.si/imena.asp> smo 31. 1. 2008 prepisali podatke v tabeli 3.5, ki so tudi v datoteki KATJA.XLS.

Tabela 3.5: Število oseb z imenom KATJA po razredih glede na leto rojstva

(Vir: <http://www.stat.si/imena.asp>).

Leto rojstva	Število žensk s tem imenom	Spodnja meja	Zgornja meja	Sredina razreda	Širina razreda	Gostota frekvenc	Relativna frekvenca	Kumulativna frekvenc	Kumulativna rel. frekvenc
	f_i	x_{imin}	x_{imax}	x_i	d_i	g_i	$f_i\%$	F_i	$F_i\%$
do 1920	3	1911	1921						
1921-1940	5								
1941-1960	120								
1961-1970	321								
1971-1980	2066								
1981-1990	3190								
1991-2000	2258								
2001-2006	518								

KATJA 1

Razrede v tabeli 3.5 dopolnite z njihovimi karakteristikami (spodnja (x_{imin}), zgornja meja (x_{imax}), sredina (x_i) in širina (d_i)). Spodnja meja prvega razreda naj bo 1911.

Zakaj v tem primeru računamo tudi gostoto frekvenc?

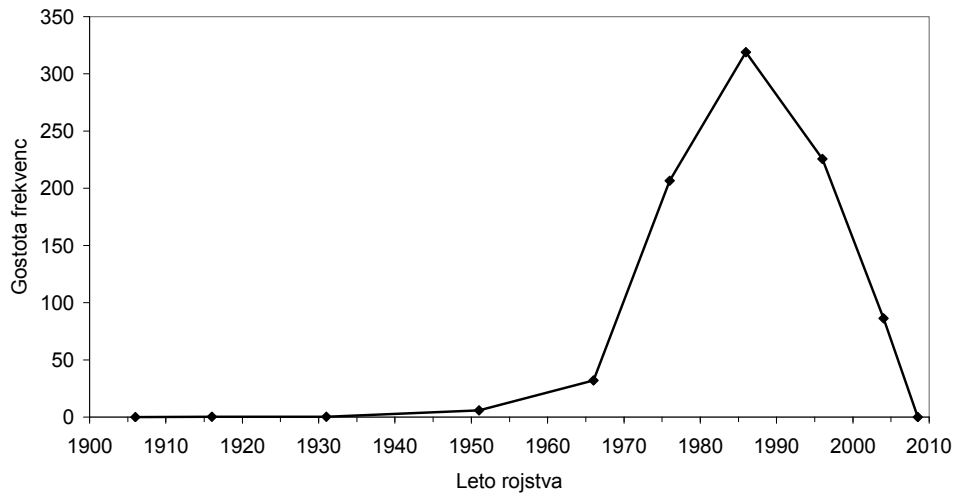
Za podatke v tabeli 3.5 izračunajte gostote frekvenc (g_i), relativne frekvence ($f_i\%$), kumulativno frekvenc (F_i) in kumulativno relativnih frekvenc ($F_i\%$). Izračune oblikujte v preglednico, kot kaže tabela 3.5. Obrazložite vrednosti za razred št. 8. ♦

KATJA 2

Podatke iz tabele 3.5 prikažite s poligonom (slika 3.4).¹⁴

Navodilo: V tabelo dodajte »ničelni« (1901-1910) in deveti razred (2007-2010). ♦

¹⁴ Excel ne zna prikazati histograma z različno širokimi razredi.

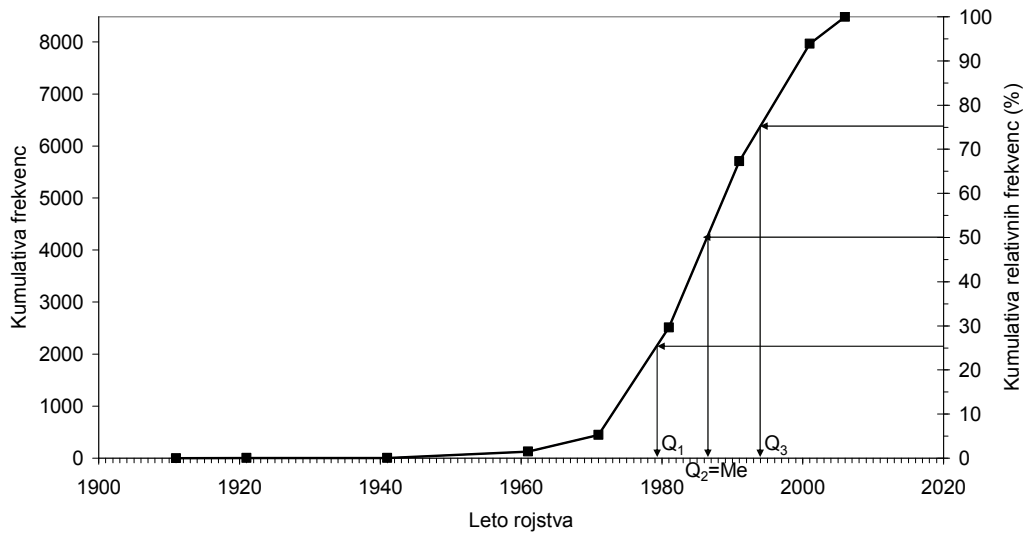


Slika 3.4: Poligon za leto rojstva žensk z imenom Katja, podatki so v tabeli 4.5.

KATJA 3

Grafično predstavite kvartile za rojstno leto oseb z imenom Katja (slika 3.5). ♦

Izračunajte približke za vse tri kvartile za rojstno leto oseb z imenom Katja. Vsebinsko obrazložite dobljene vrednosti. ♦



Slika 3.5: Kumulativna frekvenc (F) in kumulativna relativnih frekvenc ($F\%$) za leto rojstva žensk z imenom Katja. Grafični prikaz kvartilov (Q_1 , $Q_2=Me$, Q_3)

KATJA 4

Izračunajte povprečno rojstno leto (\bar{x}), varianco rojstnega leta (s^2) ter standardni odklon rojstnega leta (s). ♦

KATJA 5

Izračunajte povprečno starost oseb z imenom Katja v letu 2006. ♦

KATJA 6

Izračunajte še varianco in standardni odklon za starost oseb z imenom Katja v letu 2006. ♦

Domača naloga 3.2: Rojstno ime

Za rojstno ime matere/očeta (študentke naj izberejo ime matere, študenti pa ime očeta) naredite naloge, ki sledijo. Vse tabele s podatki in izračuni ter grafikoni shranite v datoteko ROJSTNO IME.XLS (opomba: rojstno ime, za katerega so narejeni izračuni in grafikoni).

ROJSTNO IME 1

Prepišite število oseb po razredih glede na obdobje rojstva za izbrano ime. Vir: <http://www.stat.si/imena.asp>. ♦

ROJSTNO IME 2

Izdelajte tabelo, ki za vsak razred vsebuje: spodnjo in zgornjo mejo, sredino, širino, frekvenco, relativno frekvenco (%), gostoto, kumulativo ter relativno kumulativo (%).

Navodilo: Spodnja meja prvega razreda naj bo 1911. Če dobite v določenem razredu za frekvenco oznako 'z' (statistično zaupno), jo nadomestite s številom 3. ♦

ROJSTNO IME 3

Grafično prikažite porazdelitev s poligonom.

Navodilo: Zgornjo mejo zadnjega »dodanega« razreda določite tako, da bo zajeto zadnje desetletje (npr. 2010, če vaši podatki vsebujejo podatke iz obdobja pred oz. v tem letu, sicer pa 2020, ...). ♦

ROJSTNO IME 4

Izračunajte povprečje, varianco in standardni odklon za rojstno leto. ♦

ROJSTNO IME 5

Analiziramo starost oseb s tem imenom v zadnjem letu, ki je še upoštevano v podatkih.

Izračunajte povprečno starost oseb. ♦

ROJSTNO IME 6

Izračunajte standardni odklon za starost oseb s tem imenom v zadnjem letu, ki je še upoštevano v podatkih. ♦

4 Podatkovna zbirka in vrtilne tabele

Novi pojmi:

- **podatkovna zbirka: vnos, urejanje in izbira zapisov**
- **vrtilne tabele**
- **vrtilni grafikoni**

Datoteke s podatki:

ANKETA.XLS

METEOROLOSKI PODATKI.XLS

Ustvarjene datoteke:

ANKETA_ANALIZA.XLS

MET_POD_IZBOR.XLS

ANKETA_ANALIZA.DOC

MET_POD_ANALIZA.DOC

4.1 Podatkovna zbirka

V Excelu lahko podatke organiziramo tako, da oblikujemo enostavno podatkovno zbirko. Podatkovna zbirka je sestavljena iz **zapisov**. Zapis je enota podatkovne zbirke. Podatki, ki opisujejo isto lastnost enot, so v podatkovni zbirki združeni v **polju** (spremenljivki). Primer: podatkovna zbirka je seznam študentov s podatki: ime, priimek, datum rojstva, prebivališče, letnik, število opravljenih izpitov. Podatki o študentu predstavljajo zapis, polja pa so ime, priimek, datum rojstva itd.

Vrstica na delovnem listu predstavlja zapis (enoto), stolpec pa polje (spremenljivko). Podatkovna zbirka lahko vsebuje največ toliko podatkov, kolikor je celic na enem delovnem listu, to je $65536 * 256$ (število vrstic * število stolpcev).

V prvi vrstici podatkovne zbirke morajo biti imena polj. Za pregledno delo s podatkovno zbirko je priporočljivo, da je na izbranem delovnem listu v zvezku samo podatkovna zbirka. Vse ostale informacije izpisujemo na druge delovne liste v zvezku.

Zapise v podatkovni zbirki urejamo in analiziramo z orodji v meniju **Podatki** in s funkcijami zvrsti *Zbirka podatkov*.

	A	B	C	D		H	
	Študijska smer (AG, BT ali MB)	Leto rojstva	Spol (m, ž)	Število otrok v družini	Srednja	ih točk	Štipendija (da/ne)
1							
2	AG	1982	ž	2	tehnična		ne
3	AG	1983	ž	2	gimnazija	17	ne
4	AG	1983	m		gimnazija	16	ne
5	BT	1984	ž	2	splošna		ne
6	MB	1984	m	2	Gimnazij		ne
7	BT	1984	m	3	Šolski ce	21	ne
8	AG	1984	ž	2	gimnazija	17	da
9	AG	1984	ž	2	ekonomska		ne

ANKETA 1

Odprite datoteko ANKETA.XLS in si oglejte vsebino delovnega lista »Podatki«. V datoteki so rezultati ankete, izvedene v šolskem letu 2006/07 med študenti 2. letnika univerzitetnega študija agronomije (AG), biotehnologije (BT) in mikrobiologije (MB). V nadaljevanju bomo naredili analizo teh podatkov, zato datoteko shranite z novim imenom ANKETA_ANALIZA.XLS.

Navodilo: Vse analize podatkov, ki sledijo, naredite na podatkovni zbirki v datoteki ANKETA_ANALIZA.XLS na delovnem listu »Podatki« ♦

V oblikovano podatkovno zbirko lahko poljubno dodajamo polja, tako da vstavimo prazen stolpec in napišemo ime polja v prvo vrstico. Polja lahko tudi prestavljamo in brišemo. V vsaki podatkovni zbirki pride prav, če imamo v prvem polju zaporedne številke zapisov.

ANKETA 2

Na začetek podatkovne zbirke vrnite polje **Zap. štev.** z zaporednimi števkami zapisov v podatkovni zbirki¹⁵. ♦

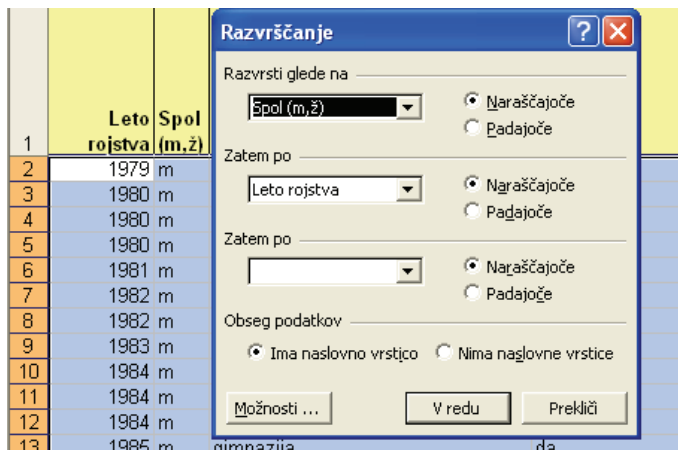
ANKETA 3

Vrnite še eno polje z imenom **Starost**. Za vsakega anketiranca izračunajte starost v dopoljenih letih v letu 2007. ♦

4.2 Urejanje zapisov

V podatkovni zbirki lahko zapise uredimo v različnem vrstnem redu. Če so podatki v polju tekstovni, jih zna Excel urediti po abecednem vrstnem redu od A do Ž ali pa od Ž do A. Številske podatke zna Excel urediti v naraščajočem ali padajočem vrstnem redu, prav tako tudi datume.

Za urejanje zapisov uporabljamo ukaz **Podatki/Razvrsti**. V pogovornem oknu **Razvrščanje** v polju **Razvrsti glede na** izberemo polje podatkovne zbirke, ki določa primarni vrstni red zapisov. Odpremo seznam imen in kliknemo na zeleno ime. Poleg imena moramo izbrati tudi naraščajoči (**Naraščajoče**) ali padajoči (**Padajoče**) vrstni red. V poljih **Zatem po** izberemo polje, ki določa sekundarni vrstni red zapisov ter polje, ki določa terciarni vrstni red zapisov.



¹⁵ V stolpec lahko zaporedne številke najhitreje napišemo tako, da v prvo celico napišemo številko 1, v celico pod njo številko 2, nato obe celici označimo v blok in križec miške postavimo v desni spodnji kot bloka. Ko se križec spremeni (postane črn), s potegom miške označimo blok, kamor naj se izpišejo zaporedne številke. Številke se izpišejo samodejno.

Zapise v podatkovni zbirki lahko uredimo tudi po več kot treh poljih, tako da ukaz **Podatki/Razvrsti** uporabimo večkrat zaporedoma.

Ukaz **Podatki/Razvrsti** spremeni vrstni red zapisov v podatkovni zbirki. Če hočemo podatke ponovno spraviti v prvotni vrstni red, jih uredimo po zaporedni številki. Zato je vedno koristno, da podatkovna zbirka vsebuje tudi zaporedne številke zapisov.

ANKETA 4

Uredite anketirance v podatkovni zbirki po starosti (letu rojstva) od najmlajšega do najstarejšega.

Opozorilo: preden uporabimo katerikoli ukaz v meniju **Podatki** z namenom urejanja ali analize podatkov, v podatkovni zbirki označimo **eno neprazno celico** na delovnem listu z zbirko podatkov. Ne smemo označiti posamezne skupine celic ali pa posameznega polja podatkovne zbirke. ♦

ANKETA 5

Uredite anketirance v podatkovni zbirki po najljubšem predmetu v 1. letniku. Kaj dobite? ♦

ANKETA 6

Uredite podatke po spolu, znotraj spola pa po starosti. ♦

4.3 Izbira zapisov

4.3.1 Samodejni filter

V podatkovni zbirki lahko izbiramo in iščemo posamezne zapise po različnih kriterijih. Uporabimo ukaz **Podatki/Filter**. Enostavne kriterije določimo s podukazom **Samodejni filter**, bolj zapletene pa s podukazom **Napredni filter**.

Ko uporabimo ukaz **Podatki/Filter, Samodejni filter**, se v naslovni vrstici poleg imen polj pojavijo gumbi s puščicami. Klik na puščico odpre podmeni, v katerem so navedene vse obstoječe vrednosti v izbranem polju in še možnosti **Razvrsti naraščajoče**, **Razvrsti padajoče**, **Vse**, **Zgornjih 10...**, **Lasten...**, in še **Prazne** in **Neprazne** v primeru, če v izbranem polju obstajajo prazne celice.

Študijska smer (AG, BT ali MB)	Leto rojstva	Spol (m, ž)	Število otrok v družini	Srednja šola, ki ste jo zaključili (polno ime, ne s kratico)	Gimnazija (da/ne)	Število doseženih točk na maturi	Štipendija (da/ne)	Najljubši predmet v 1. letniku	Povprečna ocena v 1. letniku
AG	1982	ž	2	tehnična šola - kemijski tehnik	ne		ne	botanika	7,0
AG	1983	ž	2	gimnazija	da	17	ne	botanika	7,0
AG	1983	m		gimnazija	da	16	ne	matematika	7,0

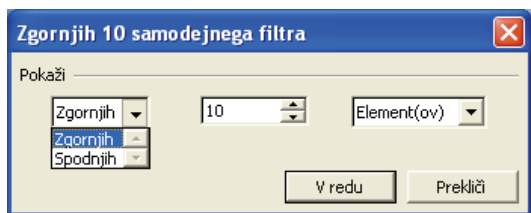
Študijska smer (AG, BT ali MB)	Leto rojstva	Spol (m, ž)
AG	1982	ž
AG	1983	ž
AG	1983	m
BT	1982	ž
MB	1982	m
BT	1983	m
AG	1984	ž
AG	1985	ž
AG	1986	ž
AG	1987	ž
AG	1984	ž

Študijska smer (AG, BT ali MB)	Leto rojstva	Spol (m, ž)	Število otrok v družini	Srednja šola, ki ste jo zaključili (polno ime, ne s kratico)
AG	1982	ž	2	tehnična šola - kemijski tehnik
AG	1983	ž	2	gimnazija
AG	1983	m		gimnazija
BT	1982	ž	2	splošna gimnazija ptuj
MB	1982	m		Gimnazija Želimlje
BT	1983	m	1	Šolski center Ljubljana - Tehniška
AG	1983	ž	2	gimnazija
AG	1984	ž	3	ekonomska
AG	1984	ž	4	kmetijska
AG	1985	ž	5	srednja šola Josipa Jurčiča Ivančna
AG	1986	ž	6	Prva gimnazija Maribor
BT	1986	ž	8	Srednješolski center Postojna
MB	1987	ž	2	

Kriterije izbire določimo na več načinov. Ko je izbira določena, Excel na mesto podatkovne zbirke izpiše samo zapise, ki ustrezajo izbranim kriterijem, druge pa skrrije.

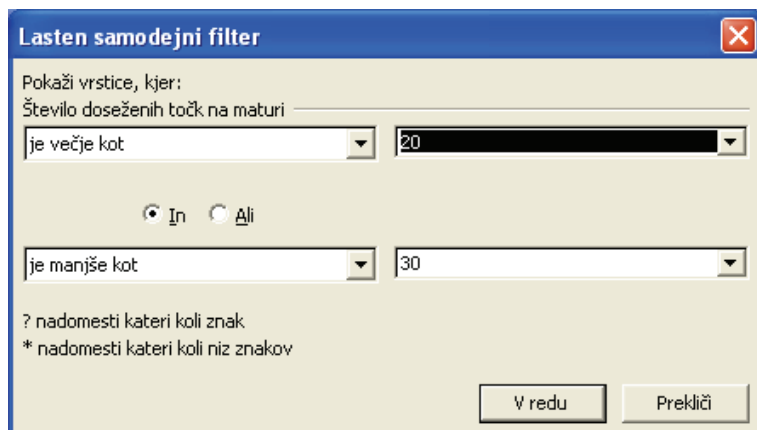
Načini izbire:

- če iščemo zapise z določeno vrednostjo izbranega polja, kliknemo to vrednost v podmeniju;
- možnost *Vse* uporabimo, če želimo preklicati predhodno uporabljeni kriterij v izbranem polju. Pri tem pa ni nujno, da se izpišejo vsi zapisi v podatkovni zbirki. Če želimo preklicati kriterije v vseh poljih hkrati, uporabimo možnost **Podatki/Filter, Pokaži vse**;
- možnost **Zgornjih 10** uporabimo, kadar želimo izpis določenega števila oz. določen odstotek zapisov z največjimi oz. najnižjimi vrednostmi izbrane spremenljivke:



- možnost **Lasten** uporabimo, kadar sami določimo kriterije. V pogovornem oknu **Lasten samodejni filter** lahko kriterij sestavimo iz enega ali dveh pogojev. V prvem polju izberemo logični operator (je enako, ni enako, je večje kot..., se začne z, se ne začne z, vsebuje,...), v sosednje polje pa napišemo vrednost. Vrednost je lahko številska ali opisna. Če določamo kriterij za numerično spremenljivko, vpišemo ustrezno število. Če določamo kriterij za opisno spremenljivko, vpišemo niz znakov.

Kadar v kriteriju uporabimo dva pogoja, ju lahko povezuje operator In ali Ali. Če želimo, da veljata oba pogoja hkrati, izberemo operator In. Če želimo, da je izpolnjen ali prvi pogoj ali drugi ali oba hkrati, izberemo operator Ali.



- možnost **Prazne** uporabimo, kadar želimo, da se izpišejo vsi zapisi, ki imajo izbrano polje prazno;
- možnost **Neprazne** uporabimo, kadar želimo, da se izpišejo vsi zapisi, v katerih izbrano polje ni prazno.

Samodejni filter izklopimo s ponovnim ukazom **Podatki/Filter, Samodejni filter** (ob tem puščice v prvi vrstici podatkovne zbirke izginejo).

ANKETA 7

Poiščite zapise za anketirance, ki stanujejo v študentskem domu ali pa so podnajemniki. ♦

ANKETA 8

Poiščite zapise za anketirance, ki so starejši od 20 let in so moškega spola.

Navodilo: samodejni filter uporabite v dveh korakih, najprej pri spremenljivki *Spol* in nato še pri spremenljivki *Starost* ali obratno. ♦

Če želimo imeti izbrane zapise za novo podatkovno zbirko, izbor zapisov kopiramo na nov delovni list.

ANKETA 9

Poiščite zapise za anketiranke (študentke), katerih telesna višina je manjša od povprečne telesne višine vseh anketiranih študentk.

Navodilo: najprej izberite študentke in njihove zapise kopirajte na delovni list z imenom »Študentke«. Na novem listu s funkcijo **AVERAGE** izračunajte povprečje telesne višine študentk (npr. v prazno celico v istem stolpcu, kot so podatki za telesno višino). Na podlagi izračunane vrednosti ponovno uporabite filter za izbor študentk s podpovprečno telesno višino na delovnem listu »Podatki«. Za kopiranje vrednosti povprečja uporabite **CTRL C** (hkratni stisk tipk CTRL in C) in v pogovornem oknu *Lasten samodejni filter* v polju za vrednost spremenljivke uporabite **CTRL V**. ♦

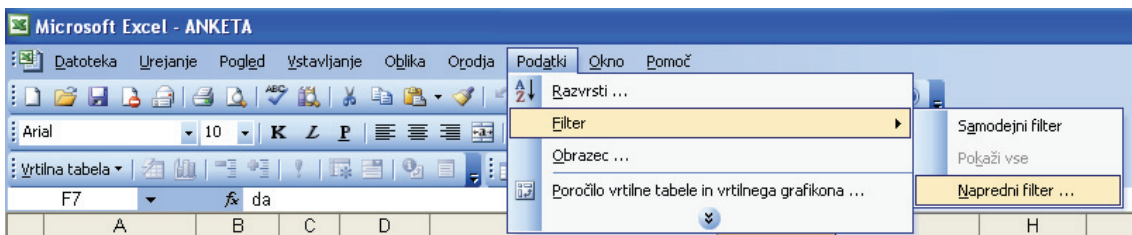
ANKETA 10

Poiščite zapise vseh, ki so nastanjeni v študentskem domu in se na fakulteto vozijo z avtomobilom. Izbrane zapise kopirajte na nov list v preglednici in ga poimenujte "Študentski dom". ♦

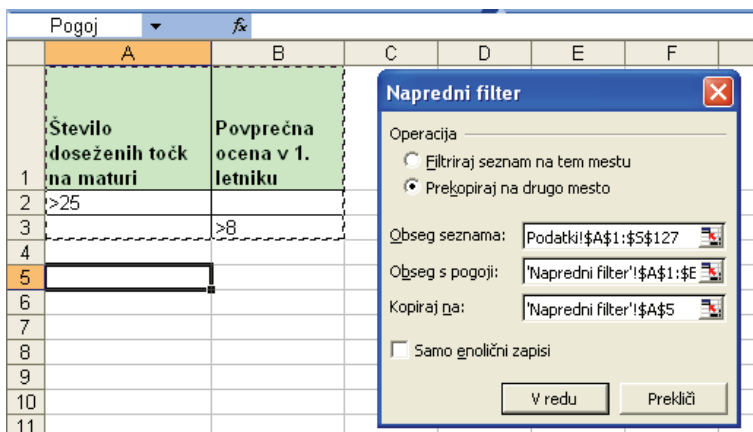
4.3.2 Napredni filter

Ukaz **Podatki/Filter, Samodejni filter**, možnost *Lasten* omogoča, da za eno polje uporabimo največ dva pogoja. Med izbranimi zapisi pa lahko iščemo še podskupine zapisov, ki ustrezajo kriterijem, določenim na osnovi več polj v podatkovni zbirki.

Kadar želimo uporabiti sestavljene, zahtevnejše kriterije, uporabimo ukaz **Podatki/Filter, Napredni filter**. Najprej moramo napisati **obseg celic s pogoji**. Obseg celic s pogoji je sestavljen iz najmanj dveh vrstic. V prvi vrstici so imena polj (oblika pisave mora biti enaka kot v naslovni vrstici), v naslednjih vrsticah pa so napisani pogoji (kriteriji). Pogoje napišemo v isto vrstico, če morajo veljati vsi hkrati oziroma v zaporedne vrstice, če mora veljati vsaj eden izmed njih.



V polje **Obseg seznama** napišemo skupino celic, v kateri je podatkovna zbirka. V polje **Obseg s pogoji** napišemo skupino celic, v kateri so napisani pogoji. Možnost **Filtriraj seznam na tem mestu** izberemo, če želimo, da se na mestu podatkovne zbirke izpišejo samo tisti zapisi, ki ustrezajo pogojem. **Prekopiraj na drugo mesto** pa izberemo, če želimo, da se izbrani zapisi izpišejo na mesto, ki ga določimo v polju **Kopiraj na**. Če je v podatkovni zbirki več enakih zapisov, možnost **Samo enolični zapisi** določi izbiro samo enega izmed njih.



ANKETA 11

Na nov delovni list z imenom »Napredni filter« napiši pogoje za izbor anketiranih, ki so imeli število točk pri maturi večje od 25 ali pa so imeli povprečno oceno v 1. letniku višjo od 8. Zapise, ki ustrezajo danemu pogoju, izpišite na delovni list »Napredni filter« v vrstice pod obsegom pogojev (glej sliko zgoraj). Koliko je takih zapisov? ♦

ANKETA 12

Kako bi izbrali zapise, ki ustrezajo pogoju: število točk pri maturi je večje od 25 in hkrati je povprečna ocena v 1. letniku višja od 8? ♦

4.4 Kontrola podatkov

Po vsakem vnosu je potrebna kontrola podatkov. V Excelu lahko to napravimo na več načinov. Z ustreznim kriterijem lahko izpišemo zapise, ki vsebujejo neustrezne vrednosti v posameznem polju. Kriterije lahko določimo tudi glede na odnose med vrednostmi v različnih poljih (logična kontrola). Pogosto kontrolo podatkov učinkovito izvedemo z vrtilnimi tabelami. Ta način si bomo ogledali kasneje.

ANKETA 13

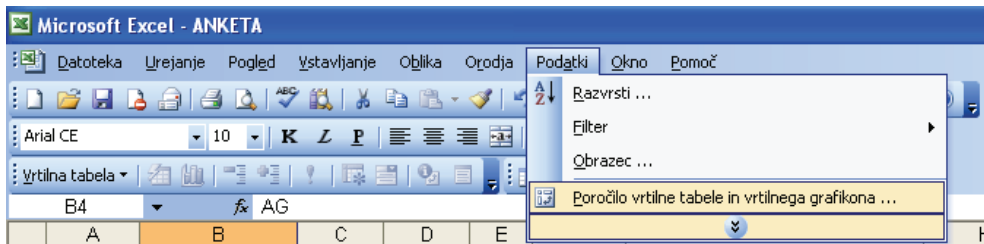
Z uporabo samodejnega filtra preglejte, katere vrednosti nastopajo za posamezne spremenljivke v podatkovni zbirki na delovnem listu »Podatki«. Ali so vse vrednosti smiselne? ♦

ANKETA 14

Preverite, ali so podatki polj *Ali se ukvarjate s športom* in *Koliko ur na teden se ukvarjate s športom* logični. ♦

4.5 Vrtilne tabele

Vrtilno tabelo v Excelu naredimo z ukazom **Podatki/Poročilo vrtilne tabele in vrtilnega grafikona**.



Vrtilno tabelo uporabimo, če želimo zapise v podatkovni zbirki razdeliti v skupine glede na vrednosti posameznih polj (spremenljivk) in za te skupine izračunati statistične vrednosti za izbrane spremenljivke (npr. prešteti, koliko je moških in koliko žensk, izračunati njihovo povprečno starost, minimalno in maksimalno starost, ...). Primeri vrtilnih tabel so na slikah od 4.1 do 4.6.

	A	B	C	D	E	F	G
1	Tule spustite polja strani						
2							
3	Štej od Zap. števil.						
4	Spol (m, ž)	Vsota					
5	m	41					
6	ž	85					
7	Skupna vsota	126					

	A	B	C	D
1	Tule spustite polja strani			
2				
3	Povprečje od Povprečna ocena v 1. letniku	Spol (m, ž)		
4	Študijska smer (AG, BT ali MB)	m	ž	Skupna vsota
5	AG	7,7	7,6	7,7
6	BT	8,0	7,5	7,7
7	MB	7,9	7,5	7,6
8	Skupna vsota	7,8	7,6	7,6

	A	B	C	D
1	Tule spustite polja strani			
2				
3		Spol (m, ž)		
4	Podatki	m	ž	Skupna vsota
5	Štej od Zap. števil.	41,0	85,0	126,0
6	Povprečje od Število doseženih točk na maturi	19,4	22,1	21,2
7	Povprečje od Povprečna ocena v 1. letniku	7,8	7,6	7,6

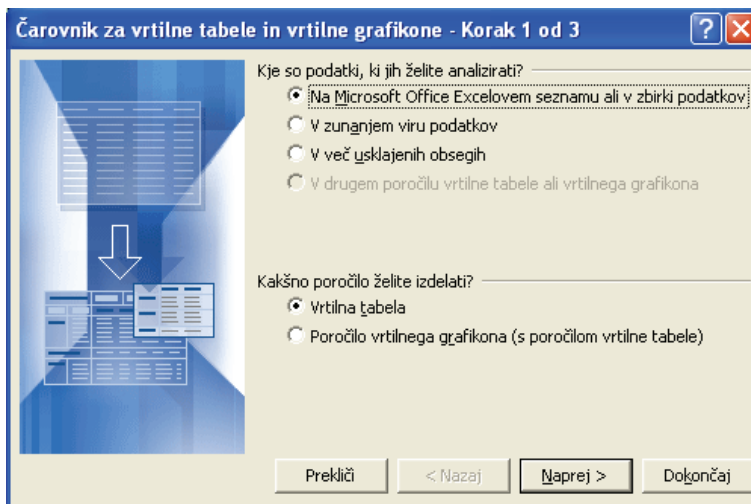
Slika 4.1: Število moških in žensk, sodelujočih v anketi (zgoraj) - enosmerna vrtilna tabela, orodna vrstica Vrtilna tabela je na delovnem listu. Povprečje *Povprečne ocene v 1. letniku* po spolu in po smeri študija (na sredi) – dvosmerna vrtilna tabela, orodna vrstica Vrtilna tabela je med ostalimi

orodnimi vrsticami na vrhu okna. Število anketiranih, povprečje *Števila doseženih točk na maturi* in povprečje *Povprečne ocene v 1. letniku* po *spolu* - enosmerna vrtilna tabela s tremi izračunanimi statistikami (spodaj). Za izraza “Vsota” oziroma “Skupna vsota” v zadnji vrstici oz. zadnjem stolpcu vrtilnih tabel glej pojasnilo na strani 86.

4.5.1 Čarovnik za vrtilne tabele

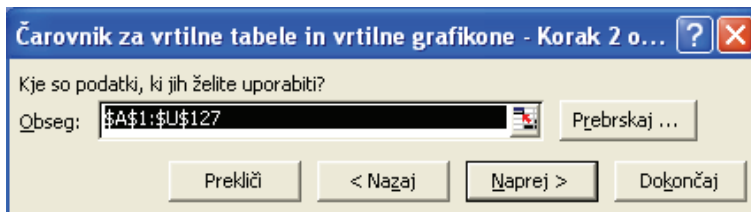
Ko uporabimo ukaz **Podatki/Poročilo vrtilne tabele in vrtilnega grafikona**, aktiviramo čarovnika za vrtilne tabele. Postopek, ki sledi, ima štiri korake:

1. korak: Čarovnik za vrtilne tabele – Korak 1 od 3



Najprej določimo podatke za vrtilno tabelo. Za podatke iz excelove podatkovne zbirke pustimo izbrani možnosti *Na Microsoft Excelovem seznamu ali v podatkovni zbirki* in *Vrtilna tabela*.

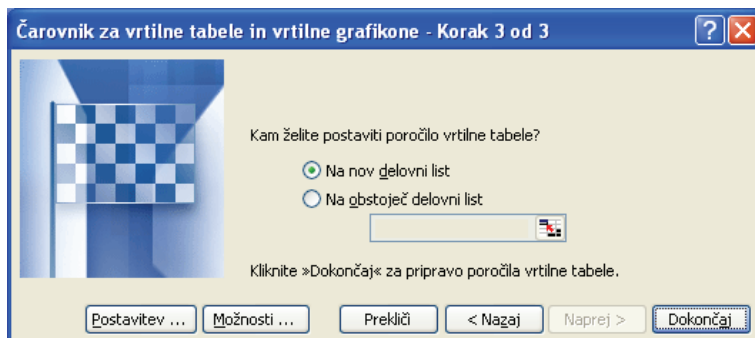
2. korak: Čarovnik za vrtilne tabele – Korak 2 od 3



V polju **Obseg** določimo skupino celic, v kateri je podatkovna zbirka. Če imamo pred uporabo ukaza **Podatki/Poročilo vrtilne tabele** označeno eno celico v podatkovni zbirki, Excel prepozna podatkovno zbirko in to polje izpolni sam. Če imamo pred

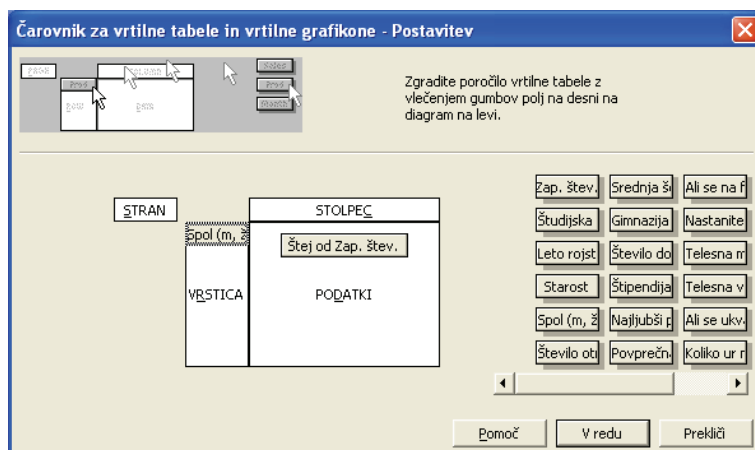
uporabo ukaza **Podatki/Poročilo vrtilne tabele in vrtilnega grafikona** označeno skupino celic v podatkovni zbirki, bo Excel upošteval podatke v označeni skupini celic.

3. korak: Čarovnik za vrtilne tabele – Korak 3 od 3



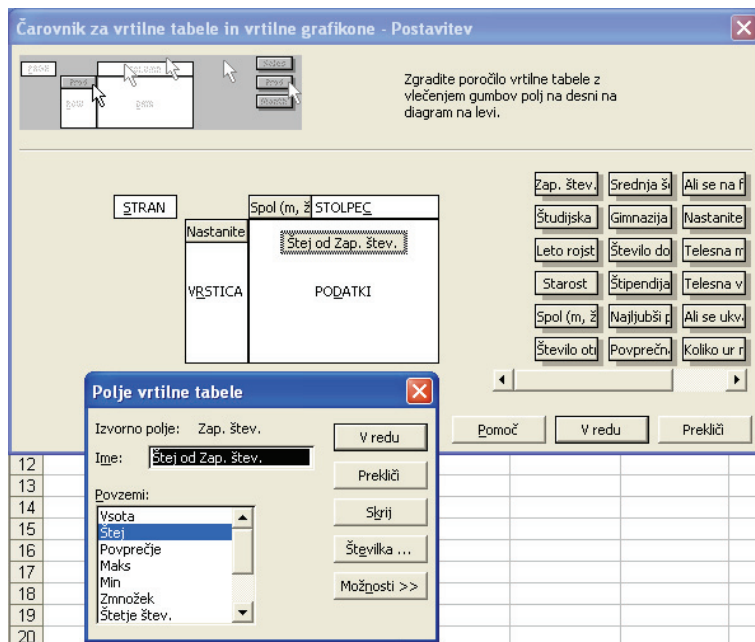
Vrtilna tabela se izpiše na nov list, če izberemo možnost *Nov delovni list* (to možnost bomo uporabljali) oz. na obstoječi delovni list, če izberemo možnost *Na obstoječi delovni list*.

S klikom na gumb **Postavitev** v tretjem koraku se odpre pogovorno okno *Čarovnik za vrtilne tabele in grafikone – Postavitev*. V pogovornem oknu je shema vrtilne tabele in gumbi z imeni polj. Shema vrtilne tabele ima štiri področja *Stran*, *Vrstica*, *Stolpec* in *Podatki*. S potegom miške vanje postavimo imena polj. Za oblikovanje vrtilne tabele na sliki 4.1 je postavitvev polj vrtilne tabele takšna:



V področja *Stran*, *Vrstica* in *Stolpec* postavimo polja, (spremenljivke) po katerih želimo zapise v podatkovni zbirki razdeliti v skupine. V področje *Vrstica* postavimo

polja, katerih vrednosti določajo vrstice tabele (npr. če izberemo polje *Spol*, bo imela vrtilna tabela dve osnovni vrstici - eno vrstico za ženske in eno za moške). V področje *Stolpec* postavimo polja, katerih vrednosti določajo stolpce tabele (npr. če izberemo polje *Nastanitev (...)*, bo imela vrtilna tabela štiri osnovne stolpce).

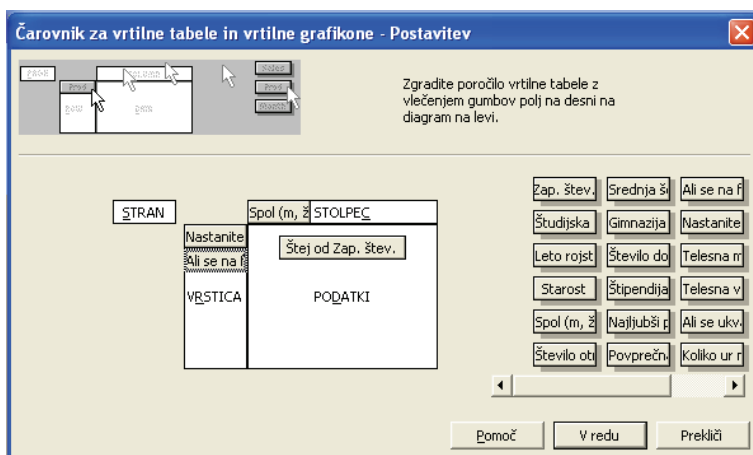


	A	B	C	D
1	Tule spustite polja strani			
2				
3	Štej od Zap. šte.	Spol (m, ž)		
4	Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž	Skupna vsota
5	doma	20	31	51
6	lastnik		1	1
7	podnajemnik	7	33	40
8	študentski dom	14	18	32
9	(prazen)		2	2
10	Skupna vsota	41	85	126

Slika 4.2: Število anketiranih po spolu in po nastanitvi - **dvosmerna vrtilna tabela**. Med podatki sta dva zapisa, v katerih ni vrednosti za spremenljivko *Nastanitev (...)*, zato se kot možna vrednost te spremenljivke v tabeli pojavi tudi vrednost "prazen".

Če v področje **Vrstica** oz. **Stolpec** postavimo več polj, bodo vrstice oz. stolpci najprej razdeljeni po kategorijah prvega polja, nato bo vsaka kategorija prvega polja razdeljena po kategorijah drugega polja itd. V področje **Stran** postavimo polja, katerih vrednosti določajo novo dimenzijo vrtilne tabele (slika 4.4). Za vsako vrednost iz tega polja se izpiše samostojna vrtilna tabela.

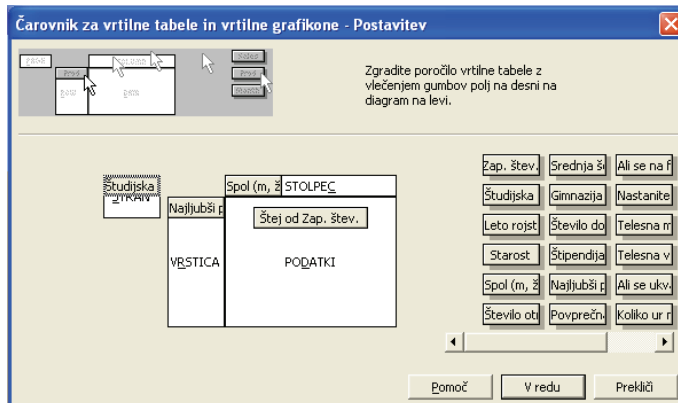
V področje **Podatki** postavimo imena polj, za katera želimo izračunati izbrane funkcije. Funkcijo izberemo z dvoklikom na ime polja v področju **Podatki**. S tem odpremo pogovorno okno **Polje vrtilne tabele**. V polju **Povzemi** izberemo funkcijo. V polju **Ime** se pojavi ime za izračune v vrtilni tabeli, ki ga Excel določi glede na izbrano funkcijo in glede na ime polja (spremenljivke): npr. *Štej od Zap. šte.* ali *Povprečje od Povprečna ocena v 1. letniku* v vrtilnih tabelah na sliki 4.1. Ime se izpiše v zgornjo levo celico vrtilne tabele.



	A	B	C	D	E
1		Tule spustite polja strani			
2					
3	Štej od Zap. šte.		Spol (m, ž)		
4	Nastanitev (doma, študentski dom, prazen)	Ali se na fakulteto vozite z osebnim avtom?	m	ž	Skupna vsota
5	doma	da	8	15	23
6		ne	3	7	10
7		občasno	9	9	18
8	Vsota doma		20	31	51
9	lastnik	ne		1	1
10	Vsota lastnik			1	1
11	podnajemnik	da	1	2	3
12		ne	5	24	29
13		občasno	1	7	8
14	Vsota podnajemnik		7	33	40
15	študentski dom	da		2	2
16		ne	12	16	28
17		občasno	2		2
18	Vsota študentski dom		14	18	32
19	(prazen)	da		2	2
20	Vsota (prazen)			2	2
21	Skupna vsota		41	85	126

Slika 4.3: Število anketiranih po spolu, nastanitvi in glede na to, ali se vozijo na fakulteto z osebnim avtomobilom - **trosmerna vrtilna tabela**

Excel v vrtilno tabelo doda kategorijo (*prazen*) za manjkajoče podatke (slike 4.2, 4.3 in 4.4).



	A	B	C	D
1	Študijska smer (AG, BT ali MB)	(Vse)		
2				
3	Štej od Zap. šte.			
4	Najljubši predmet v 1. letniku			
5	agrometeorologija z osnovami fizike okolja			11
6	angleščina			1
7	biokemija			16
8	biologija			11
9	biologija celice			19
10	botanika			14
11	fizikalna kemija			1
12	fiziologija z anatomijo domačih živali			2
13	kemija			8
14	matematika			11
15	pedologija	15	11	26
16	uvod v biotehnologijo z bioetiko	1		1
17	(prazen)		5	5
18	Skupna vsota	41	85	126

	A	B	C	D
1	Študijska smer (AG, BT ali MB)	BT		
2				
3	Štej od Zap. šte.			
4	Najljubši predmet v 1. letniku			
5	biokemija			7
6	biologija			11
7	kemija			4
8	matematika			2
9	uvod v biotehnologijo z bioetiko			1
10	Skupna vsota			25
11				
12				
13				
14				

Slika 4.4: Število anketiranih za vsako študijsko smer (*Stran*), po spolu (*Stolpec*) in po izbiri najljubšega predmeta v 1. letniku (*Vrstica*) - večplastna dvosmerna vrtilna tabela; pet študentov se ni opredelilo za najljubši predmet v 1. letniku (vrstica (prazen) v levi tabeli); če v področju *Stran* izberemo BT, se pokažejo podatki samo za biotehnologe (desna tabela).

4.5.2 Dodatno oblikovanje vrtilnih tabel

Že narejene vrtilne tabele lahko spreminjamo s klikom na ikone v orodni vrstici **Vrtilna tabela** in s premikanjem polj iz pogovornega okna *Seznam polj vrtilne tabele* v

področja vrtilne tabele. Orodna vrstica **Vrtilna tabela** in pogovorno okno **Seznam polj vrtilne tabele** se pokažeta na zaslonu, če imamo izbrano poljubno celico vrtilne tabele.¹⁶

Praktično je, da imamo v datoteki za izbrano podatkovno zbirko le eno vrtilno tabelo, ki jo poljubno spreminjamo. Z vrtilno tabelo dobljene podatke, ki jih potrebujemo za nadaljevanje analize, kopiramo z ukazom **Posebno lepljenje/Vrednosti** na drug list v delovnem zvezku ali pa v wordovo datoteko. Rekli bomo, da s tem vsebino vrtilne tabele »parkiramo«. Če za vsako analizo naredimo novo vrtilno tabelo, postane Excelova datoteka hitro zelo velika.

Excel v enosmernih vrtilnih tabelah poimenuje stolpec oz. vrstico z izračuni z imenom **Vsota** ne glede na to, katero funkcijo smo določili v področju **Podatki**, zato je večkrat vsebinsko neprimerno (npr. v tabelah na sliki 4.1). V enosmerni vrtilni tabeli tega imena ne moremo spremeniti. Ime moramo popraviti, ko tabelo kopiramo kot navadno tabelo na drugo mesto v excelovo datoteko ali pa kam drugam (npr. v wordovo datoteko). Podobno v večsmernih tabelah Excel zadnji stolpec oz. vrstico poimenuje z imenom **Skupna vsota**. Tudi to ime je enako ne glede na funkcijo v področju **Podatki**, vendar ga lahko spremenimo. Spremenjeno ime v vrtilni tabeli ostane, tudi če kasneje spreminjamo vsebino področij **Stolpec**, **Vrstica**, **Stran** ali **Podatki**. Namesto **Skupna vsota** je ponavadi veliko bolj primeren izraz **Skupaj**.

ANKETA 15

Oblikujte vrtilne tabele, prikazane na slikah 4.1-4.4 ter jih »parkirajte« na delovni list »Oblikovane tabele« v datoteki ANKETA_ANALIZA.XLS. V parkiranih tabelah podatke oblikujte na smiselno število decimalnih mest in ustrezno popravite napise.

Primeri ustrezno oblikovanih parkiranih vrtilnih tabel:

Spol	Število anketiranih
Moški	41
Ženske	85
Skupaj	126

Povprečna ocena v 1. letniku	Spol

¹⁶ Če orodne vrstice na zaslonu ni, jo postavimo nazaj na zaslon z ukazom **Pogled/Orodne vrstice**, **Vrtilna tabela**.

Študijska smer	Moški	Ženske	Skupaj
AG	7,7	7,6	7,7
BT	8,0	7,5	7,7
MB	7,9	7,5	7,6
Skupaj	7,8	7,6	7,6

Statistike	Spol		Skupaj
	m	ž	
Število anketiranih	41	85	126
Povprečno število doseženih točk na maturi	19,4	22,1	21,2
Povprečna ocena v 1. letniku	7,8	7,6	7,6

◆

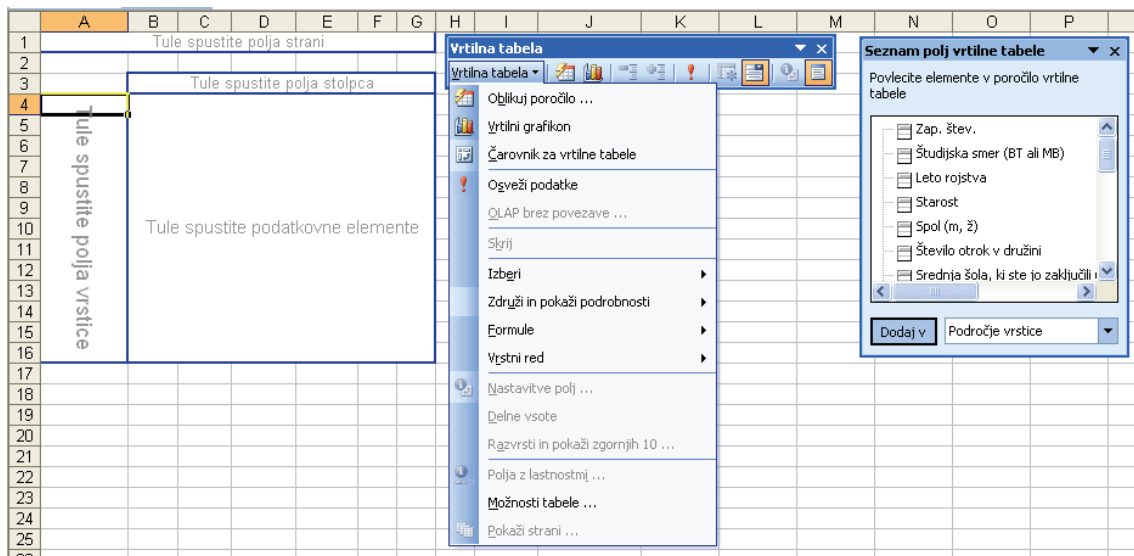
Z dvoklikom na celico vrtilne tabele, v kateri so podatki, (npr. število enot v kategoriji, skupini, ...) se odpre nov delovni list z delom podatkovne zbirke, ki vsebuje zapise enot, na podlagi katerih je bila izračunana izbrana vrednost. To aktivnost imenujemo vrtanje po tabeli.

ANKETA 16

Grafično prikažite podatke iz tabele na sliki 4.2. Grafikon oblikujte s čarovnikom za grafikone na podlagi podatkov na delovnem listu »Oblikovane tabele«. ◆

4.5.3 Orodna vrstica Vrtilna tabela

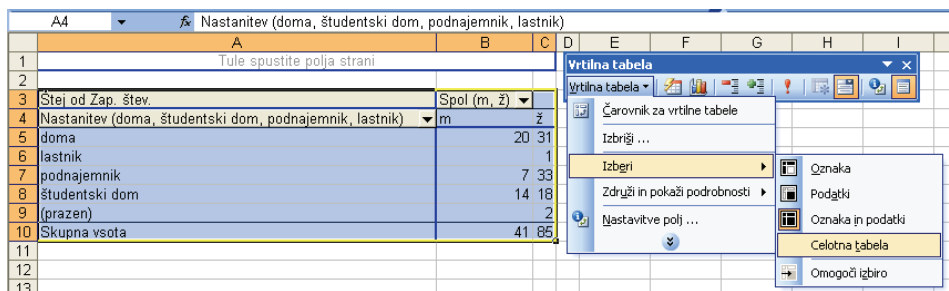
Z uporabo ukaza **Podatki/Poročilo vrtilne tabele in vrtilnega grafikona** oblikujemo tabelo, ki jo postavimo na nov delovni list v excelovem zvezku (datoteki). Ob vrtilni tabeli se pojavi še orodna vrstica **Vrtilna tabela** in okno **Seznam polj vrtilne tabele**. Če v področjih **Stran**, **Vrstica**, **Stolpec** in **Podatki** ni izbranih polj (spremenljivk), ima vrtilna tabela videz prikaza na spodnji sliki:



Prva ikona v orodni vrstici **Vrtilna tabela** odpre seznam ukazov za spreminjanje obstoječe vrtilne tabele. Nekaj ukazov s tega seznama lahko izvedemo tudi s klikom na posamezne druge ikone v orodni vrstici.


Ukaz **Čarovnik za vrtilne tabele** odpre pogovorno okno zadnjega koraka čarovnika z gumbom **Postavitev**.

Z ukazom **Izberi** označimo celo ali pa le del vrtilne tabele. Ukaz pride prav, kadar želimo celo ali del vrtilne tabele kopirati. Izbrani del vrtilne tabele se ob tem obarva svetlo modro, celica vrtilne tabele, ki je bila pred uporabo ukaza označena, in celice z imeni polj pa ostanejo bele:



V nadaljevanju si bomo podrobneje ogledali tri ukaze: **Osveži podatke**, **Skrij** in **Združi in Pokaži podrobnosti**.

4.5.3.1 Osveži podatke

Ukaz **Osveži podatke** ali klik na ikono , uporabimo, če v podatkovni zbirki podatke spremenimo in želimo, da se te spremembe odražajo tudi v obstoječi vrtilni tabeli.

ANKETA 17

Oblikujte vrtilno tabelo s povprečnim številom doseženih točk na maturi po spolu in po smereh študija (tabela 4.1). Ali se vam zdijo rezultati smiselni? ♦

Tabela 4.1: Povprečno število doseženih točk na maturi za študente in študentke 2. letnika agronomije, biotehnologije in mikrobiologije v šol. l. 2006/07 (parkirana vrtilna tabela)

Študijska smer	Povprečno število doseženih točk na maturi		
	Moški	Ženske	Skupaj
AG	16,2	15,4	15,7
BT	24,3	31,0	28,1
MB	24,7	24,3	24,3
Skupaj	19,4	22,3	21,4

ANKETA 18

Dvokliknite na celico vrtilne tabele s podatkom o povprečnem številu doseženih točk na maturi za študentke biotehnologije (vrtanje tabele). Odpre se nov delovni list z zapisi enot, ki so v tej celici. Vidimo, da se med vrednostmi *Števila doseženih točk na maturi* pojavi nenavadno visoka vrednost 86, ki je nedvomno napačna. ♦

	A	B	C	D	E	F	G	H	I	J	K	
1	Zap. štev.	Študijska smer	Leto rojstva	Starost	Spol	Število otrok	Srednja šola	Gimnazija	Število doseženih točk na maturi	Stipendija (da/ne)	Najljubši predmet	
2	93	BT	1986	21	ž	2	Gimnazija Ški	da		da	biologija	
3	90	BT	1986	21	ž	2	Škofijska klas	da		22	ne	biokemija
4	65	BT	1986	21	ž	1	Gimnazija Pol	da		24	ne	biologija
5	4	BT	1984	23	ž	2	splošna gimn	da			ne	biologija
6	63	BT	1986	21	ž	2	Gimnazija Be	da		29	ne	kemija
7	57	BT	1986	21	ž	3	Škofijska klas	da		31	da	matematika
8	55	BT	1986	21	ž	1	l. gimnazija v	da		29	ne	biologija
9	51	BT	1986	21	ž	3	škofijska gimn	da		24	da	biokemija
10	50	BT	1986	21	ž	5	Gimnazija Cel	da		28	da	biologija
11	46	BT	1986	21	ž	2	Gimnazija Or	da		86	ne	biologija
12	44	BT	1986	21	ž	2	srednja šola v	ne		28	ne	biologija
13	43	BT	1986	21	ž	3	Gimnazija Nov	da		23	da	biologija
14	39	BT	1986	21	ž	2	Gimnazija Fra	da		25	da	kemija
15	36	BT	1986	21	ž	3	srednja šola v	ne		24	da	kemija
16	35	BT	1986	21	ž	4	Gimnazija Be	da		30	da	biokemija

ANKETA 19

Na delovnem listu »podatki« s samodejnim filtrom poiščite napačni podatek in ga popravite na 26 (predpostavimo, da je bila napaka pri vnosu podatkov). ♦

ANKETA 20

Vrnite se v vrtilno tabelo in uporabite ukaz Osveži podatke. Katere vrednosti v vrtilni tabeli se s tem spremenijo? ♦

4.5.3.2 Skrij

Ukaz **Skrij** uporabimo, če želimo, da se v tabeli posamezna vrstica ali stolpec ne pokažeta. Preden uporabimo ta ukaz, moramo označiti prvo celico v vrstici oz. stolpcu. Na primer, če želimo, da se v vrtilni tabeli na sliki 4.2 ne pokaže stolpec Skupna vsota, označimo celico z besedilom Skupna vsota in uporabimo ukaz **Skrij**; če želimo, da se ne pokaže vrstica z eno študentko, ki je lastnica stanovanja, označimo celico, v kateri piše »lastnik« in uporabimo ukaz **Skrij**:

D4 Skupna vsota			
A	B	C	D
Tule spustite polja strani			
Štej od Zap. šte.	Spol (m, ž)		
Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž	Skupna vsota
doma	20	31	51
lastnik	1	1	1
podnajemnik	7	33	40
študentski dom	14	18	32
(prazen)	2	2	2
Skupna vsota	41	85	126

A6 lastnik		
A	B	C
Tule spustite polja strani		
Štej od Zap. šte.	Spol (m, ž)	
Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž
doma	20	31
lastnik	1	1
podnajemnik	7	33
študentski dom	14	18
(prazen)	2	2
Skupna vsota	41	85

A6 podnajemnik		
A	B	C
Tule spustite polja strani		
Štej od Zap. šte.	Spol (m, ž)	
Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž
doma	20	31
podnajemnik	7	33
študentski dom	14	18
(prazen)	2	2
Skupna vsota	41	84

Isti rezultat dobimo, če odpremo padajoči seznam polja *Nastanitev* (...) in odstranimo kljukico pri vrednosti »lastnik«:

	A	B	C
1	Tule spustite polja strani		
2			
3	Štej od Zap. šte.	Spol (m, ž)	
4	Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž
5		20	31
6			1
7		7	33
8		14	18
9			2
10		41	85
11			
12			
13			
14			
15			
16			
17			

4.5.3.3 Združi in pokaži podrobnosti

Ukaz **Združi in pokaži podrobnosti** smo uporabili pri oblikovanju vrtilne tabele na slikah 4.5 in 4.6. V okviru tega ukaza izbiramo med štirimi možnostmi: **Skrij podrobnosti**, **Pokaži podrobnosti**, **Združi**, **Razdruži**. Prva in zadnja možnost sta dejavni samo, če smo v vrtilni tabeli pred tem že uporabili možnosti **Pokaži podrobnosti** ali **Združi**. Ob izvajanju ukaza **Združi** Excel loči med opisnimi, diskretnimi in zveznimi številskimi spremenljivkami.

Pri združevanju kategorij opisne spremenljivke moramo v vrtilni tabeli, preden uporabimo ukaz **Združi**, označiti celice s kategorijami, ki jih želimo imeti v eni skupini: na primer, če želimo v tabeli na sliki 4.2 (spodaj) za spremenljivko *Nastanitev (...)* združiti kategoriji »doma« in »lastnik«, ti dve celici označimo in nato uporabimo ukaz **Združi**:

	A	B	C	D	E	F	G	H	I
1	Tule spustite polja strani								
2									
3	Število anketiranih								
4	Nastanitev (doma, študentski dom, podnajemnik, lastnik)	Vsota							
5	doma	51							
6	lastnik	1							
7	podnajemnik	40							
8	študentski dom	32							
9	(prazen)	2							
10	Skupna vsota	126							
11									
12									
13									

Vrtilna tabela se ob tem poveča za en stolpec, ki pripada novemu polju *Nastanitev(...)*². To polje ima za vrednosti »doma« in »lastnik« skupno kategorijo »Skupina1«:

A4			fx Nastanitev (doma, študentski dom, podnajemnik, lastnik) ²	
	A	B	C	
1	Tule spustite polja strani			
2				
3	Število anketiranih			
4	Nastanitev (doma, študentski dom, podnajemnik, lastnik) ²	Nastanitev (doma, študentski dom, podnajemnik, lastnik)	Vsota	
5	Skupina1	doma		51
6		lastnik		1
7	podnajemnik	podnajemnik		40
8	študentski dom	študentski dom		32
9	(prazen)	(prazen)		2
10	Skupna vsota			126

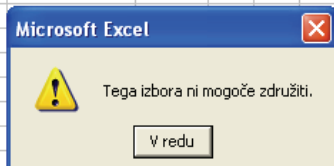
Če osnovno polje *Nastanitev(...)* odstranimo iz vrtilne tabele, dobimo naslednjo tabelo:

B4		fx Vsota	
	A	B	
1	Tule spustite polja strani		
2			
3	Število anketiranih		
4	Nastanitev (doma, študentski dom, podnajemnik, lastnik) ²	Vsota	
5	Skupina1		52
6	podnajemnik		40
7	študentski dom		32
8	(prazen)		2
9	Skupna vsota		126

Če kategorije, ki jih želimo združiti v skupino, v vrtilni tabeli ne nastopajo skupaj, jih pred uporabo ukaza **Združi** označimo tako, da med označevanjem držimo tipko CTRL.

Za številske spremenljivke možnost **Združi** omogoča, da enote podatkovne zbirke razvrstimo v razrede glede na vrednosti izbrane številske spremenljivke (polja) (podobno, kot če uporabimo orodje **Analiza podatkov/Histogram**). Številski spremenljivka ne sme imeti manjkajočih vrednosti (ni kategorije (*prazen*)). Če so med vrednostmi izbrane spremenljivke manjkajoče vrednosti, Excel vrne opozorilo: **Tega izbora ni mogoče združiti**.

	A	B	C	D	E	F	G	H
1	Tule spustite polja strani							
2								
3	Štej od Zap. šte.							
4	Število otrok v družini	Vsota						
5	1	11						
6	2	79						
7	3	21						
8	4	8						
9	5	2						
10	6	2						
11	8	2						
12	(prazen)	1						
13	Skupna vsota	126						



Preden uporabimo ukaz **Združi in pokaži podrobnosti/Združi**, mora biti v vrtilni tabeli označena ena izmed celic v vrstici oz. stolpcu z vrednostmi spremenljivke, po kateri želimo združevati (npr. ena izmed celic A5:A20 na sliki 4.5). V pogovornem oknu **Združevanje v skupine** določimo minimalno (**Začni pri:**) in maksimalno vrednost (**Končaj pri:**) ter širino razredov (**Po:**).

Za diskretne spremenljivke ukaz **Združi** naredi diskretne meje razredov, tako da točno vemo, katera vrednost sodi v posamezni razred (slika 4.5).

	A	B	C	D	E	F	G	H	I
1	Tule spustite polja strani								
2									
3	Število anketiranih								
4	Kolikokrat ste bili v zadnjem letu v kinu?	Vsota							
5	0	10							
6	1	14							
7	2	19							
8	3	17							
9	4	10							
10	5	20							
11	6	7							
12	7	3							
13	8	2							
14	9	3							
15	10	13							
16	13	1							
17	15	3							
18	20	2							
19	25	1							
20	30	1							
21	Skupna vsota	126							

	A	B	C	D	E	F	G
1	Tule spustite polja strani						
2							
3	Število anketiranih						
4	Kolikokrat ste bili v zadnjem letu v kinu?	Vsota					
5	0-4	70					
6	5-9	35					
7	10-14	14					
8	15-19	3					
9	20-24	2					
10	25-30	2					
11	Skupna vsota	126					

Slika 4.5: Število anketiranih glede na število obiskov v kinu v zadnjem letu (zgoraj), vrednosti, združene v razrede (spodaj), (ukaz **Združi in pokaži podrobnosti/Združi**, diskretna številska spremenljivka)

Za zvezne številske spremenljivke ukaz **Združi** meje razredov določi tako, da zgornja meja, ki definira interval vrednosti, ni vključena v dani razred; na primer, na sliki 4.6 so vsi, ki so bili na internetu na teden v povprečju manj kot 5 ur v prvem razredu in vsi, ki so na teden na internetu v povprečju od 5 do pod 10 ur v drugem razredu.

Pri združevanju v razrede moramo biti pozorni na to, da se v vrtilni tabeli razredi, v katerih ni nobenega podatka, ne izpišejo (slika 4.6 desno – ni razredov 35-40 ur in 40-45 ur).

	A	B
1	Tule spustite polja strani	
2		
3	Število anketiranih	
4	Koliko ur na teden v povprečju porabite za deskanje po svetovnem spletu?	Vsota
5	0	2
6	1	10
7	2	16
8	2,5	1
9	3	15
10	3,5	1
11	4	7
12	5	18
13	6	5
14	7	18
15	8	5
16	8,5	1
17	10	13
18	12	1
19	14	1
20	15	3
21	16	1
22	20	4
23	25	1
24	28	1
25	30	1
26	50	1
27	Skupna vsota	126

	A	B	C	D	E	F	G
1	Tule spustite polja strani						
2							
3	Število anketiranih						
4	Koliko ur na teden v povprečju porabite za deskanje po svetovnem spletu?	Vsota					
5	0-5	52					
6	5-10	47					
7	10-15	15					
8	15-20	4					
9	20-25	4					
10	25-30	2					
11	30-35	1					
12	45-50	1					
13	Skupna vsota	126					

Združevanje v skupine [X]

Samodejno

Začni pri: [0]

Končaj pri: [50]

Po: [5]

[V redu] [Prekliči]

Slika 4.6: Število anketiranih glede na število ur deskanja po svetovnem spletu na teden (levo), vrednosti, združene v razrede (desno), (ukaz **Združi in pokaži podrobnosti/Združi**, zvezna spremenljivka)

4.5.4 Seznam polj vrtilne tabele

Okno *Seznam polj vrtilne tabele* se pojavi ob vrtilni tabeli, če je le-ta aktivna (imamo označeno poljubno celico znotraj vrtilne tabele). Če tega okna ne vidimo, ga dobimo s klikom na desno ikono v orodni vrstici **Vrtilna tabela**:

Seznam polj vrtilne tabele [X]

Povlecite elemente v poročilo vrtilne tabele

- Zap. štev.
- Študijska smer (BT ali MB)
- Leto rojstva
- Starost
- Spol (m, ž)
- Število otrok v družini
- Srednja šola, ki ste jo zaključili

Dodaj v: Področje vrstice

- Področje vrstice
- Področje stolpca
- Področje strani
- Področje podatkov


Vrtilna tabela [X]

Vrtilna tabela [ikon] [ikon] [ikon] [ikon] [ikon] [ikon] [ikon]

Prikaži seznam polj

Obstoječo vrtilno tabelo lahko, podobno kot pri uporabi **Čarovnika za vrtilne tabele** in kliku na gumb **Postavitev**, spreminjamo tudi z vlečenjem polj vrtilne tabele iz **Seznama polj vrtilne tabele** v posamezna področja vrtilne tabele. Namesto vlečenja lahko uporabimo gumb **Dodaj v**, pred tem pa na desnem seznamu izberemo željeno področje vrtilne tabele (**Vrstica, Stolpec, Stran, Podatki**) in želena polja v seznamu polj vrtilne tabele.

4.5.5 Nastavitve polj vrtilne tabele

Pogovorno okno **Polje vrtilne tabele** dobimo z ukazom **Nastavitve polj** v meniju **Vrtilna tabela** oz. s klikom na ikono v orodni vrstici . V polju **Ime** je po osnovni nastavitvi napisano ime, ki se bo pojavilo v levi zgornji celici vrtilne tabele. To ime je sestavljeno iz imena funkcije in imena spremenljivke (polja), npr. **Štej od Zap. šte.** in ga lahko po želji spremenimo (npr. v **Število anketiranih**). Nekaterne funkcije v polju **Povzemi** smo že spoznali. Na tem mestu bomo predstavili še možnosti, ki se odpirajo v polju **Pokaži podatke kot**, ki ga v pogovorno okno **Polje vrtilne tabele** dobimo s klikom na gumb **Možnosti**.

Na primer, če želimo, da se namesto absolutnega števila (frekvence) v posamezni celici področja **Podatki** pokažejo relativne vrednosti (relativne frekvence), izberemo eno izmed možnosti **% stolpca** oz. **% vrstice**:

	A	B	C	D	E	F	G	H	I
1	Tule spustite polja strani								
2									
3	Štej od Zap. šte.								
4	Število otrok v družini	Vsota							
5	1	8,7%							
6	2	62,7%							
7	3	16,7%							
8	4	6,3%							
9	5	1,6%							
10	6	1,6%							
11	8	1,6%							
12	(prazen)	0,8%							
13	Skupna vsota	100,0%							
14									
15									
16									
17									
18									
19									
20									
21									
22									

Polje vrtilne tabele

Izvorno polje: Zap. šte. V redu

Ime: Štej od Zap. šte. Prekliči

Povzemi:

- Vsota
- Štej
- Povprečje
- Maks
- Min
- Zmnožek
- Štetje šte.


Številka ...
Možnosti >>


Pokaži podatke kot:

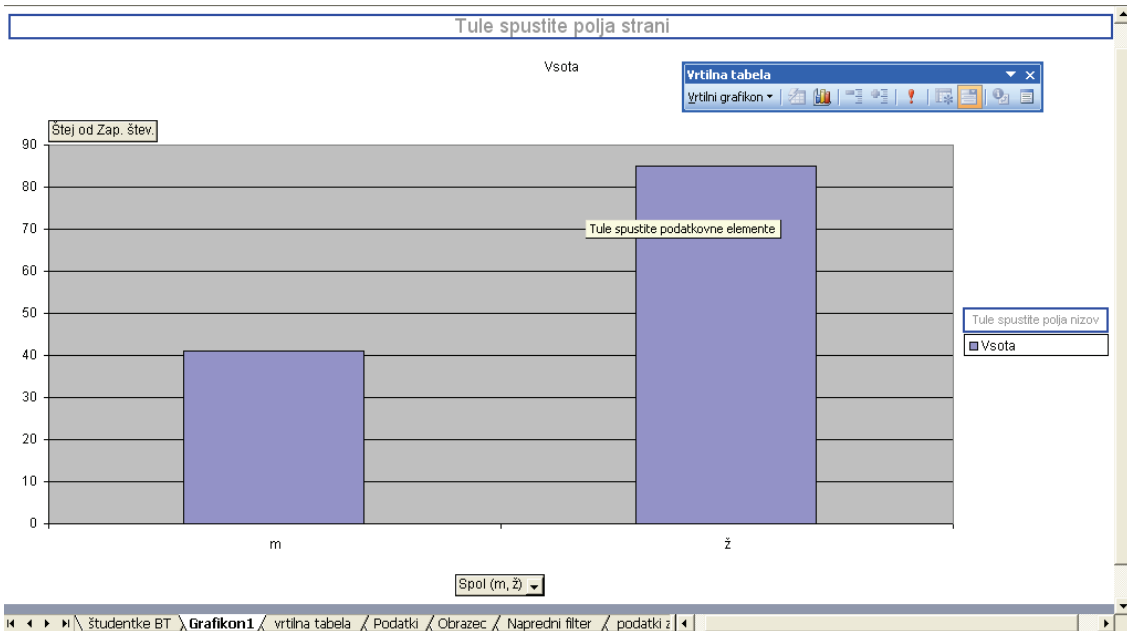
- % stolpca
- Navadno
- Razlika od
- % od
- % Razlike od
- Tekoča vsota v
- % vrstice
- % stolpca
- Število otrok v družin

V polju **Pokaži podatke kot** so na voljo še druge funkcije, ki jih v okviru te knjige ne bomo podrobneje opisali.

4.6 Vrtilni grafikon

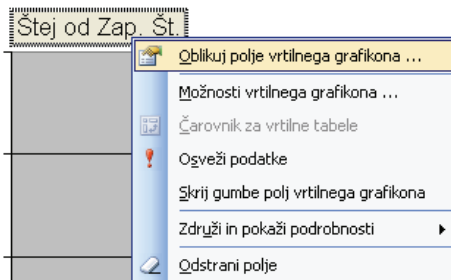
Podatke v vrtilnih tabelah grafično prikažemo s klikom na ikono  v orodni vrstici **Vrtilna tabela**. Grafikon lahko poljubno spreminjamo z ukazi in nastavitvami, ki jih že poznamo, poleg tega pa lahko izbor prikazanih vrednosti dodatno popravimo s pomočjo gumbov polj vrtilnega grafikona. Vrtilni grafikon uporabimo le za hitro grafično predstavitev podatkov, saj se spreminja s tem, ko spremenimo postavitev vrtilne tabele. Za končne grafične prikaze izbranih podatkov je bolje oblikovati grafikone na osnovi »parkiranih« vrtilnih tabel.

Poglejmo si primer vrtilnega grafikona. Če v vrtilni tabeli na sliki 4.1 (zgoraj) kliknemo na ikono , se odpre nov grafični delovni list z imenom »Grafikon 1« z naslednjim grafikonom:

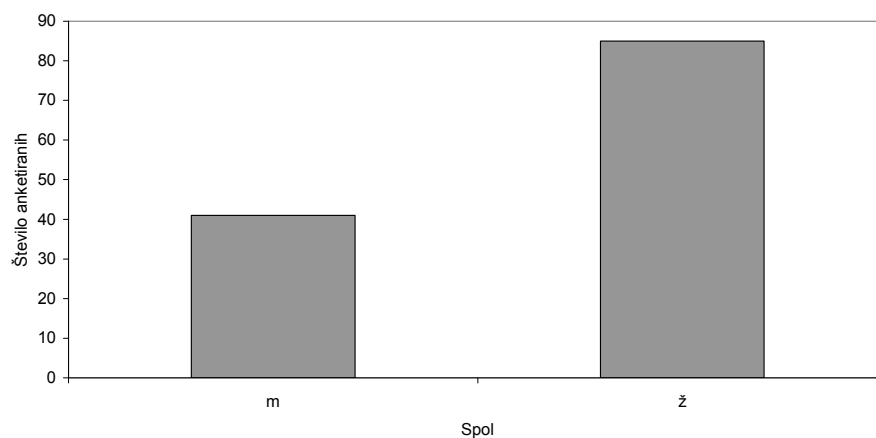


Siva gumba **Štej od Zap. Štev.** in **Spol(m,ž)** igrata enako vlogo kot v vrtilni tabeli, v vrtilnem grafikonu pa imata poleg tega tudi vlogo naslova Y oz. X osi. Z desnim klikom

na enega izmed teh dveh gumbov dobimo meni ukazov, s katerimi na različne načine spreminjamo grafikon. Gumb na vrtilnem grafikonu naredimo nevidne z ukazom **Skrrij gumb polj vrtilnega grafikona**:



V orodni vrstici **Vrtilna tabela** se prva ikona spremeni v ukaz **Vrtilni grafikon**. Če gumb vrtilnega grafikona odstranimo, jih dobimo nazaj s ponovnim klikom na ukaz **Skrrij gumb polj vrtilnega grafikona** v seznamu ukazov, ki se pokaže ob kliku na ikono **Vrtilni grafikon**. Na vrtilni grafikon brez gumbov je potrebno dodati naslove osi (**Možnosti grafikona**, kartica *Naslovi*).



Slika 4.7: Preoblikovani vrtilni grafikon, pripadajoč vrtilni tabeli na sliki 3.1.

Za vsako vrtilno tabelo Excel sicer naredi vrtilni grafikon, ki pa ni nujno ustrezen prikaz danih podatkov. Na primer, če z vrtilnim grafikonom prikažemo podatke druge vrtilne tabele na sliki 4.1, dobimo grafični prikaz z naloženimi stolpci. Taki naloženi stolpci nikakor ne predstavljajo vsebinsko primerenega načina za grafični prikaz povprečij.

S podrobnejšo razlago vrtilnih grafikonov se na tem mestu ne bomo ukvarjali, raje se bomo držali navodila, da grafikone oblikujemo na podlagi »parkiranih« vrtilnih tabel z uporabo **Čarovnika za grafikone**.

ANKETA 21

V datoteki ANKETA_ANALIZA.XLS uporabite ukaze v meniju **Podatki** in odgovorite na spodnja vprašanja. Odgovore napišite v datoteko ANKETA_ANALIZA.DOC, kamor kopirajte tudi narejene tabele in grafikone.

Katerega leta se je rodilo največ anketiranih? Koliko? Katero mero sredine predstavlja to leto rojstva?

Koliko anketiranih živi v družinah z več kot tremi otroki?

Kolikšna je povprečna, minimalna in maksimalna starost anketiranih? Kolikšne so te vrednosti posebej za moške in za ženske?

Kolikšna sta povprečje in standardni odklon za telesno maso ter telesno višino moških in žensk?

Razmislite o smiselnem številu decimalnih mest za izračunane vrednosti v tabelah. ♦

Zgled 4.1: Meteorološki podatki – časovna vrsta

METEOROLOŠKI PODATKI 1

Odprite datoteko METEOROLOSKI PODATKI.XLS s podatki, ki jih je v obdobju od 13. 4. 2004 do 21. 5. 2004 zabeležila samodejna meteorološka postaja na poskusnem polju Biotehniške fakultete na Jamnikarjevi 101. Vse odgovore in obrazložitve napišite v datoteko MET_POD_ANALIZA.DOC. V datoteko kopirajte tudi grafične prikaze in preglednice. Opremite jih s podnaslovi (slike) oz. nadnaslovi (preglednice). ♦

METEOROLOŠKI PODATKI 2

Oglejte si vrsto podatkov v datoteki (datumi, čas, številske vrednosti, ...). Spremenljivkam določite mersko lestvico. Opišite obliko celic v posameznem polju podatkovne zbirke. Kolikšen je časovni razmik med podatki? ♦

METEOROLOŠKI PODATKI 3

Datoteka METEOROLOSKI PODATKI.XLS je zelo velika. Koliko zapisov je v datoteki (podatkovni zbirki)?

Navodilo: v spodnjo desno celico zbirke podatkov se prestavimo s hkratnim pritiskom tipk CTRL in END, v zgornjo levo celico pa s hkratnim pritiskom tipk CTRL in HOME. ♦

METEOROLOŠKI PODATKI 4

Z ukazom **Podatki/Filter** izberite podatke treh zaporednih dni (5. - 7. 5. 2004) in jih spravite v novo excelovo datoteko z imenom MET_POD_IZBOR.XLS. Nadaljnjo analizo bomo delali v tej datoteki.

Navodilo: S samodejnim filtrom izbrane zapise označimo najenostavneje tako, da najprej označimo zgornjo levo celico izbora in s hkratnim pritiskom treh tipk CTRL, SHIFT in END označimo vse izbrane zapise, nato uporabimo ukaz **Kopiraj**, se postavimo na delovni list v datoteko MET_POD_IZBOR.XLS ter tam uporabimo ukaz **Prilepi**. ♦

METEOROLOŠKI PODATKI 5

Ali so med podatki manjkajoče vrednosti? ♦

METEOROLOŠKI PODATKI 6

Izračunajte povprečne, minimalne in maksimalne dnevne temperature zraka. Vsebinsko obrazložite vrednosti v stolpcu »Skupna vsota«. Ime stolpca ni vsebinsko primerno. Zakaj? ♦

	A	B	C	D	E
1	Tule spustite polja strani				
2					
3		Datum			
4	Podatki	5.5.04	6.5.04	7.5.04	Skupna vsota
5	Povprečje od Temperatura zraka (oC)	14,1	14,3	10,0	12,8
6	Min od Temperatura zraka (oC)	11,5	7,8	6,4	6,4
7	Maks od Temperatura zraka (oC)	17,3	19,1	14,9	19,1

Čarovnik za vrtilne tabele in vrtilne grafikone - Postavitve

Zgradite poročilo vrtilne tabele z vlečenjem gumbov polj na desni na diagram na levi.

STRAN

VRSTICA

Datum

STOLPEC

Povprečje od Tempe

Min od Temperatura

Maks od Temperatur

Datum

Čas

Max. hitro

Min. hitro

Padavine

Temperat

Globalna d

Relativna

Povp. hitr

Smer vetri

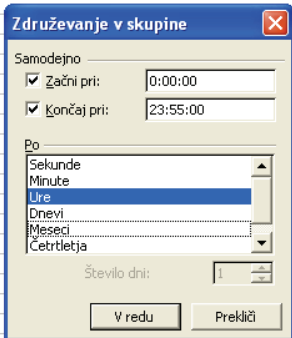
Pomoč V redu Prekliči

METEOROLOŠKI PODATKI 7

Izračunajte povprečne urne temperature zraka za celotno 3-dnevno obdobje in za vsak dan posebej. Izračune grafično prikažite. ♦

1. korak:

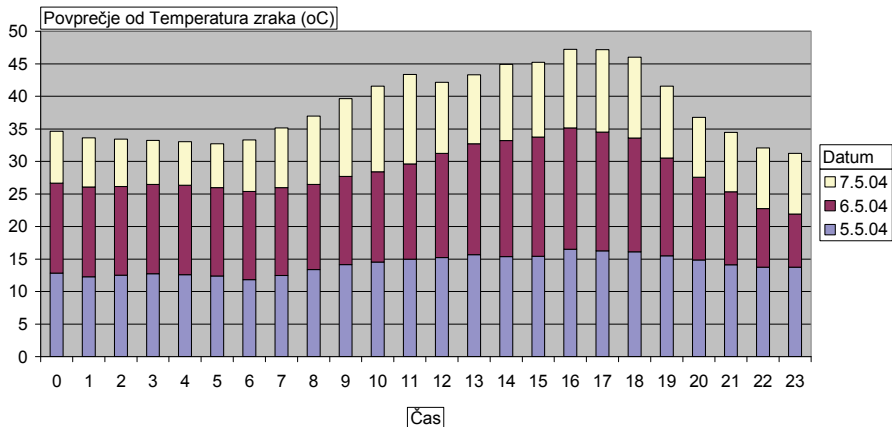
	A	B	C	D	E	
1	Tule spustite polja strani					
2						
3	Povprečje od	Temperatura zraka (oC)	Datum			
4	Čas	5.5.04	6.5.04	7.5.04	Skupna vsota	
5		0:00	12,8	13,8	8,5	11,7
6		0:05	12,9	13,8	8,3	11,7
7		0:10	13,2	13,8	8,4	11,8
8		0:15	13,2	13,8	8,2	11,7
9		0:20	13,0	13,8	8,0	11,6
10		0:25	12,9	13,8	8,1	11,6
11		0:30	12,7	13,8	7,8	11,4
12		0:35	12,6	13,9	7,8	11,4
13		0:40	12,7	13,9	8,0	11,5
14		0:45	12,6	13,9	7,9	11,5
15		0:50	12,6	13,9	7,5	11,3
16		0:55	12,6	13,8	7,2	11,2
17		1:00	12,7	13,8	7,3	11,3
18		1:05	12,7	13,8	7,0	11,2
19		1:10	12,6	13,8	6,9	11,1
20		1:15	12,5	13,9	6,8	11,1
21		1:20	12,1	13,9	7,0	11,0
22		1:25	12,0	13,9	7,5	11,1



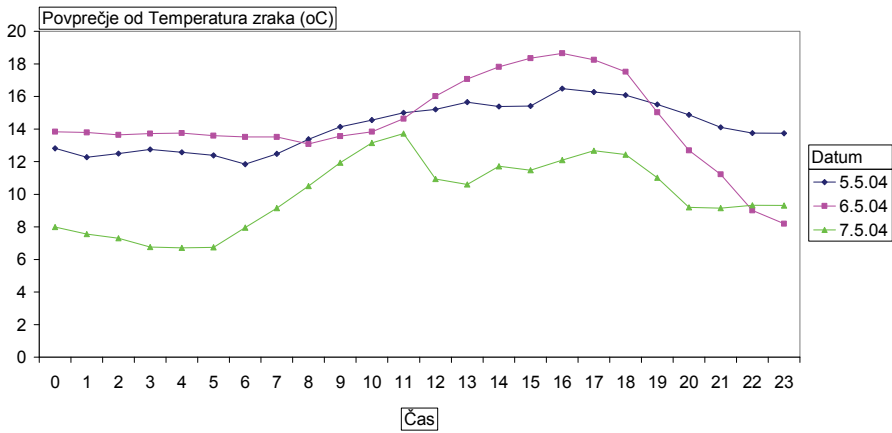
2. korak - po izvedenem ukazu **Združevanje v skupine**

Tule spustite polja strani				
Povprečje od Temperatura zraka (oC)	Datum			Skupna vsota
Čas	5.5.04	6.5.04	7.5.04	Skupna vsota
0	12,8	13,8	8,0	11,5
1	12,3	13,8	7,6	11,2
2	12,5	13,6	7,3	11,1
3	12,8	13,7	6,8	11,1
4	12,6	13,7	6,7	11,0
5	12,4	13,6	6,7	10,9
6	11,8	13,5	8,0	11,1
7	12,5	13,5	9,1	11,7
8	13,4	13,1	10,5	12,3
9	14,1	13,6	11,9	13,2
10	14,6	13,8	13,2	13,8
11	15,0	14,6	13,7	14,4
12	15,2	16,0	10,9	14,1
13	15,6	17,1	10,6	14,4
14	15,4	17,8	11,7	15,0
15	15,4	18,3	11,5	15,1
16	16,5	18,6	12,1	15,7
17	16,3	18,2	12,7	15,7
18	16,1	17,5	12,4	15,3
19	15,5	15,0	11,0	13,8
20	14,9	12,7	9,2	12,3
21	14,1	11,2	9,1	11,5
22	13,8	9,0	9,3	10,7
23	13,7	8,2	9,3	10,4
Skupna vsota	14,1	14,3	10,0	12,8

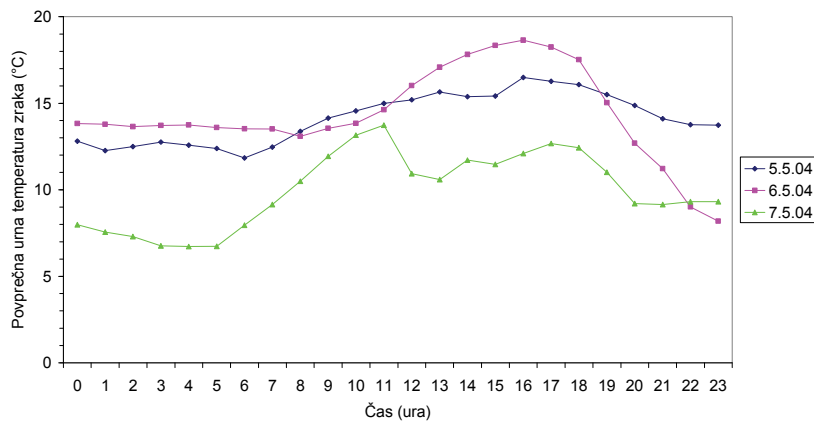
1. korak oblikovanja grafikona, ki prikazuje podatke zgornje vrtilne tabele: klik na ikono **Vrtilni grafikon**:



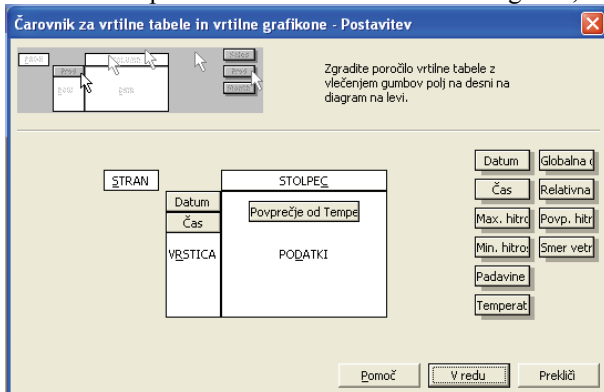
2. korak oblikovanja grafikona: vrsto grafikona spremenimo v Črtni:



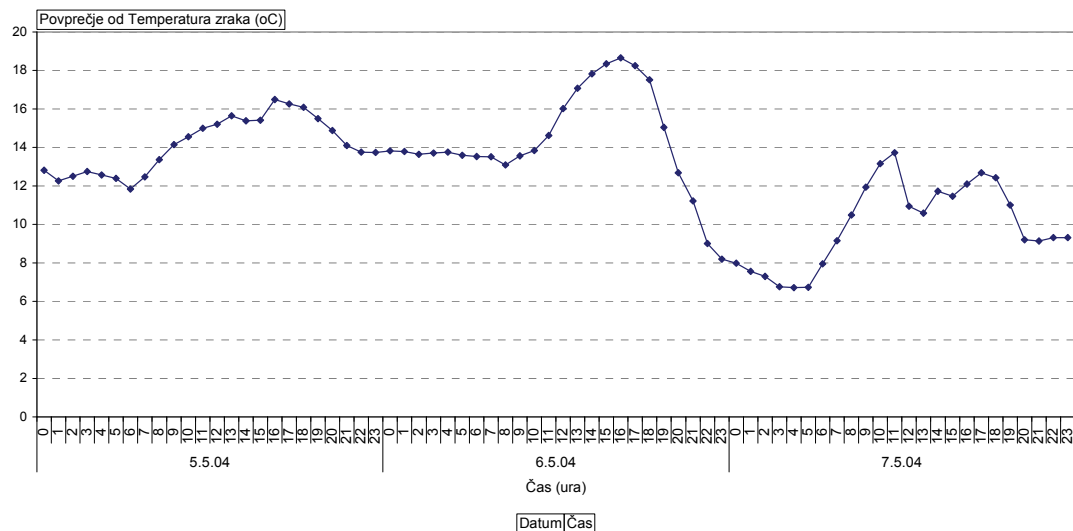
3. korak oblikovanja grafikona: odstranimo gumbe vrtilnega grafikona in napišemo naslove osi:



Če naredimo postavitev vrtilne tabele malo drugačno, sta v področju *Vrstica Datum* in *Čas*:



Vrtljni grafikon izgleda takole:



METEOROLOŠKI PODATKI 8

Izračunajte minimalno, maksimalno in povprečno dnevno hitrost vetra.

Tabela 4.2: Minimalna, maksimalna in povprečna dnevna hitrost vetra.

Podatki	Datum			Skupna vsota
	5.5.04	6.5.04	7.5.04	
Min od Min. hitrost vetra (m/s)	0,0	0,0	0,0	0,0
Maks od Max. hitrost vetra (m/s)	3,1	7,6	6,2	7,6
Povprečje od Povp. hitrost vetra (m/s)	0,2	1,4	0,5	0,7

Tabela 4.3: Minimalna, maksimalna in povprečna dnevna hitrost vetra. Vsebinsko smiselno spremenjena imena podatkov.

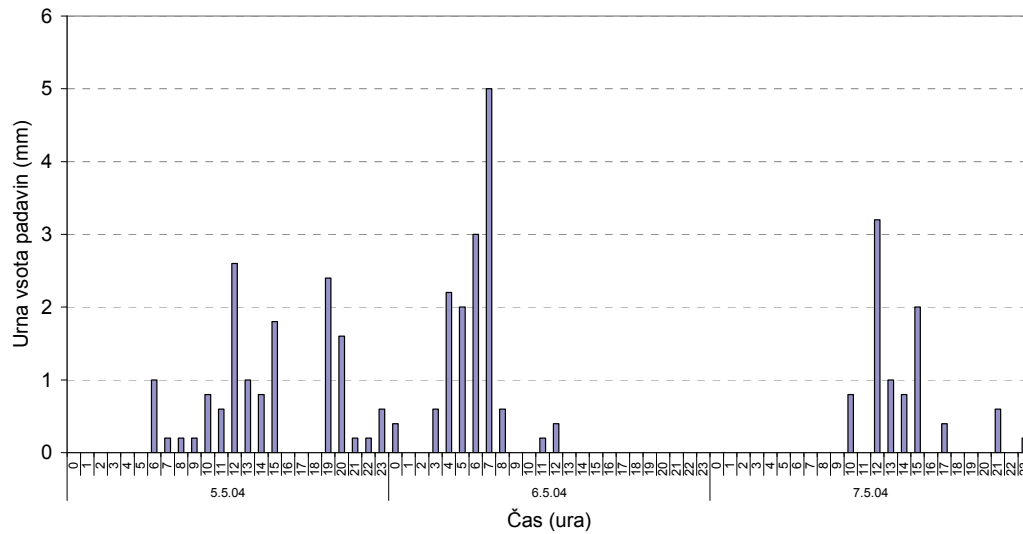
Statistike	Datum			
	5.5.04	6.5.04	7.5.04	Skupaj
Minimalna dnevna hitrost vetra (m/s)	0,0	0,0	0,0	0,0
Maksimalna dnevna hitrost vetra (sunek vetra) (m/s)	3,1	7,6	6,2	7,6
Povprečna dnevna hitrost vetra (m/s)	0,2	1,4	0,5	0,7

METEOROLOŠKI PODATKI 9

Izračunajte urne vsote padavin po dnevih in jih grafično prikažite. ♦

Tabela 4.4: Dnevne in urne vsote padavin za dano obdobje (opomba: vsot po urah, ki jih dobimo v zadnjem stolpcu vrtilne tabele, ne potrebujemo v "parkirani" vrtilni tabeli).

Vsota od Padavine (mm)	Datum		
Čas	5.5.04	6.5.04	7.5.04
0	0,0	0,4	0,0
1	0,0	0,0	0,0
2	0,0	0,0	0,0
3	0,0	0,6	0,0
4	0,0	2,2	0,0
5	0,0	2,0	0,0
6	1,0	3,0	0,0
7	0,2	5,0	0,0
8	0,2	0,6	0,0
9	0,2	0,0	0,0
10	0,8	0,0	0,8
11	0,6	0,2	0,0
12	2,6	0,4	3,2
13	1,0	0,0	1,0
14	0,8	0,0	0,8
15	1,8	0,0	2,0
16	0,0	0,0	0,0
17	0,0	0,0	0,4
18	0,0	0,0	0,0
19	2,4	0,0	0,0
20	1,6	0,0	0,0
21	0,2	0,0	0,6
22	0,2	0,0	0,0
23	0,6	0,0	0,2
Skupna vsota	14,2	14,4	9,0



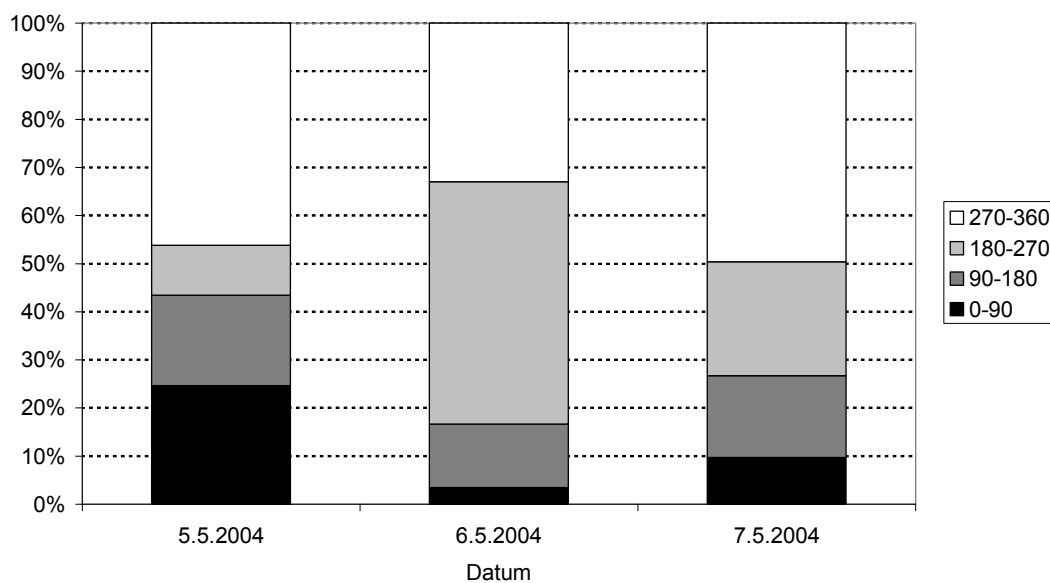
Slika 4.8: Urne vsote padavin v danem obdobju.

METEOROLOŠKI PODATKI 10

Oblikujte vrtilno tabelo, s katero boste za vsakega od izbranih dni posebej pokazali, v kolikih odstotkih primerov je veter pihal iz določene smeri neba. Smeri neba naj bodo združene v štiri razrede širine 90°: 0° do pod 90°, 90° do pod 180°, 180° do pod 270° in 270° do pod 360°. Porazdelitev grafično prikažite. ♦

Tabela 4.5: Struktura vetra glede na njegovo smer v treh dnevih meritev.

Štej od Datum	Datum			Skupna vsota
Smer vetra	5.5.04	6.5.04	7.5.04	
0-90	24,7%	3,5%	9,7%	12,6%
90-180	18,8%	13,2%	17,0%	16,3%
180-270	10,4%	50,3%	23,6%	28,1%
270-360	46,2%	33,0%	49,7%	42,9%
Skupna vsota	100,0%	100,0%	100,0%	100,0%

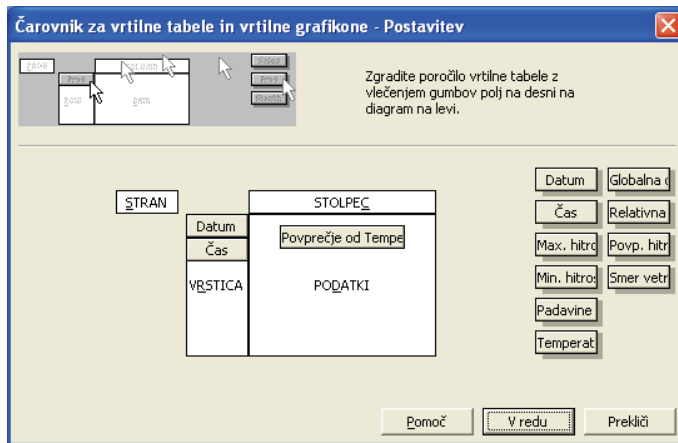
Slika 4.9: Struktura smeri vetra v treh dnevih meritev. Za oblikovanje grafikona je potrebno v vrtilni tabeli zamenjati vrstice in stolpce: *Smer vetra* gre v področje **Stolpec**, *Datum* pa v področje **Vrstica**.

METEOROLOŠKI PODATKI 11

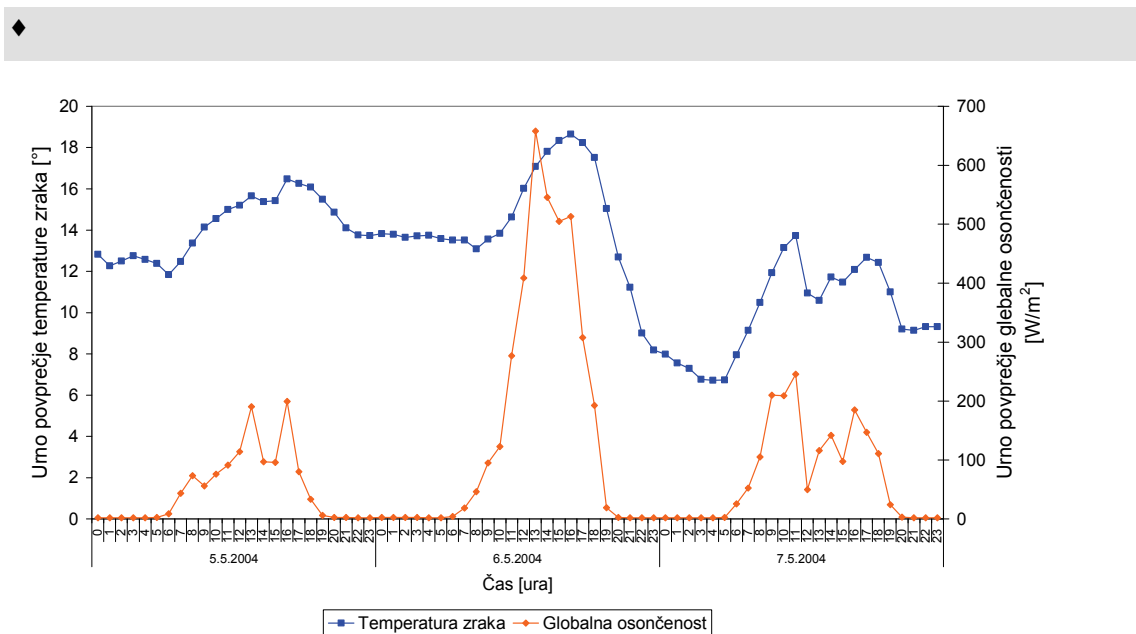
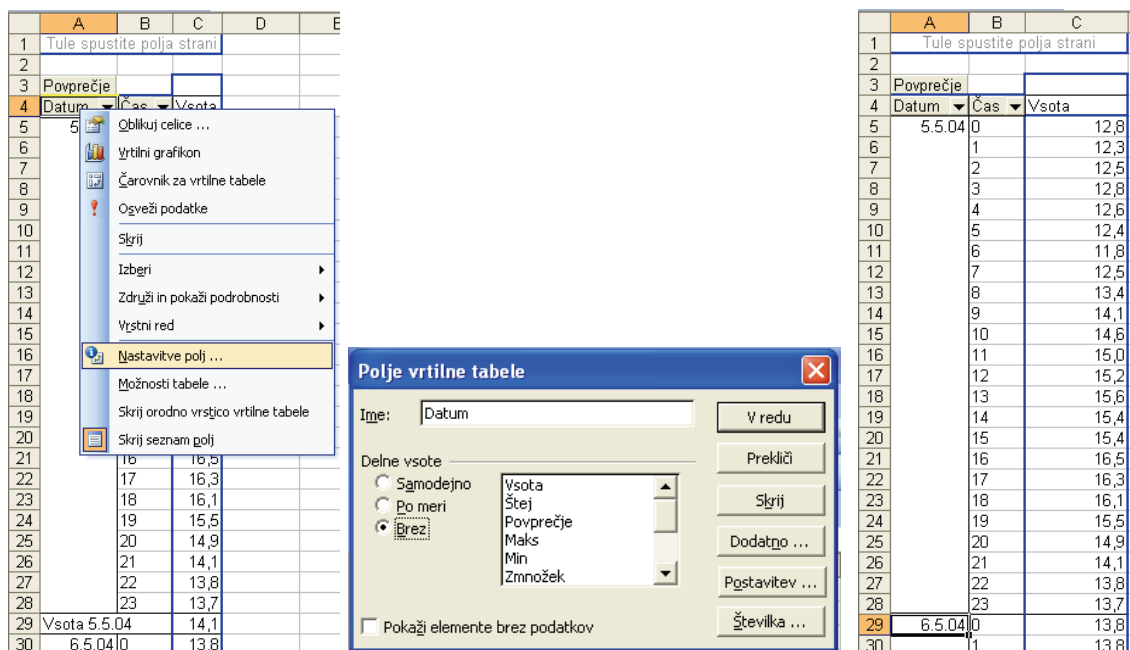
Izračunajte povprečno urno osončenost (W/m^2) in povprečno urno temperaturo zraka ($^{\circ}C$) po dnevih v obdobju 5. - 7. 5. 2004 ter ju grafično prikažite na isti sliki (slika 4.10).

Navodilo: Za izračun uporabite orodje **Vrtilna tabela**. Primerno razporejenih vrednosti obeh časovnih vrst v vrtilni tabeli ne moremo dobiti hkrati. Podatke za oblikovanje grafikona si pripravimo tako, da ustrezno »parkiramo« rezultate dveh vrtilnih tabel (enkrat za urno osončenost in drugič za temperaturo zraka).

Osnovna postavitev polj za vrtilno tabelo je sledeča:



Podatke združimo po urah, nato pa izpis vrednosti dodatno oblikujemo tako, da se v tabeli znebimo vrstic *Vsota 5.5.04*, *Vsota 6.5.04* in *Vsota 7.5.04*. V teh vrsticah so izračunana dnevna povprečja izbranih spremenljivk, ki na želeni grafikon ne sodijo. Te vrstice so zato v parkiranih tabelah nepotrebne. V vrtilni tabeli se tem vrsticam reče *Delne vsote* in se jih lahko znebimo z ukazom **Nastavitve polj** (za polje *Datum*), kjer v razdelku *Delne vsote* izberemo možnost **Brez**:



Slika 4.10: Povprečna urna temperatura zraka in povprečna urna osončenost za dano obdobje.

METEOROLOŠKI PODATKI 12

Urne vsote padavin in urno povprečje temperature zraka ustrezno grafično prikažite na istem grafikonu.

Navodilo: Padavine prikažite s stolpci, vrednosti naj bodo na sekundarni osi Y, temperaturo zraka prikažite s črto, vrednosti naj bodo na primarni osi Y. ♦

5 Opisne statistike in grafični prikazi s programom R

Novo:

- osnove programa R in grafičnega vmesnika R Commander
- opisne statistike, okvir z ročaji, histogram z neenako širokimi razredi v R-u
- slučajno vzorčenje v R

Datoteke s podatki:

ANKETA.XLS

Ustvarjene datoteke:

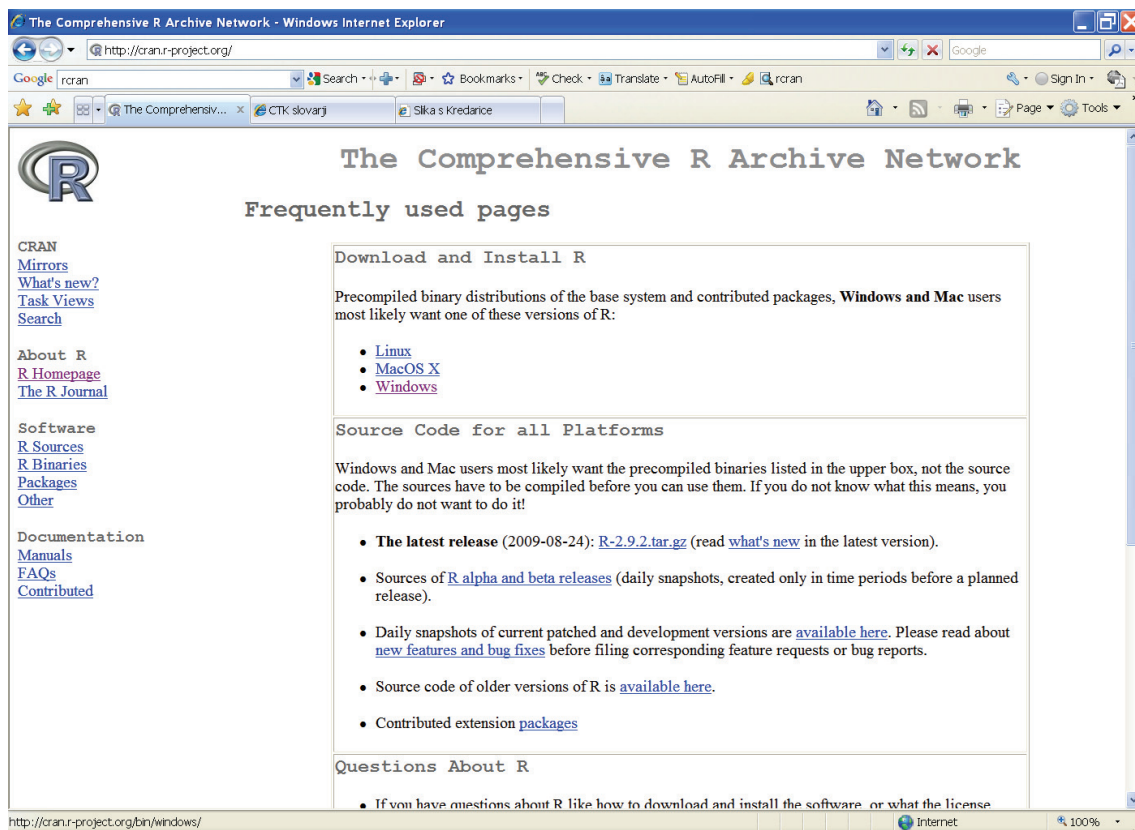
ANKETA_R.DOC

5.1 Program R in grafični vmesnik R Commander

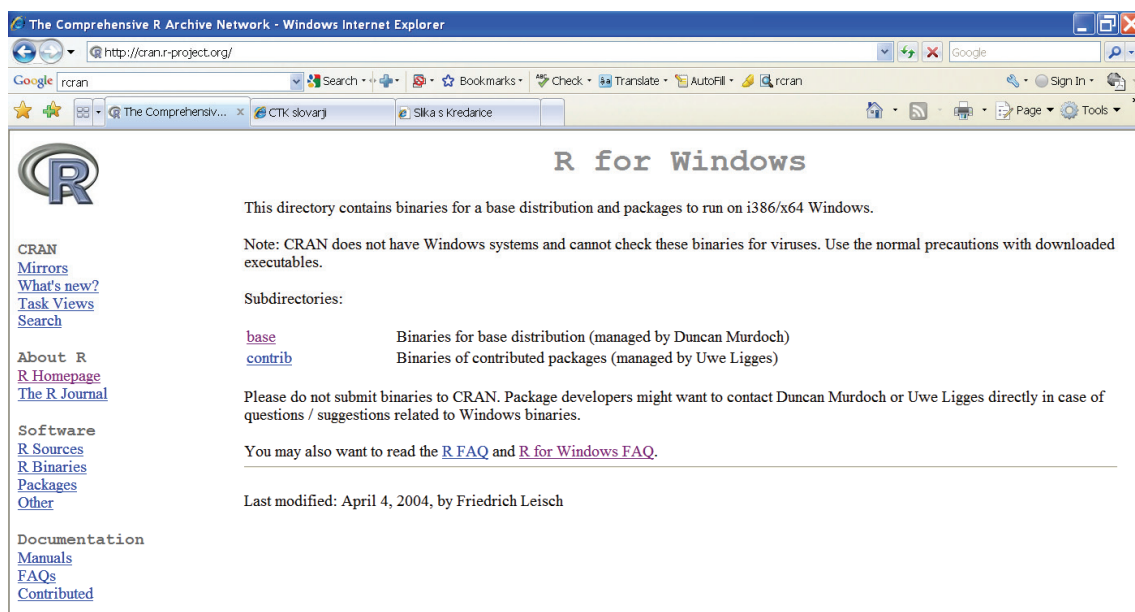
5.1.1 Kako naložiti program R?

Datoteko **R-2.9.2-win32.exe**¹⁷, ki jo potrebujemo za naložitev programa R na osebni računalnik, najdemo na spletni strani <http://cran.r-project.org>:

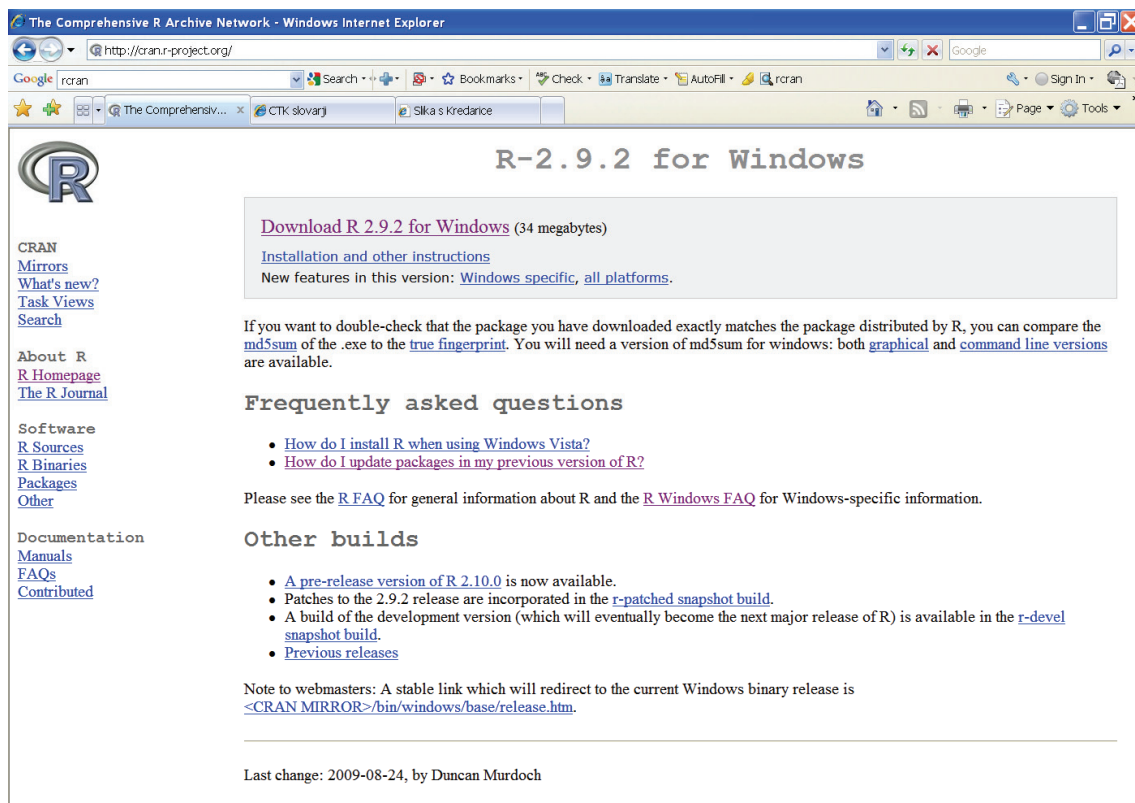
¹⁷ Program večkrat na leto dopolnijo in objavijo novo verzijo, verzija 2.9.2 je bila aktualna oktobra 2009.



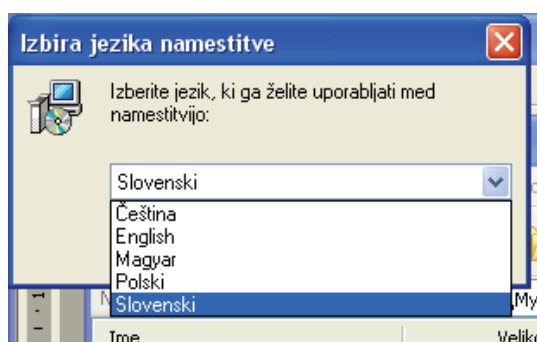
Po kliku na možnosti *Windows* se odpre pogovorno okno, v katerem izberemo možnost *base*:



In v naslednjem prikazu kliknemo ukaz **Download R 2.9.2 for Windows**. S tem na izbrano mesto v računalnik shranimo datoteko **R-2.9.2-win32.exe**.

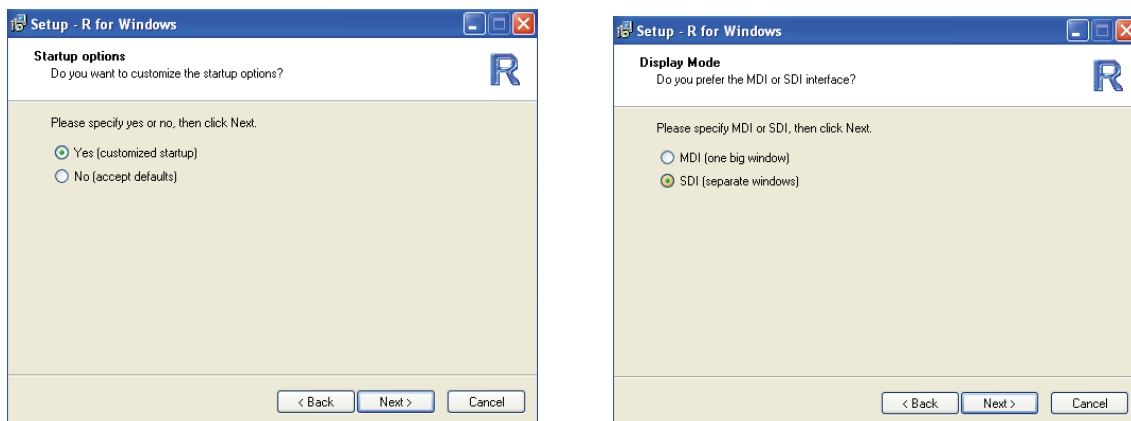



Ko zaženemo datoteko **R-2.9.2-win32.exe**, se vključi čarovnik za namestitev programa. Najprej v oknu **Izbira jezika namestitve** izberemo možnost *Slovenski*:

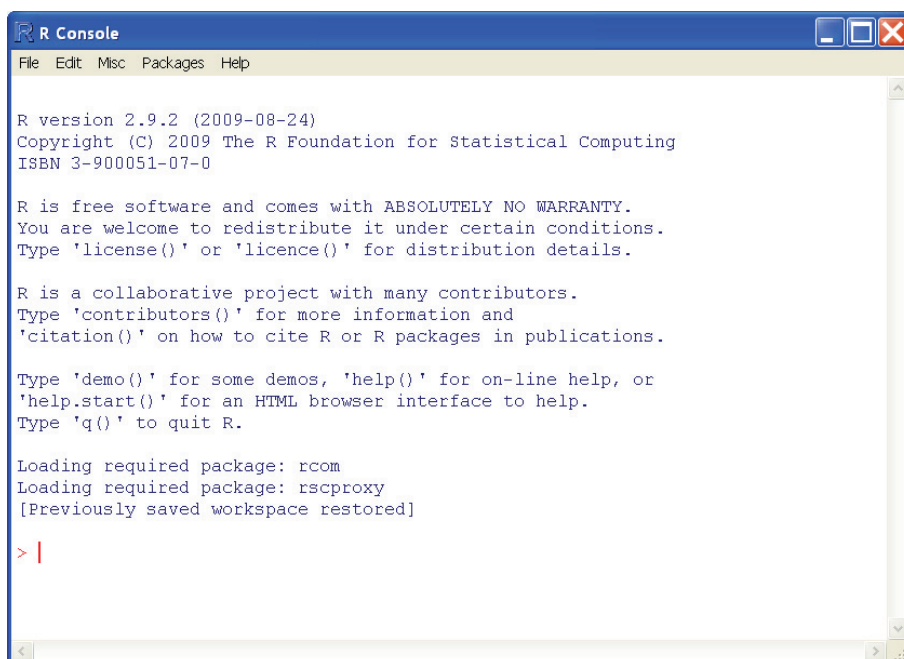


Skoraj vse naprej izbrane nastavitve v korakih čarovnika so primerne za osnovno namestitev programa. Spremembo moramo narediti le v dveh korakih: v oknu **Startup Options** izberemo možnost *Yes (customized setup)*, v naslednjem oknu **Display Mode**

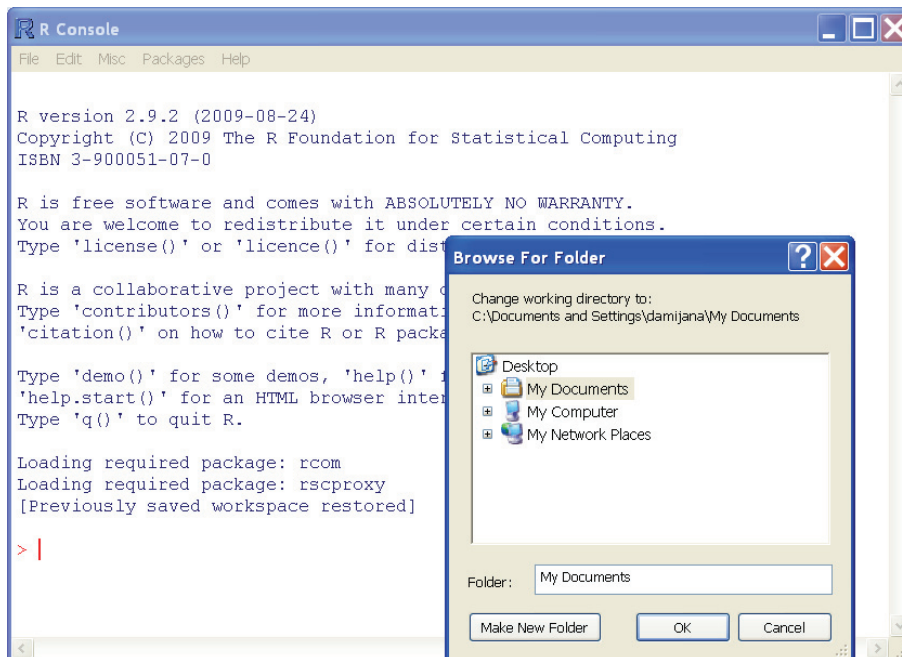
pa izberemo možnost *SDI* (*separate windows*). To izbiro pri namestitvi programa zahteva kasnejša uporaba grafičnega vmesnika **R Commander**.



Program R aktiviramo s klikom na ikono  na namizju. Ob tem se odpre okno R-ovega vmesnika **R Console** z osnovno menijsko vrstico (File, Edit, Misc, Packages, Help). To okno je namenjeno pisanju ukazov in izpisu rezultatov.



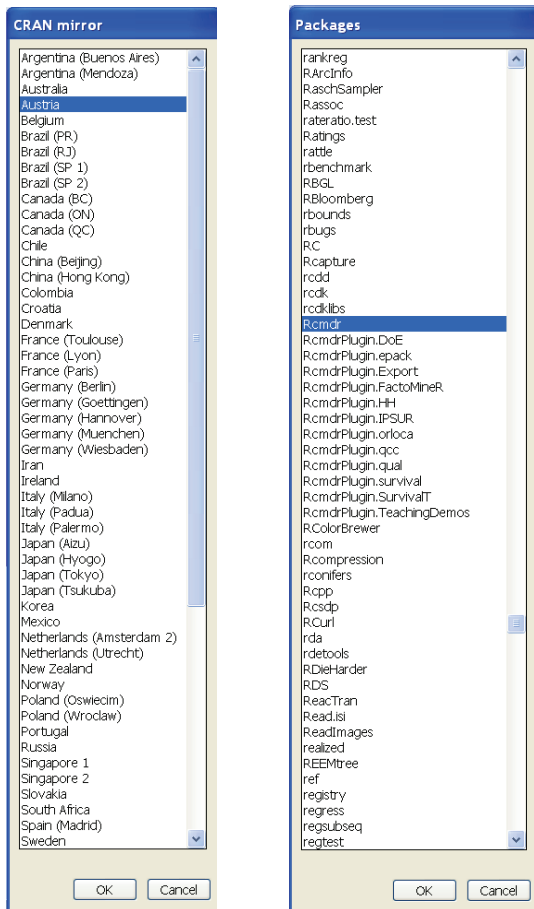
Preden nadaljujemo z delom, je praktično, da določimo delovno mapo, v kateri imamo shranjene datoteke s podatki, in hkrati bomo v to mapo shranjevali ustvarjene datoteke. V oknu **R Console** uporabimo ukaz **File/Change dir...** in izberemo želeno mapo.



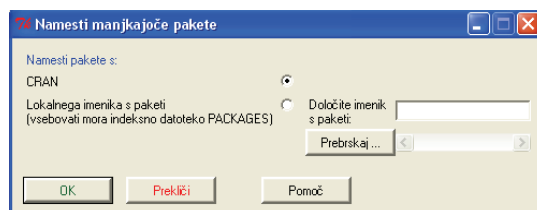
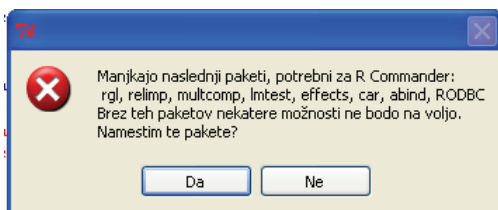
Pri vajah bomo za pisanje ukazov uporabljali grafični vmesnik **R Commander**. Tega najdemo med R-ovimi paketi, ki so na voljo na strežnikih organizacije CRAN. Paket se imenuje **Rcmdr** in jo moramo najprej shraniti v mapo Library programa R (več o tem v nadaljevanju).

5.1.2 Kako naložiti R Commander?

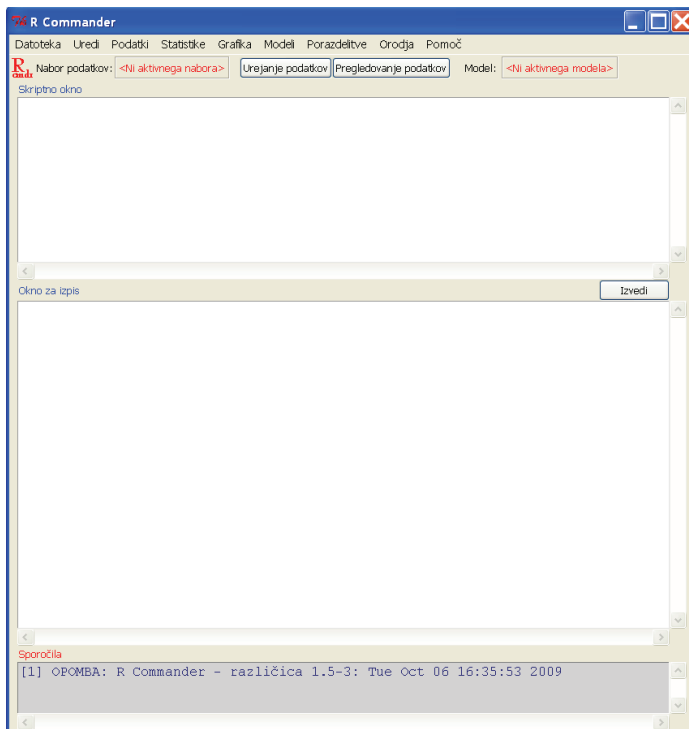
Grafični vmesnik **R Commander** je zapisan v paketu funkcij (package) **Rcmdr**. Paket moramo najprej naložiti (datoteke shraniti v mapo npr. C:\Program Files\R\R-2.9.2\library) s klikom na ukaz **Packages/Install Package(s)** v osnovni menijski vrstici okna **R Console**. V meniju **CRAN Mirror** izberemo možnost *Austria*, v naslednjem oknu **Packages** pa izberemo paket **Rcmdr**. Ob tem se sproži nalaganje vseh potrebnih paketov.



Ko se v oknu **R Console** pokaže rdeč znak **>**, lahko paket aktiviramo s klikom na ukaz **Packages/Load Package** in v meniju **Select One** ponovno izberemo možnost **Rcmdr**. Ko to na računalniku naredimo prvič, se odpre dodatno okno, ki vsebuje informacijo o tem, da nam za uporabo grafičnega vmesnika R Commanderja manjka še nekaj knjižnic. Kliknemo gumb **Da** in v naslednjem oknu kliknemo gumb **OK**:



Nalaganje paketov poteka kar nekaj minut, nato se odpre okno R Commanderja:



Če po pomoti zapremo okno R Commanderja, ga lahko v oknu R Console znova odpremo z ukazom¹⁸:

> Commander()

Podatki in analize, ki so bile trenutno aktivne v okviru programa R Commander, se ob tem ne ohranijo, kar pomeni, da moramo paziti na to, da okna R Commanderja ne zapiramo po nepotrebnem med analizo podatkov.

Okno R Commanderja sestavljajo: menijska vrstica, pod njo je vrstica z pravičnimi gumbi, pod to vrstico so tri podokna: **Skriptno okno**, **Okno za izpis** in okno za **Sporočila**.¹⁹

Vnos podatkov, grafične predstavitve in različne statistične analize delamo z izbiro ukazov v menijih. Ob uporabi posameznega ukaza se v **Skriptnem oknu** pokaže izpis ukaza v obliki, kot bi ga morali napisati "peš", če bi analizo delali v oknu R Console. V **Skriptno okno** lahko ukaz tudi napišemo brez klikanja ukazov v menijih. Izvedbo

¹⁸ S ponovnim klikom na ukaz **Packages/Load Package** ne gre.

¹⁹ Menije in ukaze programa je poslovenil Jaro Lajovic.

ukaza dosežemo s klikom na gumb **Izvedi** (ali pa s hkratnim klikom na tipki CTRL in R). Ob tem se izvede ukaz, napisan v vrstici, v kateri se nahaja utripajoči kazalec. S potegom miške lahko označimo tudi več ukazov hkrati, ki se izvedejo drug za drugim po kliku na gumb **Izvedi**. Če je ukaz dolg več kot eno vrstico, se lahko nadaljuje v naslednji vrstici, le da mora biti besedilo zamaknjeno od levega roba za en ali več presledkov ali tabulatorjev.

V **Oknu za izpis** se pokažejo ukazi v rdeči barvi in izračuni v temno modri barvi. Grafikoni se pokažejo v samostojnem oknu R Graphics: Device 2 (številka je odvisna od tega, koliko grafičnih oken odpremo).

Tretje podokno R Commanderja je namenjeno sporočilom, opozorilom, opombam in poročanju o napakah, ki jih prepozna program R.

Ukaze v glavnih menijih **Datoteka**, **Uredi**, **Podatki**, **Statistike**, **Grafika**, **Modeli**, **Porazdelitve**, **Orodja** in **Pomoč** bomo postopno opisali sproti, ko jih bomo uporabljali v nadaljevanju besedila.

Pod glavno menijsko vrstico najdemo še nekaj gumbov, ki so namenjeni delu s podatki. Levo polje *Nabor podatkov* kaže ime trenutnega aktivnega nabora podatkov. Na začetku piše <Ni aktivnega nabora>. Če imamo v obdelavi več različnih naborov podatkov, dobimo seznam s klikom na ta gumb. Večina ukazov v menijih glavne orodne vrstice se navezuje na aktivni nabor podatkov. Izjema so ukazi v menijih **Datoteka**, **Uredi** in **Porazdelitve**.

Srednja dva gumba **Urejanje podatkov** in **Pregledovanje podatkov** omogočata urejanje in pregledovanje podatkov.

Desni gumb kaže ime trenutno aktivnega statističnega modela, o tem pa več kasneje.

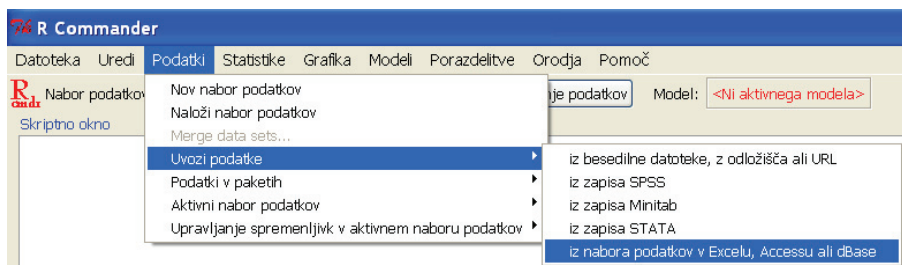
Po zagonu R Commanderja lahko okno R Console skrijemo.

Domača naloga 5.1: R na domačem računalniku

Namestite programa R in grafični vmesnik R Commander na domačem računalniku. ♦

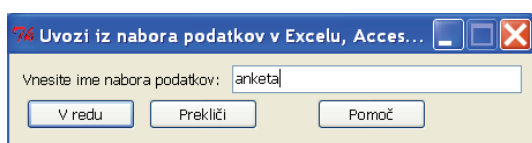
5.1.3 Vnos podatkov

Obstaja več načinov za vnos podatkov v aktivni nabor:

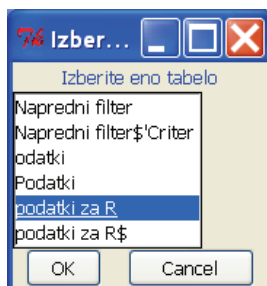


Neposredni vnos podatkov (tipkanje v R-ovo preglednico) izvedemo z ukazom **Podatki/Nov nabor podatkov**; to je smiselno za zelo majhne nabore podatkov in ga uporabljamo le izjemoma.

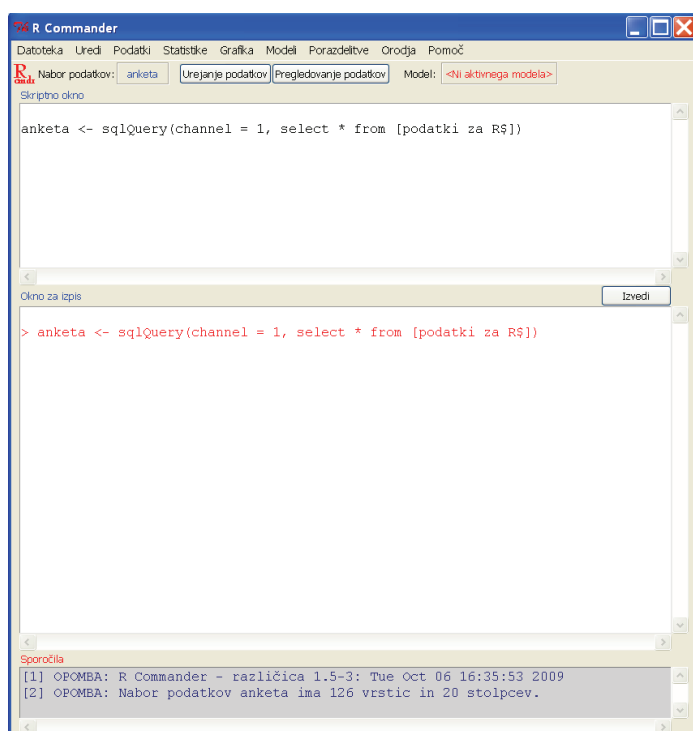
Navadno podatke vnesemo v aktivni nabor z ukazom **Podatki/Uvozi podatke**. Prva možnost je, da uvozimo podatke iz besedilne datoteke ali odlagališča (clipboard). Druge možnosti so namenjene odpiranju datotek, ustvarjenih v statističnih programih, (SPSS, Minitab ali Stata) pa tudi v programih Excel, Access ali dBase. Če želimo uvoziti podatke, ki so bili predhodno urejeni v Excelu, uporabimo ukaz **Podatki/Uvozi podatke/iz nabora podatkov v Excelu, Accessu ali dBase**. V pogovornem oknu *Uvozi iz nabora podatkov v Excelu, Accessu a...* v polje *Vnesite ime nabora podatkov* napišemo zeleno ime nabora (npr. anketa) in kliknemo na gumb **V redu**.



Za tem se odpre pogovorno okno **Open**, (okno Raziskovalca s prikazom vsebine mape, ki je bila izbrana z ukazom **File/Change dir...**) v katerem izberemo Excelovo datoteko in kliknemo na gumb **Open**. V naslednjem pogovornem oknu izberemo ime delovnega lista z izbranimi podatki v excelovi datoteki. V datoteki ANKETA.XLS so za analizo v R-u pripravljene podatki na delovnem listu »podatki za R«.



Po kliku na gumb OK v pogovornem oknu *Izberite eno tabelo* okno R Commanderja izgleda takole:



V program R smo uvozili podatke iz Excelove datoteke ANKETA.XLS. Ob tem se bomo seznanili z nekaterimi lastnostmi programa:

- R loči med velikimi in malimi črkami, zato moramo predhodno vrednosti posameznih spremenljivk v Excelovi datoteki poenotiti (npr. vrednosti spremenljivke smer studija so lahko »AG«, »BT« in »MB«, ne pa tudi »ag«, »bt« ali »mb«);

- R se ne znajde dobro, če uporabljamo šumnike, zato jih v datoteki s podatki spremenimo v sičnike tako v imenih spremenljivk, kot tudi v vrednostih spremenljivk (npr. spremenljivko višina spremenimo v visina; za spremenljivko spol sta vrednosti »z« in »m« in ne »ž« ali »Ž«,...);
- Imena spremenljivk v R-u so lahko sestavljena samo iz ene besede, sestavljene iz črk in števil, med njimi se od ločil lahko pojavi pika ali spodnji vezaj. Ime spremenljivke se ne sme začeti s številko. Temu ustrezno moramo spremeniti imena spremenljivk;
- R manjkajočim vrednostim številskih spremenljivk (prazne celice v excelovi datoteki) priredi vrednost **NA**, za opisne spremenljivke pa prazne celice ostanejo prazne tudi v R-ovem naboru podatkov;
- R za decimalno ločilo uporablja piko ne glede na to, kakšno nastavitev imamo v okolju Windows. Če imamo sicer v okolju Windows nastavljeno kot decimalno ločilo vejico, (tako je običajno v slovenskih nastavitvah) pri vnosu excelovih datotek decimalne vejice R sam spremeni v pike.

ANKETA 22

Naredite vnos podatkov iz datoteke ANKETA.XLS z delovnega lista »podatki za R« v R Commander. Nabor podatkov v R Commanderju poimenujte »anketa«. ♦

Klik na gumb Urejanje podatkov odpre okno **Data Editor**. V tem oknu podatke vidimo in jih lahko tudi popravljamo.

	ID	smer	leto_rojstva	starost	spol	st_otrok	gimnazija	tocke_matura
1	1	AG	1982	25	z	2	ne	NA
2	2	AG	1983	24	z	2	da	17
3	3	AG	1983	24	m	NA	da	16
4	4	BT	1984	23	z	2	da	NA
5	5	MB	1984	23	m	2	da	NA
6	6	BT	1984	23	m	3	da	21
7	7	AG	1984	23	z	2	da	17
8	8	AG	1984	23	z	2	ne	16
9	9	AG	1984	23	z	2	ne	NA
10	10	AG	1984	23	z	2	ne	15
11	11	BT	1985	22	m	2	da	19
12	12	MB	1985	22	z	2	ne	24
13	13	MB	1985	22	z	1	da	23
14	14	MB	1985	22	z	2	da	NA
15	15	MB	1985	22	m	2	ne	21
16	16	AG	1985	22	m	2	da	15
17	17	AG	1985	22	m	1	ne	NA
18	18	AG	1985	22	z	2	da	13
19	19	AG	1985	22	z	1	ne	12
20	20	AG	1985	22	z	3	da	15

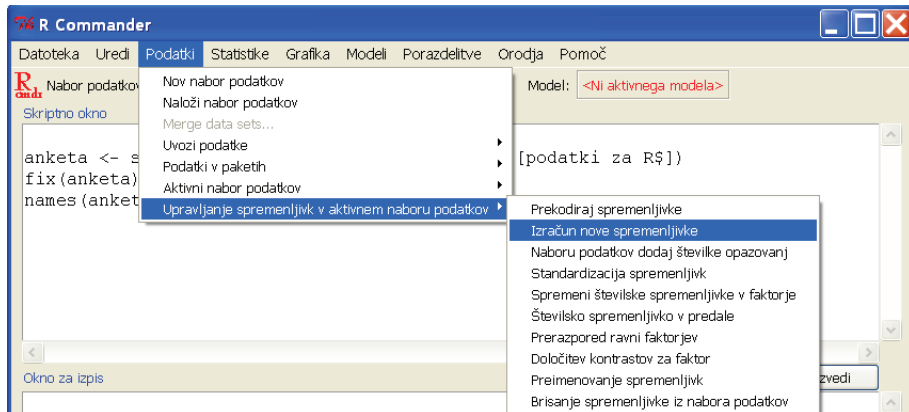
Če kliknemo na gumb **Pregledovanje podatkov**, se odpre okno, v katerem podatke le vidimo, popravljati pa jih ne moremo.

Seznam spremenljivk v aktivnem naboru podatkov se izpiše v **Okno za izpis** z ukazom **Podatki/Aktivni nabor podatkov/Spremenljivke v aktivnem naboru podatkov**.

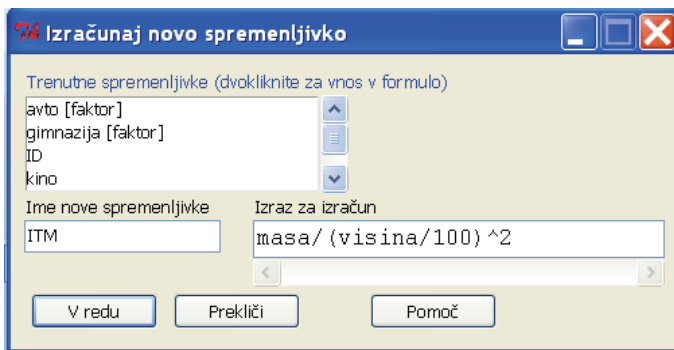
The screenshot shows the R Commander interface. The 'Podatki' menu is open, and the path 'Podatki > Aktivni nabor podatkov > Spremenljivke v aktivnem naboru podatkov' is selected. The 'Okno za izpis' window shows the following output:

```
> anketa <- sqlQuery(channel = 1, select * fr
> fix(anketa)
> names(anketa)
[1] "ID"          "smer"        "leto_rojstva" "starost"     "spol"
[6] "st_otrok"    "gimnazija"   "tocke_matura" "stipendija"  "naj_predmet"
[11] "povp_ocena" "avto"        "nastanitev"   "masa"        "visina"
[16] "sport"      "sport_ure"   "kino"         "knjige"      "splet"
```

Če želimo aktivnemu naboru podatkov dodati novo spremenljivko, izračunano na podlagi že obstoječih spremenljivk, uporabimo ukaz **Podatki/Upravljanje spremenljivk v aktivnem naboru podatkov/Izračun nove spremenljivke**:



Odpre se pogovorno okno *Izračunaj novo spremenljivko*, v polju *Ime nove spremenljivke* napišemo izbrano ime (npr. *ITM* kot kratica za indeks telesne mase), v polje *Izraz za izračun* pa napišemo formulo, ki vsebuje imena že obstoječih spremenljivk in matematične znake (=, +, -, *, /, ^). Imena spremenljivk so lahko tudi argumenti funkcij, katerih imena morajo biti napisana z malimi črkami (npr. za kvadratni koren se enako kot v Excelu uporablja funkcija `sqrt()`). Za izračun *ITM* napišemo formulo: $masa/(visina/100)^2$ (Opomba: *visina* je v izrazu za *ITM* v metrih in ne v centimetrih).



ANKETA 23

Aktivnemu naboru z imenom »anketa« dodajte spremenljivko indeks telesne mase (*ITM*).

Navodilo: $ITM = \frac{masa}{visina^2}$ z enotami kg/m^2 ♦

ANKETA 24

Z ukazom **Podatki/Aktivni nabor podatkov/Shrani aktivni nabor podatkov** shranite nabor podatkov »anketa«. Na ta način R podatke s spremenljivko *ITM* vred shrani v datoteko z imenom ANKETA.RDA. Naslednjič podatke naložimo v R z ukazom **Podatki/Naloži nabor podatkov** v pogovornem oknu, ki se odpre, ko izberemo datoteko ANKETA.RDA. ♦

5.2 Opisne statistike in grafični prikazi

Odprite novo wordovo datoteko ANKETA_R.DOC in vanjo kopirajte izračune in grafične prikaze, ki jih zahtevajo naloge v nadaljevanju tega poglavja. ♦

5.2.1 Številski povzetki

Program R različno obravnava opisne in številske spremenljivke. Ko želimo izračunati različne statistike, nam ponudi le tiste spremenljivke, ki so primerne za izračun izbrane statistike (npr. povprečje, ...)

Povprečje, standardni odklon in kvantile za posamezno spremenljivko dobimo z ukazom **Statistike/Povzetki/Številski povzetki**. Za telesno višino (*visina*) posebej za dekleta in posebej za fante moramo v pogovornem oknu **Številski povzetki** izbrati naslednje možnosti:

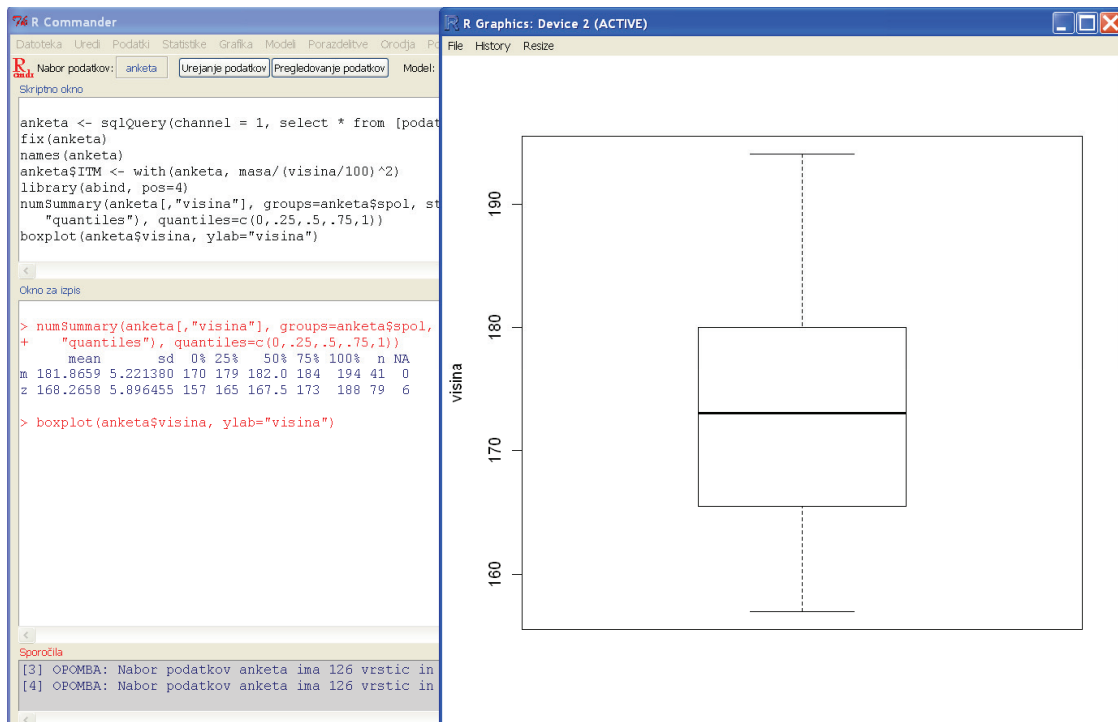
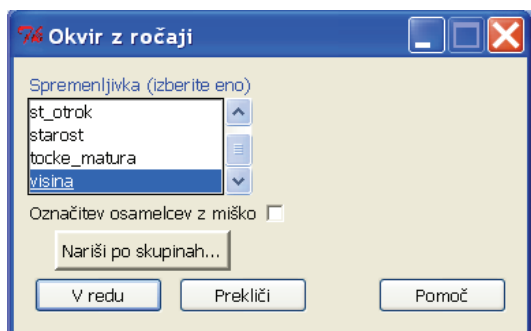
The screenshot shows the R Commander interface. The 'Številski povzetki' dialog box is open, showing the variable 'visina' selected. The 'Spremenljivke za skupine' list includes 'smer', 'spol', 'sport', and 'stipendija'. The 'Izvedi' button is visible at the bottom right of the dialog. The main R console window shows the command: > numSummary(anketa[,\"visina\"], groups=anketa\$spol, statistics=c(\"mean\", \"sd\", \"quantiles\"), quantiles=c(0, .25, .5, .75, 1)) and the resulting summary table:

	mean	sd	0%	25%	50%	75%	100%	n	NA
m	181.8659	5.221380	170	179	182.0	184	194	41	0
z	168.2658	5.896455	157	165	167.5	173	188	79	6

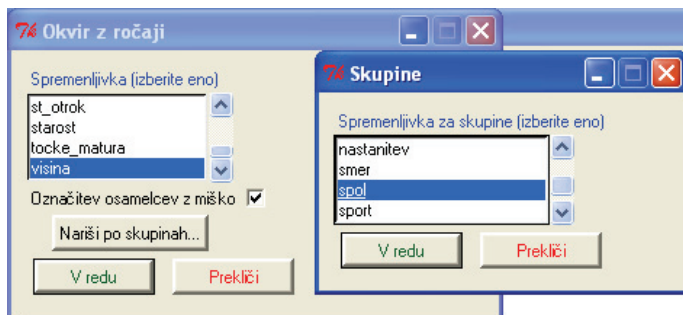
V Oknu za izpis se poleg povprečja (**mean**), standardnega odklona (**sd**), minimuma (0 %), maksimuma (100 %) ter kvartilov (25, 50 in 75 %) izpiše tudi število podatkov v vzorcu (**n**) in število manjkajočih vrednosti (**NA**).

5.2.2 Okvir z ročaji

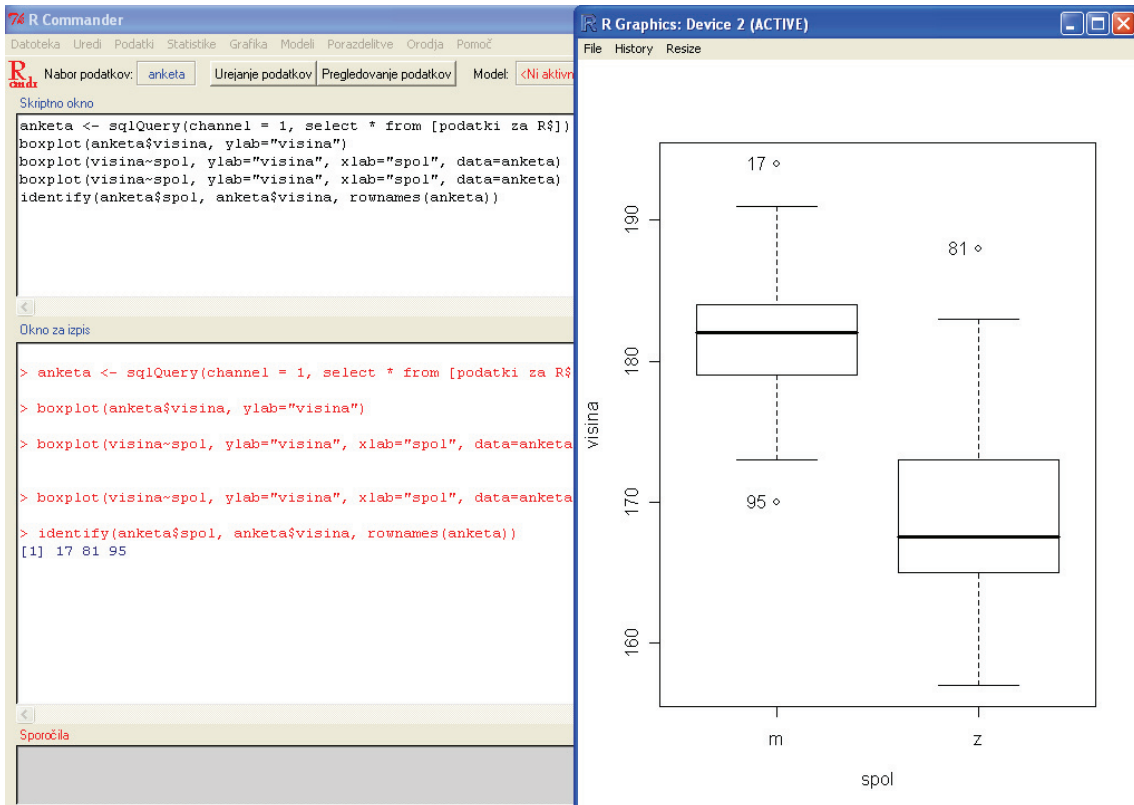
Porazdelitev vrednosti številske spremenljivke grafično prikažemo z okvirjem z ročaji z ukazom **Grafika/Okvir z ročaji**. V polju **Spremenljivka (izberite eno)** s klikom izberemo ime spremenljivke (na seznamu so samo številske spremenljivke).



Če želimo, da R nariše več okvirjev z ročaji naenkrat za različne skupine, (npr. posebej za moške in posebej za ženske) v pogovornem oknu **Okvir z ročaji** kliknemo na gumb **Nariši po skupinah** in izberemo spremenljivko, po kateri naj R razdeli podatke v skupine. Če so med podatki osamelci, jih R ob okvirju z ročaji nariše s krogcem, identificiramo pa jih z izbiro možnosti **Označitev osamelcev z miško**.

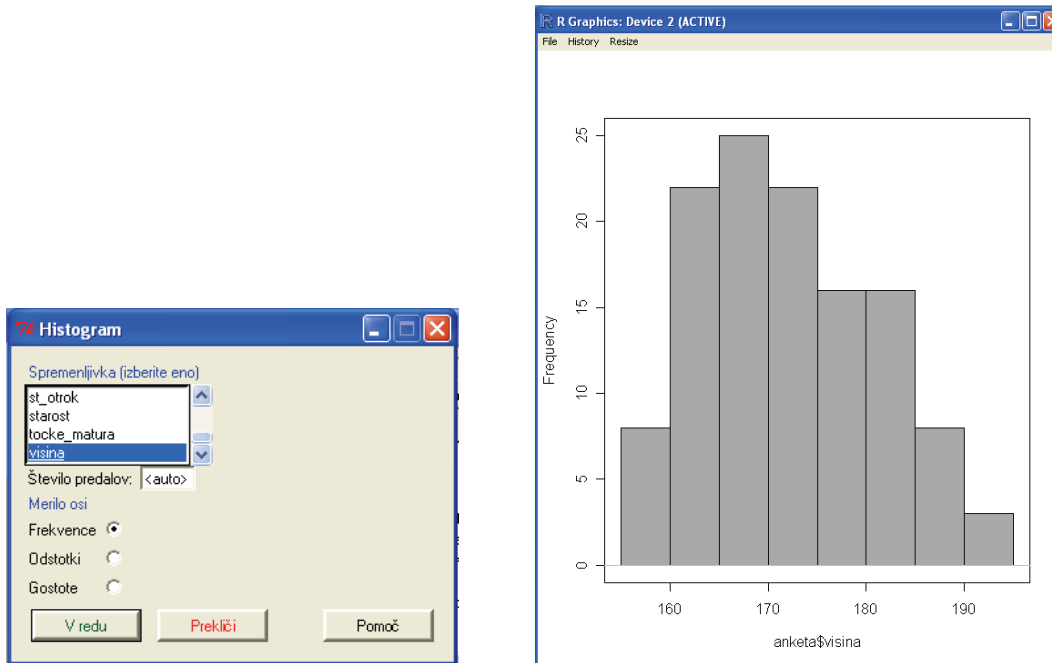


Ob izbiri možnosti **Označitev osamelcev z miško** se kazalec miške v oknu Graphics: Device 2(ACTIVE) spremeni v križec, katerega postavimo na krogec, ki prikazuje osamelec in uporabimo levi klik miške. Ob tem se ob krogec na sliko izpiše zaporedna številka podatka v naboru podatkov. Ko želimo označevanje osamelcev zaključiti, uporabimo desni klik miške in izberemo ukaz **Stop**. Seznam zaporednih številčk podatkov, ki predstavljajo osamelce, se zapiše tudi v Okno za izpis:



5.2.3 Histogram

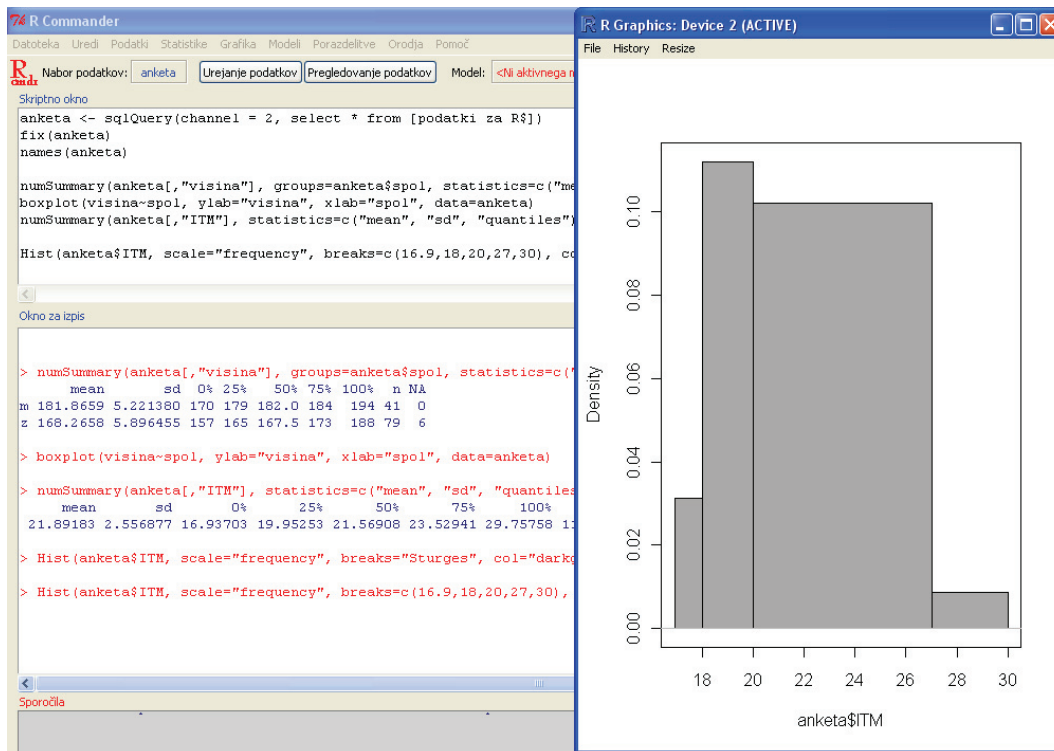
Histogram narišemo z ukazom **Grafika/Histogram**:



V polju **Število predalov** po osnovni nastavitvi piše <avto>, kar pomeni, da R sam izbere število enako širokih razredov. Če napišemo določeno število razredov, se R Commander na to ne odzove. Med možnostmi **Merilo osi** izberemo, kaj želimo prikazati: frekvence (*frequency*), relativne frekvence (odstotki, *percentage*) ali gostote (*density*) za posamezni razred. Če želimo meje razredov določiti sami (lahko so tudi različno široki), spodnje meje razredov napišemo v ukazno vrstico v Skriptnem oknu. Primer: če želimo indeks telesne mase porazdeliti v razrede *do pod 18*, *18 do pod 20*, *20 do pod 27* ter *27 do pod 30*, v ukazu **Hist** spremenimo argument **breaks**; argument **scale** mora imeti vrednost "density", saj so razredi različno široki, in rišemo gostote:

```
Hist(anketa$ITM, scale="density", breaks=c(16.9,18,20,27,30), col="darkgray")
```

R zahteva tudi spodnjo mejo prvega razreda, v zgornjem primeru smo izbrali minimum podatkov: 16,9.

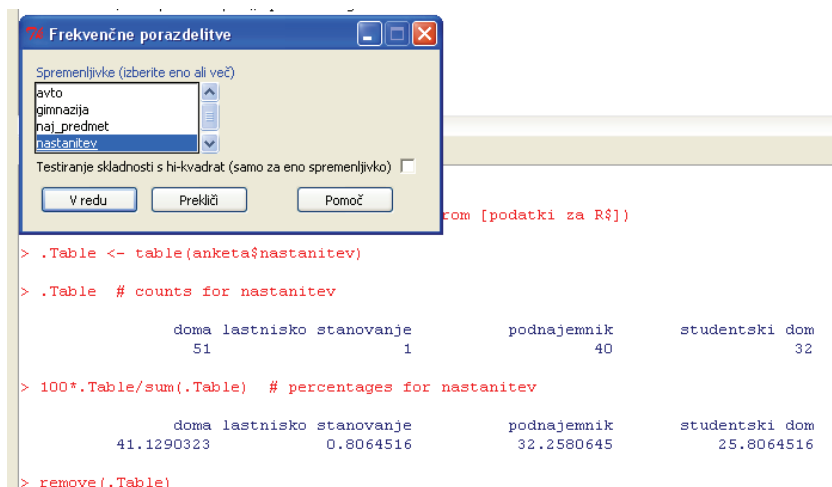


ANKETA 25

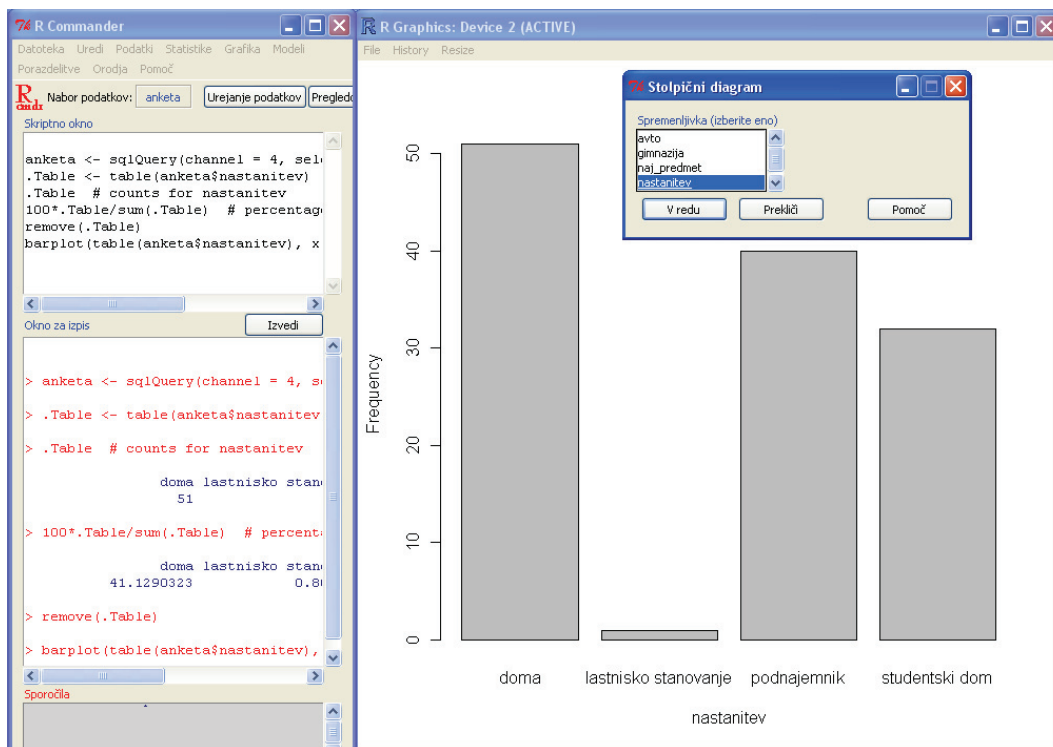
Za spremenljivko *ITM* narišite histogram s štirimi razredi, katerih meje naj določajo kvartili. ♦

5.2.4 Frekvenčna porazdelitev in stolpični diagram

Frekvenčno tabelo lahko naredimo le za opisne spremenljivke z ukazom **Statistike/Povzetki/Frekvenčne porazdelitve**:



Tu med spremenljivkami najdemo samo opisne spremenljivke. Prikaz števila posameznih vrednosti opisne spremenljivke dobimo z ukazom **Grafika/Stolpčni diagram**:



Domača naloga 5.2: Analiza ankete

ANKETA 26

Za spremenljivke *kino*, *knjige*, *splet*, *sport_ure* v aktivnem naboru »anketa« izračunajte povprečje, standardni odklon, minimum, maksimum in kvartile. Anketirane pri tem razdelite v skupine po *spolu*. ♦

ANKETA 27

Porazdelitev vrednosti spremenljivk *kino*, *knjige*, *splet*, *sport_ure* grafično prikažite z okvirji z ročaji. Anketirane pri tem razdelite v skupine po spolu. Obrazložite grafikone. ♦

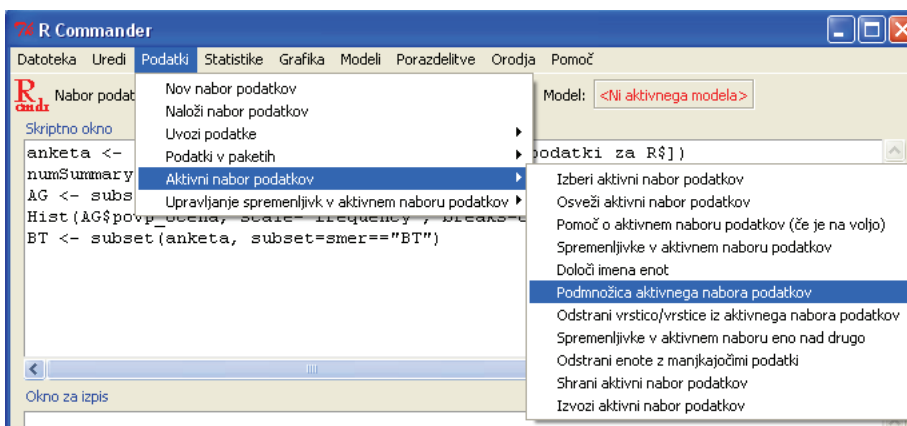
ANKETA 28

Za vsako smer študija za spremenljivki *povp_ocena* v aktivnem naboru »anketa« izračunajte povprečje, standardni odklon, minimum, maksimum in kvartile. Za vsako smer študija posebej podatke prikažite s histogramom, katerega razredi naj bodo določeni s kvartili. Obrazložite rezultate.

Navodilo: Najprej izračunamo kvartile za spremenljivko *povp_ocena* po smereh študija:

```
> numSummary(anketa[,"povp_ocena"], groups=anketa$smer, statistics=c("mean",
"sd", "quantiles"), quantiles=c( 0,.25,.5,.75,1 ))
```

Za oblikovanje histograma za posamezno smer moramo najprej narediti tri nove nabore podatkov - za vsako smer študija enega. Iz obstoječega nabora podatkov »anketa« dobimo nov nabor podatkov z ukazom **Podatki/Aktivni nabor podatkov/Podmnožica aktivnega nabora podatkov**:



V pogovorno okno *Podmnožica podatkov* v polje *Izraz za izbiro podmnožice* napišemo: smer == "AG". V polje *Ime za nov nabor podatkov* pa napišemo npr. AG.

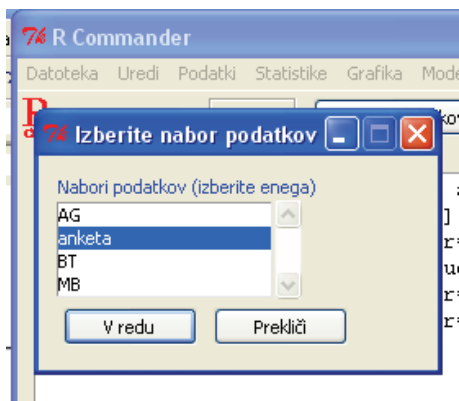
To ponovimo še za BT in MB, izpis v skriptnem oknu je naslednji:

```
AG <- subset(anketa, subset=smer=="AG")
```

```
BT <- subset(anketa, subset=smer=="BT")
```

```
MB <- subset(anketa, subset=smer=="MB")
```

Če v oknu R Commanderja kliknemo v polje *Nabor podatkov*, dobimo pogovorno okno *Izberite nabor podatkov*, v katerem je seznam naborov podatkov, med katerimi izbiramo, preden uporabimo nadaljnje ukaze.



Izberemo npr. nabor podatkov »AG« in za oblikovanje okvirja z ročaji uporabimo ukaz **Grafika/Histogram**, v drugem koraku pa v skriptnem oknu ukaz popravimo:

```
Hist(AG$povp_ocena, scale="density", breaks=c(3, 7, 7.7, 8.25, 9.9), col="darkgray")
```

◆

ANKETA 29

Oblikujte frekvenčne tabele in stolpčne grafikone za vrednosti opisnih spremenljivk v aktivnem naboru »anketa«. Obrazložite rezultate za eno opisno spremenljivko (izberite si jo sami). ◆

5.3 Slučajno vzorčenje v R

Funkcijo `sample` v programu R uporabimo za različne načine izbire enot v vzorec. Funkcija ima v splošnem štiri argumente:

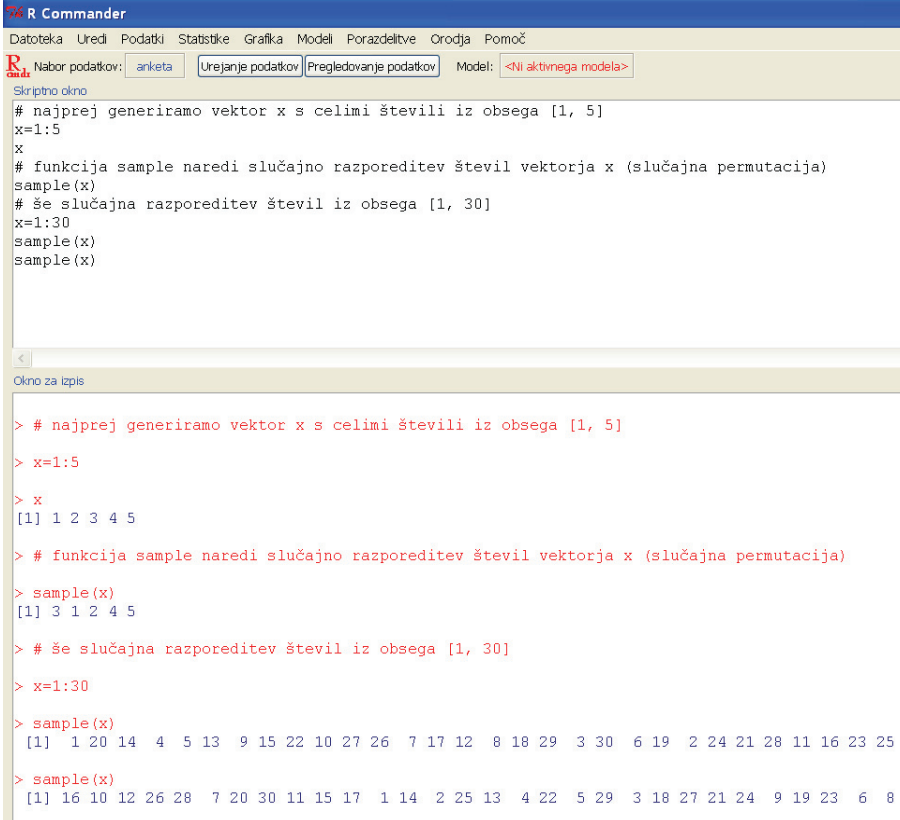
```
sample(x, size, replace = FALSE, prob = NULL)
```

`x` je vektor vrednosti, izmed katerih izbiramo; `size` je velikost vzorca - koliko enot želimo izbrati; `replace` ima po osnovni nastavitvi vrednost `FALSE`, kar pomeni, da vzorčimo brez vračanja enot v populacijo; `prob` je vektor uteži, ki predstavljajo verjetnost izbire posamezne enote v vzorec in ima po osnovni nastavitvi vrednost `NULL`, kar pomeni, da ima vsaka enota vektorja `x` enako verjetnost, da je izbrana v vzorec. V R Commanderju moramo ukaz, da se funkcija `sample` izvede, napisati v Skriptno okno (ne gre s klikanjem).

Če funkciji `sample` določimo samo argument `x` in je to vektor celoštevilčnih vrednosti iz obsega $[m, n]$, dobimo slučajno razporeditev vrednosti iz izbranega obsega. Na

primer, če želimo slučajno razporeditev celih števil iz obsega $[1, 5]$, funkcija `sample(x)` vrne slučajno permutacijo števil 1, 2, 3, 4, 5.

```
# najprej generiramo vektor x s celimi števili iz obsega [m, n]
x=m:n
x
# funkcija sample naredi slučajno razporeditev števil vektorja x (slučajna
permutacija)
sample(x)
```



The screenshot shows the R Commander interface. The script window contains the following R code:

```
# najprej generiramo vektor x s celimi števili iz obsega [1, 5]
x=1:5
x
# funkcija sample naredi slučajno razporeditev števil vektorja x (slučajna permutacija)
sample(x)
# še slučajna razporeditev števil iz obsega [1, 30]
x=1:30
sample(x)
sample(x)
```

The console window shows the following output:

```
> # najprej generiramo vektor x s celimi števili iz obsega [1, 5]
> x=1:5
> x
[1] 1 2 3 4 5
> # funkcija sample naredi slučajno razporeditev števil vektorja x (slučajna permutacija)
> sample(x)
[1] 3 1 2 4 5
> # še slučajna razporeditev števil iz obsega [1, 30]
> x=1:30
> sample(x)
[1] 1 20 14 4 5 13 9 15 22 10 27 26 7 17 12 8 18 29 3 30 6 19 2 24 21 28 11 16 23 25
> sample(x)
[1] 16 10 12 26 28 7 20 30 11 15 17 1 14 2 25 13 4 22 5 29 3 18 27 21 24 9 19 23 6 8
```

Če želimo iz populacije velikosti N izbrati vzorec velikosti n brez vračanja enot v populacijo in pri tem upoštevamo predpostavko, da ima vsaka enota v populaciji enako možnost izbire v vzorec, funkcijo uporabimo v obliki:

```
# okvir vzorčenja predstavljajo oštevilčene enote v populaciji (od 1 do N),
najprej generiramo vektor x s celimi števili na intervalu [1, N]
x = 1:N
# iz populacije slučajno vzorčimo n enot
n = n
sample(x, n)
```

The screenshot shows the R Commander window. The menu bar includes Datoteka, Uredi, Podatki, Statistike, Grafika, Modeli, Porazdelitve, Orodja, and Pomoč. Below the menu bar, there are buttons for 'Nabor podatkov: anketa', 'Urejanje podatkov', 'Pregledovanje podatkov', and 'Model: <Ni aktivnega modela>'. The main window is divided into two panes. The top pane, labeled 'Skriptno okno', contains the following R code:

```
# okvir vzorčenja predstavljajo oštevilčene enote v populaciji (od 1 do 100),
# najprej generiramo vektor x s celimi števili iz obsega [1, 100]
x = 1:100
# iz populacije slučajno vzorčimo 20 enot
n=20
sample(x, n)
```

The bottom pane, labeled 'Okno za izpis', shows the execution of the code and the resulting output:

```
> x = 1:100
> # iz populacije slučajno vzorčimo 20 enot
> n=20
> sample(x, n)
[1] 81 22 92 83 9 75 65 5 98 79 97 17 64 63 93 12 24 68 47 4
```

5.4 Slučajna izbira enot v načrtovanju poskusa

Narediti želimo načrt poskusa, v katerem bomo primerjali učinkovitost **treh različnih gojišč** (G1, G2, G3) za *in vitro* razmnoževanje rdečega radiča. Namen poskusa je ugotoviti, kako izbrana gojišča vplivajo na rast koreninskih poganjkov. **Poskusno enoto** predstavlja ena gojitvena posoda, vsako gojišče bo ponovljeno po desetkrat, torej imamo 30 gojitvenih posod. V vsako od gojitvenih posod damo po štiri izsečke rdečega radiča. Posamezen izseček predstavlja **enoto opazovanja**. Skupaj imamo torej $3 \times 10 \times 4 = 120$ izsečkov. Izid poskusa (to je število koreninskih poganjkov) štejemo na vsakem izsečku posebej.

Pogoji za izvedbo poskusa:

- variabilnost med listi znotraj iste rastline je zanemarljiva, variabilnost med rastlinami pa je pomembna;
- izseček takoj po tem, ko ga izrežemo iz rastline, položimo na gojišče v gojitveno posodo;
- zagotoviti moramo, da izsečke slučajno porazdelimo med obravnavanja G1, G2 in G3.

Postopek izvedbe poskusa

- pripravimo tri različna gojišča in damo vsakega v 10 gojitvenih posod;
- gojitvene posode postavimo v slučajni vrstni red;
- po štiri zaporedno izrezane izsečke nato po vrsti damo v gojitveno posodo;
- gojitvene posode damo v rastno komoro.

VZORČENJE 1

Predpostavite, da je vseh 120 izsečkov iz iste rastline. Določite okvir vzorčenja in izvedite slučajno razporeditev gojitvenih posod. Uporabite funkcijo `sample` v programu R.

Navodilo: Gojitvene posode oštevilčite od 1 do 30. ♦

VZORČENJE 2

Predpostavite, da lahko iz ene rastline radiča dobimo največ 12 izsečkov, kar pomeni, da za naš poskus potrebujemo deset rastlin. Različne rastline lahko v izid poskusa vnesejo dodatno variabilnost, zato moramo poskrbeti, da izvedemo vsa obravnavanja (tri gojišča) na izsečkih vsake od desetih rastlin. Rastlina v poskusu predstavlja **moteči dejavnik**.

Z gojišči napolnjene gojitvene posode tako najprej razvrstimo v deset skupin (blok), v katerih se vsako gojišče ponovi enkrat. V vsakem bloku s slučajnim izborom določimo vrstni red gojitvenih posod. Nato vanje po vrsti damo po štiri izsečke iste rastline.

Izvedite slučajno vzorčenje za tak primer.

Navodilo: Gojitvene posode za vsako gojišče oštevilčite od 1 do 10. ♦

VZORČENJE 3

Razmislite, kako bi gojitvene posode iz naloge VZORČENJE 2 razporedili v rastno komoro. ♦

Zgled 5.1: Slučajna izbira enot iz populacije

Datoteka ANKETA.XLS (nabor podatkov »anketa«) vsebuje podatke za 126 študentov 2. letnika AG, BT in MB v šol. l. 2006/07 na BF. Predpostavimo, da so na anketo odgovorili vsi študentje omenjene populacije.

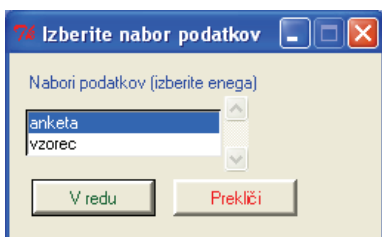
ANKETA 30

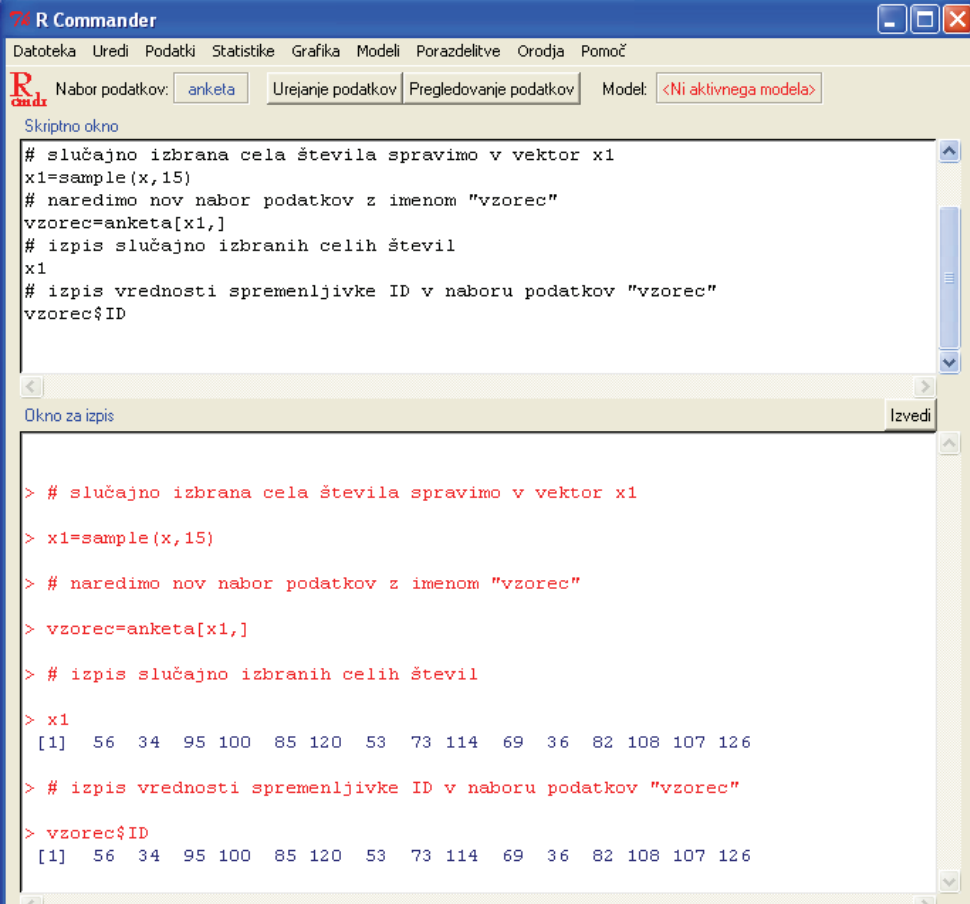
Z uporabo funkcije `sample` izvedite enostavno slučajno vzorčenje 15 študentov izmed 126. Izbrane zapise podatkovne zbirke shranite v nov nabor podatkov z imenom »vzorec«.

Navodilo: z namenom, da bomo naredili nov nabor podatkov z imenom »vzorec«, moramo rezultat funkcije `sample` shraniti v nov vektor npr. `x1`. Ukaze pišemo v Skriptno okno:

```
> x = 1:126
> x1 = sample(x, 15)
# v nov nabor podatkov »vzorec« prepišemo vrstice z zaporednimi števkami, določenimi s
števkami v vektorju x1 iz starega nabora »anketa«
> vzorec=anketa[x1,]
```

Po izvedbi zgornjega ukaza imamo v R Commanderju dva nabora podatkov. Aktivni nabor podatkov ostane »anketa«, to vidimo na gumbu Nabor podatkov. Če kliknemo na ta gumb, se pojavi pogovorno okno, v katerem lahko izbiramo med naborom »anketa« in »vzorec«:





The screenshot shows the R Commander interface. The top menu bar includes 'Datoteka', 'Uredi', 'Podatki', 'Statistike', 'Grafika', 'Modeli', 'Porazdelitve', 'Orodja', and 'Pomoč'. Below the menu, there are buttons for 'Nabor podatkov: anketa', 'Urejanje podatkov', 'Pregledovanje podatkov', and 'Model: <Ni aktivnega modela>'. The main window is titled 'Skriptno okno' and contains the following R code:

```
# slučajno izbrana cela števila spravimo v vektor x1
x1=sample(x,15)
# naredimo nov nabor podatkov z imenom "vzorec"
vzorec=anketa[x1,]
# izpis slučajno izbranih celih števil
x1
# izpis vrednosti spremenljivke ID v naboru podatkov "vzorec"
vzorec$ID
```

Below the code, there is a window titled 'Okno za izpis' with the following output:

```
> # slučajno izbrana cela števila spravimo v vektor x1
> x1=sample(x,15)
> # naredimo nov nabor podatkov z imenom "vzorec"
> vzorec=anketa[x1,]
> # izpis slučajno izbranih celih števil
> x1
[1] 56 34 95 100 85 120 53 73 114 69 36 82 108 107 126
> # izpis vrednosti spremenljivke ID v naboru podatkov "vzorec"
> vzorec$ID
[1] 56 34 95 100 85 120 53 73 114 69 36 82 108 107 126
```

ANKETA 31

Analizirajte spremenljivko **splet**: izračunajte dve primerni meri sredine in dve primerni meri variabilnosti, najprej za vseh 126 študentov, potem pa še za študente, izbrane v vzorec velikosti 15. Primerjajte rezultate. ♦

6 Normalna in binomska porazdelitev z Excelom

Novi pojmi:

- normalna porazdelitev, funkciji NORMDIST in NORMINV
- generator slučajnih števil v Excelu
- binomska porazdelitev, funkcija BINOMDIST

Datoteke s podatki:

PORAZDELITVE.XLS

V skupini statističnih funkcij so funkcije, ki za različne verjetnostne porazdelitve izračunavajo porazdelitveno funkcijo $F(x)$, gostoto verjetnosti oz. porazdelitveno shemo $p(x)$ ter vrednost slučajne spremenljivke x pri določeni verjetnosti. Vrednosti, ki jih izračunamo s temi funkcijami, potrebujemo pri statističnem sklepanju, nekatere so v statističnih tabelah.

6.1 Normalna porazdelitev, funkciji NORMDIST in NORMINV

Porazdelitveno funkcijo $F(x) = P(X \leq x)$ in gostoto verjetnosti $p(x)$ za normalno porazdeljeno slučajno spremenljivko X izračunamo s funkcijo NORMDIST(x; mean; standard_dev; cumulative). Prvi trije argumenti funkcije NORMDIST so: »x« je kvantil vrednost slučajne spremenljivke X , »mean« je pričakovana vrednost (povprečje) slučajne spremenljivke $E(X) = \mu$ in »standard_dev« je standardni odklon slučajne spremenljivke X , $\sigma = \sqrt{\text{Var}(X)}$. Za četrti argument »cumulative« vpišemo logično vrednost TRUE ali FALSE. Če izberemo logično vrednost FALSE, funkcija izračuna gostoto verjetnosti normalne porazdelitve $p(x)$:

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Če izberemo TRUE, funkcija NORMDIST izračuna porazdelitveno funkcijo $F(x)$, torej verjetnost, da je vrednost slučajne spremenljivke X manjša ali enaka vrednosti x :

$$F(x) = \int_{-\infty}^x p(t) dt = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt .$$

Normalno porazdelitev s parametroma $\mu=0$ in $\sigma=1$ imenujemo standardizirana normalna porazdelitev.

Excel poimenuje porazdelitveno funkcijo $F(x)$ kumulativna porazdelitev, gostoto verjetnosti $p(x)$ pa verjetnostna masna funkcija.

6.1.1 Gostota verjetnosti

NORMALNA 1

Odprite datoteko PORAZDELITVE.XLS. ♦

NORMALNA 2

Na delovni list »Standardizirana normalna« pripravimo podatke, na osnovi katerih bomo narisali gostoto verjetnosti standardizirane normalne porazdelitve (krivulja v obliki zvonca). V en stolpec zapišemo zalogo vrednosti slučajne spremenljivke X , ki je porazdeljena po standardizirani normalni porazdelitvi s parametroma $\mu=0$ in $\sigma=1$, in v sosednji stolpec za vsako vrednost slučajne spremenljivke X izračunamo gostoto verjetnosti normalne porazdelitve z uporabo funkcije NORMDIST. Upoštevali bomo dejstvo, da je $P(\mu - 4\sigma \leq X \leq \mu + 4\sigma) \approx 1$.

Navodilo: v prvi stolpec zapišimo vrednosti za slučajno spremenljivko X na intervalu od -4,0 do 4,0, za korak med dvema zaporednima vrednostima izberemo 0,1: napišemo -4,0 v prvo celico, v spodnjo pa -3,9, označimo obe celici, križec miške pa postavimo v desni spodnji vogal označene skupine celic. Ko križec spremeni obliko, s potegom miške navzdol dobimo zelene vrednosti.

Z uporabo funkcije NORMDIST v sosednji stolpec izračunajmo gostoto verjetnosti $p(x)$ standardizirane normalne porazdelitve $N(0, 1)$ za vrednosti x med -4,0 in 4,0. Za argumente funkcije uporabimo: NORMDIST(referenca celice, v kateri je podatek za x ; 0; 1; FALSE).

Normalna porazdelitev je zvezna porazdelitev, tu jo rišemo »po točkah«, ki jih povežemo z daljicami. Da dobimo gladko krivuljo, morajo biti točke dovolj goste. ♦

	A	B	C	D	E	F	G	H	I	J
1	x	p(x)								
2	-4	(2;0;1;false)								
3	-3,9									
4	-3,8									
5	-3,7									
6	-3,6									
7	-3,5									
8	-3,4									
9	-3,3									
10	-3,2									
11	-3,1									
12	-3									
13	-2,9									
14	-2,8									
15	-2,7									
16	-2,6									
17	-2,5									
18	-2,4									
19	-2,3									
20	-2,2									
21	-2,1									
22	-2									

Argumenti funkcije

NORMDIST

x | A2 = -4

Mean | 0 = 0

Standard_dev | 1 = 1

Cumulative | false = FALSE

= 0,00013383

Vrne normalno kumulativno porazdelitev za navedeno srednjo vrednost in standardni odklon.

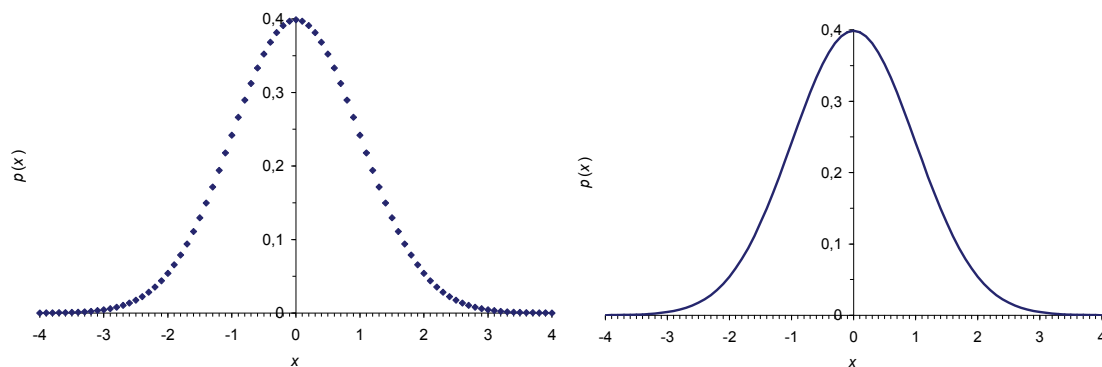
Cumulative je logična vrednost: za kumulativno porazdelitveno funkcijo uporabite TRUE; za verjetnostno masno funkcijo pa uporabite FALSE.

Rezultat formule = 0,00013383

[Pomoč za to funkcijo](#) V redu Prekliči

NORMALNA 3

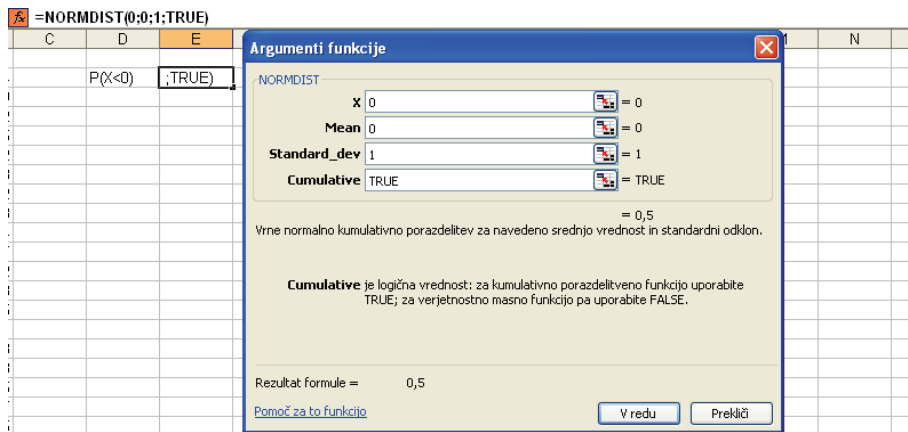
Grafično prikazite izračunane vrednosti gostote verjetnosti $N(0,1)$ - vrsta grafikona XY (*Raztreseni*) (slika 6.1 levo). Točke povežite z daljicami in odstranite predstavitelje točk, da dobite sliko 6.1 (desno) ♦



Slika 6.1: Grafični prikaz izračunanih vrednosti gostote verjetnosti $p(x)$ standardizirane normalne porazdelitve $N(0, 1)$ za izbrane vrednosti x (levo). To je vmesni korak do končnega grafičnega prikaza gostote verjetnosti standardizirane normalne porazdelitve (desno).

6.1.2 Izračun verjetnosti

Za izračun verjetnosti $P(X < x)$ v funkciji NORMDIST kot zadnji argument uporabimo vrednost *TRUE*. Na primer za $P(X < 0)$ pri standardizirani normalni porazdelitvi je postopek takšen:



NORMALNA 4

Za standardizirano normalno porazdelitev z uporabo funkcije NORMDIST(x ; 0; 1; *TRUE*) izračunajte verjetnosti:

$$\begin{array}{cccc}
 P(X \leq 0) & P(X \geq 1) & P(-1 \leq X \leq 1) & P(-2 \leq X \leq 2) \\
 P(X > 2) & P(X > -1,5) & P(X \leq 2,1) & P(-1,5 \leq X \leq 2,1)
 \end{array}$$

Navodilo: v datoteki PORAZDELITVE.XLS tabelo 6.1 izpolnite z ustreznimi vrednostmi. Končni rezultati so v celicah vijolične barve. ♦

Tabela 6.1: Izpis formul v tabeli za izračun verjetnosti standardizirane normalne porazdelitve (zgoraj) in rezultati (spodaj)

	A	B	C	D	E	F
1	Verjetnosti	x_1	$P(X < x)$	$P(X > x)$	x_2	$P(x_1 < X < x_2)$
2	$P(X < 0)$	0	=NORMDIST(B2;0;1;TRUE)			
3	$P(X > 1)$	1	=NORMDIST(B3;0;1;TRUE)	=1-C3		
4	$P(-1 < X < 1)$	-1	=NORMDIST(B4;0;1;TRUE)		1	=C3-C4
5	$P(-2 < X < 2)$	-2	=NORMDIST(B5;0;1;TRUE)		2	=1-2*C5
6	$P(X > 2)$	2	=NORMDIST(B6;0;1;TRUE)	=1-C6		
7	$P(X > -1,5)$	-1,5	=NORMDIST(B7;0;1;TRUE)	=1-C7		
8	$P(X < 2,1)$	2,1	=NORMDIST(B8;0;1;TRUE)			
9	$P(-1,5 < X < 2,1)$	-1,5	=NORMDIST(B9;0;1;TRUE)		2,1	=C8-C9

	A	B	C	D	E	F
1	Verjetnosti	x_1	$P(X < x)$	$P(X > x)$	x_2	$P(x_1 < X < x_2)$
2	$P(X < 0)$	0	0,5000			
3	$P(X > 1)$	1	0,8413	0,1587		
4	$P(-1 < X < 1)$	-1	0,1587		1	0,6827
5	$P(-2 < X < 2)$	-2	0,0228		2	0,9545
6	$P(X > 2)$	2	0,9772	0,0228		
7	$P(X > -1,5)$	-1,5	0,0668	0,9332		
8	$P(X < 2,1)$	2,1	0,9821			
9	$P(-1,5 < X < 2,1)$	-1,5	0,0668		2,1	0,9153

NORMALNA 5

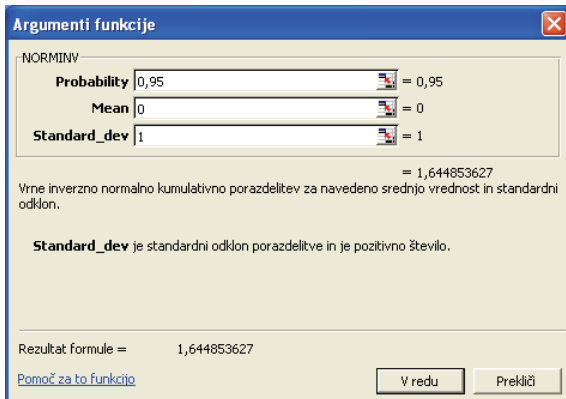
Izračune v tabeli 6.1 preverite z uporabo tabele za standardizirano normalno porazdelitev. ♦

6.1.3 Izračun kvantilov

Vrednost x pri določeni verjetnosti $P(X \leq x)$ imenujemo kvantil. Na primer, če je $P(X \leq x) = 1/2$, je x kvantil, ki ga imenujemo mediana oz. 50. centil. Za normalno porazdelitev kvantile izračunamo s funkcijo NORMINV(probability; mean; standard_dev); »probability« je verjetnost $P(X \leq x)$, druga dva argumenta sta enaka kot pri funkciji NORMDIST.

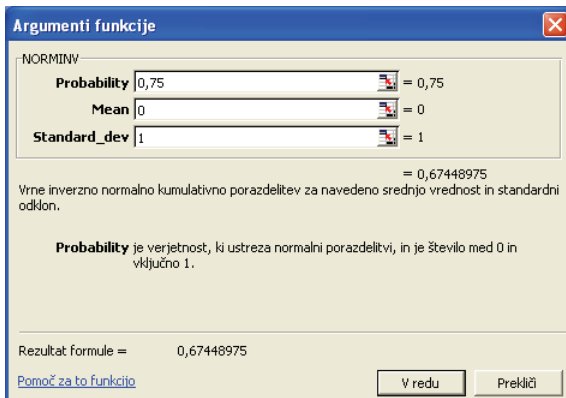
NORMALNA 6

Za standardizirano normalno porazdelitev izračunajte kvantil x_1 , ki ustreza enačbi $P(X < x_1) = 0,95$. Izračun naredite v pripravljeno tabelo na delovnem listu »Standardizirana normalna«. Kateri centil ste izračunali? ♦



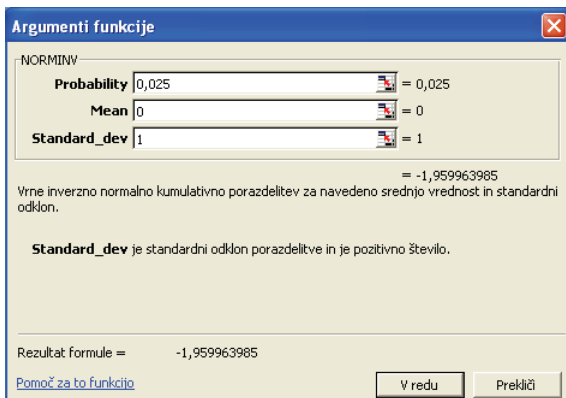
NORMALNA 7

Za standardizirano normalno porazdelitev izračunajte kvantil x_2 , ki ustreza enačbi $P(X > x_2) = 0,25$. Kateri centil ste izračunali? ♦



NORMALNA 8

Za standardizirano normalno porazdelitev izračunajte vrednost x_3 , ki ustreza enačbi $P(-x_3 \leq X \leq +x_3) = 0,95$ in vrednost x_4 , ki ustreza enačbi $P(-x_4 \leq X \leq +x_4) = 0,99$. ♦



Zgled 6.1.1: Starost ob prvi zaposlitvi

STAROST 1

Predpostavimo, da v nekem območju velja, da je starost ob prvi zaposlitvi X porazdeljena normalno s parametroma $\mu = 21$ let, $\sigma = 2$ leti. Narišite gostoto verjetnosti te normalne porazdelitve (slika 6.2, brez senčitev). Katero območje za starost boste pri risanju grafa upoštevali? Kolikšen korak med dvema starostma je potreben, da bo krivulja gladka?

Delovni list s podatki za sliko poimenujte »Starost ob prvi zaposlitvi«. ♦

STAROST 2

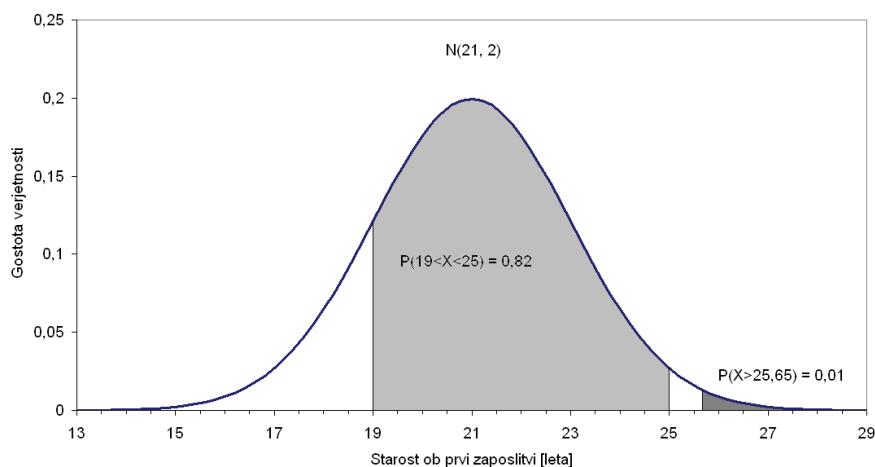
Izračunajte verjetnost, da je starost ob prvi zaposlitvi med 19 in 25 let; $P(19 \leq X \leq 25)$ (slika 6.2). ♦

STAROST 3

Izračunajte starost ob prvi zaposlitvi, za katero velja, da je prekoračena pri 1 % oseb (slika 6.2). Kateri centil ste izračunali? ♦

STAROST 4

Izračunajte starost ob prvi zaposlitvi, za katero velja, da je prekoračena pri 90 % oseb. Kateri centil ste izračunali? ♦

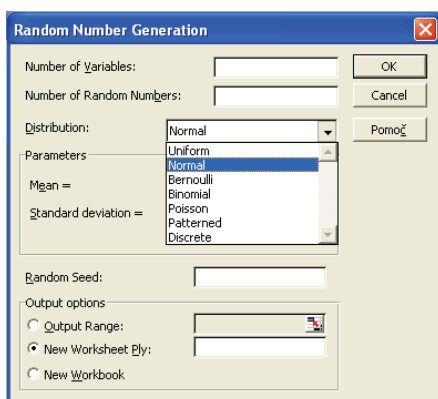


Slika 6.2: Normalna porazdelitev s parametroma $\mu = 21$ let, $\sigma = 2$ leti za starost ob prvi zaposlitvi. Svetleje osončena je prikazana verjetnost, da je oseba ob prvi zaposlitvi stara med 19 in 25 let, temneje osončena pa je prikazana verjetnost (njena vrednost je 0,01), da je oseba ob prvi zaposlitvi stara 25,65 let ali več.

6.2 Generator slučajnih števil v Excelu

V Excelu generator slučajnih števil omogoča generiranje vrednosti slučajnih spremenljivk, porazdeljenih po različnih porazdelitvah. Generator slučajnih števil zaženemo z ukazom **Orodja/Analiza podatkov/Random number generation**. V njegovem pogovornem oknu v polje *Number of variables* napišemo število slučajnih spremenljivk, ki jih želimo generirati; v polje *Number of Random Numbers* pa napišemo število vrednosti za vsako od slučajnih spremenljivk. V polju *Distribution* izberemo obliko porazdelitve slučajnih števil: *Uniform, Normal, Discrete,...*

V polje *Random Seed* napišemo celo pozitivno število med 1 in 32767, to je t. i. "brst", ki sproži generator slučajnih števil. Če za *Random Seed* uporabimo vedno isto številko, dobimo pri izbrani porazdelitvi vedno ista slučajna števila.²⁰ To polje lahko pustimo prazno in tedaj vedno dobimo druga slučajna števila.



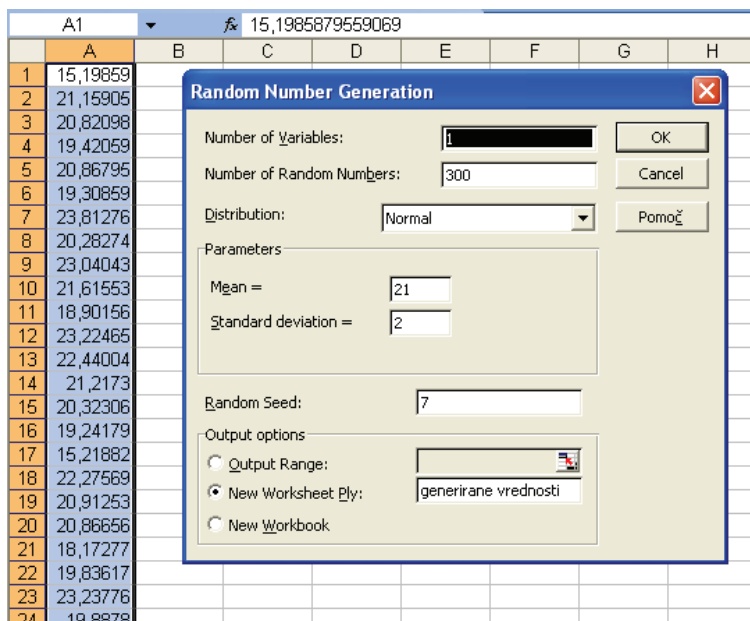
6.2.1 Generiranje vrednosti, porazdeljenih po normalni porazdelitvi

Za generiranje vrednosti spremenljivk, porazdeljenih po normalni porazdelitvi v polju *Distribution*, izberemo možnost *Normal*.

NORMALNA 9

Generirajte 300 vrednosti slučajne spremenljivke, porazdeljene po $N(21, 2)$. Za *Random Seed* izberite številko dneva vašega rojstnega dne. Vrednosti naj se izpišejo na nov delovni list z imenom »Generirane vrednosti«. ♦

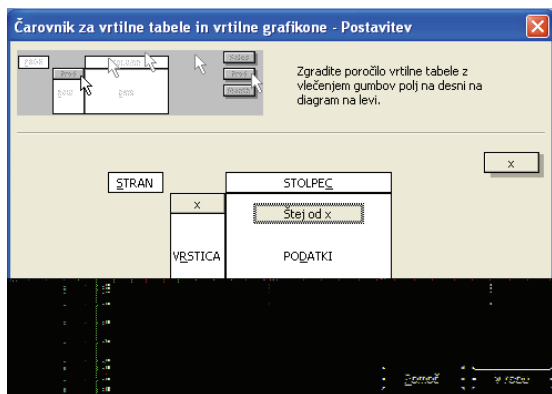
²⁰ Taka števila imenujemo pseudoslučajna števila.



NORMALNA 10

Generirane vrednosti razporedite v razrede širine 2 s pomočjo vrtilne tabele in narišite histogram. Kakšna je oblika empirične porazdelitve?

Navodilo: vrinite eno vrstico na začetek generiranih vrednosti in v celico A1 napišite ime vrednosti slučajne spremenljivke (npr. x). Uporabite čarovnika za oblikovanje vrtilnih tabel:



Vrednosti spremenljivke x združimo v razrede z ukazom **Združi in pokaži podrobnosti**:

	A	B	C	D	E	F
1	Tule spustite polja strani					
2						
3	Štej od x					
4	x	Vsota				
5	15,12248144	1				
6	15,19858796	1				
7	15,21881512	1				
8	15,72580384	1				
9	16,16774551	1				
10	16,51797373	1				
11	16,53478119	1				
12	16,63115125	1				
13	16,80228177	1				
14	16,85870682	1				
15	17,05344783	1				
16	17,29942977	1				
17	17,34474967	1				

	A	B	C
1	Tule spustite polja strani		
2			
3	Štej od x		
4	x	Vsota	
5	15-17	10	
6	17-19	39	
7	19-21	95	
8	21-23	98	
9	23-25	48	
10	25-27	8	
11	27-29	2	
12	Skupna vsota	300	
13			

Združevanje v skupine

Samodejno

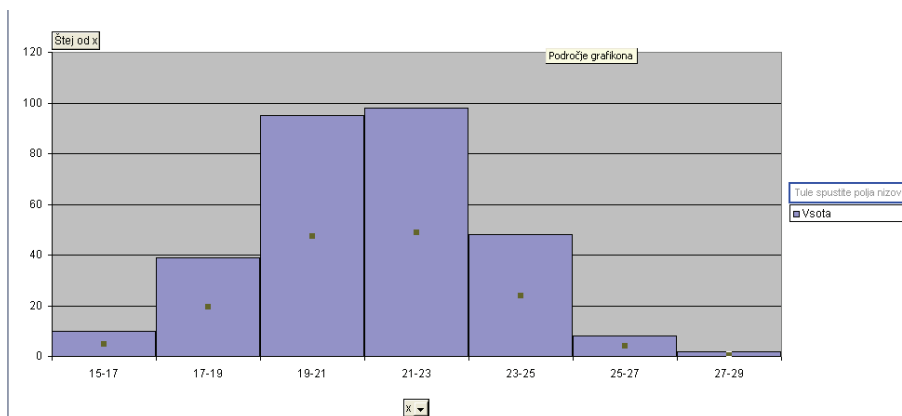
Začni pri: 15

Končaj pri: 29

Bo: 2

V redu Prekliči

Naredite še vrtilni grafikon, na katerem poskrbite, da se stolpci držijo skupaj:



Sliko ustrezno dodelajte: odstranite gumbe vrtilnega grafikona, popravite naslove osi, odstranite legendo, itd. ♦

6.3 Binomska porazdelitev, funkcija BINOMDIST

Binomska porazdelitev $b(n, p)$ je diskretna porazdelitev z dvema parametroma n in p , n je število Bernoullijevih poskusov (v vsakem poskusu se lahko zgodi dogodek A ali pa njemu nasprotni dogodek \bar{A}) in p je verjetnost, da se v posameznem poskusu zgodi dogodek A . Verjetnost, da se v zaporedju n -tih poskusov dogodek A zgodi x -krat, izračunamo po Bernoullijevi formuli:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

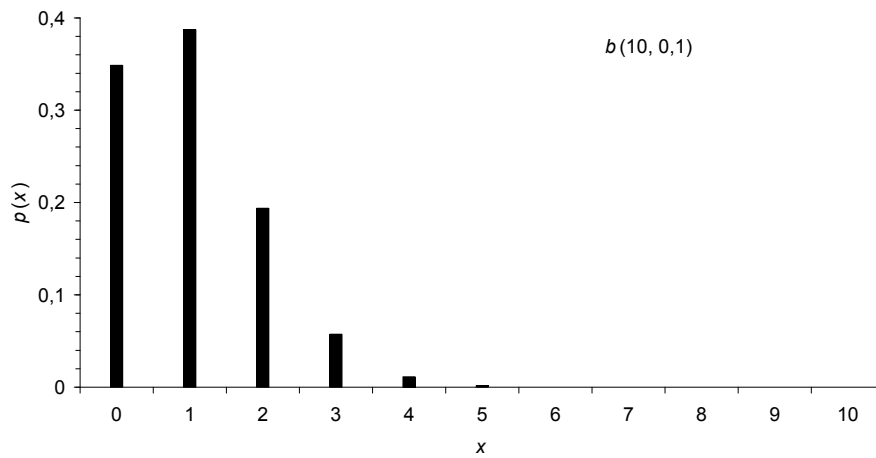
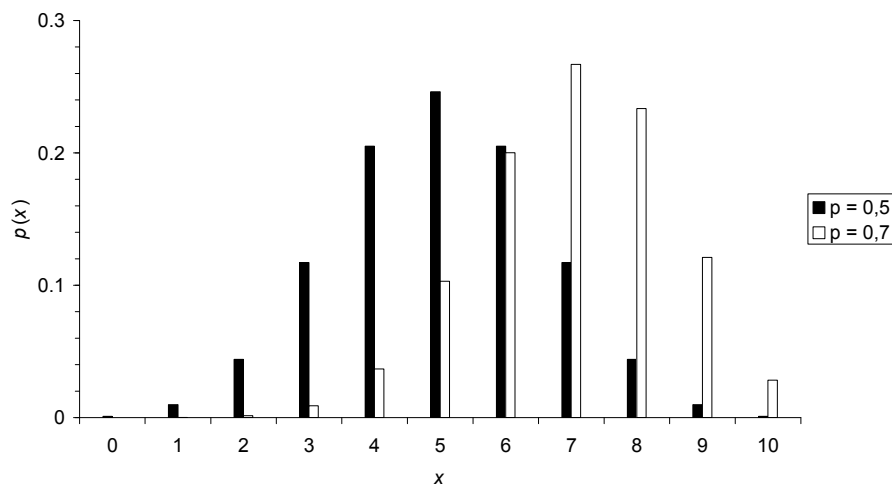
V Excelu $p(x)$ izračunamo z uporabo funkcije BINOMDIST(number_s; trials; probability_s; cumulative). Funkcija ima štiri argumente: »number_s« je število dogodkov A , ki smo ga v formuli označili z x ; »trials« je število poskusov n ; »probability_s« je verjetnost za dogodek A , ki je v formuli označena s p . Za argument »cumulative« vpišemo logično vrednost *TRUE* ali *FALSE*. Če izberemo logično vrednost *FALSE*, funkcija izračuna verjetnost $p(x)$. Če izberemo *TRUE*, funkcija BINOMDIST izračuna kumulativno verjetnost $P(X \leq x) = \sum_{i=0}^x p(x)$. Excel poimenuje izračunano verjetnost $p(x)$ binomska porazdelitvena verjetnost.

BINOMDIST		=BINOMDIST(A3;10;0,1;FALSE)	
A	B	C	D
1	b (n = 10, p)		
2	x	p(x), p = 0,1	
3	0	1, FALSE)	
4	1	0,3874205	
5	2	0,1937102	
6	3	0,0573956	
7	4	0,0111603	
8	5	0,001488	
9	6	0,0001378	
10	7	8,748E-06	
11	8	3,645E-07	
12	9	9E-09	
13	10	1E-10	
14	Vsota p(x)	1	
15			
16			
17			

Argumenti funkcije	
Number_s	A3 = 0
Trials	10 = 10
Probability_s	0,1 = 0,1
Cumulative	FALSE = FALSE
Vrne posamezno binomsko porazdelitveno verjetnost.	
= 0,34867844	
Number_s je število uspešnih preizkusov.	
Rezultat formule = 0,34867844	
Pomoč za to funkcijo	
V redu Prekliči	

BINOMSKA 1

Z uporabo funkcije BINOMDIST izračunajte verjetnosti $p(x)$ za binomsko porazdelitev s parametroma $n=10$ in $p=0,1$. Izračune naredite v pripravljeno tabelo na delovnem listu »Binomska«. Verjetnostno porazdelitev prikažite grafično, kot kaže slika 6.3. ♦

Slika 6.3: Binomska porazdelitev za parametra $n = 10$ in $p = 0,1$ Slika 6.4: Binomski porazdelitvi za parametra $n = 10$ in $p = 0,5$ ter $n = 10$ in $p = 0,7$

BINOMSKA 2

Z uporabo funkcije BINOMDIST izračunajte verjetnosti $p(x)$ še za binomske porazdelitve s parametrom $n=10$ in tremi različnimi vrednostmi za p : $p=0,3$, $p=0,5$ in $p=0,7$. Verjetnostni porazdelitvi $b(10, 0,5)$ in $b(10, 0,7)$ prikažite grafično, kot kaže slika 6.4. ♦

BINOMSKA 3

Za binomsko porazdelitev $b(10, 0,3)$ izračunajte verjetnost, da se bo dogodek A zgodil največ trikrat. ♦

BINOMSKA 4

Za binomsko porazdelitev $b(10, 0,3)$ izračunajte verjetnost, da se bo dogodek A zgodil več kot trikrat. ♦

BINOMSKA 5

Za binomsko porazdelitev $b(10, 0,3)$ izračunajte verjetnost, da se bo dogodek A zgodil štirikrat ali petkrat. ♦

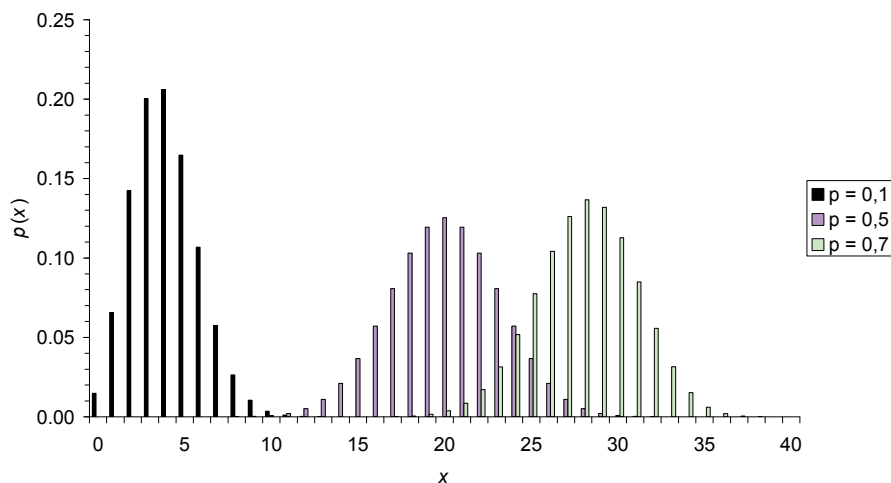
6.4 Aproximacija binomske porazdelitve z normalno porazdelitvijo

Pogoja za aproksimacijo binomske porazdelitve $b(n, p)$ z normalno porazdelitvijo $N(np, \sqrt{np(1-p)})$ sta²¹:

$$n \cdot p \geq 5 \quad \text{in hkrati} \quad n \cdot (1-p) \geq 5$$

BINOMSKA 6

Z uporabo funkcije BINOMDIST izračunajte verjetnosti $p(x)$ še za binomske porazdelitve s parametrom $n = 40$ in tremi različnimi vrednostmi za p : $p = 0,1$, $p = 0,5$ in $p = 0,7$. Verjetnostne porazdelitve prikažite grafično, kot kaže slika 6.5. ♦



Slika 6.5: Binomske porazdelitve za parametra $n = 40$ in $p = 0,1$; $n = 40$ in $p = 0,5$; $n = 40$ in $p = 0,7$

²¹ V literaturi za aproksimacijo najdemo tudi drugačne pogoje

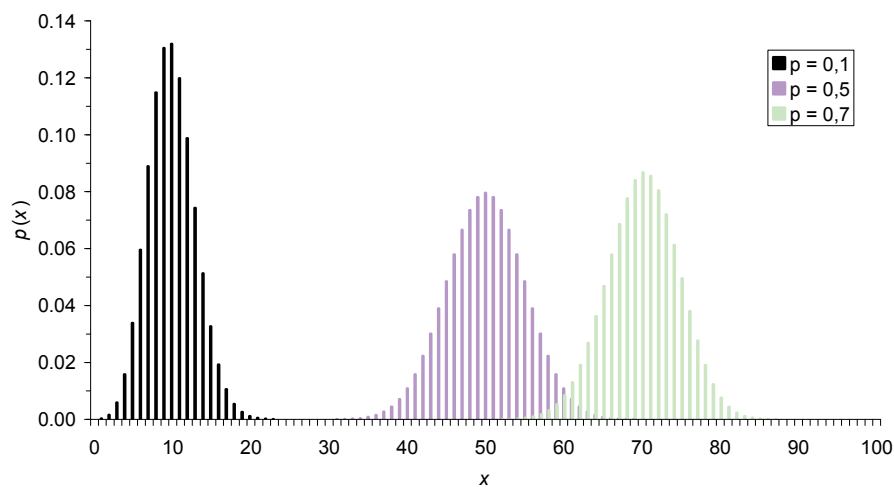
BINOMSKA 7

Ali je v teh primerih aproksimacija binomske porazdelitve z normalno upravičena? Utemeljite odgovor. ♦

BINOMSKA 8

Z uporabo funkcije BINOMDIST izračunajte verjetnosti $p(x)$ še za binomske porazdelitve s parametrom $n=100$ in tremi različnimi vrednostmi za p : $p=0,1$, $p=0,5$ in $p=0,7$.

Verjetnostne porazdelitve prikažite grafično, kot kaže slika 6.6. ♦



Slika 6.6: Binomske porazdelitve za parametra $n = 100$ in $p = 0,1$; $n = 100$ in $p = 0,5$; $n = 100$ in $p = 0,7$

BINOMSKA 9

Ali je v teh primerih aproksimacija binomske porazdelitve z normalno upravičena? Utemeljite odgovor. ♦

BINOMSKA 10

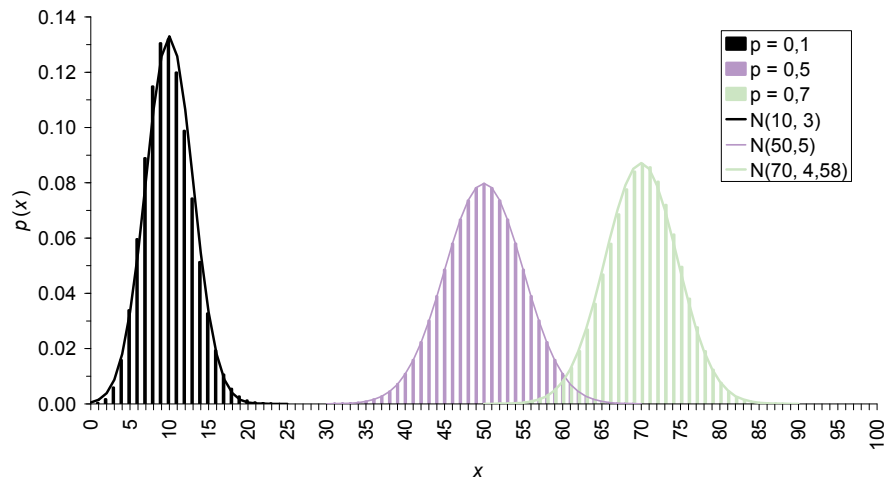
Na sliko 6.6 dodajte še nize za gostote verjetnosti pripadajočih normalnih porazdelitev (slika 6.7).

Navodilo:

$$N(100 \cdot 0,1, \sqrt{100 \cdot 0,1 \cdot 0,9}) = N(10, 3),$$

$$N(100 \cdot 0,5, \sqrt{100 \cdot 0,5 \cdot 0,5}) = N(50, 5),$$

$$N(100 \cdot 0,7, \sqrt{100 \cdot 0,7 \cdot 0,3}) = N(70, 4,58). \quad \blacklozenge$$



Slika 6.7: Binomske porazdelitve in njihove aproksimacije z normalno porazdelitvijo

7 Verjetnostne porazdelitve s programom R

Novi pojmi:

- porazdelitve v R-u
- generiranje slučajnih števil, porazdeljenih po izbrani porazdelitvi (normalna, Studentova, hi-kvadrat, binomska)

Datoteke s podatki:

v tem poglavju jih ni

Ustvarjene datoteke:

PORAZDELITVE_R.DOC

V predhodnih poglavjih smo se ukvarjali zgolj z računalniško analizo podatkov in uporabno statistiko. Matematične teorije, ki stoji v ozadju, nismo predstavljali. S tem področjem se ukvarja matematična statistika. V tem poglavju pa bomo z uporabo programa R na zgledih ilustrirali nekatere izreke iz matematične statistike. Zaradi lažje povezave med teorijo in prakso te izreke navajamo, sicer brez dokazov. Ukvarjali se bomo s porazdelitvami vzorčnih statistik: povprečja, reskalirane variance in deleža na populaciji vseh možnih vzorcev velikosti n .

7.1 Porazdelitev vzorčnih aritmetičnih sredin

Izrek

Če je slučajna spremenljivka X na **osnovni populaciji** porazdeljena po normalni porazdelitvi s parametroma μ in σ - $N(\mu, \sigma)$, potem je slučajna spremenljivka \bar{X} na **populaciji vseh vzorcev velikosti n** porazdeljena po normalni porazdelitvi s parametroma μ in σ/\sqrt{n} - $N(\mu, \sigma/\sqrt{n})$.

V Wordu odprite datoteko z imenom PORAZDELITVE_R.DOC. Vanjo napišite odgovore na vprašanja, kopirajte rezultate izračunov in oblikovane grafikone v nalogah tega poglavja. ♦

Zgled 7.1: Porazdelitev zgornjega krvnega tlaka

Raziskovalci so ugotovili, da je zgornji krvni tlak (spremenljivka X) pri ženskah, starih med 25 in 35 let, porazdeljen po normalni porazdelitvi $N(125 \text{ mm}, 15 \text{ mm})$.

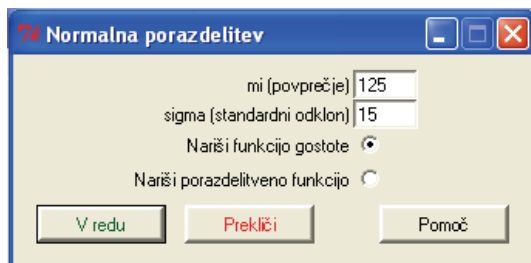
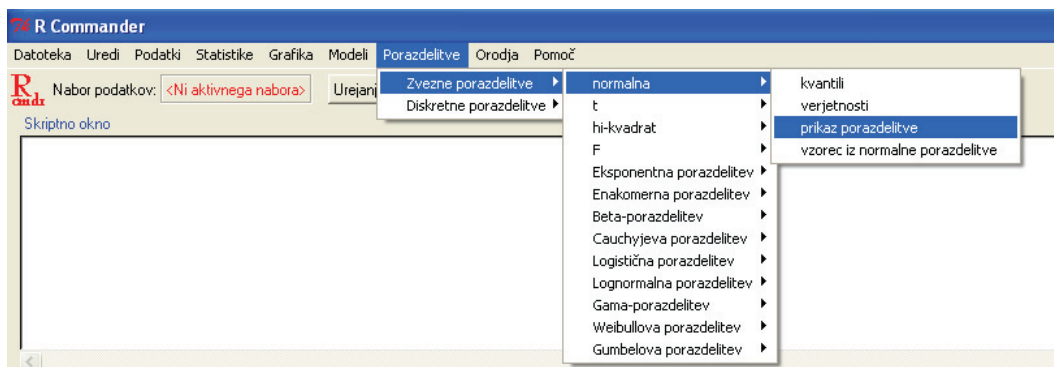
KRVNI TLAK 1

Opredelite proučevano populacijo. Kaj je enota populacije? Katera je proučevana spremenljivka? Določite ji mersko lestvico. Ali je populacija dejanska ali hipotetična? ♦

Za reševanje naloge bomo tokrat uporabili program R z grafičnim vmesnikom R-Commander.

KRVNI TLAK 2

Narišite gostoto verjetnosti normalne porazdelitve $N(125 \text{ mm}, 15 \text{ mm})$ za zgornji krvni tlak pri ženskah, starih med 25 in 35 let. Sliko kopirajte v wordovo datoteko z imenom PORAZDELITVE_R.DOC z oznako Slika 7.1 (levo). Sliki napišite podnaslov.



Navodilo: s klikanjem naredimo sliko gostote verjetnosti, pri tem se v skriptno okno izpišejo štiri ukazne vrstice. V skriptnem oknu lahko zamenjamo angleško besedilo v naslovih grafikona s slovenskimi izrazi:

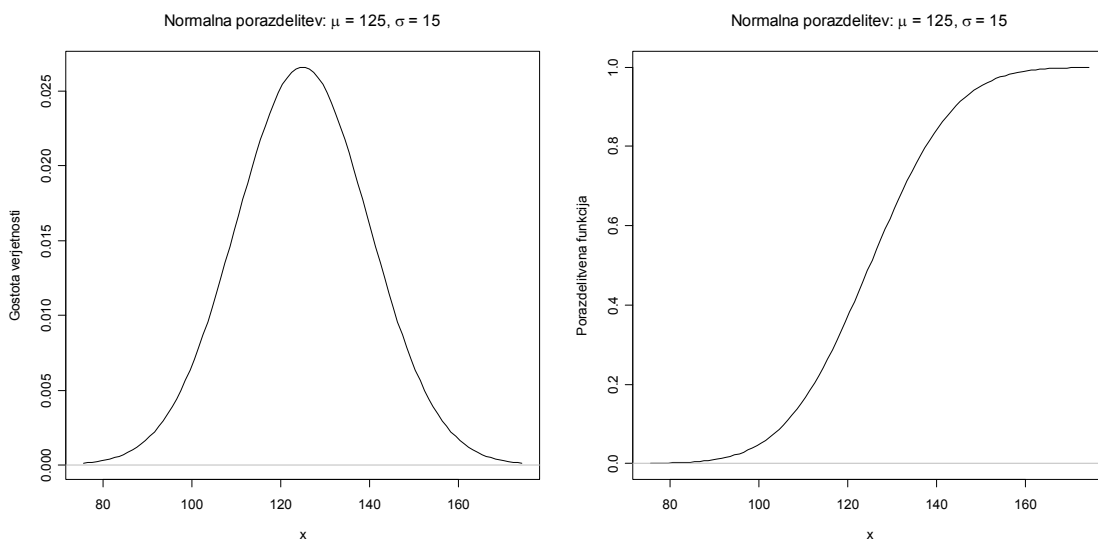
```
.x <- seq(75.642, 174.358, length=100)
# "Density" zamenjamo z "Gostota verjetnosti"; "Normal Distribution" pa z "Normalna
porazdelitev"
plot(.x, dnorm(.x, mean=125, sd=15), xlab="x", ylab="Gostota verjetnosti",
main=expression(paste("Normalna porazdelitev: ", mu, " = 125, ", sigma, " = 15")), type="l")
abline(h=0, col="gray")
remove(.x)
```

Opomba: preden kliknemo gumb **Izvedi**, je potrebno označiti vse štiri ukazne vrstice. ♦

KRVNI TLAK 3

Za dano porazdelitev narišite še porazdelitveno funkcijo. Naslove na grafikonu poslovenite.

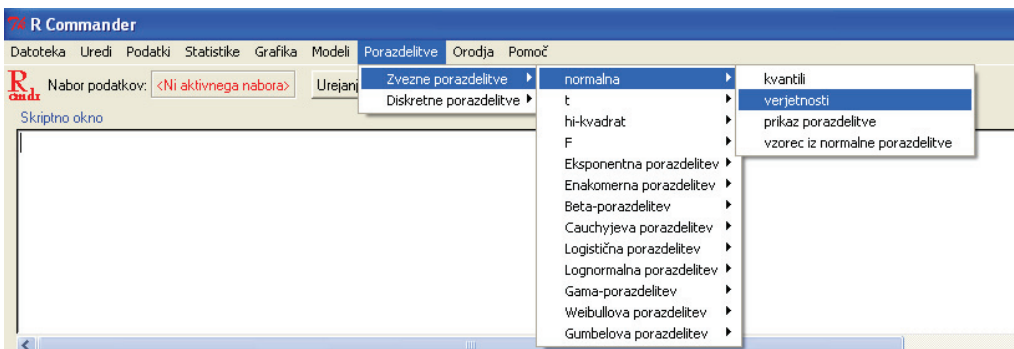
Sliko kopirajte v wordovo datoteko z oznako Slika 7.1 (desno). ♦

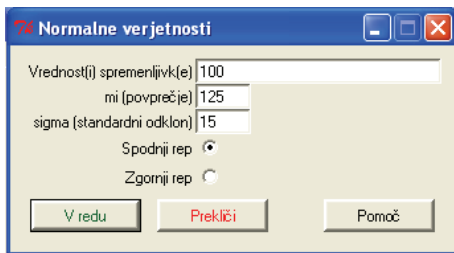


Slika 7.1: Gostota verjetnosti (levo) in porazdelitvena funkcija (desno) za normalno porazdelitev $N(125,15)$

KRVNI TLAK 4

Izračunajte, kolikšen odstotek žensk v tej starostni skupini ima zgornji krvni tlak pod 100 mm.



**Rešitev:**

```
> pnorm(c(100), mean=125, sd=15, lower.tail=TRUE)
```

```
[1] 0.04779035 ◆
```

KRVNI TLAK 5

Katero verjetnost dobimo, če v pogovornem oknu *Verjetnosti normalne porazdelitve* izberemo namesto možnosti *Spodnji rep*, možnost *Zgornji rep*? Kolikšna je ta verjetnost? ◆

KRVNI TLAK 6

Izračunajte, kolikšen odstotek žensk v tej starostni skupini ima krvni tlak med 120 mm in 140 mm.

Navodilo: v pogovornem oknu *Verjetnosti normalne porazdelitve* lahko v polje *Vrednost(i) spremenljivk(e)* napišemo hkrati tudi več vrednosti, med številkami pa pustimo presledek (ne uporabljamo ločil). Rezultat funkcije `pnorm` lahko shranimo v spremenljivko z izbranim imenom, na primer:

```
a = pnorm(c(120,140), mean=125, sd=15, lower.tail=TRUE).
```

Verjetnost $P(120 < X < 140)$ nato izračunamo kot razliko:

```
a[2]-a[1] ◆
```

Skriptno okno

```
a=pnorm(c(120,140), mean=125, sd=15, lower.tail=TRUE)
a[2]-a[1]
```

Dkno za izpis

```
> pnorm(c(120,140), mean=125, sd=15, lower.tail=TRUE)
[1] 0.3694413 0.8413447

> a=pnorm(c(120,140), mean=125, sd=15, lower.tail=TRUE)

> a
[1] 0.3694413 0.8413447

> a[2]-a[1]
[1] 0.4719034
```

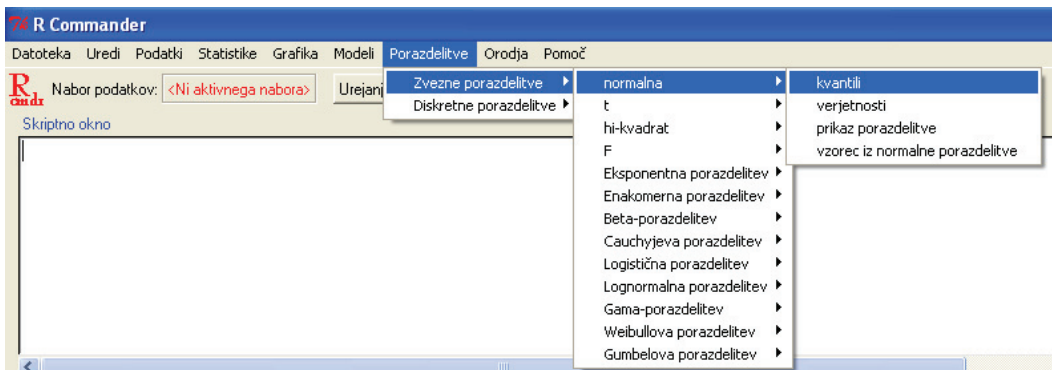
74 Verjetnosti normalne porazdelitve

Vrednost(i) spremenljivk(e) 120 140
 mi (povprečje) 125
 sigma (standardni odklon) 15
 Spodnji rep
 Zgornji rep

V redu Prekliči Pomoč

KRVNI TLAK 7

Za katero višino zgornjega krvnega tlaka velja, da je prekoračena pri 10 % žensk v izbrani starostni skupini? Kako imenujemo kvantil, ki ste ga izračunali?



Rešitev:

```
> qnorm(c(0.1), mean=125, sd=15, lower.tail=FALSE)
```

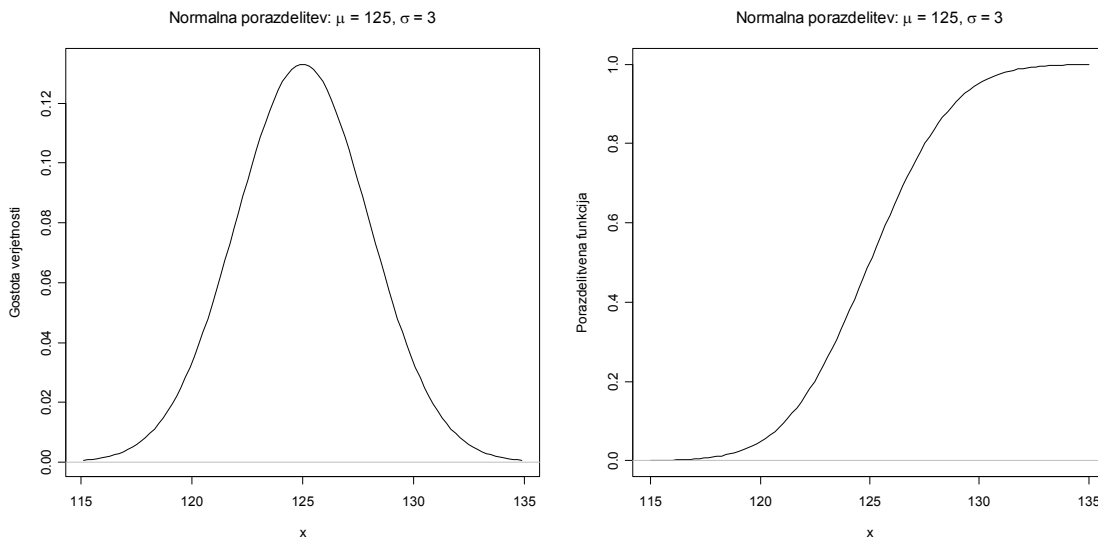
```
[1] 144.2233 ♦
```

KRVNI TLAK 8

Iz omenjene populacije žensk vzorčimo vzorce velikosti 25 žensk. Kaj je enota populacije vseh možnih vzorcev velikosti 25? ♦

KRVNI TLAK 9

Določite porazdelitev vzorčnih aritmetičnih sredin v vzorcih velikosti 25. Narišite njeno gostoto verjetnosti in porazdelitveno funkcijo. Oznaka slike naj bo Slika 7.2. Sliki napišite podnaslov. ♦



Slika 7.2: Gostota verjetnosti (levo) in porazdelitvena funkcija (desno) za porazdelitev aritmetičnih sredin vzorcev velikosti $n = 25$

KRVNI TLAK 10

Izračunajte, za koliko od 100 vzorcev velikosti 25 lahko pričakujemo, da bo vzorčno povprečje nad 124. ♦

KRVNI TLAK 11

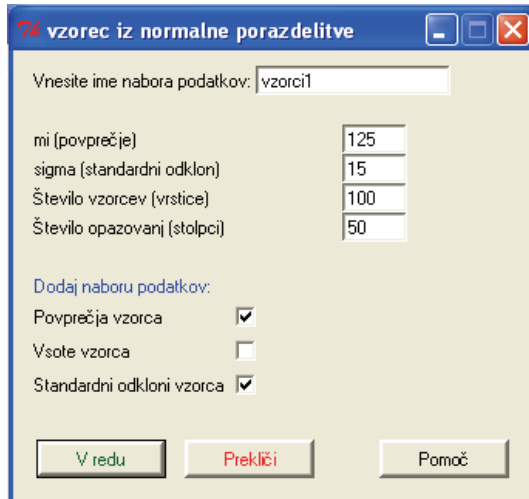
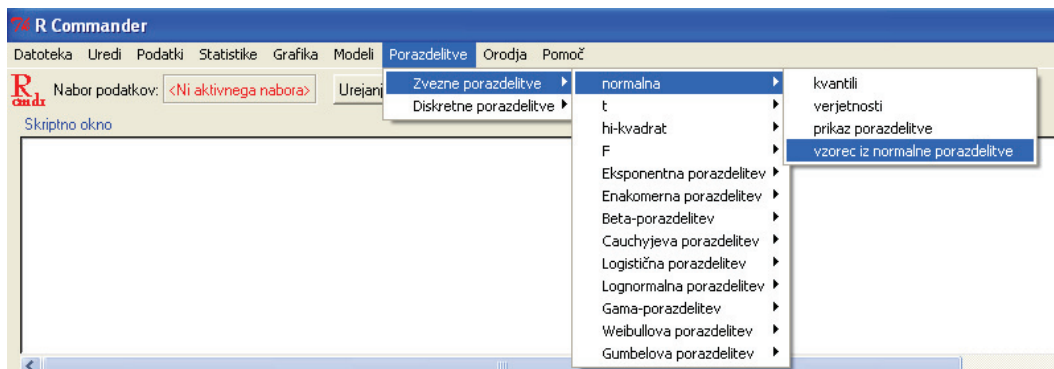
Izračunajte še, za koliko od 100 vzorcev velikosti 25 lahko pričakujemo, da bo vzorčno povprečje med 124 in 126. ♦

KRVNI TLAK 12

Za vzorčne aritmetične sredine vzorcev velikosti $n = 25$ določite vrednosti \bar{x}_1 in \bar{x}_2 , za kateri velja, da je $P(\bar{X} \leq \bar{x}_1) = 0,025$ in $P(\bar{X} \geq \bar{x}_2) = 0,025$. ♦

Zgled 7.2: Generiranje vrednosti slučajne spremenljivke X , ki je na osnovni populaciji porazdeljena po $N(125, 15)$, porazdelitev vzorčnih povprečij

Generirati želimo 100 vzorcev velikosti $n = 50$ z vrednostmi slučajne spremenljivke X , ki je porazdeljena po $N(125, 15)$. Uporabili bomo generator slučajnih števil v programu R. Generirane vrednosti se shranijo v nov nabor podatkov, ki ga bomo poimenovali »vzorci1«:



V pogovornem oknu *Vzorec iz normalne porazdelitve* v prvem polju določimo ime nabora podatkov, definiramo parametra normalne porazdelitve *mi (povprečje)* in *sigma (standardni odklon)*, določimo *Število vzorcev (vrstice)* in njihovo velikost *Število opazovanj (stolpci)*.

NORMALNA 11

S programom R generirajte 100 vzorcev velikosti $n = 50$ z vrednostmi slučajne spremenljivke X , ki je porazdeljena po $N(125, 15)$. ♦

Za vsak vzorec R izračuna povprečje in vzorčni standardni odklon, če izberemo možnosti *Povprečja vzorca* in *Standardni odkloni vzorca*. V naboru podatkov »vzorci1« so generirane vrednosti slučajne spremenljivke X urejene tako, da ena vrstica predstavlja en vzorec. Vsaka vrednost se v preglednici ob uporabi gumba **Pregledovanje podatkov** pokaže na pet decimalnih mest natančno. V zadnjih dveh stolpcih sta za vsak vzorec posebej izračunana povprečje (*mean*) in standardni odklon (*sd*). S klikom na gumb **Pregledovanje podatkov** in s pomikom drsnika skrajno desno dobimo sledečo obliko prikaza podatkov v naboru »vzorci1« (vse vrednosti pa so odvisne od slučaja):

	44	obs45	obs46	obs47	obs48	obs49	obs50	mean	sd
sample1	72	121.03831	168.40079	135.18590	117.08180	120.92672	141.04580	127.5262	16.15180
sample2	47	122.55165	138.12547	123.16567	120.15218	116.36341	123.89522	121.3750	14.82342
sample3	70	121.98188	93.05556	130.17366	153.97046	122.92869	127.86049	125.7054	15.86802
sample4	64	110.10039	129.81329	131.08674	114.99548	158.21945	142.60499	125.9894	13.53030
sample5	49	100.25677	112.62678	109.04344	122.27840	130.82071	145.70141	126.7697	15.40268
sample6	68	153.53060	133.18427	123.04152	111.96590	151.16361	109.94346	125.4466	15.64397
sample7	30	124.37625	138.28140	130.16085	128.37087	131.45637	121.37184	125.5083	13.26749
sample8	65	132.57638	98.81807	101.51517	117.49848	112.16035	129.16075	127.1260	13.14323
sample9	47	109.95702	134.56148	122.24372	100.26716	124.36407	151.74842	129.1131	16.25490
sample10	93	130.63621	110.28800	129.91327	100.31206	112.71840	123.70489	126.5974	10.71886
sample11	03	138.84270	120.02605	128.14218	98.01613	121.71990	120.53128	127.0008	15.23096
sample12	37	104.69059	104.57315	134.47767	125.05787	138.85965	143.91957	122.0471	13.42908
sample13	72	115.68823	110.03544	117.85504	136.14518	130.67233	126.89795	127.7229	13.15589
sample14	34	115.47941	117.70415	125.28324	138.37636	125.07540	108.53968	123.5248	14.48346
sample15	25	146.43609	114.93578	143.46310	137.37922	119.92065	102.98772	125.1900	18.23242
sample16	90	128.58588	122.37727	133.13063	127.04436	120.98150	132.17611	127.8523	13.11319
sample17	08	128.70806	117.81825	132.70800	135.88047	132.37347	100.41341	126.7147	16.90352
sample18	36	109.55711	130.42675	167.39849	135.86212	102.54079	134.41541	125.0774	16.78443
sample19	83	123.96007	140.02794	135.90405	141.72259	153.42018	145.31906	123.4297	13.98565
sample20	37	106.85839	110.62465	119.03545	114.21782	115.34339	131.12031	124.5095	12.63476
sample21	21	129.19711	139.03003	121.50020	110.14948	117.89712	143.16401	121.9143	14.40104
sample22	93	127.28799	126.02331	148.24414	131.07471	130.91732	150.27568	128.3287	14.58527
sample23	64	108.27567	141.30793	112.25790	138.20272	121.85231	104.73333	123.9049	12.63650
sample24	08	125.72254	130.52936	124.13268	142.83800	99.83130	136.78509	124.6430	13.67190
sample25	04	125.73427	127.54771	130.64426	136.06456	133.13191	127.27365	126.4031	14.24464
sample26	44	110.14351	132.30535	128.91258	129.82141	125.94462	120.99663	125.9316	15.32447
sample27	34	112.06869	120.10747	137.76121	117.49625	128.37701	140.20887	124.8277	15.18806
sample28	08	130.59048	112.46796	121.86733	148.74411	132.29156	142.44414	124.8313	14.38532
sample29	13	123.53603	152.93403	132.00355	130.14706	132.26951	124.30498	126.6818	15.75618
sample30	65	148.09838	148.33262	124.40154	143.29935	145.81855	124.93914	130.0258	15.16359

Poglejmo, kakšno obliko ima porazdelitev tako dobljenih vzorčnih aritmetičnih sredin.

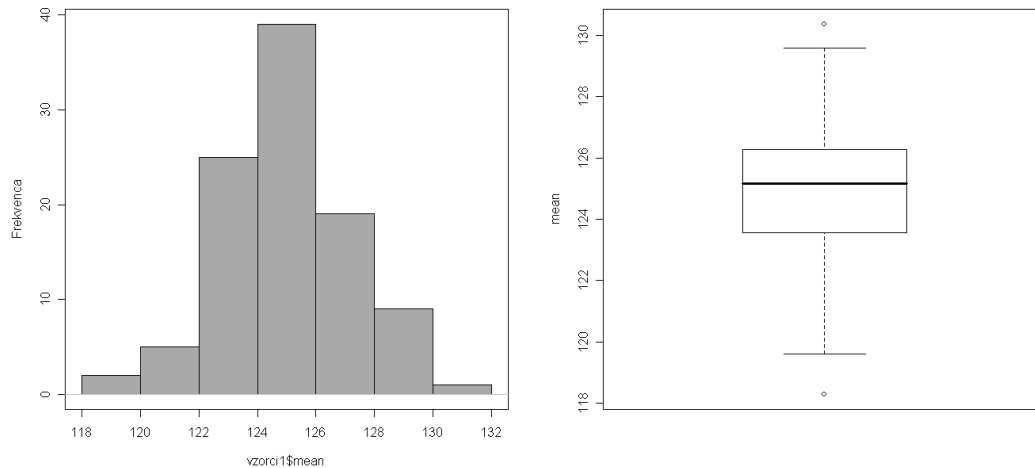
NORMALNA 12

S programom R narišite histogram in okvir z ročaji za spremenljivko *mean* v naboru podatkov »vzorci1« (slika 7.3).

Navodilo: da se naslov y-osi (»Frekvenca«) izpiše v slovenskem jeziku, moramo ukaz v skriptnem oknu popraviti:

```
#najprej moramo dodati v ukaz Hist argument ylab="", s katerim povemo, da se samodejni
naslov osi y ne izpiše:
Hist(vzorci1$mean, scale="frequency", ylab="", breaks="Sturges", col="darkgray")
# naslov osi y moramo nato na sliko dodati z ukazom title:
title(ylab="Frekvenca")
```

◆



Slika 7.3: Empirična porazdelitev vzorčnih aritmetičnih sredin vzorcev velikosti $n = 50$, porazdelitev slučajne spremenljivke na osnovni populaciji je normalna $N(125,15)$ (opomba: vsako generiranje podatkov generira druga slučajna števila, primerjajte svoj histogram s tem).

NORMALNA 13

Izračunajte povprečje in standardni odklon za spremenljivko *mean* (vzorčne aritmetične sredine). Ali so izračuni v skladu s pričakovanji o obliki porazdelitve vzorčnih aritmetičnih sredin? Kateri dve vrednosti pričakujemo za povprečje in standardni odklon vzorčnih povprečij (mean)? ◆

7.2 Porazdelitev vzorčnih varianc

Izrek

Če je slučajna spremenljivka X porazdeljena po $N(\mu, \sigma)$, potem je na populaciji vzorcev velikosti n slučajna spremenljivka $((n-1)/\sigma^2) \cdot S^2$, ki jo priredimo vzorčnim variancam (s^2), porazdeljena po χ^2 -porazdelitvi z $n-1$ stopinjami prostosti:

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(SP = n-1).$$

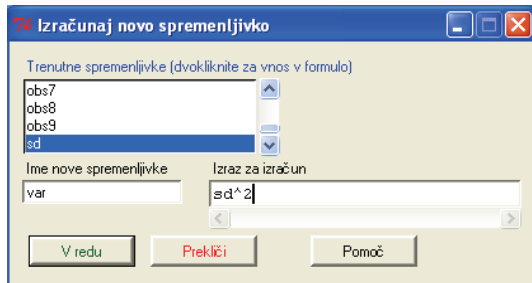
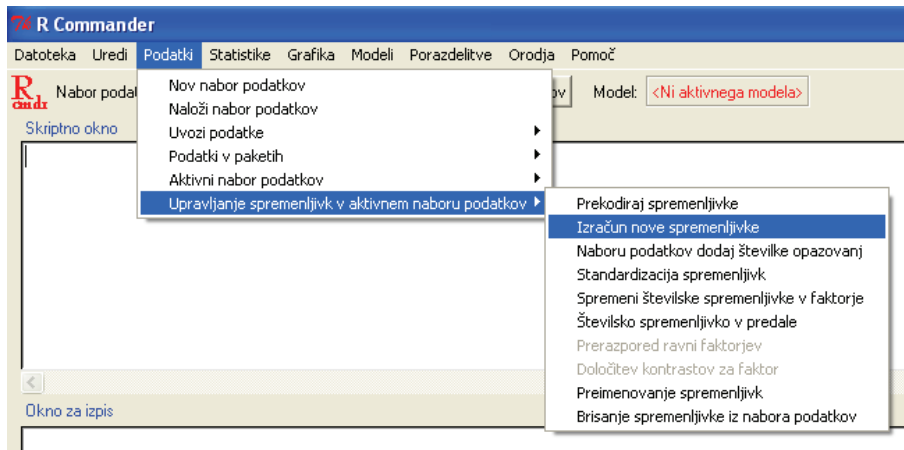
Opomba: količino $((n-1)/\sigma^2) \cdot S^2$ bomo imenovali reskalirana varianca.

Zgled 7.2 (nadaljevanje): Porazdelitev reskaliranih vzorčnih varianc

PORAZDELITEV VARIANC 1

Na osnovi vzorčnih standardnih odklonov (sd) v podatkovnem naboru »vzorci1« izračunajmo še vzorčne variance vseh 100 vzorcev (nova spremenljivka var).

Navodilo: uporabimo ukaz **Podatki/Upravljanje spremenljivk v aktivnem naboru podatkov/Izračun nove spremenljivke:**

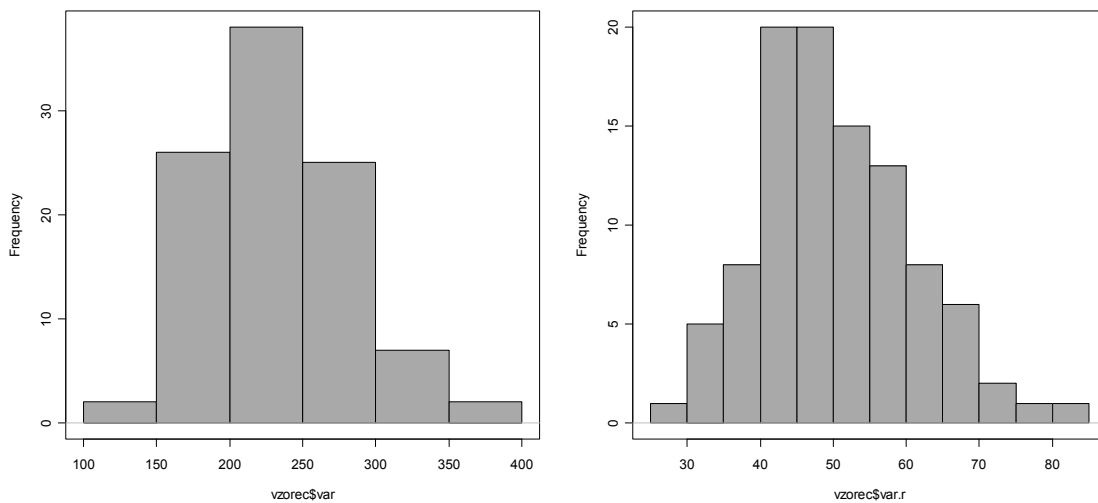


PORAZDELITEV VARIANC 2

Narišite empirično porazdelitev vzorčnih varianc (slika 7.4 levo). ♦

PORAZDELITEV VARIANC 3

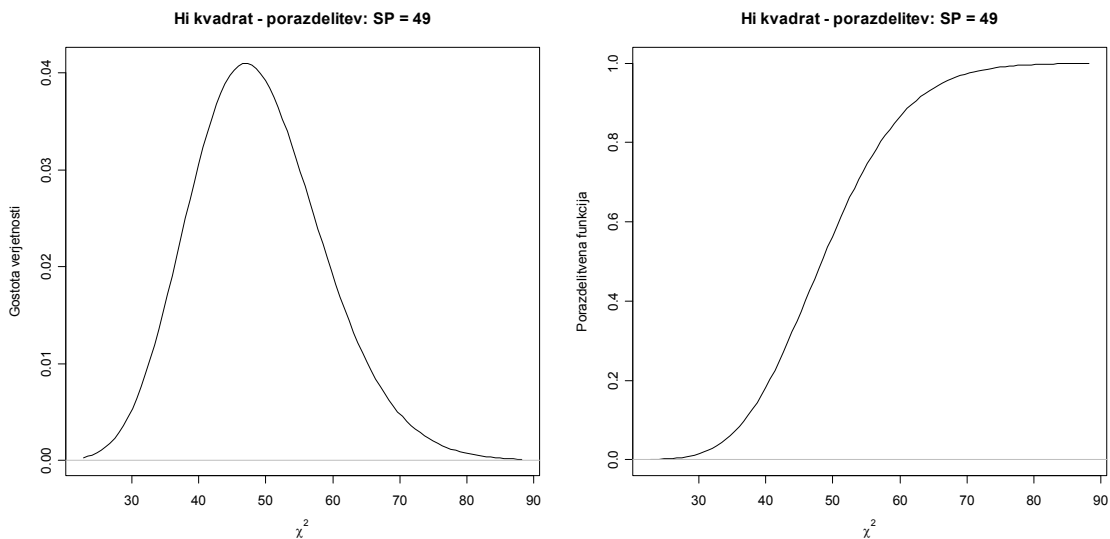
Za vse vzorce izračunajte še vrednosti reskalirane vzorčne variance $\left(\frac{n-1}{\sigma^2}\right) \cdot s^2$ in grafično prikažite njeno empirično porazdelitev s histogramom (slika 7.4 desno). ♦



Slika 7.4: Empirična porazdelitev vzorčnih varianc s^2 (levo) in reskaliranih vzorčnih varianc $\left(\frac{n-1}{\sigma^2}\right) \cdot s^2$ (desno) za 100 vzorcev velikosti $n = 50$, porazdelitev slučajne spremenljivke na osnovni populaciji je normalna $N(125,15)$

PORAZDELITEV VARIANC 4

Katera je verjetnostna porazdelitev statistik $\left(\frac{n-1}{\sigma^2}\right) \cdot s^2$ na populaciji vseh vzorcev velikosti 50? Narišite gostoto verjetnosti in porazdelitveno funkcijo te porazdelitve (slika 7.5). ♦



Slika 7.5: Gostota verjetnosti (levo) in porazdelitvena funkcija (desno) za χ^2 - porazdelitev s stopinjami prostosti $SP = 49$

Zgled 7.3: Ilustracija centralnega limitnega izreka za porazdelitev vzorčnih povprečij**Centralni limitni izrek**

\bar{X} se pri velikih vzorcih porazdeljuje približno normalno tudi tedaj, ko verjetnostna porazdelitev slučajne spremenljivke X na osnovni populaciji ni normalna. Velja:

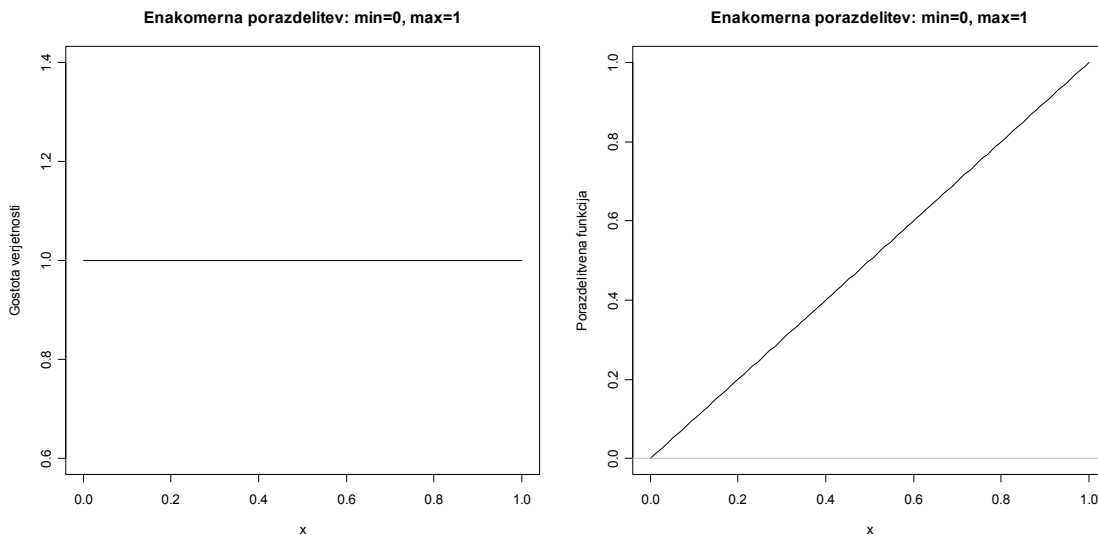
$$\bar{X} \approx N\left(E(X), \sqrt{\frac{\text{Var}(X)}{n}}\right), \quad \text{če je } n > 30$$

ENAKOMERNA PORAZDELITEV 1

Narišite gostoto verjetnosti enakomerne zvezne porazdelitve na intervalu od 0 do 1 (slika 7.6 levo). ♦

ENAKOMERNA PORAZDELITEV 2

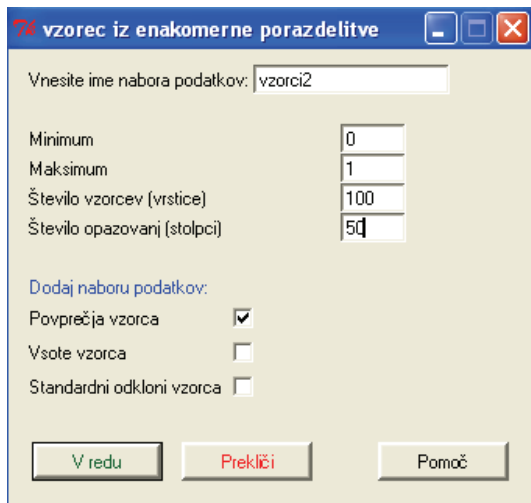
Narišite še porazdelitveno funkcijo za to porazdelitev (slika 7.6 desno). ♦



Slika 7.6: Gostota verjetnosti (levo) in porazdelitvena funkcija (desno) za enakomerno zvezno porazdelitev na intervalu od 0 do 1.

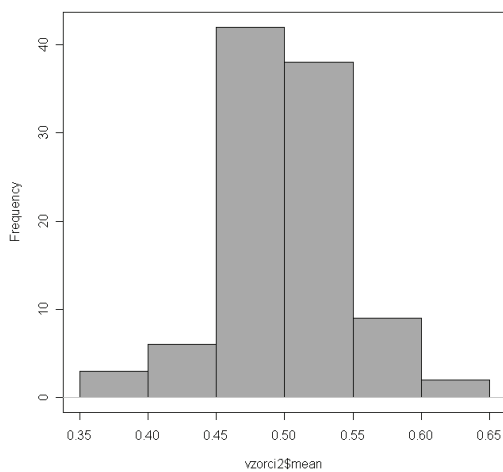
ENAKOMERNA PORAZDELITEV 3

V nabor podatkov »vzorci2« generirajte 100 vzorcev velikosti $n = 50$ z vrednostmi slučajne spremenljivke, porazdeljene po enakomerni zvezni porazdelitvi na intervalu $[0, 1]$. ♦



ENAKOMERNA PORAZDELITEV 4

Narišite empirično porazdelitev vzorčnih aritmetičnih sredin generiranih vzorcev (slika 7.7). Kakšno obliko porazdelitve pričakujemo? Izračunajte še povprečje vzorčnih aritmetičnih sredin. Katero vrednost pričakujete? ♦



Slika 7.7: Empirična porazdelitev vzorčnih povprečij 100 vzorcev velikosti $n = 50$, porazdelitev slučajne spremenljivke na osnovni populaciji je enakomerna zvezna na intervalu $[0, 1]$.

7.3 Studentova porazdelitev (t-porazdelitev)

William Gosset (s psevdonomom Student) je uvedel t-statistiko:
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Izrek

Slučajna spremenljivka X je porazdeljena $N(\mu, \sigma)$. Potem je na populaciji vzorcev velikosti n slučajna spremenljivka

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \text{ porazdeljena po Studentovi porazdelitvi s } SP = n - 1.$$

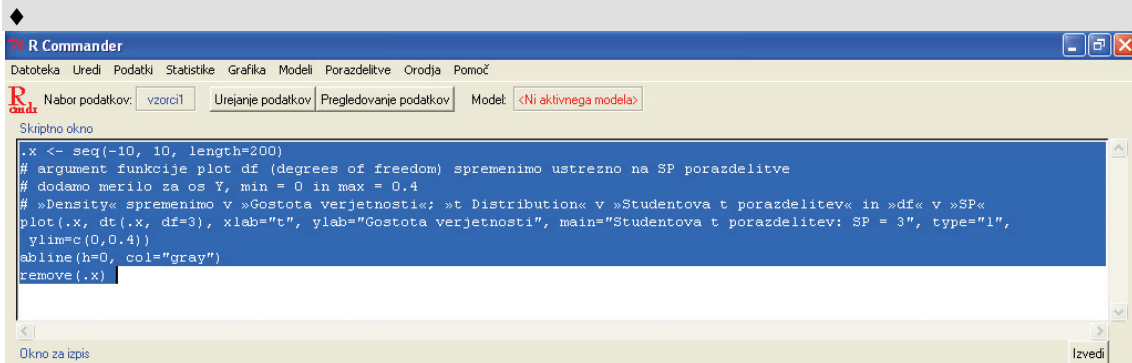
STUDENTOVA PORAZDELITEV 1

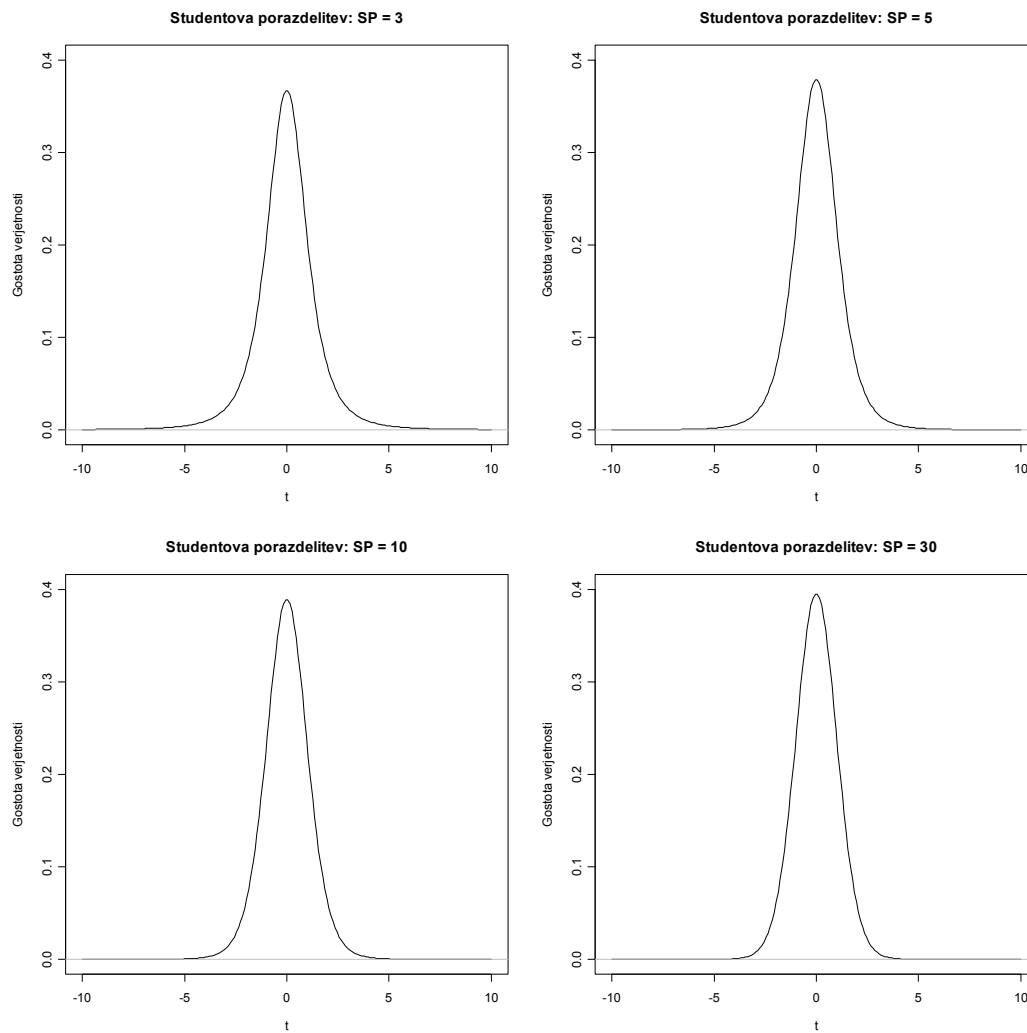
Narišite gostoto verjetnosti Studentove porazdelitve za $SP=3$, $SP=5$, $SP=10$ in $SP=30$ (slika 7.8). Vse štiri porazdelitve naj imajo isto merilo na oseh X in Y .

Navodilo: Če želimo v R-u dobiti štiri slike z istim merilom na oseh X in Y , v R Commanderju naredimo prvo sliko ($SP=3$). Pri tem se v skriptno okno izpišejo štiri ukazne vrstice. Nekaj vrednosti v prvih dveh ukaznih vrsticah (rdeče besedilo) bomo spremenili, nekaj dodali, angleško besedilo v naslovih grafikona bomo spremenili v slovensko:

```
# generiramo 200 vrednosti spremenljivke X iz obsega [-10, 10]
.x <- seq(-10, 10, length=200)
# argument funkcije plot df (degrees of freedom) spremenimo ustrezno na SP porazdelitve
# dodamo merilo za os Y, min = 0 in max = 0.4
# »Density« spremenimo v »Gostota verjetnosti«; »t Distribution« v »Studentova t porazdelitev«
in »df« v »SP«
plot(.x, dt(.x, df=3), xlab="t", ylab="Gostota verjetnosti", main="Studentova t porazdelitev: SP =
3", type="l", ylim=c(0,0.4))
abline(h=0, col="gray")
remove(.x)
```

Ko vsebino v skriptnem oknu spremenimo, označimo vse štiri vrstice in kliknemo gumb **Izvedi**.





Slika 7.8: Gostota verjetnosti za štiri različne Studentove porazdelitve

STUDENTOVA PORAZDELITEV 2*

Na isto sliko narišite gostoto verjetnosti Studentove porazdelitve s $SP=3$ in $SP=10$ ter gostoto verjetnosti standardizirane normalne porazdelitve. Iz teorije vemo, da je Studentova porazdelitev s $SP = neskončno$ enaka standardizirani normalni porazdelitvi (slika 7.9).

Navodilo: Za prikaz take slike moramo napisati ukaze v Skriptno okno R Commanderja:

```
# generiramo vrednosti spremenljivke x na intervalu [-10,10] s korakom 0,1
```

```
x <- seq(-10.0, 10.0, by=0.1)
```

```
# izračun gostote verjetnosti za vrednosti x za tri različne porazdelitve
```

```
y1=dt(x=x,df=3)
```

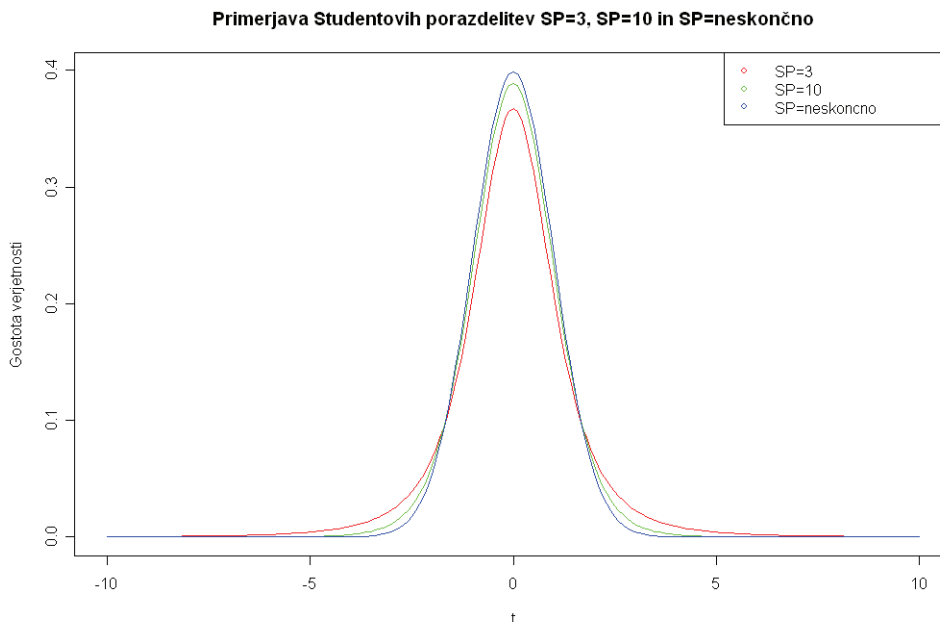
```
y2=dt(x=x,df=10)
```

```

y3=dnorm(x=x,mean=0,sd=1)
# število krivulj, ki jih rišemo, potrebujemo ga na koncu za določitev barvne palete
n.krivulj=3
# določitev minimalne in maksimalne vrednosti za x-os na grafu
xos=c(min(x),max(x))
# določitev minimalne in maksimalne vrednosti za y-os na grafu
yos=c(min(y1,y2,y3),max(y1,y2,y3))
# risanje y1
plot(data.frame(x=x,y=y1),xlim=xos,ylim=yos,xlab="",ylab="",type="l",pch=1,col=rainbow(3)[1])
# dodamo y2
lines(data.frame(x=x,y=y2),xlim=xos,ylim=yos,xlab="",ylab="",type="l",pch=2,col=rainbow(3)[2])
# dodamo y3
lines(data.frame(x=x,y=y3),xlim=xos,ylim=yos,xlab="",ylab="",type="l",pch=3,col=rainbow(3)[3])
# dodamo naslov in legendo
title(main="Primerjava Studentovih porazdelitev SP=3, SP=10 in SP=neskončno", xlab="t",
ylab="Gostota verjetnosti")
legend(x="topright",legend=c("SP=3","SP=10","SP=neskončno"),col=rainbow(n.krivulj),pch=1)

```

Skupino ukazov za grafični prikaz več Studentovih porazdelitev na istem grafikonu v skriptnem oknu označite in jo kopirajte v wordovo datoteko PORAZDELITVE_R.DOC. Če želite ukaze drugič ponoviti, jih lahko kopirate iz wordove datoteke. ♦



Slika 7.9: Gostota verjetnosti za tri različne Studentove porazdelitve.

STUDENTOVA PORAZDELITEV 3

Za Studentove porazdelitve s parametrom $SP_1 = 3$, $SP_2 = 5$, $SP_3 = 10$, $SP_4 = 30$ določite vrednosti t_1 , t_2 , t_3 in t_4 , za katere velja $P(-t_i \leq T \leq t_i) = 0,95$, za $i = 1, \dots, 4$. Dobljene vrednosti primerjajte z vrednostjo z , za katero velja $P(-z \leq Z \leq z) = 0,95$. Spremenljivka Z je porazdeljena po standardizirani normalni porazdelitvi. ♦

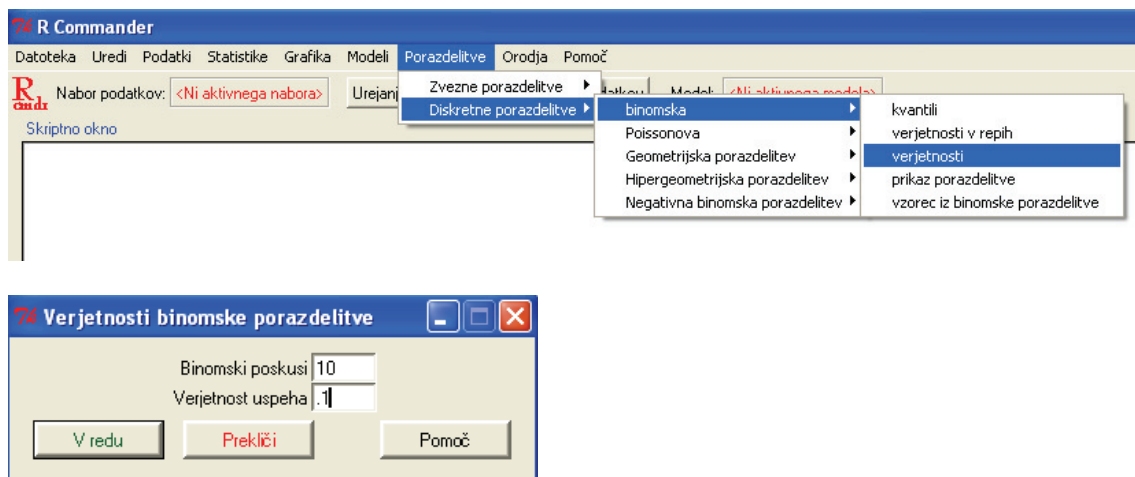
STUDENTOVA PORAZDELITEV 4

Vrednosti t_i določite še z uporabo tabel. ♦

7.4 Binomska porazdelitev

BINOMSKA 11

Izračunajte verjetnosti $p(x)$ za binomsko porazdelitev s parametroma $n = 10$ in $p = 0,1$. Verjetnosti in porazdelitveno funkcijo prikažite grafično, kot kaže slika 7.10.



Navodilo:

v skriptnem oknu popravimo izraz `.x <- 0:5`, da dobimo sliko 7.10 (levo)

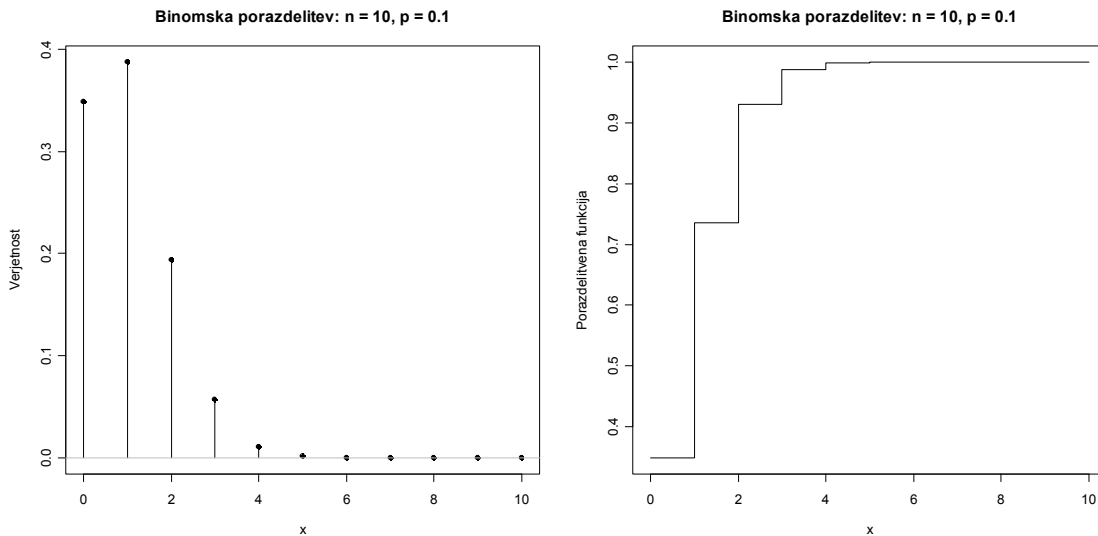
```
.x <- 0:10
```

```
plot(.x, dbinom(.x, size=10, prob=0.1), xlab="x", ylab="Verjetnost", main="Binomska porazdelitev: n = 10, p = 0.1", type="h")
```

```
points(.x, dbinom(.x, size=10, prob=0.1), pch=16)
```

```
abline(h=0, col="gray")
```

```
remove(.x) ♦
```



Slika 7.10: Verjetnosti (levo) in porazdelitvena funkcija (desno) za binomsko porazdelitev $b(n = 10, p = 0,1)$

BINOMSKA 12

Izračunajte verjetnosti $p(x)$ še za binomske porazdelitve s parametrom $n = 10$ in tremi različnimi vrednostmi za $p = 0,3$, $p = 0,5$ in $p = 0,7$. Verjetnostne porazdelitve grafično prikažite. ♦

BINOMSKA 13

Za binomsko porazdelitev $b(10, 0,3)$ izračunajte verjetnost, da se bo dogodek A zgodil največ trikrat.

Rešitev:

```
pbinom(c(3), size=10, prob=0.3, lower.tail=TRUE)
```

```
[1] 0.6496107 ♦
```

BINOMSKA 14

Za binomsko porazdelitev $b(10, 0,3)$ izračunajte verjetnost, da se bo dogodek A zgodil več kot trikrat.

Rešitev:

```
pbinom(c(3), size=10, prob=0.3, lower.tail=FALSE)
```

```
[1] 0.3503893 ♦
```

BINOMSKA 15

Za binomsko porazdelitev $b(10, 0,3)$ izračunajte verjetnost, da se bo dogodek A zgodil trikrat.

Rešitev:

```
.Table <- data.frame(Pr=dbinom(0:10, size=10, prob=0.3))
rownames(.Table) <- 0:10
.Table
remove(.Table)
      Pr
0 0.0282475249
1 0.1210608210
2 0.2334744405
3 0.2668279320
4 0.2001209490
5 0.1029193452
6 0.0367569090
7 0.0090016920
8 0.0014467005
9 0.0001377810
10 0.0000059049
```

Verjetnost, da se bo dogodek zgodil trikrat, je 0,2668. ♦

BINOMSKA 16

Za binomsko porazdelitev $b(10, 0,3)$ izračunajte verjetnost, da se bo dogodek A zgodil štirikrat ali petkrat.

Rešitev:

```
# izraz napišemo v skriptno okno, številke kopiramo iz zgornje tabele
P4ali5=0.2001209490+0.1029193452
P4ali5
[1] 0.3030403 ♦
```


8 Intervali zaupanja

Novi pojmi:

- interval zaupanja za povprečje μ za normalno porazdelitev $N(\mu, \sigma)$; σ znana, σ ni znana
- interval zaupanja za p za binomsko porazdelitev $b(n, p)$
- izračun velikosti vzorca
- funkciji IF in COUNTIF
- funkcija ROUNDUP

Datoteke s podatki:

INTERVALI ZAUPANJA.XLS

8.1 Interval zaupanja za povprečje

8.1.1 σ je znana

Pri vnaprej določeni verjetnosti α takole izračunamo spodnjo mejo (l_1) in zgornjo mejo (l_2) za $(1 - \alpha) \cdot 100\%$ interval zaupanja za povprečje v primeru znane σ :

$$l_{1,2} = \bar{x} \pm d,$$

$d = z_{\alpha/2} \cdot \sigma / \sqrt{n}$ je odklon intervala zaupanja.

Zgled 8.1: Interval zaupanja za zgornji krvni tlak pri ženskah

Raziskovalci so ugotovili, da je za zgornji krvni tlak (spremenljivka X) pri ženskah, starih od 25 do 35 let, primerna normalna porazdelitev.

KRVNI TLAK 13

V vzorcu velikosti 25 smo dobili vzorčno aritmetično sredino zgornjega krvnega tlaka 123,7 mm. Izračunajte spodnjo in zgornjo mejo 95 % intervala zaupanja za povprečni zgornji krvni tlak. Predpostavite, da je σ znana, $\sigma = 15$ mm.

Navodilo: izračune najprej naredite z računalom in uporabo tabel, nato pa še v Excelu v datoteki INTERVALI ZAUPANJA.XLS na delovnem listu »krvni tlak« (stolpec B v tabeli 8.1).

◆

Tabela 8.1: Prikaz dela delovnega lista »KRVNI TLAK« v datoteki INTERVALI ZAUPANJA.XLS. V stolpcu B (levo) so izpisane excelove formule za izračun intervalov zaupanja za povprečje, če je sigma znana. Uporabljeno je relativno naslavljanje celic z namenom, da lahko formule kopiramo v stolpce C, D in E. V desni tabeli so izpisani rezultati formul.

	A	B	C	D	E
2	Intervali zaupanja za povprečje, sigma znana				
3	Zaupanje				
4		0,95	0,99	0,95	0,99
5	n	25	25	250	250
6	\bar{x}_{povp}	123,7			
7	sigma	15			
8	alfa	=1-B4			
9	$Z_{\alpha/2}$	=NORMINV(1-B8/2;0;1)			
10	d	=B9*B7/SQRT(B5)			
11	l_1	=B6-B10			
12	l_2	=B6+B10			
13	\bar{x}_{povp}	=B6			

	A	B	C	D	E
2	Intervali zaupanja za povprečje, sigma znana				
3	Zaupanje				
4		95%	99%	95%	99%
5	n	25	25	250	250
6	\bar{x}_{povp}	123,7	123,7	123,7	123,7
7	sigma	15	15	15	15
8	alfa	0,05	0,01	0,05	0,01
9	$Z_{\alpha/2}$	1,960	2,576	1,960	2,576
10	d	5,88	7,73	1,86	2,44
11	l_1	117,8	116,0	121,8	121,3
12	l_2	129,6	131,4	125,6	126,1
13	\bar{x}_{povp}	123,7	123,7	123,7	123,7

KRVNI TLAK 14

Ponovite izračun pri 99 % zaupanju (tabela 8.1, stolpec C). ◆

KRVNI TLAK 15

Izračunajte še 95 % in 99 % interval zaupanja za povprečni zgornji krvni tlak ob predpostavki, da je σ znana in da je velikost vzorca $n = 250$ (tabela 8.1, stolpca D in E). ◆

KRVNI TLAK 16

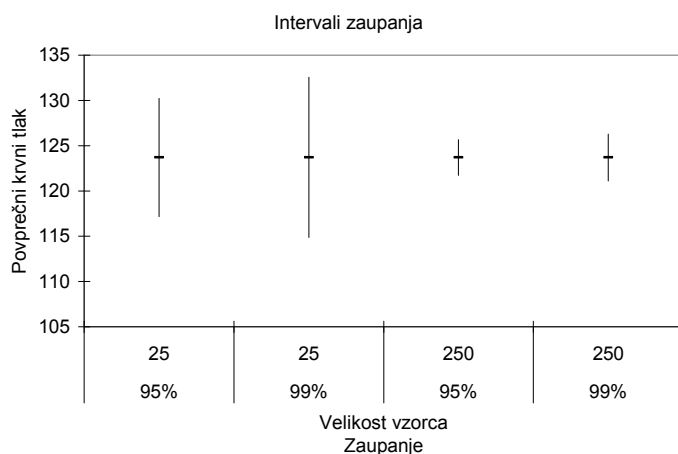
Izračunane intervale zaupanja grafično prikažite, kot kaže slika 8.1.

Navodilo: najprej iz pete v dvanajsto vrstico kopiramo vzorčno povprečje, da primerno pripravimo razporeditev podatkov za grafični prikaz intervalov zaupanja. Nato izberemo (označimo) tri vrste podatkov, urejene v sledečem vrstnem redu: l_1 , l_2 in \bar{x}_{povp} . Te tri vrste podatkov predstavljajo tri nize, ki so ob uporabi vrste grafikona **Borzni, prva podvrsta** prikazani na poseben način. Vsak IZ je na borznem grafikonu prikazan z navpično črto. Vzorčna

povprečja so na sredini teh črt po osnovni nastavitvi označena s kratko vodoravno črtico. Spodnji konci črt predstavljajo spodnje meje IZ l_1 in zgornji konci črt predstavljajo zgornje meje IZ l_2 .

Če želimo znak (predstavitelj) zamenjati ali dodatno oblikovati, z levim klikom na to črtico izberemo celoten niz vzorčnih povprečij, nato z desnim klikom miške odpremo pomožni meni in izberemo ukaz **Oblikuj nize podatkov**. Če želimo spremeniti obliko posameznega predstavitelja, po izbiri niza podatkov z levim gumbom miške ponovno kliknemo predstavitelja in na pomožnem meniju dobimo ukaz **Oblikuj podatkovno točko**.

Črto, ki označuje pravo povprečje, narišemo z orodjem **Risanje**. ♦



Slika 8.1: Grafični prikaz različnih intervalov zaupanja za povprečni zgornji krvni tlak žensk, starih od 25 do 35 let, ob predpostavki, da je sigma znana

8.1.2 σ ni znana

Pri vnaprej določeni verjetnosti α takole izračunamo spodnjo mejo (l_1) in zgornjo mejo (l_2) za $(1 - \alpha) \cdot 100\%$ interval zaupanja za povprečje v primeru neznane σ :

$$l_{1,2} = \bar{x} \pm d,$$

odklon intervala zaupanja tokrat izračunamo drugače: $d = t_{\alpha/2}(SP) \cdot s / \sqrt{n}$, $SP = n - 1$.

V Excelu izračunamo vrednost $t_{\alpha/2}(SP)$ s funkcijo TINV(probability; deg_freedom).

Funkcija ima dva argumenta: »probability« je verjetnost α in »deg_freedom« so

stopinje prostosti SP. Opozarjamo, da pri funkciji TINV verjetnost α ne razpolovimo, saj to Excel naredi sam.

KRVNI TLAK 17

V vzorcu velikosti 25 smo dobili vzorčno aritmetično sredino zgornjega krvnega tlaka 123,7 mm in vzorčni standardni odklon 15,2 mm. Izračunajte spodnjo in zgornjo mejo 95 % intervala zaupanja za povprečni zgornji krvni tlak ob predpostavki, da σ ne poznate. Intervale zaupanja obrazložite.

Navodilo: izračune najprej naredite z računalom in uporabo tabel, nato pa še v Excelu v datoteki INTERVALI ZAUPANJA.XLS na delovnem listu »krvni tlak« (tabela 8.2, stolpec B).

Rešitev:

Z 95 % verjetnostjo trdimo, da je interval (117,4, 130,0) eden tistih, ki vsebujejo μ , to je povprečni zgornji krvni tlak v populaciji žensk, starih od 25 do 35 let.

Ali pa:

Pri 95 % zaupanju interval (117,4, 130,0) vsebuje vrednost za povprečni krvni tlak v populaciji žensk v starostni skupini med 25 in 35 let. ♦

KRVNI TLAK 18

Ponovite izračun pri 99 % zaupanju (tabela 8.2, stolpec C). ♦

Tabela 8.2: Prikaz dela delovnega lista »krvni tlak« v datoteki INTERVALI ZAUPANJA.XLS. V stolpcu B (levo) so izpisane excelove formule za izračun intervalov zaupanja za povprečje, če sigma ne poznamo. Uporabljeno je relativno naslavljanje celic z namenom, da lahko formule kopiramo v stolpce C, D in E. V desni tabeli so podani rezultati formul.

	A	B	C	D	E
16	Intervali zaupanja za povprečje, sigma ni znana				
17	Zaupanje				
18		0,95	0,99	0,95	0,99
19	n	25	25	250	250
20	\bar{x}_{povp}	123,7			
21	s	15,2			
22	alfa	=1-B18			
23	SP=n-1	=B19-1			
24	$t_{\alpha/2}(SP)$	=TINV(B22;B23)			
25	d	=B24*B21/SQRT(B19)			
26	l_1	=B20-B25			
27	l_2	=B20+B25			
28	\bar{x}_{povp}	=B20			

	A	B	C	D	E
16	Intervali zaupanja za povprečje, sigma ni znana				
17	Zaupanje				
18		95%	99%	95%	99%
19	n	25	25	250	250
20	\bar{x}_{povp}	123,7	123,7	123,7	123,7
21	s	15,2	15,2	15,2	15,2
22	alfa	0,05	0,01	0,05	0,01
23	SP=n-1	24	24	249	249
24	$t_{\alpha/2}(SP)$	2,0639	2,7969	1,9695	2,5957
25	d	6,27	8,50	1,89	2,50
26	l_1	117,4	115,2	121,8	121,2
27	l_2	130,0	132,2	125,6	126,2
28	\bar{x}_{povp}	123,7	123,7	123,7	123,7

KRVNI TLAK 19

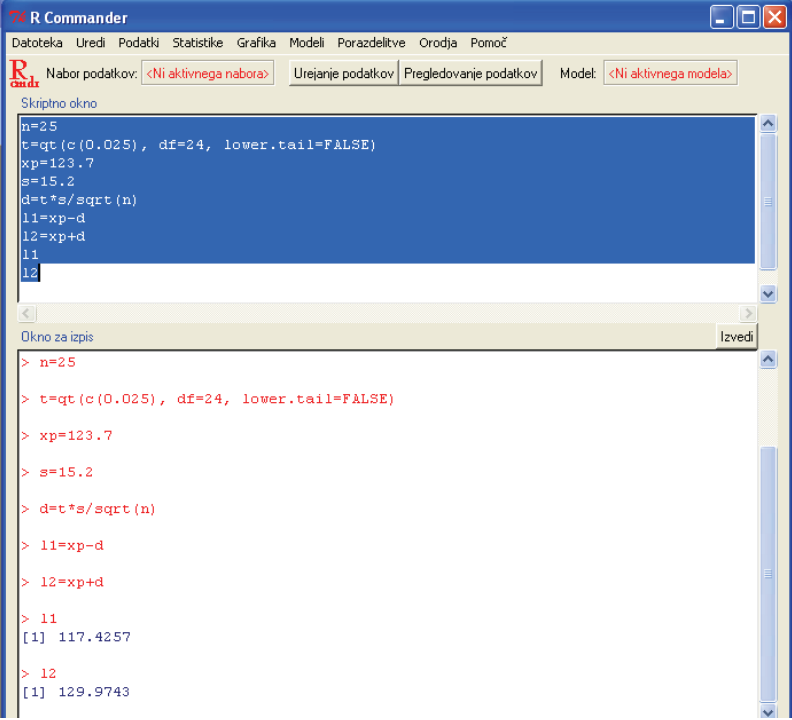
Izračunajte še 95 % in 99 % interval zaupanja za povprečni zgornji krvni tlak ob predpostavki, da σ ne poznate in da je velikost vzorca $n = 250$ (tabela 8.2, stolpca D in E). ♦

KRVNI TLAK 20

Izračunane intervale zaupanja grafično prikažite. ♦

KRVNI TLAK 21

Za vajo ponovite izračune intervalov zaupanja za povprečni zgornji krvni tlak s programom R v skriptnem oknu R Commanderja. Ukaze, ki so na sliki označeni, najprej napišemo v skriptno okno (kot bi formule pisali v zvezek), nato vse vrstice z ukazi označimo in kliknemo gumb **Izvedi**. Rezultat je prikazan v oknu za izpis. ♦



```

R Commander
Datoteka Uredi Podatki Statistike Grafika Modeli Porazdelitve Orodja Pomoč
Nabor podatkov: <Ni aktivnega nabora> Urejanje podatkov Pregledovanje podatkov Model: <Ni aktivnega modela>

Skriptno okno
n=25
t=qt(c(0.025), df=24, lower.tail=FALSE)
xp=123.7
s=15.2
d=t*s/sqrt(n)
l1=xp-d
l2=xp+d
l1
l2

Okno za izpis
Izvedi
> n=25
> t=qt(c(0.025), df=24, lower.tail=FALSE)
> xp=123.7
> s=15.2
> d=t*s/sqrt(n)
> l1=xp-d
> l2=xp+d
> l1
[1] 117.4257
> l2
[1] 129.9743

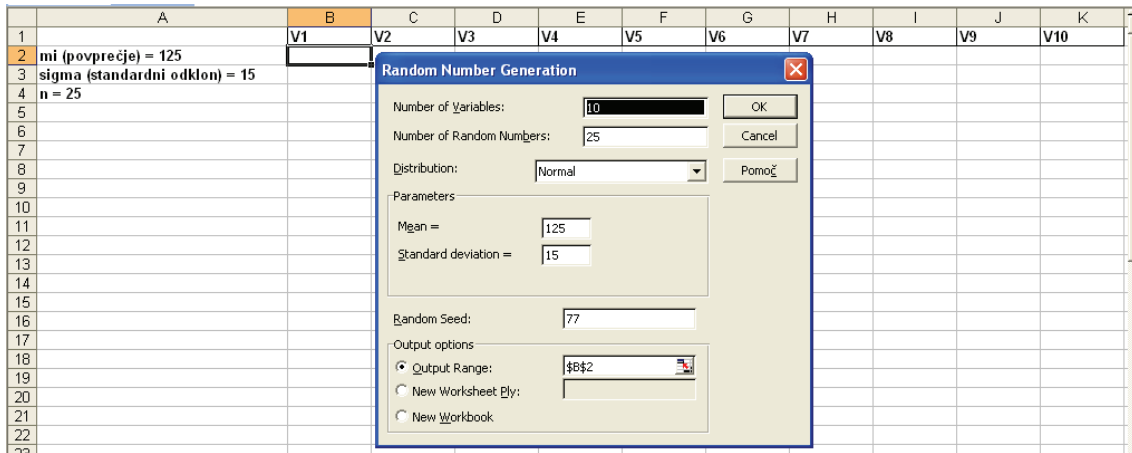
```

Zgled 8.2: Deset intervalov zaupanja

INTERVALI ZAUPANJA 1

V datoteki INTERVALI ZAUPANJA.XLS odprite delovni list z imenom »deset vzorcev«. V prvi vrstici so oznake vzorcev V1 do V10, prvi stolpec je prazen. V celice B2:K26 generirajte vrednosti slučajne spremenljivke X , ki je porazdeljena po $N(125, 15)$ za 10 vzorcev velikosti $n = 25$.

Navodilo: uporabite ukaz **Orodja/Analiza podatkov/Random Number Generation, Distribution: Normal**, za **Random Seed** napišite 77: ♦



	A	B	C	D	E	F	G	H	I	J	K
1		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
2	mi (povprečje) = 125	89,4227501	124,727459	129,832282	129,568335	125,300082	128,307827	127,611603	106,327892	107,278632	124,517399
3	sigma (standardni odklon) = 15	139,376621	151,451971	101,421259	143,401272	136,336681	113,810862	116,757437	105,667462	120,942335	134,080429
4	n = 25	123,679277	119,267796	148,778193	115,982889	142,48607	96,0701569	150,176337	136,778286	136,522286	143,160108
5		107,849124	132,279561	135,669874	84,3390385	111,215255	126,352987	126,933677	126,275805	108,804707	114,341927
6		150,041277	172,907815	109,153817	137,318543	114,681862	113,326078	115,194885	151,761791	138,339218	127,156048
7		118,735776	136,681708	115,920952	130,798915	125,757086	112,316062	144,596177	114,784845	116,368986	109,303406
8		130,733432	119,01889	132,356823	136,178531	123,441797	115,029471	137,494615	121,403483	116,45009	135,825636
9		137,519001	116,784125	124,162492	103,043274	142,234947	105,371422	143,474532	128,420803	119,576456	128,034415
10		123,306754	119,298662	111,368562	120,207588	125,677824	106,951623	110,510351	113,394291	133,904647	140,493697
11		111,931243	129,151781	133,423308	135,72018	135,037888	113,958848	122,293992	125,258763	147,978134	106,729525
12		117,352912	134,470807	122,117681	142,11561	107,70049	131,369544	120,746788	87,5395603	138,615852	136,714354
13		81,0888682	127,624427	170,810157	120,759919	123,408339	113,204524	125,094661	104,89421	105,20819	150,960749
14		133,706968	147,297309	129,927864	127,849595	168,51059	117,426001	101,087838	102,657535	144,715344	117,597554
15		112,87813	102,896221	129,490266	132,248593	127,070288	129,583973	113,853256	93,9207925	113,008108	112,3683217
16		115,238882	119,350435	135,681708	123,189255	108,255361	140,820126	140,928617	123,72764	147,507334	133,104126
17		110,392787	102,667972	111,288583	142,384605	132,741323	157,691878	148,169196	146,718324	92,0568286	117,998028
18		97,0069809	124,712537	120,509734	113,812397	130,074639	130,050356	120,608027	112,480999	97,4341251	131,124253
19		96,2234967	142,719049	140,341891	130,427228	142,014793	136,486884	123,075583	113,234197	113,514651	139,790805
20		146,46287	141,906063	144,518347	114,147557	117,24425	119,18026	131,058224	117,518394	120,211203	119,238158
21		132,222609	101,005711	140,662897	110,087436	120,40762	101,939944	110,45353	120,219662	135,764688	117,959624
22		132,274411	120,666689	131,726754	136,85025	129,927864	135,936105	113,428328	108,293833	123,097564	113,649574
23		127,196634	127,814551	162,512291	116,019979	124,489842	110,895306	111,705018	116,75609	146,517337	126,789135
24		126,979959	104,944823	136,840871	120,049114	118,037676	133,035164	116,348659	128,107078	122,792929	124,864019
25		112,214051	133,037807	114,54479	116,610252	108,996588	111,20649	100,733749	130,063708	119,330722	108,087423
26		126,518981	145,05238	95,1139586	123,127629	102,792322	120,8518	135,046517	130,058848	140,159117	147,242057

INTERVALI ZAUPANJA 2

Za vzorce velikosti $n = 25$, generirane po normalni porazdelitvi, izračunajte 95 % intervale zaupanja za povprečje (tabeli 8.3 in 8.4, do vrstice 39). Računajte s predpostavko, da σ ocenimo z vzorčnim standardnim odklonom s , izračunanim na osnovi vrednosti v vzorca. Formule napišite le za prvi vzorec (stolpec B), nato jih za ostale vzorce kopirajte. ♦

Tabela 8.3: Prikaz dela delovnega lista »deset vzorcev« v datoteki INTERVALI ZAUPANJA.XLS. V stolpcu B so izpisane excelove formule za izračun intervalov zaupanja za povprečje. Uporabljeno je relativno naslavljanje celic z namenom, da se vsi izračuni samodejno preračunajo, če spremenimo zaupanje ali če na mesto generiranih podatkov (v vrstice 2:26) ponovno generiramo vrednosti z drugačnim »brstom«.

	A	B
28	zaupanje	0,95
29	vzorčne vrednosti	V1
30	\bar{x}_{povp}	=AVERAGE(B2:B26)
31	s	=STDEV(B2:B26)
32	n	=COUNT(B2:B26)
33	alfa	=1-\$B\$28
34	SP	=B32-1
35	$t_{\alpha/2}(SP=n-1)$	=TINV(B33,B34)
36	d	=B35*B31/SQRT(B32)
37	I1	=B30-B36
38	I2	=B30+B36
39	\bar{x}_{povp}	=B30
40	Ali je populacijsko povprečje pod spodnjo mejo, $I_1 > 125$? DA - 1, NE - 0	=IF(B37>125,1;0)
41	Ali je populacijsko povprečje nad zgornjo mejo, $I_2 < 125$? DA - 1, NE - 0	=IF(B38<125,1;0)
42	Ali 95 % IZ vsebuje populacijsko povprečje? DA - 0, NE - 1	=SUM(B40:B41)
43	Število intervalov, ki vsebujejo populacijsko povprečje:	=10-SUM(B42:K42)

INTERVALI ZAUPANJA 3

Preštejte IZ, ki vsebujejo populacijsko povprečje ($\mu = 125$) (tabeli 8,3 in 8.4, vrstice 40-43).

Navodilo: uporabite logično funkcijo IF.

Funkcija IF(logical_test; value_if_true; value_if_false) vrne eno izmed dveh vrednosti, ki jih določimo kot argumenta te funkcije. Funkcija ima skupaj tri argumente: »logical_test« je logični izraz. Če je pravilen, funkcija vrne vrednost, ki nastopa kot drugi argument »Value_if_true«; če logični izraz ni pravilen, funkcija vrne vrednost, ki jo zapišemo kot tretji argument »Value_if_false« (tabela 8.3, vrstici 40-41):

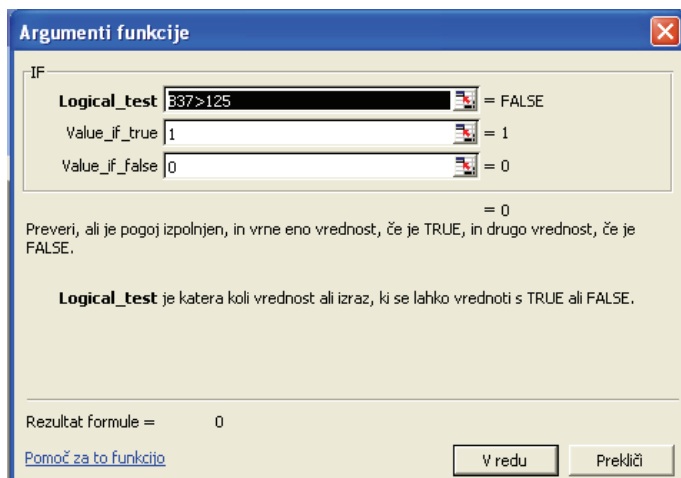


Tabela 8.4: Prikaz izračunov za 95 % interval zaupanja za povprečje na delovnem listu »deset vzorcev« v datoteki INTERVALI ZAUPANJA.XLS. Brst generatorja slučajnih števil je 77.

	A	B	C	D	E	F	G	H	I	J	K
28	zaupanje	0,95									
29	vzorčne vrednosti	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
30	\bar{x}_{povp}	120,01	127,87	129,13	124,25	125,75	120,85	124,30	118,65	124,24	126,45
31	s	17,13	16,86	17,57	13,70	14,25	13,94	14,06	14,95	15,99	12,61
32	n	25	25	25	25	25	25	25	25	25	25
33	alfa	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
34	SP	24	24	24	24	24	24	24	24	24	24
35	$t_{alfa/2}(SP=n-1)$	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639
36	d	7,07	6,96	7,25	5,65	5,88	5,75	5,80	6,17	6,60	5,21
37	I1	112,94	120,91	121,87	118,60	119,87	115,09	118,49	112,48	117,64	121,24
38	I2	127,09	134,83	136,38	129,90	131,64	126,60	130,10	124,82	130,85	131,65
39	\bar{x}_{povp}	120,01	127,87	129,13	124,25	125,75	120,85	124,30	118,65	124,24	126,45
40	Ali je populacijsko povprečje pod spodnjo mejo, $I_1 > 125$? DA - 1, NE - 0	0	0	0	0	0	0	0	0	0	0
41	Ali je populacijsko povprečje nad zgornjo mejo, $I_2 < 125$? DA - 1, NE - 0	0	0	0	0	0	0	0	1	0	0
42	Ali 95 % IZ vsebuje populacijsko povprečje? DA - 0, NE - 1	0	0	0	0	0	0	0	1	0	0
43	Število intervalov, ki vsebujejo populacijsko povprečje:	9									

INTERVALI ZAUPANJA 4

Intervale zaupanja grafično prikažite, kot kaže slika 8.2 (zgoraj). Grafikon shranite na grafičnem delovnem listu »slika IZ, deset vzorcev«. ♦

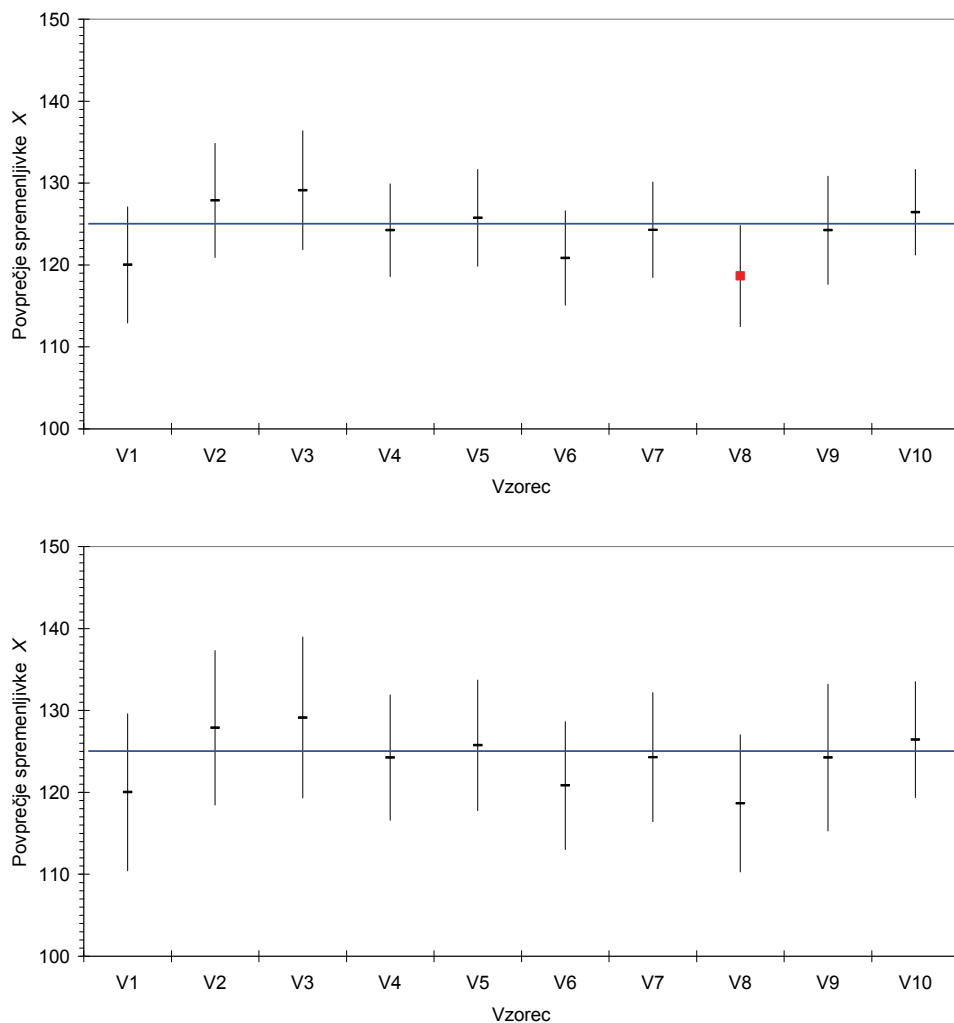
INTERVALI ZAUPANJA 5

Izračunajte še 99 % IZ za povprečje za 10 generiranih vzorcev na delovnem listu »deset vzorcev«. Koliko IZ vsebuje populacijsko povprečje $\mu = 125$? Intervale zaupanja grafično prikažite (slika 8.2 spodaj).

Navodilo: v tabeli z izračuni za 95 % interval zaupanja ustrezno popravite zaupanje. Ponovno preglejte formule in ugotovite, ali so s tem izračuni ustrezni za 99 % interval zaupanja.

Ali grafikon na grafičnem delovnem listu »slika IZ deset vzorcev« po spremembi zaupanja prikazuje 99 % IZ ali 95 % IZ?

Kako se s povečanjem zaupanja spremeni širina intervala zaupanja? ♦



Slika 8.2: Grafični prikaz 95 % IZ (zgoraj) in 99 % IZ (spodaj) za populacijsko povprečje za vzorce velikosti $n = 25$. Podatki so bili generirani po normalni porazdelitvi $N(125, 15)$. Z rdečim kvadratom je označeno vzorčno povprečje, katerega IZ ne vsebuje populacijskega povprečja $\mu = 125$. Modra črta je dodana z orodjem **Risanje**.

INTERVALI ZAUPANJA 6

Ponovite generiranje vrednosti slučajne spremenljivke X , ki je porazdeljena po $N(125, 15)$, za 10 vzorcev velikosti $n = 25$ v celice B2:K26, tako da si izberete za brst generatorja slučajnih števil, različen od 77. Koliko 95 % IZ vsebuje populacijsko povprečje? Kaj pa pri 99 % zaupanju? Oglejte si grafični prikaz na delovnem listu »slika IZ deset vzorcev«. ♦

Zgled 8.3: Sto intervalov zaupanja za povprečje

INTERVALI ZAUPANJA 7

V datoteki INTERVALI ZAUPANJA.XLS odprite delovni list z imenom »sto vzorcev«. V celice B2:CW26 generirajte vrednosti slučajne spremenljivke X , ki je porazdeljena po $N(125, 15)$ za 100 vzorcev velikosti $n = 25$.

Navodilo: uporabite ukaz **Orodja/Analiza podatkov/Random Number Generation, Distribution: Normal**, polje **Random Seed** pustite prazno. ♦

INTERVALI ZAUPANJA 8

Za vzorce velikosti $n = 25$, generirane po normalni porazdelitvi, izračunajte 95 % in 99 % intervale zaupanja za povprečje. Računajte ob predpostavki, da σ ocenimo z vzorčnim standardnim odklonom s , izračunanim na osnovi vrednosti v vzorcu.

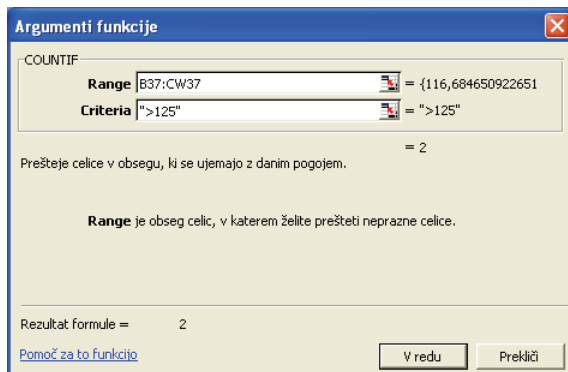
Navodilo: izračune uredite enako, kot v tabeli 8.3. Ali lahko formule iz tabele 8.3 na delovnem listu »deset vzorcev« kopiramo na delovni list »sto vzorcev«? Katero formulo moramo popraviti? ♦

INTERVALI ZAUPANJA 9

Preštejte IZ, ki ne vsebujejo populacijskega povprečja ($\mu = 125$). Ali je rezultat v skladu s pričakovanji, vezanimi na definicijo intervala zaupanja?

Navodilo: namesto dvakratne uporabe funkcije IF in seštevanja (vrstice 40-43 v tabeli 8.3) uporabite logično funkcijo COUNTIF.

Funkcijo COUNTIF(range; criteria) najdemo v skupini statističnih funkcij. Preštejte celice v obsegu, ki ustrezajo danim pogojem. Funkcija ima dva argumenta: »range« je obseg podatkov in »criteria« je logični izraz.



	A	B
40	Ali je populacijsko povprečje pod spodnjo mejo, $I_1 > 125$? DA - 1, NE - 0	=IF(B37>125;1;0)
41	Ali je populacijsko povprečje nad zg. mejo, $I_2 < 125$? DA - 1, NE - 0	=IF(B38<125;1;0)
42	Ali 95 % IZ vsebuje populacijsko povprečje? DA - 1, NE - 0	=SUM(B40:B41)
43	Število intervalov, ki vsebujejo populacijsko povprečje:	=100-SUM(B42:CW42)
44	Število intervalov, ki ne vsebujejo populacijskega povprečja	=COUNTIF(B37:CW37;">125")+COUNTIF(B38:CW38;"<125")



INTERVALI ZAUPANJA 10

Intervale zaupanja grafično prikažite v istem stilu, kot na sliki 8.2. Grafični delovni list poimenujte »slika IZ sto vzorcev«. ◆

INTERVALI ZAUPANJA 11

Primerjajte širine intervalov zaupanja. Kaj opazite? Katere količine vplivajo na širino intervala zaupanja? Kako? ◆

INTERVALI ZAUPANJA 12

Ponovite generiranje vrednosti slučajne spremenljivke X , ki je porazdeljena po $N(125, 15)$, za 100 vzorcev velikosti $n = 25$ v celice B2:CW26. Koliko 95 % IZ vsebuje populacijsko povprečje? Kaj pa pri 99 % zaupanju? Oglejte si grafični prikaz na delovnem listu »slika IZ sto vzorcev«. ◆

8.2 Velikost vzorca za povprečje pri normalni porazdelitvi

8.2.1 Standardni odklon je znan

Zanima nas, kako velik vzorec moramo imeti, če želimo pri zaupanju $(1 - \alpha) \cdot 100\%$ oceniti populacijsko povprečje, tako da je razlika med oceno povprečja (\bar{x}) in

populacijskim povprečjem (μ) največ d . Pri tem predpostavimo, da je σ znana. V takem primeru velikost vzorca izračunamo z izrazom

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{d} \right)^2 .$$

Da dobimo celo število, rezultat vedno zaokrožimo navzgor.

VELIKOST VZORCA 1

Kako velik vzorec moramo imeti, če hočemo, da bo pri 95 % zaupanju razlika med oceno povprečja in populacijskim povprečjem zgornjega krvnega tlaka največ 10 mm? Kaj pa če je največja dopustna razlika 5 mm? Predpostavite, da je σ znana, njena vrednost je 15 mm.

Navodilo: v datoteki INTERVALI ZAUPANJA.xls na delovnem listu »velikost vzorca, sigma znana« izpolnite tabelo 8.5. Formule napišite tako, da se vam ob spreminjanju vrednosti za σ za zaupanje $1 - \alpha$ in d samodejno izračuna potrebna velikost vzorca n .

Za zaokrožitev števil navzgor uporabimo funkcijo ROUNDUP(number; num_digits), ki ima dva argumenta, »number« predstavlja vrednost oz. referenco celice, v kateri je vrednost, ki jo želimo zaokrožiti navzgor na izbrano število decimalnih mest. Drugi argument funkcije »num_digits« je število zelenih decimalnih mest navzgor zaokroženega števila. ♦

Tabela 8.5: Prikaz izračuna potrebne velikosti vzorca, če je sigma znana, na delovnem listu »velikost vzorca, sigma znana« v datoteki INTERVALI ZAUPANJA.XLS

sigma	napišemo vrednost za σ
zaupanje ($1 - \alpha$)	napišemo želeno zaupanje (npr. 0,95)
α	napišemo formulo za izračun α
$z_{\alpha/2}$	funkcija NORMINV
d	napišemo želeno maksimalno odstopanje od povprečja
Velikost vzorca	
n	napišemo formulo
navzgor zaokroženo število n	funkcija ROUNDUP

8.2.2 Standardni odklon ni znan – iterativni postopek

Zanima nas, kako velik vzorec moramo imeti, če želimo pri zaupanju $(1 - \alpha) \cdot 100\%$ oceniti populacijsko povprečje, tako da je razlika med oceno povprečja (\bar{x}) in populacijskim povprečjem (μ) največ d . Pri tem predpostavimo, da σ ni znana,

imamo pa njeno vzorčno oceno s . V takem primeru velikost vzorca izračunamo z iterativnim postopkom. Prvi približek za velikost vzorca izračunamo z izrazom:

$$n_0 = \left(\frac{z_{\alpha/2} \cdot s}{d} \right)^2$$

V drugem koraku velikost vzorca izračunamo z izrazom:

$$n_1 = \left(\frac{t_{\alpha/2}(SP = n_0 - 1) \cdot s}{d} \right)^2$$

Če je n_1 večje od n_0 , iterativni postopek nadaljujemo:

$$n_2 = \left(\frac{t_{\alpha/2}(SP = n_1 - 1) \cdot s}{d} \right)^2, \text{ sicer je } n_1 \text{ iskana rešitev.}$$

Da dobimo celo število, rezultat izračuna v vsakem koraku iteracije zaokrožimo navzgor. Ko je na nekem koraku izračunani n enak ali celo manjši kot v prejšnjem koraku, je velikost vzorca izračunana.

VELIKOST VZORCA 2

Za generirane vzorce v datoteki INTERVALI ZAUPANJA.XLS predpostavite, da σ ni znana, za vsak vzorec posebej imamo le izračunane vzorčne standardne odklone s . Za prvih pet vzorcev (pet različnih vzorčnih standardnih odklonov) izračunajte potrebno velikost vzorca, če hočemo, da bo pri 95 % zaupanju razlika med oceno in populacijskim povprečjem največ 5 mm.

Navodilo: izračune uredite na delovnem listu »velikost vzorca, sigma ni znana«, kot kaže tabela 8.6. ♦

Tabela 8.6: Prikaz izračuna potrebne velikosti vzorca, če sigme ne poznamo in uporabimo iterativni postopek na delovnem listu »velikost vzorca, sigma ni znana« v datoteki INTERVALI ZAUPANJA.XLS

	A	B	C	D	E	F
2	Izračun velikosti vzorca, če sigme ne poznamo	V1	V2	V3	V4	V5
3	d	napišemo vrednost				
4	s	kopiramo vzorčni standardni odklon za V1				
5	1. korak					
6	alfa	napišemo vrednost				
7	$Z_{\alpha/2}$	funkcija NORMINV				
8	n_0	napišemo formulo				
9	navzgor zaokroženo število n_0	funkcija ROUNDUP				
10	2. korak					
11	$SP=n_0-1$	napišemo formulo				
12	$t_{\alpha/2}(SP)$	funkcija TINV				
13	n_1	napišemo formulo				
14	navzgor zaokroženo število n_1	funkcija ROUNDUP				
15	3. korak					
16	$SP=n_1-1$	napišemo formulo				
17	$t_{\alpha/2}(SP)$	funkcija TINV				
18	n_2	napišemo formulo				
19	navzgor zaokroženo število n_2	funkcija ROUNDUP				
20	itd. dokler se velikost vzorca večja					

VELIKOST VZORCA 3

Kaj pa, če je zaupanje 99 %? Ponovite izračune še za tak primer. Katero vrednost v preglednici morate spremeniti in se vse samodejno preračuna? Ali pričakujete, da bo velikost vzorca pri večjem zaupanju večja ali manjša? ♦

8.3 Interval zaupanja in velikost vzorca za p pri binomski porazdelitvi $b(n, p)$

Izračun spodnje in zgornje meje intervala zaupanja za verjetnost p temelji na porazdelitvi vzorčnih deležev $\hat{p} = \frac{x}{n}$. Če je slučajna spremenljivka X porazdeljena po binomski porazdelitvi s parametroma n in p in če lahko binomsko porazdelitev aproksimiramo z normalno porazdelitvijo $N(np, \sqrt{np(1-p)})$, je porazdelitev slučajne

spremenljivke $\frac{X}{n}$ (delež) približno normalna $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. Približni interval

zaupanja izračunamo po formuli

$$l_{1,2} = \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{in je primeren, če je } n > 40.$$

INTERVAL ZAUPANJA IN VELIKOST VZORCA ZA p 1

Od 250 slučajno izbranih oseb jih je za gripo zbolelo 45. Izračunajte točkovno oceno za verjetnost, da oseba zboli za gripo. Za izračun te in naslednjih nalog uporabite računalno, lahko pa za vajo izračunate tudi v Excelu (delovni list »IZ za p«, tabela 8.7) ali v R Commanderju. ♦

INTERVAL ZAUPANJA IN VELIKOST VZORCA ZA p 2

Izračunajte 95 % in 99 % interval zaupanja za verjetnost, da oseba zboli za gripo (tabela 8.7). Izračunane vrednosti obrazložite. ♦

Tabela 8.7: Prikaz izračunov za 95 % interval zaupanja za verjetnost²², da oseba zboli za gripo na delovnem listu »IZ za p« v datoteki INTERVALI ZAUPANJA.XLS; izpis formul (levo); izpis rezultatov (desno)

	A	B
2	Približni interval zaupanja za p	
3	zaupanje	0,95
4	n	250
5	x	45
6	p*	=B5/B4
7	1-p*	=1-B6
8	alfa	=1-B3
9	Z _{alfa/2}	=NORMINV(1-B8/2;0;1)
10	d	=B9*SQRT(B6*(1-B6)/B4)
11	l ₁	=B6-B10
12	l ₂	=B6+B10

	A	B
2	Približni interval zaupanja za p	
3	zaupanje	95%
4	n	250
5	x	45
6	p*	0,18
7	1-p*	0,82
8	alfa	0,05
9	Z _{alfa/2}	1,960
10	d	0,05
11	l ₁	0,13
12	l ₂	0,23

Zanima nas, kako velik vzorec moramo imeti, če želimo pri zaupanju $(1-\alpha) \cdot 100\%$ oceniti verjetnost p , tako da je razlika med oceno verjetnosti (\hat{p}) in populacijsko verjetnostjo (p) največ d . V takem primeru velikost vzorca izračunamo z izrazom

$$n = \hat{p}(1-\hat{p}) \cdot \left(\frac{z_{\alpha/2}}{d}\right)^2.$$

²² S p* v excelovih tabelah označujemo vzorčno oceno verjetnosti \hat{p} .

Da dobimo celo število, rezultat izračuna vedno zaokrožimo navzgor.

INTERVAL ZAUPANJA IN VELIKOST VZORCA ZA p 3

Kako velik vzorec oseb bi potrebovali, če želimo pri 95 % zaupanju oceniti verjetnost, da oseba zboli za gripo, tako da bo razlika med oceno in pravo vrednostjo največ 1 odstotna točka (delovni list »velikost vzorca za p«, tabela 8.8)? ♦

Tabela 8.8: Prikaz izračunov za potrebno velikost vzorca ob vnaprej določeni maksimalni vrednosti za odklon intervala zaupanja $d = 0,01$, danem zaupanju in znani oceni za verjetnost, da oseba zboli za gripo $\hat{p} = 0,18$; izpis formul (levo); izpis rezultatov (desno)

	A	B
2	zaupanje	0,95
3	n	250
4	x	45
5	p*	=B4/B3
6	1-p*	=1-B5
7	alfa	=1-B2
8	Z _{alfa/2}	=NORMINV(1-B7/2;0;1)
9	d	0,01
10	Velikost vzorca	
11	n	=B5*B6*(B8/B9)^2
12	navzgor zaokroženo število n	=ROUNDUP(B11;0)

	A	B
2	zaupanje	95%
3	n	250
4	x	45
5	p*	0,18
6	1-p*	0,82
7	alfa	0,05
8	Z _{alfa/2}	1,960
9	d	0,01
10	Velikost vzorca	
11	n	5670,0
12	navzgor zaokroženo število n	5670

INTERVAL ZAUPANJA IN VELIKOST VZORCA ZA p 4

Ponovi izračun potrebne velikosti vzorca še za primere, ko je pri 95 % zaupanju razlika med oceno in pravo vrednostjo največ:

- 5 odstotnih točk,
- 10 odstotnih točk.

Navodilo: spremenite samo podatek za d v tabeli 8.8. Ali se izračuni s tem ustrezno spremenijo? ♦

INTERVAL ZAUPANJA IN VELIKOST VZORCA ZA p 5

Ponovite izračune potrebne velikosti vzorca še pri 99 % zaupanju. ♦

9 Preverjanje statističnih domnev za en vzorec

Novi pojmi:

- ničelna domneva $H_0 : \mu = \mu_0$; σ je znana, σ ni znana
- ničelna domneva $H_0 : p = p_0$
- alternativna domneva dvostranska/enostranska

Datoteke s podatki:

DOMNEVE_EN VZOREC.XLS

Ustvarjene datoteke:

DOMNEVE_EN VZOREC.DOC

Odprite novo wordovo datoteko DOMENE_EN VZOREC.DOC. Vanjo kopirajte vse izračune in grafične prikaze ter napišite obrazložitve rezultatov nalog v tem poglavju. ♦

9.1 Povprečje

Za preverjanje ničelne domneve $H_0 : \mu = \mu_0$ izračunamo testno statistiko $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,

kjer je \bar{x} vzorčna aritmetična sredina; s vzorčni standardni odklon, n velikost vzorca in μ_0 pričakovana vrednost populacijskega povprečja v ničelni domnevi.

Pri dvostranski alternativni domnevi $H_1 : \mu \neq \mu_0$ sta kritični vrednosti $\pm t_{\alpha/2}(SP = n - 1)$, pri enostranski alternativni domnevi $H_1 : \mu > \mu_0$ je $t_{\alpha}(SP = n - 1)$, pri enostranski alternativni domnevi $H_1 : \mu < \mu_0$ pa $-t_{\alpha}(SP = n - 1)$.

Računalniški program namesto kritične vrednosti izračuna p -vrednost in jo napiše poleg vrednosti testne statistike; p -vrednost je verjetnost, ki kaže na izbiro H_0 ali H_1 . Če je p -vrednost »velika« ($p \geq \alpha$), obdržimo H_0 , če pa je »majhna« ($p < \alpha$), zavrnamo H_0 v korist H_1 .

Zgled 9.1: Čakalna doba v zdravstvenem domu

Podatki o čakalni dobi v zdravstvenem domu za vzorec 20 pacientov so v datoteki DOMNEVE_EN VZOREC.XLS na delovnem listu »čakalna doba« in v tabeli 9.1.

Tabela 9.1: Podatki o čakalni dobi v zdravstvenem domu (min) za 20 pacientov.

10	41	38	46	48	34	27	40	35	28
20	16	46	21	36	31	23	54	46	42

ČAKALNA DOBA 1

Na osnovi vzorca preverite ničelno domnevo, da je povprečna čakalna doba v populaciji vseh pacientov tega zdravstvenega doma 30 min ($\alpha = 0,05$). Izračunajte tudi p -vrednost.

Navodila: izračune naredite na dva načina, najprej z računalom in tabelami, nato v Excelu s formulami (tabela 9.2).

Za izračun p -vrednosti v Excelu uporabimo funkcijo TDIST(x; deg_freedom; tails), ki ima tri argumente: »x« je absolutna vrednost t statistike, »deg_freedom« so stopinje prostosti in »tails« je število upoštevanih repov porazdelitve; 1 napišemo, če imamo enostransko alternativno domnevo in 2 napišemo, če imamo dvostransko alternativno domnevo.

Tabela 9.2: Izračuni za t-preizkus v Excelu, formule (levo), rezultati (desno)

	A	B		A	B
23	t-preizkus		23	t-preizkus	
24	\bar{x}_{povp}	=AVERAGE(A2:A21)	24	\bar{x}_{povp}	34,1
25	s	=STDEV(A2:A21)	25	s	11,9
26	n	=COUNT(A2:A21)	26	n	20
27	alfa	0,05	27	alfa	0,05
28	$H_0: \mu_i = \mu_{i0}$	$\mu_{i0} = 30$	28	$H_0: \mu_i = \mu_{i0}$	$\mu_{i0} = 30$
29	test (enostranski, dvostranski)	2	29	test (enostranski, dvostranski)	2
30	SP	=B26-1	30	SP	19
31	$t_{\text{števec}}$	=B24-B28	31	$t_{\text{števec}}$	4,1000
32	$t_{\text{imenovalec}}$	=B25/SQRT(B26)	32	$t_{\text{imenovalec}}$	2,6546
33	t	=B31/B32	33	t	1,5445
34	mejna vrednost $t_{\text{alfa}/2}(\text{SP})$	=TINV(B27;B30)	34	mejna vrednost $t_{\text{alfa}/2}(\text{SP})$	2,0930
35	p-vrednost	=TDIST(ABS(B33);B30;B29)	35	p-vrednost	0,1390

Rešitev:

$$H_0: \mu = \mu_0 = 30 \text{ min in } H_1: \mu \neq 30 \text{ min}$$

$$t = 1,5445 \text{ in } p = 0,1390$$

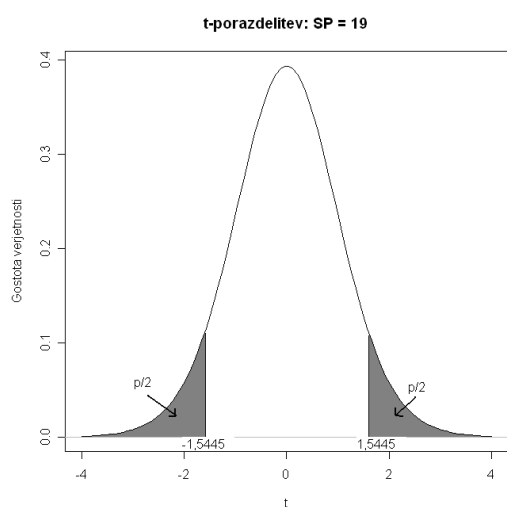
Statistični sklep: ničelno domnevo obdržimo ($p = 0,1390$). Rezultati niso statistično značilni.

Vsebinski sklep: podatki v vzorcu ne nasprotujejo domnevi, da je povprečna čakalna doba v populaciji vseh pacientov tega zdravstvenega doma 30 min. ♦

ČAKALNA DOBA 2

p -vrednost obrazložite in jo shematično grafično prikažite (na papir, slika 9.1). Ali lahko p -vrednost dobite iz tabel Studentove porazdelitve?

Rešitev: verjetnost, da pri dani ničelni porazdelitvi dobimo vrednost t -statistike po absolutni vrednosti večjo od $t = 1,5445$, je $p = 0,1390$.



Slika 9.1: Grafični prikaz p -vrednosti za vrednost testne statistike $t = 1,5445$ in $SP = 19$ pri dvostranski alternativni domnevi

p -vrednost lahko na podlagi tabel Studentove porazdelitve le ocenimo z intervalom verjetnosti (izsek iz tabele):

	$\alpha/2$		
SP	0,1	0,05	0,025
⋮	⋮	⋮	⋮
19	1,328	1,729	2,093
⋮	⋮	⋮	⋮

$$(t_1 = 1,328) < (t = 1,5445) < (t_2 = 1,729) \quad (p_2 = 2 \cdot 0,05) < p < (p_1 = 2 \cdot 0,10)$$

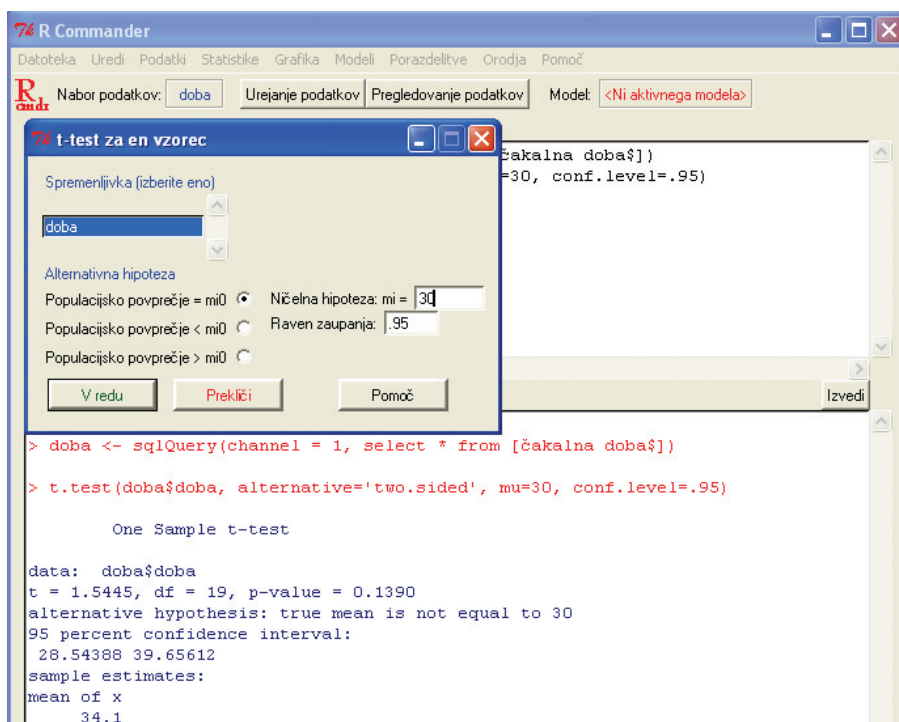
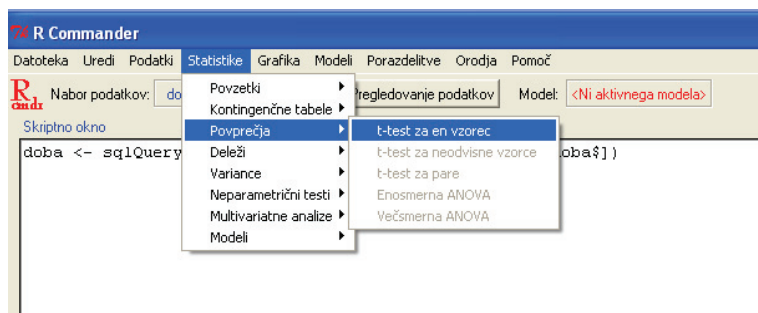
$$0,10 < p < 0,20 \text{ . } \blacklozenge$$

ČAKALNA DOBA 3

Ponovite preverjanje domneve še z uporabo programa R.

Navodila: pred uvozom podatkov iz excelove datoteke v R Commander, jih moramo kopirati na nov delovni list z imenom »doba za R«. Paziti moramo, da kopiramo samo podatke (A1:A21) brez izračunov.

Uporabite ukaz **Statistike/Povprečja/t-test za en vzorec** (glej sliko spodaj). R v tem primeru p -vrednost izračuna samodejno. Izračuna tudi pripadajoč interval zaupanja za populacijsko povprečje.²³ ♦



²³ V pogovornem oknu **t-test za en vzorec** je napaka pri izpisu prve alternativne domneve (hipoteze), pisati bi moralo **Populacijsko povprečje $\neq \mu_0$** .

ČAKALNA DOBA 4

Na osnovi vzorca preverite še ničelno domnevo, da je povprečna čakalna doba 45 min ($\alpha = 0,05$). Izračunajte p -vrednost ter obrazložite rezultate.

Navodilo: uporabite izračune v tabeli 9.2. Katero vrednost morate spremeniti, da se formule preračunajo ustrezno tej nalogi?

	A	B
23	t-preizkus	
24	\bar{x}_{povp}	34,1
25	s	11,9
26	n	20
27	alfa	0,05
28	$H_0: \mu = \mu_0$ $\mu_0 =$	45
29	test (enostranski, dvostranski)	2
30	SP	19
31	t števec	-10,9000
32	t_imenovalec	2,6546
33	t	-4,1061
34	mejna vrednost $t_{\text{alfa}/2}(\text{SP})$	2,0930
35	p-vrednost	0,0006

Rešitev:

$$H_0 : \mu = \mu_0 = 45 \text{ min} \quad \text{in} \quad H_1 : \mu \neq 45 \text{ min}$$

$$t = -4,1061 \quad \text{in} \quad p = 0,0006$$

Statistični sklep: ničelno domnevo zavrnamo pri $\alpha = 0,05$. Povprečna čakalna doba v zdravstvenem domu se statistično značilno razlikuje od 45 min.

Vsebinski sklep: s tveganjem, manjšim od $\alpha = 0,05$, trdimo, da povprečna čakalna doba v populaciji vseh pacientov tega zdravstvenega doma ni 45 min ($p = 0,0006$). ♦

Domača naloga 9.1: Čakalna doba v zdravstvenem domu - ponovitev snovi

ČAKALNA DOBA 5

Vse izračune, grafikone in obrazložitve te domače naloge uredite v wordovi datoteki z imenom DOMNEVE_EN_VZOREC.DOC. ♦

ČAKALNA DOBA 6

Izračunajte vrednosti, ki določajo okvir z ročaji. Ne pozabite preveriti, ali med podatki obstajajo osamelci. Izračunane vrednosti uredite v tabelo v wordovi datoteki in jih vsebinsko obrazložite.



ČAKALNA DOBA 7

S programom R s pomočjo grafičnega vmesnika R Commander podatke grafično prikažite z okvirjem z ročaji. Sliko kopirajte v wordovo datoteko in sliki napišite podnaslov. ◆

ČAKALNA DOBA 8

Izračunajte 2 primerni meri sredine in 2 primerni meri variabilnosti za navedene podatke. Izračune vpišite v tabelo v wordovo datoteko in vrednosti obrazložite. ◆

ČAKALNA DOBA 9

Za čakalno dobo v zdravstvenem domu lahko privzamemo normalno porazdelitev. Na osnovi vzorčnih podatkov izračunajte 95 % interval zaupanja za povprečno čakalno dobo v zdravstvenem domu in ga obrazložite. Izračune z obrazložitvijo vpišite v wordovo datoteko. ◆

ČAKALNA DOBA 10

Kaj bi morali narediti, če bi želeli, da se širina intervala zaupanja 2-krat zmanjša? ◆

ČAKALNA DOBA 11

Kako velik vzorec potrebujemo, če želimo, da pri 95 % zaupanju vzorčno povprečje ne odstopa od populacijske vrednosti za več kot 4 min? Pri izračunu velikosti vzorca uporabite oceno standardnega odklona, ki ste jo dobili na osnovi zgornjega vzorca. ◆

Domača naloga 9.2: Povprečna telesna višina študentk

Med študenti 2. letnika Agronomije, Biotehnologije in Mikrobiologije na BF so v študijskem letu 2006/07 izvedli anketo. Med drugim so študente vprašali o njihovi telesni višini. V datoteki DOMNEVE_EN VZOREC.XLS na delovnem listu »telesna višina« so podatki za študentke. Le-te bomo šteli za reprezentativni vzorec populacije vseh študentk.

TELESNA VIŠINA 1

Za telesno višino študentk lahko predpostavimo, da je porazdeljena po normalni porazdelitvi. Pri stopnji značilnosti 0,05 preverite domnevo, da je povprečna telesna višina študentk 165 cm. Izračunajte tudi p -vrednost. Obrazložite rezultate.

Rešitev: $t = 4,9228$ in $p = 0,0000$. ♦

TELESNA VIŠINA 2

p -vrednost shematično grafično prikažite (na list papirja). ♦

Zgled 9.2: Debelina debel dreves

V tabeli 4.3 in v datoteki DOMNEVE_EN VZOREC.XLS na delovnem listu »debelina debel« (tudi v DEBELINA DEBEL.XLS) so podatki o številu dreves v posameznem razredu glede na debelino (premer) debla.

DEBELINA DEBEL 8

Na podlagi približka za povprečje in ob predpostavki, da je znana $\sigma = 6$ cm, preverite domnevo, da je povprečna debelina debla 22 cm ($\alpha = 0,01$). Izračunajte p -vrednost. Obrazložite rezultate.

Navodila: izračune naredite v Excelu (datoteka DOMNEVE_EN VZOREC.XLS, delovni list »debelina debel«, izračun približka za povprečje je prikazan v zgledu 4.3).

Rešitev:

$H_0 : \mu = \mu_0 = 22$ cm in $H_1 : \mu \neq 22$ cm (tabela 9.3)

$z = 0,4845$ in $p = 0,6280$ ♦

Tabela 9.3: Izračuni za z -preizkus za preverjanje ničelne domneve o povprečni debelini dreves; izpisi formul (levo); izpis rezultatov (desno)

	A	B		A	B
13	približek za povprečje	22,13	13	približek za povprečje	22,13
14	sigma	6	14	sigma	6,00
15	n	500	15	n	500
16	alfa	0,01	16	alfa	0,01
17	$H_0: \mu = \mu_0$	22	17	$H_0: \mu = \mu_0$	22
18	test(enostranski, dvostranski)	2	18	test(enostranski, dvostranski)	2
19	z števec	=B13-B17	19	z števec	0,1300
20	z imenovalec	=B14/SQRT(B15)	20	z imenovalec	0,2683
21	z	=B19/B20	21	z	0,4845
22	z mejna	=NORMINV(1-B16/2;0;1)	22	z mejna	2,5758
23	p-vrednost	=B18*NORMDIST(-ABS(B21);0;1;TRUE)	23	p-vrednost	0,6280

DEBELINA DEBEL 9

Isto ničelno domnevo preverite ob predpostavki, da sigma ni znana (izračun približka za standardni odklon je prikazan v zgledu 4.3).

Rešitev: $t = 0,4614$ in $p = 0,6447$ (tabela 9.4). ♦

Tabela 9.4: Izračuni za t-preizkus za preverjanje ničelne domneve o povprečni debelini dreves; izpis formul (levo); izpis rezultatov (desno)

	F	G		F	G
13	približek za povprečje	22,13	13	približek za povprečje	22,13
14	približek za standardni odklon	6,3	14	približek za standardni odklon	6,30
15	n	500	15	n	500
16	alfa	0,01	16	alfa	0,01
17	H ₀ : m _i = m ₀	22	17	H ₀ : m _i = m ₀	22
18	test(enostranski, dvostranski)	2	18	test(enostranski, dvostranski)	2
19	t števec	=G13-G17	19	t števec	0,1300
20	t imenovalec	=G14/SQRT(G15)	20	t imenovalec	0,2818
21	t	=G19/G20	21	t	0,4614
22	t mejna	=TINV(B16;G15-1)	22	t mejna	2,5857
23	p-vrednost	=TDIST(ABS(G21);G15-1;G18)	23	p-vrednost	0,6447

Domača naloga 9.3: Debelina debel dreves - ponovitev snovi

DEBELINA DEBEL 10

Predpostavljamo, da je debelina debla slučajna spremenljivka, porazdeljena po normalni porazdelitvi s parametroma $\mu = 22$ cm in $\sigma = 6$ cm. Narišite gostoto verjetnosti te normalne porazdelitve. ♦

Domača naloga 9.4: Starost gledalcev

Pri stopnji značilnosti 0,05 preverite domnevo, da je povprečna starost gledalcev določene televizijske oddaje 35 let. V vzorec je bilo vključenih 150 slučajno izbranih gledalcev. Podatki o njihovi starosti (dopolnjena leta) so v frekvenčni porazdelitvi:

Tabela 9.5: Število gledalcev po starosti, ki je izražena v dopoljenih letih

Starost	Število
20-29	20
30-39	75
40-49	45
50 in več	10

STAROST GLEDALCEV 1

Podatke iz preglednice 9.5 vpišite v datoteko DOMNEVE_EN VZOREC.XLS. Delovni list poimenujte »starost gledalcev«. ♦

STAROST GLEDALCEV 2

Izračunajte približek za povprečje in za standardni odklon za starost gledalcev.

Rešitev: $\bar{x} = 38,0$ let $s = 7,84$ let, če so vsi razredi enako široki. ♦

STAROST GLEDALCEV 3

Preverite ničelno domnevo, da je povprečna starost gledalcev 35 let (stopnja značilnosti je 0,05). Izračunajte p -vrednost in obrazložite rezultate.

Rešitev: $t = 4,6887$ in $p = 0,0000$ ♦

9.2 Bernoullijeva verjetnost

Če so izpolnjeni pogoji za aproksimacijo binomske porazdelitve $b(n, p)$ z normalno $N(np, \sqrt{np(1-p)})$, ničelno domnevo $H_0: p = p_0$ testiramo s testno statistiko

$z = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$. Njena ničelna porazdelitev je normalna $N(0, 1)$.

Zgled 9.4: Učinkovitost dveh vrst zdravil

Proučevali so učinkovitost dveh vrst zdravil proti bakterijskemu vnetju. Na voljo so imeli 300 bolnikov z laboratorijsko potrjenim bakterijskim vnetjem. S slučajno izbiro so bolnike razdelili na polovico; 150 bolnikom so dali zdravilo A in 150 bolnikom zdravilo B. Stanje se je izboljšalo 110 bolnikom, ki so prejeli zdravilo A in 130 bolnikom, ki so prejeli zdravilo B.

UČINKOVITOST ZDRAVIL 1

Za vsako skupino bolnikov izračunajte točkovno oceno za verjetnost, da je zdravilo učinkovito. Izračune naredite z računalom in v Excelu (tabela 9.6).

Rešitev: $\hat{p}_A = 0,73$ in $\hat{p}_B = 0,87$ ♦

UČINKOVITOST ZDRAVIL 2

Standardno zdravilo proti bakterijskemu vnetju S ima učinkovitost 0,80. Za zdravilo A trdijo, da je boljše od standardnega zdravila S. Napišite ničelno in alternativno domnevo in na osnovi podatkov preverite to domnevo ($\alpha = 0,05$) (enostranska alternativna domneva).

Izračunajte p -vrednost ter jo obrazložite (tabela 9.6).

Rešitev:

$$H_0 : p = p_0 = 0,80 \text{ in } H_1 : p > 0,80$$

$$z = -2,041 \text{ in } p = 0,9794$$

Ničelno domnevo obdržimo ($p = 0,9794$). Rezultati niso statistično značilni. Podatki nakazujejo, da je učinkovitost zdravila A enaka učinkovitosti standardnega zdravila S. ♦

UČINKOVITOST ZDRAVIL 3

Standardno zdravilo proti bakterijskemu vnetju S ima učinkovitost 0,80. Tudi za zdravilo B trdijo, da je boljše od standardnega zdravila S. Napišite ničelno in alternativno domnevo in na osnovi podatkov preverite to domnevo ($\alpha = 0,05$)

Izračunajte p -vrednost ter jo obrazložite (tabela 9.6).

Rešitev: $z = 2,041$ in $p = 0,0206$

Ničelno domnevo zavrnemo ($p = 0,0206$). Rezultati so statistično značilni. S tveganjem, manjšim od $\alpha = 0,05$, trdimo, da je zdravilo B učinkovitejše od standardnega zdravila. ♦

Tabela 9.6: Izračun verjetnosti, da zdravili A in B učinkujeta, preverjanje ničelne domneve ob enostranski alternativni domnevi ter izračun 95 % intervala zaupanja za verjetnost; izpis formul (zgoraj); izpis rezultatov (spodaj)

	A	B		A	B	C
2		Zdravilo	2		Zdravilo	
3		A	3		A	B
4	n	150	4	n	150	150
5	x	110	5	x	110	130
6	p*	=B5/B4	6	p*	0,73	0,87
7	1-p*	=1-B6	7	1-p*	0,27	0,13
8	alfa	0,05	8	alfa	0,05	0,05
9	H ₀ : p=p ₀	0,8	9	H ₀ : p=p ₀	0,8	0,8
10	H ₁ : p>p ₀		10	H ₁ : p>p ₀		
11	test (enostranski, dvostranski)	1	11	test (enostranski, dvostranski)	1	1
12	z števec	=B5-B4*B9	12	z števec	-10,000	10,000
13	z imenovalec	=SQRT(B4*B9*(1-B9))	13	z imenovalec	4,899	4,899
14	z-statistika	=B12/B13	14	z-statistika	-2,041	2,041
15	Zmejna	=NORMINV((1-B8/B11);0;1)	15	Zmejna	1,645	1,645
16	p-vrednost	=B11*(1-NORMDIST(B14;0;1;TRUE))	16	p-vrednost	0,9794	0,0206
17	95 % IZ		17	95 % IZ		
18	Z _{alfa/2}	=NORMINV(1-B8/2;0;1)	18	Z _{alfa/2}	1,960	1,960
19	d	=B18*SQRT(B6*(1-B6)/B4)	19	d	0,071	0,054
20	l ₁	=B6-B19	20	l ₁	0,66	0,81
21	l ₂	=B6+B19	21	l ₂	0,80	0,92

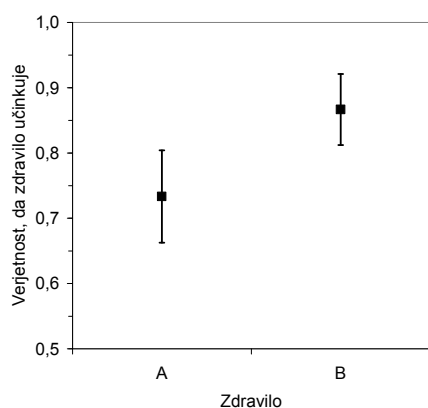
Domača naloga 9.5: Učinkovitost zdravil - ponovitev snovi

UČINKOVITOST ZDRAVIL 4

Opišite postopek slučajne izbire, po katerem so izmed 300 bolnikov izbrali 150 za zdravilo A, preostalim 150-im pa zdravilo B. ♦

UČINKOVITOST ZDRAVIL 5

Za vsako zdravilo posebej izračunajte 95 % interval zaupanja za verjetnost, da je zdravilo učinkovito (tabela 9.6). Oba intervala grafično prikažite na isti sliki (slika 9.2). Obrazložite rezultate. ♦



Slika 9.2: 95 % intervala zaupanja za verjetnost, da zdravilo učinkuje za zdravili A in B

10 Preverjanje statističnih domnev za dva vzorca in interval zaupanja za razliko

Novi pojmi:

- ničelna domneva $H_0 : \mu_1 = \mu_2$ za dva neodvisna in za dva odvisna vzorca
- ničelna domneva $H_0 : p_1 = p_2$ za dva neodvisna in za dva odvisna vzorca
- interval zaupanja za razliko $\mu_1 - \mu_2$ za dva neodvisna in za dva odvisna vzorca
- interval zaupanja za razliko $p_1 - p_2$ za dva neodvisna in za dva odvisna vzorca

Datoteke s podatki:

DOMNEVE_DVA VZORCA.XLS
ANKETA.XLS

Ustvarjene datoteke:

DOMNEVE_DVA VZORCA.DOC

Odprite novo wordovo datoteko DOMENE_DVA VZORCA.DOC. Vanjo kopirajte vse izračune in grafične prikaze ter napišite obrazložitve rezultatov nalog v tem poglavju. ♦

10.1 Razlika povprečij za neodvisna vzorca

Imamo dva neodvisna vzorca, v katerih analiziramo slučajno spremenljivko X , ki je v prvi populaciji porazdeljena $X_1 \sim N(\mu_1, \sigma)$, v drugi populaciji pa $X_2 \sim N(\mu_2, \sigma)$.

Preverjamo ničelno domnevo $H_0 : \delta = \mu_2 - \mu_1 = \delta_0$. Testna statistika je

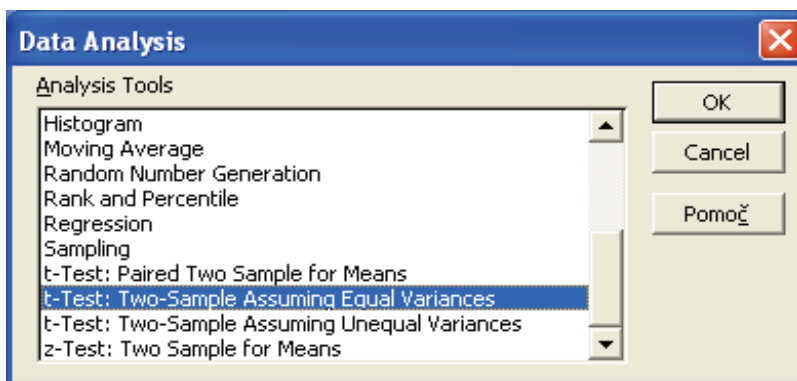
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_{sk}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

njena ničelna porazdelitev je Studentova porazdelitev: $t \sim t(SP = n_1 + n_2 - 2)$. V

t -statistiki je s_{sk}^2 skupna varianca, ki se izračuna $s_{sk}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.

Opozorjamo, da sta varianci obeh normalnih porazdelitev enaki.²⁴

Za preverjanje ničelne domneve $H_0: \delta = \mu_1 - \mu_2 = \delta_0$ ob pogoju, da sta vzorca neodvisna in ob pogoju, da lahko predpostavimo, da sta varianci populacij enaki, v Excelu uporabimo ukaz **Orodja/Analiza podatkov/t-Test: Two-Sample Assuming Equal Variances**. Če pa predpostavke o enakosti varianc ne moremo privzeti, uporabimo **t-Test: Two-Sample Assuming Unequal Variances**.



Pri obeh omenjenih ukazih se odpre pogovorno okno z enakimi polji, razlika je samo v naslovu okna. V polje **Variable 1 Range** vpišemo ime skupine celic s podatki prvega vzorca. Podatke lahko označimo tudi z miško. V polje **Variable 2 Range** vpišemo ime skupine celic s podatki drugega vzorca. V polje **Hypothesized Mean Difference** vpišemo vrednost razlike $\mu_1 - \mu_2 = \delta_0$. Možnost **Labels** izberemo, če je v prvi vrstici skupine celic s podatki ime pripadajočega vzorca. V polje **Alpha** vpišemo vrednost za

²⁴ Za preverjanje predpostavke o enakosti varianc v obeh populacijah lahko uporabimo ohlapno »pravilo palca« na vzorčnih podatkih. Če velja: $\frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)} < 2$, je uporaba zgornjega t -preizkusa vedno dopustna.

stopnjo značilnosti α . Pri osnovni nastavitvi je $\alpha = 0,05$. V poljih razdelka *Output options* določimo, kam naj se izpiše preglednica z rezultati (tabela 10.2).

Zgled 10.1: Višina rastlin tobaka

Strokovnjaki so ugotavljali, ali okuženost tobakovih rastlin z virusom vpliva na njihovo višino. Vzorčili so po 10 okuženih in neokuženih rastlin in izmerili njihovo višino. Podatki so v tabeli 10.1 (datoteka DOMNEVE_DVA VZORCA.XLS, delovni list »tobak«).

Tabela 10.1: Višine okuženih in neokuženih rastlin tobaka

Višina rastlin tobaka (cm)	
okužene rastline	neokužene rastline
66	99
67	77
64	88
77	91
86	86
72	83
93	99
80	93
83	90
74	101

TOBAK 1

Preverite ničelno domnevo, da je povprečna višina rastlin tobaka enaka za okužene in neokužene rastline ($\alpha = 0,05$).

Navodilo: najprej napišemo ničelno in alternativno domnevo.

$H_0 : \mu_{Ne} - \mu_{Ok} = 0$ Povprečna višina rastlin tobaka je enaka za okužene in neokužene rastline.

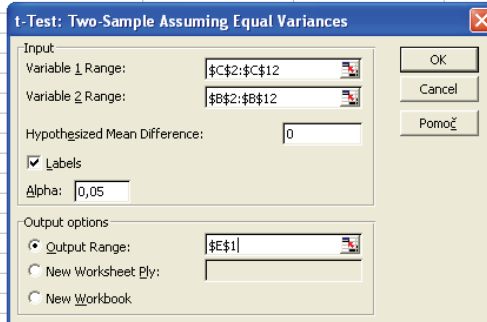
$H_1 : \mu_{Ne} - \mu_{Ok} \neq 0$ Povprečna višina rastlin tobaka za okužene in neokužene rastline ni enaka.

Nato izračunamo vzorčno povprečje in vzorčno varianco za vsak vzorec posebej, skupno varianco ter vzorčno t -statistiko. Izračunamo še mejno vrednost t -statistike in p -vrednost.

Izračune najprej naredite s formulami v pripravljene tabeli na delovnem listu »tobak« v datoteki DOMNEVE_DVA VZORCA.XLS, potem pa uporabite omenjeno excelovo orodje **Orodja/Analiza podatkov/t-Test: Two-Sample Assuming Equal Variances**. Izpis rezultatov

preverjanja naj bo na istem delovnem listu, kot so podatki. Rezultati in razlaga posameznih vrednosti so v preglednici 10.2. ♦

	A	B	C	D	E	F
1		Višina rastlin tobaka (cm)				
2		Okužene	Neokužene			
3		66	99			
4		67	77			
5		64	88			
6		77	91			
7		86	86			
8		72	83			
9		93	99			
10		80	93			
11		83	90			
12		74	101			
13						
14	n	=COUNT(B3:B12)	=COUNT(C3:C12)			
15	SP	=B14-1	=C14-1			
16	vzorčno povprečje	=AVERAGE(B3:B12)	=AVERAGE(C3:C12)			
17	vzorčna varianca	=VAR(B3:B12)	=VAR(C3:C12)			
18	H ₀ : delta=mi ₁ -mi ₂ =delta ₀	0				
19	skupna varianca	=(B15*B17+C15*C17)/(B15+C15)				
20	SP za ničelno porazdelitev	=B15+C15				
21	števec t	=(C16-B16)-B18				
22	imenovalec t	=SQRT(B19*(1/B14+1/C14))				
23	t-statistika	=B21/B22				
24	alfa	0,05				
25	test(eno, dvostranski)	2				
26	mejna vrednost t-statistike	=TINV(B24*2/B25,B20)				
27	p-vrednost	=TDIST(ABS(B23),B20,B25)				



TOBAK 2

Napišite statistični in vsebinski sklep.

Rešitev:

$$t = 3,7780 \text{ in } p = 0,0014$$

Statistični sklep: H₀ zavrnemo pri $\alpha = 0,05$. Rezultati so statistično značilni.

Vsebinski sklep: s tveganjem, manjšim od 0,05, trdimo, da povprečni višini okuženih in neokuženih rastlin tobaka nista enaki ($p = 0,0014$). ♦

Tabela 10.2: Rezultat ukaza **Orodja/Analiza podatkov/t-Test: Two-Sample Assuming Equal Variances** za podatke v tabeli 10.1.

	<i>Neokužene</i>	<i>Okužene</i>	Razlaga izračunov
Mean	90,7	76,2	vzorčni aritmetični sredini \bar{x}_{Ne} in \bar{x}_{Ok}
Variance	58,46	88,84	vzorčni varianci s_{Ne}^2 in s_{Ok}^2
Observations	10	10	število enot v vzorcu
Pooled Variance	73,65		skupna varianca s_{sk}^2
Hypothesized Mean Difference	0		$H_0 : \delta = \mu_{Ne} - \mu_{Ok} = 0$
df	18		stopinje prostosti SP
t Stat	3,7780		izračunana t statistika
P(T<=t) one-tail	0,0007		p -vrednost (enostransko)
t Critical one-tail	1,7341		$t_{\alpha}(SP)$
P(T<=t) two-tail	0,0014		p -vrednost (dvostransko)
t Critical two-tail	2,1009		$t_{\alpha/2}(SP)$

TOBAK 3

Iz predhodnih študij je znano, da okuženost z virusom zavira rast rastlin tobaka. Preverite H_0 iz prejšnje naloge ob predpostavki, da je alternativna domneva enostranska (tabela 10.2). Zapišite ničelno in alternativno domnevo. Izvedite preizkus, izračunajte p -vrednost in obrazložite rezultate.

Navodilo: najprej napišemo ničelno in alternativno domnevo.

$H_0 : \mu_{Ne} - \mu_{Ok} = 0$ Povprečna višina rastlin tobaka je enaka za okužene in neokužene rastline.

$H_1 : \mu_{Ne} - \mu_{Ok} > 0$ Povprečna višina rastlin tobaka za neokužene je večja od povprečne višine rastlin za neokužene rastline.

Rešitev:

$$t = 3,7780 \text{ in } p = 0,0007$$

Statistični sklep: H_0 zavrnemo pri $\alpha = 0,05$. Rezultati so statistično značilni.

Vsebinski sklep: s tveganjem, manjšim od 0,05, trdimo, da je povprečna višina okuženih rastlin manjša od povprečne višine neokuženih rastlin tobaka ($p = 0,0007$). ♦

TOBAK 4

Preverite domnevo, da je povprečna višina neokuženih rastlin tobaka višja od povprečne višine okuženih rastlin za več kot 7 cm ($\alpha = 0,05$). Zapišite ničelno in alternativno domnevo ter statistični in vsebinski sklep.

Navodilo: uporabite excelovo orodje.

$H_0 : \mu_{Ne} - \mu_{Ok} = 7$ Razlika med povprečno višino neokuženih in okuženih rastlin tobaka je 7 cm.

$H_1 : \mu_{Ne} - \mu_{Ok} > 7$ Razlika med povprečno višino neokuženih in okuženih rastlin tobaka je večja od 7 cm.

	E	F	G
t-Test: Two-Sample Assuming Equal Variances			
		<i>Neokužene</i>	<i>Okužene</i>
Mean		90,7	76,2
Variance		58,46	88,84
Observations		10	10
Pooled Variance		73,65	
Hypothesized Mean Difference		7	
df		18	
t Stat		1,9542	
P(T<=t) one-tail		0,0332	
t Critical one-tail		1,7341	
P(T<=t) two-tail		0,0664	
t Critical two-tail		2,1009	

Rešitev:

$$t = 1,9542 \text{ in } p = 0,0332$$

Statistični sklep: H_0 zavrnamo pri $\alpha = 0,05$. Rezultati so statistično značilni.

Vsebinski sklep: s tveganjem, manjšim od 0,05, trdimo, da je razlika med povprečno višino neokuženih in okuženih rastlin tobaka večja od 7 cm ($p = 0,0332$). ♦

TOBAK 5

Kakšen bi bil statističen in kakšen vsebinski sklep, če bi predpostavili stopnjo značilnosti $\alpha = 0,01$? ♦

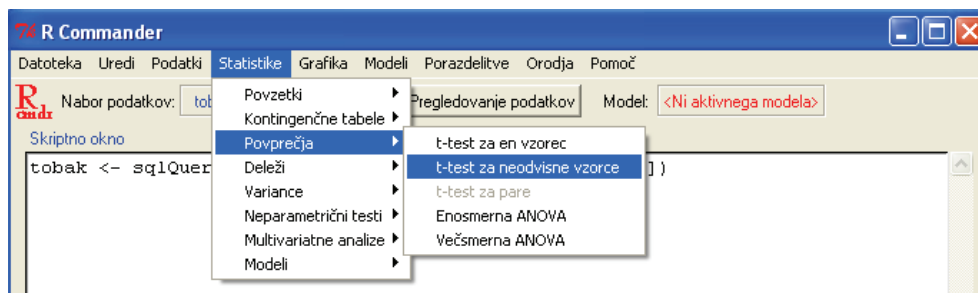
TOBAK 6

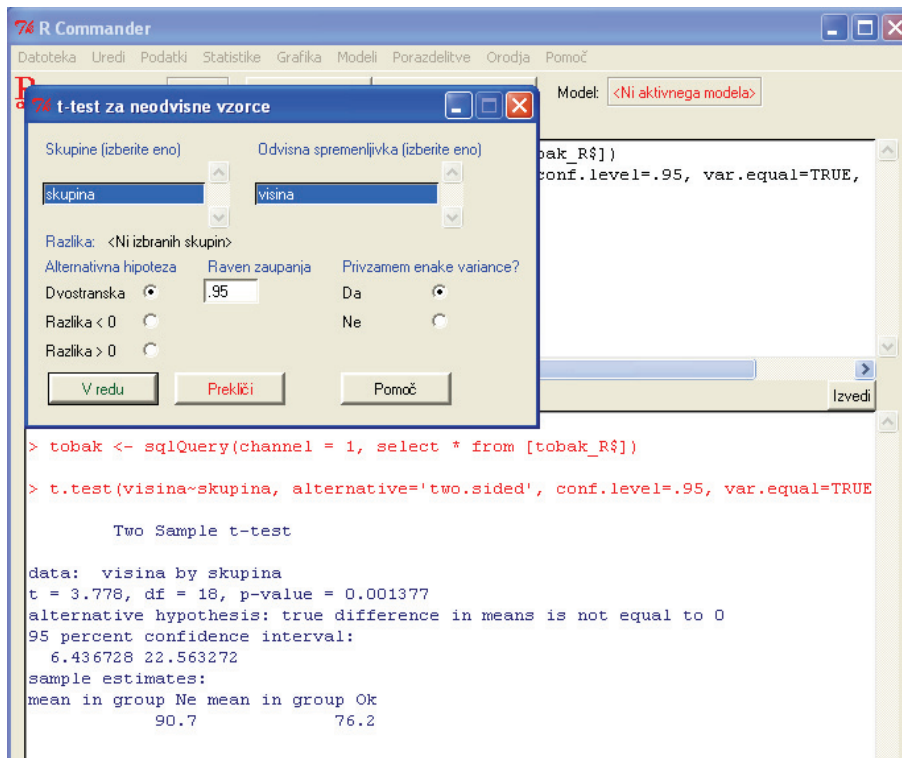
Preverite domnevo o enaki povprečni višini za neokužene in okužene rastline še s programom R.

Navodilo: pri uvozu podatkov v R-ov nabor podatkov bodite pozorni na obliko podatkov v excelovi preglednici. Če uvozite excelov delovni list »tobak«, R ne ponudi možnosti *t-test za neodvisne vzorce*. Podatke z delovnega lista »tobak« najprej kopirajte na nov delovni list z imenom »tobak_R« in jih tam preuredite v dva stolpca; v prvem stolpcu so oznake skupine (Ok - okužene, Ne - neokužene), v drugem pa vrednosti za višino:

	A	B	C
1	skupina	visina	
2	Ok	66	
3	Ok	67	
4	Ok	64	
5	Ok	77	
6	Ok	86	
7	Ok	72	
8	Ok	93	
9	Ok	80	
10	Ok	83	
11	Ok	74	
12	Ne	99	
13	Ne	77	
14	Ne	88	
15	Ne	91	
16	Ne	86	
17	Ne	83	
18	Ne	99	
19	Ne	93	
20	Ne	90	
21	Ne	101	
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			

V R-u uporabite ukaz **Statistike/Povprečja/t-test za neodvisne vzorce**. Bodite pozorni na izbiro možnosti pri vprašanju Privzamem enake variance?.





TOBAK 7

V programu R preverite še domnevo, da je povprečna višina neokuženih rastlin tobaka višja od povprečne višine okuženih rastlin za več kot 7 cm ($\alpha = 0,05$).

Navodilo: za preverjanje $H_0 : \mu_{Ne} - \mu_{Ok} = \delta_0 = 7$ cm ob alternativni domnevi $H_1 : \mu_{Ne} - \mu_{Ok} > 7$ cm moramo ukaz v skriptnem oknu dopolniti (ne gre samo z izpolnjevanjem pogovornega okna v R Commanderju):

alternativna domneva je enostranska (alternative='greater'), in sicer pričakujemo, da je razlika med neokuženimi in okuženimi rastlinami tobaka več kot 7 cm; dodati moramo izraz mu=7, ki povzema zgornjo ničelno domnevo.

```
t.test(visina~skupina, alternative='greater', mu=7, conf.level=.95, var.equal=TRUE, data=tobak)
```

Opomba: R pri postavljanju ničelne domneve deluje »po abecedi«: ker je »N« pred »O«, je ničelna domneva $H_0 : \mu_{Ne} - \mu_{Ok} = \delta_0 = 7$ cm. Če bi bili oznaki skupin v naboru podatkov namesto »Ok« »tobak1« in namesto »Ne« »tobak2«, bi R postavil ničelno domnevo takole: $H_0 : \mu_{tobak1} - \mu_{tobak2} = \delta_0$. Vrednost za δ_0 je v tem primeru -7.

Poleg preverjanja ničelne domneve program R izračuna tudi interval zaupnja za razliko povprečij. Raven zaupanja določimo v pogovornem oknu ukaza **Statistike/Povprečja/t-test za neodvisne vzorce**. ♦

```
Okno za izpis
> t.test(visina~skupina, alternative='greater', mu=7, conf.level=.95, var.equal=TRUE, data=tobak)

Two Sample t-test

data: visina by skupina
t = 1.9542, df = 18, p-value = 0.0332
alternative hypothesis: true difference in means is greater than 7
95 percent confidence interval:
 7.844719      Inf
sample estimates:
mean in group Ne mean in group Ok
          90.7          76.2
```

$(1 - \alpha)100\%$ interval zaupanja za $\delta = \mu_1 - \mu_2$, če lahko predpostavimo enakost varianc, izračunamo z izrazom

$$l_{1,2} = (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(SP = n_1 + n_2 - 2) \sqrt{s_{sk}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}.$$

TOBAK 8

Izračunajte 95 % interval zaupanja za razliko povprečij višine okuženih in neokuženih rastlin tobaka. Obrazložite rezultate.

Navodila: v Excelu z vrednostmi in formulami izpolnite pripravljeno tabelo na delovnem listu »tobak« (tabela 10.3).

Tabela 10.3: Izračuni za interval zaupanja za razliko povprečij dveh neodvisnih vzorcev; izpis formul (levo); izpis rezultatov (desno)

	A	B
30	zaupanje	0,95
31	alfa	=1-B30
32	SP	=B20
33	$t_{\alpha/2}(SP=n_1+n_2-2)$	=TINV(B31;B32)
34	d	=B33*B22
35	$\bar{x}_{Ne} - \bar{x}_{Ok}$	=C16-B16
36	l_1	=B35-B34
37	l_2	=B35+B34

	A	B
30	zaupanje	0,95
31	alfa	0,05
32	SP	18
33	$t_{\alpha/2}(SP=n_1+n_2-2)$	2,1009
34	d	8,06
35	$\bar{x}_{Ne} - \bar{x}_{Ok}$	14,50
36	l_1	6,44
37	l_2	22,56

Rešitev:

Pri 95 % zaupanju interval (6,4 cm, 22, 6 cm) vsebuje vrednost za razliko povprečja višine v populaciji neokuženih in povprečja višine v populaciji okuženih rastlin tobaka. ♦

TOBAK 9

Ponovite izračun pri 99 % zaupanju.

Navodilo: ustrezno spremenite vrednost za zaupanje v tabeli z izračuni na delovnem listu »tobak«. ♦

Zgled 10.2 (nadaljevanje zgleda 9.2): Povprečna telesna višina študentk

Med študenti 2. letnika Agronomije, Biotehnologije in Mikrobiologije na BF so v študijskem letu 2006/07 izvedli anketo. Med drugim so študente vprašali o njihovi telesni višini. V datoteki DOMNEVE_DVA VZORCA.XLS na delovnem listu »telesna višina« so podatki za študentke.

TELESNA VIŠINA 3

S pomočjo orodja vrtilne tabele oblikujte tabelo 10.4. Kaj predstavljajo izračuni v stolpcu »Skupaj«? ♦

The screenshot shows an Excel spreadsheet with two columns: 'Smer študija' (Study Direction) and 'Telesna višina (cm)' (Body Height in cm). The data rows are numbered 1 to 31. Two dialog boxes are overlaid on the spreadsheet:

- Čarovnik za vrtilne tabele in vrtilne grafikone - Korak 3 od 3:** This dialog asks 'Kam želite postaviti poročilo vrtilne tabele?' (Where do you want to place the pivot table report?) and has a radio button selected for 'Na nov delovni list' (On a new worksheet).
- Čarovnik za vrtilne tabele in vrtilne grafikone - Postavitev:** This dialog shows the pivot table layout. The 'STRAN' (Rows) field is set to 'Smer študija' and the 'STOLPEC' (Columns) field is set to 'Telesna v'. The 'VRSTICA' (Columns) field is set to 'Šteje od Telesna višin', 'Povprečje od Telesni', and 'Variance od Telesna'. There are buttons for 'Pomoč', 'V redu', and 'Prekliči'.

Tabela 10.4: Število študentk, vzorčno povprečje (cm) in vzorčna varianca (cm²) za vsako smer študija posebej

Podatki	Smer študija			
	AG	BT	MB	Skupaj
Štej od Telesna višina (cm)	33	14	32	79
Povprečje od Telesna višina (cm)	168,8	167,1	168,2	168,3
Varianca od Telesna višina (cm)	40,6	26,0	33,8	34,8

TELESNA VIŠINA 4

Preverite ničelno domnevo, da je povprečna telesna višina študentk BT enaka povprečni telesni višini študentk MB ($\alpha = 0,05$). Izračunajte p -vrednost ter obrazložite rezultate.

Navodilo: najprej izračune naredite v Excelu s formulami in pri tem uporabite oceni povprečja in variance, ki ste jih izračunali z uporabo vrtilne tabele (tabela 10.4). Preden se lotite izračunov za preverjanje ničelne domneve, parkirajte vrtilno tabelo na nov delovni list z imenom »BT in MB s formulami«. Kako si lahko pomagate s tabelo na delovnem listu »tobak«?

Drugič izračune naredite z uporabo excelovega orodja. V ta namen je potrebno podatke preurediti. Na delovnem listu morajo biti vrednosti telesne višine študentk iste smeri študija (BT in MB) zapisane skupaj, ena za drugo v stolpcu. Podatke na delovnem listu »telesna višina« najhitreje preuredimo z ukazom **Podatki /Filter/Samodejni filter**, tako da v padajočem seznamu ob spremenljivki *Smer študija* izberemo ukaz **Razvrsti naraščajoče**:

	A	B	C	D
1	Smer študija	Telesna višina (cm)		
2	Razvrsti naraščajoče	165		
3	Razvrsti padajoče	157		
4	(Vse)	169		
5	(Zgornjih 10 ...)	165		
6	(Lasten ...)	170		
7	AG	166		
8	BT	162,5		
9	MB	165		
10	MB	165		
11	MB	168		
12	AG	179		
13	AG	163		

S tem ukazom se zapisi (en zapis sestavljata vrednosti dveh spremenljivk *Smer študija* in *Telesna višina*) razvrstijo po abecednem redu: najprej AG, sledijo jim BT in MB. Vrednosti telesne višine za študentke BT in MB kopirajte na nov delovni list z imenom »višina za BT in MB«, za vsako smer študija naj bodo v svojem stolpcu. Izpis ukaza **Orodja/Analiza podatkov/t-Test: Two-Sample Assuming Equal Variances** naj bo na delovnem listu »višina za BT in MB« poleg podatkov.

	A	B	C	D	E	F
1	BT	MB		t-Test: Two-Sample Assuming Equal Variances		
2	169	165				
3	174	165			BT	MB
4	167	168		Mean	167,1	168,2
5	159	163		Variance	26,0	33,8
6	167,5	165		Observations	14	32
7	161	158		Pooled Variance	31,5	
8	170	169		Hypothesized Mean Difference	0	
9	164,5	175		df	44	
10	160	172		t Stat	-0,5895	
11	167	165		P(T<=t) one-tail	0,2793	
12	176	164		t Critical one-tail	1,6802	
13	167	159		P(T<=t) two-tail	0,5585	
14	173	176		t Critical two-tail	2,0154	
15	165	165				
16		168				
17		173				
18		167				
19		174				
20		172				
21		165				
22		157				
23		168				
24		178				
25		173				
26		168				
27		170				
28		163,5				
29		164				
30		165				
31		174				

Rešitev:

$$H_0 : \mu_{BT} - \mu_{MB} = 0 \text{ cm}$$

$$H_1 : \mu_{BT} - \mu_{MB} \neq 0 \text{ cm}$$

$$t = -0,5895 \text{ in } p = 0,5585 \blacklozenge$$

TELESNA VIŠINA 5

Preverite še domnevo, da so v povprečju študentke BT za 3 cm višje od študentk MB ($\alpha = 0,05$).

Rešitev:

$$H_0 : \mu_{BT} - \mu_{MB} = 3 \text{ cm}$$

$$H_1 : \mu_{BT} - \mu_{MB} \neq 3 \text{ cm}$$

$$t = -2,2576 \text{ in } p = 0,0290 \blacklozenge$$

TELESNA VIŠINA 6

Ponovite preverjanje ničelne domneve z uporabo programa R in grafičnega vmesnika R Commander, **Statistike/Povprečja/t-test za neodvisne vzorce**. ♦

10.2 Razlika povprečij za odvisna vzorca

Imamo dva odvisna vzorca, v katerih analiziramo spremenljivko X , ki je v prvi populaciji porazdeljena $X_1 \sim N(\mu_1, \sigma_1)$, v drugi populaciji pa $X_2 \sim N(\mu_2, \sigma_2)$.

Tvorimo njuno razliko $D = X_1 - X_2$, njena porazdelitev je $D \sim N(\mu_1 - \mu_2 = \mu_D, \sigma_D)$.

Preverjamo ničelno domnevo $H_0: \mu_D = \mu_{D_0}$. Imamo n parov podatkov, za vsakega izračunamo razliko $d_i = x_{i1} - x_{i2}$. Testna statistika je

$$t = \frac{\bar{d} - \mu_{D_0}}{\frac{s_d}{\sqrt{n}}},$$

kjer je $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i2})$ aritmetična sredina razlik in $s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2}$

standardni odklon razlik. Ničelna porazdelitev testne statistike je Studentova porazdelitev $t \sim t(SP = n - 1)$.

Zgled 10.3: Plavanje

Prostovoljci so se udeležili trimesečnega tečaja plavanja. Njihove plavalne sposobnosti so preverili na začetku in na koncu tečaja, tako da so neprekinjeno plavali 12 minut v poljubni plavalni tehniki. Merili so število metrov, ki so jih plavalci preplavali v 12 minutah. Njihovi rezultati so podani v tabeli 10.5.

Tabela 10.5: Preplavana dolžina ob začetku in koncu tečaja (Vir: Košmelj, Uporabna statistika, Ljubljana 2007, str. 138)

Plavalec	Na začetku tečaja	Ob koncu tečaja
1	375	400
2	500	600
3	475	525
4	550	575
5	375	400
6	300	435
7	500	575
8	425	450
9	500	550
10	480	525
11	475	560
12	400	500
13	500	575
14	575	600
15	400	425

PLAVANJE 1

Trenerka domneva, da udeleženci takega tečaja v povprečju izboljšajo rezultat neprekinjenega plavanja na 12 minut za 50 metrov. Ali so rezultati v skladu s trditvijo trenerke?

Navodilo: nalogo rešite najprej v Excelu s formulami (pripravljena tabela v datoteki), nato uporabite še Excelovo orodje **Orodja/Analiza podatkov/t-Test: Paired Two Sample for Means** (tabela 10.6).

Rešitev:

$$H_0 : \mu_{\text{konec}} - \mu_{\text{zacetek}} = 50 \text{ m} \qquad H_1 : \mu_{\text{konec}} - \mu_{\text{zacetek}} \neq 50 \text{ m}$$

$$t = 0,8376 \text{ in } p = 0,4163$$

Statistični sklep: H_0 obdržimo ($\alpha = 0,05$).

Vsebinski sklep: rezultati so v skladu s trditvijo trenerke, da se rezultat plavanja po opravljenem plavalnem tečaju v povprečju izboljša za 50 m ($p = 0,4163$). ♦

	A	B	C	D
1	Plavalec	Na začetku tečaja	Ob koncu tečaja	Razlika, d_i
2	1	375	400	=C2-B2
3	2	500	600	=C3-B3
4	3	475	525	=C4-B4
5	4	550	575	=C5-B5
6	5	375	400	=C6-B6
7	6	300	435	=C7-B7
8	7	500	575	=C8-B8
9	8	425	450	=C9-B9
10	9	500	550	=C10-B10
11	10	480	525	=C11-B11
12	11	475	560	=C12-B12
13	12	400	500	=C13-B13
14	13	500	575	=C14-B14
15	14	575	600	=C15-B15
16	15	400	425	=C16-B16
17			n	=COUNT(D2:D16)
18			povprečna razlika	=AVERAGE(D2:D16)
19			st. odklon razlik s_d	=STDEV(D2:D16)
20			$H_0: \delta = \mu_k - \mu_z$	50
21			SP	=D17-1
22			t števec	=D18-D20
23			t imenovalec	=D19/SQRT(D17)
24			t-statistika	=D22/D23
25			test (eno, dvostranski)	2
26			alfa	0,05
27			mejna vrednost t statistike	=TINV(D26*2/D25,D21)
28			p-vrednost	=TDIST(D24,D21,D25)

t-Test: Paired Two Sample for Means

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha:

Output options

Output Range:

New Worksheet Ply:

New Workbook

OK Cancel Pomoč

Tabela 10.6 Rezultat ukaza **Orodja/Analiza podatkov/t-Test: Paired Two Sample for Means** za podatke iz tabele 10.4

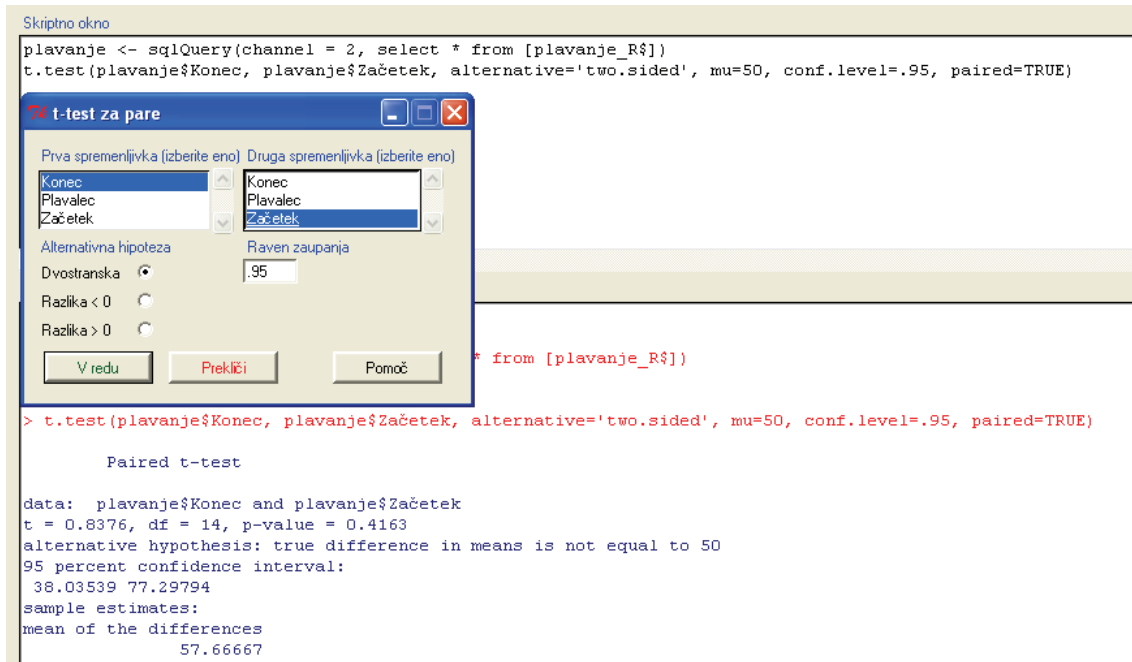
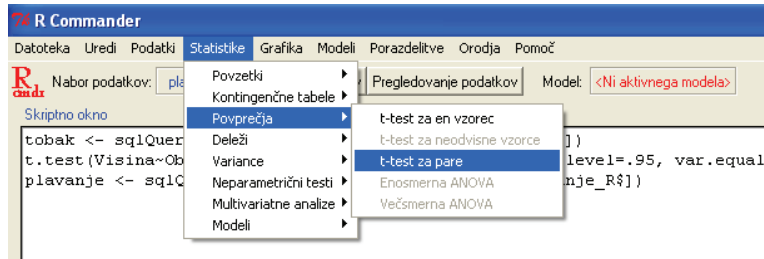
	<i>Ob koncu tečaja</i>	<i>Na začetku tečaja</i>	Razlaga izračunov
Mean	513,0	455,3	vzorčni aritmetični sredini
Variance	5288,6	5480,2	vzorčni varianci
Observations	15	15	število enot v vzorcu
Pearson Correlation	0,88		Pearsonov koeficient korelacije
Hypothesized Mean Difference	50		$H_0 : \delta = \mu_K - \mu_Z = 0$
df	14		stopinje prostosti SP
t Stat	0,8376		izračunana t vrednost
P(T<=t) one-tail	0,2082		p -vrednost (enostranska)
t Critical one-tail	1,7613		$t_{\alpha}(SP = 14)$
P(T<=t) two-tail	0,4163		p -vrednost (dvostranska)
t Critical two-tail	2,1448		$t_{\alpha/2}(SP = 14)$

PLAVANJE 2

Analizo naredite še v R-u z ukazom **Statistike/Povprečja/t-test za pare**.

Navodilo: preden podatke iz excelove datoteke uvozite v R z R Commanderjem, jih brez tabele za izračune (samo celice A1:C16) kopirajte na nov list z imenom »plavanje za R«. V R Commanderju lahko z ukazom **Statistike/Povprečja/t-test za pare** preverite le ničelno domnevo $H_0 : \delta = 0$; če je delta različna od 0, moramo to popraviti v skriptnem oknu (podobno kot pri t-preizkusu za neodvisna vzorca). V izraz moramo dodati **mu=50**:

```
t.test(plavanje$Konec, plavanje$Zacetek, alternative='two.sided', mu=50, conf.level=.95,
paired=TRUE)
```



Poleg preverjanja ničelne domneve program R izračuna tudi interval zaupnja za razliko povprečij. Raven zaupanja določimo v pogovornem oknu ukaza **Statistike/Povprečja/t-test za pare**. ♦

$(1 - \alpha)$ 100% interval zaupanja za δ izračunamo z izrazom

$$l_{1,2} = \bar{d} \pm t_{\alpha/2}(SP = n - 1) \frac{S_d}{\sqrt{n}}.$$

PLAVANJE 3

Izračunajte 95 % interval zaupanja za povprečno razliko preplavane dolžine na začetku in na koncu tečaja. Obrazložite izračunane vrednosti. Izračunajte po formulah v Excelu (pripravljena tabela na delovnem listu »plavanje«) ali z računalom.

Rešitev:

Pri 95 % zaupanju interval (38,0, 77,3) vsebuje vrednost za povprečno razliko v preplavanih metrih ob neprekinjenem plavanju 12 minut pred in po koncu plavalnega tečaja. ♦

PLAVANJE 4

Ponovite izračune pri 99 % zaupanju. ♦

Zgled 10.4: Kationska izmenjalna kapaciteta tal, določena po dveh različnih metodah

V datoteki DOMNEVE_DVA VZORCA.XLS so na delovnem listu »CEC« med drugim podatki o količini organske snovi [%] in kationski izmenjalni kapaciteti tal (CEC) [mmol_c/100 g tal] za 50 talnih vzorcev. Za vsak vzorec je CEC določena po dveh različnih metodah (AA in EFE). Pri primerjavi rezultatov, dobljenih po dveh metodah, bomo upoštevali samo vzorce tal, ki imajo vsebnost organske snovi manjšo od 20 %. Zanima nas, ali metodi AA in EFE v povprečju dajeta enake rezultate za kationsko izmenjalno kapaciteto tal (CEC).

CEC 1

V excelovi datoteki najprej izberite talne vzorce s količino organske snovi pod 20 %.

Navodilo: uporabite ukaz **Podatki/Filter**. Nato tri stolpce podatkovne zbirke: **oznaka vzorcev**, **metoda** in **CEC** izbranih talnih vzorcev kopirajte na delovni list »izbor CEC«. Podatke uredite tako, da bodo v enem stolpcu podatki za CEC po metodi AA, v sosednjem pa podatki za CEC po metodi EFE. Pazite, da bodo v vrsticah podatki za isti talni vzorec (ista oznaka vzorca). Tako ureditev podatkov najlažje naredite z uporabo vrtilne tabele:

10 Preverjanje statističnih domnev za dva vzorca, interval zaupanja za razliko

	A	B	C	D	E	F	G	H	I	J	K
1	oznaka vzorca	metoda	CEC								
2		35 EFE	0,06								
3		76 EFE	23,30								
4		78 EFE	19,52								
5		82 EFE	0,98								
6		83 EFE	6,15								
7		85 EFE	23,98								
8		87 EFE	14,33								
9		88 EFE	13,75								
10		89 EFE	15,33								
11		90 EFE	3,64								
12		91 EFE	2,61								
13		93 EFE	5,47								
14		99 EFE	2,75								
15		100 EFE	2,71								
16		101 EFE	14,85								
17		104 EFE	25,47								
18		107 EFE	18,27								
19		112 EFE	27,39								
20		113 EFE	30,12								
21		118 EFE	5,59								
22		119 EFE	5,04								
23		125 EFE	34,34								
24		129 EFE	40,12								
25		131 EFE	8,86								
26		132 EFE	5,86								
27		133 EFE	15,64								
28		137 EFE	21,06								
29		138 EFE	16,47								
30		139 EFE	5,83								
31		146 EFE	21,14								

Čarovnik za vrtilne tabele in vrtilne grafikone - Postavitev

Zgradite poročilo vrtilne tabele z vlečenjem gumbov polj na desni na diagram na levi.

STRAN

metoda

STOLPEC

oznaka vz

metoda

CEC

VRSTICA

Vsota od CEC

PODATKI

Pomoč

V redu

Prekliči

Del vrtilne tabele s podatki kopirajte na nov delovni list z imenom »CEC za R«:

	A	B	C	D	E	F	G
1	oznaka vzorca	AA	EFE				
2		35	9,71	0,06			
3		76	28,77	23,30			
4		78	20,59	19,52			
5		82	14,29	0,98			
6		83	13,66	6,15			
7		85	32,30	23,98			
8		87	20,70	14,33			
9		88	20,68	13,75			
10		89	32,90	15,33			
11		90	23,30	3,64			
12		91	20,18	2,61			
13		93	12,73	5,47			
14		99	17,60	2,75			
15		100	17,12	2,71			
16		101	23,93	14,85			
17		104	31,25	25,47			
18		107	35,67	18,27			
19		112	32,38	27,39			
20		113	32,77	30,12			
21		118	18,90	5,59			
22		119	14,68	5,04			
23		125	38,43	34,34			
24		129	58,14	40,12			
25		131	14,08	8,86			
26		132	11,73	5,86			
27		133	19,77	15,64			
28		137	31,14	21,06			
29		138	38,66	16,47			
30		139	24,02	5,83			
31		146	24,70	21,14			

\\ \tobak \ telesna višina \ plavanje_R \ plavanje \ CEC za R / izbor CEC / CEC

Koliko talnih vzorcev ste dobili? ♦

CEC 2

Kaj je statistični vzorec? Koliko jih je? Kaj je enota proučevanja? Koliko enot je v statističnem vzorcu? ♦

CEC 3

Preverite ničelno domnevo, ki pravi, da metodi AA in EFE dajeta v povprečju enake rezultate za CEC ($\alpha = 0,05$). Najprej uporabite ukaz **Orodja/Analiza podatkov/t-Test: Paired Two Sample for Means** v Excelu. Izhodno tabelo shranite na delovni list »izbor CEC«. Obrazložite rezultate.

Rešitev:

$$H_0 : \mu_{AA} - \mu_{EFE} = 0$$

$$H_1 : \mu_{AA} - \mu_{EFE} \neq 0$$

$$t = 10,3626 \text{ in } p = 0,0000$$

CEC 4

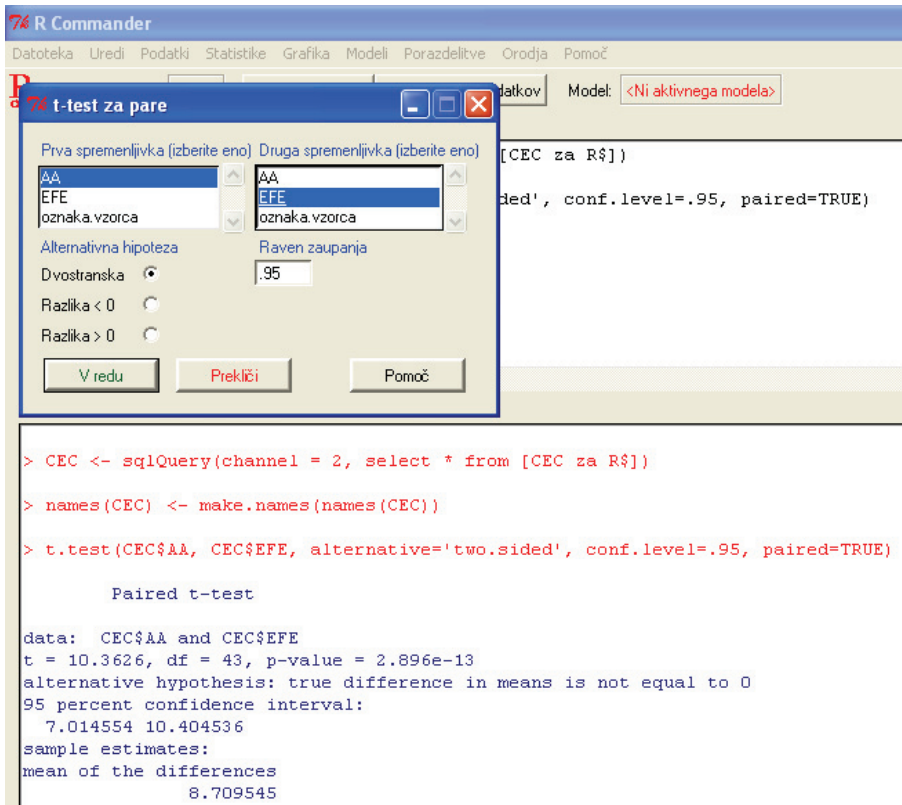
Izračunajte 95 % interval zaupanja za povprečno razliko CEC med metodama AA in EFE.

Rešitev:

Pri 95 % zaupanju interval (7,02, 10,40) vsebuje vrednost za povprečno razliko CEC (mmol_c/100 g tal), izmerjeno po metodi AA in EFE. ♦

CEC 5

Analizo naredite še v programu R.



The screenshot shows the R Commander interface. A dialog box titled "t-test za pare" is open, with "Prva spremenljivka (izberite eno)" set to "AA" and "Druga spremenljivka (izberite eno)" set to "EFE". The "Alternativna hipoteza" is "Dvostranska" and "Raven zaupanja" is ".95". The R console shows the following code and output:

```
> CEC <- sqlQuery(channel = 2, select * from [CEC za R$])
> names(CEC) <- make.names(names(CEC))
> t.test(CEC$AA, CEC$EFE, alternative='two.sided', conf.level=.95, paired=TRUE)

Paired t-test

data: CEC$AA and CEC$EFE
t = 10.3626, df = 43, p-value = 2.896e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 7.014554 10.404536
sample estimates:
mean of the differences
      8.709545
```

10.3 Razlika Bernoullijevih verjetnosti za neodvisna vzorca

Imamo dva neodvisna vzorca, v katerih analiziramo slučajno spremenljivko X , ki je v prvi populaciji porazdeljena $X_1 \sim b(n_1, p_1)$ in v drugi populaciji $X_2 \sim b(n_2, p_2)$. Aproksimacija binomske porazdelitve z normalno je utemeljena: $b(n_1, p_1) \approx N(n_1 p_1, \sqrt{n_1 p_1 (1 - p_1)})$, $b(n_2, p_2) \approx N(n_2 p_2, \sqrt{n_2 p_2 (1 - p_2)})$. Preverjamo ničelno domnevo $H_0 : p_1 = p_2 = p_{sk}$. Testna statistika je

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_{sk} \cdot (1 - \hat{p}_{sk}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

njena ničelna porazdelitev je $N(0,1)$. V izrazu za z je $\hat{p}_1 = \frac{x_1}{n_1}$, $\hat{p}_2 = \frac{x_2}{n_2}$ in

$$\hat{p}_{sk} = \frac{x_1 + x_2}{n_1 + n_2}.$$

Zgled 10.5 (nadaljevanje zgleda 9.4): Učinkovitost dveh vrst zdravil

Proučevali so učinkovitost dveh vrst zdravil proti bakterijskemu vnetju. Na voljo so imeli 300 bolnikov z laboratorijsko potrjenim bakterijskim vnetjem. S slučajno izbiro so bolnike razdelili na polovico, 150 bolnikom so dali zdravilo A in 150 bolnikom zdravilo B. Stanje se je izboljšalo 110 bolnikom, ki so prejeli zdravilo A in 130 bolnikom, ki so prejeli zdravilo B.

UČINKOVITOST ZDRAVIL 6

Preverite ničelno domnevo, da je učinkovitost zdravil A in B enaka ($\alpha = 0,05$). Obrazložite rezultate.

Navodilo: nalogo rešite s pisanjem formul na delovnem listu »učinkovitost zdravil« v datoteki DOMNEVE_DVA VZORCA.XLS (tabeli 10.7 in 10.8).

Tabela 10.7: Izpis formul za preverjanje domneve o razliki Bernoullijevih verjetnosti za neodvisna vzorca

	A	B	C
2		Zdravilo	
3		A	B
4	n	150	150
5	x	110	130
6	p*	=B5/B4	=C5/C4
7	1-p*	=1-B6	=1-C6
8			
9	H ₀ : p _A = p _B = p _{sk}		
10	H ₁ : p _A ni enako p _B		
11	test (enostranski, dvostranski)	2	
12	alfa	0,05	
13	p _{sk}	=(B5+C5)/(B4+C4)	
14	z števec	=B6-C6	
15	z imenovalec	=SQRT(B13*(1-B13)*(1/B4+1/C4))	
16	z-statistika	=B14/B15	
17	zmejna	=NORMINV(1-B12/B11;0;1)	
18	p-vrednost	=B11*(1-NORMDIST(ABS(B16);0;1;TRUE))	

Tabela 10.8: Rezultati formul iz tabele 10.7, preverjanje domneve o razliki Bernoullijevih verjetnosti za neodvisna vzorca

	A	B	C
2		Zdravilo	
3		A	B
4	n	150	150
5	x	110	130
6	p*	0,73	0,87
7	1-p*	0,27	0,13
8			
9	$H_0: p_A = p_B = p_{sk}$		
10	$H_1: p_A$ ni enako p_B		
11	test (enostranski, dvostranski)	2	
12	alfa	0,05	
13	p_{sk}	0,80	
14	z števec	-0,13	
15	z imenovalec	0,046	
16	z-statistika	-2,887	
17	z-mejna	1,960	
18	p-vrednost	0,0039	

Rešitev:

$$H_0: p_A = p_B = p_{sk} \quad H_1: p_A \neq p_B$$

$$z = -2,887 \quad p = 0,0039$$

UČINKOVITOST ZDRAVIL 7

Izračune ponovite še s pisanjem formul v programu R. ♦

Približni $(1 - \alpha)$ 100 % interval zaupanja za $p_1 - p_2$ izračunamo z izrazom

$$l_{1,2} \approx (\hat{p}_1 - \hat{p}_2) \mp z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

UČINKOVITOST ZDRAVIL 8

Izračunajte 95 % interval zaupanja za razliko učinkovitosti zdravil A in B. Obrazložite rezultate.

Navodilo: izračune naredite s pisanjem formul v pripravljeno tabelo na delovnem listu »učinkovitost zdravil« (tabela 10.9).

Tabela 10.9: Izračun intervala zaupanja za razliko Bernoullijevih verjetnosti za dva neodvisna vzorca; izpis formul (levo); rezultati (desno)

	A	B
22	zaupanje	0,95
23	alfa	=1-B22
24	$z_{\alpha/2}$	=NORMINV(1-B23/2;0;1)
25	$p^*_B - p^*_A$	=C6-B6
26	d	=B\$24*SQRT(B6*(1-B6)/B4+C6*(1-C6)/C4)
27	l_1	=B25-B26
28	l_2	=B25+B26

	A	B
22	zaupanje	0,95
23	alfa	0,05
24	$z_{\alpha/2}$	1,960
25	$p^*_B - p^*_A$	0,133
26	d	0,089
27	l_1	0,04
28	l_2	0,22

Rešitev:

S 95 % verjetnostjo trdimo, da je interval (0,04, 0,22) eden tistih, ki vsebujejo pravo vrednost za razliko učinkovitosti zdravil B in A. ♦

UČINKOVITOST ZDRAVIL 9

Ponovite izračune intervala zaupanja za razliko učinkovitosti zdravil A in B pri 99 % zaupanju. Katero vrednost v tabeli z izračuni prejšnje naloge moramo spremeniti, da dobimo ustrezen rezultat? ♦

Domača naloga 10.1: Športniki med študenti in študentkami na BF

ANKETA 32

Na podlagi podatkov v datoteki ANKETA.XLS preverite domnevo, da je delež študentk in študentov, ki se ukvarjajo s športom, enak ($\alpha = 0,05$).

Navodilo: najprej z vrtilno tabelo preštejte, koliko študentov in koliko študentk se ukvarja oz. ne ukvarja s športom. Tiste študente, ki na vprašanje niso odgovorili, izločite iz analize.

Pri izračunih si lahko pomagamo z izpolnjeno tabelo (vrednosti, formule) na delovnem listu »učinkovitost zdravil«. Katere vrednosti moramo ustrezno spremeniti? Ali je potrebno spremeniti kakšno formulo?

Rešitev:

$$z = 1,652 \quad p = 0,0985 \blacklozenge$$

Domača naloga 10.2: Štipendisti med študenti in študentkami BT in AG

ANKETA 33

Na podlagi podatkov v datoteki ANKETA.XLS preverite domnevo, da je delež štipendistov in štipendistk (ne glede na spol) pri smereh študija AG in BT enak ($\alpha = 0,05$).

Rešitev:

$$z = -2,413 \quad p = 0,0158 \blacklozenge$$

10.4 Razlika Bernoullijevih verjetnosti za odvisna vzorca

Imamo dva odvisna vzorca, v katerih analiziramo slučajno spremenljivko X , ki je v prvi populaciji porazdeljena $X_1 \sim b(n_1, p_1)$ in v drugi populaciji $X_2 \sim b(n_2, p_2)$. Aproximacija binomske porazdelitve z normalno je utemeljena. Preverjamo ničelno domnevo $H_0 : p_1 = p_2$. Testna statistika je

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{21} + f_{12}}},$$

njena ničelna porazdelitev je $N(0,1)$.²⁵ V izrazu za z sta f_{12} in f_{21} frekvenci iz frekvenčne tabele:

	$X_2 = A$	$X_2 = \bar{A}$	Skupaj
$X_1 = A$	f_{11}	f_{12}	$f_{11} + f_{12}$
$X_1 = \bar{A}$	f_{21}	f_{22}	$f_{21} + f_{22}$
Skupaj	$f_{11} + f_{21}$	$f_{12} + f_{22}$	n

²⁵ McNemarjev preizkus simetrije

Zgled 10.6: Vpliv kajenja

Zdravniki so izvedli t. i. študijo primerov s kontrolami. V vzorcu je bilo 490 parov bolnik – kontrola. Bolnik je bil izbran iz populacije moških, obolelih za rakom na pljučih. Njegova kontrola je bila izbrana iz populacije zdravih, vendar z bolnikom usklajena po spolu in po starosti. Namen študije je bil ugotavljati vpliv kajenja na pojavnost raka na pljučih. Podatki so v tabeli 10.10 in tudi v datoteki DOMNEVE_DVA VZORCA.XLS na delovnem listu »vpliv kajenja«.

Tabela 10.10: Podatki o pojavu raka na pljučih pri kadilcih in nekadilcih za vzorca velikosti 490 (Vir: Košmelj, Uporabna statistika, 2007, str. 143)

	Bolnik kadi	Bolnik ne kadi	Skupaj
Kontrola kadi	20	29	49
Kontrola ne kadi	431	10	441
Skupaj	451	39	490

VPLIV KAJENJA 1

Preverite ničelno domnevo, da je delež kadilcev v populaciji bolnikov enak deležu kadilcev v populaciji zdravih ($\alpha = 0,01$). Obrazložite rezultate.

Navodilo: nalogo rešite s formulami v pripravljene tabeli na delovnem listu »vpliv kajenja« v datoteki DOMNEVE_DVA VZORCA.XLS (tabela 10.11).

Tabela 10.11: Izračuni za preverjanje ničelne domneve o razliki Bernoullijevih verjetnosti za dva odvisna vzorca; izpis formul (zgoraj); rezultati (spodaj)

	A	B	C	D
1		Bolnik kadi	Bolnik ne kadi	Skupaj
2	Kontrola kadi	20	29	49
3	Kontrola ne kadi	431	10	441
4	Skupaj	451	39	490
5				
6	$H_0: p_{\text{kadi zdravi}} = p_{\text{kadi bolni}}$			
7	test(enostranski, dvostranski)	2		
8	alfa	0,01		
9	z_števec	=C2-B3		
10	z_imenovalec	=SQRT(B3+C2)		
11	z-statistika	=B9/B10		
12	zmejna	=NORMINV(1-B8/2;0;1)		
13	p-vrednost	=B7*NORMDIST((1-ABS(B11));0;1;TRUE)		

10 Preverjanje statističnih domnev za dva vzorca, interval zaupanja za razliko

	A	B
6	H ₀ : p _{kadi zdravi} = p _{kadi bolni}	
7	test(enostranski,	2
8	alfa	0,01
9	z_števec	-402,00
10	z_imenovalec	21,45
11	z-statistika	-18,74
12	z.mejna	2,5758
13	p-vrednost	0,0000

Rešitev:

$$z = -18,74 \quad p = 0,0000 \blacklozenge$$

Približni $(1 - \alpha)$ 100% interval zaupanja za $p_1 - p_2$ izračunamo z izrazom:

$$l_{1,2} \approx \left(\frac{f_{12} - f_{21}}{n} \right) \pm z_{\alpha/2} \sqrt{\frac{f_{12} + f_{21}}{n^2}}$$

VPLIV KAJENJA 2

Izračunajte 95 % interval zaupanja za razliko deležev kadilcev pri bolnikih in zdravih (tabela 10.12). \blacklozenge

Tabela 10.12: Izračuni intervala zaupanja za razliko Bernoullijevih verjetnosti za dva odvisna vzorca; izpis formul (levo); rezultati (desno)

	A	B
16	zaupanje	0,95
17	alfa	=1-B16
18	z _{alfa/2}	=NORMINV(1-B17/2,0,1)
19	(f ₁₂ -f ₂₁)/n	=(B3-C2)/D4
20	d	=B18*SQRT((C2+B3)/D4^2)
21	l ₁	=B19-B20
22	l ₂	=B19+B20

	A	B
16	zaupanje	0,95
17	alfa	0,05
18	z _{alfa/2}	1,960
19	(f ₁₂ -f ₂₁)/n	0,820
20	d	0,086
21	l ₁	0,73
22	l ₂	0,91

11 Linearna regresija in korelacija

Novi pojmi:

- linearna regresija
- Pearsonov koeficient korelacije
- Spearmanov koeficient korelacije

Datoteke s podatki:

REGRESIJSKA ANALIZA.XLS
ANKETA.XLS
KORELACIJA.XLS

Ustvarjene datoteke:

REGRESIJA_KORELACIJA.DOC

Odprite novo wordovo datoteko REGRESIJA_KORELACIJA.DOC in vanjo kopirajte vse izračune, grafične prikaze in napišite obrazložitve izračunanih vrednosti za naloge, ki sledijo. ♦

11.1 Linearna regresija

Model linearne regresije je

$$Y = \alpha + \beta X + \varepsilon$$

Na podlagi vzorčnih vrednosti odvisne ($y_i, i = 1, \dots, n$) in neodvisne spremenljivke ($x_i, i = 1, \dots, n$) po metodi najmanjših kvadratov izračunamo oceno parametrov regresijskega modela a in b .

Zgled 11.1: Telesna višina in starost otrok

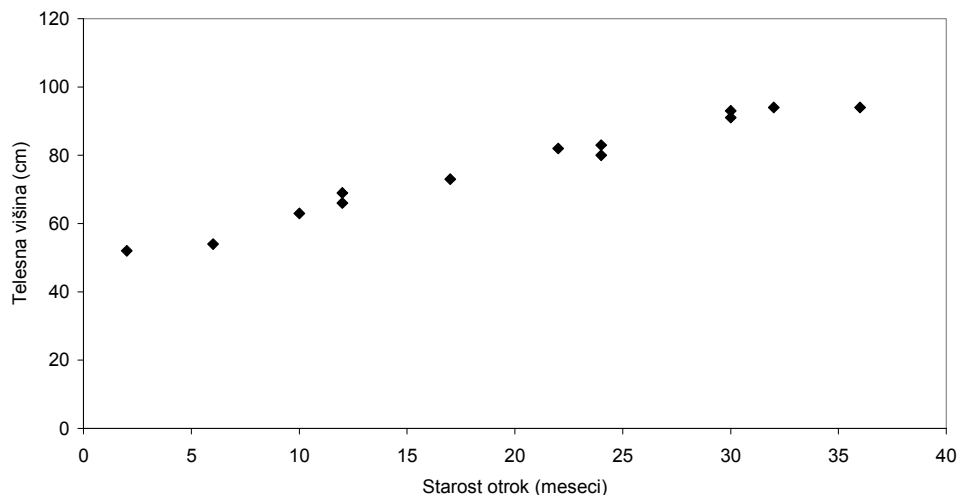
V tabeli 11.1 (datoteka REGRESIJSKA ANALIZA.XLS, delovni list »otroci«) so podatki o starosti (v mesecih) in telesni višini (cm) za 13 otrok.

OTROCI 1

Narišite sliko, ki prikazuje odvisnost telesne višine od starosti otrok (slika 11.1). ♦

Tabela 11.1: Podatki o starosti (v mesecih) in telesni višini (v cm) za 13 otrok

otrok	starost	višina
1	2	52
2	6	54
3	10	63
4	12	66
5	12	69
6	17	73
7	22	82
8	24	83
9	24	80
10	30	91
11	30	93
12	36	94
13	32	94



Slika 11.1: Odvisnost telesne višine od starosti otrok

OTROCI 2

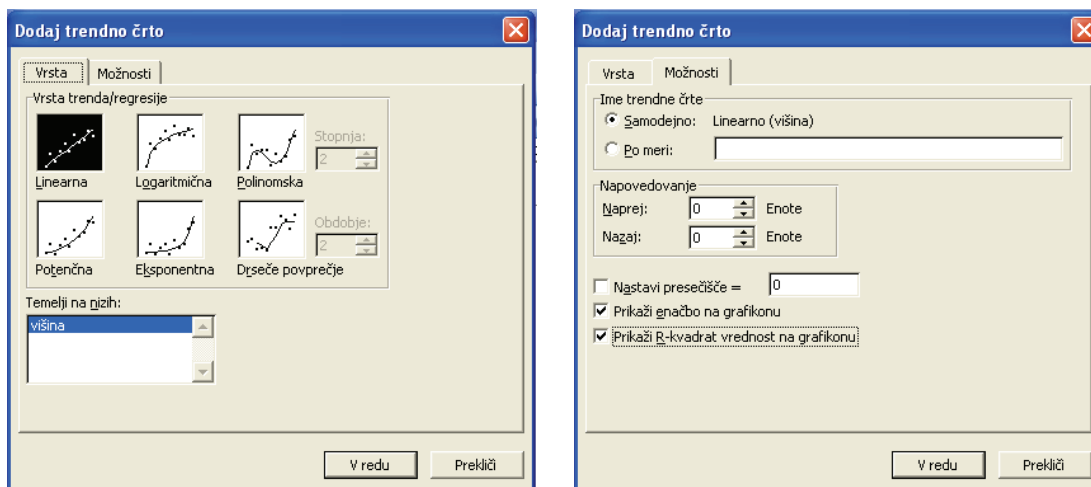
Regresijsko premico dodajte na sliko 11.1. Izračunajte oceni parametrov enostavnega linearnega regresijskega modela in ju obrazložite. Izračunajte in obrazložite koeficient determinacije.

Navodilo: vsaka premica je določena z natanko dvema točkama, ki ležita na njej. Če rišemo premico »peš«, najprej narišemo dve točki in nato skozi njiju potegnemo premico. Dve točki, za kateri vemo, da ležita na premici in zanju ne potrebujemo veliko računanja, sta $A(\bar{x}, \bar{y})$ in $B(0, a)$. Če so podatki taki, da točke $B(0, a)$ ni na sliki, izberemo smiselno vrednost x_0 in zanju izračunamo vrednost y_0 , tako da točka $C(x_0, y_0)$ leži na premici.

Če želimo v Excelu na razsevni grafikon (vrsta grafikona *XY (Raztreseni)*) dodati premico za izbrani niz podatkov, ga najprej označimo z levim klikom miške, nato z desnim klikom odpremo seznam ukazov, ki so v dani situaciji na voljo in izberemo ukaz **Dodaj trendno črto**. V pogovornem oknu med *Vrsta trenda/regresije* izberemo možnost **Linearna**. V polju *Temelji na nizih* je izbran niz z imenom »višina« in to je hkrati edini niz, prikazan na sliki. Če bi bilo nizov več, bi v tem polju lahko izbrali, kateremu nizu želimo dodati trendno črto (premico). Na kartici *Možnosti* lahko z izbiro možnosti **Prikaži enačbo na grafikonu** in **Prikaži R-kvadrat vrednost na grafikonu** določimo, da se enačba premice in koeficient determinacije izpišeta na grafikon (slika 11.2). Na grafikonu prilagodimo oceni parametrov regresijskega modela na primerno število decimalnih mest.

Rešitev: $a = 49,5$, $b = 1,36$, $r^2 = 98\%$.

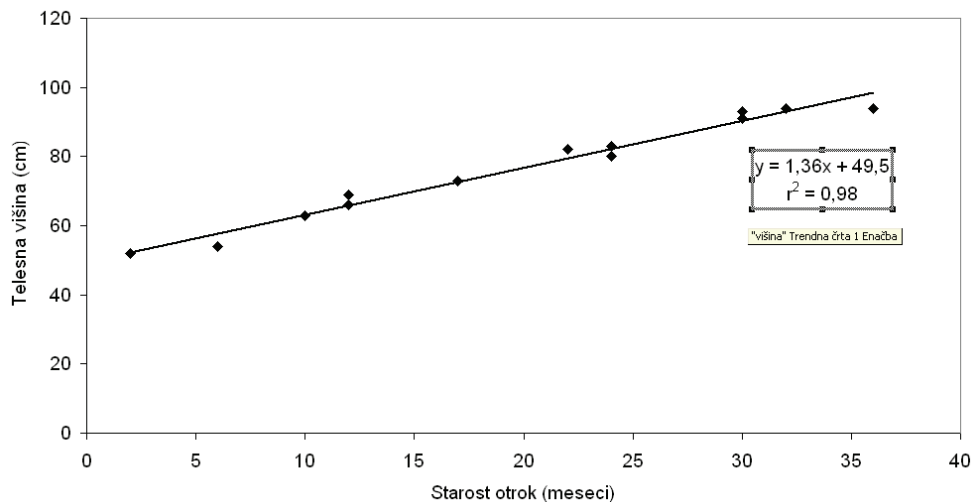
V povprečju otroci v tem starostnem obdobju (2 do 36 mesecev) zrastejo za 1,4 cm na mesec. Model ocenjuje, da je pri starosti 0 mesecev povprečna višina otroka 49,5 cm; za izbrane podatke je to manjša ekstrapolacija, saj je najmlajši otrok star 2 meseca. 98 % variabilnosti telesne višine otrok je pojasnjen z modelom oz. s starostjo otrok. ♦



OTROCI 3

Ali lahko na osnovi linearnega regresijskega modela napovemo višino otrok pri starosti 2 let in 5 let? Če je napoved upravičena, jo izračunajte.

Rešitev: dve leti stari otroci so v povprečju visoki 82,2 cm. Za otroke, stare 5 let z dobljenim modelom, ne smemo izračunati povprečne višine, ker bi to bila ekstrapolacija. ♦



Slika 11.2: Odvisnost telesne višine od starosti otrok; podatki z dodano regresijsko premico (ukaz **Dodaj trendno črto**)

11.1.1 Excelovo orodje za linearno regresijo

Model linearne regresije v Excelu izračunamo z ukazom **Orodja/Analiza podatkov/Regression**. V pogovornem oknu **Regression** so naslednja polja za določitev vhodnih podatkov:

	A	B	C	D	E	F	G	H	I	J
1	otrok	starost	višina							
2	1	2	52							
3	2	6	54							
4	3	10	63							
5	4	12	66							
6	5	12	69							
7	6	17	73							
8	7	22	82							
9	8	24	83							
10	9	24	80							
11	10	30	91							
12	11	30	93							
13	12	36	94							
14	13	32	94							
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										

Regression

Input

Input Y Range:

Input X Range:

Labels Constant is Zero

Confidence Level: %

Output options

Output Range:

New Worksheet Ply:

New Workbook

Residuals

Residuals Residual Plots

Standardized Residuals Line Fit Plots

Normal Probability

Normal Probability Plots

OK Cancel Pomoč

- v polju **Input Y Range** navedemo ime skupine celic z vrednostmi odvisne spremenljivke Y . Vrednosti morajo biti napisane v stolpcu, v izbranih celicah ne sme biti manjkajočih podatkov (praznih celic);
- v polju **Input X Range** navedemo ime skupine celic z vrednostmi neodvisne spremenljivke X . Vrednosti morajo biti napisane v stolpcu, v izbranih celicah ne sme biti manjkajočih podatkov (praznih celic);
- možnost **Labels** izberemo, če imamo v prvi vrstici stolpcev oznaki odvisne in neodvisne spremenljivke;
- možnost **Constant is Zero** izberemo, če želimo oceniti regresijski model za premico, ki gre skozi koordinatno izhodišče;
- v polju **Confidence Level** določimo verjetnost za interval zaupanja.

Mesto izpisa rezultatov določimo kot pri že opisanih ukazih iz skupine **Orodja/Analiza podatkov**. Ostale možnosti v pogovornem oknu **Regression** se navezujejo na analizo ostankov v modelu in jih na tem mestu ne bomo razlagali (*Residuals*). Pri regresijski analizi ponavadi podatke najprej grafično prikažemo z razsevnim grafikonom, da vidimo, ali je uporaba linearnega modela sploh upravičena, zato je možnost **Line Fit Plots** v večini primerov nepotrebna.

OTROCI 4

Naredite regresijsko analizo odvisnosti višine od starosti otrok še z orodjem **Orodja/Analiza podatkov/Regression** (tabela 11.2).

Rešitev: $a = 49,5$, $b = 1,36$, $r^2 = 98\%$.

V povprečju otroci v tem starostnem obdobju (2 do 36 mesecev) zrastejo za 1,4 cm na mesec. Model ocenjuje, da je pri starosti 0 mesecev povprečna višina otroka 49,5 cm; za izbrane podatke je to ekstrapolacija, saj je najmlajši otrok star 2 meseca.

98 % variabilnosti telesne višine otrok je pojasnjen z modelom oz. s starostjo otrok.

S 95 % zaupanjem trdimo, da interval (1,22 cm/mesec, 1,50 cm/mesec) pokrije naklon premice (parameter β).

Ničelno domnevo $H_0 : \beta = 0$ na podlagi F -preizkusa iz tabele ANOVA pri $\alpha = 0,05$ zavrneemo v korist alternativni domnevi $H_1 : \beta \neq 0$, ki pravi, da odvisnost telesne višine od starosti otrok ne obstaja ($p = 0,0000$). ♦

Tabela 11.2: Rezultati regresijske analize, narejene z ukazom **Orodja/Analiza podatkov/Regression** za podatke v datoteki REGRESIJSKA ANALIZA.XLS na delovnem listu "otroci". Sivo obarvani stolpci in vrstice s slovenskimi izrazi za izračune v tabelah so dodane.

Razlaga izračunov	Regression Statistics	
koeficient korelacije, r	Multiple R	0,99
koeficient determinacije, r ²	R Square	0,98
prilagojen koeficient determinacije	Adjusted R Square	0,97
standardna napaka regresije, s	Standard Error	2,42
število enot, n	Observations	13

		stopinje prostosti	vsota kvadriranih odklonov	srednji kvadrirani odklon	F statistika	p-vrednost
Vir variabilnosti	ANOVA	df	SS	MS	F	Significance F
model	Regression	1	2582,96	2582,96	442,07	3,12E-10
ostanek	Residual	11	64,27	5,84		
skupaj	Total	12	2647,23			

		ocena parametra modela	standardna napaka ocene	t - statistika	p-vrednost	spodnja meja intervala zaupanja	zgornja meja intervala zaupanja
		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
ocena za α	Intercept	49,53	1,45	34,268	1,57E-12	46,35	52,72
ocena za β	starost	1,36	0,06	21,026	3,12E-10	1,22	1,50

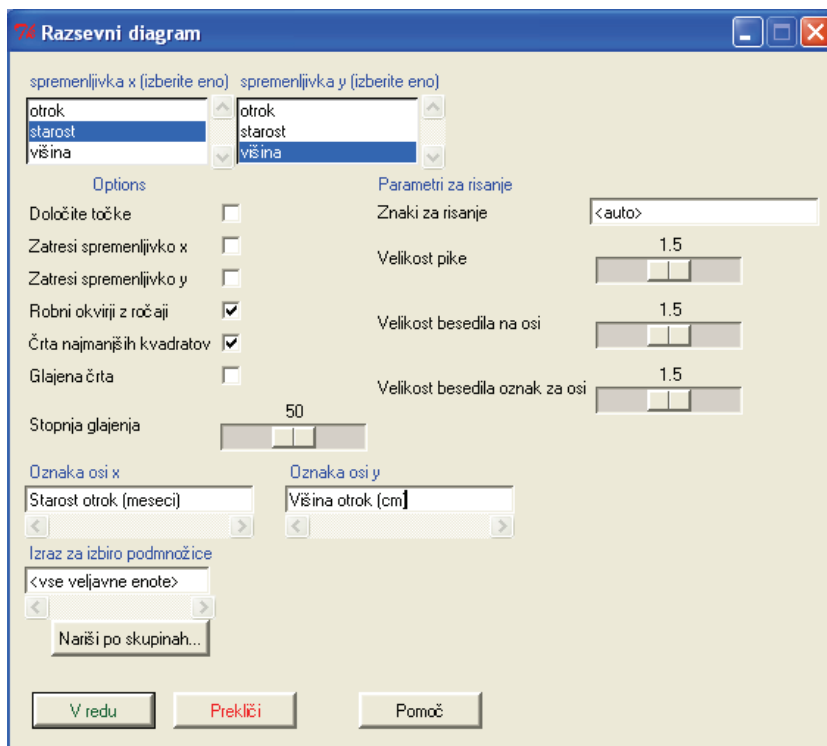
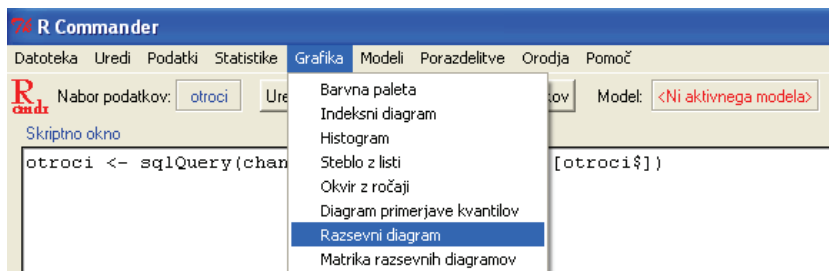
11.1.2 Linearna regresija s programom R

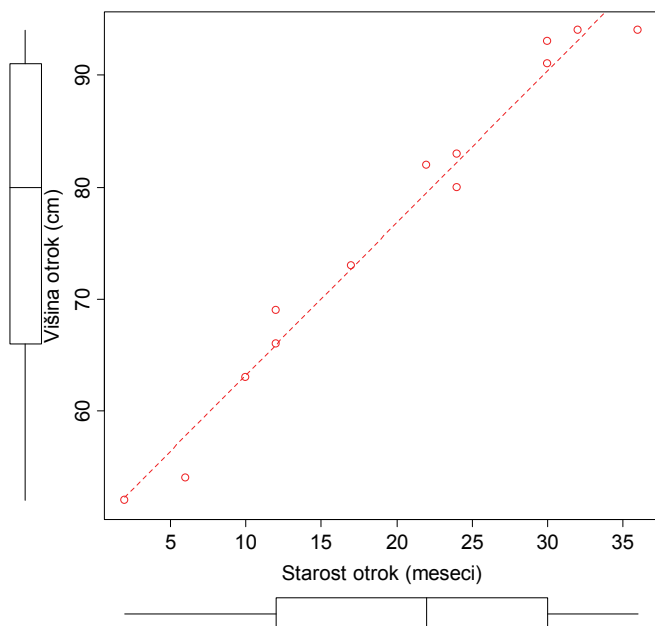
Regresijsko analizo s programom R v grafičnem vmesniku R Commander naredimo z ukazi **Statistike/Modeli/Linearna regresija**, **Modeli/Intervali zaupanja** in **Modeli/Preizkušanje hipotez/Tabela ANOVA**. Za oblikovanje razsevnega grafikona uporabimo ukaz **Grafika/Razsevni diagram**.

OTROCI 5

Izvedite regresijsko analizo odvisnosti višine otrok od starosti še s programom R v grafičnem vmesniku R Commander. Sledite spodnjim prikazom ukazov in pogovornih oken v R Commanderju:

Najprej podatke grafično prikažemo z razsevnim grafikonom. V pogovornem oknu **Razsevni diagram** najprej s klikom na ime izberemo neodvisno spremenljivko x , nato odvisno spremenljivko y . Nato izbiramo med različnimi možnostmi, poglejmo si samo dve: možnost **Robni okvirji z ročaji** omogoča, da vrednosti odvisne in neodvisne spremenljivke ob oseh prikažemo tudi z okvirjem z ročaji; možnost **Črta najmanjših kvadratov** omogoča grafični prikaz premice. Druga polja v pogovornem oknu omogočajo ustrezno prilagoditev velikosti predstavitev (krogcev) in velikost ter vsebino besedila ob oseh (slika 11.3).

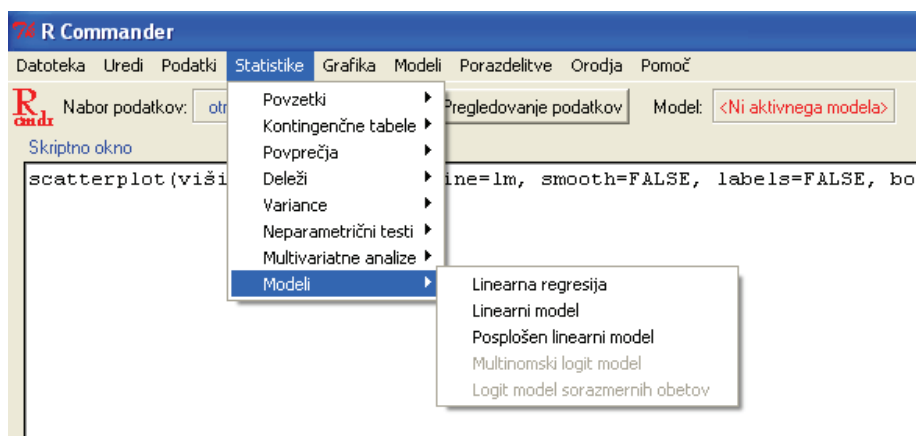




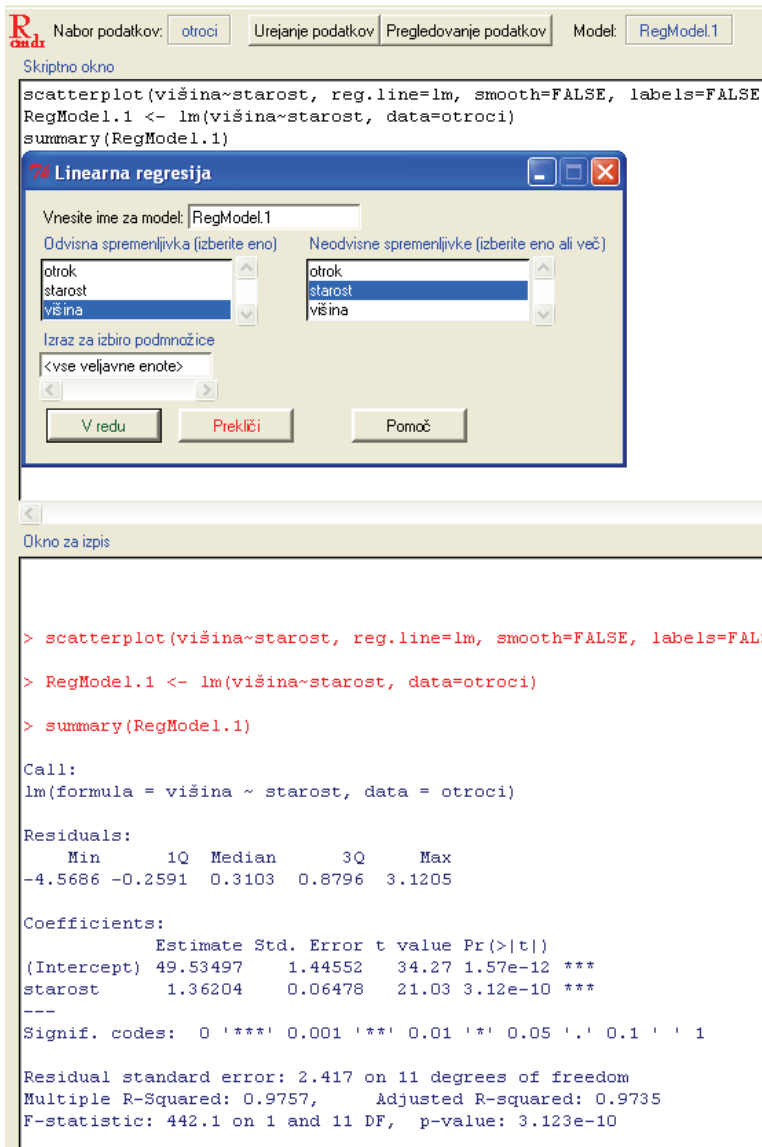
Slika 11.3: Razsevni grafikon za odvisnost telesne višine od starosti otrok z dodano regresijsko premico in okvirji z ročaji (program R, R Commander, ukaz **Grafika/Razsevni diagram**)

Oceni parametra linearnega regresijskega modela dobimo z ukazom

Statistike/Modeli/Linearna regresija:



V pogovornem oknu **Linearna regresija** v prvo polje vnesemo ime za model; po osnovni nastavitvi prvi model dobi ime »RegModel.1«, to ime lahko spremenimo po lastni izbiri. S klikom na ime spremenljivk nato izberemo odvisno in neodvisno spremenljivko:



Skriptno okno

```
scatterplot(višina~starost, reg.line=lm, smooth=FALSE, labels=FALSE)
RegModel.1 <- lm(višina~starost, data=otroci)
summary(RegModel.1)
```

Linearna regresija

Vnesite ime za model: RegModel.1

Odvisna spremenljivka (izberite eno):
otrok
starost
višina

Neodvisne spremenljivke (izberite eno ali več):
otrok
starost
višina

Izraz za izbiro podmnožice
<vse veljavne enote>

V redu Prekliči Pomôč

Okno za izpis

```
> scatterplot(višina~starost, reg.line=lm, smooth=FALSE, labels=FAL:
> RegModel.1 <- lm(višina~starost, data=otroci)
> summary(RegModel.1)

Call:
lm(formula = višina ~ starost, data = otroci)

Residuals:
    Min       1Q   Median       3Q      Max
-4.5686 -0.2591  0.3103  0.8796  3.1205

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  49.53497    1.44552   34.27 1.57e-12 ***
starost       1.36204    0.06478   21.03 3.12e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

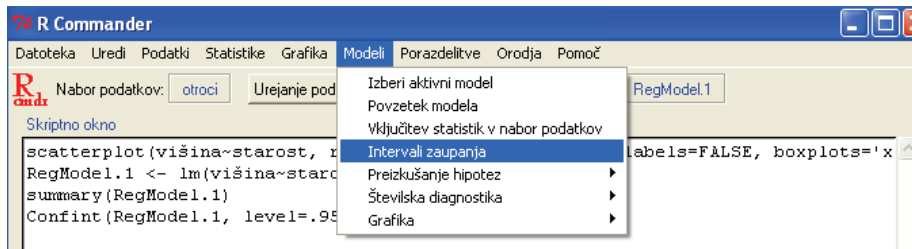
Residual standard error: 2.417 on 11 degrees of freedom
Multiple R-Squared:  0.9757,    Adjusted R-squared:  0.9735
F-statistic: 442.1 on 1 and 11 DF,  p-value: 3.123e-10
```

V izpisu modela z ukazom `summary(regModel.1)` se najprej izpiše ukazna vrstica za linearni model (`lm(...)`), nato se izpišejo opisne statistike ostankov²⁶, sledi tabela »Coefficients« z ocenami parametrov regresijskega modela (Estimate), standardno napako ocen parametrov (Std. Error), t -statistiko (t value) in p -vrednostjo ($Pr>|t|$). Ta tabela je podobna izpisu excelovega orodja. Pod tabelo se izpiše še Residual standard error, ki je standardna napaka

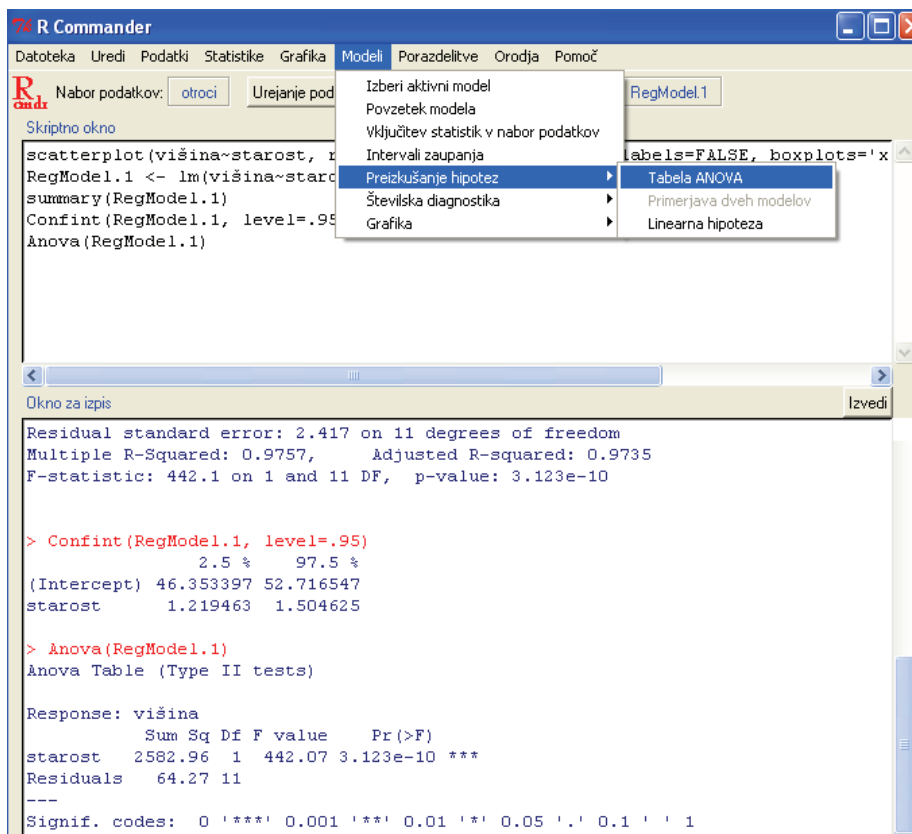
²⁶ Ostanek je razlika med vrednostjo podatka (y_i) in z modelom napovedano vrednostjo (\hat{y}_i), torej $(y_i - \hat{y}_i)$.

regresije z ustreznimi stopinjami prostosti (degrees of freedom). Multiple R-Squared je r^2 . F -statistika in njena p -vrednost sta iz tabele ANOVA.

Z ukazom **Modeli/Intervali zaupanja** dobimo spodnjo in zgornjo mejo intervalov zaupanja za parametre modela, zaupanje izberemo sami:



Klasično obliko tabele ANOVA za regresijski model dobimo z ukazom **Modeli/Preizkušanje hipotez/Tabela ANOVA**:



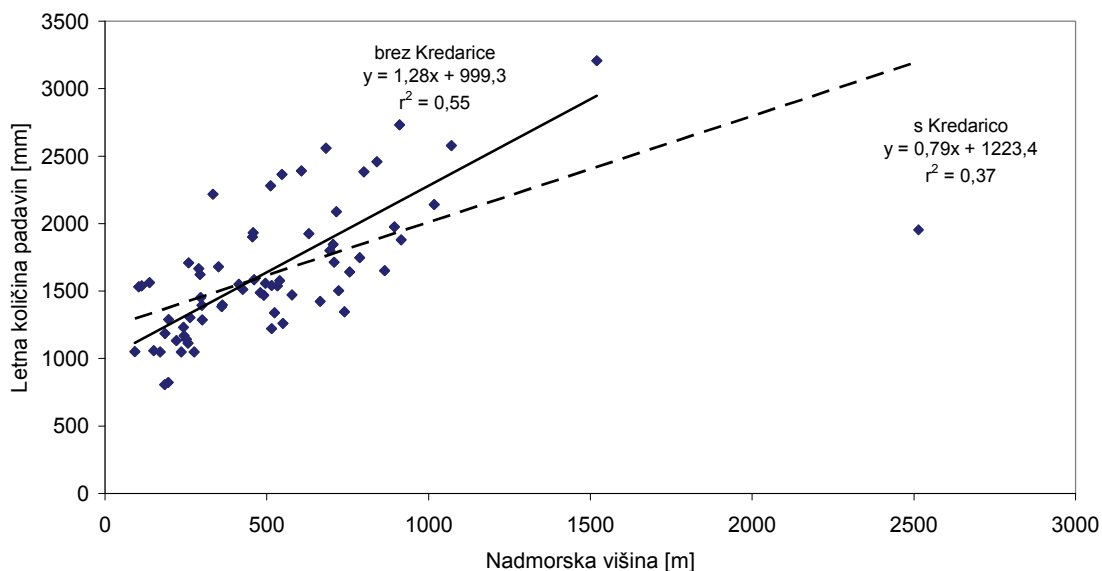
◆

Zgled 11.2: Padavine v Sloveniji

V datoteki REGRESIJSKA ANALIZA.XLS na delovnem listu »padavine« so podatki za 67 meteoroloških postaj v Sloveniji za leto 1992: ime postaje, nadmorska višina in letna količina padavin. Postaje so urejene po naraščajoči nadmorski višini. Analizirati želimo odvisnost količine padavin od nadmorske višine.

LETNE PADAVINE 1

Grafično prikažite odvisnost letne količine padavin od nadmorske višine (slika 11.4, vsi podatki). Ali je uporaba modela linearne regresije za te podatke upravičena? Obrazložite. ♦



Slika 11.4: Odvisnost letne količine padavin od nadmorske višine za vse meteorološke postaje in odvisnost za postaje brez Kredarice

LETNE PADAVINE 2

Regresijsko analizo ponovite še enkrat, tako da izločite postajo Kredarica. Primerjajte regresijska modela. ♦

LETNE PADAVINE 3

Analizirajte vpliv Kredarice na regresijski model. Njen vpliv grafično predstavite, kot kaže slika 11.4.

Navodilo: na sliko 11.4 dodajte še en niz podatkov, ta naj ima za neodvisno spremenljivko (nadmorska višina) vse vrednosti, razen podatka za Kredarico, pri izbiri vrednosti odvisne spremenljivke (količina padavin) pa prav tako izpustite podatek za Kredarico. V nadaljevanju grafikone oblikujte tako, da bosta oba niza odvisne spremenljivke prikazana z enakimi oznakami (**Oblikuj nize podatkov, Predstavitelj**). Premici narišite z uporabo ukaza **Dodaj trendno črto** za vsak niz odvisne spremenljivke posebej. V pogovornem oknu **Oblikuj trendno črto** na kartici *Vrsta* izberite možnost **Linearna**, na kartici *Možnosti* pa izberite možnosti **Prikaži enačbo na grafikonu** in **Prikaži R-kvadrat vrednost na grafikonu**. ♦

LETNE PADAVINE 4

Regresijsko analizo ponovite še enkrat, tako da izločite tudi postajo Dom na Komni. Primerjajte regresijske modele. ♦

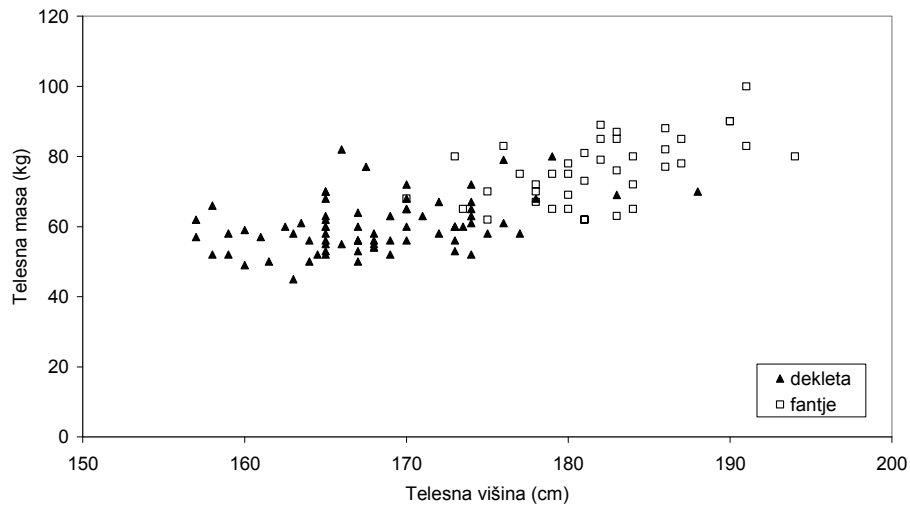
Domača naloga 11.1: Odvisnost telesne mase od telesne višine

Proučujemo odvisnost telesne mase študentk in študentov AG, BT in MB od telesne višine. Podatki so v datoteki ANKETA.XLS.

ANKETA 34

Podatke spremenljivk **spol**, **telesna masa** in **telesna višina** kopirajte v datoteko REGREGRESIJSKA ANALIZA.XLS na delovni list »Tmasa_Tvišina«. Grafično prikažite odvisnost telesne mase od višine posebej za študentke in posebej za študente na istem grafikonu (slika 11.5).

Navodilo: kopirane zapise (**spol**, **telesna masa** in **telesna višina**) najprej uredite po spolu (ukaz **Podatki/Razvrsti**). Vrstni red spremenljivk **telesna masa** in **telesna višina** v datoteki ni ustrezen za direktno uporabo grafikona XY(Raztreseni). Zato pred uporabo čarovnika za grafikone označite samo celice s **telesno maso** za moške, v drugem koraku čarovnika na kartici *Niz* pa določite še vrednosti za os X (**telesna višina**). Nato na grafikon dodajte še en niz, ki prikazuje podatke za dekleta. ♦



Domača naloga 11.2: Gozdna površina v Sloveniji

V tabeli 11.3 so podatki za gozdno površino po izbranih letih (Vir: Gozd in gozdarstvo Slovenije, 2004).

Tabela 11.3: Gozdna površina Slovenije po izbranih letih (Vir: Gozd in gozdarstvo Slovenije, 2004)

Leto	Gozdna površina (1000 ha)
1875	737
1947	897
1961	961
1970	1026
1980	1045
1990	1077
2000	1134
2003	1158

GOZDNA POVRŠINA 1

Podatke iz tabele 11.3 prepisite na delovni list »gozdna površina« v datoteko REGRESIJSKA ANALIZA.XLS. Izračunajte % gozda v Sloveniji po letih. Za velikost Slovenije upoštevajte vrednost 20 000 km², kar je 2 milijona ha. Rezultate grafično prikažite. ♦

GOZDNA POVRŠINA 2

Za gozdno površino izračunajte indekse z osnovo 1990 in jih grafično prikažite. ♦

GOZDNA POVRŠINA 3

Izračunajte povprečno letno stopnjo rasti za gozdno površino za obdobja:

1875-1947

1961-1980

1990-2003. ♦

GOZDNA POVRŠINA 4

Izračunajte oceni parametrov regresijskega modela za odvisnost gozdne površine od časa (leta).

Obrazložite rezultate. ♦

11.2 Korelacija

Analizo korelacije oz. povezanosti dveh številskih spremenljivk (X , Y) naredimo takrat, kadar je odnos med spremenljivkama obojestranski, kar pomeni, da prva spremenljivka vpliva na vrednosti druge in obratno. Kot mero povezanosti dveh številskih spremenljivk bomo spoznali Pearsonov in Spearmanov koeficient korelacije. Pearsonov koeficient korelacije uporabljamo kot mero za linearno povezanost dveh številskih spremenljivk.

Če povezava med spremenljivkami ni linearna, je pa monotona in če so podatki ene ali obeh spremenljivk rangi ali če so podatki nezanesljivi, kot mero povezanosti dveh številskih spremenljivk uporabimo Spearmanov koeficient korelacije.

Če želimo preveriti, ali je povezanost med dvema spremenljivkama statistično značilna, moramo preveriti ničelno domnevo, ki pravi, da povezanosti ni ($\rho = 0$). Najenostavneje to naredimo »peš« z uporabo statističnih tabel, ki podajajo mejne vrednosti Pearsonovega in Spearmanovega koeficienta korelacije za različno število podatkov (n) in za različne stopnje značilnosti (α).

Če tabel nimamo, v primeru preverjanja Pearsonovega koeficienta korelacije, izračunamo testno statistiko:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}, \text{ katere ničelna porazdelitev je Studentova porazdelitev } t(SP = n - 2).$$

11.2.1 Pearsonov in Spearmanov koeficient korelacije v Excelu

Pearsonov koeficient korelacije Excel izračuna s funkcijo `CORREL(array1; array2)`. Za argumenta funkcije določimo obseg celic, v katerih imamo podatke za spremenljivki, npr. »array1« je obseg celic A1:A10 za spremenljivko X in »array2« je obseg celic B1:B10 za spremenljivko Y . Ko računamo Pearsonov koeficient korelacije, vrstni red spremenljivk ni pomemben.

Če analiziramo povezanost med več kot dvema spremenljivkama, namesto funkcije CORREL uporabimo orodje **Orodja/Analiza podatkov/Correlation**. Na ta način dobimo izračun Pearsonovih koeficientov korelacije za vse pare spremenljivk naenkrat.

Za izračun Spearmanovega koeficienta korelacije v Excelu uporabimo isto funkcijo kot za Pearsonov koeficient korelacije CORREL, le da za argumente funkcije določimo podatke, ki predstavljajo range. Če računamo Spearmanov koeficient korelacije za spremenljivki, katerih vrednosti niso rangi, moramo podatkom predhodno prirediti range. Če imata dve enoti isto vrednost izbrane spremenljivke, jima priredimo njun povprečni rang.

Zgled 11.3: Vsebnost žvepla in dušika v lišajih

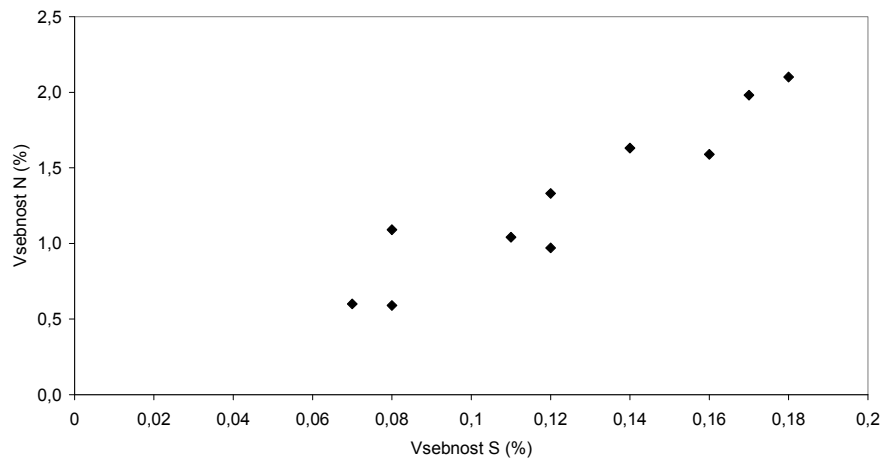
Dva izmed pokazateljev splošne onesnaženosti zraka sta vsebnost dušika in vsebnost žvepla v drevesnih lišajih (tabela 11.4, datoteka KORELACIJA.XLS, delovni list »lišaji«).

Tabela 11.4: Vsebnost dušika (N) in žvepla (S) v drevesnih lišajih za deset lokacij v zahodni Sloveniji

Lokacija	S (%)	N (%)
Učeja	0,17	1,98
Mangart	0,12	0,97
Lepena	0,14	1,63
Fužinske planine	0,07	0,60
Gorenja Trebuša	0,11	1,04
Kotel	0,12	1,33
Križ	0,08	1,09
Martuljek	0,08	0,59
Planina pod Golico	0,16	1,59
Kanji dol	0,18	2,10

LIŠAJI 1

V Excelu grafično prikažite povezanost N in S v lišajih (slika 11.6). ♦



Slika 11.6: Razsevni grafikon za povezanost vsebnosti S in N v lišajih

LIŠAJI 2

Katera mera povezanosti je primerna za take podatke? Izračunajte jo najprej z uporabo računalna, nato pa še v Excelu (funkcija CORREL ali orodje **Orodja/Analiza podatkov/Correlation**). ♦

LIŠAJI 3

Preverite domnevo, da ni povezave med vsebnostjo žvepla in vsebnostjo dušika v lišajih (stopnja značilnosti je 0,01). Najprej uporabite tabele, nato izračunajte ustrezno testno statistiko in mejno vrednost zanjo (tabela 11.5). Obrazložite rezultate.

Rešitev: $H_0 : \rho = 0$ in $H_1 : \rho \neq 0$

$r = 0,94$, iz tabel odčitana mejna vrednost Pearsonovega koeficienta korelacije pri $\alpha = 0,01$ in $n = 10$ je $r_{mejna} = 0,7646$. Ničelno domnevo pri $\alpha = 0,01$ zavrnamo. Linearna povezanost med vsebnostjo žvepla in dušika v lišajih je statistično značilna ($p < 0,01$).

Obstaja močna pozitivna linearna povezanost med vsebnostjo žvepla in dušika v lišajih ($p < 0,01$).

Tabela 11.5: Izračun Pearsonovega koeficienta korelacije in preverjanje njegove statistične značilnosti; izpis formul (levo); rezultati (desno)

	A	B		A	B
13	Pearsonov koeficient korelacije	=CORREL(B2:B11;C2:C11)	13	Pearsonov koeficient korelacije	0,94
14	alfa	0,01	14	alfa	0,01
15	$H_0: \rho = 0$		15	$H_0: \rho = 0$	
16	$H_1: \rho$ ni enako 0		16	$H_1: \rho$ ni enako 0	
17	n	=COUNT(B2:B11)	17	n	10
18	SP	=B17-2	18	SP	8
19	t	=B13*SQRT(B17-2)/SQRT(1-B13^2)	19	t	7,7240
20	t_{mejna}	=TINV(B14;B18)	20	t_{mejna}	3,3554
21	p-vrednost	=TDIST(ABS(B19);B18;2)	21	p-vrednost	0,0001

t -preizkus: $t = 7,724$ in $p = 0,0001$.

Ničelno domnevo pri $\alpha = 0,01$ zavrnamo. Linearna povezanost med vsebnostjo žvepla in dušika v lišajih je statistično značilna ($p = 0,0001$).

Obstaja močna pozitivna linearna povezanost med vsebnostjo žvepla in dušika v lišajih ($p = 0,0001$). ♦

Zgled 11.5: Vsebnost težkih kovin v tleh

Raziskovalci so proučevali vsebnost težkih kovin v tleh. V okolici Celja so vzorčili tla na 15 lokacijah in v laboratoriju izmerili vsebnost cinka (Zn), kadmija (Cd), svinca (Pb) in bakra (Cu). Podatki so v tabeli 11.6 in v datoteki KORELACIJA.XLS, delovni list »kovine«.

Tabela 11.6: Vsebnost težkih kovin (mg/g) v talnih vzorcih iz okolice Celja

vzorec	Cd	Zn	Cu	Pb
1	0,5	128,0	30,7	29,9
2	0,4	150,0	25,8	24,8
3	0,6	97,0	15,6	36,7
4	0,3	96,0	24,0	23,6
5	0,5	68,0	9,6	45,3
6	1,0	144,0	21,8	47,9
7	0,9	132,0	17,1	41,4
8	1,3	172,0	11,2	85,8
9	0,7	130,0	13,3	46,2
10	1,4	294,0	22,9	90,5
11	1,9	431,0	17,5	108,0
12	5,2	497,0	26,7	80,7
13	2,1	293,0	27,9	64,0
14	1,9	223,0	10,8	130,1
15	1,0	119,0	12,0	52,1

KOVINE 1

Oblikujte razsevni grafikon za Cd in Pb. Ali je Pearsonov koeficient korelacije v tem primeru primerna mera povezanosti? Ali v podatkih obstajajo osamelci? Kakšno povezanost lahko predpostavimo, če osamelca izločimo iz analize povezanosti? ♦

KOVINE 2

V excelovi datoteki podatkom za Cd in Pb priredite range (določite vrednosti dveh novih spremenljivk). Bodite pozorni na vezane range.

Navodilo: za prirejanje rangov ima Excel pripravljeno funkcijo RANK, ki pa ne zna prirejati povprečnih rangov enotam z enako vrednostjo, kar pomeni, da deluje prav samo v primeru, ko imajo vse enote v vzorcu različne vrednosti. Range je zato varneje prirediti "peš". Pri tem si lahko pomagamo tako, da podatke za vsako spremenljivko posebej uredimo po velikosti od najmanjšega do največjega z ukazom **Podatki/Razvrsti**. Ko so podatki urejeni, naredimo novo spremenljivko z zaporedno vrednostjo v razvrstitvi (samodejno polnjenje celic z zaporednimi števili). Za enote, ki imajo iste vrednosti, moramo nato izračunati povprečni rang:

	A	B	C	D	E	F	G
1	vzorec	Cd	Zn	Cu	Pb	Rang Cd	Rang Pb
2	4	0,3	96	24	23,6	1	1
3	2	0,4	150	25,8	24,8	2	2
4	1	0,5	128	30,7	29,9	3,5	3
5	5	0,5	68	9,6	45,3	3,5	6
6	3	0,6	97	15,6	36,7	5	4
7	9	0,7	130	13,3	46,2	6	7
8	7	0,9	132	17,1	41,4	7	5
9	6	1	144	21,8	47,9	8,5	8
10	15	1	119	12	52,1	8,5	9
11	8	1,3	172	11,2	85,8	10	12
12	10	1,4	294	22,9	90,5	11	13
13	11	1,9	431	17,5	108	12,5	14
14	14	1,9	223	10,8	130,1	12,5	15
15	13	2,1	293	27,9	64	14	10
16	12	5,2	497	26,7	80,7	15	11
17							
18	Spearmanov koeficient korelacije		=CORREL(F2:F16;G2:G16)				

♦

KOVINE 3

Izračunajte Spearmanov koeficient korelacije za Cd in Pb. Obrazložite rezultate. Z uporabo tabel ugotovite, ali je povezanost med Cd in Pb statistično značilna ($\alpha = 0,05$).

Rešitev: $r_s = 0,89$, pri $\alpha = 0,05$ iz tabel odčitamo mejno vrednost $r_{s_mejna} = 0,5214$, kar pomeni, da je povezanost med vsebnostjo Cd in Pb statistično značilna ($\alpha = 0,05$).

Obstaja pozitivna povezanost med vsebnostjo Cd in Pb v tleh ($p < 0,05$). ♦

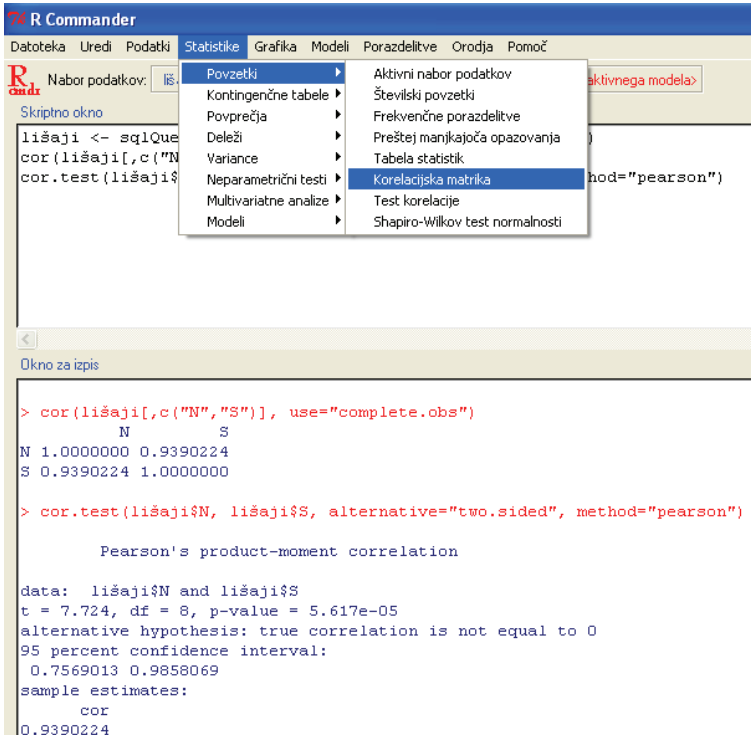
11.2.2 Pearsonov in Spearmanov koeficient korelacije v programu R

S programom R in grafičnim vmesnikom R Commander Pearsonov in Spearmanov koeficient korelacije izračunamo z ukazom **Statistike/Povzetki/Korelacijska matrika**, statistično značilnost koeficientov korelacije pa ugotovimo z ukazom **Statistike/Povzetki/Test korelacije**. S tem ukazom se izračunata tudi spodnja in zgornja meja intervala zaupanja za Pearsonov koeficient korelacije ρ .

LIŠAJI 4

Analizo povezanosti N in S v lišajih naredite še s programom R.

Navodilo: pred uvozom podatkov v R v Excelu na nov delovni list »lišaji za R« pripravite podatke v primerni obliki (brez tabel z izračuni, brez znaka za % v imenih spremenljivk). ♦



```

R Commander
Datoteka Uredi Podatki Statistike Grafika Modeli Porazdelitve Orodja Pomoč
R Nabor podatkov: lišaji
Statistike
  Povzetki
  Kontingenčne tabele
  Povprečja
  Deleži
  Variance
  Neparometrični testi
  Multivariatne analize
  Modeli
  Aktivni nabor podatkov
  Številski povzetki
  Frekvenčne porazdelitve
  Preštej manjkajoča opazovanja
  Tabela statistik
  Korelacijska matrika
  Test korelacije
  Shapiro-Wilkov test normalnosti
  Aktivnega modela:

Skriptno okno
lišaji <- sqlQue
cor(lišaji[,c("N","S")])
cor.test(lišaji$N, lišaji$S, alternative="two.sided", method="pearson")

Okno za izpis
> cor(lišaji[,c("N","S")], use="complete.obs")
      N      S
N 1.0000000 0.9390224
S 0.9390224 1.0000000

> cor.test(lišaji$N, lišaji$S, alternative="two.sided", method="pearson")

Pearson's product-moment correlation

data: lišaji$N and lišaji$S
t = 7.724, df = 8, p-value = 5.617e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7569013 0.9858069
sample estimates:
      cor
0.9390224

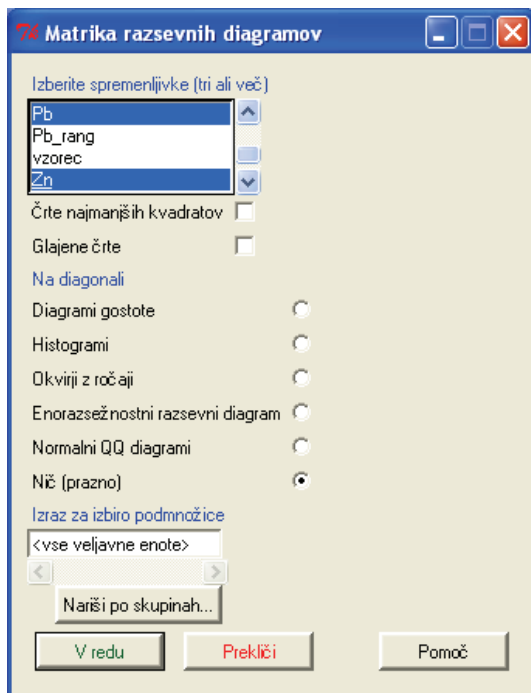
```

KOVINE 4

Oblikujte razsevne grafikone za vse pare težkih kovin. Katera mera povezanosti je primerna za posamezen par kovin?

Navodilo: v programu R dobimo vse slike naenkrat z ukazom **Grafika/Matrika razsevnih diagramov** (slika 11.7 zgoraj). V pogovornem oknu *Matrika razsevnih grafikonov* izberemo več spremenljivk naenkrat, tako da med klikanjem na imena spremenljivk držimo tipko CTRL:

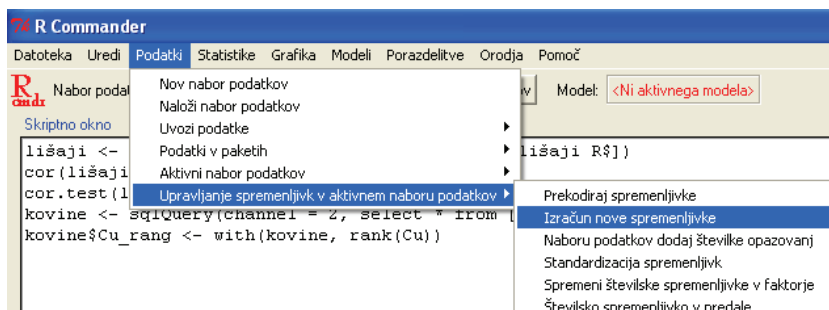
◆

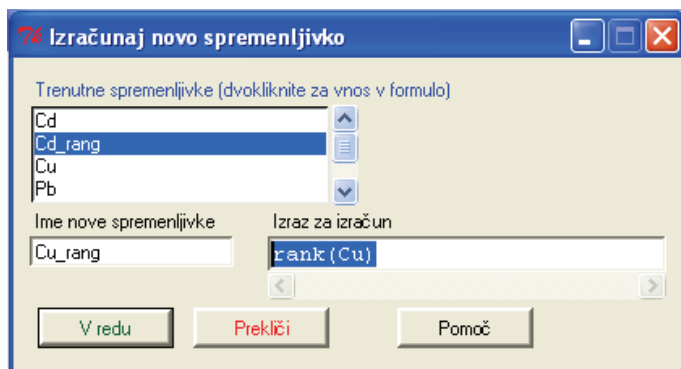


KOVINE 5

Da bi grafično prikazali povezanost vsebnosti težkih kovin na podlagi rangov, podatkom za vse štiri težke kovine priredite range.

Navodilo: uporabite ukaz **Podatki/Upravljanje spremenljivk v aktivnem naboru podatkov/Izračun nove spremenljivke**. V programu R range izračuna funkcijo rank():





```

R Commander
Datoteka Uredi Podatki Statistike Grafika Modeli Porazdelitve Orodja Pomoč
Rcmdr Nabor podatkov: kovine Urejanje podatkov Pregledovanje podatkov Model: <Ni aktivnega modela>
Skriptno okno
kovine$Cd_rang <- with(kovine, rank(Cd))
kovine$Cu_rang <- with(kovine, rank(Cu))
kovine$Cu_rang
kovine$Cd_rang

kovine$Pb_rang <- with(kovine, rank(Pb))
kovine$Zn_rang <- with(kovine, rank(Zn))
kovine$Pb_rang
kovine$Zn_rang

Okno za izpis
> kovine$Cu_rang <- with(kovine, rank(Cu))
> kovine$Cu_rang
[1] 15 12 6 11 1 9 7 3 5 10 8 13 14 2 4
> kovine$Cd_rang
[1] 3.5 2.0 5.0 1.0 3.5 8.5 7.0 10.0 6.0 11.0 12.5 15.0 14.0 12.5 8.5
> kovine$Pb_rang <- with(kovine, rank(Pb))
> kovine$Zn_rang <- with(kovine, rank(Zn))
> kovine$Pb_rang
[1] 3 2 4 1 6 8 5 12 7 13 14 11 10 15 9
> kovine$Zn_rang
[1] 5 9 3 2 1 8 7 10 6 13 14 15 12 11 4

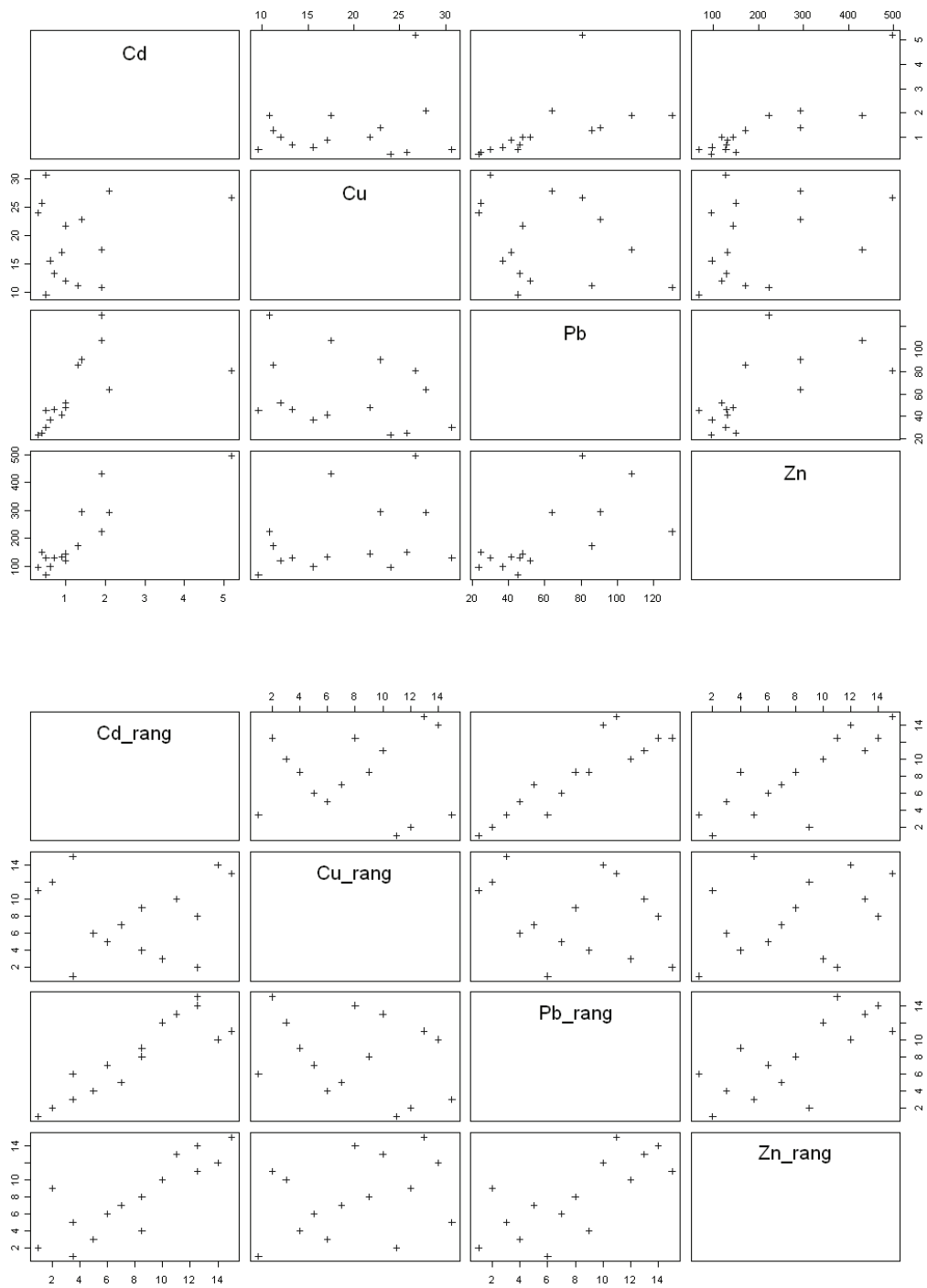
```

Funkcija `rank`, v nasprotju z excelovo funkcijo `RANK`, pravilno priredi vezane range. ♦

KOVINE 6

Ponovite grafični prikaz povezanosti vsebnosti težkih kovin v tleh na rangih (slika 11.7 spodaj).

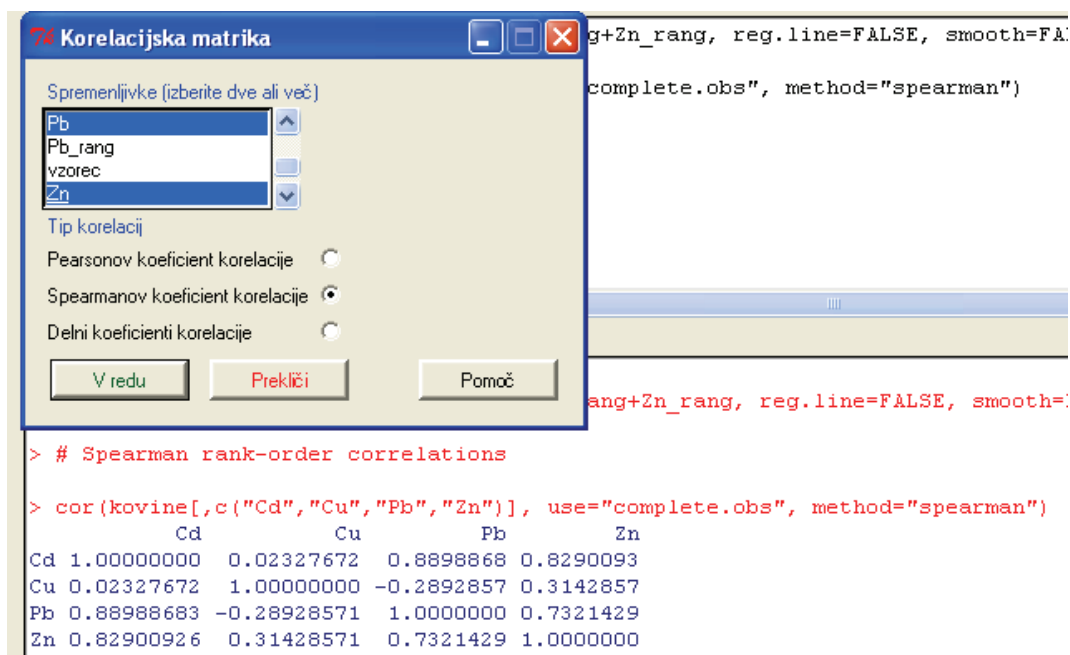
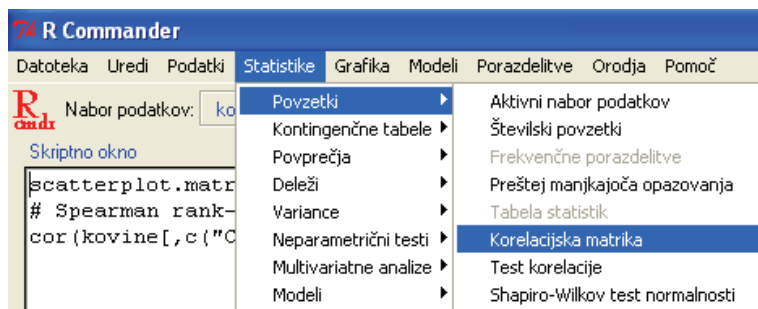
♦



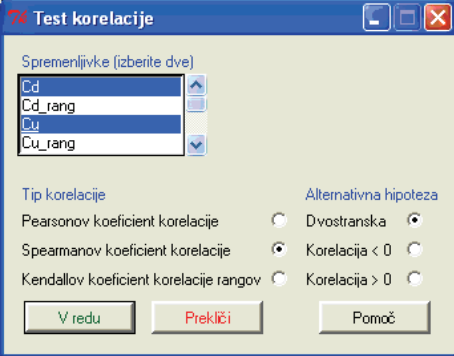
Slika 11.7: Matrika razsevnih grafikonov za talne vzorce glede na vsebnost težkih kovin (zgoraj) in glede na range osnovnih spremenljivk (spodaj)

KOVINE 7

Analizo povezanosti vsebnosti kovin v tleh ponovite še v programu R. Uporabite ukaz **Statistike/Povzetki/Korelacijska matrika**, možnost *Spearmanov koeficient korelacije*:



Statistično značilnost Spearmanovega koeficienta korelacije za posamezen par spremenljivk lahko preverimo z ukazom **Statistike/Povzetki/Test korelacije**: ♦



```

> cor.test(kovine$Cd, kovine$Cu, alternative="two.sided", method="spearman")

Spearman's rank correlation rho

data: kovine$Cd and kovine$Cu
S = 546.965, p-value = 0.9344
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.02327672

Warning message:
In cor.test.default(kovine$Cd, kovine$Cu, alternative = "two.sided", :
  Cannot compute exact p-values with ties

```

KOVINE 8

Za vse pare težkih kovin izračunajte Pearsonove koeficiente korelacije na rangih. Primerjajte rezultate z rezultati iz prejšnje naloge. ♦

KOVINE 9

Vzorec 12 glede vsebnosti Cd predstavlja osamelec. Analizo povezanosti kovin v talnih vzorcih ponovite še brez podatkov za vzorec 12. Ali je v tem primeru primerna mera povezanosti Pearsonov koeficient korelacije? ♦

12 Hi-kvadrat preizkusi

Novi pojmi:

- hi-kvadrat preizkus za preverjanje domneve o porazdelitvi
- hi-kvadrat preizkus za preverjanje povezanosti dveh opisnih spremenljivk
- hi-kvadrat preizkus za preverjanje homogenosti struktur

Datoteke s podatki:

HI KVADRAT.XLS
ANKETA.XLS

Ustvarjene datoteke:

HI KVADRAT.DOC

Poročilo o opravljenih nalogah, ki sledijo, oblikujte v wordovi datoteki z imenom HI KVADRAT.DOC. ♦

12.1 Hi-kvadrat preizkus za preverjanje domneve o porazdelitvi

Preverjamo ničelno domnevo:

H_0 : verjetnostna porazdelitev slučajne spremenljivke X je:

$$X: \begin{bmatrix} x_1 & x_2 & \cdots & x_k \\ p_{10} & p_{20} & \cdots & p_{k0} \end{bmatrix}.$$

Tu x_i , $i=1, \dots, k$ označuje vrednost ali razred spremenljivke X ; $p_{10}, p_{20}, \dots, p_{k0}$ so teoretične oz. pričakovane verjetnosti, ki izhajajo iz teorije; k je število vrednosti (npr. število možnih izidov) ali razredov spremenljivke.

Alternativna domneva je:

H_1 : H_0 ne velja.

Imamo vzorec velikosti n . Ugotovimo, koliko enot v vzorcu ima posamezno vrednost oz. sodi v posamezen razred x_i . Pripadajoče frekvence imenujemo dejanske frekvence

$f_i, i = 1, \dots, k$. Na podlagi teoretičnih verjetnosti izračunamo teoretične oz. pričakovane frekvence $f'_i = n \cdot p_{i0}, i = 1, \dots, k$. Ujemanje dejanskih in pričakovanih frekvenc ovrednotimo s Pearsonovo hi-kvadrat statistiko:

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - f'_i)^2}{f'_i}.$$

Ničelna porazdelitev je $\chi^2 \approx \chi^2_{\alpha}(SP = k - 1)$. H_0 obdržimo, če je $\chi^2 \leq \chi^2_{\alpha}(SP = k - 1)$ in zavrnemo v korist H_1 , če je $\chi^2 > \chi^2_{\alpha}(SP = k - 1)$. Alternativna domneva je torej enostranska.

Pogoj za uporabo Pearsonove hi-kvadrat statistike je: $f'_i \geq 5, i = 1, \dots, k$.

Zgled 12.1: Vhod v trgovino

V trgovini s tremi vhodi so želeli ugotoviti, ali kupci uporabljajo vse tri vhode enako pogosto. Izvedli so opazovanje, v katerem so šteli, skozi kateri vhod je vstopilo prvih 200 kupcev v izbranem dnevu. Dobili so sledeče rezultate: 83 skozi vhod A, 61 skozi vhod B in 56 skozi vhod C.

VHOD 1

Podatke vnesite v pripravljeno tabelo v datoteki HI KVADRAT.XLS na delovnem listu »vhod«. Koliko od 200 obiskovalcev trgovine bi vstopilo skozi vsaka od vrat, če predpostavimo, da je verjetnost, da vstopijo skozi katerikoli vhod, enaka (tabela 12.1)? ♦

VHOD 2

Preverite ničelno domnevo, da je verjetnost, da vstopijo skozi katerikoli vhod, enaka ($\alpha = 0,05$). Obrazložite rezultate.

Navodilo: nalogo rešite v Excelu v pripravljene tabeli na delovnem listu »vhod«. Pri izračunu uporabite funkcije CHIINV, CHIDIST in CHITEST (tabela 12.1). Vse tri funkcije najdemo v skupini oz. zvrsti funkcij Statistika.

Funkcija CHIINV(probability; deg_freedom) ima dva argumenta: »probability« je verjetnost v repu porazdelitve (pri preverjanju domnev je to stopnja značilnosti α); »deg_freedom« so stopinje prostosti SP (parameter hi-kvadrat porazdelitve). Funkcija vrne vrednost hi-kvadrat statistike χ_1^2 , za katero velja $P(\chi^2 > \chi_1^2) = \alpha$.

Funkcija CHIDIST(x; deg_freedom) ima dva argumenta: »x« je vrednost hi-kvadrat statistike, za katero funkcija vrne verjetnost $P(\chi^2 > x) = p$; »deg_freedom« so stopinje prostosti SP .

Funkcija CHITEST(actual_range;expected_range) vrne p -vrednost hi-kvadrat preizkusa. Prvi argument »actual_range« je obseg celic z dejanskimi frekvencami in drugi argument »expected_range« je obseg celic s pričakovanimi frekvencami.

Rešitev:

$$H_0 : p_A = p_B = p_C = 1/3$$

H_1 : ničelna domneva ne velja.

$$\chi^2 = 6,19 \quad SP = 2 \quad p = 0,0453 \text{ (tabela 12.1)}$$

Ničelno domnevo pri $\alpha = 0,05$ zavrnemo. Rezultati so statistično značilni.

Število obiskovalcev, ki pridejo v trgovino skozi enega izmed treh vhodov, ni enakomerno porazdeljeno po vhodih ($p = 0,0453$). Največja odstopanja so pri vhodu A. Podatki nakazujejo, da skozi vhod A vstopi več obiskovalcev, kot bi pričakovali pri enakomerni porazdelitvi. ♦

Tabela 12.1: Izračuni pričakovanih frekvenc in hi-kvadrat preizkus za preverjanje ničelne domneve, da je verjetnost, da obiskovalec vstopi v trgovino skozi katerikoli vhod, enaka; izpis formul (levo); izpis rezultatov (desno)

	A	B	C	D
1	vhod	frekvenca	pričakovana frekvenca	prispevek k hi-kvadrat statistiki
2	A	83	=\$B\$5/3	=(B2-C2)*2/C2
3	B	61	=\$B\$5/3	=(B3-C3)*2/C3
4	C	56	=\$B\$5/3	=(B4-C4)*2/C4
5	Vsota	=SUM(B2:B4)	=SUM(C2:C4)	=SUM(D2:D4)
6				
7	hi-kvadrat	=D5		
8	alfa	0,05		
9	SP	=COUNT(B2:B4)-1		
10	hi-kvadrat _{mejna}	=CHIINV(B8;B9)		
11	p-vrednost CHIDIST	=CHIDIST(B7;B9)		
12	p-vrednost CHITEST	=CHITEST(B2:B4;C2:C4)		

	A	B	C	D
1	vhod	frekvenca	pričakovana frekvenca	prispevek k hi-kvadrat statistiki
2	A	83	66,67	4,00
3	B	61	66,67	0,48
4	C	56	66,67	1,71
5	Vsota	200	200	6,19
6				
7	hi-kvadrat	6,190		
8	alfa	0,05		
9	SP	2		
10	hi-kvadrat _{mejna}	5,991		
11	p-vrednost CHIDIST	0,0453		
12	p-vrednost CHITEST	0,0453		

VHOD 3

Kakšen bi bil statistični sklep pri stopnji značilnosti 0,01? ♦

Zgled 12.2: Porazdelitev debeline debel dreves

V datoteki DEBELINA DEBEL.XLS in v datoteki HI KVADRAT.XLS na delovnem listu »debelina debel« je frekvenčna tabela za debelino debel dreves. Prvi in zadnji razred sta v tej nalogi odprta navzdol oz. navzgor.

DEBELINA DEBEL 1

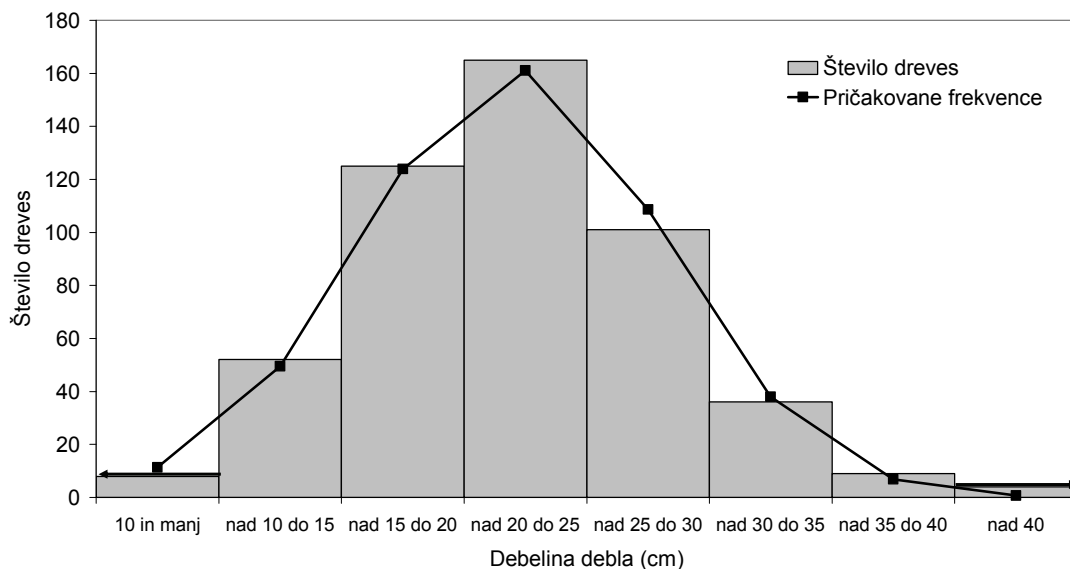
Ob predpostavki, da je debelina debel dreves porazdeljena normalno $N(22, 6)$, izračunajte verjetnost, da je debelina debela v posameznem razredu empirične porazdelitve (tabela 12.2). ♦

DEBELINA DEBEL 2

Koliko od 500 dreves bi bilo v posameznem razredu, če bi bila debelina debela porazdeljena po normalni porazdelitvi $N(22, 6)$? Dobljene pričakovane frekvence primerjajte z dejanskimi (tabela 12.2). ♦

DEBELINA DEBEL 3

Grafično prikažite dejanske in pričakovane frekvence s histogramom in poligonom, kot kaže slika 12.1. ♦



Slika 12.1: Empirična porazdelitev dreves glede na debelino debel (histogram); poligon predstavlja pričakovane frekvence po normalni porazdelitvi $N(22, 6)$ za posamezni razred

DEBELINA DEBEL 4

Ali empirična porazdelitev debeline debel dreves potrjuje predpostavko, da je debelina debel porazdeljena normalno s parametroma $\mu = 22$ cm in $\sigma = 6$ cm, $\alpha = 0,05$ (tabeli 12.2 in 12.3)? Napišite ničelno in alternativno domnevo ter naredite statistični preizkus. Obrazložite rezultate.

Navodilo: pazite na predpostavko o potrebni velikosti pričakovanih frekvenc. Za pravilno uporabo hi-kvadrat preizkusa je potrebno v frekvenčni tabeli zadnja dva razreda združiti (tabela 12.3).

Rešitev:

H_0 : verjetnostna porazdelitev debeline debel je normalna $N(22, 6)$:

$$X: \begin{bmatrix} <10 & 10 \text{ do pod } 15 & 15 \text{ do pod } 20 & 20 \text{ do pod } 25 & 25 \text{ do pod } 30 & 30 \text{ do pod } 35 & \geq 35 \\ 0,0228 & 0,0989 & 0,2478 & 0,3220 & 0,2173 & 0,0761 & 0,0151 \end{bmatrix}$$

$$k = 7, \chi^2 = 5,795 \text{ in } p = 0,4465$$

Ničelno domnevo obdržimo. Empirična porazdelitev debeline debel dreves je v skladu s predpostavko, da za debelino debel dreves lahko privzamemo normalno porazdelitev $N(22, 6)$.

Kakšen bi bil statistični sklep, če razredov ne bi združili (kar bi bilo narobe)? ♦

DEBELINA DEBEL 5

Kakšni bi bili rezultati hi-kvadrat preizkusa, če bi predpostavili, da je debelina debel dreves porazdeljena normalno s parametroma $\mu = 23$ cm in $\sigma = 7$ cm?

Navodilo: katere vrednosti v tabelah 12.2 in 12.3 moramo spremeniti, da se formule pravilno preračunajo? Ali je potrebno tudi v tem primeru razrede združiti?

Rešitev: $k = 7 \quad \chi^2 = 18,255 \text{ in } p = 0,0056 \quad \blacklozenge$

Tabela 12.2: Izračuni za hi-kvadrat preizkus ničelne domneve o porazdelitvi debeline debel; izpis formul (zgoraj); izpis rezultatov (spodaj)

	B	C	D	E	F	G	H	I	J
1	Debelina debela (cm)	Število dreves	Sp. meja	Zg. meja	P(X<sp. meja)	P(X<zg. meja)	P(razred)	Pričakovane frekvence	Prispevek k hi-kvadrat statistiki
2	10 in manj	8		10		=NORMDIST(E2;\$C\$12;\$C\$13;TRUE)	=G2-F2	=\$C\$10*H2	=(I2-C2)^2/I2
3	nad 10 do 15	52	10	15	=NORMDIST(D3;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E3;\$C\$12;\$C\$13;TRUE)	=G3-F3	=\$C\$10*H3	=(I3-C3)^2/I3
4	nad 15 do 20	125	15	20	=NORMDIST(D4;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E4;\$C\$12;\$C\$13;TRUE)	=G4-F4	=\$C\$10*H4	=(I4-C4)^2/I4
5	nad 20 do 25	165	20	25	=NORMDIST(D5;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E5;\$C\$12;\$C\$13;TRUE)	=G5-F5	=\$C\$10*H5	=(I5-C5)^2/I5
6	nad 25 do 30	101	25	30	=NORMDIST(D6;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E6;\$C\$12;\$C\$13;TRUE)	=G6-F6	=\$C\$10*H6	=(I6-C6)^2/I6
7	nad 30 do 35	36	30	35	=NORMDIST(D7;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E7;\$C\$12;\$C\$13;TRUE)	=G7-F7	=\$C\$10*H7	=(I7-C7)^2/I7
8	nad 35 do 40	9	35	40	=NORMDIST(D8;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E8;\$C\$12;\$C\$13;TRUE)	=G8-F8	=\$C\$10*H8	=(I8-C8)^2/I8
9	nad 40	4	40		=NORMDIST(D9;\$C\$12;\$C\$13;TRUE)	1	=G9-F9	=\$C\$10*H9	=(I9-C9)^2/I9
10		=SUM(C2:C9)						=SUM(H2:H9)	=SUM(J2:J9)
11						hi-kvadrat statistika			=SUM(J2:J9)
12	mi	22				alfa			0,05
13	sigma	6				SP			7
14						mejna vrednost hi-kvadrat statistike			=CHINV(0,05;J13)
15						p-vrednost izračunana s funkcijo CHIDIST			=CHIDIST(J11;J13)
16						p-vrednost izračunana s funkcijo CHITEST			=CHITEST(C2:C9;I2:I9)

	B	C	D	E	F	G	H	I	J
1	Debelina debela (cm)	Število dreves	Sp. meja	Zg. meja	P(X<sp. meja)	P(X<zg. meja)	P(razred)	Pričakovane frekvence	Prispevek k hi-kvadrat statistiki
2	10 in manj	8		10		0,0228	0,0228	11,38	1,00
3	nad 10 do 15	52	10	15	0,0228	0,1217	0,0989	49,46	0,13
4	nad 15 do 20	125	15	20	0,1217	0,3694	0,2478	123,88	0,01
5	nad 20 do 25	165	20	25	0,3694	0,6915	0,3220	161,01	0,10
6	nad 25 do 30	101	25	30	0,6915	0,9088	0,2173	108,66	0,54
7	nad 30 do 35	36	30	35	0,9088	0,9849	0,0761	38,04	0,11
8	nad 35 do 40	9	35	40	0,9849	0,9987	0,0138	6,89	0,65
9	nad 40	4	40		0,9987	1,0000	0,0013	0,67	18,38
10		500					1,0000	500,0	18,92
11						hi-kvadrat statistika			18,917
12	mi	22				alfa			0,050
13	sigma	6				SP			7
14						mejna vrednost hi-kvadrat statistike			14,067
15						p-vrednost izračunana s funkcijo CHIDIST			0,0085
16						p-vrednost izračunana s funkcijo CHITEST			0,0085

Tabela 12.3: Izpis rezultatov hi-kvadrat preizkusa ničelne domneve o prileganju normalne porazdelitve N(22, 6) empirični porazdelitvi debeline debel; zadnja dva razreda sta združena

	A	B	C	D	E	F	G	H	I	J
18	Štev. razreda	Debelina debela (cm)	Število dreves	Sp. meja	Zg. meja	P(X<sp. meja)	P(X<zg. meja)	P(razred)	Pričakovane frekvence	Prispevek k hi-kvadrat statistiki
19	1	10 in manj	8		10		0,0228	0,0228	11,38	1,00
20	2	nad 10 do 15	52	10	15	0,0228	0,1217	0,0989	49,46	0,13
21	3	nad 15 do 20	125	15	20	0,1217	0,3694	0,2478	123,88	0,01
22	4	nad 20 do 25	165	20	25	0,3694	0,6915	0,3220	161,01	0,10
23	5	nad 25 do 30	101	25	30	0,6915	0,9088	0,2173	108,66	0,54
24	6	nad 30 do 35	36	30	35	0,9088	0,9849	0,0761	38,04	0,11
25	7	nad 35	13	35	40	0,9849	1,0000	0,0151	7,57	3,90
26			500					1,0000	500,0	5,80
27						hi-kvadrat statistika				5,795
28						alfa				0,050
29						SP				6
30						mejna vrednost hi-kvadrat statistike				12,592
31						p-vrednost izračunana s funkcijo CHIDIST				0,4465
32						p-vrednost izračunana s funkcijo CHITEST				0,4465

DEBELINA DEBEL 6

Po spremembi vrednosti za μ in σ v prejšnji nalogi si oglejte ujemanje dejanskih in pričakovanih frekvenc na sliki 12.1. ♦

Domača naloga 12.1: Porazdelitev mase lešnikov

V datoteki HI KVADRAT.XLS na delovnem listu »lešniki« so podatki o masi (g) za 70 lešnikov sorte 'Istrska dolgoplodna leska'.

LEŠNIKI 1

Podatke o masi lešnikov razporedite v razrede širine 0,25 g. Pri tem uporabite orodje **Vrtilna tabela**. Dobljene frekvence kopirajte na nov delovni list z imenom »lešniki-hi kvadrat«. ♦

	A	B	C	D	E	F	G	H	I	J
1	Tule spustite polja strani									
2										
3	Štej od Masa (g)									
4	Masa (g)	Vsota								
5	1,7-1,95	2								
6	1,95-2,2	7								
7	2,2-2,45	23								
8	2,45-2,7	23								
9	2,7-2,95	14								
10	2,95-3,2	1								
11	Skupna vsota	70								
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										

LEŠNIKI 2

Ob predpostavki, da je masa lešnikov porazdeljena normalno $N(2,5, 0,25)$, izračunajte verjetnost, da je masa lešnikov v posameznem razredu empirične porazdelitve (prvi razred naj bo odprt navzdol in zadnji razred naj bo odprt navzgor). ♦

LEŠNIKI 3

Koliko od 70 lešnikov bi bilo v posameznem razredu, če bi bila masa lešnikov porazdeljena po normalni porazdelitvi $N(2,5, 0,25)$? Dobljene pričakovane frekvence primerjajte z dejanskimi. ♦

LEŠNIKI 4

Grafično prikažite dejanske in pričakovane frekvence s histogramom in poligonom. ♦

LEŠNIKI 5

Ali empirična porazdelitev lešnikov glede na maso potrjuje predpostavko, da je masa lešnikov porazdeljena normalno s parametroma $\mu = 2,5$ g in $\sigma = 0,25$ cm ($\alpha = 0,05$)? Napišite ničelno in alternativno domnevo ter naredite statistični preizkus. Obrazložite rezultate.

Rešitev:

Po združitvi razredov: $k = 4$, $\chi^2 = 0,520$ in $p = 0,9144$. ♦

Domača naloga 12.2: Rumen in rdeč paradižnik

Če rdeč paradižnik križamo z rumenim, dobimo v F1 generaciji same rdeče paradižnike. V F2 generaciji pričakujemo $\frac{3}{4}$ rdečih in $\frac{1}{4}$ rumenih paradižnikov. V poskus je bilo vključenih 400 rastlin F2 generacije. Od teh je bilo 90 rumenih.

Ali rezultati poskusa potrjujejo zgornjo domnevo o razmerju med rdečimi in rumenimi paradižniki v F2 generaciji ($\alpha = 0,01$)? ♦

Zgled 12.3: Kakovost izdelanih zgoščenk (CD-jev)

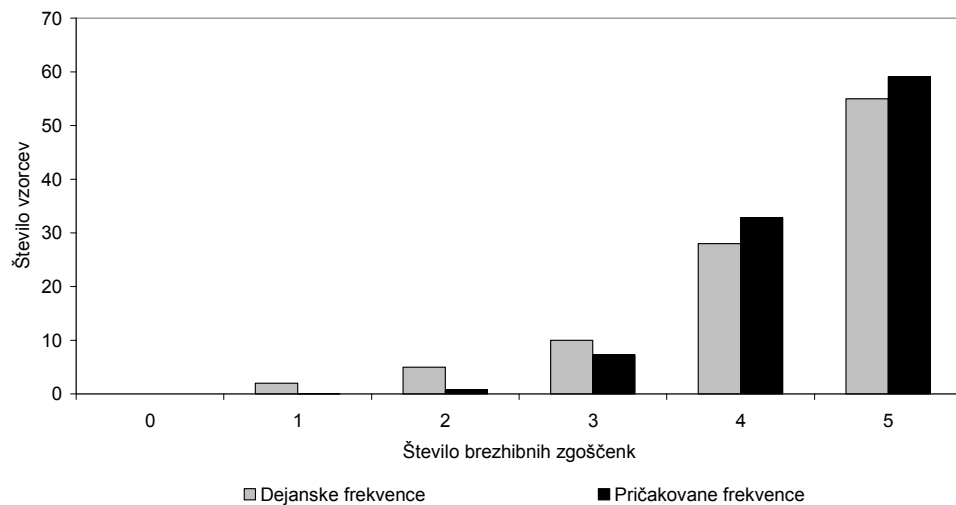
V tovarni zgoščenk trdijo, da je verjetnost, da je izdelana zgoščenska brezhibna, enaka 0,9. V oddelku za kontrolo kakovosti izberejo s slučajno izbiro v kontrolni vzorec 5 zgoščenk. Za vsako zgoščenko ugotovijo, ali je brezhibna ali ne. V kontrolnem vzorcu je lahko 0, 1, 2, 3, 4, 5 brezhibnih zgoščenk. Pregledali so 100 kontrolnih vzorcev in ugotovili naslednje stanje:

Število brezhibnih zgoščenk v kontrolnem vzorcu velikosti 5	Število vzorcev
0	0
1	2
2	5
3	10
4	28
5	55

CD 1

Podatke vpišite v pripravljeno tabelo na delovnem listu »kakovost zgoščenk« in grafično prikažite porazdelitev vzorcev glede na število brezhibnih zgoščenk (sivi stolpci na sliki 12.2).

♦



Slika 12.2: Porazdelitev vzorcev velikosti $n = 5$ glede na število brezhibnih zgoščenk

CD 2

Za spremenljivko X , ki označuje število brezhibnih zgoščenk v vzorcu velikosti $n = 5$, lahko privzamemo binomsko porazdelitev $b(5, p = 0,9)$. Izračunajte verjetnostno shemo za X (tabela 12.4).

Navodilo: uporabite funkcijo BINOMDIST. ♦

CD 3

Izračunajte pričakovano število vzorcev velikosti $n = 5$ z 0, 1, ..., 5 brezhibnimi zgoščenkami, če imamo skupaj 100 vzorcev (tabela 12.4). Pričakovane frekvence dodajte na sliko 12.2 (črni stolpci).

CD 4

Pri stopnji značilnosti 0,01 preverite trditev, da je verjetnost brezhibne zgoščenske 0,9. Zapišite ničelno in alternativno domnevo. Bodite pozorni na velikost pričakovanih frekvenc. Obrazložite rezultate (tabela 12.4).

Tabela 12.4: Podatki in izračuni za preverjanje trditve, da je verjetnost za brezhibnost zgoščenke enaka 0,9; izpis formul (zgoraj); izpis rezultatov (spodaj). Pričakovane frekvence prvih treh možnih izidov ($x = 0, 1$ ali 2) so premajhne, zato v drugem koraku prve štiri izide združimo v en razred: “tri brezhibne zgoščenke ali manj”.

	A	B	C	D	E
1	p_0	0,9			
2	n	5			
3	število vzorcev	100			
4	x	Dejanske frekvence	$p(x)$	Pričakovane frekvence	Prispevek k hi-kvadrat statistiki
5	0	0	=BINOMDIST(A5,\$B\$2,\$B\$1,FALSE)	=C5*\$B\$3	=(B5-D5)^2/D5
6	1	2	=BINOMDIST(A6,\$B\$2,\$B\$1,FALSE)	=C6*\$B\$3	=(B6-D6)^2/D6
7	2	5	=BINOMDIST(A7,\$B\$2,\$B\$1,FALSE)	=C7*\$B\$3	=(B7-D7)^2/D7
8	3	10	=BINOMDIST(A8,\$B\$2,\$B\$1,FALSE)	=C8*\$B\$3	=(B8-D8)^2/D8
9	4	28	=BINOMDIST(A9,\$B\$2,\$B\$1,FALSE)	=C9*\$B\$3	=(B9-D9)^2/D9
10	5	55	=BINOMDIST(A10,\$B\$2,\$B\$1,FALSE)	=C10*\$B\$3	=(B10-D10)^2/D10
11	Vsota	=SUM(B5:B10)	=SUM(C5:C10)	=SUM(D5:D10)	=SUM(E5:E10)

	A	B	C	D	E
1	p_0	0,9			
2	n	5			
3	število vzorcev	100			
4	x	Dejanske frekvence	$p(x)$	Pričakovane frekvence	Prispevek k hi-kvadrat statistiki
5	0	0	0,00001	0,00	0,001
6	1	2	0,00045	0,05	84,934
7	2	5	0,00810	0,81	21,674
8	3	10	0,07290	7,29	1,007
9	4	28	0,32805	32,81	0,704
10	5	55	0,59049	59,05	0,278
11	Vsota	100	1	100	108,598
12					
13	Združimo prve štiri razrede				
14	x	Dejanske frekvence	$p(x)$	Pričakovane frekvence	Prispevek k hi-kvadrat statistiki
15	0, 1, 2, 3	17	0,08146	8,15	9,624
16	4	28	0,32805	32,81	0,704
17	5	55	0,59049	59,05	0,278
18	Vsota	100	1	100	10,605
19		hi-kvadrat statistika			10,605
20		alfa			0,01
21		SP			2
22		mejna vrednost hi-kvadrat statistike			5,991
23		p-vrednost izračunana s funkcijo CHIDIST			0,0050
24		p-vrednost izračunana s funkcijo CHITEST			0,0050

Rešitev:

$$H_0 : p_0 = 0,9 \text{ in } H_1 : p_0 \neq 0,9$$

$$k = 3, \chi^2 = 10,605 \text{ in } p = 0,0050$$

Ničelno domnevo zavrnamo. Pri stopnji značilnosti $\alpha = 0,01$ trdimo, da verjetnost brezhibne zgoščenke ni 0,9 ($p = 0,0050$). ♦

CD 5

Izračunajte oceno za verjetnost brezhibne zgoščenke.

Rešitev:
$$\hat{p} = \frac{1 \cdot 2 + 2 \cdot 5 + 3 \cdot 10 + 4 \cdot 28 + 5 \cdot 55}{500} = 0,858$$

Podatki nakazujejo, da je verjetnost brezhibne zgoščenke precenjena. ♦

12.2 Hi-kvadrat za preverjanje povezanosti opisnih spremenljivk

Imamo dve opisni spremenljivki, X ima k vrednosti, Y pa r vrednosti. V vzorcu imamo n enot, za vsako enoto imamo vrednost za X in vrednost za Y . Podatke uredimo v dvorazsežno kontingenčno tabelo:

X	Y				Skupaj
	y_1	...	y_j	...	
x_1	f_{11}		f_{1j}		$f_{1\bullet}$
...					
x_i	f_{i1}		f_{ij}		$f_{i\bullet}$
...					
x_k	f_{k1}		f_{kj}		$f_{k\bullet}$
Skupaj	$f_{\bullet 1}$		$f_{\bullet j}$		$f_{\bullet r}$
					n

f_{ij} je število enot v i -ti vrstici in v j -tem stolpcu oz. število enot, ki imajo vrednost spremenljivke X enako x_i in vrednost spremenljivke Y enako y_j . $f_{i\bullet}$ je robna vsota i -te vrstice, $f_{i\bullet} = \sum_{j=1}^r f_{ij}$, $i = 1, \dots, k$; $f_{\bullet j}$ je robna vsota j -tega stolpca, $f_{\bullet j} = \sum_{i=1}^k f_{ij}$,

$j = 1, \dots, r$. Vsota vseh frekvenc je enaka številu enot v vzorcu: $n = \sum_{i=1}^k \sum_{j=1}^r f_{ij}$.

$n = \sum_{i=1}^k \sum_{j=1}^r f_{ij}$.

H_0 : X in Y nista povezani

H_1 : X in Y sta povezani

Pričakovano frekvenco f'_{ij} za primer, ko ničelna domneva velja, izrazimo z robnimi frekvencami iz kontingenčne tabele:

$$f'_{ij} = \frac{f_{i\cdot} \cdot f_{\cdot j}}{n}, \quad i = 1, \dots, k, \quad j = 1, \dots, r.$$

Mera za ujemanje dejanskih in pričakovanih frekvenc je tudi tukaj Pearsonova hi-kvadrat statistika:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^r \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}}.$$

Njena ničelna porazdelitev je $\chi^2 \approx \chi^2_{\alpha}(SP = (k-1)(r-1))$. H_0 obdržimo, če je $\chi^2 \leq \chi^2_{\alpha}(SP = (k-1)(r-1))$ in zavrnemo v korist H_1 , če je $\chi^2 > \chi^2_{\alpha}(SP = (k-1)(r-1))$.

Pogoj za uporabo Pearsonove hi-kvadrat statistike je: $f'_{ij} \geq 5$, $i = 1, \dots, k$, $j = 1, \dots, r$.

Zgled 12.4: Nastanitev in vožnja z avtom za študente BF

V datoteki ANKETA.XLS imamo za študentke in študente drugega letnika Agronomije, Biotehnologije in Mikrobiologije med drugim tudi podatke o nastanitvi in o tem, ali se na fakulteto vozijo z avtom. Zanima nas, ali sta spremenljivki *nastanitev* in *avto* povezani?

ANKETA 39

Z uporabo vrtilne tabele v datoteki ANKETA.XLS na delovnem listu »podatki za R« oblikujte kontingenčno tabelo na podlagi spremenljivk *nastanitev* in *avto*. Koliko različnih vrednosti imata spremenljivki? ♦

	A	B	C	D	E
1	Tule spustite polja strani				
2					
3	Štej od ID	avto			
4	nastanitev	da	ne	obcasno	Skupna vsota
5	doma	23	10	18	51
6	podnajemnik	3	29	8	40
7	studentski dom	2	28	2	32
8	(prazen)	2			2
9	lastnik		1		1
10	Skupna vsota	30	68	28	126

ANKETA 40

Za spremenljivko *nastanitev* združite vrednosti »doma« in »lastnik« v razred »Skupina1« ter izločite zapise, ki ne vsebujejo podatka o nastanitvi:

	A	B	C	D	E
1	Tule spustite polja strani				
2					
3	Štej od ID	avto			
4	nastanitev2	da	ne	obcasno	Skupna vsota
5	Skupina1	23	11	18	52
6	podnajemnik	3	29	8	40
7	studentski dom	2	28	2	32
8	Skupna vsota	28	68	28	124

Vrednosti tako oblikovane vrtilne tabele parkirajte v datoteko HI KVADRAT.XLS na delovni list »nastanitev_avto«. ♦

ANKETA 41

Ugotovite, ali sta spremenljivki *nastanitev* in *avto* povezani ($\alpha = 0,05$).

Navodilo: izračune naredite v pripravljeno tabelo na delovnem listu »nastanitev_avto«. Najprej izračunajte pričakovane frekvence, nato pa prispevke k hi-kvadrat statistiki za vsako celico posebej.

Kako mora biti napisana prva formula v tabelah (celici B10 in B17), da jo lahko kopiramo v ostale celice tabele in dobimo ustrezne izračune (tabela 12.5)?

Rešitev:

H_0 : *nastanitev* in *avto* nista povezani spremenljivki

H_1 : *nastanitev* in *avto* sta povezani spremenljivki

$$\chi^2 = 45,09 \quad p = 0,0000 \text{ (tabela 12.6)}$$

Ničelno domnevo pri $\alpha = 0,05$ zavrnemo v korist alternativni domnevi. Brez tveganja trdimo, da sta spremenljivki *nastanitev* in *avto* povezani ($p = 0,0000$). Povezava se najbolj kaže v tem, da se več študentov, ki stanujejo doma ali so lastniki stanovanja, vozi z avtom, kot bi pričakovali, če povezave med spremenljivkama *nastanitev* in *avto* ne bi bilo. Največji prispevek k hi-kvadrat statistiki doprinesejo študentje, ki stanujejo doma oz. so lastniki stanovanja. Med njimi je več, kot bi pričakovali, če povezave med spremenljivkama ne bi bilo, takih, ki se vozijo z avtom redno (10,79), po drugi strani pa je manj - kot bi pričakovali - takih, ki se ne vozijo z avtom (10,76). Obratna situacija pa je pri tistih, ki stanujejo v študentskem domu. Pri teh je več takih, ki se ne vozijo z avtom, kot bi pričakovali, če povezave med spremenljivkama ne bi bilo. Prispevek te skupine študentov k hi-kvadrat statistiki je 6,22.

Tabela 12.5: Kontingenčna tabela, pričakovane frekvence in prispevki k hi-kvadrat statistiki za spremenljivki *nastanitev* in *avto*; izpis formul(zgoraj); izpis rezultatov (spodaj)

	A	B	C	D	E
1	Dejanske frekvence	Ali se vozite na fakulteto z avtom?			
2	Nastanitev	da	ne	občasno	Skupaj
3	doma ali lastnik	23	11	18	52
4	podnajemnik	3	29	8	40
5	studentski dom	2	28	2	32
6	Skupaj	28	68	28	124
7					
8	Pričakovane frekvence	Ali se vozite na fakulteto z avtom?			
9	Nastanitev	da	ne	občasno	Skupaj
10	doma ali lastnik	=E3*B\$6/\$E\$6	=E3*C\$6/\$E\$6	=E3*D\$6/\$E\$6	=SUM(B10:D10)
11	podnajemnik	=E4*B\$6/\$E\$6	=E4*C\$6/\$E\$6	=E4*D\$6/\$E\$6	=SUM(B11:D11)
12	studentski dom	=E5*B\$6/\$E\$6	=E5*C\$6/\$E\$6	=E5*D\$6/\$E\$6	=SUM(B12:D12)
13	Skupaj	=SUM(B10:B12)	=SUM(C10:C12)	=SUM(D10:D12)	=SUM(E10:E12)
14					
15	Prispevki k hi-kvadrat statistiki	Ali se vozite na fakulteto z avtom?			
16	Nastanitev	da	ne	občasno	Skupaj
17	doma ali lastnik	=(B3-B10)^2/B10	=(C3-C10)^2/C10	=(D3-D10)^2/D10	=SUM(B17:D17)
18	podnajemnik	=(B4-B11)^2/B11	=(C4-C11)^2/C11	=(D4-D11)^2/D11	=SUM(B18:D18)
19	studentski dom	=(B5-B12)^2/B12	=(C5-C12)^2/C12	=(D5-D12)^2/D12	=SUM(B19:D19)
20	Skupaj	=SUM(B17:B19)	=SUM(C17:C19)	=SUM(D17:D19)	=SUM(E17:E19)

	A	B	C	D	E
8	Pričakovane frekvence	Ali se vozite na fakulteto z avtom?			
9	Nastanitev	da	ne	občasno	Skupaj
10	doma ali lastnik	11,7	28,5	11,7	52,0
11	podnajemnik	9,0	21,9	9,0	40,0
12	studentski dom	7,2	17,5	7,2	32,0
13	Skupaj	28,0	68,0	28,0	124,0
14					
15	Prispevki k hi-kvadrat statistiki	Ali se vozite na fakulteto z avtom?			
16	Nastanitev	da	ne	občasno	Skupaj
17	doma ali lastnik	10,79	10,76	3,34	24,89
18	podnajemnik	4,03	2,28	0,12	6,42
19	studentski dom	3,78	6,22	3,78	13,78
20	Skupaj	18,60	19,26	7,23	45,09

 Tabela 12.6: Hi-kvadrat preizkus za povezanost spremenljivk *nastanitev* in *avto*; izpis formul (levo); izpis rezultatov (desno)

	A	B
23	H_0 : <i>nastanitev</i> in <i>avto</i> nista povezani spremenljivki	
24	H_1 : <i>nastanitev</i> in <i>avto</i> sta povezani spremenljivki	
25	alfa	0.05
26	hi-kvadrat statistika	=E20
27	število vrednosti spremenljivke <i>nastanitev</i> (k)	3
28	število vrednosti spremenljivke <i>avto</i> (r)	3
29	SP	=(B27-1)*(B28-1)
30	hi_kvadrat _{mejna}	=CHIINV(B25;B29)
31	p-vrednost CHIDIST	=CHIDIST(B26;B29)
32	p-vrednost CHITEST	=CHITEST(B3:D5;B10:D12)

	B
23	
24	
25	0.05
26	45.09
27	3
28	3
29	4
30	9.49
31	3.8E-09
32	3.8E-09

12.3 Hi-kvadrat za preverjanje homogenosti struktur

Kontingenčna tabela ima v primeru preverjanja homogenosti struktur posebno obliko. Z zasnovo študije je dana robna vrstica, vsaka njena celica predstavlja velikost slučajnega vzorca določene velikosti. Imamo r vzorcev, ki so med sabo neodvisni. Za enote vseh vzorcev imamo vrednosti opisne spremenljivke X . Vrednosti te spremenljivke določajo ostale vrstice kontingenčne tabele. Podatek v j -tem stolpcu in i -ti vrstici je število enot v j -tem vzorcu, katerih vrednost spremenljivke X je x_i . Ničelno in alternativno domnevo zapišemo kot:

H_0 : struktura po stolpcih je enaka.

H_1 : struktura po stolpcih ni enaka.

Pričakovane frekvence za primer, da ničelna domneva velja, izrazimo iz robnih vsot kontingenčne tabele (tako kot pri preverjanju povezanosti dveh opisnih spremenljivk):

$$f'_{ij} = \frac{f_{i\cdot} \cdot f_{\cdot j}}{n}, \quad i = 1, \dots, k, \quad j = 1, \dots, r.$$

Mera za ujemanje dejanskih in pričakovanih frekvenc je tudi tu Pearsonova hi-kvadrat

statistika
$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^r \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}}.$$

Njena ničelna porazdelitev je $\chi^2 \approx \chi^2_{\alpha}(SP = (k-1)(r-1))$. H_0 obdržimo, če je $\chi^2 \leq \chi^2_{\alpha}(SP = (k-1)(r-1))$ in zavrnemo v korist H_1 , če je $\chi^2 > \chi^2_{\alpha}(SP = (k-1)(r-1))$.

Pogoj za uporabo Pearsonove hi-kvadrat statistike je: $f'_{ij} \geq 5$, $i = 1, \dots, k$, $j = 1, \dots, r$.

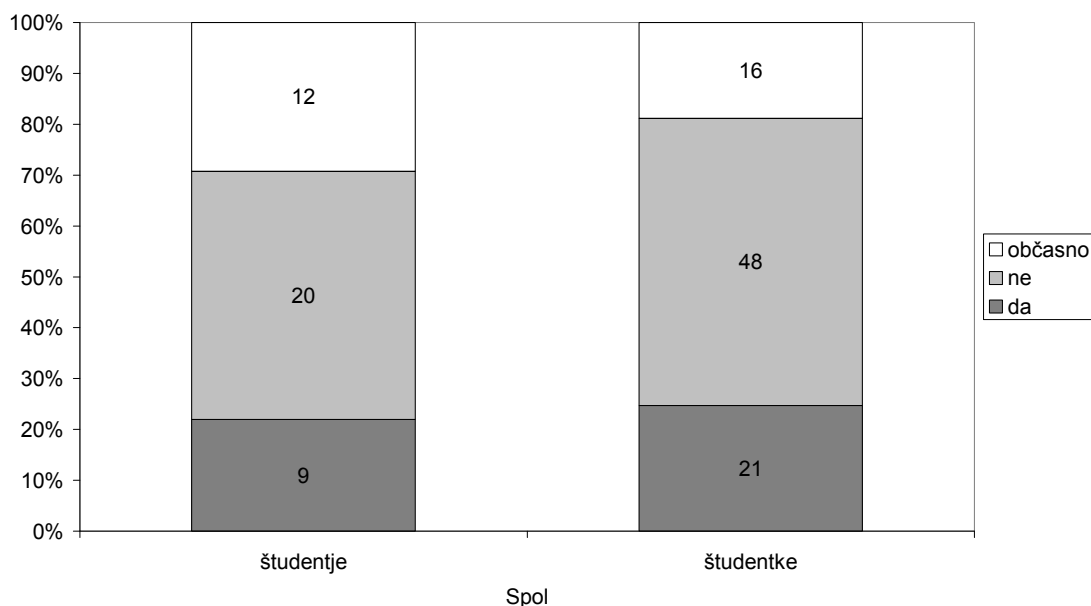
Zgled 12.5: Struktura študentov in študentk glede na prevoz z avtom

ANKETA 42

Na podlagi podatkov v datoteki ANKETA.XLS, delovni list »podatki za R« oblikujte kontingenčno tabelo za spremenljivki *spol* in *avto*. Stolpce naj določa spremenljivka *spol*, vrstice pa *avto*. Vrednosti iz vrtilne tabele parkirajte v tabelo na delovnem listu »spol_avto« v datoteki HI KVADRAT.XLS (tabela 12.7). ♦

ANKETA 43

S strukturnima stolpcema prikažite strukturo študentk in študentov glede na prevoz na fakulteto z avtom (slika 12.3). ♦



Slika 12.3: Struktura študentov in študentk glede na to, ali se na fakulteto vozijo z avtom. Številke na stolpcih predstavljajo frekvence.

ANKETA 44

Privzeli bomo, da imamo opravka s slučajnima vzorcema študentov in študentk. Preverite, ali sta strukturi študentov in študentk glede na to, ali se vozijo na fakulteto z osebnim avtomobilom, enaki ($\alpha = 0,05$) (tabela 12.7). Obrazložite rezultate.

Rešitev: $\chi^2 = 1,75$ $p = 0,4171$

Ničelno domnevo obdržimo ($p = 0,4171$). Podatki nakazujejo, da sta strukturi študentov in študentk glede na prevoz homogeni (enaki). ♦

Tabela 12.7: Izračuni hi-kvadrat preizkusa za homogenost struktur študentk in študentov glede na to, ali se na fakulteto vozijo z avtom

	A	B	C	D
1	Dejanske frekvence	spol		
2	avto	študentje	študentke	Skupaj
3	da	9	21	30
4	ne	20	48	68
5	občasno	12	16	28
6	Skupaj	41	85	126
7	Pričakovane frekvence	spol		
8	avto	študentje	študentke	Skupaj
9	da	9.8	20.2	30.0
10	ne	22.1	45.9	68.0
11	občasno	9.1	18.9	28.0
12	Skupaj	41.0	85.0	126.0
13	Prispevki k hi-kvadrat statistiki	spol		
14	avto	študentje	študentke	Skupaj
15	da	0.06	0.03	0.09
16	ne	0.20	0.10	0.30
17	občasno	0.92	0.44	1.36
18	Skupaj	1.18	0.57	1.75
19	H_0 : za študente in študentke je struktura po tem, ali se na fakulteto vozijo z avtom, enaka			
20	H_1 : za študente in študentke struktura po tem, ali se na fakulteto vozijo z avtom, ni enaka			
21	alfa	0.05		
22	hi-kvadrat statistika	1.75		
23	število vrednosti spremenljivke <i>avto</i> (k)	3		
24	število vrednosti spremenljivke <i>spol</i> (r)	2		
25	SP	2		
26	hi_kvadrat _{mejna}	5.99		
27	p-vrednost CHIDIST	0.4171		
28	p-vrednost CHITEST	0.4171		

Domača naloga 12.3: Struktura študentov AG, BT in MB glede na to, ali prejemajo štipendijo.

ANKETA 45

Na podlagi podatkov v datoteki ANKETA.XLS s hi-kvadrat preizkusom preverite domnevo, da je delež štipendistov pri vseh treh smereh študija (AG, BT in MB) enak ($\alpha = 0,05$). Obrazložite rezultate.

Navodila: z vrtilno tabelo oblikovano kontingenčno tabelo kopirajte v datoteko HI KVADRAT.XLS na nov delovni list z imenom »smer študija_štipendija«. Na tem delovnem listu naredite vse potrebne izračune. Ali si lahko pomagate s kopiranjem formul iz prejšnjih zgledov v tem poglavju?

Rešitev: $\chi^2 = 6,15$ $p = 0,0462$ ♦

ANKETA 46

Strukturo treh smeri študija glede na to, ali imajo štipendijo ali ne, grafično predstavite. Delovni list v datoteki HI KVADRAT.XLS poimenujte »Slika 12.4«. ♦

13 DODATEK

13.1 Uporaba kalkulatorja za izračun povprečja in standardnega odklona

S kalkulatorjem enostavno izračunamo vzorčno povprečje \bar{x} , vzorčni standardni odklon s in populacijski standardni odklon σ .

Opozoriti moramo, da različni proizvajalci kalkulatorjev uporabljajo različne oznake, kar zahteva, da se seznanimo z vsakim kalkulatorjem posebej.

Postopek:

- na kalkulatorju izberemo tipko MODE in vključimo osnovni statistični način: SD, STAT ali STAT1;
- v spomin kalkulatorja zaporedoma shranimo n vrednosti izbrane spremenljivke (n je velikost vzorca): natipkamo prvo število in jo s tipko **DATA** ali **M+** shranimo v spomin, natipkamo drugo število in nato tipko **DATA**, ... Vnos podatkov lahko prekinemo z izpisom delnih rezultatov in kasneje nadaljujemo z vnosom podatkov;
- vzorčno povprečje dobimo s pritiskom na tipko, ki vrne rezultat funkcije $\frac{1}{n} \sum_{i=1}^n x_i$ in je nad ali pod tipko označena z \bar{x} .
- vzorčni standardni odklon $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ dobimo s tipko, označeno na različnih kalkulatorjih na različne načine: sx , s_x , s , $\sigma x n - 1$, σx_{n-1} , σ_{n-1}, \dots
- populacijski standardni odklon $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ dobimo s tipko, označeno na različnih kalkulatorjih na različne načine: σx , σ_x , σ , $\sigma x n$, σx_n , σ_n, \dots

- vzorčno oz. populacijsko varianco dobimo s kvadriranjem vzorčnega oz. populacijskega standardnega odklona.

Preden začnemo z izračuni za novi niz podatkov je potrebno sprazniti spomin statističnega načina računanja. To naredimo tako, da se vrnemo v navaden način (CALC, COMP, 0, ...), nekateri kalkulatorji pa za to uporabljajo funkcijo *Scl*.

13.2 Uporaba kalkulatorja za linearno regresijo in korelacijo

Za izračun ocen parametrov enostavnega linearnega regresijskega modela a , b in za izračun Pearsonovega koeficienta korelacije r uporabimo statistični način (MODE) za regresijo.²⁷

Postopek:

- na kalkulatorju vključimo regresijski statistični način (MODE), običajne oznake tega načina so $\boxed{a+bx}$, \boxed{REG} , $\boxed{STAT 2}$, ...
- najprej vnesemo pare podatkov (x_i, y_i) , $i = 1, \dots, n$; postopek vnosa podatkov je naslednji: vnesemo vrednost za x_i , pritisnemo tipko $\boxed{=}$ ali \boxed{STO} ali $\boxed{(x,y)}$, ..., nato vnesemo še y_i ter na koncu par podatkov shranimo v spomin kalkulatorja s pritiskom na tipko \boxed{DATA} , \boxed{DT} ;
- s pritiskom na izbrane tipke kalkulator izračuna oceni parametrov a in b in koeficient korelacije r . Koeficient determinacije r^2 v linearni regresiji je kvadrat koeficienta korelacije r . Dobimo tudi naslednje vsote: $\sum x$, $\sum x^2$, $\sum y$, $\sum y^2$ in $\sum xy$.

Spomin statističnega načina računanja spraznimo tako, da se vrnemo v navaden način (CALC), nekateri kalkulatorji pa za to uporabljajo funkcijo *Scl*.

²⁷ Enostavni kalkulatorji tega načina nimajo.