

Damijana Kastelec in Katarina Košmelj

# **OSNOVE STATISTIKE Z EXCELOM 2007**

September 2010



Recenzenta: prof. dr. Andrej Blejec, doc. dr. Gaj Vidmar

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

311:004.42(0.034.2)

KASTELEC, Damijana

Osnove statistike z Excelom 2007 [Elektronski vir] / Damijana  
Kastelec in Katarina Košmelj. - El. knjiga. - Ljubljana :  
Biotehniška fakulteta, 2010

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/osnove-statistike-z-excelom/>

ISBN 978-961-6379-15-1

1. Košmelj, Katarina

252837888

Copyright ©Damijana Kastelec in Katarina Košmelj, 2010

## **PREDGOVOR**

Knjiga je nastala na podlagi predhodnega učbenika Statistična analiza podatkov s programoma Excel 2003 in R in je zasnovana kot učbenik, vezan na pouk statistike na različnih študijskih programih. Nekatera poglavja so vsebinsko dopolnjena, predvsem pa so na novo napisana navodila za delo z Excelom 2007, ki se v razporeditvi ukazov in njihovi grafični predstavitvi precej razlikuje od predhodnih verzij. V knjigi za statistično analizo podatkov obravnavamo program Excel 2007 brez dodatkov, ki jih za posamezne sklope statistične analize lahko najdemo na različnih spletnih straneh, nekateri so zastoj, je pa tudi več plačljivih.

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 grafičnih prikazih podatkov in rezultatov.

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**

<b>1</b>	<b>Grafični prikazi in računanje.....</b>	<b>4</b>
1.1	Uvod.....	4
1.2	Datoteka v Excelu - zvezek.....	5
1.2.1	Vnos podatkov v celico.....	7
1.2.2	Spreminjanje vsebine v celicah.....	8
1.3	Kartica <i>Osnovno</i> .....	9
1.3.1	Kopiranje in lepljenje.....	9
1.3.2	Oblikovanje vsebine celice.....	10
1.3.3	Celice.....	11
1.3.4	Urejanje vsebine celic.....	14
1.4	Grafično prikazovanje podatkov.....	15
1.4.1	Orodja za grafikone, kartica <i>Načrt</i> .....	17
1.4.2	Orodja za grafikone, kartici <i>Postavitve</i> in <i>Oblika</i> .....	19
1.5	Formule.....	25
1.5.1	Kopiranje formul.....	26
1.5.2	Vstavljanje funkcij.....	27
1.5.3	Lepljenje na poseben način.....	29
<b>2</b>	<b>Relativna števila in grafični prikazi.....</b>	<b>31</b>
2.1	Oblikovanje osi grafikona.....	31
2.2	Oblikovanje niza podatkov.....	39
2.3	Koeficienti in njihovi grafični prikazi, povprečni koeficient.....	41
2.4	Časovne vrste, indeksi, stopnje rasti in gostote ter njihovi grafični prikazi.....	45
<b>3</b>	<b>Opisne statistike in frekvenčna porazdelitev.....</b>	<b>67</b>
3.1	Uporaba vgrajenih funkcij.....	67
3.1.1	Geometrijska sredina.....	69
3.2	Napake pri uporabi formul.....	71
3.3	Opisne statistike.....	71
3.3.1	Uporaba funkcij.....	71
3.3.2	Uporaba orodja Descriptive Statistics.....	73
3.4	Frekvenčna porazdelitev in kvantili.....	80

<b>4</b>	<b>Podatkovna zbirka in vrtilne tabele.....</b>	<b>90</b>
4.1	Podatkovna zbirka .....	90
4.2	Razvrščanje enot.....	92
4.3	Izbira zapisov .....	93
4.3.1	Filter .....	93
4.3.2	Napredni filter .....	98
4.4	Vrtilne tabele .....	99
4.4.1	Seznam polj vrtilne tabele .....	102
4.4.2	Primeri oblikovanja vrtilnih tabel.....	105
4.4.3	Orodja za vrtilne tabele .....	111
4.5	Vrtilni grafikon.....	125
4.6	Pogojno oblikovanje celic .....	127
<b>5</b>	<b>Normalna in binomska porazdelitev.....</b>	<b>149</b>
5.1	Normalna porazdelitev, funkciji NORMDIST in NORMINV .....	149
5.1.1	Gostota verjetnosti.....	150
5.1.2	Izračun verjetnosti .....	151
5.1.3	Izračun kvantilov .....	153
5.2	Generator slučajnih števil v Excelu.....	155
5.2.1	Generiranje vrednosti, porazdeljenih po normalni porazdelitvi .....	155
5.3	Binomska porazdelitev, funkcija BINOMDIST .....	158
5.4	Aproksimacija binomske porazdelitve z normalno porazdelitvijo.....	161
<b>6</b>	<b>Intervali zaupanja .....</b>	<b>164</b>
6.1	Interval zaupanja za povprečje .....	164
6.1.1	$\sigma$ je znana.....	164
6.1.2	$\sigma$ ni znana.....	166
6.2	Velikost vzorca za povprečje pri normalni porazdelitvi .....	174
6.2.1	Standardni odklon je znan .....	174
6.2.2	Standardni odklon ni znan – iterativni postopek.....	175
6.3	Interval zaupanja in velikost vzorca za $p$ pri binomski porazdelitvi $b(n, p)$ ...	177
<b>7</b>	<b>Preverjanje statističnih domnev za en vzorec.....</b>	<b>180</b>
7.1	Povprečje .....	180
7.2	Bernoullijeva verjetnost .....	187

<b>8</b>	<b>Preverjanje statističnih domnev za dva vzorca in interval zaupanja za razliko .....</b>	<b>192</b>
8.1	Razlika povprečij za neodvisna vzorca .....	192
8.2	Razlika povprečij za odvisna vzorca .....	201
8.3	Razlika Bernoullijevih verjetnosti za neodvisna vzorca .....	207
8.4	Razlika Bernoullijevih verjetnosti za odvisna vzorca .....	211
<b>9</b>	<b>Linearna regresija in korelacija.....</b>	<b>214</b>
9.1	Linearna regresija .....	214
9.1.1	Excelovo orodje za linearno regresijo .....	217
9.2	Korelacija .....	223
<b>10</b>	<b>Hi-kvadrat preizkusi .....</b>	<b>228</b>
10.1	Hi-kvadrat preizkus za preverjanje domneve o porazdelitvi.....	228
10.2	Hi-kvadrat za preverjanje povezanosti opisnih spremenljivk .....	241
10.3	Hi-kvadrat za preverjanje homogenosti struktur .....	245
<b>11</b>	<b>VIRI .....</b>	<b>249</b>
<b>12</b>	<b>DODATEK A.....</b>	<b>250</b>

## 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

#### Delovne datoteke:

PREHODNOST.xlsx

OKRASNE RASTLINE.xlsx

VPIS.xlsx

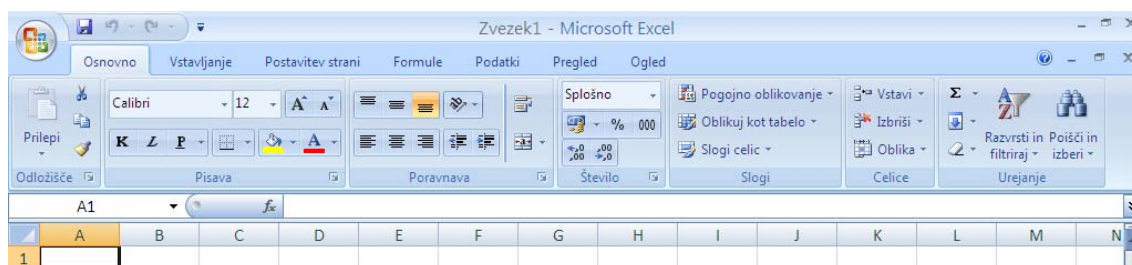
### 1.1 Uvod




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. V tej knjigi je predstavljen Microsoft Excel 2007, ki se od prejšnjih različic Excela kar precej razlikuje predvsem v načinu dostopa do posameznih ukazov. Spremembe so naredili tudi v taki smeri, da bi naredili ukaze v programu čim lažje dostopne. Menije ukazov in dve standardni orodni vrstici so zamenjali z **opravilnim trakom**, na katerem so različne skupine ikon, ki predstavljajo ukaze. Ko se z miško postavimo na posamezno ikono, se prikaže okence z razširjenim imenom ukaza in njegovo kratko obrazložitvijo. V nadaljevanju te knjige bomo pri predstavitvi ikone uporabljali izraz ukaz.

Skupine ikon so pri osnovni nastavitvi razdeljene na več kartic: Osnovno, Vstavljanje, Postavitev strani, Formule, Podatki, Pregled in Ogled. V posamezni skupini ukazov je na zaslonu vidnih toliko ikon, kolikor jih omogočajo grafične lastnosti zaslona. Ukazi,




ki niso vidni v obliki ikon, se pokažejo v pogovornem oknu, ki se odpre ob kliku puščice v naslovni vrstici skupine.



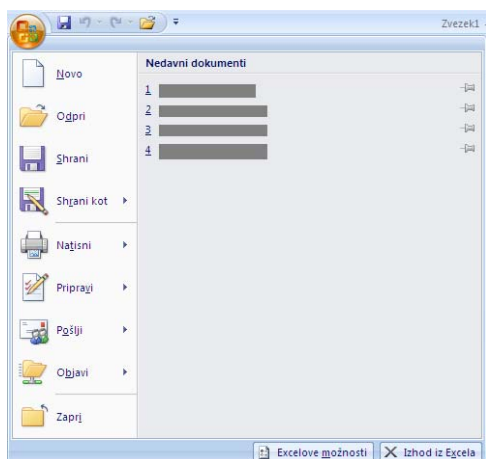
V zgornjem levem kotu opravljalnega traka je **gumb** , s klikom nanj se odpre meni z ukazi za delo z datotekami. Desno od tega gumba je **orodna vrstica za hitri dostop** , ki po osnovnih nastavitvah vsebuje le tri ukaze **Shrani**, **Razveljavi**, **Uveljavi**. To orodno vrstico lahko preprosto dopolnimo z ukazi, ki jih najpogosteje uporabljamo. Pod trakom s karticami je vnosna vrstica z imenom označene celice v preglednici (npr. A1), in ukazom **Vstavi funkcijo** . Pod vnosno vrstico je na več listih **preglednica** sestavljena iz celic, ki so poimenovane s črkami (stolpci) in številkami (vrstice), npr. A1. Ob spodnjem robu okna je **vrstica stanja**, ki podaja informacijo o trenutno izvajajoči se operaciji (npr. Pripravljen).

## 1.2 Datoteka v Excelu - zvezek

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 okna. Delovne liste lahko poljubno dodajamo, 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*...).

Delo z datotekami poteka z ukazi **Novo**, **Odpri**, **Zapri**, **Shrani**, **Shrani kot...**, ki jih najdemo v meniju gumba .

Datoteke, napisane v Excelu 2007, imajo končnico .xlsx (prejšnje verzije Excela pa .xls)<sup>1</sup>. Naenkrat je lahko odprtih več excelovih datotek. Tedaj se iz datoteke v datoteko prestavljamo s klikom na gumb z imenom datoteke v opravilni vrstici okolja Okna na dnu zaslona.



## PREHODNOST 1

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

Preglednico oz. tabelo na delovnem listu Excela sestavlja veliko število celic, ki imajo naslove določene z imeni stolpcev in vrstic: celica A1 stoji v prvem stolpcu in v prvi vrstici. Stolpci so označeni s črkami v **naslovni vrstici**, vrstice pa s številkami v **naslovnem stolpcu**. 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. Skupino celic označimo s potegom miške, pri tem se označene celice pobarvajo svetlo modro, samo celica, s katero smo začeli označevanje, ostane bela, skupina označenih celic dobi odebeljen črni okvir. 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, ima ime A1:C3. Celo vrstico oz. stolpec označimo s klikom polja naslova vrstice oz. stolpca. Naenkrat lahko označimo tudi celice, ki se ne

<sup>1</sup> V Excelu 2007 lahko datoteko shranimo tudi v obliki, ki jo lahko beremo in popravljamo v starejših različicah programa: v pogovornem oknu **Shrani kot** v polju **Shrani kot vrsto** izberemo možnost *Excelov delovni zvezek 97-2003 (\*.xls)*. Ob imenu tako shranjene datoteke se v naslovni vrstici okna v oklepaju izpiše opomba [Združljivostni način].

stikajo. Najprej označimo prvo skupino celic, pritisnemo tipko CTRL in jo držimo, medtem ko označujemo še preostale skupine celic.

Če v označeni celici hkrati vidimo tudi utripajoči kazalec, je **celica aktivna**. Celico aktiviramo z dvoklikom. Celico aktiviramo, če želimo popraviti njeno vsebino.

### 1.2.1 Vnos podatkov v celico

Podatek vnesemo v celico tako, da jo označimo, natipkamo njeno vsebino in vnos potrdimo s pritiskom na tipko ENTER. Če želimo vnos že natipkanega podatka preprečiti, pritisnemo na tipko ESC. V celice preglednice lahko vnašamo konstante ali formule. Konstante so: števila, datumi, čas, denarne vrednosti, odstotki, ulomki, 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*.<sup>2</sup> Po osnovni nastavitvi se ob vnosu podatkov, s katerimi zna Excel računati (številke, formule, datumi), le-ti poravnajo ob desni rob celice, opisne vrednosti pa ob levi rob celice. Tak način poravnave omogoča hitro kontrolo vrste vnesenih podatkov (število, besedilo). Če odtipkamo napačni znak za decimalno ločilo, Excel vneseni podatek razume kot besedilo in ne kot število<sup>3</sup>.

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 vnosni vrstici.

Če v celico vnašamo numerične vrednosti in se po kliku na gumb ENTER vsebina celice spremeni v #####, to pomeni, da je celica preozka za izpis in jo je potrebno razširiti.

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;

---

<sup>2</sup> Program Nadzorna plošča (angl. Control panel), podprogram Področne in jezikovne nastavitve (angl. Regional and Language options), v pogovornem oknu kliknemo na gumb Nastavi (angl. Customize).

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

- črn križec se pojavi, ko se z miško postavimo v desni spodnji vogal označenih celic v preglednici;
- črn križec s puščicami na koncih se pojavi, ko se z miško postavimo na rob označenih celic v preglednici (npr. kadar želimo vsebino označenih celic prestaviti drugam v preglednico).

Excel omogoča preklic zadnjih nekaj zaporednih ukazov oz. vnosov podatkov ali formul v celice z ukazom **Razveljavi**  v orodni vrstici za hitri dostop. Obstaja nekaj ukazov, ki jih ne moremo razveljaviti.

### 1.2.2 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 prekrije 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...). 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 popravimo.

#### PREHODNOST 2

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

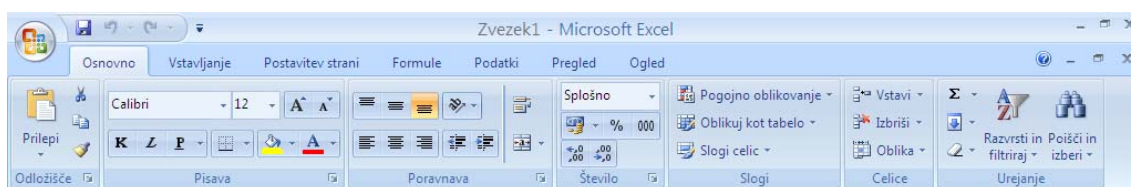
Tabela 1.1: Prehodnost generacije študentov 1. letnika v 2. letnik na univerzitetnih študijskih programih Biotehniške fakultete v šolskem letu 2002/2003 (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003)

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

### PREHODNOST 3

Označite tabelo 1.1, ki ste jo vnesli na delovni list v nalogi PREHODNOST 2. ♦

## 1.3 Kartica *Osnovno*



Kartica *Osnovno* vsebuje sedem skupin ukazov: Odložišče, Pisava, Poravnava, Število, Slogi, Celice in Urejanje. Ukazi so razporejeni glede na pričakovano pogostost njihove uporabe. Predvideno je, da se najpogosteje uporablja ukaze v skupini Odložišče in najmanj pogosto ukaze v skupini Urejanje. Znotraj skupin so ikone predvidoma pogosteje uporabljenih ukazov večje. Podrobneje bomo opisali samo ukaze, ki so pomembnejši v nadaljevanju te knjige, za ostale ukaze pa bralca vabimo, da si pomaga z razlago, ki se pojavi, če se z miško postavimo na ikono.

### 1.3.1 Kopiranje in lepljenje

V skupini Odložišče najdemo ukaze **Prilepi**, **Izreži**, **Kopiraj** in **Preslikovalnik oblik**. S prvimi tremi ukazi označene celice prilepimo, izrežemo, kopiramo - prestavljamo po preglednici, med delovnimi listi zvezka in med različnimi zvezki. Ko uporabimo ukaza **Izreži** ali **Kopiraj**, se okoli označenih celic pojavi črtkana utripajoča črta. Ta črta je prisotna, dokler je vsebina označenih celic v odložišču. Z ukazom **Prilepi** jo postavimo

na izbrano mesto v preglednici. To mesto določimo tako, da označimo celico, ki predstavlja položaj levega zgornjega vogala vsebine odložišča.

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

**Preslikovalnik oblik** uporabimo, kadar želimo posebne nastavitve oblik izbrane celice preslikati na druge celice.



PREHODNOST 4

Prvi in tretji stolpec v tabeli 1.1 kopirajte v celice pod tabelo na isti delovni list. Oglejte si položaj kopiranih celic, nato uporabite ukaz **Razveljavi**. ♦

PREHODNOST 5

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

### 1.3.2 Oblikovanje vsebine celice

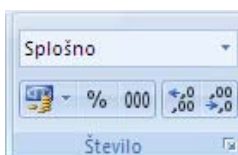
Na kartici **Osnovno** so v skupinah Pisava in Poravnava ukazi, ki omogočajo izbiro oblike, velikosti, barvo pisave, poravnavo vsebine znotraj celice, barvo ozadja vsebine v celici, obliko obrobe celic. V skupini Poravnava večkrat prideta prav ukaza **Prelomi besedilo**  in **Združi na sredino** .

PREHODNOST 6

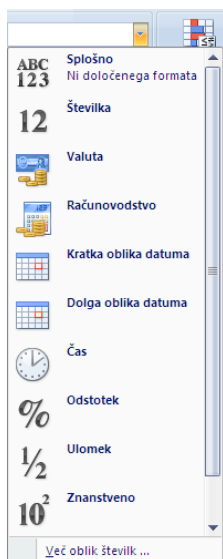
Besedilo v prvi vrstici tabele 1.1 naj bo izpisano krepko, ozadje celic naj bo sivo. Določite, naj se besedilo v celicah prve vrstice izpiše v več vrstic, če je stolpec preozek.

V tabeli naj bodo vse celice obrobljene s črto. ♦


Skupina ukazov **Število** je namenjena oblikovanju številskih vrednosti.



Oblika *Splošno* pomeni, da se vse številske vrednosti v celicah samodejno poravnajo ob desni rob, besedilo pa ob levi rob. Če odpremo padajoč seznam v tem polju, se odpre meni različnih možnih oblik vsebine celic:



Z ukazom **Več oblik števil...** se odpre pogovorno okno, ki vsebuje šest kartic in je enako pogovornemu oknu, ki smo ga v Excelu 2003 dobili z ukazom **Oblika/Celice**.

V tej skupini prideta večkrat prav ukaza **Povečaj** oz. **Zmanjšaj število decimalnih mest** .

Skupina ukazov **Slogi** je namenjena zahtevnejšemu načinu oblikovanja vsebine in oblike celic. V tej knjigi se s tem ne bomo ukvarjali.

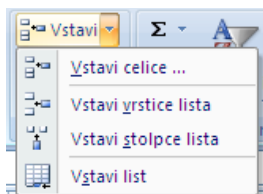
### 1.3.3 Celice



Na kartici **Osnovno** je skupina ukazov **Celice** namenjena določanju oblike, velikosti in vidljivosti celic, organiziranosti delovnih listov ter zaščiti. Tu sta tudi ukaza za vstavljanje praznih celic, vrstic in stolpcev (**Vstavi**), za njihovo brisanje (**Izbriši**).

### 1.3.3.1 Vstavljanje celic

V preglednico lahko vstavljamo posamezne prazne celice, vrstice, stolpce, skupine celic ali delovne liste z ukazom **Vstavi**.



Prazno vrstico vstavimo tako, da se v preglednici postavimo v vrstico, nad katero želimo vstaviti prazno vrstico in uporabimo ukaz **Vstavi vrstice lista**. S tem se vsebina celic v označeni in vseh naslednjih vrsticah premakne za eno vrstico navzdol. Hkrati lahko vrinemo tudi več vrstic. Označimo več vrstic, nato uporabimo ukaz **Vstavi vrstice lista**. Hitreje ta ukaz izvedemo, če na označene celice oz. vrstice kliknemo z desnim gumbom miške in v tako dobljenem meniju kliknemo ukaz **Vstavi**.

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

#### PREHODNOST 7

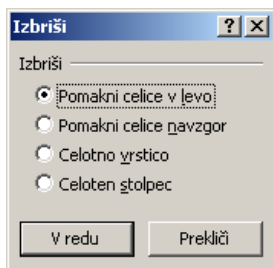
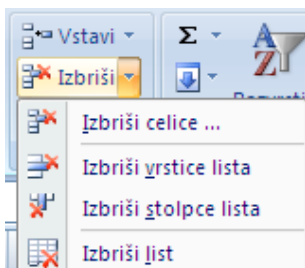
Vrinite v tabelo 1.1 vrstico med podatke za študijski program Biologija in Krajinska arhitektura in vnesite v vrinjeno vrstico podatke za študijski program Gozdarstvo (20 vpisanih v 2. letnik, 10 ponovno vpisanih v 1. letnik in 11 neznan). ♦

#### PREHODNOST 8

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

### 1.3.3.2 Izbris celic

Ukaz **Izbriši** ne pobriše samo vsebine, temveč tudi same celice, vrstice, stolpce ali delovne liste.





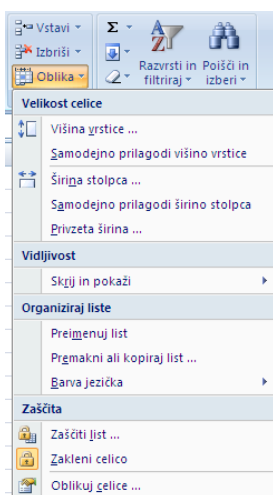
Ko uporabimo ta ukaz, se odpre pogovorno okno z naslednjimi možnostmi: **Pomakni celice v levo** - celice desno od izbranih se pomaknejo v levo; **Pomakni celice navzgor** - celice pod izbranimi se pomaknejo navzgor; **Celotno vrstico** - pobriše označene vrstice; **Celoten stolpec** - pobriše označene stolpce. Če imamo predhodno označeno celotno vrstico oziroma stolpec, ju ukaz samodejno izbriše brez prikaza omenjenega pogovornega okna.

#### PREHODNOST 9

Celice, ki vsebujejo podatke za študijski program Biologija v tabeli 1.1, pobrišite z ukazom **Izbriši**, **Premakni celice navzgor**. Nato uporabite ukaz **Razveljavi**, da bo na koncu ostala tabela nespremenjena. ♦

#### 1.3.3.3 Oblikovanje celic

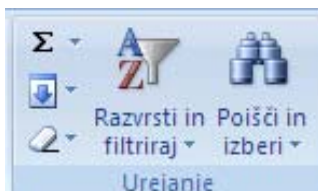
S klikom ukaza **Oblika** *Višina vrstice oz. Širina stolpca* se odpre pogovorno okno s trenutno višino oz. širino označenih vrstic oz. stolpcev, vrednosti lahko ustrezno spremenimo. Enostavneje višino vrstic in širino stolpcev spreminjamo tako, da se s križcem miške postavimo na mejo med dvema vrsticama oz. stolpcema v naslovni vrstici oz. stolpcu. Ko se križec spremeni v dvosmerno puščico, s potegom miške prestavimo mejo vrstice/stolpca.



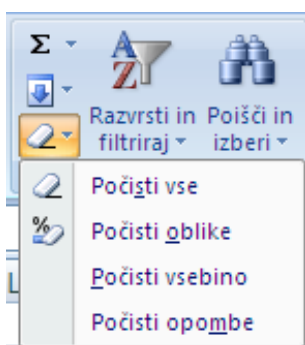
Kadar je v celici že neka vsebina, potrebno višino vrstice oz. širino stolpca določimo z dvoklikom na mejo med stolpcem in njegovim desnim sosedom v naslovni vrstici ali

uporabimo ukaz **Samodejno prilagodi višino vrstice** oz. **Samodejno prilagodi širino stolpca** v meniju, ki se pokaže, ko odpremo padajoč seznam ukaza **Oblika**.

### 1.3.4 Urejanje vsebine celic



Skupina ukazov **Urejanje** na kartici **Osnovno** je namenjena dodatnemu urejanju podatkov. Ukaze **Sum** (Vsota), **Polnilo** in **Razvrsti in filtriraj** bomo spoznali v naslednjih poglavjih.

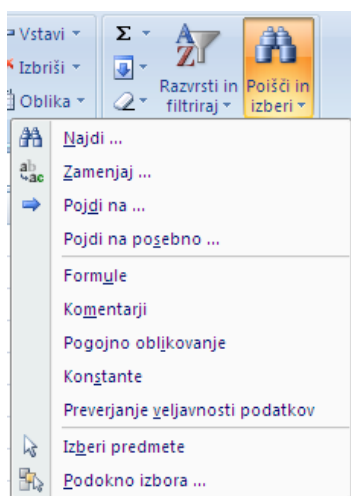


Z ukazom **Počisti** v celicah pobrišemo različne dele vsebine celic: **Počisti vse** - pobriše vse; **Počisti oblike** - pobriše oblikovne nastavitve v celici (meje, vzorci, osenčitve, oblika črk...); **Počisti vsebino** - pobriše podatek v celici, oblikovni parametri pa ostanejo za nove podatke; **Počisti opombe** – pobriše vstavljene komentarje<sup>4</sup>. Tipka DELETE deluje enako kot ukaz **Počisti vsebino**.

Klik na ikono **Poišči in izberi** odpre meni ukazov za iskanje in menjavo izbranega besedila, za premikanje na celice določene oblike, na celice z določeno vsebino in podobno.

---

<sup>4</sup> Komentarje v celico dodamo z ukazom **Nov komentar** na kartici **Pregled** na opravljenem traku.

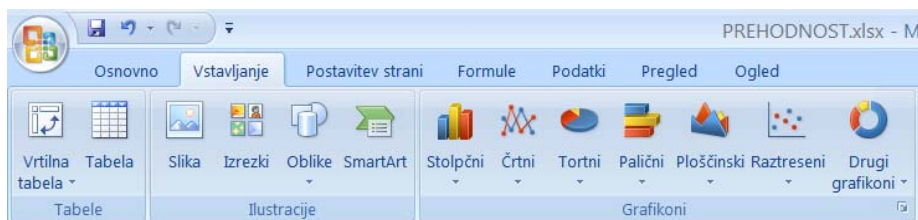


#### PREHODNOST 10

Z ukazom **Počisti vse** pobrišete vsebino celic, ki vsebujejo podatke za študijski program Biologija v tabeli 1.1, nato uporabite ukaz **Razveljavi**. Postopek ponovite z ukazom **Počisti vsebino** in nato še z ukazom **Počisti obliko**. Na koncu naj bo tabela nespremenjena. ♦

### 1.4 Grafično prikazovanje podatkov

Podatke lahko grafično prikažemo v Excelu na veliko različnih načinov. Pri tem delu nam pomagajo ukazi v skupini Grafikoni na kartici *Vstavljanje*.

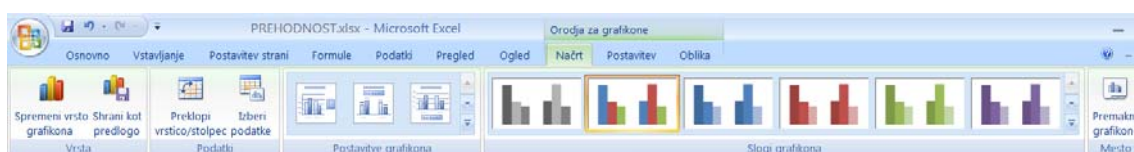


Podatke lahko prikazujemo v dveh ali v treh dimenzijah. Grafikoni, ki jih prikazujemo v dveh dimenzijah, imajo vodoravno in navpično os. Vrednosti na vodoravni osi so lahko številske ali opisne (kategorične), na navpični osi pa so vrednosti vedno številske.

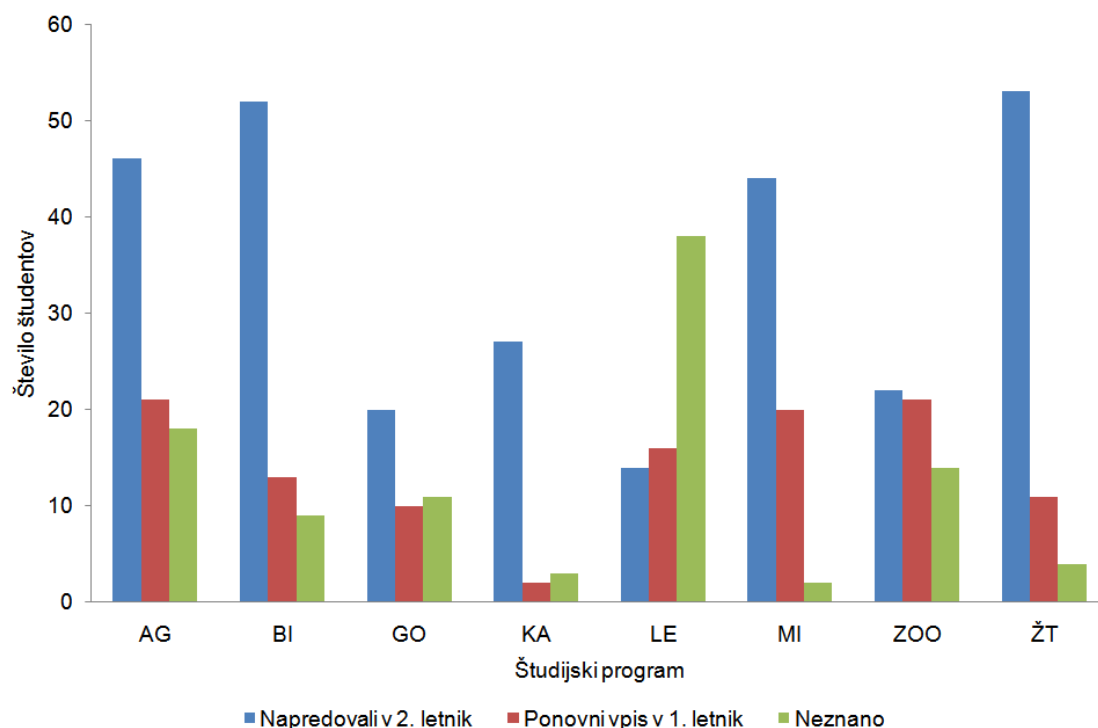
Podatki za grafikon so: vrednosti na vodoravni osi, niz ali več nizov vrednosti na navpični osi, oznake nizov za izdelavo legende. Pri izdelovanju grafikonov pogosto uporabljamo deljene skupine označenih celic.

Prvi korak oblikovanja grafikonov je izbira podatkov v preglednici: označimo skupino celic s podatki, ki jih želimo prikazati na grafikonu. V drugem koraku na kartici

**Vstavljanje** v skupini Grafikoni s klikom na ikono izberemo vrsto (stolpčni, črtni, tortni, palični,...) in nato še podvrsto grafikona (2D, 3D,...). Za tem se v preglednici ob podatkih pokaže prvi približek grafikona za izbrane podatke, hkrati se prikaže na pravilnem traku gumb **Orodja za grafikone** s tremi novimi karticami: **Načrt**, **Postavitev** in **Oblika**. Na vsaki od teh kartic so ukazi za dodatno oblikovanje in spreminjanje grafikona.



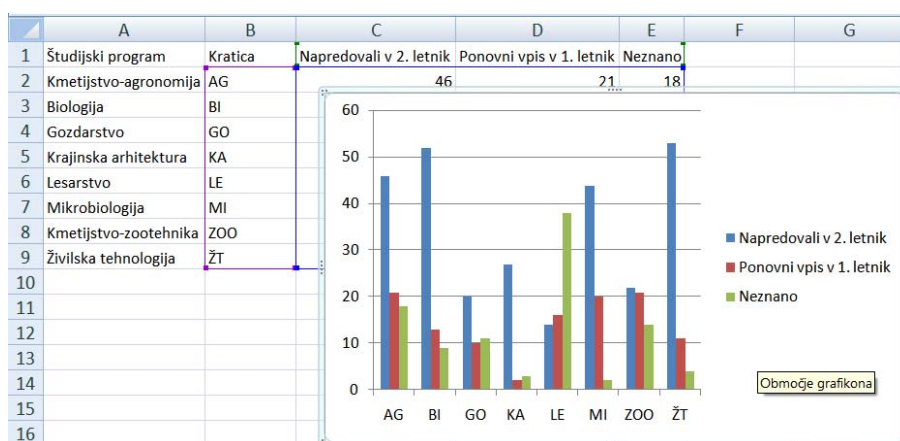
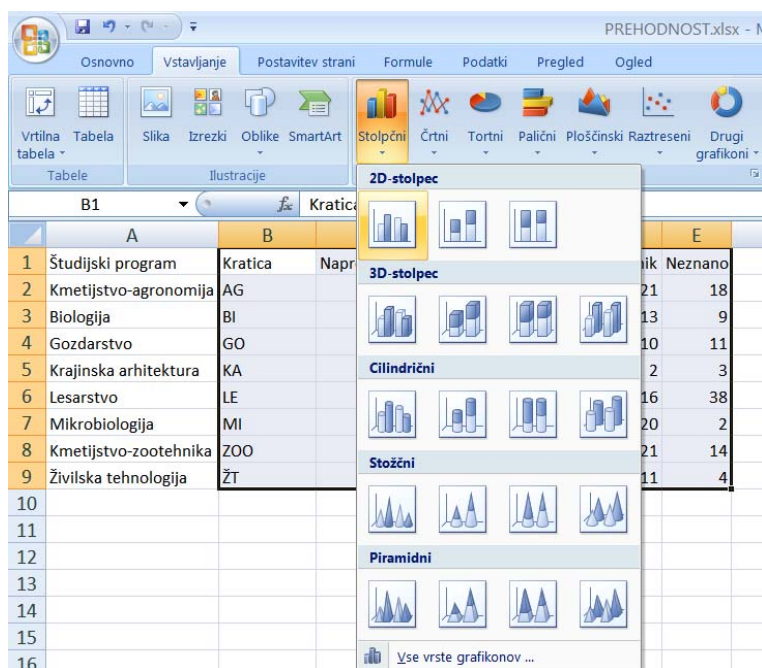
V nadaljevanju bomo oblikovali grafikon na sliki 1.1.



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

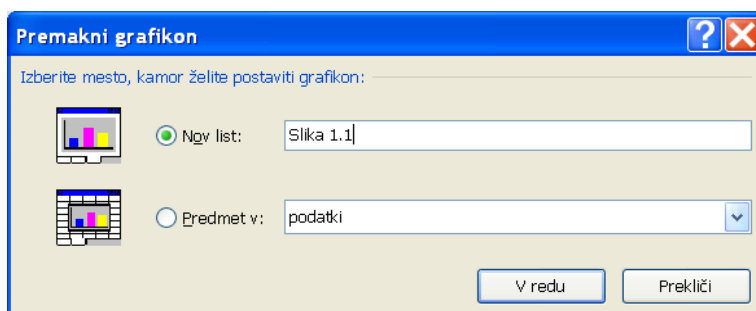
#### PREHODNOST 11

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. Na kartici **Vstavljanje** v skupini Grafikoni kliknite na ikono **Stolpčni** in nato izberite prvo podvrsto 2D-stolpcev. ♦



#### 1.4.1 Orodja za grafikone, kartica *Načrt*

Na kartici *Načrt* je pet skupin: Vrsta, Podatki, Postavitve grafikona, Slogi grafikona in Mesto. Najprej omenimo zadnjo skupino Mesto z ikono **Premakni grafikon**. S klikom na ikono se odpre pogovorno okno, v katerem grafikon iz preglednice prestavimo na samostojni grafični list ali obratno. Če v pogovornem oknu **Premakni grafikon** izberemo možnost *Nov list*, se grafikon pojavi na novem grafičnem listu v zvezku. Ime **Grafikon 1** lahko spremenimo po lastni izbiri (npr. Slika 1.1 na spodnji sliki). Možnost **Predmet** v je izbrana ob osnovni nastavitvi, ko je grafikon postavljen kot objekt na delovnem listu s podatki.

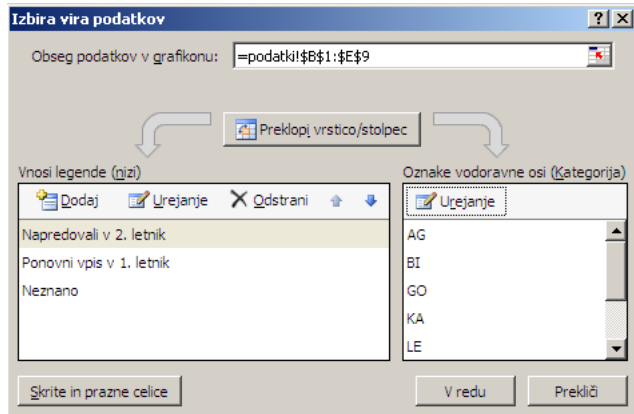


Priporočamo, da so grafični prikazi na grafičnih delovnih listih in ne kot objekt v preglednici. Grafične liste vsebinsko smiselno poimenujemo.

#### PREHODNOST 12

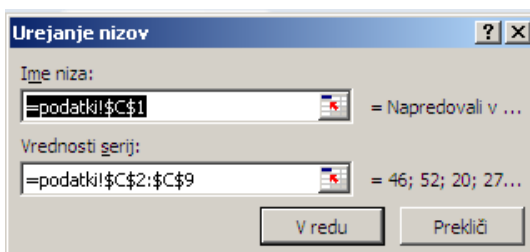
Postavite grafikon na nov grafični list z imenom »Slika1.1«. ♦

Prva skupina na kartici *Načrt* je Vrsta z dvema ukazoma: **Spremeni vrsto grafikona** - spremenimo vrsto že obstoječega grafikona; ukaz **Shrani kot predlogo** pa oblikovna določila že oblikovanega grafikona shrani kot predlogo, ki jo lahko uporabimo na drugih podatkih in si s tem prihranimo delo z dodatnim oblikovanjem grafikona.



Klik ukaza **Izberi podatke** v skupini Podatki odpre pogovorno okno **Izbira vira podatkov**. V polju *Obseg podatkov v grafikonu* se izpiše ime izbrane skupine označenih celic. Če podatkov predhodno nismo označili, je to polje prazno in ga moramo izponiti. Z gumbom **Preklopi vrstico/stolpec** določimo, ali so nizi podatkov v vrsticah ali v stolpcih. V polju *Vnosi legende (nizi)* lahko posamezne nize spremenimo, odstranimo ali dodamo na grafikon. V seznamu so navedena imena nizov, ki so na grafikonu vidna v legendi. Obseg podatkov za posamezen niz ali ime niza spremenimo

tako, da kliknemo ime niza v seznamu, odpre se pogovorno okno **Urejanje nizov**, v katerem popravimo vsebino polj **Ime niza** in **Vrednosti serij**:



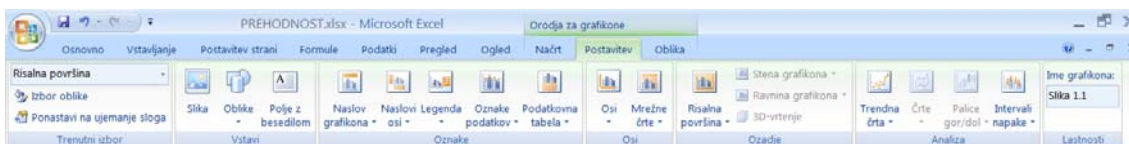
Vsebina celice, ki jo določimo v polju **Ime niza**, je v legendi na grafikonu. V polje **Vrednosti serij** označimo skupino celic z vrednostmi izbranega niza. Posamezen niz odstranimo z grafikona tako, da označimo njegovo ime v seznamu in kliknemo na gumb **Odstrani**. Nov niz dodamo na grafikon s klikom na gumb **Dodaj** in določitvijo polj **Ime niza** in **Vrednosti serij**.

V polje **Oznake vodoravne osi (kategorije)** napišemo (ali označimo) ime skupine celic z oznakami kategorij oz. vrednosti na osi X.

Z ukazi v skupinah **Postavitve grafikona** in **Slogi grafikona** dodatno oblikujemo grafikon, na primer položaj naslova grafikona, položaj naslovov osi, položaj legende, mrežnih črt.

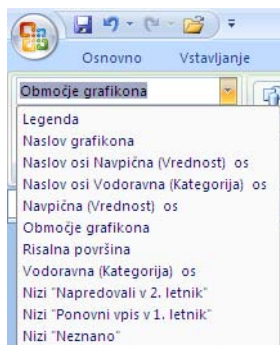
#### 1.4.2 Orodja za grafikone, kartici **Postavitev** in **Oblika**

Vsak grafikon lahko dodatno oblikujemo in spreminjamo z ukazi na karticah **Postavitev** in **Oblika**. Na kartici **Postavitev** so skupine ukazov **Trenutni izbor**, **Vstavi**, **Oznake**, **Osi**, **Ozadje**, **Analiza** in **Lastnosti**:



Vsak grafikon v Excelu ima več delov, ki jih lahko ločeno spreminjamo z ukazi v omenjenih skupinah. Dele aktivnega grafikona najdemo napisane na seznamu, ki se odpre s klikom puščice v prvem polju skupine **Trenutni izbor**. Če dela grafikona npr.

*Naslov grafikona* na grafikonu ni, ga tudi na seznamu ne vidimo, lahko pa ga na grafikon dodamo z ukazi v ostalih skupinah kartice **Postavitve**.



Glavni deli grafikona so:

- **Legenda** je sestavljena iz oznak in imen nizov;
- **Naslov grafikona** je nad grafom;
- **Naslov osi Navpična (Vrednost) os** je levo od navpične osi;
- **Naslov osi Vodoravna (Kategorija) os** je pod vodoravno osjo;
- **Navpična (Vrednost) os** je črta, ki ponazarja navpično os s črticami, ki določajo skalo na osi (glavne in pomožne črtice) ter z oznakami teh črtic;
- **Območje grafikona** je celotno območje grafikona, lahko mu določimo/spremenimo obliko znakov (naslovi, oznake na oseh), dodamo obrobo, barvo ozadja itd.;
- **Risalna površina** je med vodoravno in navpično osjo;
- **Mrežne črte** so črte na risalni površini v ozadju prikazanih nizov podatkov; ob osnovni nastavitvi so na grafikonu primarne/glavne vodoravne mrežne črte, lahko pa dodamo tudi navpične ali pomožne mrežne črte;
- **Vodoravna (Kategorija) os** je črta, ki ponazarja navpično os s črticami, ki določajo skalo na osi ali pa ločijo posamezne kategorije ter z oznakami teh črtic;
- **Nizi »ime spremenljivke«** ima obliko določeno z vrsto grafikona: npr. točke, krivulja, stolpci, krožni izseki itd..



Posamezni del grafikona označimo s klikom. S tem se ime tega dela grafikona pokaže v polju v skupini Trenutni izbor. Del grafikona, ki je trenutno izbran, je obdan z okvirjem, ki ima na vogalih kroge, pri delih, katerih velikost lahko spreminjamo s potegom z miško v različne smeri, pa so na sredini stranic še kvadratki. Ukaz **Izbior oblike** omogoča dodatno oblikovanje izbranega dela grafikona. Če želimo na grafikon dodati katerega od delov, ki ga še ni, za to uporabimo ukaze v skupinah Vstavi, Oznake, Osi ali Analiza.

Omenjene ukaze lahko izberemo tudi na meniju, ki se pokaže ob desnem kliku na posamezni del grafikona. Bolj natančno bomo nekatere od omenjenih ukazov opisali v naslednjih poglavjih knjige.

#### PREHODNOST 13

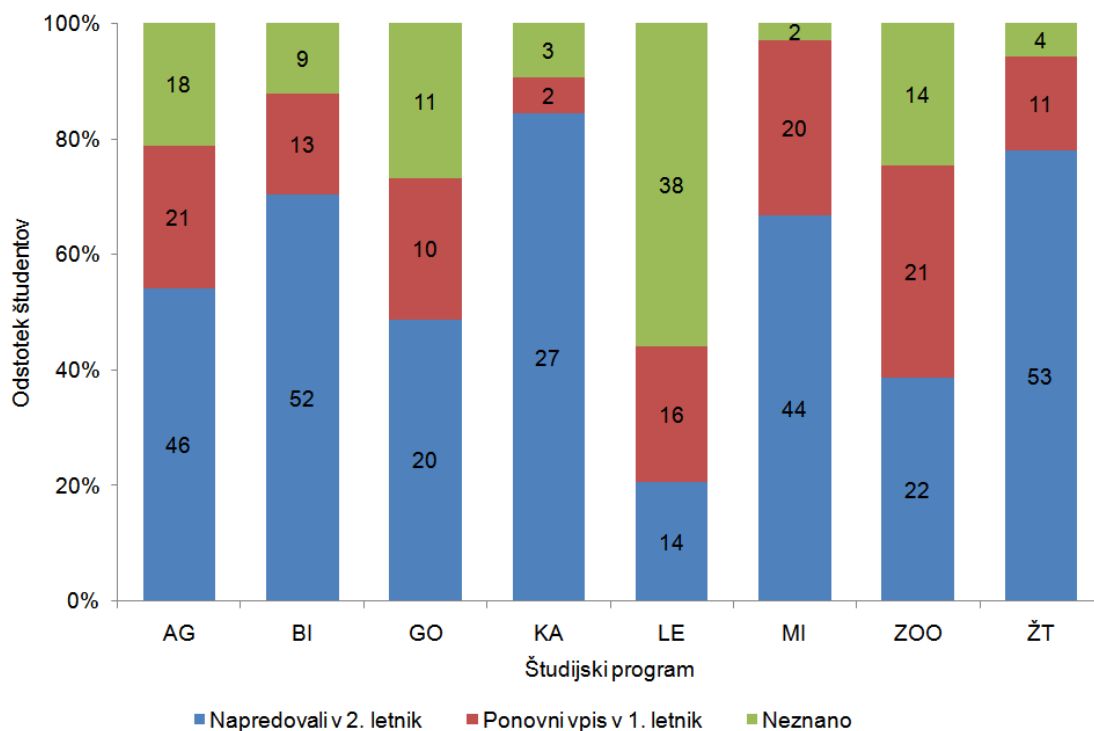
Preizkusite različne postavitve in sloge grafikona s slike 1.1 in opazujte, kako se grafikon spreminja. Končna izbira možnosti naj bo taka, da bo grafikon čim bolj podoben sliki 1.1. Opremite grafikon z naslovi, izbrišite mrežne črte, legendo postavite spodaj, odstranite obrobo območja grafikona. ♦

#### PREHODNOST 14

Za celotno področje grafikona določite pisavo Times New Roman. ♦

#### PREHODNOST 15

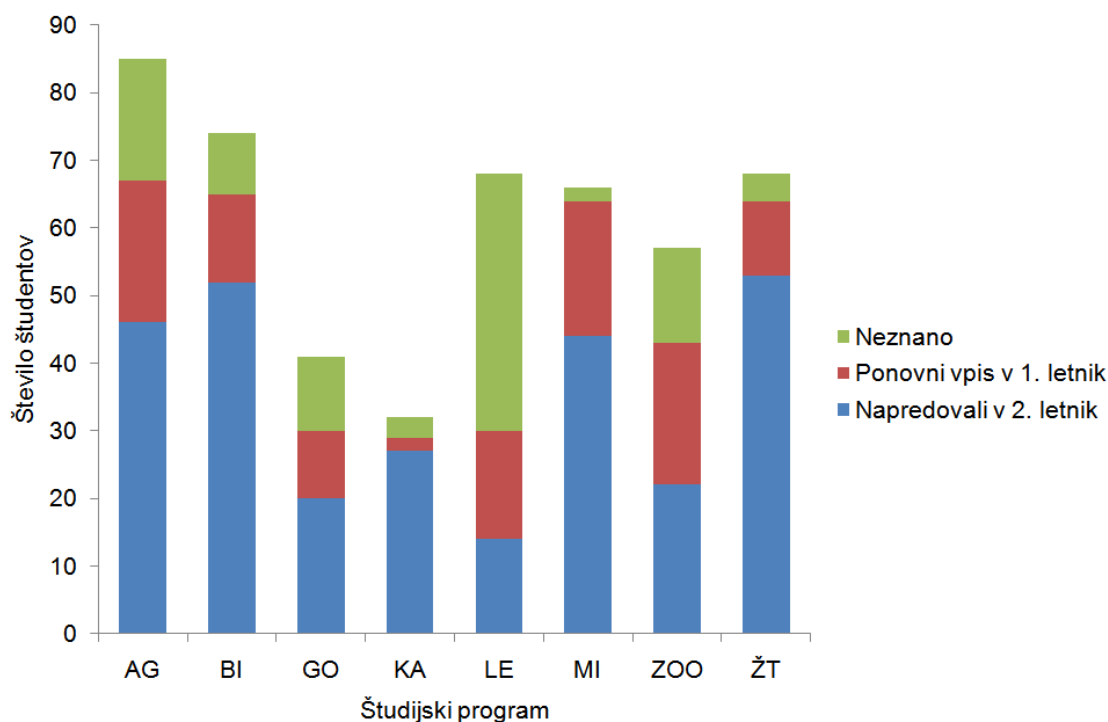
Na podlagi podatkov v tabeli 1.1 oblikujte še en grafikon, prikazan na sliki 1.2. Grafikon naj bo na svojem grafičnem listu z imenom »Slika 1.2«. V tem primeru je za vsak študijski program s strukturnimi stolpci prikazana struktura prehodnosti študentov vpisanih v 1. letnik v šol. l. 2002/03. ♦



Slika 1.2: Struktura prehodnosti študentov iz 1. letnika v 2. letnik na Biotehniški fakulteti v Ljubljani v šol. l. 2002/03, po študijskih programih (%); **grafični prikaz s strukturnimi stolpci** (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003). Na strukturnih stolpcih je prikazano absolutno število študentov.

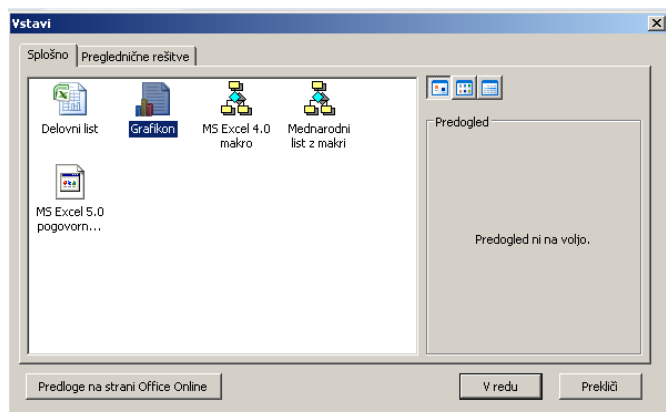
#### PREHODNOST 16

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 posamezen študijski program prikazani z naloženimi stolpci. Katero vrednost predstavlja višina stolpca na sliki 1.3? ♦



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

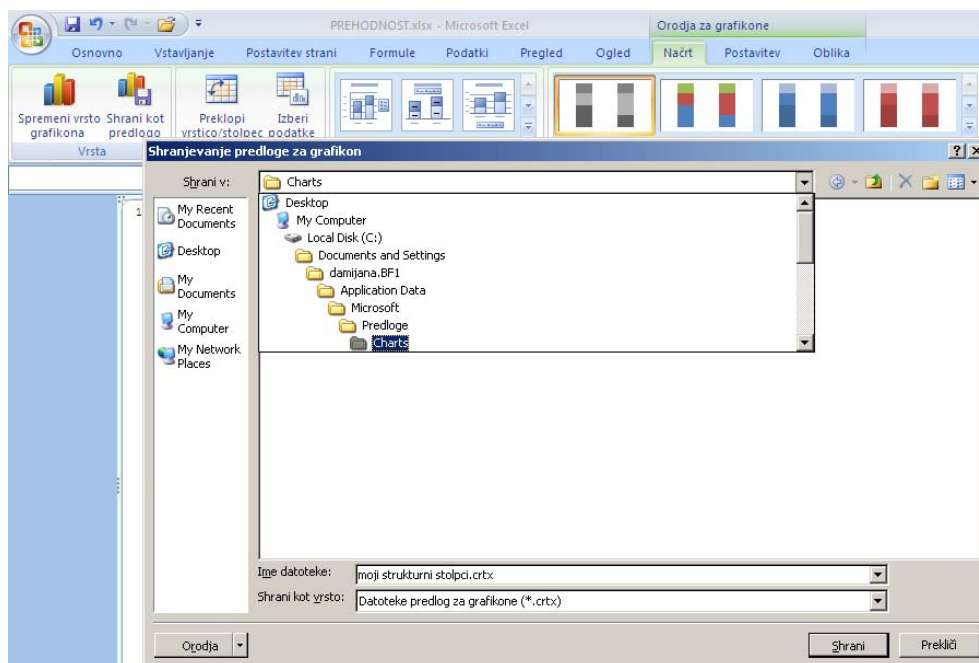
Če potrebujemo prazen grafični list, na katerega želimo kopirati že oblikovani grafičnik ali pa želimo grafičnik oblikovati z orodjem Risanje, kliknemo enega izmed jezičkov obstoječih delovnih listov v zvezku, z desnim klikom odpremo meni in izberemo ukaz Vstavi:



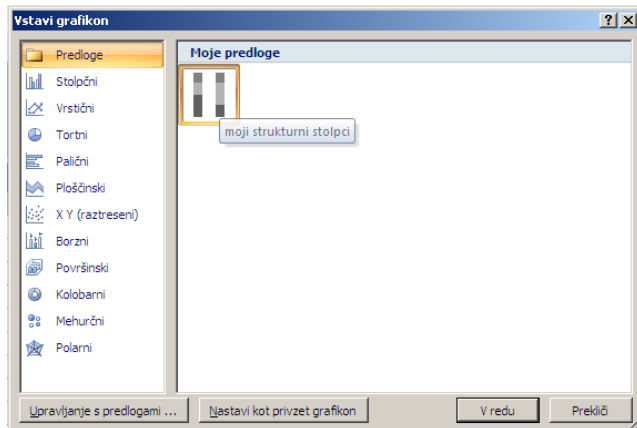
V pogovornem oknu *Vstavi* izberemo ikono **Grafikon** in kliknemo na gumb **V redu**. S tem se odpre nov delovni list z imenom »Grafikon #« (# je številka, ki je odvisna od

tega, koliko grafičnih listov v zvezku že obstaja). Če smo imeli pred izvedbo tega ukaza v preglednici označene podatke, jih tako vstavljen grafični list prikazuje. Da se grafikona znebimo, z desnim klikom izberemo **Območje grafikona** in pritisnemo tipko Delete.

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 shranimo grafikon kot predlogo. S klikom ukaza **Shrani kot predlogo** v skupini Vrsta na kartici **Načrt** se odpre okno **Shranjevanje predloge za grafikon**, v katerem izberemo mesto, kamor to predlogo shranimo, Excel samodejno ponudi mapo Charts:



V polju **Ime datoteke** predlogo poimenujem (npr. »moji strukturni stolpci«), datoteka dobi končnico .crtx. Predlogo na drugih podatkih uporabimo tako, da na kartici **Vstavljanje** v skupini Grafikoni kliknemo na desno spodnjo puščico, s tem se odpre pogovorno okno **Vstavi grafikon**, kjer v mapi Predloge izberemo eno izmed lastno oblikovanih predlog.



## 1.5 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 začnemo pisati z znakom =, + ali -, 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), \* (množenje) in ^ (potenca). Logični operatorji so = (enako), < (manjše), > (večje), >= (večje ali enako), <= (manjše ali enako), <> (neenako). Vrstni red računanja v zapletenejših formulah določamo z oklepaji.

### FORMULE 1

V datoteko PREHODNOST.xlsx se postavite na prazen delovni list in ga poimenujte »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 formuli. 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  $=A3*10$ . Pri tem ni pomembno, ali napišemo veliko ali majhno črko (a3 ali A3), po pritisku tipke **ENTER** Excel samodejno izpiše reference celic z velikimi črkami. 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 te celice: npr. najprej natipkamo znak =, nato z miško kliknemo celico A3, ob tem se referenca celice izpiše v formulo, nadaljujemo s tipkanjem znaka \* in številke 10. ♦

### 1.5.1 Kopiranje formul

Formule lahko kopiramo in prestavljamo 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. ♦

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**. Npr. za formulo  $=A3*10$  najprej natipkamo znak =, nato natipkamo ali kliknemo 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 pritisnemo 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

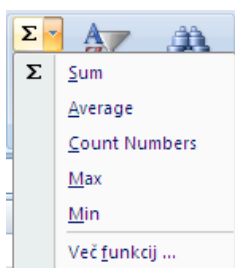
Popravite formulo v celici B3 na  $=\$A3*10$  in jo kopirajte v skupino celic B4:B10 in v skupino celic C3:F3. V katerih celicah se rezultati spremenijo? ♦

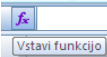
#### FORMULE 8

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? ♦

### 1.5.2 Vstavljanje funkcij

Excel ima veliko vgrajenih funkcij. Na kartici *Osnovno* v skupini Urejanje stoji ikona  (*Sum*), ki omogoča hitro seštevanje po stolpcih ali vrsticah. Klik puščice na ikoni  odpre seznam ostalih hitro dostopnih funkcij: *Average* povprečje izbranih vrednosti, *Count Numbers* prešteje koliko številčnih vrednosti je v izboru celic, *Max* in *Min* poiščeta maksimalno in minimalno vrednost. Ostale funkcije najdemo na seznamu, ki se odpre s klikom ukaza **Več funkcij**.



Ukaz **Več funkcij** izvedemo tudi s klikom ikone  v vrstici za vnos. Preden uporabimo ukaz **Vstavi funkcijo**, moramo označiti celico, v kateri želimo imeti izračun. Če ime funkcije poznamo, jo lahko vnesemo tudi z vpisom v celico ali pa z vpisom v vrstico za vnos. Natančneje se bomo z načini uporabe zahtevnejših formul seznanili v prihodnjih poglavjih.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Študijski program	Kratika	Napredovali v 2. letnik	Ponovni vpis v 1. letnik	Neznano	Skupaj	% napredovali	% ponovni v	% neznano	% BF		
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							

#### PREHODNOST 17

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

- v skupino celic F2:F9 izračunajte število vpisanih študentov v 1. letnik v šol. l. 2002/03 za vsak študijski program;
- 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 vsak študijski program;
- v skupino celic J2:J9 izračunajte strukturo vpisanih v 1. letnik v šol. l. 2002/03 na BF po študijskih programih (%).

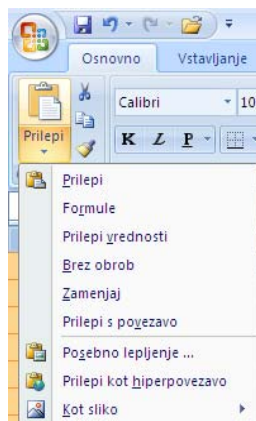
**Navodilo:** Formule vpišite samo za prvi študijski program, nato jih iz prve vrstice kopirajte v ostale vrstice. Pazite na relativno in absolutno naslavljanje celic. Hitro kopiranje lahko izvedemo na kartici **Osnovno** v skupini Urejanje z ukazom **Polnilo**: označimo celice F2:I9 in kliknemo ikono **Polnilo/Dol**. Drug način hitrega kopiranja je s potegom miške: z miško se postavite v desni spodnji vogal celice s formulo, ko se oblika in barva križca spremenita, potegnite z miško čez celice, v katere želite kopirati formulo. Naenkrat lahko označite tudi več celic (npr. celice F2:I2) in izvedete poteg z miško. Preverite, ali so izračuni pravilni. ♦



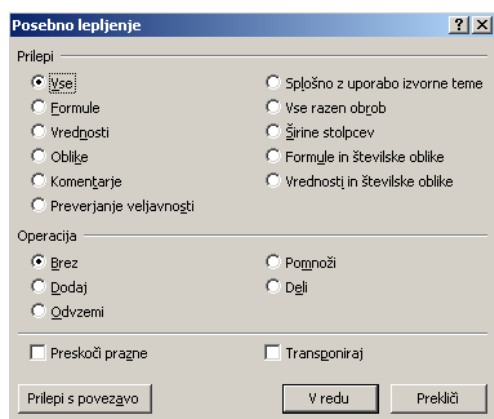
	A	B	C	D	E	F	G	H	I	J
1	Študijski program	Kratica	Napredovali v 2. letnik	Ponovni vpis v 1. letnik	Neznano	Skupaj	% napredovali v 2. letnik	% ponovni vpis	% neznano	% BF
2	Kmetijstvo-agronomija	AG	46	21	18	85	54,1	24,7	21,2	17,3
3	Biologija	BI	52	13	9	74				
4	Gozdarstvo	GO	20	10	11	41				
5	Krajinska arhitektura	KA	27	2	3	32				
6	Lesarstvo	LF	14	16	38	68				
7	Mikrobiologija	MI	44	20	2	66				
8	Kmetijstvo-zootehnika	ZOO	22	21	14	57				
9	Živilska tehnologija	ZT	53	11	4	68				
10						491				

### 1.5.3 Lepljenje na poseben način

Poleg običajnega načina lepljenja z ukazom **Prilepi**, ki ga že poznamo, Excel omogoča tudi posebno lepljenje z ukazi **Formule**, **Prilepi vrednosti**, **Brez obrob**, **Zamenjaj**, **Prilepi s povezavo**. Te ukaze najdemo tudi v pogovornem oknu, ki ga odpre ukaz **Posebno lepljenje**.



V pogovornem oknu **Posebno lepljenje** imamo dve skupini možnosti, **Prilepi** in **Operacija**. V skupini **Prilepi** so naslednje možnosti:



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

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, da 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 odložišča tako, da vrstice postavi v stolpce: prva vrstica postane prvi stolpec.

PREHODNOST 18

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

PREHODNOST 19

Celice F1:J9 kopirajte na nov delovni list kot vrednosti. ♦

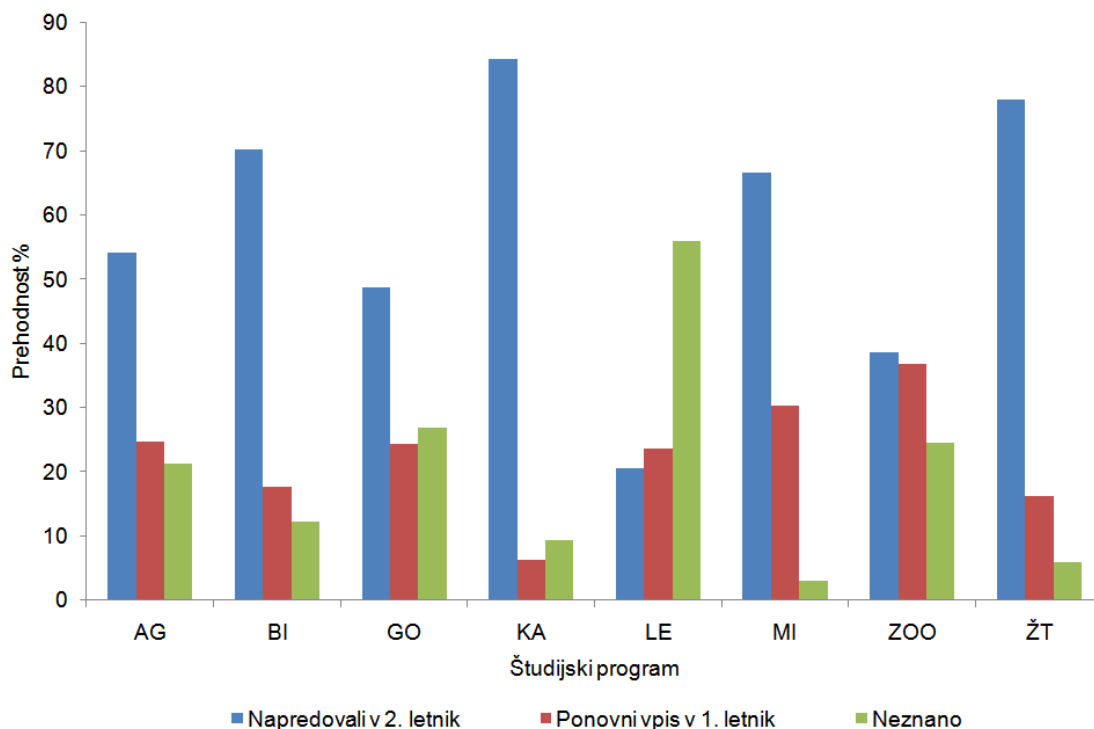
PREHODNOST 20

Tabelo A1:J9 kopirajte na nov delovni list tako, da stolpce spremenite v vrstice. ♦

## PREHODNOST 21

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? ♦

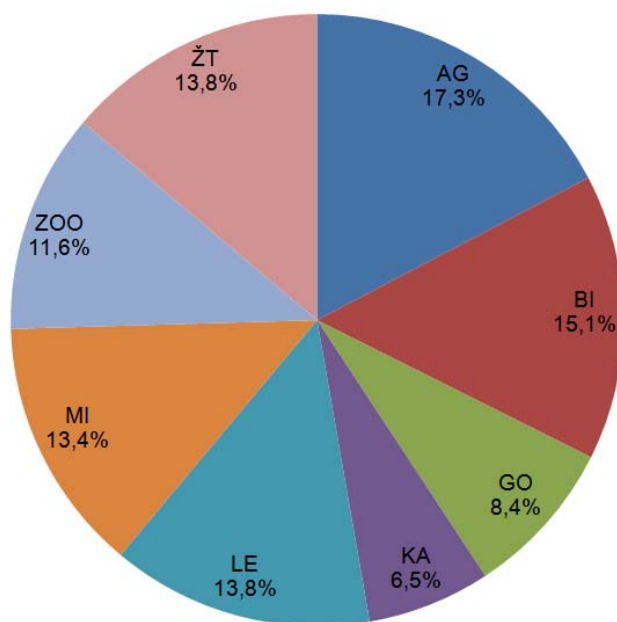


Slika 1.4: Prehodnost generacije študentov 1. letnika v 2. letnik na univerzitetnih študijskih programih Biotehniške fakultete v šolskem letu 2002/2003, grafični prikaz z **razrezanimi strukturnimi stolpci** (Vir: Univerza v Ljubljani, BF, Poročilo za leto 2003)

## PREHODNOST 22

S strukturnim krogom prikažite strukturo po študijskih programih za vse vpisane v 1. letnik na BF (slika 1.5). Delovni list z grafikonom v preglednici poimenujte z imenom »Slika 1.5«.

**Navodilo:** grafikon lahko oblikujete tako na podlagi absolutnih podatkov (števila študentov po programih), kot tudi na podlagi relativnih vrednosti- izračunane strukture. Naredite oba primera: enkrat za vhodne podatke grafikona izberite kratice študijskega programa in število vpisanih v 1. letnik v šol. l. 2002/03 za posamezen program, drugič pa za vhodne podatke grafikona izberite kratice študijskih programov in izračunano strukturo vpisanih v 1. letnik v šol. l. 2002/03 na BF po študijskih programih. ♦



Slika 1.5: Struktura po študijskih programih (%) 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 prepišite v excelovo datoteko z imenom OKRASNE RASTLINE.xlsx 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 univerzitetni študij v šol. l. 2006/07 po študijskih programih 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	0
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

\* Pri študiju Biotehnologije absolventov v letu 2006/07 še ni bilo.

## VPIS 1

Podatke iz tabele 1.3 prepišite v datoteko VPIS.xlsx. ♦

## VPIS 2

Za vsak študijski program izračunajte skupno število študentov v šolskem letu 2006/2007.

Izračune grafično prikažite na eni sliki. ♦

## VPIS 3

Izračunajte strukturo vseh študentov na BF po študijskih programih in jo grafično prikažite z razrezanimi strukturnimi stolpci. ♦

## VPIS 4

Za vsak študijski program izračunajte strukturo študentov po letnikih študija (brez absolventov).

Izračune grafično prikažite s strukturnimi stolpci na eni sliki. ♦

## VPIS 5

Za vsak študijski program 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:

PROMETNE NESRECE.xlsx

PREBIVALSTVO.xlsx

GOSTOTA PREBIVALSTVA.xlsx

#### Delovne datoteke:

PREBIVALSTVO\_analiza.xlsx

VPIS.xlsx

OKRASNE RASTLINE.xlsx

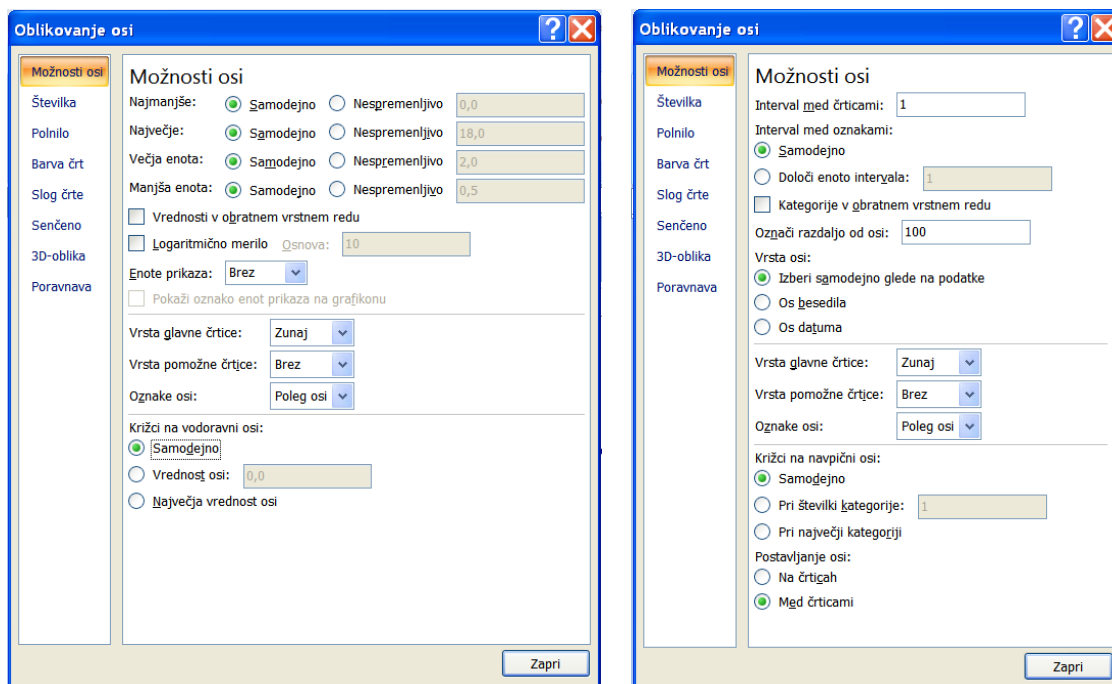
V tem poglavju bomo natančneje opisali oblikovanje nekaterih delov grafikona. Ukaze za to najdemo na kartici *Postavitev* v **Orodju za grafikone**, ki je v opravljenem traku vidno, če je grafikon v datoteki aktiven (označen).

### 2.1 Oblikovanje osi grafikona

Med deli grafikona, ki jih lahko dodatno oblikujemo in spreminjamo njihovo postavitev, sta vodoravna in navpična os. Dodatno oblikovanje obeh osi omogočajo skupine ukazov oz. možnosti v pogovornem oknu **Oblikovanje osi**. Pogovorno okno **Oblikovanje osi** odpremo na tri načine:

- na območje osi na grafikonu kliknemo z desnim klikom in na meniju izberemo zadnji ukaz **Oblikuj os**;
- na karticah *Postavitev* ali *Oblika* v skupini Trenutni izbor v prvem polju izberemo ustrezno os in kliknemo ukaz **Izbor oblike**;
- na kartici *Postavitev* v skupini Osi izvedemo ukaz **Osi**, izberemo eno izmed možnosti *Primarna vodoravna os* ali *Primarna navpična os* ter na meniju izberemo zadnji ukaz **Več možnosti za**.

V levem, ožjem delu tega okna je seznam skupin ukazov: Možnosti osi, Številka, Polnilo, Barva črt, Slog črte, Senčeno, 3D-oblika, Poravnava. V desnem delu se pokažejo ukazi oz. možnosti, ki jih ponuja posamezna skupina. Spodnje levo okno prikazuje oblikovanje številske, desno pa kategorične osi. Poglejmo ju podrobneje.



Prva skupina ukazov Možnosti osi ponuja dva različna nabora možnosti, odvisno od tega, ali je izbrana os številska ali kategorična.

Če oblikujemo številsko os (*Navpična (Vrednost) os* ali *Vodoravna (Vrednost) os*), so *Možnosti osi* naslednje:

- **Najmanjše:** najmanjša prikazana vrednost, ki je po navadi na spodnjem koncu osi, pri nekaterih vrstah grafikona pa je na zgornjem koncu osi; Excel jo, tako kot tudi naslednje tri vrednosti, ob osnovni nastavitvi določi samodejno glede na vrednosti podatkov. Če jo želimo spremeniti, izberemo možnost *Nespremenljivo* in v polje napišemo zeleno vrednost;
- **Največje:** največja prikazana vrednost, ki je po navadi na zgornjem, izjemoma tudi na spodnjem koncu osi;

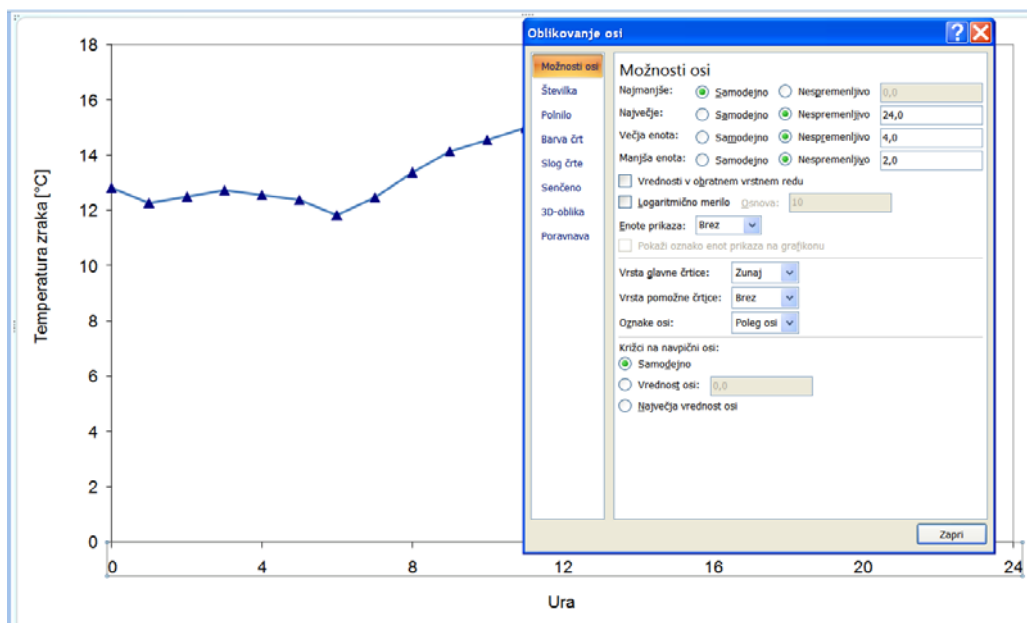


- *Večja enota*: razmak med glavnimi črticami na osi, ob njih so oznake napisane;
- *Manjša enota*: razmak med pomožnimi črticami na osi, ob njih ni oznak;
- *Vrednosti v obratnem vrstnem redu*: največja vrednost na osi je prikazana na spodnjem oz. desnem koncu osi in obratno;
- *Logaritmično merilo*: Excel vrednosti na osi logaritmirajo; osnovo logaritma določimo v desnem polju, ob osnovni nastavitvi izračuna desetiški logaritem;
- *Enote prikaza*: v tem polju lahko izberemo enote številke ob osi (*Brez*, *Stotice*, *Tisočice*, *Milijoni*, *Bilijoni*, *Triljoni*).

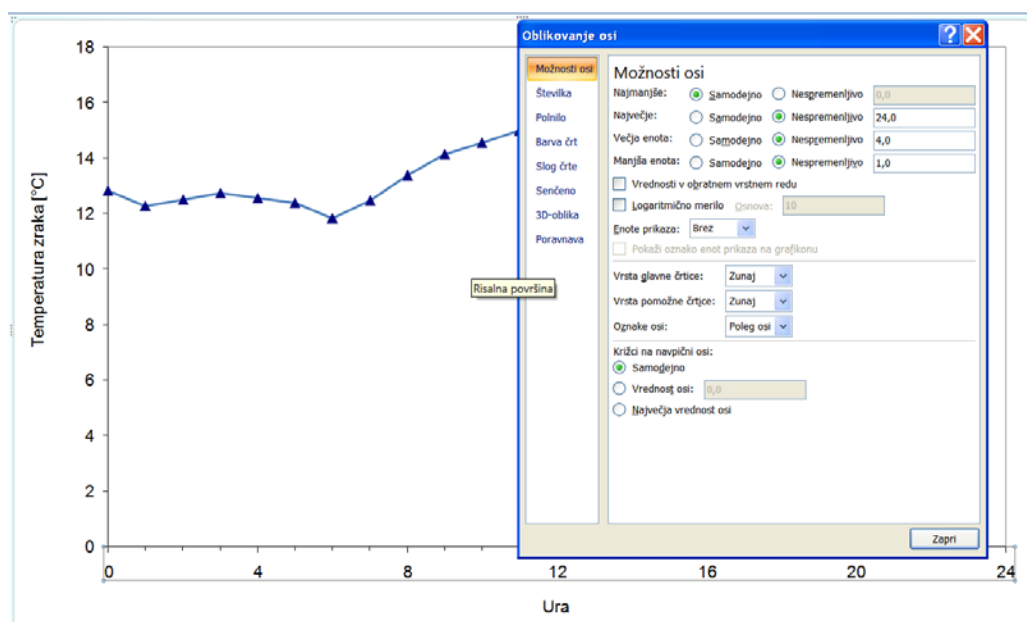
V drugem razdelku so možnosti, ki določajo položaj črtic in oznak na osi:

- *Vrsta glavne črtice*: izberemo položaj črtic na osi. Ob osnovni nastavitvi je tu izbrana možnost *Zunaj*;
- *Vrsta pomožne črtice*: izberemo položaj pomožnih (krajših) črtic, ki so med glavnimi črticami na osi, ob njih se ne izpisujejo oznake. Ob osnovni nastavitvi pomožne črtice niso vidne, izbrana je možnost *Brez*;
- *Oznake črtic*: določimo položaj oznak vrednosti ob glavnih črticah na osi. Ob osnovni nastavitvi so oznake črtic *Poleg osi*.

Na spodnji sliki je izbrana *Vodoravna (Vrednost) os*, vrsta grafikona je *Raztreseni*, najmanjša vrednost je določena samodejno, največja vrednost je določena nespremenljivo 24, večja enota je 4, manjša enota je 2, vendar je na sliki ne zaznamo, ker je *Vrsta pomožne črtice: Brez*.



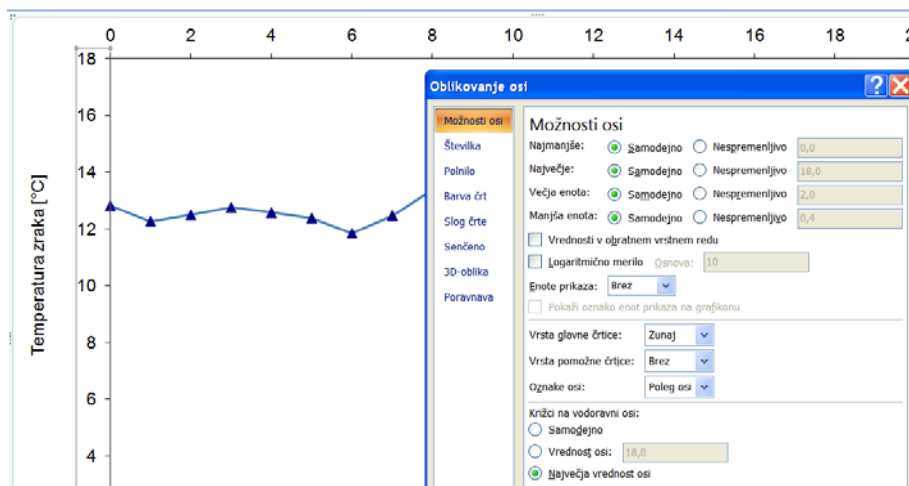
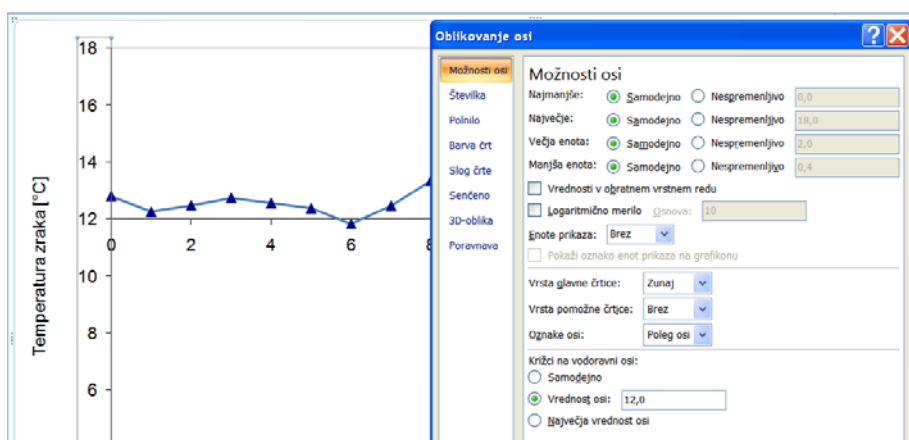
Če je *Manjša enota* 1 in za možnost *Vrsta pomožne črtice* izberemo vrednost *Zunaj*, je os taka:



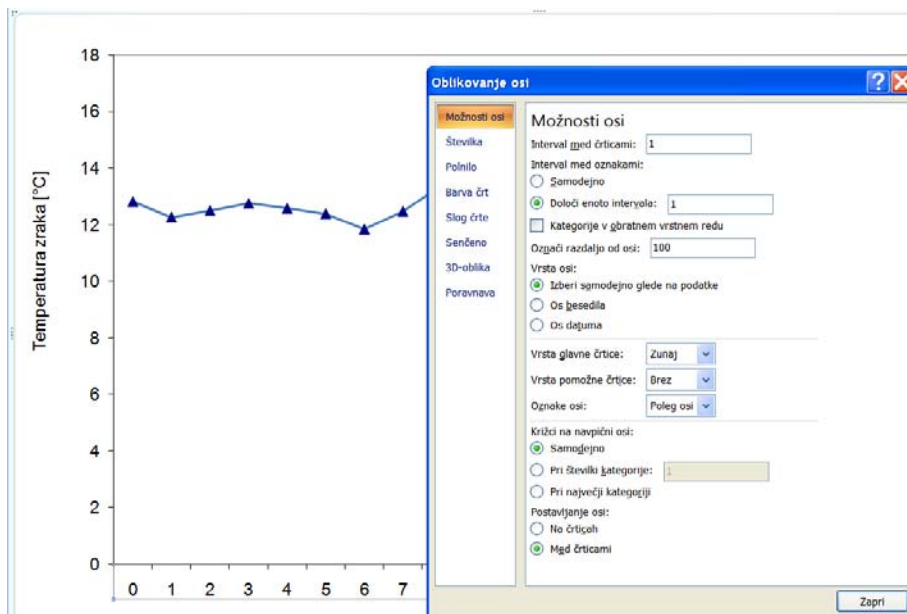
V tretjem razdelku pogovornega okna **Oblikovanje osi** je še ena možnost, ki je odvisna od tega, ali je izbrana vodoravna oz. navpična os, in določa, pri kateri vrednosti navpična/vodoravna os grafikona seka vodoravno/navpično os:

## 2 Relativna števila in grafični prikazi

- Križci na navpični/vodoravni osi: *Samodejno* - osi se sekata pri najmanjših vrednostih; *Vrednost osi* - napišemo vrednost na vodoravni/navpični osi, pri kateri jo seka navpična/vodoravna os (na primer 12 na spodnji sliki); *Največja vrednost osi* - navpična/vodoravna os seka vodoravno/navpično os pri največji namesto pri najmanjši vrednosti (glej drugo sliko spodaj). Če je izbrana vodoravna os, ta možnost določa, da je navpična os na desnem robu grafikona; če je izbrana navpična os, pa je vodoravna os na zgornjem robu grafikona.



Oblikovanje kategorične osi je vezano na določene vrste grafikona, na primer *Stolpčni*, *Palični*, *Črtni*. Če oblikujemo *Vodoravno (Kategorija) os*, so v pogovornem oknu **Oblikovanje osi Možnosti osi** naslednje:



- *Interval med črticami*: številski razmak za glavne črtice, vmesne glavne črtice se na osi ne narišejo<sup>5</sup>;
- *Interval med oznakami*: številski razmak za oznake glavnih črtic, k vmesnim glavnim črticam se oznake kategorij ne izpišejo;
- *Kategorije v obratnem vrstnem redu*: kategorije se na vodoravni osi pokažejo v obratnem vrstnem redu, kot so navedene v preglednici;
- *Označi razdaljo od osi*: velikost razdalje oznak od osi, ob osnovni nastavitvi je 100;
- *Vrsta osi*: če želimo posebej poudariti, da so oznake osi v bistvu oznake kategorij izberemo možnost *Os besedila*, če so datumi pa *Os datuma*. V zadnjem primeru se odpre novo pogovorno okno s posebnimi določili za datume.

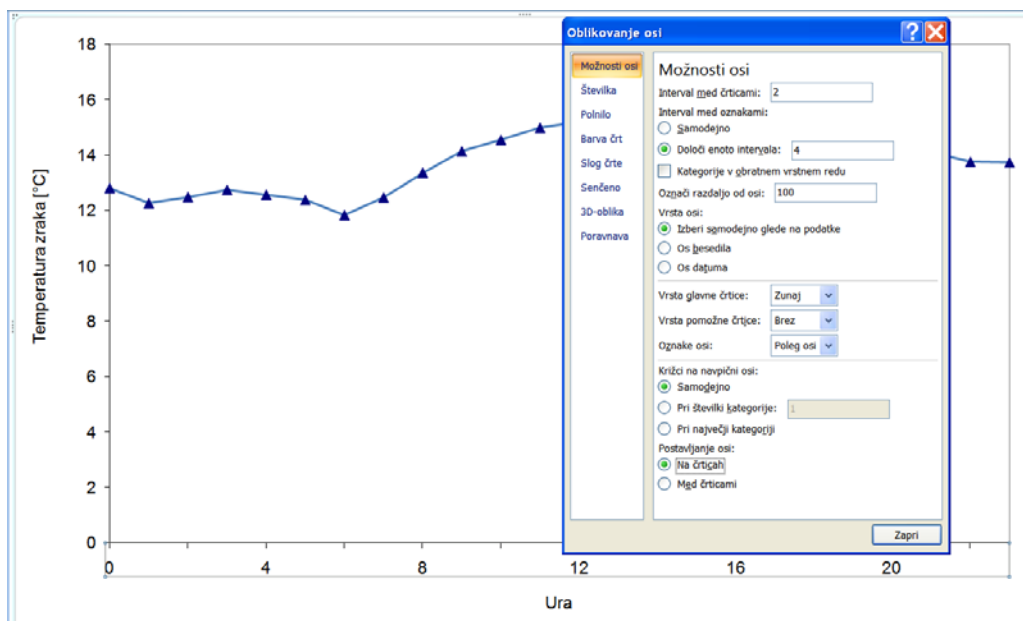
<sup>5</sup> Na kategorični osi pomožne črtice niso smiselne.

Drugi razdelek možnosti okna **Oblikovanje osi** je enak kot v primeru, če je os številka.

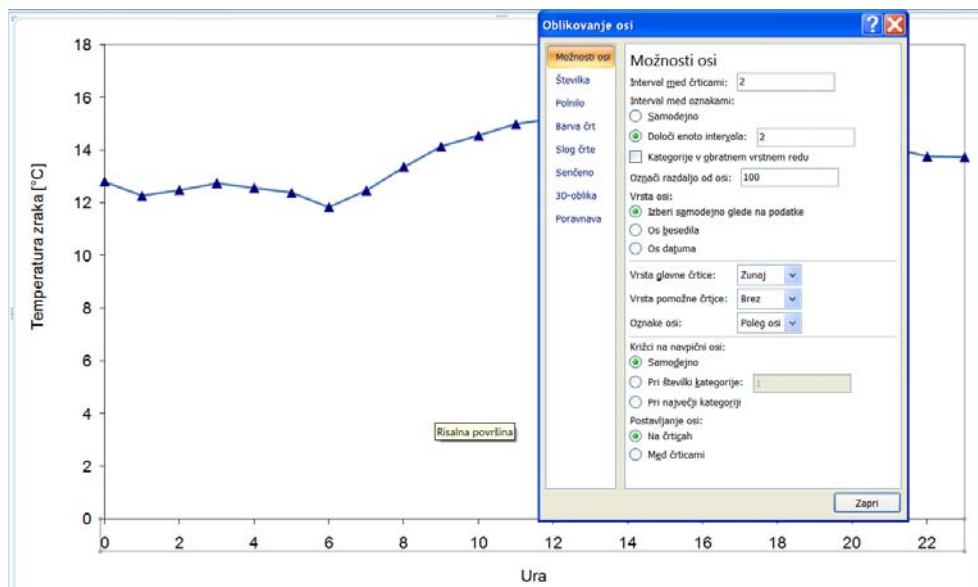
V tretjem razdelku pogovornega okna **Oblikovanje osi** pa sta dve možnosti: najprej izbiramo med tremi možnostmi, ki določajo, pri kateri kategoriji navpična os seka vodoravno os:

- *Križci na navpični osi: Samodejno* - osi se sekata pri prvi kategoriji; *Pri številki kategorije* - napišemo številko kategorije, pri kateri jo seka navpična os; *Pri največji kategoriji* - navpična os seka vodoravno pri zadnji namesto pri prvi kategoriji;
- *Postavljanje osi: Na črticah* - oznake na vodoravni osi so pod glavno črtico in oznake nizov so na risalni površini navpično nad glavno črtico; *Med črticami* – oznake na vodoravni osi so med glavnimi črticami in oznake nizov so na risalni površini med glavnimi črticami (glej primer v nadaljevanju).

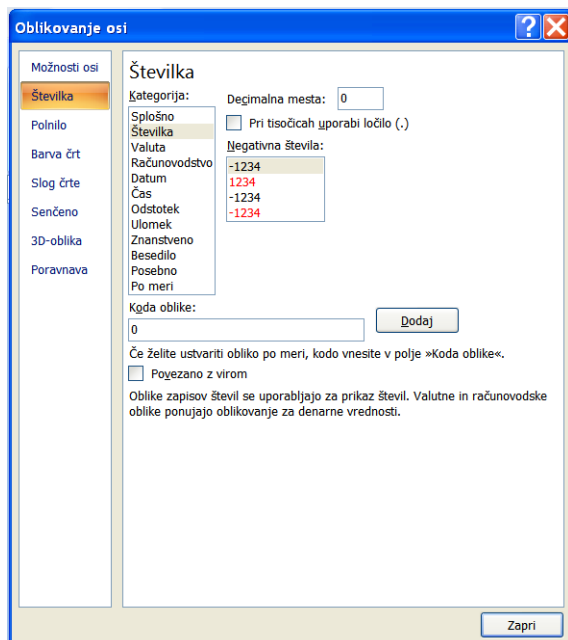
Na spodnji sliki je izbrana *Vodoravna (Kategorija) os*, vrsta grafikona je *Črtni* (na nekaterih menijih je poimenovan *Vrstični*), oznake osi so ob samodejni nastavitvi med črticami, tu smo ji postavili na črtice.



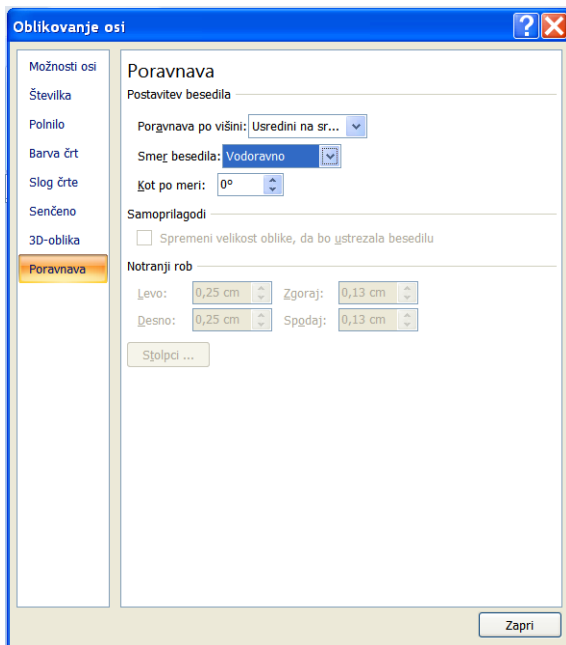
Če izberemo za možnost *Postavljanje osi Na črticah* ter določimo, da je *Interval med črticami 2* in *Interval med oznakami 2*, je vodoravna os taka:



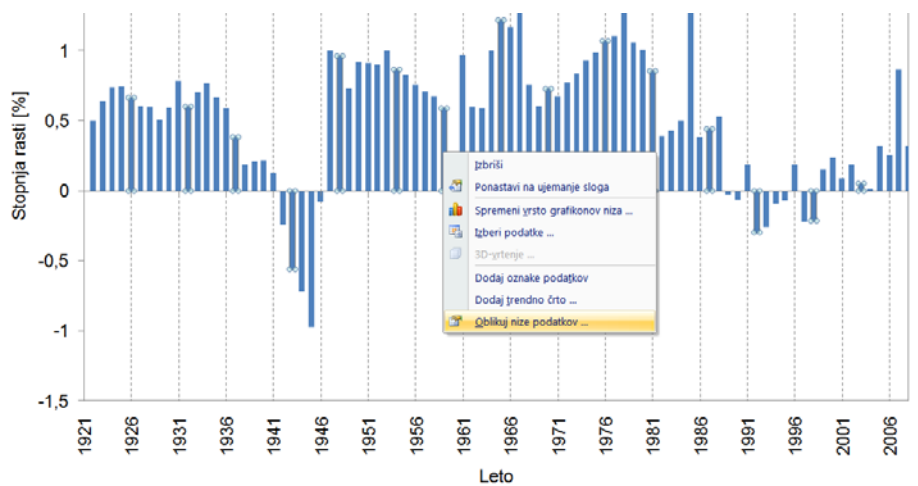
Od ostalih skupin ukazov v pogovornem oknu **Oblikovanje osi** omenjamo le še skupini **Številka** in **Poravnava**. V skupini možnosti **Številka** določimo obliko vrednosti števil na osi:



V skupini možnosti Poravnava določimo položaj oznak na osi (vodoraven, navpičen izpis).

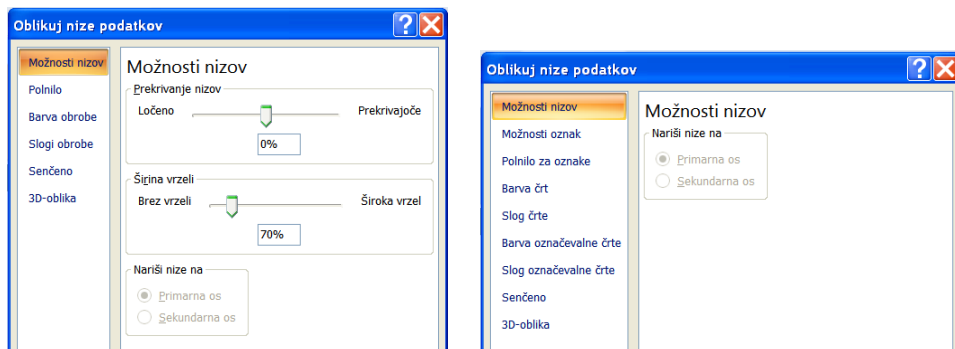


### 2.2 Oblikovanje niza podatkov



Z desnim klikom predhodno izbranega objekta – niza podatkov (npr. krivulje, stolpca) v grafikonu, se pokaže meni z ukazi za oblikovanje tega objekta. Z ukazom **Izbriši** pobrišemo prikaz izbranega objekta. Z ukazom **Oblikuj nize podatkov...** se odpre pogovorno okno **Oblikuj nize podatkov**, ki vsebuje več skupin ukazov (*Možnosti*

*nizov, Polnilo, Barva obrobe, Možnosti oznak, ...*). Koliko in katere skupine ukazov so prisotne v pogovornem oknu, je odvisno od vrste grafikona oz. vrste grafičnega prikaza niza podatkov. Na primer, če je niz podatkov prikazan s stolpci, ima pogovorno okno vsebino, kot kaže spodnja leva slika. Če pa je niz podatkov prikazan s črto v grafikonu vrste *Raztreseni*, je pogovorno okno kot kaže spodnja desna slika.

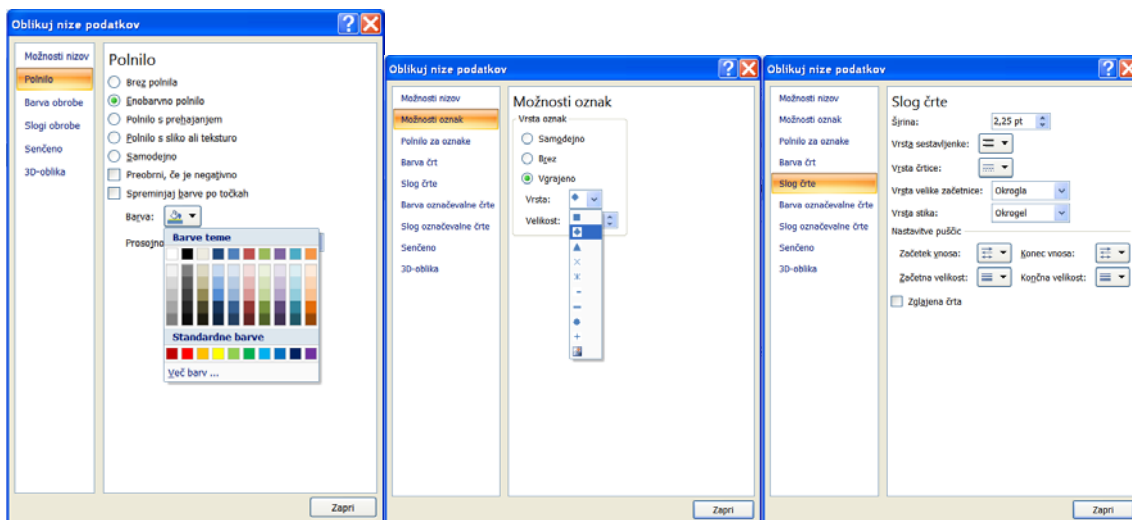


Prva skupina ukazov v pogovornem oknu **Oblikuj nize podatkov** je vedno *Možnosti nizov*, to so možnosti, ki jih najpogosteje spreminjamo. Na *Stolpčnem* grafikonu tu nastavimo širino vrzeli med stolpci. V osnovni nastavitvi je *Širina vrzeli* 150; če želimo, da se stolpci na grafikonu držijo skupaj, mora biti v polju *Širina vrzeli* vrednost 0; v polju *Prekrivanje nizov* določimo širino, za katero se zaporedni stolpci prekrivajo.

Tako na *Stolpčnem*, kot na *Črtnem* ali *Raztresenem* grafikonu, med *Možnostmi nizov* izberemo, ali je posamezen niz prikazan na primarni (levi) oziroma sekundarni (desni) navpični osi. Excel pri osnovni nastavitvi vse nize podatkov prikaže na primarni navpični osi. Če grafikon prikazuje le en niz podatkov, izbira sekundarne navpične osi ni mogoča.

Ostale skupine ukazov v pogovornem oknu **Oblikuj nize podatkov** omogočajo različne grafične nastavitve, kot so oblika črt (npr. neprekinjena, prekinjena, odebeljena črta), oblika oznak (npr. krog, kvadrat, trikotnik,...), barva črt, oznak, površin idr... Ti ukazi so v pogovornem oknu zelo nazorno predstavljeni. Bralca vabimo, da jih preizkusi sam.





### 2.3 Koeficienti in njihovi grafični prikazi, povprečni koeficient

#### Zgled 2.1: 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 študijskih programih na Biotehniški fakulteti v Ljubljani (Vir: Univerza v Ljubljani BF, Poročilo za leto 2007)

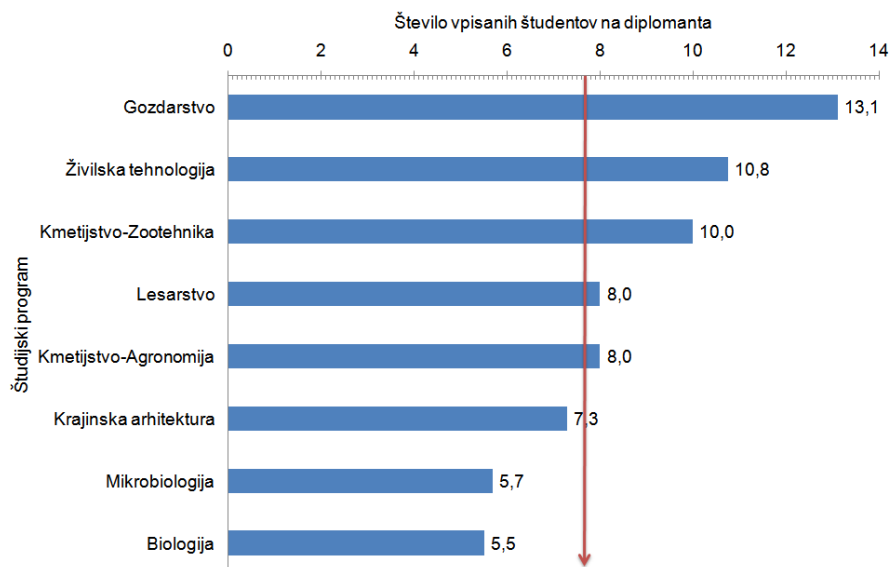
Študijski program	Število vpisanih 2006/07	Število diplomantov 2007
Biologija	429	78
Biotehnologija <sup>6</sup>	167	
Gozdarstvo	210	16
Kmetijstvo-Agronomija	368	46
Kmetijstvo-Zootehnika	270	27
Krajinska arhitektura	182	25
Lesarstvo	168	21
Mikrobiologija	267	47
Živilska tehnologija	344	32

#### VPIS 6

V datoteko VPIS.xlsx dodajte podatke iz tabele 2.1 k podatkom iz tabele 1.3. Za vsak program študija izračunajte razmerje število vpisanih v šol. l. 2006/07 na število diplomantov v letu 2007 na BF. Kako imenujemo izračunana števila?

<sup>6</sup> Na študiju Biotehnologije v letu 2007 še ni bilo diplomantov.

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



Slika 2.1: Grafični prikaz koeficientov: število vpisanih v šol. 1. 2006/07 na diplomanta v letu 2007, rdeča puščica prikazuje povprečni koeficient

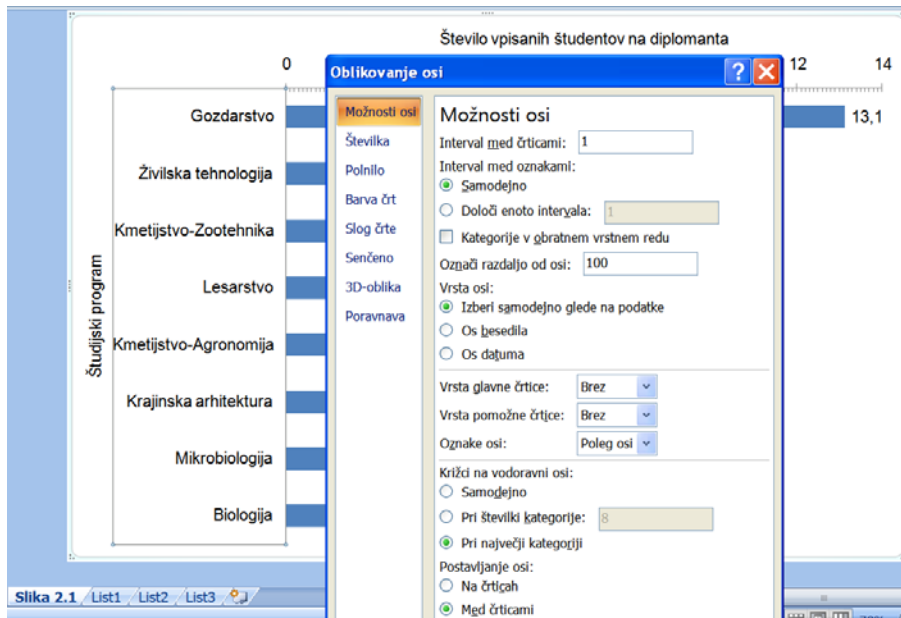
VPIS 7

Grafično prikažite število vpisanih v šol. 1. 2006/07 na diplomanta v letu 2007 po študijskih programih na BF (slika 2.1).

**Navodilo:** najprej tabelo s podatki in izračuni uredite tako, da bo vrstni red študijskih programov na BF določen z velikostjo koeficienta in sicer v naraščajočem vrstnem redu (kartica **Osnovno**, skupina Urejanje ukaz **Razvrsti in filtriraj/Razvrsti od najmanjšega do največjega**).

Študijski program	Število vpisanih 2006/07	Diplomanti 2007	Vpisani/diplomanta
1 Biologija	429	78	5,5
3 Mikrobiologija	267	47	5,7
4 Krajinska arhitektura	182	25	7,3
5 Kmetijstvo-Agronomija	368	46	8,0
6 Lesarstvo	168	21	8,0
7 Kmetijstvo-Zootehnika	270	27	10,0
8 Živilska tehnologija	344	32	10,8
9 Gozdarstvo	210	16	13,1
10 Biotehnologija	167		

Za grafični prikaz izberite grafikon vrste Palični. Navpično os oblikujte tako, da bo izbrana možnost *Križci na vodoravni osi: Pri največji kategoriji*.




Ob »palice« naj se izpišejo tudi vrednosti koeficientov (desni klik »palice«, ukaz **Dodaj oznake podatkov**). ♦


VPIS 8

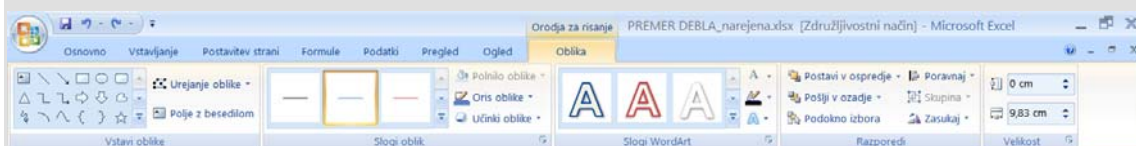
Izračunajte povprečno število vpisanih študentov v šol. l. 2006/07 na diplomanta v letu 2007 na BF ( $\bar{k}$ ) in ga grafično prikažite na sliki 2.1.

**Rešitev:** 
$$\bar{k} = \frac{\sum_{i=1}^K \text{Število vpisanih}_i}{\sum_{i=1}^K \text{Število diplomantov}_i}, \quad K = 8 \text{ je število študijskih programov na BF}$$

**Navodilo:** grafikon dopolnimo z navpično puščico, ki označuje vrednost povprečnega koeficienta. Puščico na grafikon dodamo z klikom na ikono *Puščica* , ki jo najdemo na kartici *Vstavljanje*, v skupini *Ilustracije*, s klikom na ikono **Oblike**:



Klik na ikono  in poteg z miško na grafikonu nariše puščico kot samostojen objekt. Da je puščica postavljena navpično, ob potegu z miško držimo tipko **Shift**. Ko na grafikon dodamo nek objekt, in je ta aktiven (označen), se med karticami v opravljenem traku pojavi **Orodje za risanje** s kartico **Oblika**. Tu najdemo ukaze za vstavljanje in oblikovanje različnih objektov. Klik ukaza **Oris oblike** ponudi široko paleto možnosti za spreminjanje označenega objekta (barva, debelina, slog črte...).



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 študijske programe? ♦

VPIS 10

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

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

to je **tehtana harmonična sredina**. ♦

VPIS 11

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

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

to je **tehtana aritmetična sredina**. ♦

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

### Zgled 2.2: Prometne nesreče v Sloveniji

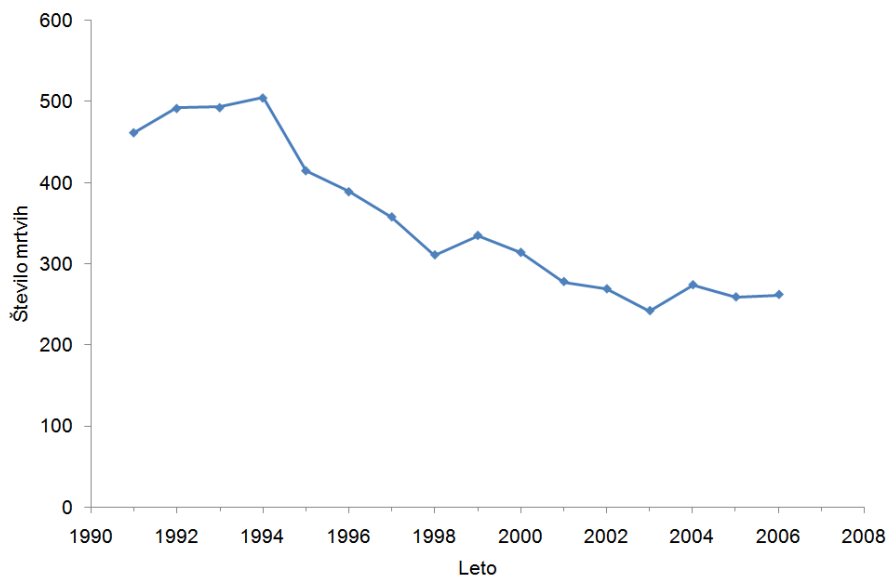
V datoteki PROMETNE NESRECE.xlsx so podatki o številu prometnih nesreč, o številu umrlih v prometnih nesrečah in o številu poškodovanih v prometnih nesrečah po letih v obdobju 1954-2006.

#### PROMETNE NESREČE 1

Prekopirajte podatke za število mrtvih za obdobje 1991-2006 na nov delovni list in ga poimenujete »mrtvi 91-06«. Podatke o številu mrtvih v tem obdobju grafično predstavite (slika 2.2).

**Navodilo:** Časovne vrste običajno prikazujemo z vrsto grafikona *Raztreseni*.

Na delovnem listu najprej označimo podatke za leta in za število mrtvih (celice A1:B17), nato na kartici *Vstavljanje* v skupini *Grafikoni* izberemo vrsto grafikona *Raztreseni*. Excel na tem grafikonu samodejno priredi vodoravni osi podatke iz levega stolpca (*Leto*), podatki v desnem stolpcu (*Število mrtvih*) pa se narišejo kot niz z vrednostmi na navpični osi. ♦



Slika 2.2: Število mrtvih v prometnih nesrečah v Sloveniji v obdobju 1991-2006

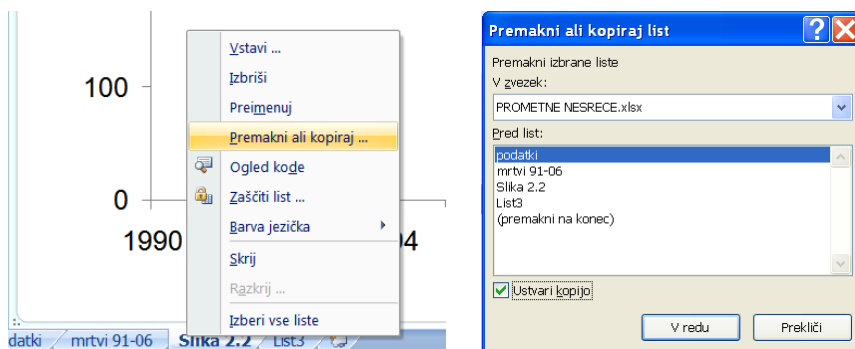
## PROMETNE NESREČE 2

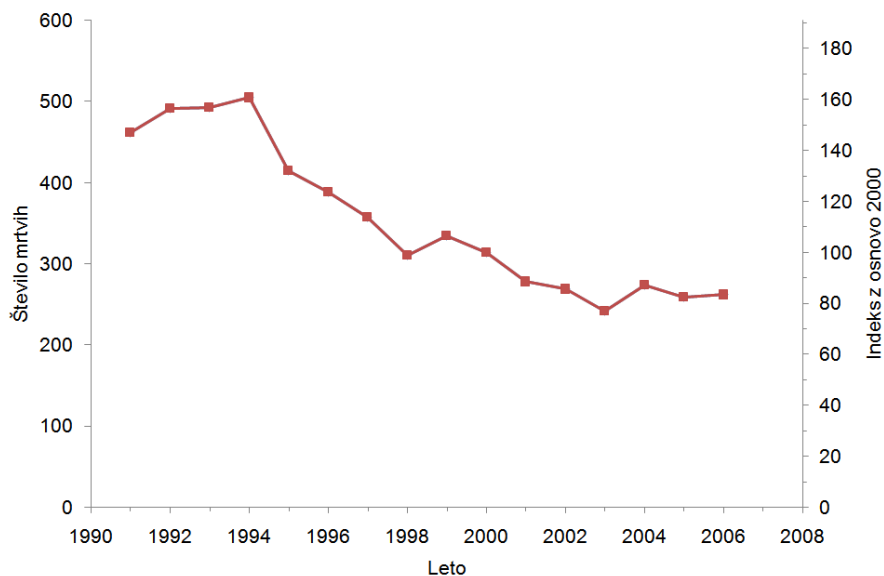
Za število mrtvih izračunajte indekse z osnovo 2000. Izračunane vrednosti za leto 2006 obrazložite. ♦

## PROMETNE NESREČE 3

Število mrtvih in indekse z osnovo 2000 za število mrtvih grafično prikažite na istem grafikonu (Slika 2.3). Vrednosti indeksov z osnovo 2000 naj bodo prikazani na sekundarni osi.

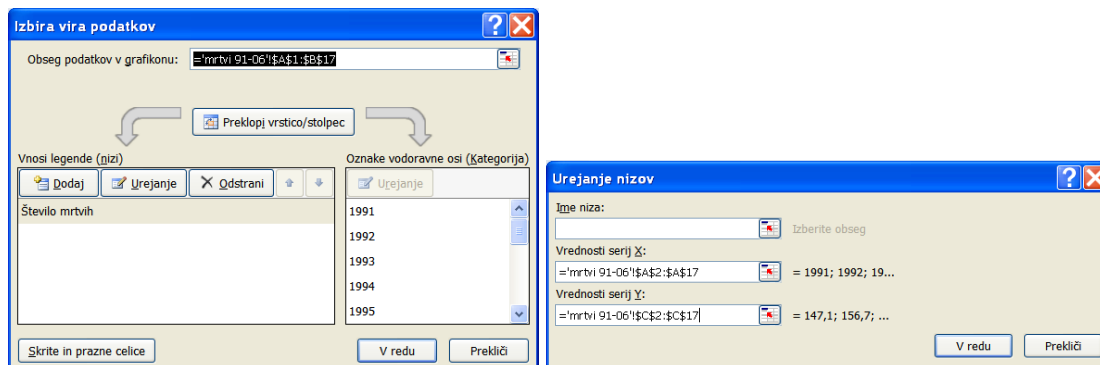
**Navodilo:** najprej naredite kopijo slike 2.2 in jo poimenujte »Slika 2.3«. To naredimo z ukazom **Premakni ali kopiraj**, ki ga najdemo na meniju, ki se odpre z desnim klikom zavihka grafičnega delovnega lista »Slika 2.2«. V pogovornem oknu tega ukaza izberemo možnost *Ustvari kopijo*. S tem se v datoteki pojavi nov grafični delovni list z imenom »Slika 2.2 (2)«, preimenujemo ga v »Slika 2.3«.



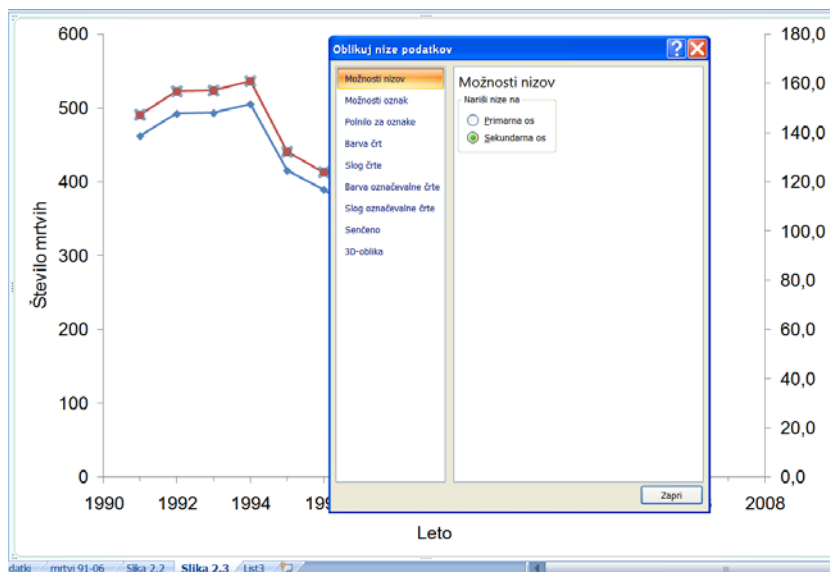


Slika 2.3 Število mrtvih in pripadajoči indeks z osnovo 2000 v Sloveniji v obdobju 1991-2006

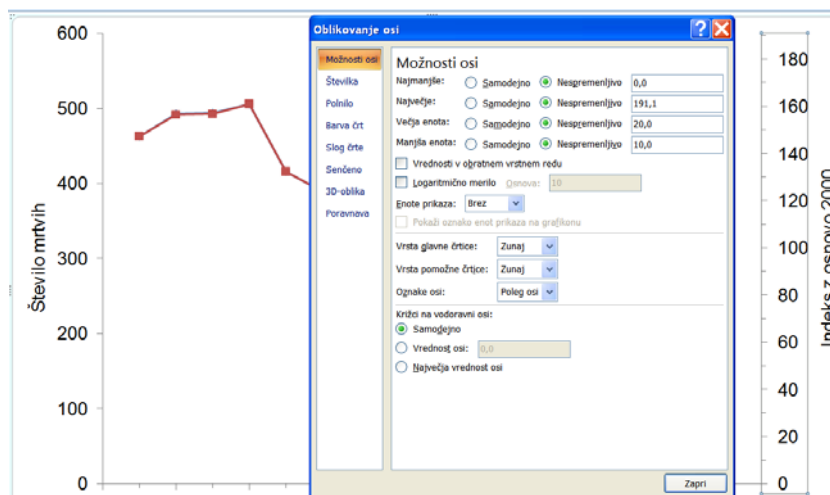
Na sliko 2.3 dodamo indekse z osnovo 2000 z ukazom **Izberi podatke** na kartici *Načrt*. V pogovornem oknu **Izbira vira podatkov** kliknemo gumb **Dodaj** in v pogovornem oknu **Urejanje nizov** določimo vrednosti serij (nizov) X in Y:



Da se vrednosti indeksa z osnovo 2000 pokažejo na sekundarni navpični osi, v pogovornem oknu **Oblikuj nize podatkov** izberemo možnost *Sekundarna os*.

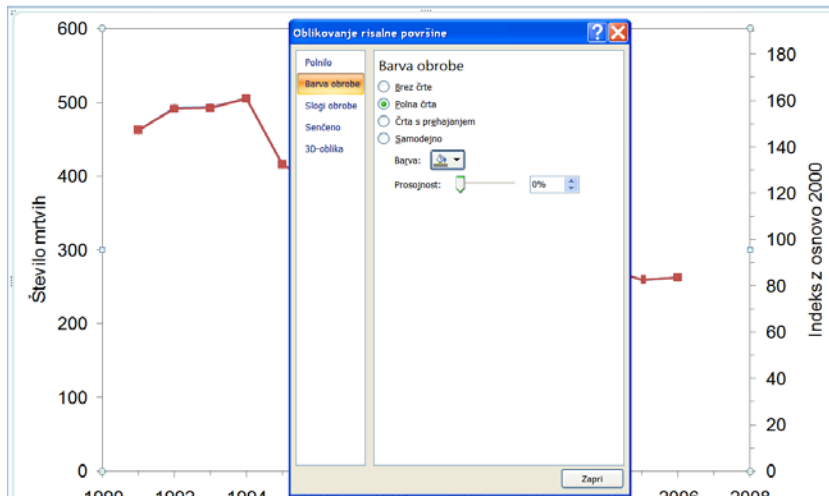


V naslednjem koraku primarno in sekundarno os uskladimo tako, da se niza prekrivata. Za sekundarno navpično os pri nastavitvah v pogovornem oknu **Oblikovanje osi** upoštevamo, da predstavlja 314 celoto, torej 100 % oziroma, da je indeks z osnovo 2000 za 600 mrtvih (maksimalna vrednost na primarni navpični osi) enak 191,1.



Če želimo, da se nariše tudi zgornja obroba risalne površine grafikona, uporabimo ukaz:





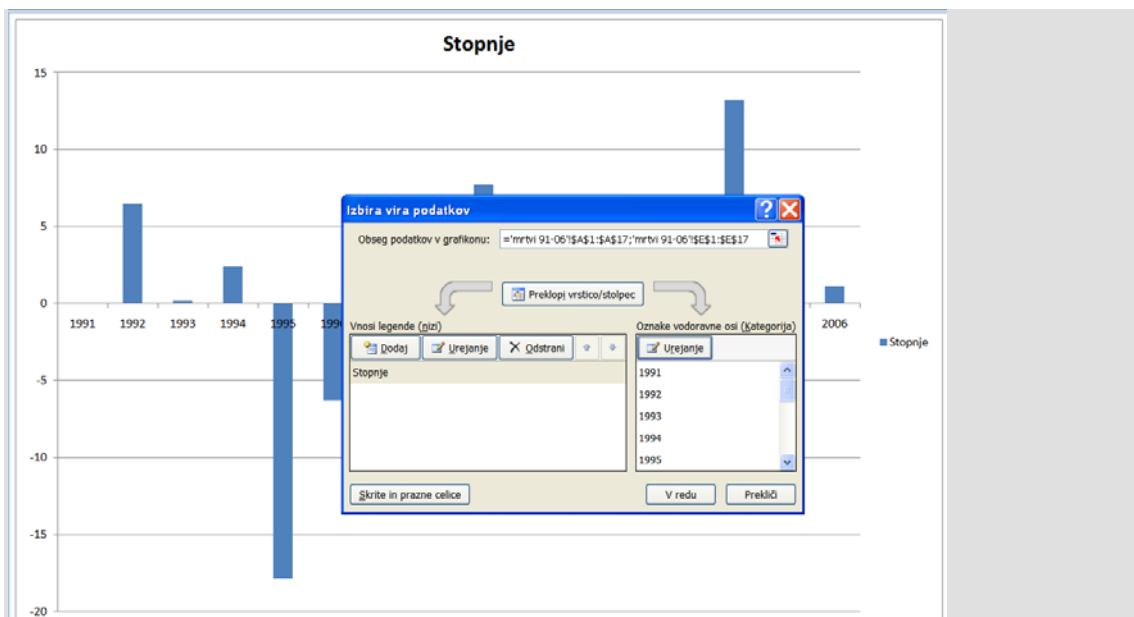
### PROMETNE NESREČE 4

Za število mrtvih izračunajte verižne indekse in letne stopnje rasti. Obrazložite stopnjo rasti za leto 2006. ♦

### PROMETNE NESREČE 5

Letne stopnje rasti grafično predstavite (Slika 2.4).

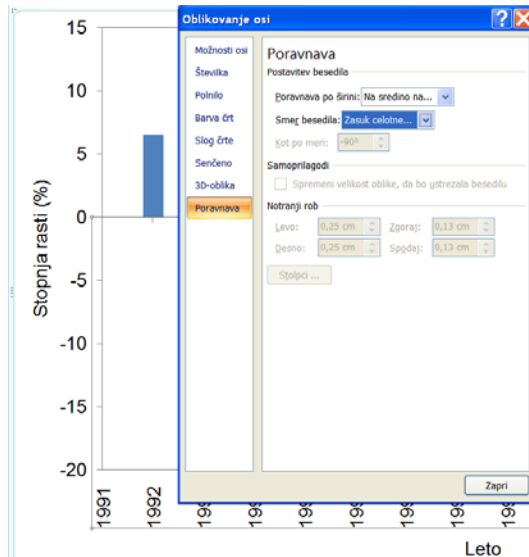
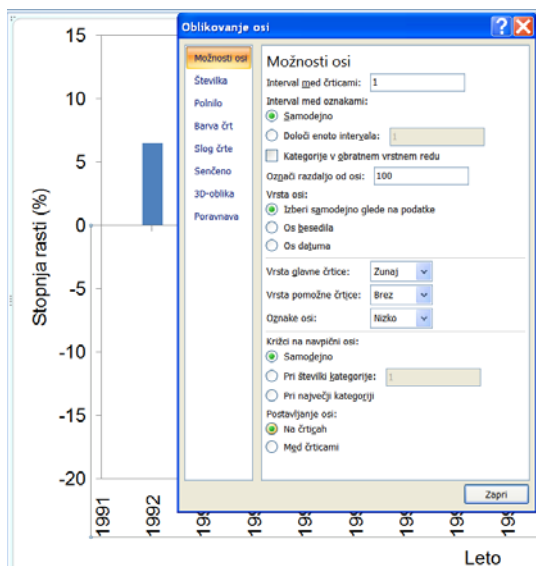
**Navodilo:** Za grafični prikaz stopnje rasti je primeren prikaz s stolpci (vrsta grafikona *Stolpčni*). Označimo celice s podatki o stopnji rasti (E1:E17) in izberemo vrsto grafikona *Stolpčni*. Za oznake vodoravne osi določimo letnice (A2:A17) (**Orodja za grafikone**, kartica *Načrt*, skupina Podatki ukaz **Izberi podatke**, Oznake vodoravne osi, gumb **Urejanje**).

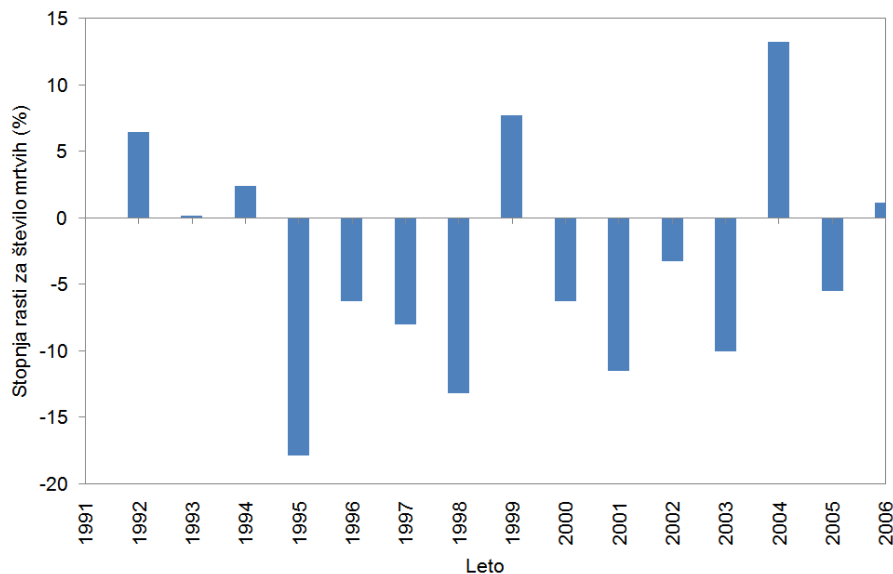


Excel samodejno nariše grafikon tako, da vodoravna os seka navpično os pri vrednosti 0.

Oznake črtic vodoravne osi (letnice) postavite na spodnji rob risalne površine grafikona.

Določite, da so oznake vodoravne osi »na črticah« in da se številke izpišejo pokonci.





Slika 2.4: Letna stopnja rasti števila mrtvih v prometnih nesrečah v Sloveniji v obdobju 1991-2006

### PROMETNE NESREČE 6

Stopnja rasti razvrstite od najmanjše do največje. Kdaj je bilo relativno letno zmanjšanje največje, kdaj najmanjše?

**Navodilo:** razvrščanje ne da ustreznih rezultatov, ker je stopnja rasti izračunana s formulo, ki vrednosti pobira iz različni vrstic delovnega lista. Preden vrednosti razvrstimo, moramo podatke za stopnjo rasti spremeniti v vrednosti. To naredimo s posebnim lepljenjem *Leto* in *Stopnje rasti*:

	A	B	C	D	E
1	Leto	Število mrtvih	Indeks 2000	Verižni indeks	Stopnja rasti
2	1991	462	147,1		
3	1992			106,5	6,5
4	1993			100,2	0,2
5	1994			102,4	2,4
6	1995			82,2	-17,8
7	1996			93,7	-6,3
8	1997			92,0	-8,0
9	1998			86,9	-13,1
10	1999			107,7	7,7
11	2000			93,7	-6,3
12	2001			88,5	-11,5
13	2002			96,8	-3,2
14	2003			90,0	-10,0
15	2004	274	87,3	113,2	13,2
16	2005	259	82,5	94,5	-5,5
17	2006	262	83,4	101,2	1,2

Leto	Stopnja rasti
1991	13,2
1992	6,5
1993	0,2
1994	2,4
1995	-17,8
1996	-6,3
1997	-8,0
1998	-13,1
1999	7,7
2000	-6,3
2001	-11,5
2002	-3,2
2003	-10,0
2004	13,2
2005	-5,5
2006	1,2

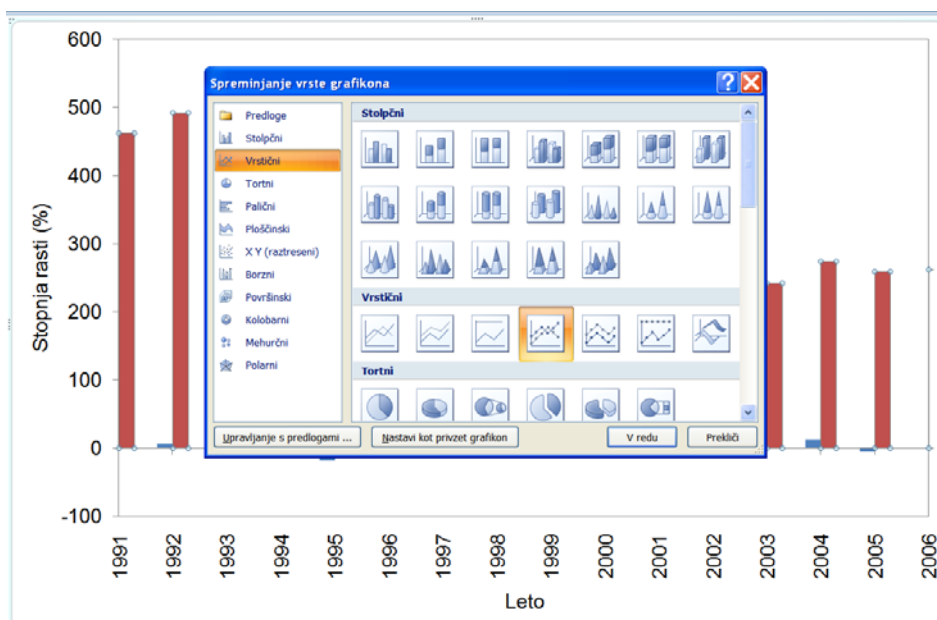
  

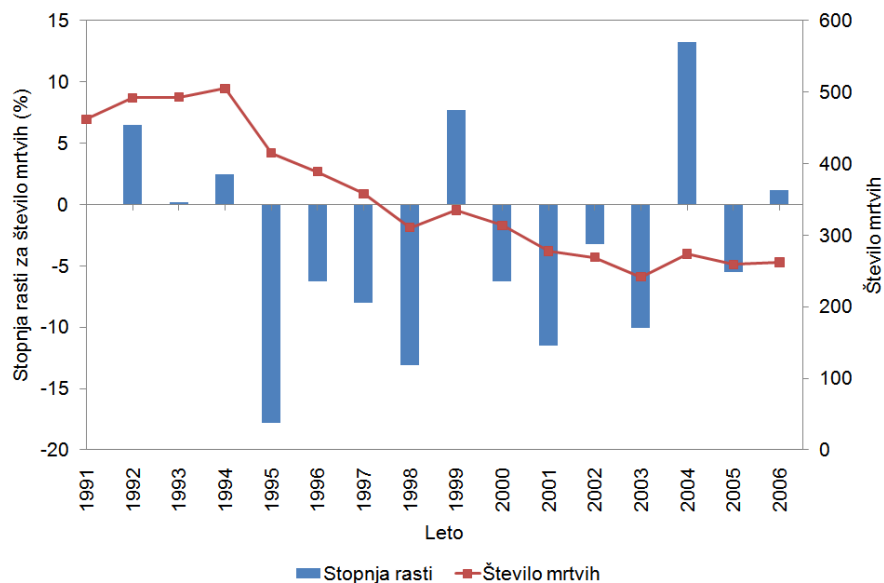
Leto	Stopnja rasti
1995	-17,8
1998	-13,1
2001	-11,5
2003	-10,0
1997	-8,0
2000	-6,3
1996	-6,3
2005	-5,5
2002	-3,2
1993	0,2
2006	1,2
1994	2,4
1992	6,5
1999	7,7
2004	13,2
1991	13,2

#### PROMETNE NESREČE 7

Grafično prikažite število mrtvih in stopnjo rasti na isti sliki (Slika 2.5).

**Navodilo:** naredite kopijo slike 2.4 in jo poimenujte »Slika 2.5«. Dodajte število mrtvih na sliko. Nizu podatkov, ki prikazuje število mrtvih, spremenite vrsto grafikona na Vrstični in vrednosti prikažite na sekundarni osi.





Slika 2.5: Letna stopnja rasti za število mrtvih in število mrtvih v prometnih nesrečah v Sloveniji v obdobju 1991-2006

### Zgled 2.3: Naravno gibanje prebivalstva v Sloveniji

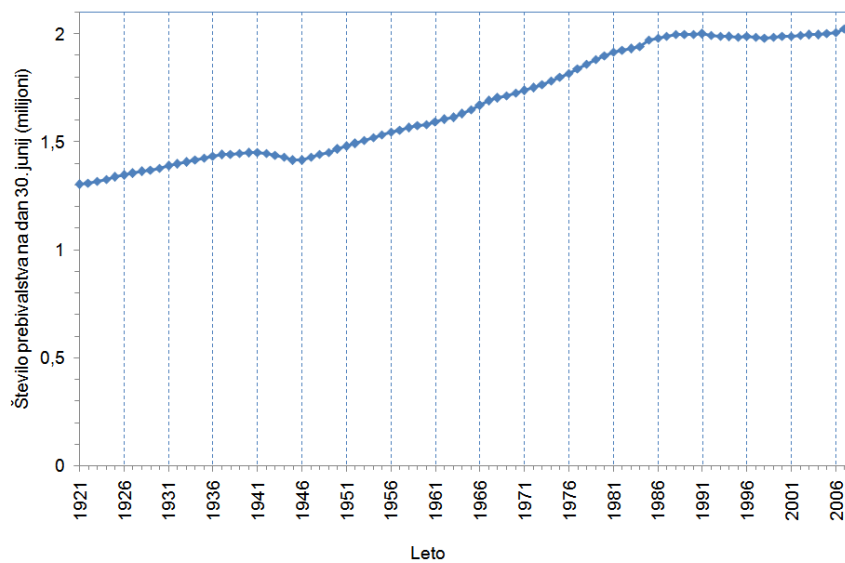
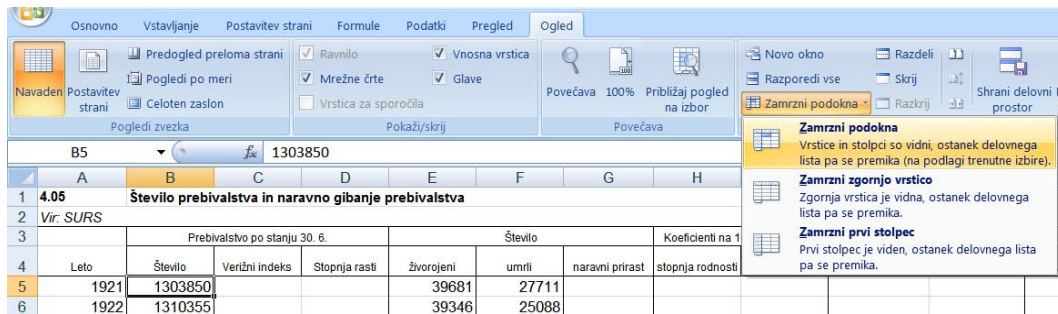
V datoteki PREBIVALSTVO.xlsx so podatki o številu stalnega prebivalstva, o številu živorojenih, o številu umrlih po letih v Sloveniji v obdobju 1921-2008 (Vir: spletna stran SURS). Število prebivalstva je ocenjeno stanje na dan 30.6., število živorojenih in število umrlih se nanašata na celotno leto.

#### PREBIVALSTVO 1

Datoteko PREBIVALSTVO.xlsx shranite z delovnim imenom PREBIVALSTVO analiza.xlsx. Vse izračune in grafične prikaze naredite v datoteki PREBIVALSTVO analiza.xlsx. Za boljši pregled podatkov na delovnem listu »Podatki in izračuni« 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 **Zamrzni podokna** na kartici **Ogled** v skupini Okno. Rezultat tega ukaza se odraža na videzu zaslona ob tem, ko se na delovnem listu predstavite do spodnjega desnega roba tabele s podatki.<sup>7</sup> ♦

<sup>7</sup> Z enkratno uporabo omenjenega ukaza na izbranem delovnem listu v datoteki se na kartici **Ogled** v skupini Okno pojavi ukaz **Odmrzni podokna**.

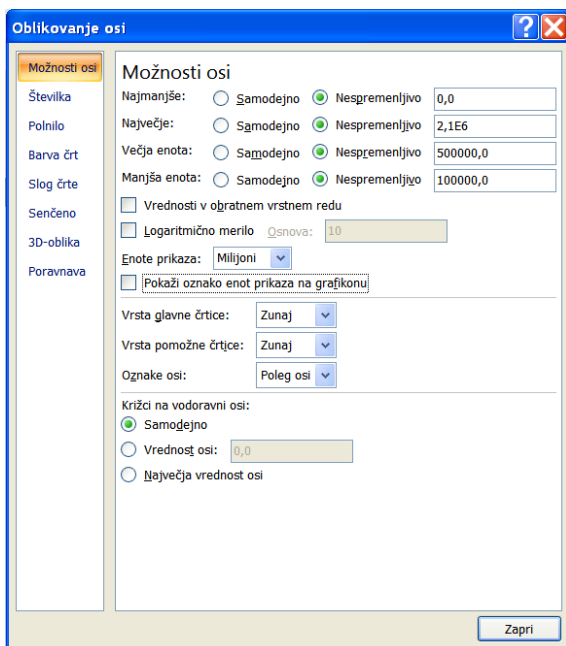


Slika 2.6: Število prebivalstva v Sloveniji na dan 30. junija v obdobju 1921-2008

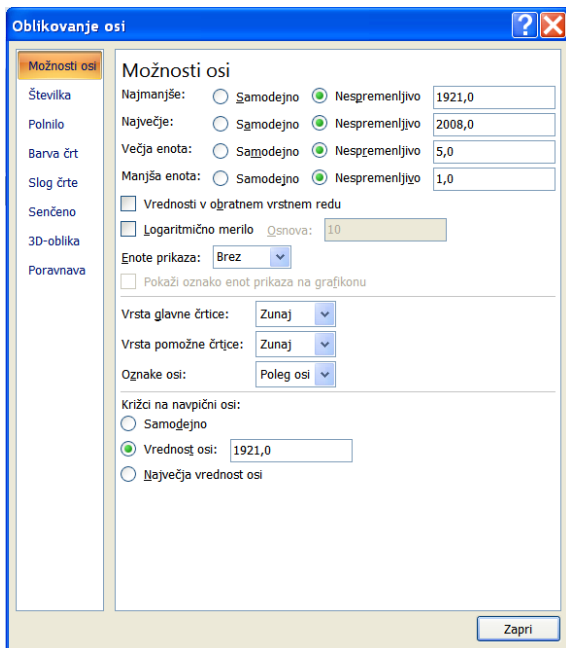
## PREBIVALSTVO 2

Grafično predstavite število prebivalstva v Sloveniji v obdobju 1921-2008 (slika 2.6).

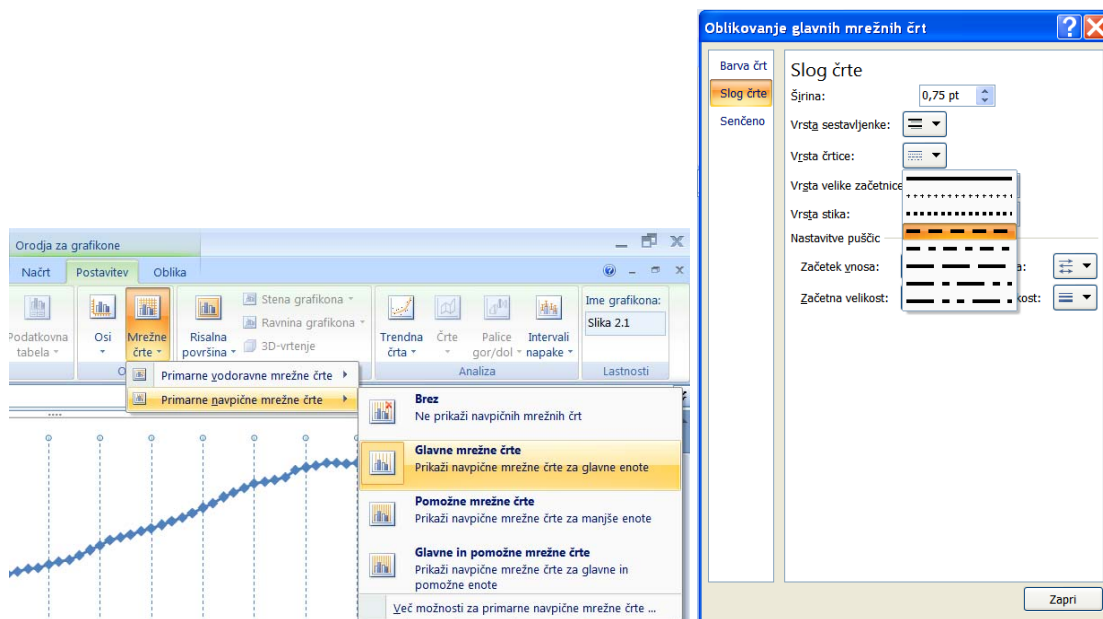
**Navodilo:** navpično os oblikujte tako, da bodo oznake črtic na osi izpisane v milijonih in da bo maksimalna prikazana vrednost 2100000 (Excel včasih samodejno izpiše kot 2,1E6). Poleg tega naj bodo glavne črtice na osi 500000 vsak sebi, vidne pa morajo biti tudi pomožne črtice 100000 vsak sebi:



Na vodoravni osi naj se letnice izpišejo vsako peto leto, v obdobju 1921-2008. Ko določate vrednosti v skupini ukazov *Možnosti osi*, možnosti za samodejno nastavitve ne smejo biti izbrane:



V okviru **Orodja za grafikone** na kartici *Postavitev* ustrezno določite položaj mrežnih črt:

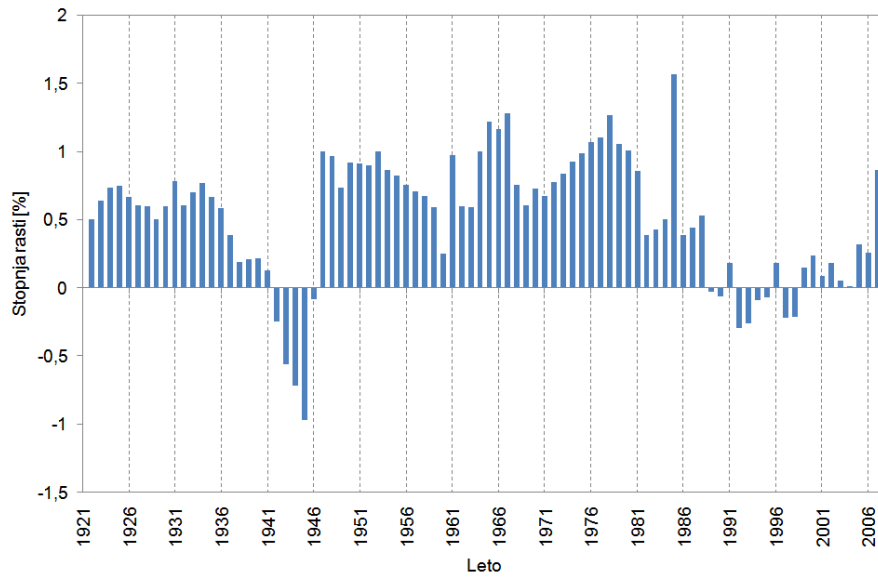


### PREBIVALSTVO 3

Izračunajte verižne indekse in stopnje rasti za število prebivalstva v Sloveniji v obdobju 1921-2008.

**Navodilo:** Formulo napišite samo za leto 1922, nato izvedite »hitro kopiranje«: z miško se postavite v desni spodnji vogal celice s formulo, ko se oblika in barva križca spremenita, križec dvokliknite. S tem dobimo celotno časovno vrsto verižnih indeksov. ◆





Slika 2.7: Letna stopnja rasti za število prebivalstva na dan 30. junij v Sloveniji v obdobju 1921-2008

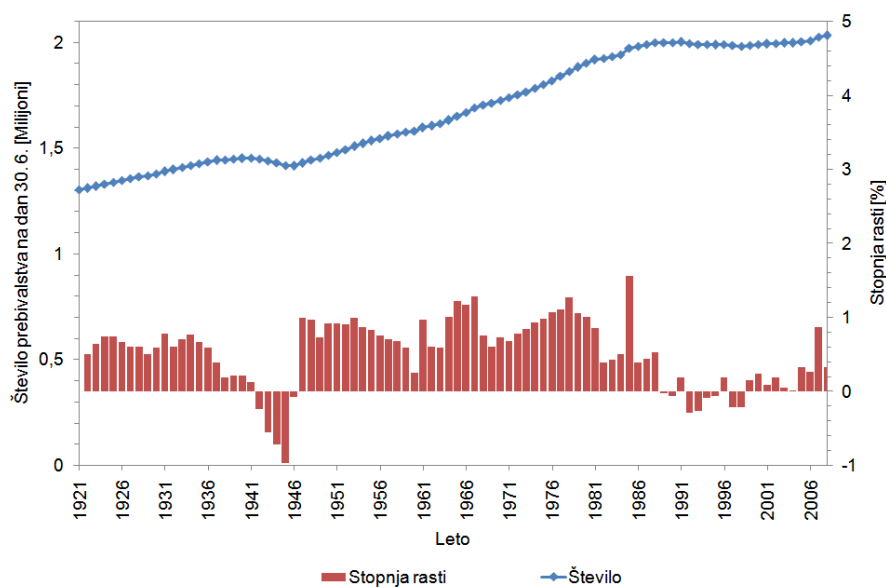
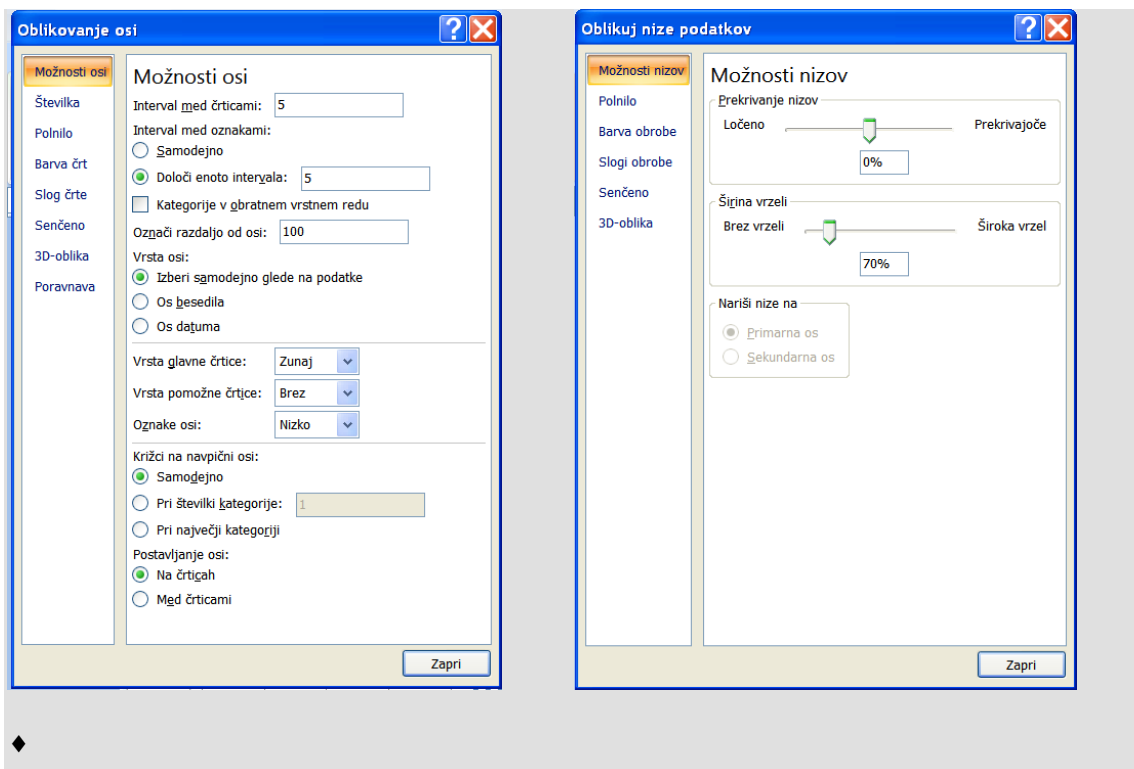
#### PREBIVALSTVO 4

Letno stopnjo rasti za število prebivalstva v Sloveniji grafično prikažite (slika 2.7).

**Navodilo:** pokončne mrežne črte nastavite na 5 let narazen, prav tako naj bodo 5 let narazen oznake črtic na vodoravni osi.

Določite, da so oznake vodoravne osi »na črticah«.

Širina vrzeli med stolpci naj bo 70 % (ukaz **Oblikuj nize podatkov**, kartica ***Možnosti nizov***, možnost *Širina vrzeli*)



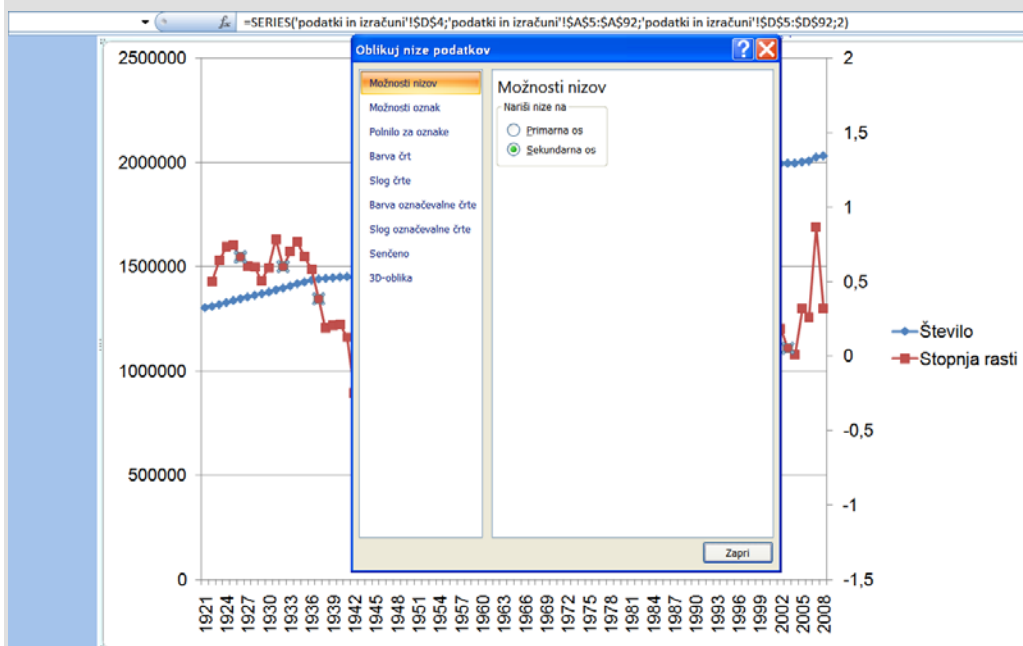
Slika 2.8: Število prebivalstva v Sloveniji na dan 30. junija v obdobju 1921-2008 (primarna os) in pripadajoča stopnja rasti (sekundarna os)

#### PREBIVALSTVO 5

Grafično predstavite število prebivalstva in njegovo letno stopnjo rasti v Sloveniji v obdobju 1921-2008 na istem grafikonu (slika 2.8).

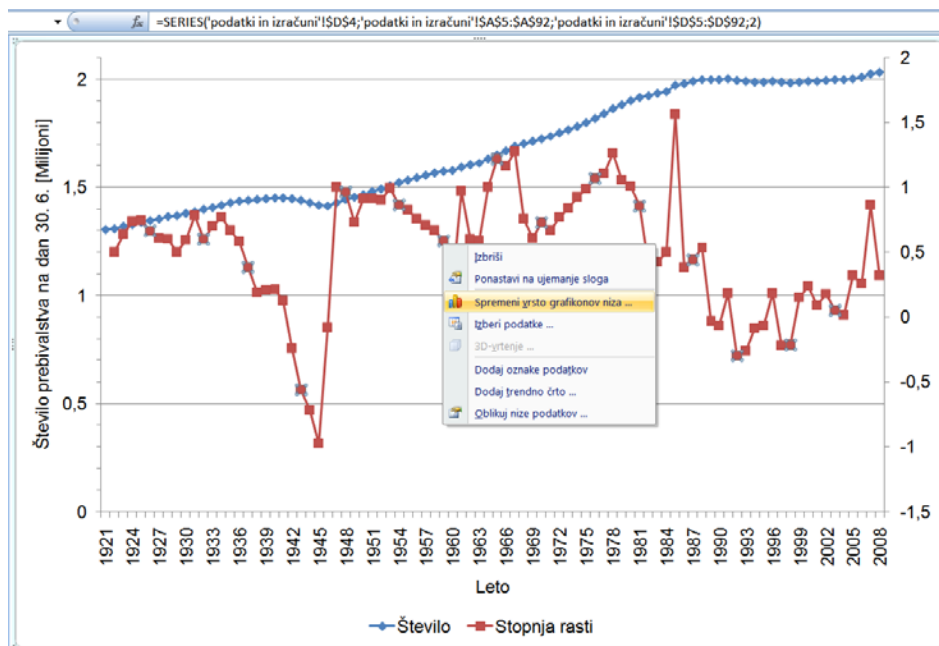
**Navodilo:** označimo celice s podatki o srednjem številu prebivalstva in o stopnji rasti (B4:B92; D4:D92, celica D5 mora biti prazna) in izberemo *Črtni grafikon*<sup>8</sup>, ki omogoča, da kasneje za prikaz stopnje rasti izberemo stolpce. Tudi v tem primeru določimo letnice (A5:A92) za oznake vodoravne osi.

V naslednjem koraku določimo, da se vrednosti za stopnjo rasti pokažejo na sekundarni navpični osi: s klikom izberemo drugi niz podatkov, uporabimo ukaz **Oblikuj niz podatkov** in v skupini *Možnosti nizov*, izberemo možnost *Sekundarna os*.

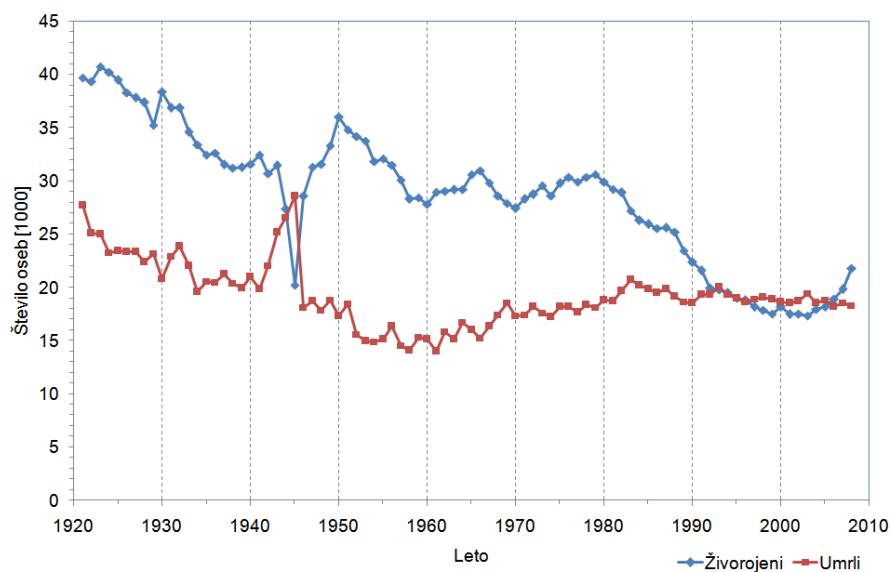


Črto, ki prikazuje stopnje rasti, spremenimo v stolpec tako, da ta niz najprej izberemo in uporabimo ukaz **Spremeni vrsto grafikonov niza** na pomožnem meniju, ki se pokaže ob desnem kliku miške ter izberemo vrsto *Stolpčni*.

<sup>8</sup> Sicer za grafični prikaz časovne vrste v splošnem uporabljamo grafikon vrste *Raztreseni*; tu zaradi zahteve, da mora biti ena vrsta prikazana s stolpci, uporabimo črtni grafikon.



Vodoravno, primarno in sekundarno navpično os oblikujte tako, kot kaže slika 2.8. ♦



Slika 2.9: Število živorojenih in umrlih v Sloveniji v obdobju 1921-2008

#### PREBIVALSTVO 6

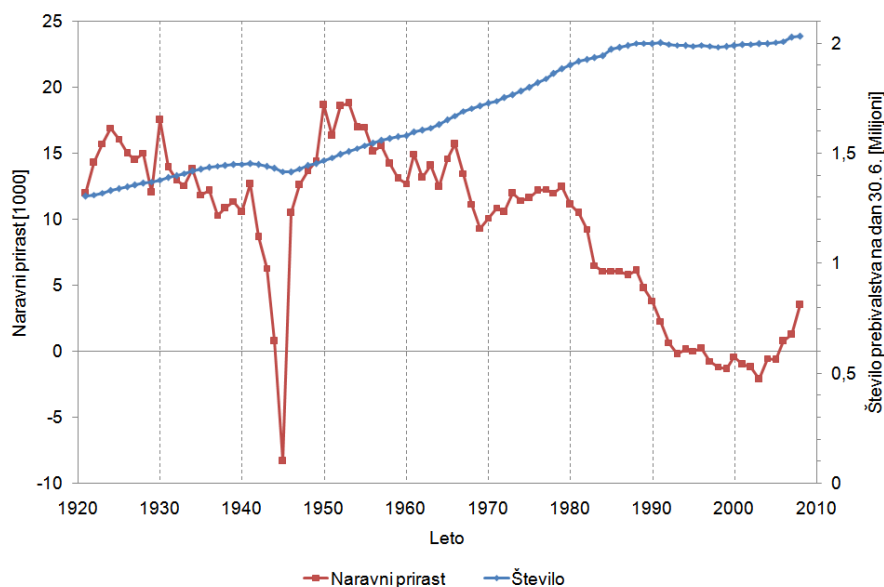
Na istem grafikonu prikažite podatke o številu živorojenih in umrlih (grafični prikaz dveh časovnih vrst) (slika 2.9).

**Navodilo:** Če na delovnem listu izberemo tri stolpce podatkov: A4:A92, E4:F92, grafikon vrste *Raztreseni* podatke A5:A92 (leta) prikaže na vodoravni osi, podatke E5:E92 kot prvi niz in podatke F5:F92 kot drugi niz. Podatek v celici A4 predstavlja naslov vodoravne osi, a na grafikonu ni prikazan; podatka v celicah E4 in F4 pa predstavljata imeni nizov, ki se pojavita v legendi.

Grafikon oblikujte tako, da bo čim bolj podoben sliki 2.9. Bodite pozorni na oblikovanje vodoravne in navpične osi. ♦

### PREBIVALSTVO 7

Izračunajte naravni prirast v posameznih letih. Naravni prirast je razlika med številom živorojenih in številom umrlih. Vrednosti za leti 1945 in 2008 obrazložite. ♦



Slika 2.10: Naravni prirast (primarna os) in število prebivalcev na dan 30. 6. (sekundarna os) v Sloveniji po letih (1921-2008)

### PREBIVALSTVO 8

Na istem grafikonu predstavite naravni prirast (primarna navpična os) in število prebivalstva (sekundarna navpična os) v Sloveniji v obdobju 1921-2008 (slika 2.10).

**Navodilo:** Označimo celice A4:B92 in G4:G92 ter izberemo *Raztreseni* grafikon. Vrednosti za število prebivalcev postavimo na sekundarno navpično os. ♦

## PREBIVALSTVO 9

Izračunajte verižne indekse in stopnjo rasti za število živorojenih in za število umrlih za obdobje 1921-2008.

Izračunane stopnje rasti vsebinsko smiselno grafično prikažite (dva grafikona na delovnih listih z imeni »Slika 2.7a« in »Slika 2.7b«). ♦

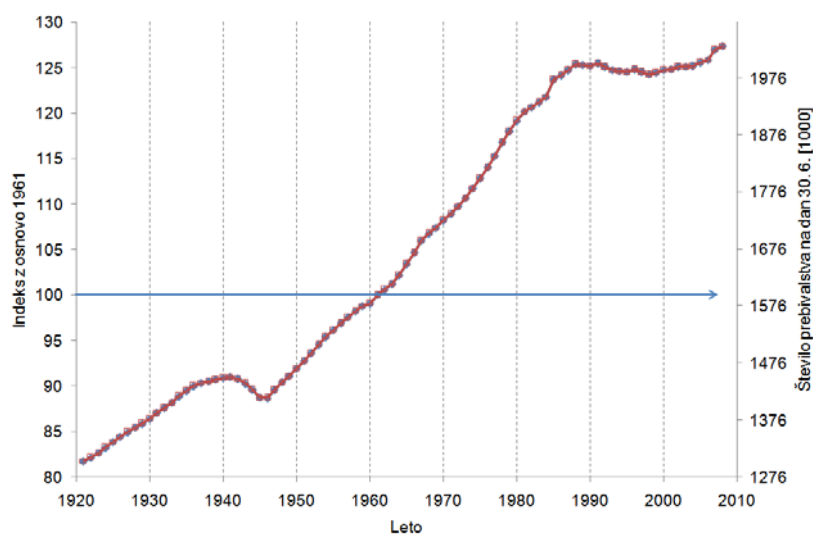
## PREBIVALSTVO 10

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 2.9a«). Sliko primerjajte s sliko 2.9. 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 2008. Na sliko dodajte število prebivalstva z vrednostmi na sekundarni osi (slika 2.11). ♦



Slika 2.11: Indeksi z osnovo 1961 za število prebivalstva (primarna os) in število prebivalstva na dan 30. 6. (sekundarna os) v Sloveniji v obdobju 1921-2008. Na grafikonu se prekrivata črti za dva niza.

**Navodilo:** pazite na usklajenost primarne in sekundarne navpične osi. Če skale na primarni navpični osi ne spreminjamo, potem minimum in maksimum sekundarne navpične osi izračunamo takole:

	Minimum	Maksimum
Primarna os	80	130
Sekundarna os	$1595450 * 0,8 = 1276360$	$1595450 * 1,3 = 2074085$
Sekundarna os (zaokroženo na tisoč)	1276000	2074000

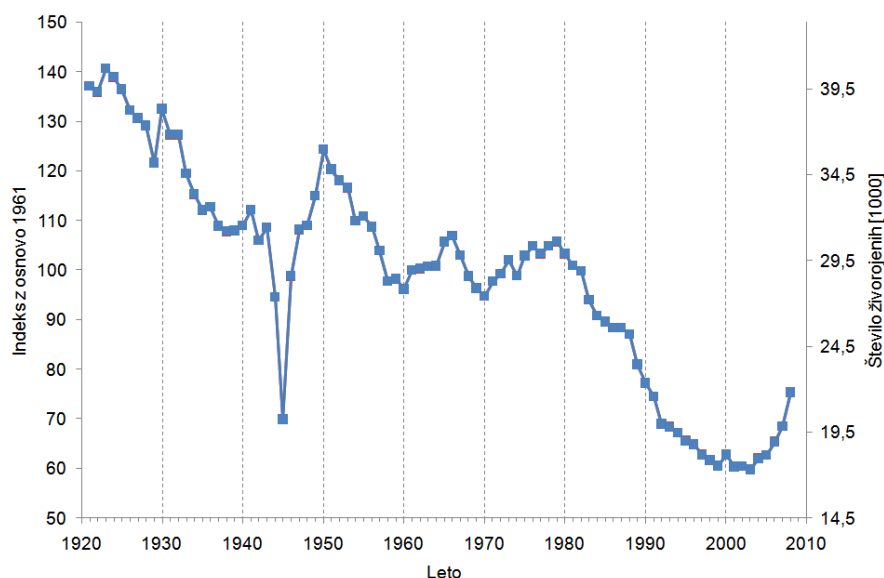
Izračunane vrednosti za maksimum in minimum na sekundarni navpični osi v tem primeru lahko zaokrožimo na tisočice natančno, da dobimo za oznake črtic cela števila, črti obeh nizov pa sta pri tej grafični natančnosti prekriti. ♦

### PREBIVALSTVO 12

Izračunajte indekse z osnovo 1961 za število živorojenih in za število umrlih v Sloveniji v danem obdobju. Izračune za leti 1945 in 2008 obrazložite. ♦

### PREBIVALSTVO 13

Izračunane indekse z osnovo 1961 za število živorojenih grafično prikažite skupaj z originalnimi podatki, kot kaže slika 2.12. ♦



Slika 2.12: Indeksi z osnovo 1961 za število živorojenih (primarna os) in število živorojenih (sekundarna os) v Sloveniji v obdobju 1921-2008

### PREBIVALSTVO 14

Izračunane indekse z osnovo 1961 za število umrlih grafično prikažite skupaj z originalnimi podatki. Grafikon postavite na grafični delovni list z imenom »Slika 2.12a«. ♦

**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	
	Leto 2003	Leto 2006	Leto 2003	Leto 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.xlsx. 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.xlsx so podatki o številu gospodinjstev in o številu moških in žensk, po popisih od leta 1921 dalje na ozemlju Slovenije (tabela 2.3).

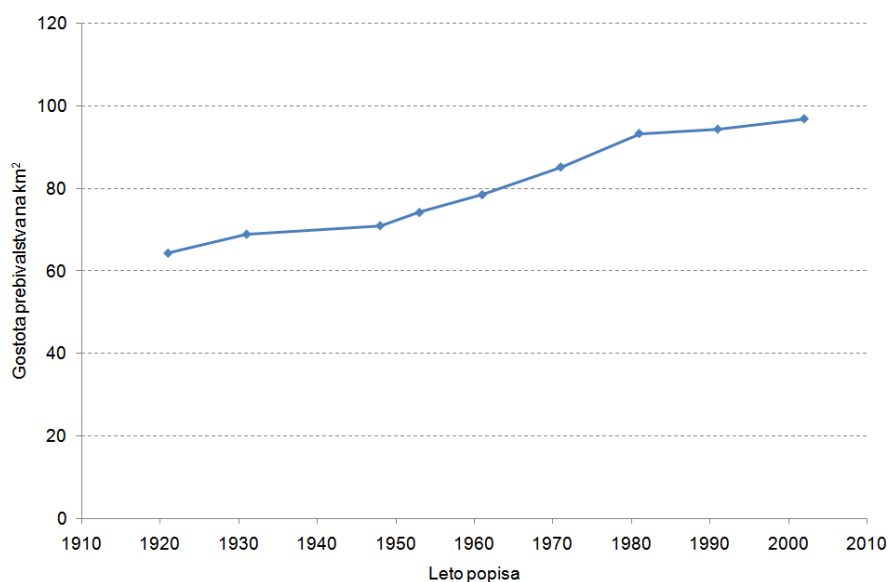


Tabela 2.3: Gospodinjstva in prebivalstvo v Sloveniji, popisi 1921 – 2002.

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 1

Izračunajte število prebivalcev za posamezni popis. ♦



Slika 2.13: Gostota prebivalstva v Sloveniji ob popisih, upoštevana je današnja površina Slovenije

### GOSTOTA PREBIVALSTVA 2

Izračunajte gostoto prebivalstva za posamezni popis, ob tem upoštevajte površino sedanje Slovenije (20273 km<sup>2</sup>). Izračune grafično prikažite (slika 2.13). ♦

### GOSTOTA PREBIVALSTVA 3

Izračunajte povprečno število članov gospodinjstva za posamezni popis. ♦

GOSTOTA PREBIVALSTVA 4

Izračunajte število žensk na enega moškega za posamezni popis. ♦

### 3 Opisne statistike in frekvenčna porazdelitev

#### Novi pojmi:

- ukaz **Vstavi funkcijo** in drugi ukazi na kartici *Formule*
- orodje **Analiza podatkov/Descriptive statistics**

#### Datoteke s podatki:

PADAVINE.xlsx

LJUBLJANSKI MARATON.xlsx

PREMER DEBLA.xlsx

KATJA.xlsx

#### Delovne datoteke:

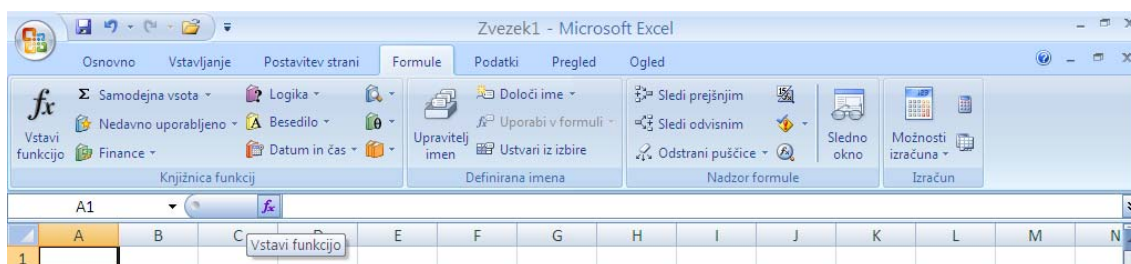
PREBIVALSTVO analiza.xlsx



SPLET.xlsx

ROJSTNO IME.xlsx

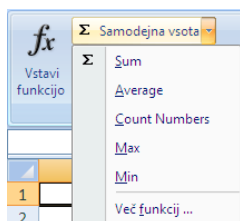
### 3.1 Uporaba vgrajenih funkcij

V Excelu lahko uporabimo različne že vnaprej vgrajene funkcije, ki so razporejene v večje skupine (zvrsti). Ukazi, ki omogočajo uporabo vgrajenih funkcij so na kartici *Formule*:



Ob uporabi posamezne funkcije najprej označimo celico na delovnem listu, v katero se bo izpisal rezultat, nato uporabimo ukaz **Vstavi funkcijo** 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 Sum, Average, Count Numbers, Max, Min

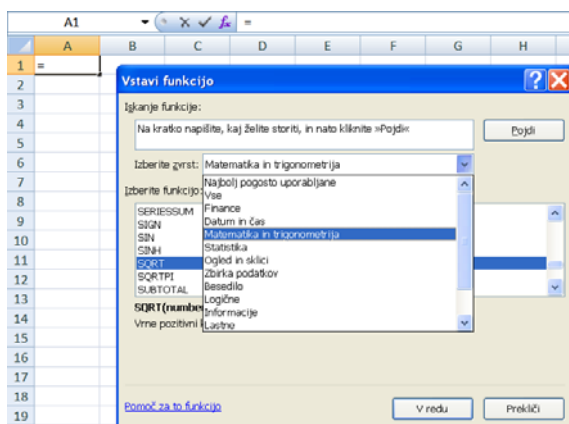
(vsota, povprečje, število številskih vrednosti, maksimum, minimum), ali ukaz **Več funkcij**.



V pogovornem oknu *Vstavi funkcijo* določimo funkcijo v dveh korakih.

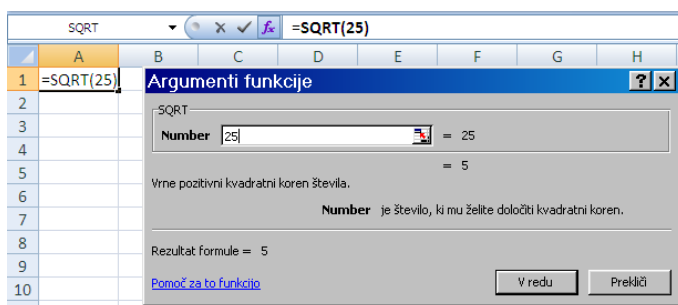
### 1. korak

V pogovornem oknu sta dva seznama: *Izberite zvrst* - seznam zvrsti 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. Na primer kvadratni koren najdemo v skupini funkcij *Matematika in trigonometrija*, ime funkcije je **SQRT**.



S klikom na gumb **V redu** se odpre pogovorno okno **Argumenti funkcije**.

## 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.

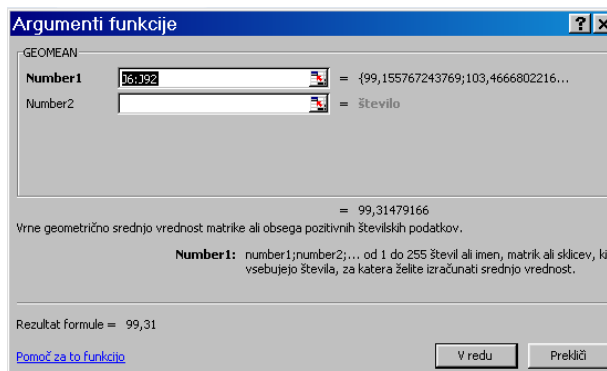
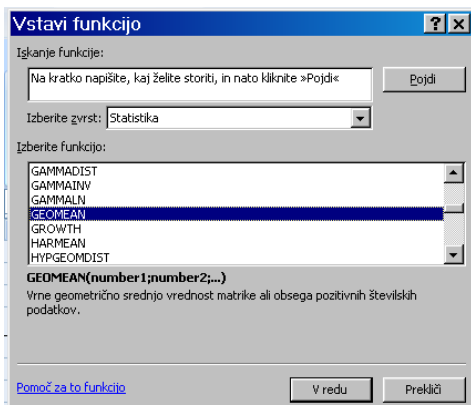
V polju *Rezultat formule* se po vnosu argumentov izpiše rezultat. Z gumbom *V redu* se rezultat pokaže v označeni celici.

### 3.1.1 Geometrijska sredina

Povprečni verižni indeks ( $\bar{I}$ ) izračunamo kot geometrijsko sredino verižnih indeksov ( $I_{i/i-1}$ ) ali pa na podlagi prvega ( $y_1$ ) in zadnjega ( $y_T$ ) podatka v časovni vrsti dolžine  $T$ :

$$\bar{I} = \sqrt[T]{I_{2/1} \cdot I_{3/2} \cdot \dots \cdot I_{T/T-1}} = \sqrt[T]{\frac{y_T}{y_1}} \cdot 100$$

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



J94		=GEOMEAN(J6:J92)			
	A	J	K	L	M
1	4,05				
2	Vir: SURS				
3		Verižni indeksi		Stopnje rasti	
4	Leto	živorojeni	umrli	živorojeni	umrli
88	2004	103,7	95,2	3,7	-4,8
89	2005	101,1	101,6	1,1	1,6
90	2006	104,3	96,6	4,3	-3,4
91	2007	104,7	102,2	4,7	2,2
92	2008	110,1	98,5	10,1	-1,5
93		Povprečni verižni indeks:		Povprečna stopnja rasti:	
94		99,31	99,52	-0,69	-0,48

## PREBIVALSTVO 15

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

**Rešitev:** povprečno stopnjo rasti ( $\bar{S}$ ) izračunamo tako, da od povprečnega verižnega indeksa odštejemo 100:

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

**Navodilo:** izračun povprečnega verižnega indeksa na podlagi prvega ( $y_1$ ) in zadnjega ( $y_T$ ) podatka v časovni vrsti dolžine  $T$  naredimo s funkcijo POWER(number; power), ki ima dva argumenta: »number« je število, katerega potenciramo in »power« je potenca (na primer 1/2, če računamo 2. koren).

**Rešitev:**  $T = 88$

=POWER(E92/E5;1/(87))\*100.

Za izračun pri številu živorojenih smo za argument funkcije »number« napisali formulo E92/E5, za potenco smo uporabili 87. koren. ♦

## 3.2 Napake pri uporabi formul

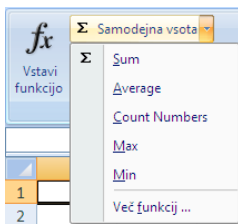
Če pri računanju s formulami naredimo napako, Excel v celico namesto rezultata napiše oznako za vrsto napake, npr.:

#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 (na primer 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  $\Sigma$ :



Funkcija **AVERAGE** izračuna aritmetično sredino, **COUNT** prešteje številske podatke, **MAX** poišče maksimum in funkcija **MIN** poišče minimum. Spodaj je navedenih še nekaj drugih funkcij v skupini **Statistika**<sup>9</sup>, ki jih uporabljamo pri računanju opisnih statistik:

<sup>9</sup> 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.

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.		

### Zgled 3.1: 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.xlsx. ♦

#### SPLET 2

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

**Navodilo:** bodite pozorni na izračun modusa, med podatki imamo namreč dve vrednosti, ki se ponovita trikrat: 7 in 2. Funkcija MODE za podatke, ko je modulusov več, vrne samo en modus. V našem primeru vrne kot rezultat samo vrednost 7. V primeru, če so v podatkih same različne vrednosti, ta funkcija vrne vrednost #N/V, kar pomeni, da modusa danim podatkom ne moremo določiti. ♦

#### 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 »quart« 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:

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

#### 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 formul<sup>10</sup> (levo); izpis rezultatov (desno)

	A	B	C	D
1	<i>DSS</i>		Število podatkov	=COUNT(A2:B15)
2	10		Minimum	=MIN(A2:A15)
3	7		Maksimum	=MAX(A2:A15)
4	14		Modus	=MODE(A2:A15)
5	20		Mediana, $Q_2 = Me$	=MEDIAN(A2:A15)
6	2		Povprečje	=AVERAGE(A2:A15)
7	2		$Q_1$	=QUARTILE(A2:A15;1)
8	5		$Q_2 = Me$	=QUARTILE(A2:A15;2)
9	2		$Q_3$	=QUARTILE(A2:A15;3)
10	7		Q	=D9-D7
11	4		Meja za sp. osamelce	=D7-1,5*D10
12	0		Meja za zg. osamelce	=D9+1,5*D10
13	7		$C_{33,3}$	=PERCENTILE(A2:A15;1/3)
14	5		$C_{66,7}$	=PERCENTILE(A2:A15;2/3)
15	15		Varianca	=VAR(A2:A15)
16			Standardni odklon	=STDEV(A2:A15)
17			KV%	=D16/D6*100

#### SPLET 6

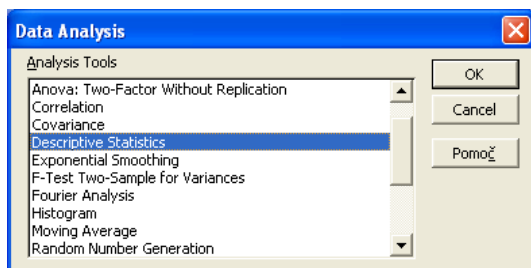
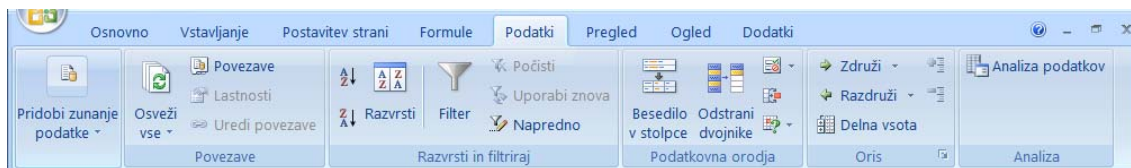
Za *DSS* izračunajte še varianco ( $s^2$ ), standardni odklon ( $s$ ) in koeficient variacije ( $KV\%$ ) (tabela 3.1). ♦

### 3.3.2 Uporaba orodja Descriptive Statistics

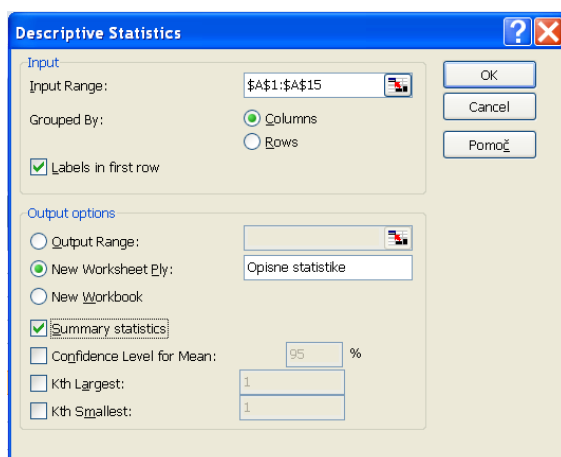
Ogledali si bomo uporabo orodja **Descriptive Statistics**, ki združuje več funkcij. V splošnem orodju, ki jih uporabljamo pri statistični analizi podatkov, najdemo na kartici **Podatki** v zadnji skupini Analiza pod ukazom **Analiza podatkov**<sup>11</sup>.

<sup>10</sup> Izhis formul dosežemo s tem, da na kartico **Formule** v skupini Nadzor formule kliknemo na ikono **Pokaži formule**.

<sup>11</sup> Če skupine Analiza in ukaza **Analiza podatkov** na kartici **Podatki** ni, pomeni, da to orodje v trenutni nastavitvi Excela še ni nameščeno. Za namestitev orodja sledite navodilom v Dodatku A.



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



- 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**).

<sup>12</sup> Ukazi v pogovornem oknu niso prevedeni v slovenščino.

- 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 (Opisne statistike)** 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 **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 delovni list »Opisne statistike« 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 uporabo ikone  (zmanjšanje) ali  (povečanje) na kartici **Osnovno** v skupini Število.

## 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*)

<i>DSS</i>		Razlaga izračunov
Mean	7,1	aritmetična sredina
Standard Error	1,5	standardna napaka povprečja
Median	6	mediana
Mode	7	modus <sup>13</sup>
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

<sup>13</sup> Glej navodilo pri nalogi SPLET 2.

### Domača naloga 3.1: Število dni s padavinami

Za meteorološko postajo Ljubljana Bežigrad imamo podatke o številu dni s padavinami v obdobju 1951-2008 v datoteki PADAVINE.xlsx.

PADAVINE 1

Izračunajte opisne statistike za število dni s padavinami v Ljubljani v obdobju 1951-2008. ♦

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 1951-2008. 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 1979-2008. Primerjajte rezultate. ♦

### Zgled 3.2: Maraton

Datoteka LJUBLJANSKI MARATON.xlsx vsebuje podatke za 403 moške udeležence, ki so bili uvrščeni na Ljubljanskem maratonu 2005. Uvrščeni so bili samo tekači, ki so na cilj prišli v štirih urah ali prej.

MARATON 1

Za vsakega maratonca na podlagi podatka o letu rojstva ( $LR$ ) izračunajte njegovo **starost** 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 za starost obstajajo osamelci? Izračunajte mejo za spodnje/zgornje osamelce. Ali lahko na podlagi osamelcev odkrijemo napako v podatkih? Obrazložite svoj odgovor.

#### MARATON 2

Vrednosti spremenljivke *starost* kopirajte v prazen stolpec in ga poimenujte *starost\_1*. V tem stolpcu izbrišite vrednost pri tekmovalcu, katerega izračunana starost je 6 let (osamelec). Za spremenljivko *starost\_1* ponovite izračune iz naloge MARATON 1. Primerjajte rezultate. ♦

#### MARATON 3

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

**Navodilo:** Za izračun spremenljivke *čas teka* uporabimo funkciji, ki ju najdemo v skupini funkcij *Datum in čas*, HOUR in MINUTE: HOUR(H2)\*60+MINUTE(H2). V primeru, da se rezultat formule izpiše v obliki 00:00:00, obliko celice spremenimo: kartica *Osnovno*, skupina Število; obliko *Po meri* zamenjamo z obliko *Splošno*.

#### MARATON 4

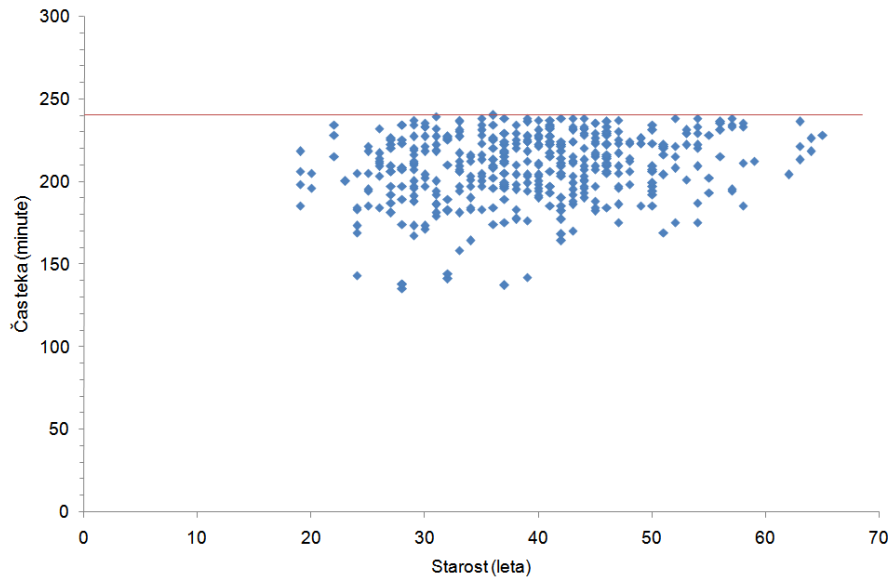
Primerjajte variabilnost *starosti* in *časa teka*.

**Navodilo:** Za vsako spremenljivko posebej izračunajte koeficient variabilnosti. ♦

#### MARATON 5

Grafično prikažite odvisnost časa teka od starosti.

**Navodilo:** izberite vrsto grafikona *Raztreseni*, prva podvrsta (*Raztreseni grafikon samo z oznakami*), na navpični osi je *čas teka* in na vodoravni osi je *starost*. ♦



Slika 3.1: Odvisnost časa teka od starosti tekačev na Ljubljanskem maratonu 2005. Uvrščeni so bili samo tekači, ki so na cilj prišli v štirih urah ali prej. To mejo prikazuje vzporednica vodoravne osi.

#### MARATON 6

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?

**Rešitev:** izračunati je potrebno število kombinacij brez ponavljanja:

$$C_N^n = \binom{N}{n},$$

$N$  je velikost populacije in  $n$  je velikost vzorca. V Excelu to izračuna funkcija **COMBIN(number; number\_chosen)**, ki jo najdemo v zvrsti *Matematika in trigonometrija*. Argument »number« je velikost populacije  $N$ , »number\_chosen« pa velikost vzorca  $n$ .

$$=\text{COMBIN}(403;15) = 7,056 \cdot 10^{26} \blacklozenge$$

#### MARATON 7

Na izbranem vzorcu tekačev na delovnem listu »vzorec 1« ponovite analizo *starosti* in *časa teka*.  $\blacklozenge$

## 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  $|ocena - prava\ vrednost|$  in relativno napako izračunamo z izrazom  $|ocena - prava\ vrednost| / prava\ vrednost$ . Relativno napako ponavadi izražamo v %. ♦

## MARATON 9

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

## MARATON 10

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

## MARATON 11

Analizo v nalogah MARATON 6 do 9 ponovite še na podatkih na delovnem listu »vzorec 2«. Primerjajte rezultate. ♦

### 3.4 Frekvenčna porazdelitev in kvantili

#### Zgled 3.3: Premer debla dreves

V tabeli 3.3 in v datoteki PREMIER DEBLA.xlsx so podatki o številu dreves v posameznem razredu glede na premer debla.

## PREMIER DEBLA 1

Razrede v tabeli 3.3 dopolnite z njihovimi karakteristikami (spodnja ( $x_{imin}$ ), zgornja meja ( $x_{imax}$ ), sredina ( $x_i$ ) in širina ( $d_i$ )).

Za vsak razred izračunajte relativno frekvenco ( $f_i\%$ ), kumulativo frekvenc ( $F_i$ ) in kumulativo 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 premer debla

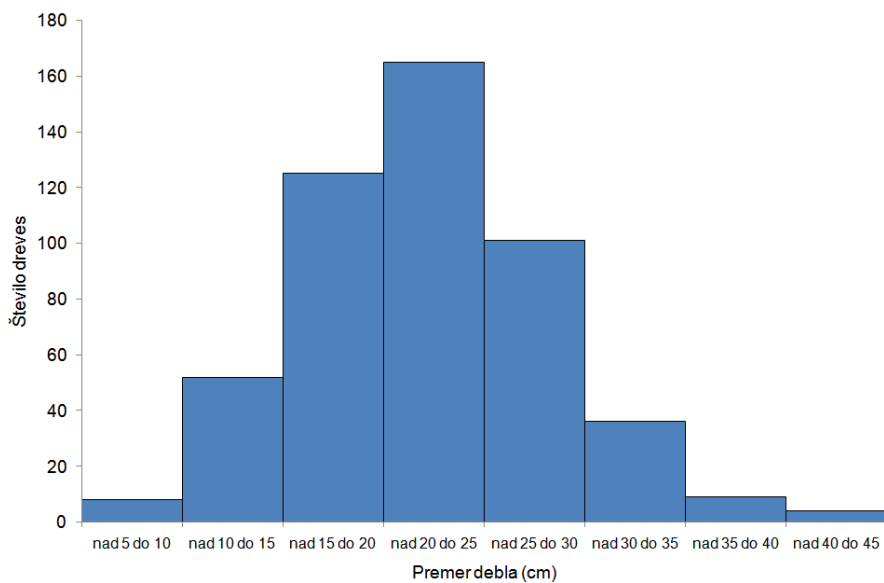
Razred	Premer debla (cm)	Število dreves	Spodnja meja	Zgornja meja	Sredina razreda	Širina razreda	Relativna frekvenca	Kumulativna frekvenc	Kumulativna rel. frekvenc
$i$		$f_i$	$x_{\min}$	$x_{\max}$	$x_i$	$d_i$	$f_i\%$	$F_i$	$F_i\%$
1	nad 5 do 10	8							
2	nad 10 do 15	52							
3	nad 15 do 20	125							
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							

## PREMER DEBLA 2

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

## PREMER DEBLA 3

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



Slika 3.2: Histogram za premer debla za 500 dreves

## PREMER DEBLA 4

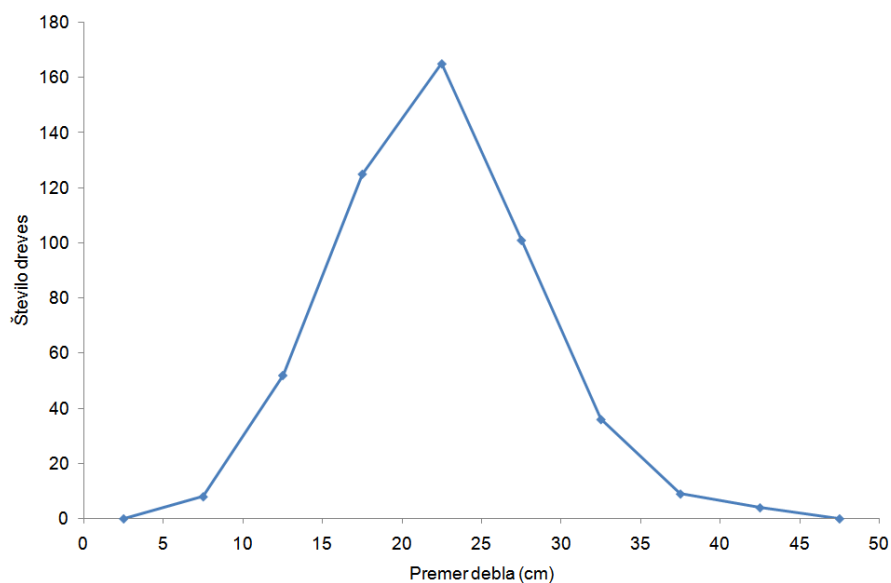
Grafično prikažite podatke iz tabele 3.3 s poligonom (slika 3.3).

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

Razred	Premer debela (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 *Raztreseni*. Na vodoravni osi so sredine razredov, na navpični osi so frekvence.

Preden začnemo delo s čarovnikom za grafikone, označimo samo stolpec s frekvencami (podatki za navpično os) in izberemo *Raztreseni* grafikon. Nato na kartici *Načrt*, v skupini Podatki, z ukazom **Izberi podatke**, za *Niz1* kliknemo na gumb Urejanje in izberemo za *Vrednosti serij X* sredine razredov. Tak postopek je potreben, ker *Raztreseni* grafikon vzame podatke iz levega označenega stolpca za vodoravno os in iz desnega označenega stolpca za navpično os. Vrstni red stolpcev v tabeli 3.3 pa ne ustreza excelovim zahtevam za risanje poligona. ♦



Slika 3.3: Poligon za premer debela za 500 dreves

#### PREMER DEBLA 5


Grafično predstavite kvartile za premer dreves (slika 3.4).

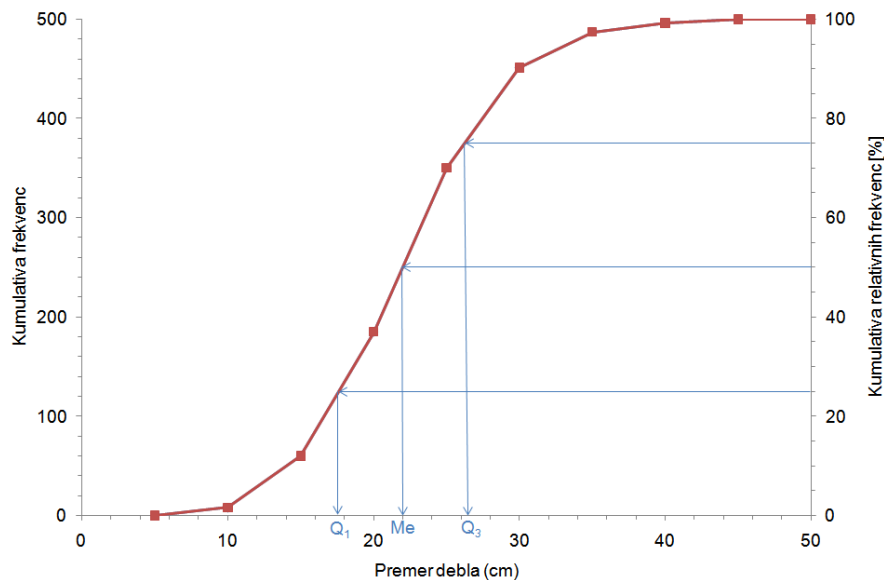
**Navodilo:** Najprej moramo grafično prikazati kumulativno relativnih frekvenc (grafični prikaz se imenuje ogiva). Ogivo narišemo tako, da bo na primarni navpični osi skala za kumulativno frekvenc, na sekundarni navpični osi pa skala za kumulativno relativnih frekvenc (slika 3.4).

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

Nizu  $F_i\%$  določimo skalo na sekundarni navpični osi in ustrezno popravimo merila obeh navpičnih osi (500 dreves predstavlja 100%). S tem dosežemo, da se črti, ki prikazujeta omenjena dva niza podatkov, prekrivata.

Za grafični prikaz kvartilov na grafikon dodamo puščice z ukazom **Oblike**, na kartici **Vstavljanje**, v skupini ilustracije. Za večjo natančnost, pred risanjem puščic, primerno oblikujemo merilo na oseh (slika 3.4).

Oznake kvartilov dodamo na grafikon v *Polje z besedilom*. Klik na ikono *Polje z besedilom*  in poteg z miško na področju grafikona, nanj postavi prazno polje, v katerega napišemo oznako posameznega kvartila. ♦



Slika 3.4: Kumulativna frekvenc ( $F$ ) in kumulativna relativnih frekvenc ( $F\%$ ) za premer debela za podatke v tabeli 3.3. Grafični prikaz kvartilov ( $Q_1$ ,  $Q_2 = Me$ ,  $Q_3$ )

## PREMER DEBLA 6

Izračunajte približke za vse tri kvartile premera debla. Vsebinsko obrazložite dobljene vrednosti.

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

$Q_i$  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 ( $R = P \cdot (n - 1) + 1$ ;  $P$  je relativni rang),  $F(x_{0\min})$  je kumulativna frekvenc pri  $x_{0\min}$  (razreda pred »ničelnim« razredom),  $f_0$  je frekvenca in  $d_0$  širina »ničelnega razreda«.

	A	B	C	D	E	F	G	H
15	Kvartil	Rang kvartila	Štev. razreda v katerem je kvartil	Spodnja meja razreda ( $x_{0\min}$ )	Kumulativa pri $x_{0\min}$	Frekvenca razreda $f_0$	Širina razreda $d_0$	Približek kvartila
16	$Q_1$	=0,25*(C13-1)+1	3	15	60	125	5	=D16+(B16-E16)/F16*G16
17	$Q_2=Me$	=0,5*(C13-1)+1	4	20	185	165	5	=D17+(B17-E17)/F17*G17
18	$Q_3$	=0,75*(C13-1)+1	5	25	350	101	5	=D18+(B18-E18)/F18*G18

	A	B	C	D	E	F	G	H
15	Kvartil	Rang kvartila	Štev. razreda v katerem je kvartil	Spodnja meja razreda ( $x_{0\min}$ )	Kumulativa pri $x_{0\min}$	Frekvenca razreda $f_0$	Širina razreda $d_0$	Približek kvartila
16	$Q_1$	125,75	3	15	60	125	5	17,6
17	$Q_2=Me$	250,50	4	20	185	165	5	22,0
18	$Q_3$	375,25	5	25	350	101	5	26,3

◆

## PREMER DEBLA 7

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

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	Premer debla (cm)	Število dreves	Sredina razreda	Pomožni izračuni	
2	<i>i</i>		$f_i$	$x_i$	$f_i \cdot x_i$	$f_i \cdot (x_i - \bar{x}_p)^2$
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	Premer debla (cm)	Število dreves	Sredina razreda	Pomožni izračuni	
2	<i>i</i>		$f_i$	$x_i$	$f_i \cdot x_i$	$f_i \cdot (x_i - \bar{x}_p)^2$
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	2187,5	2679,61
6	4	nad 20 do 25	165	22,5	3712,5	22,59
7	5	nad 25 do 30	101	27,5	2777,5	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.xlsx.

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 frekvenca	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?

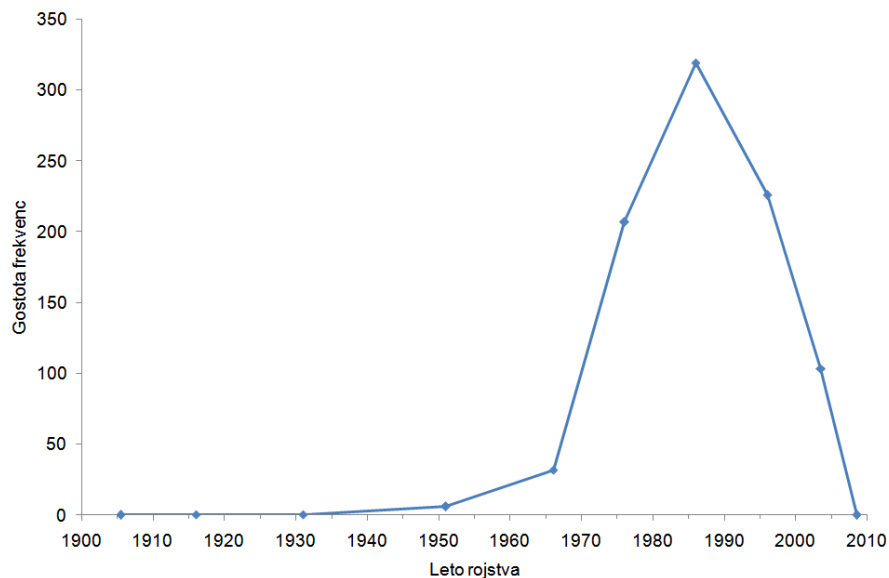
Za podatke v tabeli 3.5 izračunajte gostote frekvenc ( $g_i$ ), relativne frekvence ( $f_i\%$ ), kumulativno frekvenc ( $F_i$ ) in kumulativno relativnih frekvenc ( $F_i\%$ ). Izračune oblikujte v preglednico, kot kaže tabela 3.5. Obrazložite vrednosti za razred št. 8. ♦

**KATJA 2**

Podatke iz tabele 3.5 prikažite s poligonom (slika 3.5).<sup>14</sup>

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

<sup>14</sup> Za grafično predstavitev bomo uporabili poligon, ker Excel ne zna prikazati histograma z različno širokimi razredi.

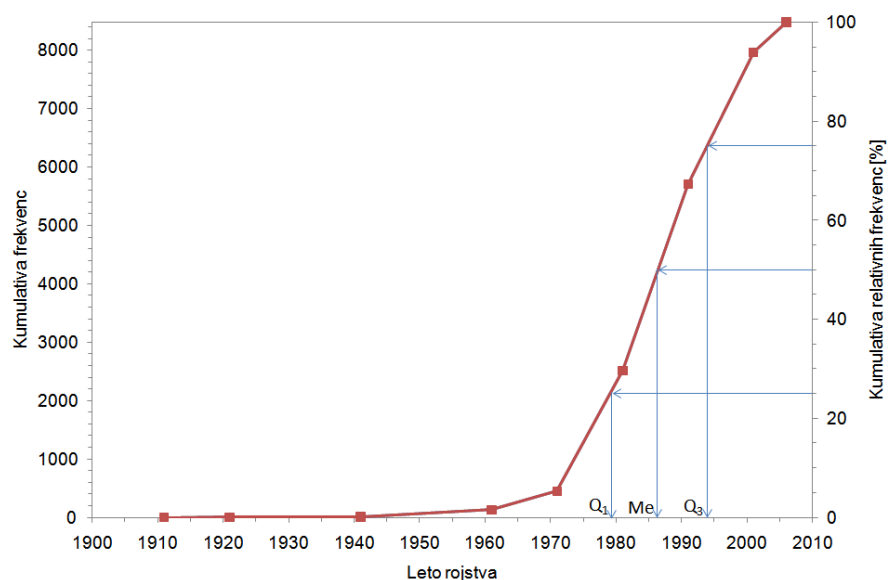


Slika 3.5: Poligon za leto rojstva žensk z imenom Katja, podatki so v tabeli 4.5.

KATJA 3

Grafično predstavite kvartile za rojstno leto oseb z imenom Katja (slika 3.6). ♦

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



Slika 3.6: 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.xlsx (opomba: »rojstno ime« naj bo 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

Grafično prikažite deseti in devetdeseti centil starosti oseb z izbranim rojstnim imenom v zadnjem letu, ki je še upoštevano v podatkih. ♦

#### ROJSTNO IME 5

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

#### ROJSTNO IME 6

Analiziramo starost oseb s tem imenom v zadnjem letu, ki je še upoštevano v podatkih. Izračunajte povprečno starost oseb. Primerjajte jo s povprečnim letom rojstva. ♦

#### ROJSTNO IME 7

Izračunajte standardni odklon za starost oseb s tem imenom v zadnjem letu, ki je še upoštevano v podatkih. Primerjajte ga s standardnim odklonom za rojstno leto. ♦

## 4 Podatkovna zbirka in vrtilne tabele

### Novi pojmi:

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

#### Datoteke s podatki:

ANKETA.xlsx

METEOROLOSKI PODATKI.xlsx

#### Ustvarjene datoteke:

ANKETA\_ANALIZA.xlsx

MET\_POD\_IZBOR.xlsx

### 4.1 Podatkovna zbirka

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

V excelovi datoteki vrstica na delovnem listu predstavlja enoto, stolpec pa spremenljivko. Podatkovna zbirka lahko vsebuje največ toliko podatkov, kolikor je celic na enem delovnem listu, to je  $1048576 * 16384$  (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.

Enote v podatkovni zbirki urejamo in analiziramo z orodji na kartici **Podatki**.

	A	B	C	D	E	F	G	H
	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)
1	AG	1982	ž	2	tehnična šola - kemijski tehnik	ne		ne
2	AG	1983	ž	2	gimnazija	da	17	ne
3	AG	1983	m		gimnazija	da	16	ne
4	BT	1984	ž	2	splošna gimnazija ptuj	da		ne
5	MB	1984	m	2	Gimnazija Želimlje	da		ne
6	BT	1984	m	3	Šolski center Ljubljana - Tehniška g	da		21
7								

## ANKETA 1

Odprite datoteko ANKETA.xlsx 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, v kateri bomo v podatkovno zbirko dodali tri nove spremenljivke, zato datoteko shranite z novim imenom ANKETA\_ANALIZA.xlsx. ♦

V oblikovano podatkovno zbirko lahko poljubno dodajamo spremenljivke, tako da vstavimo prazen stolpec in napišemo ime spremenljivke v prvo vrstico. Podatke običajno izračunamo iz vrednosti drugih spremenljivk. Spremenljivke lahko tudi prestavljamo in brišemo. V vsaki podatkovni zbirki pride prav, če prva spremenljivka predstavlja zaporedne številke enot.

## ANKETA 2

Na začetek podatkovne zbirke vrnite spremenljivko *Zap. štev.* z zaporednimi številkami enot<sup>15</sup>. ♦

## ANKETA 3

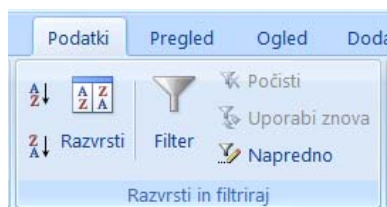
Za spremenljivko *Leto rojstva* vrnite še eno spremenljivko z imenom *Starost*. Za vsakega anketiranca izračunajte starost v dopoljenih letih v letu 2007. ♦





<sup>15</sup> V stolpec lahko zaporedne številke najhitreje napišemo tako, da napišemo prvi dve vrednosti (1 in 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.

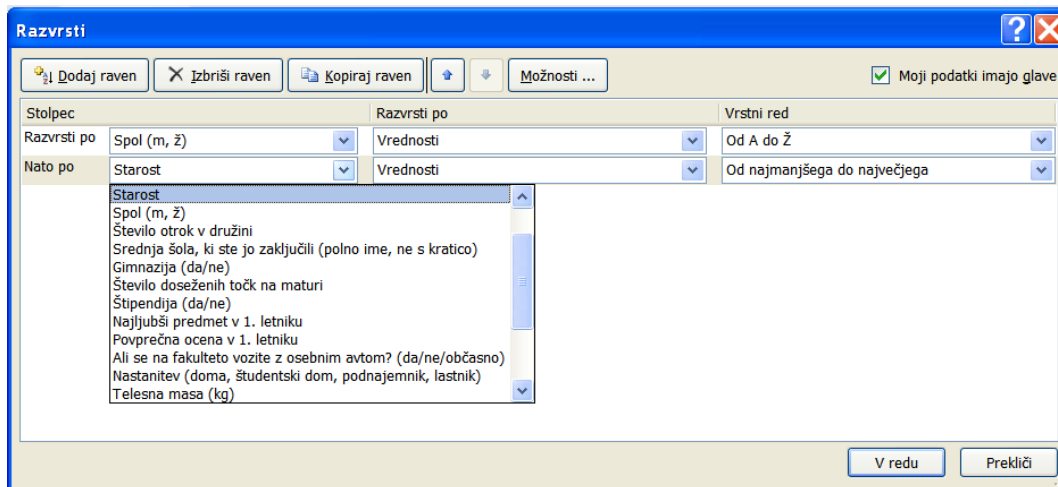
## 4.2 Razvrščanje enot

V podatkovni zbirki lahko enote razvrstimo v različnem vrstnem redu glede na vrednosti posameznih spremenljivk. Če je spremenljivka opisna (besedilna), Excel razvrsti enote glede na vrednosti te spremenljivke po abecednem vrstnem redu od A do Ž ali pa od Ž do A. Glede na vrednosti številskih spremenljivk Excel enote razvrsti v naraščajočem ali padajočem vrstnem redu. Datumi tu sodijo med številke spremenljivke.

Za razvrščanje enot uporabljamo ukaze v skupini Razvrsti in filtriraj na kartici **Podatki**.



Če želimo enote razvrstiti glede na vrednosti ene spremenljivke, označimo poljubno celico v pripadajočem stolpcu in kliknemo eno od ikon  (naraščajoč vrstni red) ali  (padajoč vrstni red). Če želimo enote razvrstiti glede na vrednosti več spremenljivk, uporabimo ukaz **Razvrsti** . V pogovornem oknu **Razvrsti** v polju *Razvrsti po* izberemo spremenljivko podatkovne zbirke, ki določa primarni vrstni red enot. Odpremo seznam imen in kliknemo na želeno ime. Poleg imena moramo izbrati tudi vrednost v poljih **Razvrsti po** (*Vrednosti, Barva celice, Barva pisave, Ikona celice*) in **Vrstni red** (*Od najmanjšega do največjega, od največjega do najmanjšega, Seznam po meri*). Sekundarni vrstni red enot določimo s klikom na gumb  in izbiro spremenljivke v polju **Zatem po**. Dodamo lahko poljubno število ravni razvrščanja enot.



Ukaz **Razvrsti** spremeni vrstni red enot 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 enot.

ANKETA 4

Uredite anketirance v podatkovni zbirki po **Starosti** od najmlajšega do najstarejšega.

**Opozorilo:** preden uporabimo katerikoli ukaz na kartici **Podatki** z namenom urejanja ali analize podatkov, v podatkovni zbirki označimo **eno neprazno celico**. Ne smemo označiti posamezne skupine celic ali posameznih stolpcev podatkovne zbirke. ♦

ANKETA 5

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

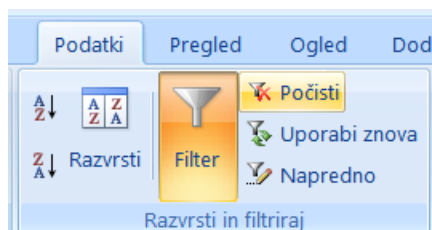
ANKETA 6

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

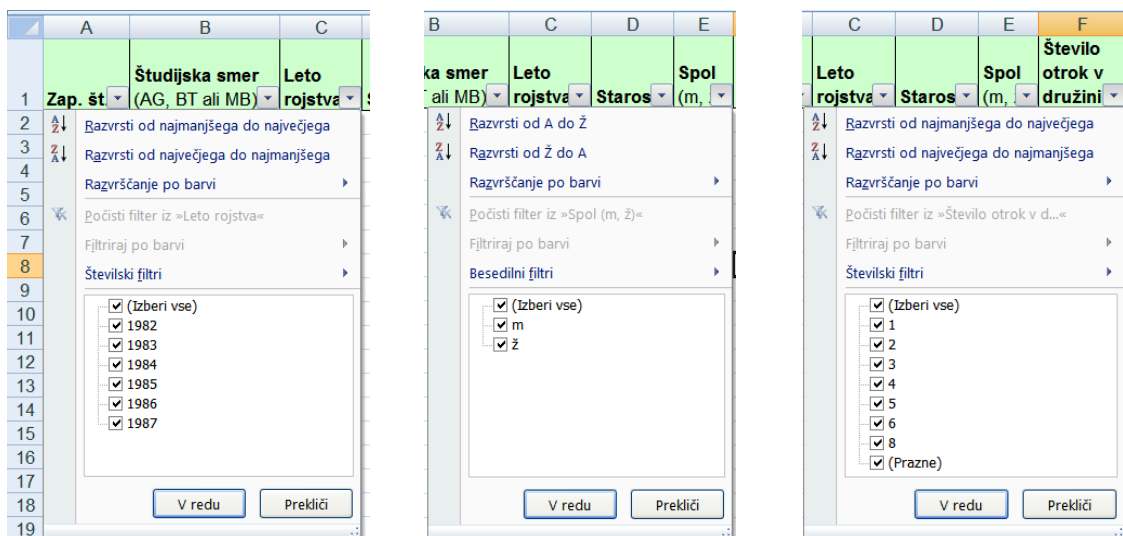
## 4.3 Izbira zapisov

### 4.3.1 Filter

V podatkovni zbirki lahko izbiramo in iščemo posamezne zapise po različnih kriterijih. Uporabimo ukaz **Filter** na kartici **Podatki** v skupini Razvrsti in filtriraj. Bolj zapletene (kompleksne) kriterije določimo z ukazom **Napredno**.



Ko uporabimo ukaz **Filter**, se v naslovni vrstici poleg imen spremenljivk pojavijo gumbi s puščicami. Klik na puščico odpre podmeni, v katerem so na vrhu navedeni ukazi za razvrščanje, ki smo jih opisali v prejšnjem poglavju. Sledijo ukazi za izbiro enot: **Počisti filter iz »ime spremenljivke«** (uporaba možna samo, če je bil pred tem filter že uporabljen); **Filtriraj po barvi** (uporaba možna, če so celice izbrane spremenljivke različno obarvane); z ukazom **Številski filtri/Besedilni filtri** določimo kriterij izbora. V spodnjem delu menija so navedene vse obstoječe vrednosti izbrane spremenljivke in še vrednost *Prazne* v primeru, če izbrana spremenljivka nima določene vrednosti za vse enote v podatkovni zbirki (manjkajoči podatki - prazne celice).



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

Načini izbire:

- s klikom odstranimo kljukice v kvadratih pred vrednostmi, katerih izpisa ne želimo;

- s klikom odstranimo kljukico v kvadratu pred možnostjo (*Izberi vse*), nato odkljukamo samo tiste vrednosti, ki ustrezajo kriteriju;

Zap. št.	Študijska smer (AG, BT ali MB)	Leto rojstva
34	33 MB	1986
35	34 MB	1986
36	35 BT	1986
37	36 BT	1986
38	37 MB	1986
39	38 MB	1986
75	74 AG	1986
76	75 AG	1986
77	76 AG	1986
78	77 AG	1986
79	78 AG	1986
80	79 AG	1986

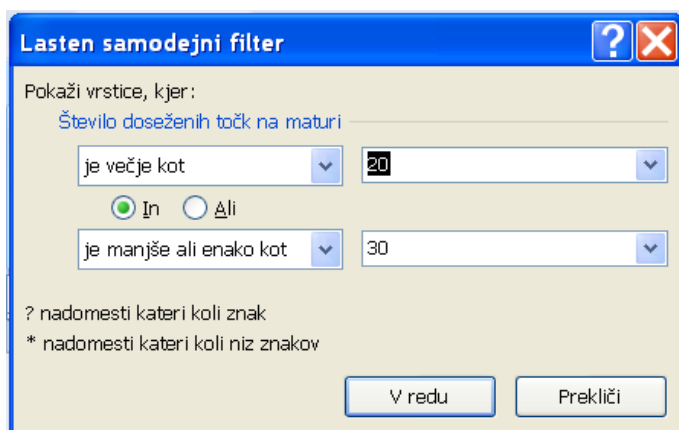
čili	Gimnazija (da/ne)	Število doseženih točk na maturi
da		24
utome da		25
da		29
da		26
da		24
da		23

- ukaz **Besedilni filtri** oz. **Številski filtri** uporabimo, kadar določimo kompleksnejši kriterij. Odpre se seznam možnosti za določitev kriterija. Vsebina seznama je odvisna od tega, ali je spremenljivka številska ali opisna;

izjaja	Število doseženih točk na maturi	Štipendija (da/ne)	Najljubši predmet v 1. letniku
ne		12	botanika
da		15	botanika
			matematika
			biologija
			fizikalna kemija
			uvod v biologijo

S klikom katerekoli možnosti s seznama se odpre pogovorno okno **Lasten samodejni filter**, v katerem lahko kriterij sestavimo iz enega ali dveh pogojev. V prvem polju izberemo logični operator (*Je enako, Ni enako, Večje kot..., Začne se z, Konča se z, Vsebuje, Zgornjih 10, Nadpovprečno, Podpovprečno,...*), v sosednje polje pa napišemo vrednost ali jo izberemo s seznama. Če določamo kriterij za številsko 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 (*Izberi vse*) uporabimo, če želimo preklicati predhodno uporabljeni kriterij za izbrano spremenljivko. Enak rezultat dobimo tudi, če uporabimo ukaz **Počisti filter iz »ime spremenljivke«**. Pri tem pa ni nujno, da se izpišejo vse enote v podatkovni zbirki, ker imamo lahko hkrati kriterij določen tudi na podlagi vrednosti katere druge spremenljivke. Če želimo preklicati vse kriterije hkrati, uporabimo ukaz **Počisti** na kartici **Podatki** v skupini Razvrsti in filtriraj;

Samodejni filter izklopimo s ponovnim ukazom **Filter** na kartici **Podatki** v skupini Razvrsti in filtriraj (ob tem puščice v prvi vrstici podatkovne zbirke izginejo).

#### ANKETA 7

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



ANKETA 8

Poiščite 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 enote kot novo podatkovno zbirko, izbor enot kopiramo na nov delovni list.

ANKETA 9

Poiščite anketirane, ki izhajajo iz družin z več kot dvema otrokoma. ♦

ANKETA 10

Poiščite anketirane, ki so v zadnjem letu prebrali nadpovprečno število leposlovnih knjig. Izračunajte povprečje števila prebranih leposlovnih knjig za vse anketirane.

**Navodilo:** za izbor uporabite ukaz **Številski filtri/Nadpovprečno**. Izračun povprečnega števila prebranih leposlovnih knjig naredite na delovnem listu »Podatki«, v celico, ki se ne dotika celic podatkovne zbirke (na primer v celico T129, vrstica 128 mora biti prazna). ♦

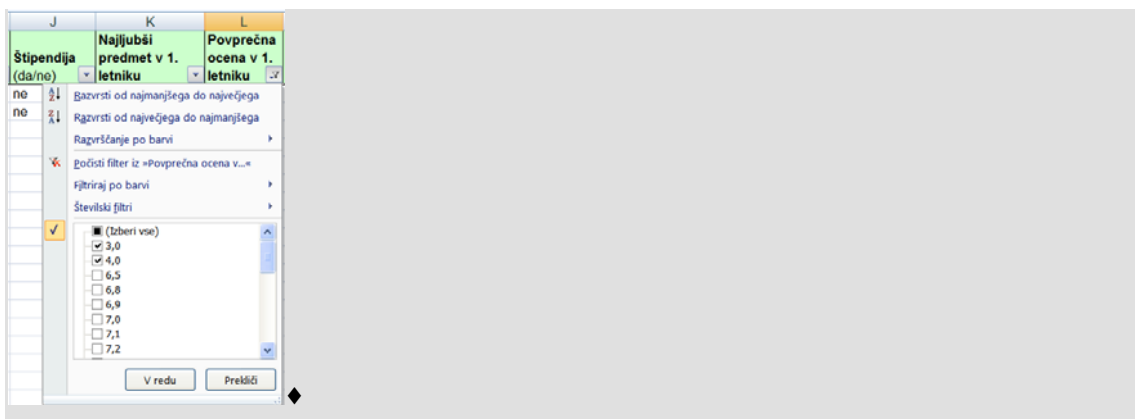
ANKETA 11

Poiščite anketirance, ki so nastanjeni v študentskem domu in se na fakulteto vozijo z avtomobilom. Izbrane enote kopirajte na nov list in ga poimenujte »Študentski dom\_avto«. ♦

ANKETA 12

S pomočjo filtra si oglejte vrednosti spremenljivke *Povprečna ocena v 1. letniku*. Izberite enote, pri katerih ima ta spremenljivka nesprejemljivo vrednost. Zakaj so te vrednosti za študente, ki so vpisani v 2. letnik, nesprejemljive? Koliko jih je?

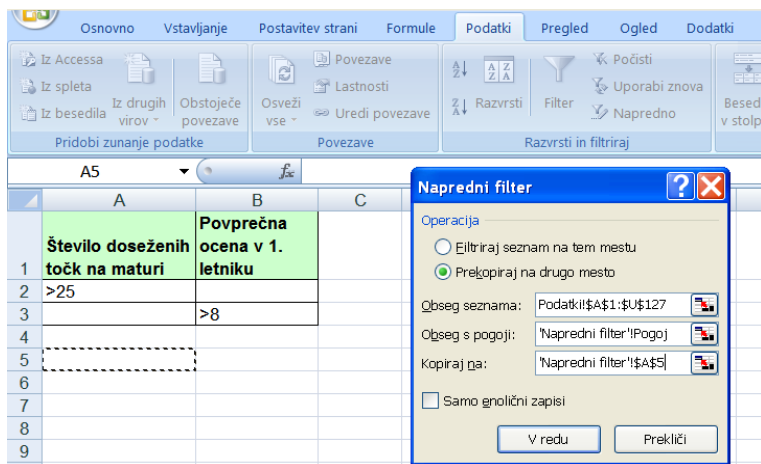
Napačne vrednosti v podatkovni zbirki pobrišite. Analiza podatkov v nadaljevanju bo potekala, kot da izbrani študentje na to vprašanje niso odgovorili.



### 4.3.2 Napredni filter

Ukaz **Filter** na kartici **Podatki** v skupini Razvrsti in filtriraj s podukazom **Številski filtri** oz. **Besedilni filtri** omogoča, da za eno spremenljivko uporabimo največ dva pogoja. Med izbranimi zapisi pa lahko iščemo še podskupine zapisov, ki ustrezajo kriterijem, določenim na osnovi več spremenljivk v podatkovni zbirki.

Kadar želimo uporabiti zahtevnejše sestavljene kriterije, za več spremenljivk hkrati, uporabimo ukaz **Napredno** na kartici **Podatki** v skupini Razvrsti in filtriraj. Najprej moramo pripraviti **obseg celic s pogoji**. Obseg celic s pogoji je sestavljen iz najmanj dveh vrstic. V prvi vrstici so imena spremenljivk (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 enot z enakimi vrednostmi spremenljivk, ki določajo kriterij, možnost **Samo enolični zapisi** določi izbiro samo ene izmed njih.

#### ANKETA 13

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. Enote, ki ustrezajo danemu pogoju, izpišite na delovni list »Napredni filter« v vrstice pod obsegom pogojev (glej sliko zgoraj). Koliko je takih enot? ♦

## 4.4 Vrtilne tabele

Vrtilno tabelo v Excelu naredimo z ukazom **Vrtilna tabela** na kartici **Vstavljanje** v skupini Tabele.



Vrtilno tabelo uporabimo, če želimo enote v podatkovni zbirki razdeliti v skupine glede na vrednosti posameznih 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 sliki od 4.1.

The image shows three screenshots of Excel pivot tables. The first screenshot shows a pivot table with columns A and B, and rows 1 to 7. The data is as follows:

	A	B
1		
2		
3	Oznake vrstic	Štej od Zap. št.
4	m	41
5	ž	85
6	Skupna vsota	126
7		

The second screenshot shows a pivot table with columns A, B, C, and D, and rows 1 to 9. The data is as follows:

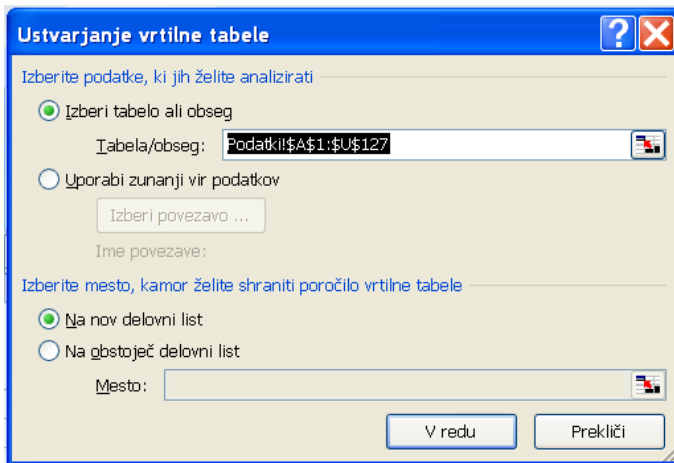
	A	B	C	D
1				
2				
3	Povprečje od Povprečna ocena v 1. letniku	Oznake stolpcev		
4	Oznake vrstic	m	ž	Skupna vsota
5	AG	7,7	7,8	7,7
6	BT	8,0	7,8	7,9
7	MB	7,9	7,5	7,6
8	Skupna vsota	7,8	7,7	7,7
9				

The third screenshot shows a pivot table with columns A, B, C, and D, and rows 1 to 8. The data is as follows:

	A	B	C	D
1				
2				
3		Oznake stolpcev		
4	Vrednosti	m	ž	Skupna vsota
5	Štej od Zap. št.	41	85	126
6	Povprečje od Povprečna ocena v 1. letniku	7,8	7,7	7,7
7	Povprečje od Število doseženih točk na maturi	19,4	21,4	20,8
8				

Slika 4.1: Število moških in žensk, sodelujočih v anketi (zgoraj) - **enosmerna vrtilna tabela**; povprečje *Povprečne ocene v 1. letniku* po spolu in po smeri študija (na sredi) – **dvosmerna vrtilna tabela**; število anketiranih, povprečje *Povprečne ocene v 1. letniku* in povprečje *Števila doseženih točk na maturi* po *spolu* - **enosmerna vrtilna tabela s tremi izračunanimi Vrednostmi - statistikami** (spodaj). Za izraz “Skupna vsota” v zadnji vrstici oz. zadnjem stolpcu vrtilnih tabel glej pojasnilo v sprotni opombi na strani 108.

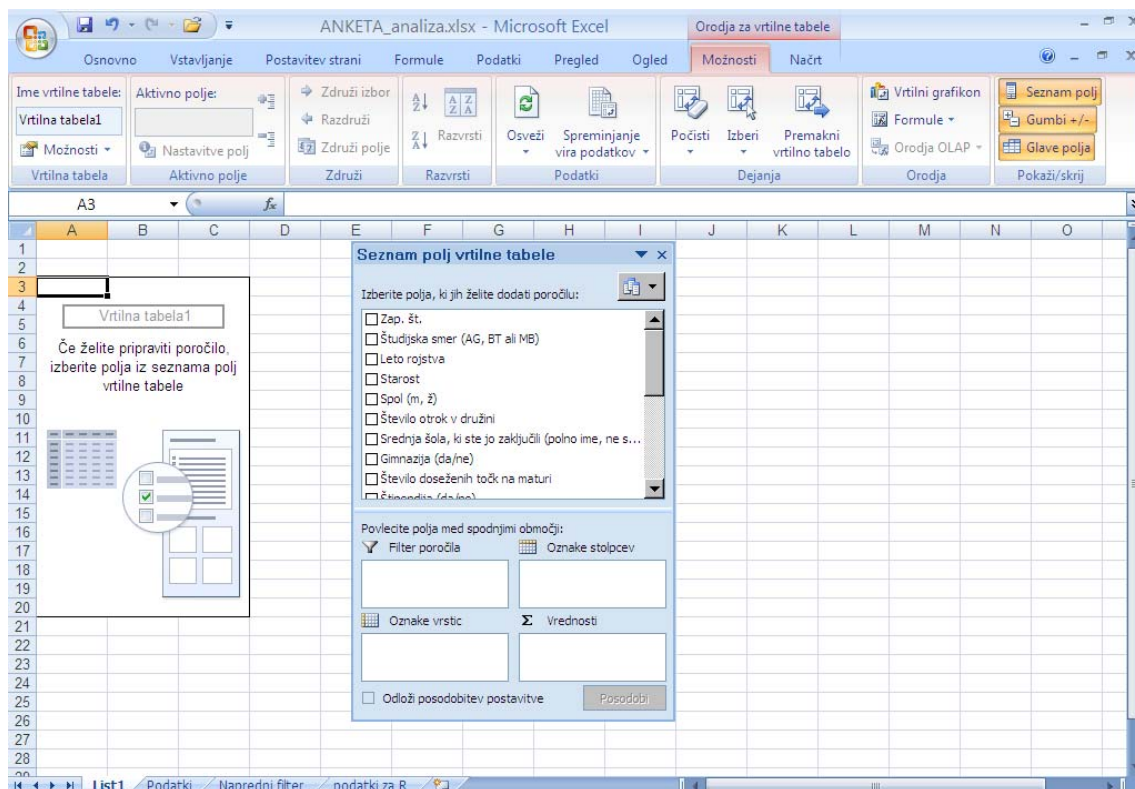
Po kliku ukaza **Vrtilna tabela** na kartici **Vstavljanje** v skupini Tabele se odpre pogovorno okno **Ustvarjanje vrtilne tabele**. Tu določimo obseg podatkov, na podlagi katerih želimo oblikovati vrtilno tabelo, in mesto v datoteki, kamor se vrtilna tabela izpiše. V primerih te knjige bomo imeli podatke vedno urejene v Excelovi datoteki, zato bomo vedno izbrali že vnaprej ponujeno možnost **Izberi tabelo ali obseg**.



V polju **Tabela/Obseg** določimo obseg celic podatkovne zbirke. Če imamo pred uporabo ukaza **Vrtilna tabela** označeno eno celico v podatkovni zbirki, Excel prepozna podatkovno zbirko in to polje izpolni sam. Če imamo pred uporabo ukaza **Vrtilna tabela** označeno skupino celic v podatkovni zbirki, bo Excel upošteval podatke v označeni skupini celic. Preden nadaljujemo z oblikovanjem vrtilne tabele, se moramo prepričati, da je v polju **Tabela/Obseg** izbran ustrezen obseg celic.

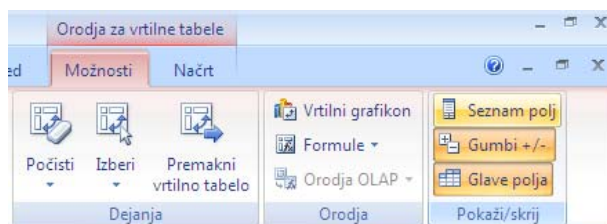
Vrtilna tabela se izpiše na nov list, če izberemo možnost *Na nov delovni list* (to možnost bomo uporabljali) oz. na obstoječi delovni list, če izberemo možnost *Na obstoječ delovni list*.

Po kliku gumba **V redu**, se v datoteki pojavi nov delovni list s shemo prazne vrtilne tabele in pogovornim oknom **Seznam polj vrtilne tabele**. Poleg tega se na opravljenem traku pojavi **Orodje za vrtilne tabele** s karticama *Možnosti* in *Načrt*.



#### 4.4.1 Seznam polj vrtilne tabele

Okno **Seznam polj vrtilne tabele** je na zaslonu vidno ob vrtilni tabeli, če je le-ta aktivna (imamo označeno poljubno celico znotraj vrtilne tabele). Če tega okna, kljub aktivni vrtilni tabeli ne vidimo, se pokaže s klikom ukaza **Seznam polj** na kartici **Možnosti** v zadnji skupini Pokaži/Skrij.



V oknu **Seznam polj vrtilne tabele** je v zgornjem delu seznam spremenljivk (polj), v spodnjem delu pa so območja vrtilne tabele, v katera z miško potegnemo spremenljivke, na podlagi katerih želimo oblikovati tabelo. Vrtilna tabela ima štiri območja **Filter sporočila**, **Oznake stolpcev**, **Oznake vrstic** in **Vrednosti**. V območja **Filter sporočila**, **Oznake stolpcev**, **Oznake vrstic** postavimo spremenljivke, po katerih želimo enote v

podatkovni zbirki razdeliti v skupine. V območje **Oznake vrstic** postavimo spremenljivke, katerih vrednosti določajo vrstice tabele (npr. če izberemo **Spol**, bo imela vrtilna tabela dve osnovni vrstici - eno vrstico za ženske in eno za moške). V območje **Oznake stolpcev** postavimo spremenljivke, katerih vrednosti določajo stolpce tabele (npr. če izberemo **Nastanitev**, bo imela vrtilna tabela štiri osnovne stolpce). Ko ime spremenljivke s seznama postavimo v posamezno območje, se v območju pojavi gumb z imenom te spremenljivke. Za vsako skupino enot iz podatkovne zbirke Excel izračuna različne funkcije na podlagi spremenljivk, ki jih postavimo v območje **Vrednosti**.

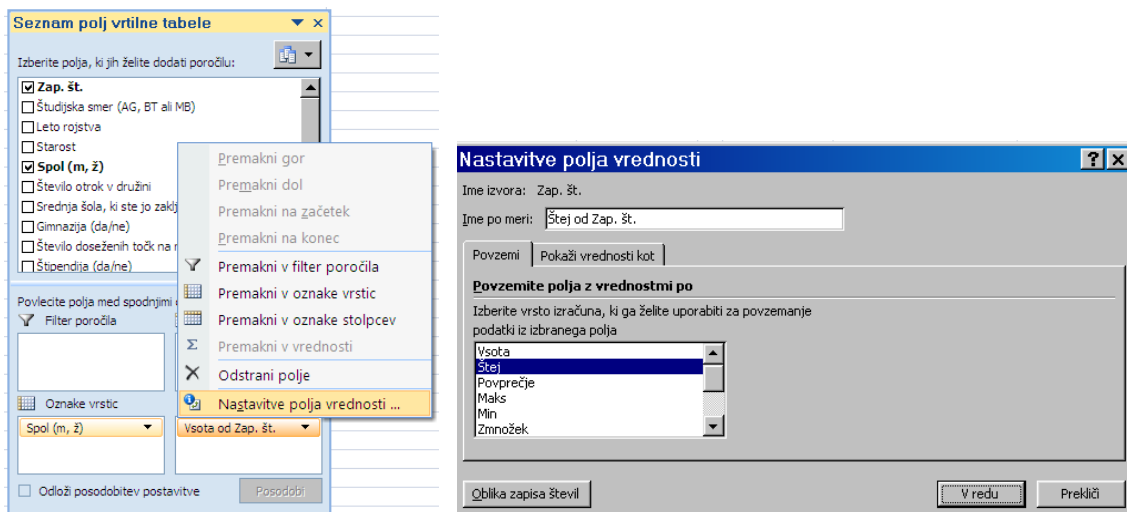
Ko številsko spremenljivko prvič postavimo v območje **Vrednosti**, Excel samodejno izbere funkcijo **Vsota (SUM)**, s katero sešteje vse vrednosti izbrane spremenljivke za enote v posameznih skupinah. Če je spremenljivka opisna, Excel samodejno izbere funkcijo **Štej (COUNTA)**, s katero prešteje, koliko enot je v posamezni skupini ima vrednost za izbrano spremenljivko (prešeteje neprazne celice).

Ime gumba v območju **Vrednosti** je sestavljeno iz imena funkcije in imena spremenljivke. Poleg funkcij **Vsota** in **Štej** v vrtilni tabeli lahko izbiramo še med funkcijami **Povprečje (AVERAGE)**, **Maks (MAX)**, **Min (MIN)**, **Zmnožek (PRODUCT)**, **Štetje števil (COUNT)**, **StdOdk (STDEV)**, **StdOdkp (STDEVP)**, **Varianca (VAR)**, **Variancap (VARP)**.<sup>16</sup>

Funkcijo izberemo v pogovornem oknu **Nastavitve polja vrednosti**. To okno se odpre s klikom istoimenskega ukaza na seznamu, ki se odpre s klikom na puščico na gumbu v območju **Vrednosti**.

---

<sup>16</sup> V orodju za vrtilne tabele so te funkcije prevedene v slovenski jezik, v oklepaju navajamo imena funkcij v ukazu **Vstavi funkcijo**.



V oknu **Nastavitve polja vrednosti** na seznamu kartice **Povzemi** izberemo funkcijo. V polju **Ime po meri** je ime, ki se v shemi okna **Seznam polj vrtilne tabele**, pojavi kot ime gumba v območju **Vrednosti**. Excel ga samodejno določi glede na izbrano funkcijo in glede na ime spremenljivke: na primer **Štej od Zap. št.** ali **Povprečje od Povprečna ocena v 1. letniku** v vrtilnih tabelah na sliki 4.1. Če imamo enote razdeljene v skupine samo na podlagi vrednosti ene spremenljivke, se to ime izpiše kot naslov vrstice oz. stolpca, v katerem so izračuni, sicer se izpiše v zgornjo levo celico vrtilne tabele (slika 4.1).

Že narejene vrtilne tabele lahko spreminjamo s premikanjem (vlečenjem z miško) gumbov spremenljivk v shemi območij vrtilne tabele v pogovornem oknu **Seznam polj vrtilne tabele**. Če želimo spremeniti način razvrščanja enot podatkovne zbirke v skupine, spremenimo vsebino območij **Oznake stolpcev**, **Oznake vrstic** in **Filter poročila**. Če želimo, da se za skupine izračunajo druge funkcije na osnovi drugih ali pa istih spremenljivk, spremenimo vsebino območja **Vrednosti**. Posamezen gumb iz sheme odstranimo tako, da ga enostavno povlečemo ven iz območja **Vrednosti**. Gumbje lahko s potegom tudi prestavimo v drugo območje.

Z dvoklikom na celico vrtilne tabele, v kateri je izračunana vrednost za posamezno skupino enot, se odpre nov delovni list z delom podatkovne zbirke, ki vsebuje zapise



enot, na podlagi katerih je bila izračunana izbrana vrednost. To aktivnost imenujemo vrtnanje po tabeli.

Izračune v vrtilnih tabelah lahko preuredimo in nadgradimo na veliko načinov, ukaze za to najdemo v **Orodju za vrtilne tabele** na karticah **Možnosti** in **Načrt**.

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, prilepimo z ukazom **Prilepi/Posebno lepljenje**, možnost **Vrednosti** na drug list v delovnem zvezku. Rekli bomo, da s tem vsebino vrtilne tabele »parkiramo«. Če za vsako analizo naredimo novo vrtilno tabelo, postane Excelova datoteka hitro zelo velika in nepregledna.

#### 4.4.2 Primeri oblikovanja vrtilnih tabel

Za oblikovanje vrtilne tabele, ki pove koliko je bilo med anketiranimi moških in koliko žensk (zgornja tabela na sliki 4.1), je postavitve spremenljivk v območja vrtilne tabele takšna:

Oznake vrstic	Štej od Zap. št.
m	41
ž	85
<b>Skupna vsota</b>	<b>126</b>

Seznam polj vrtilne tabele

Izberite polja, ki jih želite dodati poročilu:

- Zap. št.
- Studijska smer (AG, BT ali MB)
- Leto rojstva
- Starost
- 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
- Ali se na fakulteto vozite z osebnim avtom? (da/ne/obč...)
- Nastanitev (dosa... študentki, dosa... podastopnik, lastnik)

Povlecite polja med spodnjimi območji:

Filter poročila: [ ] Oznake stolpcev: [ ]

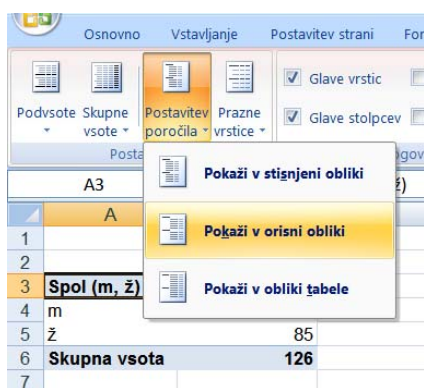
Oznake vrstic: Spol (m, ž) Vrednosti: Štej od Zap. št.

Odloži posodobitev postavitve Posodobi

Za štetje enot v posamezni skupini smo v tem primeru uporabili spremenljivko **Zap. št.** zato, ker za to spremenljivko obstaja vrednost za vsako enoto v podatkovni zbirki (ni

manjkajočih podatkov). Enak rezultat bi dobili, če bi šteli po spremenljivkah, ki nimajo manjkajočih podatkov (na primer *Študijska smer*, *Leto rojstva*, *Starost*, *Spol*,...).

Ob osnovni nastavitvi Excel vrtilno tabelo prikaže v **stisnjeni obliki**, kar pomeni, da v tabeli ne vidimo imen spremenljivk, po katerih so enote razvrščene v skupine. Izgled vrtilne tabele spremenimo z ukazom **Postavitev poročila** na kartici *Načrt* v skupini Postavitev. Če izberemo ukaz **Postavitev poročila/Pokaži v orisni obliki**, se izgled vrtilne tabele spremeni tako, da se namesto izraza *Oznake vrstic* oz. *stolpcev* izpišejo imena spremenljivk, ki določajo vrstice oz. stolpce. V nadaljevanju bomo uporabljali **orisno obliko** prikaza vrtilnih tabel.



Če enote razvrstimo v skupine tako po *Spolu* kot po *Nastanitvi*, je postavitve vrtilne tabele takšna, kot kaže slika 4.2 (zgoraj). V tem primeru Excel v vrtilno tabelo doda kategorijo (*prazen*) za manjkajoče podatke spremenljivke *Nastanitev*.

Štej od Zap. št.	Spol (m, ž)		Skupna vsota
Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž	
doma	20	31	51
lastnik		1	1
podnajemnik	7	33	40
študentski dom	14	18	32
(prazen)		2	2
<b>Skupna vsota</b>	<b>41</b>	<b>85</b>	<b>126</b>

Štej od Nastanitev (doma, študentski dom, podnajemnik, lastnik)	Spol (m, ž)		Skupna vsota
Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž	
doma	20	31	51
lastnik		1	1
podnajemnik	7	33	40
študentski dom	14	18	32
(prazen)		2	2
<b>Skupna vsota</b>	<b>41</b>	<b>83</b>	<b>124</b>

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 vrednost te spremenljivke v tabeli pojavi tudi vrednost »prazen«. Če namesto po *Zap. št.* štejemo po spremenljivki *Nastanitev*, se podatkovna zbirka zmanjša za 2 manjkajoča podatka, pri vrednosti »prazen« pa ostanejo celice tabele prazne (spodaj). Za okno **Seznam polj vrtilne tabele** smo izbrali prikaz brez seznama spremenljivk: *Samo odsek območij (2 krat 2)* (zgoraj).

Če v območji *Oznake stolpcev* oz. *Oznake vrstic* postavimo več spremenljivk, bodo stolpci oz. vrstice najprej razdeljeni po vrednostih prve spremenljivke, nato bo vsaka vrednost prve spremenljivke razdeljena po vrednostih druge spremenljivke itd (slika 4.3).

Štej od Zap. št.	Oznake vrstic	m	ž	Skupna vsota
5	doma	20	31	51
6	da	8	15	23
7	ne	3	7	10
8	občasno	9	9	18
9	lastnik	1	1	1
10	ne	1	1	1
11	podnajemnik	7	33	40
12	da	1	2	3
13	ne	5	24	29
14	občasno	1	7	8
15	študentski dom	14	18	32
16	da	2	2	2
17	ne	12	16	28
18	občasno	2	2	2
19	(prazen)		2	2
20	da		2	2
21	Skupna vsota	41	85	126

**Seznam polj vrtilne tabele**

Povlecite polja med spodnjimi območji:

Filter poročila: Oznake stolpcev: Spol (m, ž)

Oznake vrstic: Nastanitev (doma, ...): Štej od Zap. št.

Ali se na fakulteto ...

Odloži posodobitev postavitev Posodobi

ANKETA\_analiza.xlsx - Microsoft Excel

Orodja za vrtilne tabele

Osnovno Vstavljanje Postavitve strani Formule Podatki Pregled Ogled Možnosti Načrt

Podsvote Skupne vsote Postavitve poročila Prazne vrstice

Glave vrstic Združene vrstice Glave stolpcev Združeni stolpci

Postavitve vrtilne tabele Slogi vrtilne tabele

Pokaži v stisnjeni obliki

Pokaži v orisni obliki

Pokaži v obliki tabele

Štej od Zap. št.	Nastanitev (doma, ...)	Spol (m, ž)	m	ž	Skupna vsota
8	doma		20	31	51
9	da		8	15	23
10	ne		3	7	10
11	občasno		9	9	18
12	lastnik		1	1	1
13	ne		1	1	1
14	podnajemnik		7	33	40
15	da		1	2	3
16	ne		5	24	29
17	občasno		1	7	8
18	študentski dom		14	18	32
19	da		2	2	2
20	ne		12	16	28
21	občasno		2	2	2
22	(prazen)			2	2
23	da			2	2
24	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 - triserna vrtilna tabela. Zgornja vrtilna tabela je prikazana v **stisnjeni obliki**, spodnja pa v **orisni obliki**. Obliko prikaza določimo z ukazom **Postavitve poročila** na kartici **Načrt**.

V območje **Filter poročila** postavimo spremenljivke, katerih vrednosti določajo novo dimenzijo vrtilne tabele (slika 4.4). Za vsako vrednost te spremenljivke se izpiše samostojna vrtilna tabela.

The image shows two screenshots of an Excel spreadsheet. The top screenshot shows a pivot table with the following data:

Študijska smer	Spol (m, ž)	Štež od Zap. št.
AG	m	6
AG	ž	1
BT	m	10
BT	ž	8
MB	m	18
MB	ž	12
<b>Skupna vsota</b>	<b>m</b>	<b>41</b>
<b>Skupna vsota</b>	<b>ž</b>	<b>85</b>
<b>Skupna vsota</b>	<b>Vse</b>	<b>126</b>

The bottom screenshot shows the same pivot table but with the filter 'Najljubši predmet' set to 'BT'. The data is now filtered to show only biotechnology students:

Študijska smer	Spol (m, ž)	Štež od Zap. št.
BT	m	3
BT	ž	8
<b>Skupna vsota</b>	<b>m</b>	<b>3</b>
<b>Skupna vsota</b>	<b>ž</b>	<b>15</b>
<b>Skupna vsota</b>	<b>Vse</b>	<b>25</b>

Slika 4.4: Število anketiranih po spolu (*Oznake stolpcev*) in po izbiri najljubšega predmeta v 1. letniku (*Oznake vrstic*) za vsako študijsko smer (*Filter poročila*) - večplastna dvosmerna vrtilna tabela; pet študentov se ni opredelilo za najljubši predmet v 1. letniku (vrstica (prazen) v zgornji tabeli); če v območju *Filter poročila* izberemo BT, se pokažejo podatki samo za biotehnologe (spodnja tabela).

Tudi v območje *Vrednosti* lahko postavimo več spremenljivk hkrati. V tem primeru se v območju *Oznake vrstic* ali *Oznake stolpcev* pojavi tudi gumb *Vrednosti*. Na primer za spodnjo vrtilno tabelo na sliki 4.1 je postavitve v območjih vrtilne tabele takšna:

	A	B	C	D
1				
2				
3		Spol (m, ž)		
4	Vrednosti	m	ž	Skupna vsota
5	Štej od Zap. št.	41	85	126
6	Povprečje od Povprečna ocena v 1. letniku	7,8	7,7	7,7
7	Povprečje od Število doseženih točk na maturi	19,4	21,4	20,8
8				
9				
10				
11				
12				
13				
14				
15				
16				

Gumb *Vrednosti* v območju *Oznake vrstic* lahko s potegom miške prestavimo v območje *Oznake stolpcev*, spremenljivko *Spol* pa v območje *Oznake vrstic*. S tem vrtilna tabela dobi naslednjo obliko<sup>17</sup>:

	A	B	C	D	E
1					
2					
3		Vrednosti			
4	Spol (m, ž)	Štej od Zap. št.	Povprečje od Povprečna ocena v 1. letniku	Povprečje od Število doseženih točk na maturi	
5	m	41	7,8	19,4	
6	ž	85	7,7	21,4	
7	Skupna vsota	126	7,7	20,8	
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					

#### ANKETA 14

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

Primeri ustrezno oblikovanih parkiranih vrtilnih tabel:

<sup>17</sup> Excel v večsmernih tabelah zadnji stolpec oz. vrstico poimenuje z imenom *Skupna vsota* ne glede na funkcijo v območju *Vrednosti*. Izraz *Skupna vsota* ponavadi spremenimo v izraz *Skupaj*. Spremenjeno ime zadnje vrstice oz. stolpca v vrtilni tabeli ostane, tudi če kasneje spreminjamo vsebino območij *Oznake stolpcev*, *Oznake vrstic*, *Filter poročila* ali *Vrednosti*.

Spol	Število anketiranih
Moški	41
Ženske	85
Skupaj	126

Povprečna ocena v 1. letniku	Spol		
	Moški	Ženske	Skupaj
Študijska smer			
AG	7,7	7,8	7,7
BT	8,0	7,8	7,9
MB	7,9	7,5	7,6
Skupaj	7,8	7,7	7,7

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	7,7	7,7

#### ANKETA 15

Grafično prikažite podatke iz tabele na sliki 4.2 (zgoraj). Grafikon oblikujte na podlagi podatkov na delovnem listu »Oblikovane tabele«. ♦

#### 4.4.3 Orodja za vrtilne tabele

Šteje od Zap. št.	Spol (m, ž)		Skupna vsota
Študijska smer (AG, BT ali MB)	m	ž	
AG	27	35	62
BT	10	15	25
MB	4	35	39
Skupna vsota	41	85	126

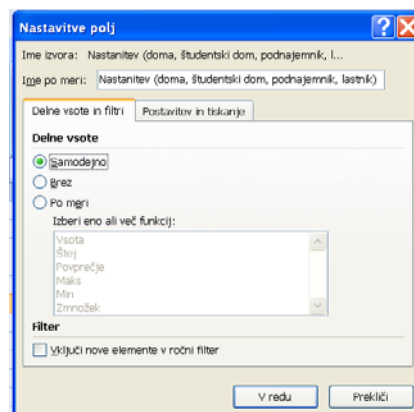
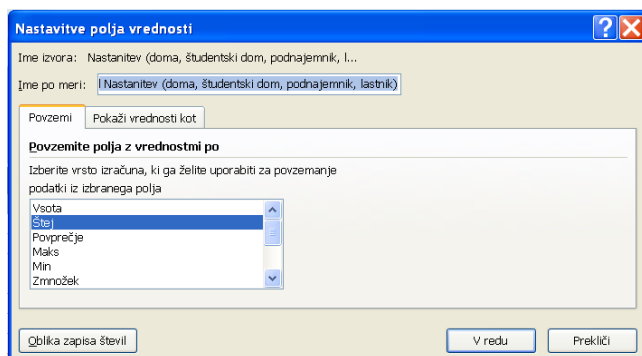
Orodja za vrtilne tabele so na dveh karticah *Možnosti* in *Načrt*. Na kartici *Možnosti* so ukazi, s katerimi spreminjamo vsebino vrtilne tabele, na kartici *Načrt* pa tisti, s katerimi spreminjamo izgled vrtilne tabele. Predstavili bomo samo tiste ukaze na kartici *Možnosti*, ki pogosto pridejo prav pri statistični analizi podatkov.

V skupini Vrtilna tabela je polje z imenom vrtilne tabele in ukazom **Možnosti**, ki določajo splošni izgled vrtilne tabele. Osnovne nastavitve teh možnosti v nadaljevanju ne bomo spreminjali.

#### 4.4.3.1 Nastavitve polj

V skupini ukazov Aktivno polje je ime spremenljivke, na podlagi katere je določena vsebina označene celice v vrtilni tabeli. V tej skupini so še trije ukazi **Nastavitve polj**, **Razširi celotno polje** in **Strni celotno polje**.

Ukaz **Nastavitve polj** odpre dve različni pogovorni okni, odvisno od tega, katera celica v vrtilni tabeli je označena. Če je v vrtilni tabeli izbrana celica iz območja **Vrednosti**, se odpre pogovorno okno **Nastavitve polja vrednosti**, sicer pa pogovorno okno **Nastavitve polj**:



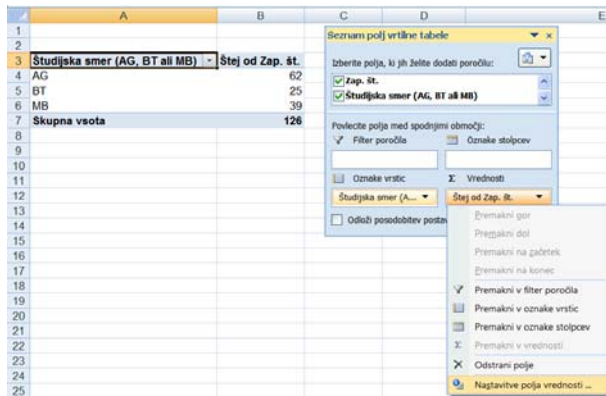
Pogovorno okno **Nastavitve polja vrednosti** lahko odpremo na več načinov:

- klik ukaza **Nastavitve polj** na kartici **Možnosti** v skupini Aktivno polje;

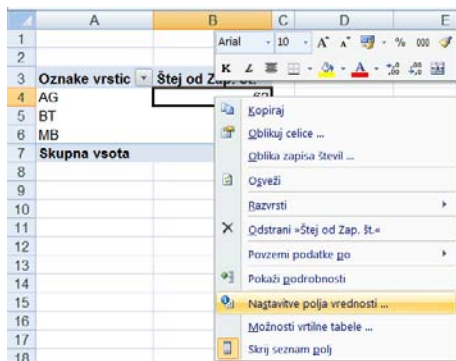
	A	B	C	D	E
1					
2					
3	Študijska smer (AG, BT ali MB)	Šteji od Zap. št.			
4	AG	62			
5	BT	25			
6	MB	39			
7	Skupna vsota	126			



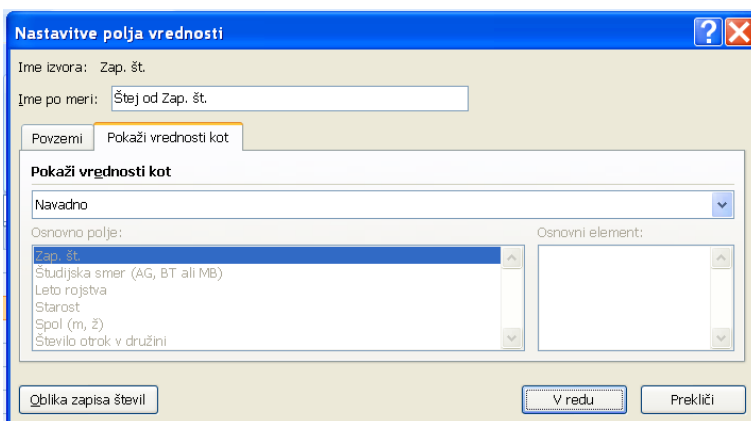
- klik ukaza **Nastavitve polj** na seznamu, ki se odpre s klikom na puščico v desnem delu izbranega gumba v območju **Vrednosti**;



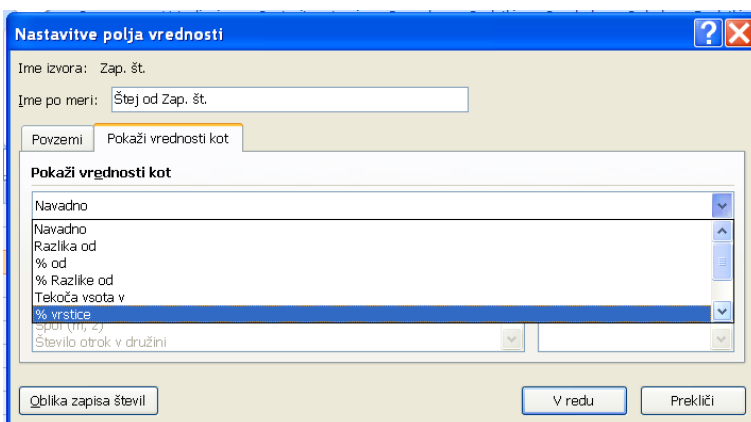
- desni klik na gumb vrednosti ali pa na celico z izračunano vrednostjo v vrtilni tabeli in izbira ukaza na pomožnem meniju.



- dvoklik na gumb vrednosti v vrtilni tabeli (na primer na celico B3, *Štej od Zap-št.*, v zgornji vrtilni tabeli). Ta način deluje samo v enosmerni vrtilni tabeli<sup>18</sup>;

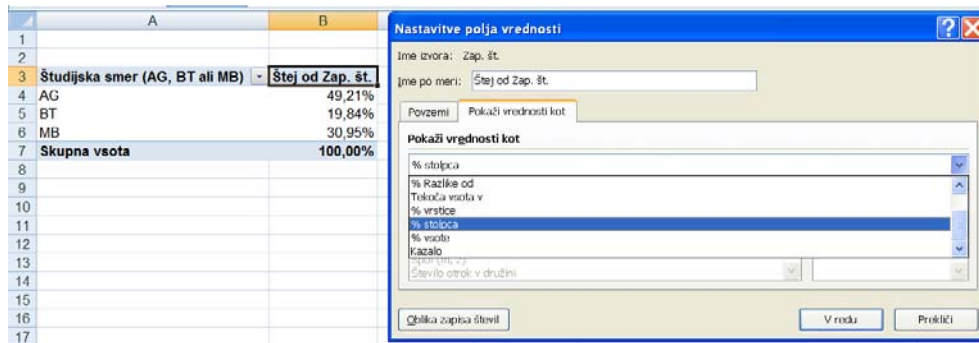


Omenili smo že izbiro funkcij za izračun vrednosti v vrtilni tabeli na kartici *Povzemi*. Tu bomo predstavili še možnosti, ki jih ponuja kartica *Pokaži vrednost kot*. Do zdaj smo imeli izbrano možnost *Navadno* (osnovna nastavitvev). Izbiramo lahko še med *Razlika od*, *% od*, *% Razlike od*, *% vrstice*, *% stolpca*, *% vsote* in *Kazalo*.



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

<sup>18</sup> Če imamo za prikaz vrtilne tabele v pogovornem oknu **Možnosti vrtilne tabele** na kartici *Prikaz* izbrano možnost *Videz klasične vrtilne tabele*, tak način deluje tudi v večsmernih tabelah.



Za druge možnosti bralca vabimo, da jih preizkusi sam.

#### ANKETA 16

Oblikujte vrtilno tabelo, v kateri so anketirani razvrščeni v skupine po *Spolu* in *Študijski smeri*. Podatke predstavite z relativnimi frekvencami za vsak spol posebej. Obrazložite vrednosti v tabeli.

**Rešitev:**

Štej od Zap. št.	Študijska smer (AG, BT ali MB)	Spol (m, ž)		Skupna vsota
		m	ž	
5	AG	65,9%	41,2%	49,2%
6	BT	24,4%	17,6%	19,8%
7	MB	9,8%	41,2%	31,0%
8	Skupna vsota	100,0%	100,0%	100,0%



#### ANKETA 17

Vrtilno tabelo iz prejšnje naloge preoblikujte tako, da podatke predstavite z relativnimi frekvencami za vsako študijsko smer posebej. Obrazložite vrednosti v tabeli.

Štej od Zap. št.	Študijska smer (AG, BT ali MB)	Spol (m, ž)		Skupna vsota
		m	ž	
5	AG	43,5%	56,5%	100,0%
6	BT	40,0%	60,0%	100,0%
7	MB	10,3%	89,7%	100,0%
8	Skupna vsota	32,5%	67,5%	100,0%



#### ANKETA 18

Oblikujte vrtilno tabelo z odstotkom anketiranih po številu otrok v družini. ◆

#### 4.4.3.2 Osveži podatke

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



#### ANKETA 19

Oblikujte vrtilno tabelo s povprečnim številom doseženih točk na maturi po spolu in po smereh študija (tabela 4.1). ♦

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)

Povprečno število doseženih točk na maturi	Spol		Skupaj
	Moški	Ženske	
Študijska smer			
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 20

Ali se vam zdijo rezultati pri (BT, ženske) smiselni?

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 *Številna doseženih točk na maturi* pojavi nenavadno visoka vrednost 86, ki je nedvomno napačna. Vrednost bomo popravili, vendar pa to ni možno na tem delovnem listu.

	A	B	C	D	E	F	G	H	I	J		
1	Zap. šte.	Študijska smer	Leto rojstva	Starost	Spol	Število otrok	Srednja šola	Gimnazija	Število doseženih točk na maturi	Štipendija (da/ne)	Najljubši študijski predmet	
2	93	BT	1986	21	ž		2	Gimnazija Škrlada		da	biologija	
3	90	BT	1986	21	ž		2	Škofjška klas	da	22	ne	biokemija
4	65	BT	1986	21	ž		1	Gimnazija Polda		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	Škofjška 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	Škofjška 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 Orr	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 21

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

## ANKETA 22

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

## 4.4.3.3 Združi

Ukaz **Združi izbor** na kartici **Možnosti** v skupini **Združi** smo uporabili pri oblikovanju vrtilne tabele na slikah 4.5 in 4.6. Ukaz **Razdruži** v isti skupini je dejaven le, če smo v vrtilni tabeli pred tem že uporabili ukaz **Združi izbor**. Ob izvajanju ukaza **Združi** Excel loči med opisnimi, celoštevilskimi in neceloštevilskimi spremenljivkami.

Opisne spremenljivke

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

	Štej od Zap.
doma	51
lastnik	1
podnajemnik	40
študentski dom	32
(prazen)	2
<b>Skupna vsota</b>	<b>126</b>

Vrtilna tabela se ob tem poveča za en stolpec, ki pripada novi spremenljivki *Nastanitev2*. Ta spremenljivka ima za vrednosti »doma« in »lastnik« skupno kategorijo »Skupina1«:

	Nastanitev (doma, študentski dom, podnajemnik, lastnik)2	Štej od Zap.
Skupina1	doma	51
	lastnik	1
podnajemnik	podnajemnik	40
študentski dom	študentski dom	32
(prazen)	(prazen)	2
<b>Skupna vsota</b>		<b>126</b>

Če osnovno spremenljivko *Nastanitev* odstranimo iz vrtilne tabele, dobimo naslednjo tabelo:

	Štej od Zap.
Skupina1	52
podnajemnik	40
študentski dom	32
(prazen)	2
<b>Skupna vsota</b>	<b>126</b>

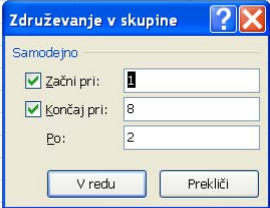
Če kategorije, ki jih želimo združiti v skupino, v vrtilni tabeli ne nastopajo skupaj, jih pred uporabo ukaza **Združi izbor** označimo tako, da med označevanjem držimo tipko CTRL.

### Številske spremenljivke

Za številske spremenljivke možnost **Združi izbor** omogoča, da enote podatkovne zbirke razvrstimo v razrede glede na vrednosti izbrane številske spremenljivke. Če ima

številka spremenljivka manjkajoče vrednosti, poleg razredov v vrtilni tabeli ostane tudi kategorija »prazen«.

	A	B	C	D	E
1					
2					
3	Število otrok v družini	Štej od Zap. št.			
4	(prazen)	1			
5	1-2	90			
6	3-4	29			
7	5-6	4			
8	7-8	2			
9	<b>Skupna vsota</b>	<b>126</b>			
10					
11					
12					



Preden uporabimo ukaz **Združi izbor** 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 spodnjo mejo prvega razreda (**Začni pri:**) in zgornjo mejo zadnjega razreda (**Končaj pri:**) ter širino razredov (**Po:**).

Za celoštevilске spremenljivke ukaz **Združi izbor** naredi meje razredov tako, da se točno vidi, katera vrednost sodi v posamezni razred (slika 4.5).

	A	B	C	D	E
1					
2					
3	Kolikokrat ste bili v zadnjem letu v kinu?	Štej od Zap. št.			
4		0	10		
5		1	14		
6		2	19		
7		3	17		
8		4	10		
9		5	20		
10		6	7		
11		7	3		
12		8	2		
13		9	3		
14		10	13		
15		13	1		
16		15	3		
17		20	2		
18		25	1		
19		30	1		
20	<b>Skupna vsota</b>		<b>126</b>		

	A	B
1		
2		
3	Kolikokrat ste bili v zadnjem letu v kinu?	Štej od Zap. št.
4	0-9	105
5	10-19	17
6	20-30	4
7	<b>Skupna vsota</b>	<b>126</b>

	A	B	C	D	E
1					
2					
3	Kolikokrat ste bili v zadnjem letu v kinu?	Štej od Zap. št.			
4	0-4	70			
5	5-9	35			
6	10-14	14			
7	15-19	3			
8	20-24	2			
9	25-30	2			
10	<b>Skupna vsota</b>	<b>126</b>			
11					

Slika 4.5: Število anketiranih glede na število obiskov v kinu v zadnjem letu (zgoraj), vrednosti, združene v razrede širine 10 (sredina) in 5 (spodaj), (ukaz **Združi izbor**, celoštevilaska spremenljivka)

Za neceloštevilske spremenljivke ukaz **Združi izbor** 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).

Koliko ur na teden v povprečju porabite za deskanje po svetovnem spletu?	Štej od Zap. št.
0	2
1	10
2	16
2,5	1
3	15
3,5	1
4	7
5	18
6	5
7	18
8	5
8,5	1
10	13
12	1
14	1
15	3
16	1
20	4
25	1
28	1
30	1
50	1
<b>Skupna vsota</b>	<b>126</b>

Koliko ur na teden v povprečju porabite za deskanje po svetovnem spletu?	Štej od Zap. št.
0-5	52
5-10	47
10-15	15
15-20	4
20-25	4
25-30	2
30-35	1
45-50	1
<b>Skupna vsota</b>	<b>126</b>

Slika 4.6: Število anketiranih glede na število ur deskanja po svetovnem spletu na teden (levo), vrednosti, združene v razrede (desno), (ukaz **Združi izbor**, neceloštevilska spremenljivka)

Če so med številskimi vrednostmi tudi opisne vrednosti, Excel vrne opozorilo: *Tega izbora ni mogoče združiti.*

Število doseženih točk na maturi	Štej od Zap. št.
10	2
11	1
12	4
13	6
14	5
15	7
16	3
17	8
18	2
19	2
20	5
21	6
22	8
23	7
24	12
25	5
26	3
27,5	1
28	2
29	7
30	2
31	1
32	2
34	1
NE VEM	1
(prazen)	23
<b>Skupna vsota</b>	<b>126</b>

## ANKETA 23

Na delovnem listu »podatki« pobrišite vrednosti »NE VEM« pri spremenljivki **Število doseženih točk na maturi**. Za vsako smer študija anketirane razvrstite v frekvenčno tabelo glede na **Število doseženih točk na maturi**. Širina razredov naj bo 4 točke. Ne pozabite »osvežiti« vrtilne tabele.

Vrtilno tabelo parkirajte na delovni list »Oblikovane tabele« in podatke iz tabele grafično prikažite s histogramom za vsako smer študija posebej. Na sliki spustite tiste, ki nimajo podatkov.

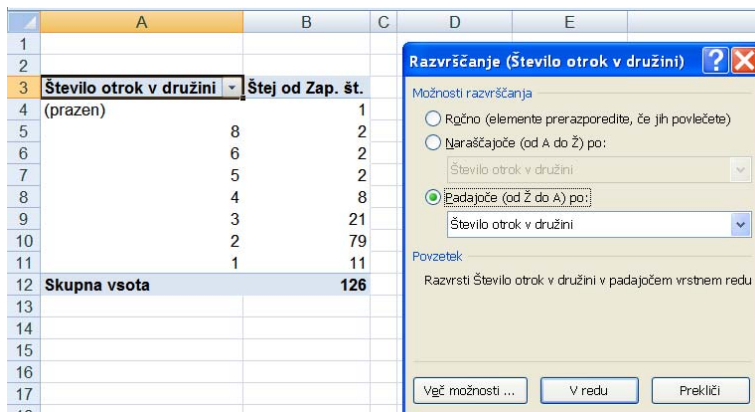
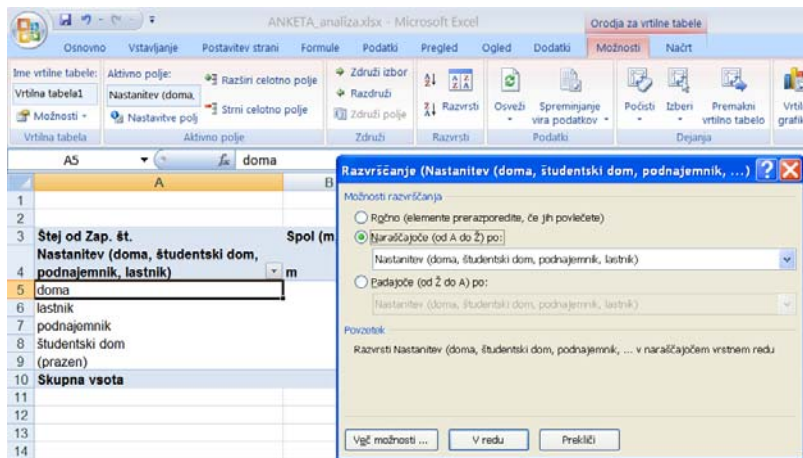
**Rešitev:**

Število doseženih točk na maturi	Študijska smer			Skupaj
	AG	BT	MB	
Ni podatka	17	2	5	24
10-14	13			13
14-18	23			23
18-22	6	4	5	15
22-26	2	9	21	32
26-30	1	7	5	13
30-34		3	3	6
Skupaj	62	25	39	126

**4.4.3.4 Razvrščanje in filtriranje podatkov v vrtilni tabeli**

Razvrščanje in filtriranje v osnovni podatkovni zbirki smo že opisali, podobno ti dve orodji delujeta v vrtilnih tabelah, ukaz **Razvrsti** na kartici **Možnosti** v **Orodju za vrtilne tabele**. Ukaz razvršča vrstice oz. stolpce glede na vsebino celice, ki je izbrana ob njegovi uporabi. Z razvrščanjem določimo vrstni red vrednosti v vrsticah oz. stolpcih. Po osnovni nastavitvi se opisne vrednosti spremenljivk razvrstijo po abecedi; na primer, če v območje **Oznake vrstic** postavimo spremenljivko **Nastanitev**, prvo vrstico vrtilne tabele določa vrednost »**doma**«, drugo »**lastnik**« itd.

Vrednosti številskih spremenljivk se po osnovni nastavitvi razvrstijo od najmanjše do največje.



Če v pogovornem oknu **Razvrščanje** izberemo možnost *Ročno*, vrednosti izbrane spremenljivke prerazporedimo z vlečenjem z miško. Pri tem moramo pogovorno okno najprej zapreti s klikom na gumb **V redu**, nato izberemo celico z vrednostjo, ki jo želimo prestaviti. Miško postavimo na rob označene celice tako, da se bel križec spremeni v tanki prekrizani puščici, ki omogočata, da s potegom z miško vsebino celice prestavimo.

V vrtilnih tabelah lahko vrstni red stolpcev oz. vrstic določimo tudi na podlagi izračunanih vrednosti.

Štej od Zap. št.	Spol (m, ž)		Skupna vsota
Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž	
5 doma	20	31	51
6 podnajemnik	7	33	40
7 študentski dom	14	18	32
8 (prazen)		2	2
9 lastnik		1	1
10 Skupna vsota	41	85	126

**Razvrsti po vrednosti**

Možnosti razvrščanja:  Od najmanjšega do največjega  Od največjega do najmanjšega

Smer razvrščanja:  Od vrha proti dnu  Od leve proti desni

Povzetek: Razvrsti Nastanitev (doma, študentski dom, podnajemnik, ... po Štej od Zap. št. v padajočem vrstnem redu uporaba vrednosti v tem stolpcu: Skupna vsota

V redu Prekliči

Številno otrok v družini	Štej od Zap. št.
2	79
3	21
1	11
4	8
8	2
6	2
5	2
(prazen)	1
12 Skupna vsota	126

**Razvrsti po vrednosti**

Možnosti razvrščanja:  Od najmanjšega do največjega  Od največjega do najmanjšega

Smer razvrščanja:  Od vrha proti dnu  Od leve proti desni

Povzetek: Razvrsti Številno otrok v družini po Štej od Zap. št. v padajočem vrstnem redu

V redu Prekliči

Če želimo, da se v vrtilni tabeli posamezna vrstica ali stolpec ne pokažeta, uporabimo filter bri izbrani spremenljivki, ki določa vrstice oz. stolpce vrtilne tabele. Na primer, če ne želimo, da se izpiše število enot, za katere nimamo vrednosti spremenljivke *Nastanitev* v tabeli 4.2, odstranimo kljukico pri vrednosti »prazen«:

Štej od Zap. št.	Spol (m, ž)		Skupna vsota
Nastanitev (doma, študentski dom, podnajemnik, lastnik)	m	ž	
5 doma	20	31	51
6 lastnik	7	33	40
7 študentski dom	14	18	32
9 Skupna vsota	41	83	124

Bočisti filter iz »Nastanitev (doma,...«

Filtri oznak

Filtri vrednosti

(Izberi vse)

doma

lastnik

podnajemnik

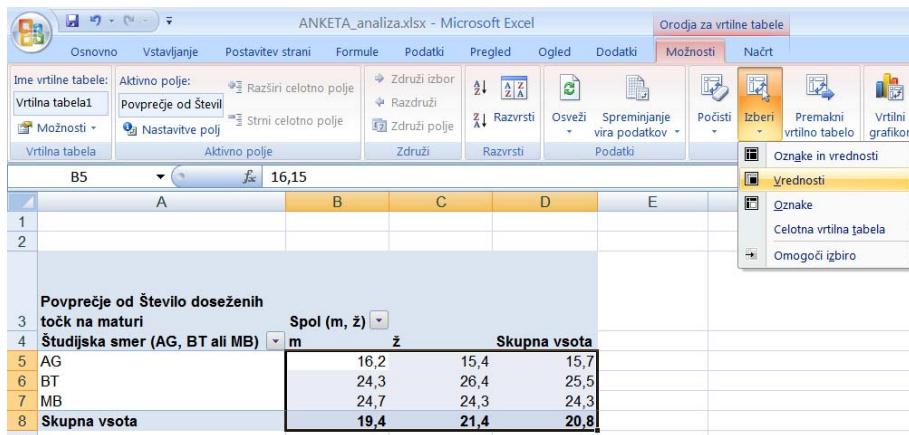
študentski dom

(prazen)

V redu Prekliči

#### 4.4.3.5 Izberi

Z ukazom **Izberi** na kartici **Možnosti** v skupini Dejanja označimo celo ali pa le del vrtilne tabele. Ukaz pride prav, kadar želimo celo ali del vrtilne tabele kopirati. Isto dosežemo tudi z običajnim potegom z miško.



#### ANKETA 24

Na podlagi podatkov v datoteki ANKETA\_ANALIZA.xlsx odgovorite na spodnja vprašanja. Odgovore napišite v datoteko ANKETA\_ANALIZA.docx, kamor kopirajte tudi narejene tabele in grafikone.

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

Koliko odstotkov anketiranih živi v družinah z več kot tremi otroki?

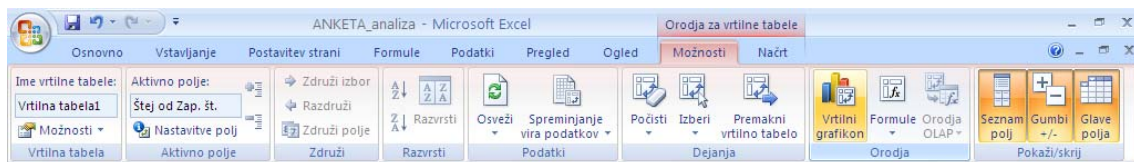
Kolikšna je povprečna, minimalna in maksimalna starost anketiranih? Kolikšne so te vrednosti posebej za moške in za ženske?

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. ♦

### 4.5 Vrtilni grafikon

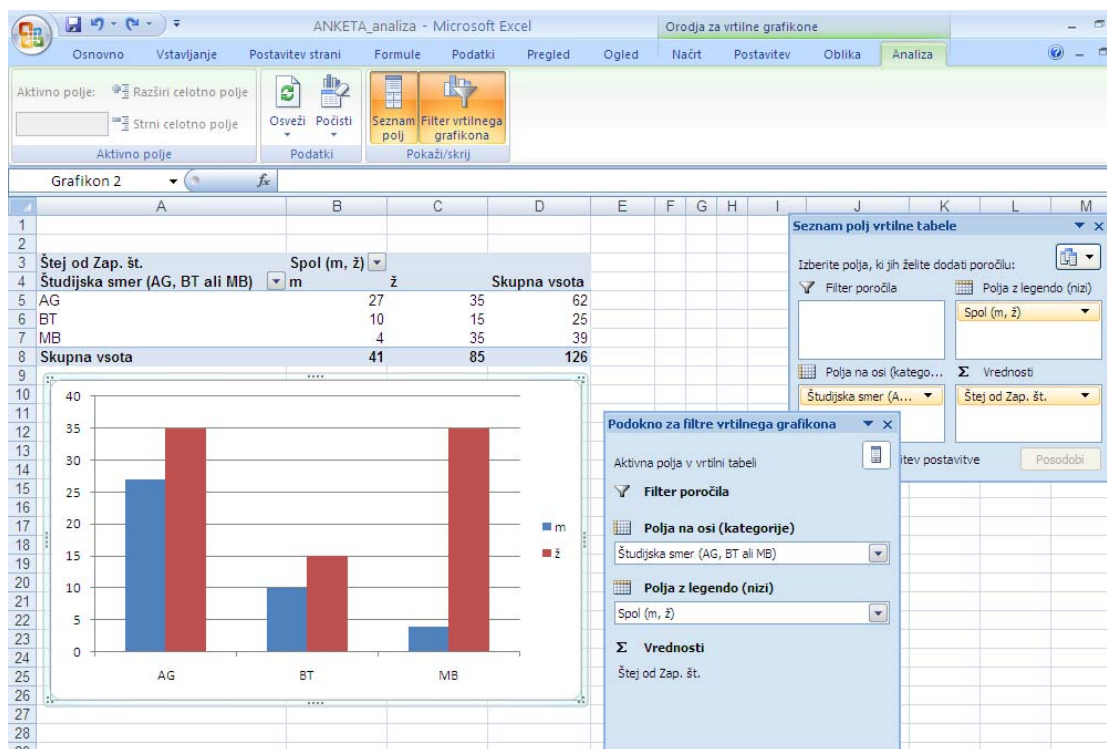
Med orodji za vrtilne tabele na kartici **Možnosti** v skupini Orodja najdemo tudi ukaz **Vrtilni grafikon**, ki grafično prikaže podatke v vrtilni tabeli.



Po kliku ukaza **Vrtilni grafikon** se odpre pogovorno okno **Vstavi grafikon**, v katerem

izberemo vrsto grafikona. Vrtilni grafikon se pojavi na delovnem listu ob vrtilni tabeli, poleg njega pa pogovorno okno **Podokno za filtre vrtilnega grafikona**.

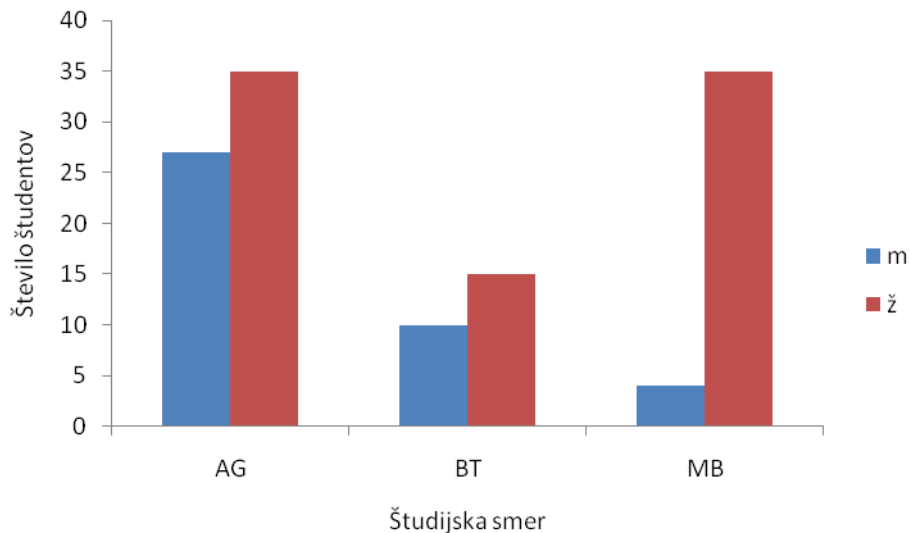
Poglejmo primer vrtilnega grafikona za vrtilno tabelo, v kateri preštejemo, koliko študentov in koliko študentk iz posamezne študijske smeri je bilo anketiranih.



Vrtilni grafikon je po osnovni nastavitvi brez naslovov osi. Vrednosti spremenljivke, ki določa stolpce (**Oznake stolpcev**), nastopajo v legendi. Oznake na vodoravni osi predstavljajo vrednosti spremenljivke, ki določa vrstice vrtilne tabele (**Oznake vrstic**). Skladno s tem se v pogovornem oknu **Seznam polj vrtilne tabele** namesto izraza **Oznake stolpcev** pojavi izraz **Polja z legendo (nizi)** in namesto izraza **Oznake vrstic** se pojavi izraz **Polja na osi (kategorije)**.

Grafikon lahko v **Orodju za vrtilne grafikone** poljubno spreminjamo z ukazi in nastavitvami, ki jih že poznamo, poleg tega pa lahko izbor prikazanih vrednosti dodatno spreminimo z ukazi na dodatni kartici **Analiza** ali nastavitvami v **Podoknu za filtre vrtilnega grafikona**. Če postavitev vrtilne tabele spreminimo, se ustrezno spremeni

tudi vrtilni grafikon, in obratno, če postavitev v oknu **Podokna za filtre vrtilnega grafikona** spremenimo, se spremeni tudi vrtilna tabela.



Slika 4.7: Število anketiranih študentov po spolu in študijski smeri. Dopolnjeni vrtilni grafikon.

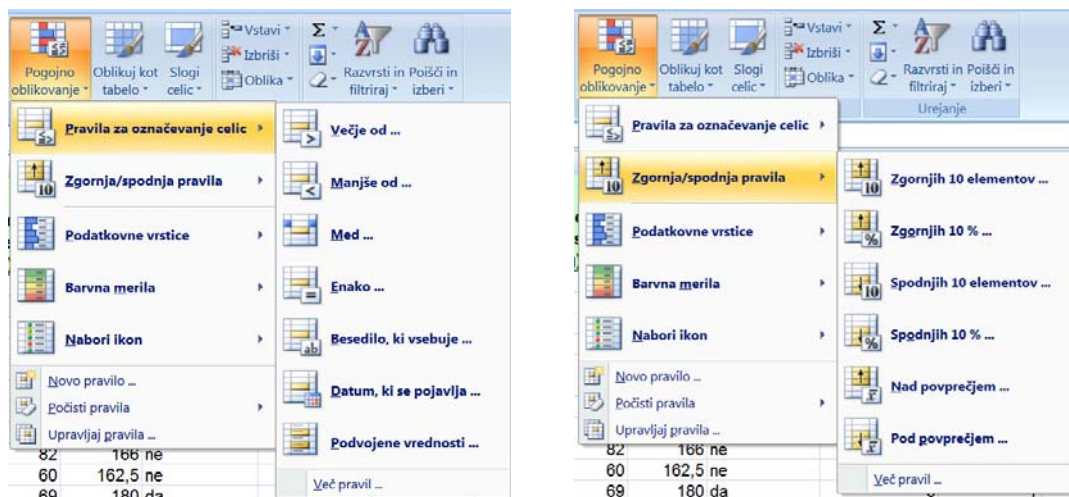
S podrobnejšo razlago vrtilnih grafikonov se na tem mestu ne bomo ukvarjali. V splošnem se bomo še naprej držali navodila, da grafikone oblikujemo na podlagi »parkiranih« vrtilnih tabel. Vrtilni grafikon je priročen za hitri grafični prikaz podatkov, zato si bomo nekaj primerov njegove uporabe ogledali v naslednjem zgledu. Določenih vrst grafikonov vrtilni grafikon ne omogoča (npr. *Raztreseni*).

## 4.6 Pogojno oblikovanje celic

Pogojno oblikovanje celic uporabimo za vizualno raziskovanje številskih vrednosti v označenem obsegu celic oz. v podatkovni zbirki. Ukaz **Pogojno oblikovanje** v skupini Slogi na kartici **Osnovno** olajša označevanje obsegov celic na tak način, da lahko hitro prepoznamo nenavadne vrednosti, najmanjše in največje vrednosti in relativni položaj vrednosti v označenem obsegu celic. Pogojna oblika spremeni podobo obsega celic na podlagi izbranega pogoja. Pogojno lahko oblikujemo tako obsege celic v podatkovni zbirki kot tudi v vrtilnih tabelah.

	H	I	J	K	L
	Gimnazija (da/ne)	Število doseženih točk na maturi	Štipendija (da/ne)	Najljubši predmet v 1. letniku	Povprečna ocena v 1. letniku
1	ne		ne	botanika	7,0
2	da		17 ne	botanika	7,0
3	da		16 ne	matematika	7,0
4	da		ne	biologija	7,5
5	da		ne	fizikalna kemija	7,0
6	da		21 ne	uvod v biotehnolog	7,1
7	da		17 da	botanika	7,9
8	ne		16 ne		
9	ne		da	botanika	8,0
10	ne			fiziologija z anatoni	7,8
11	ne		15		

Preden uporabimo katerikoli ukaz pogojnega oblikovanja, moramo označiti obseg celic, v katerem naj se to oblikovanje izvede. Prva dva ukaza pogojnega oblikovanja: **Pravila za označevanje celic** in **Zgornja/spodnja pravila** delujeta podobno kot ukaz **Filter**, le da izbrane celice le označita (pobarvata); vidne, a nepobarvane ostanejo tudi celice, ki ne ustrezajo pogoju.



ANKETA 25

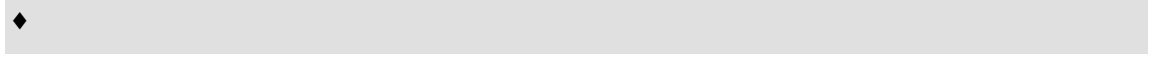
Na delovnem listu »Podatki« za spremenljivko *Povprečna ocena v 1. letniku* označite vse celice z vrednostjo večjo od 7,5.



**Navodilo:** najprej označimo stolpec s podatki o povprečni oceni v 1. letniku, nato uporabimo ukaz **Pogojno oblikovanje/Pravila za označevanje celic/Večje od**. V pogovornem oknu **Večje kot** v levo polje napišemo vrednost 7,5 in pustimo v desnem polju samodejno izbrano barvo označitve.

V spošnem v desnem polju izbiramo med vnaprej določenimi barvami za ozadje celic in pisavo v njih. Z možnostjo *Oblikovanje po meri* barve lahko oblikujemo celice po svoje.

	K	L	M	N	O	P	Q	R
1	Najljubši predmet v 1. letniku	Povprečna ocena v 1. letniku	vozite z osebnim avtom?	Nastanitev (doma, študentski dom,	Telesna masa	Telesna višina	ukvarjate s športom	teden se ukvarjate s
2	botanika	7,0	občasno	študentski dom				
3	botanika	7,0	ne	študentski dom				
4	matematika	7,0	občasno	študentski dom				
5	biologija	7,5	da	študentski dom				
6	fizikalna kemija	7,0	da	študentski dom				
7	uvod v biotehnologijo	7,1	da	študentski dom				
8	botanika	7,9	da	študentski dom				
9	botanika	8,0	ne	študentski dom				
10	botanika	8,0	ne	študentski dom				
11	fiziologija z anatomijo	7,8	da	doma	60	162,5	ne	
12	biokemija	8,0	ne	študentski dom	69	180	da	
13	biologija celice	7,1	občasno	podnajemnik	62	165	da	
14	biokemija	6,8	da	podnajemnik	60	165	da	
15	biologija celice	7,5	ne	podnajemnik	55	168	da	
16	biologija celice	7,9	občasno	doma	65	179	da	
17	pedologija	7,5	ne	študentski dom	85	183	da	
18	botanika	6,5	občasno	študentski dom	80	194	ne	
19	botanika	7,0	občasno	doma	80	179	da	



Na izbranem obsegu celic lahko uporabimo več zaporednih ukazov pogojnega oblikovanja. Celice, izbrane pod različnimi pogoji, lahko označimo z različnimi barvami.

#### ANKETA 26

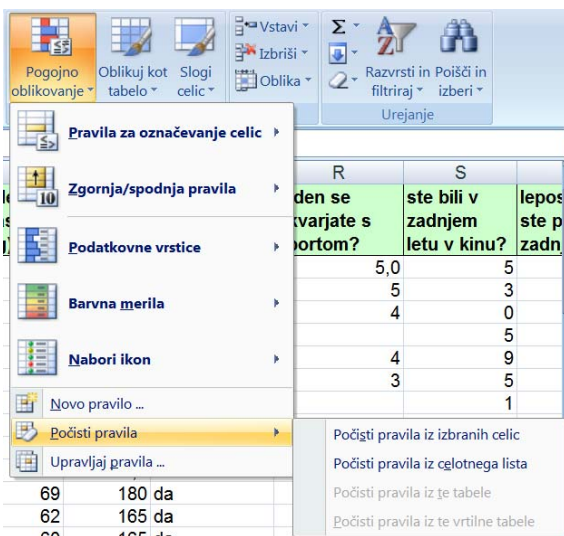
Na označenem obsegu celic iz prejšnje naloge (*Povprečna ocena v 1. letniku*) označite še vse celice z vrednostjo manjšo kot 7,0 z zelenim polnilom in temno zelenim besedilom<sup>19</sup>. ♦

<sup>19</sup> Prazne celice imajo v kontekstu pogojnega oblikovanja vrednost 0.

	K	L	M	N	O	P	Q	R
	Najljubši predmet v 1. letniku	Povprečna ocena v 1. letniku	vozite z osebnim avtom?	Nastanitev (doma, študentski dom,	Telesna masa	Telesna višina	ukvarjate s športom	teden se ukvarjate s
1			(da/	Manjše kot				n?
2	botanika	7,0	občasno					
3	botanika	7,0	ne					
4	matematika	7,0	občasno					
5	biologija	7,5	da					
6	fizikalna kemija	7,0	da					
7	uvod v biotehnolog	7,1	da					
8	botanika	7,9	da	študentski dom	56	165	ne	
9			ne	študentski dom	65	170	ne	
10	botanika	8,0	ne	študentski dom	82	166	ne	
11	fiziologija z anatomi	7,8	da	doma	60	162,5	ne	
12	biokemija		ne	študentski dom	69	180	da	
13	biologija celice	7,1	občasno	podnajemnik	62	165	da	
14	biokemija	6,8	da	podnajemnik	60	165	da	
15	biologija celice	7,5	ne	podnajemnik	55	168	da	
16	biologija celice	7,9	občasno	doma	65	179	da	
17	pedologija	7,5	ne	študentski dom	85	183	da	
18	botanika	6,5	občasno	študentski dom	80	194	ne	

Pri ustvarjanju pogojne oblike se lahko sklicujemo tudi na vsebino celice na delovnem listu, kar pomeni, da v levo polje pogovornega okna **Manjše kot** (**Večje kot**,...) napišemo referenco na to celico.

Pogojno oblikovanje odstranimo z eno od možnosti na meniju ukaza **Pogojno oblikovane/Počisti pravila**:



Prva dva ukaza v meniju **Pogojno oblikovanje** označita le celice v predhodno označenem obsegu celic, ki ustrezajo pogoju. Trije ukazi, ki sledijo: **Podatkovne vrstice**, **Barvna merila** in **Nabori ikon**, pa v izbranem slogu oblikujejo vse celice predhodno izbranega obsega celic.

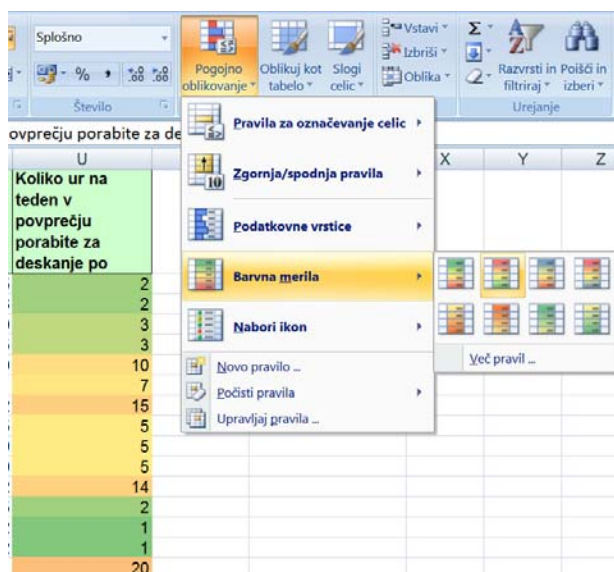
Pogojno oblikovanje celic s slogom **Podatkovne vrstice** v celice poleg vrednosti nariše vodoravne stolpce (podatkovne vrstice), katerih dolžina predstavlja vrednost v celici relativno na največjo vrednost v izbranem obsegu celic. V podatkovni zbirki je smiselno uporabiti ta slog za vsak stolpec (spremenljivko) posebej.

	R	S	T	U
1	Koliko ur na teden se ukvarjate s športom?	Kolikokrat ste bili v zadnjem letu v kinu?	Koliko leposlovnih knjig ste prebrali v zadnjem letu?	Koliko ur na teden v povprečju porabite za deskanje po
2		5,0	5	5
3		5	3	6
4		4	0	0
5			5	5
6		4	9	0
7		3	5	1
8			1	2
9			2	5
10			3	10
11			0	30
12		6	7	2
13		3	15	5
14		2	10	2
15		5	10	12
16		6	9	1
17		8	5	0
18			10	0
19		3	2	5
20		4	9	5
21		6	15	0
22			2	2
23		4	1	7

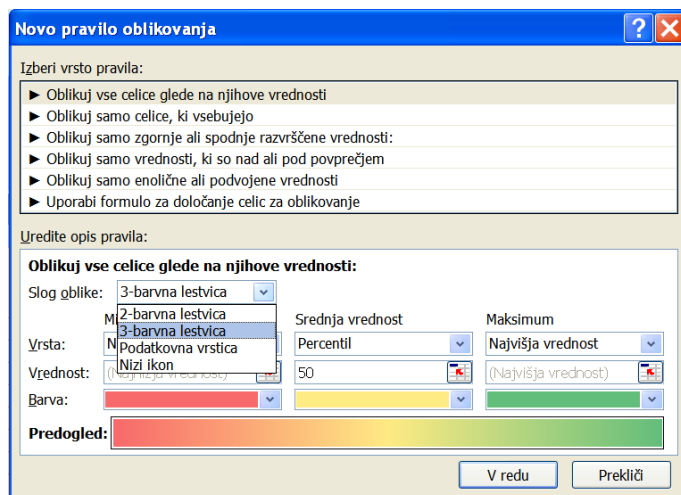
#### ANKETA 27

Za vsako od zadnjih štirih spremenljivk na delovnem listu »Podatki« (glej sliko zgoraj) uporabite slog **Podatkovna vrstica**. Od česa je odvisna dolžina podatkovnih vrstic v celicah? Spremenite širino stolpca izbrane spremenljivke in si oglejte, kako to vpliva na dolžino pripadajočih podatkovnih vrstic. ♦

Pogojno oblikovanje s slogom **Barvna merila** izbrani obseg celic pobarva z različnimi odtenki dveh ali treh barv. S stopnjevanjem dveh ali treh barv lahko primerjamo vrednosti v obsegu celic. Spodnja barva na izbrani barvni lestvici predstavlja nižje vrednosti, nato barva po odtenkih slabi in se prelije v drugo barvo, ki predstavlja višje vrednosti.



Lastno barvno lestvico sloga **Barvno merilo** določimo s klikom možnosti *Več pravil*. V pogovornem oknu **Novo pravilo oblikovanja** v razdelku *Uredite opis pravila* nastavimo želene možnosti oblikovanja. Posameznih možnosti na tem mestu ne bomo natančneje opisovali, ker delujejo podobno kot pri oblikovanju celic in grafikonov.



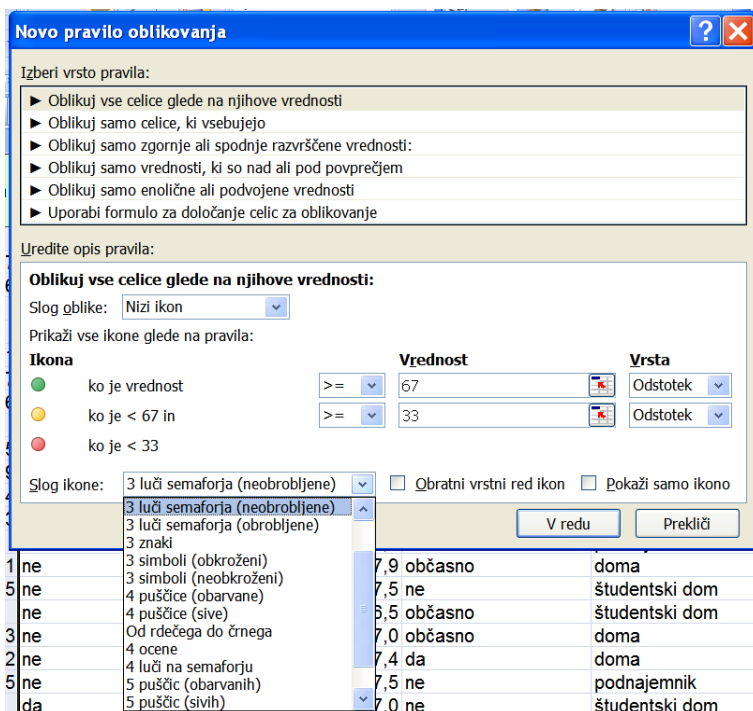
Pogovorno okno **Novo pravilo oblikovanja** se odpre tudi z ukazom **Novo pravilo** v glavnem meniju ukaza **Pogojno oblikovanje**. V prvem razdelku *Izberi vrsto pravila* izbiramo med možnostmi, ki temeljijo na različnih pogojih izbire celic (*Oblikuj vse celice glede na njihove vrednosti*, *Oblikuj samo celice, ki vsebujejo*, *Oblikuj samo zgornje ali spodnje razvrščene vrednosti*, ...). Drugi razdelek *Uredite opis pravila*, pa

ponudi možnosti, ki določajo slog označevanja in so vezane na izbiro možnosti v prvem razdelku.

Ostane nam še zadnji slog pogojnega označevanja **Nabori ikon**. V tem primeru se poleg vrednosti v celice izbranega obsega podatkov narišijo še ikone, ki na različne načine predstavljajo vrednosti v celicah.

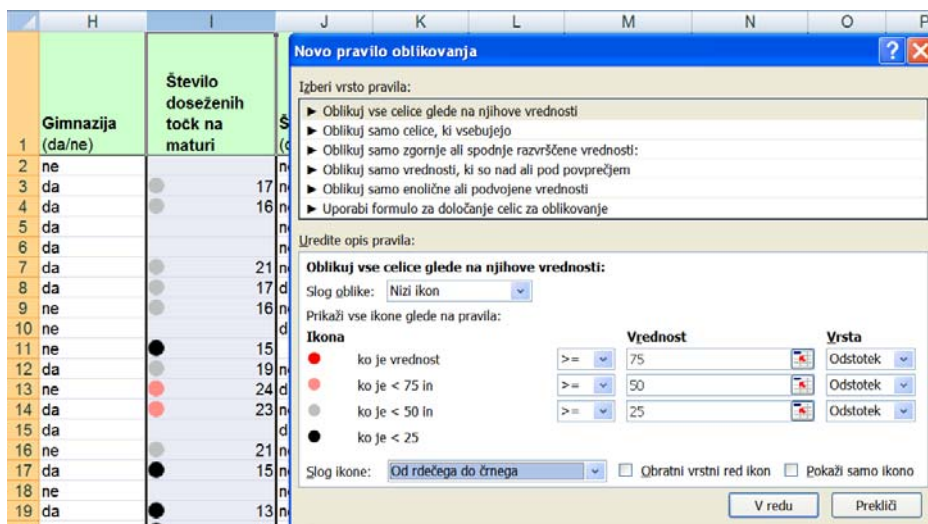
	H	I	J	K	L
	Gimnazija (da/ne)	Število doseženih točk na maturi	Štipendija (da/ne)	Najljubši predmet v 1. letniku	Povprečna ocena v 1. letniku
1					
2	ne		ne	botanika	7,0
3	da		17 da	botanika	7,0
4	da		16 ne	matematika	7,0
5	da		ne		
6	da		ne		
7	da		21 ne		
8	da		17 da		
9	ne		16 ne		
10	ne		da		
11	ne		15		
12	da		19 ne		
13	ne		24 da		
14	da		23 ne		
15	da		da		
16	ne		21 ne		
17	da		15 ne		
18	ne		ne		
19	da		13 ne		

S klikom možnosti *Več pravil* se odpre pogovorno okno **Novo pravilo oblikovanja**. V razdelku *Uredite opis pravila* razberemo pravila za izbiro ustrezne ikone:



## ANKETA 28

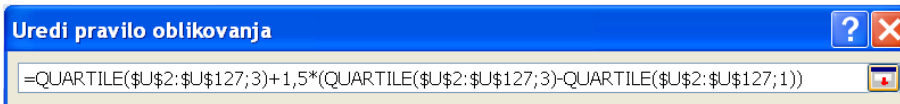
Celice z vrednostmi spremenljivke *Število doseženih točk na maturi* označite z ikonami tako, da bodo vrednosti razmejene glede na kvartile. ♦



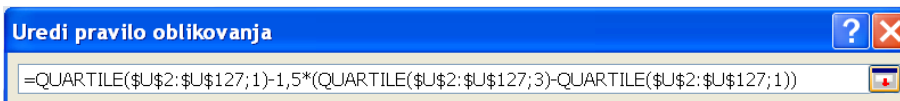
## ANKETA 29

Celice z vrednostmi spremenljivke *Koliko ur na teden porabite za deskanje po svetovnem spletu* označite z ikonami tako, da bodo spodnji osamelci označeni z zeleno piko, zgornji z rdečo, ostale vrednosti pa z rumeno piko.

**Navodilo:** v pogovornem oknu **Novo pravilo oblikovanja** v razdelku Uredite opis pravila izberite *Slog oblike: Nizi ikon*, za vrsto vrednosti pogoja izberite možnost *Formula* ter v polje *Vrednost* napišite formuli za zgornje osamelce:

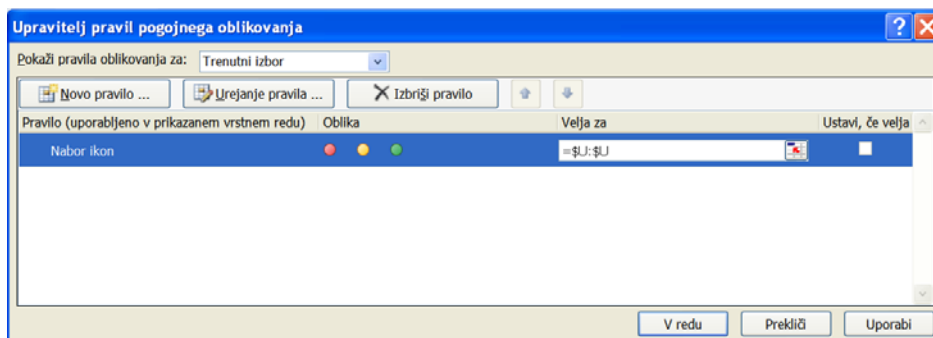


oz. za spodnje osamelce:



Izberite tudi možnost *Obratni vrstni red ikon*, ker bodo sicer zgornji osamelci označeni z zeleno piko. ♦

Obstoječa pogojna oblikovanja lahko spreminjamo v okviru ukaza **Upravljalj pravila**, ki odpre pogovorno okno **Upravitelj pravil pogojnega oblikovanja**.



### Zgled 4.1: Meteorološki podatki – časovna vrsta

#### METEOROLOŠKI PODATKI 1

Odprite datoteko METEOROLOSKI PODATKI.xlsx s podatki, ki jih je v obdobju od 14. 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.docx. 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, enote,...). Spremenljivkam določite mersko lestvico. Opišite obliko celic za posamezno spremenljivko podatkovne zbirke. Kolikšen je časovni razmik med podatki? ♦

#### METEOROLOŠKI PODATKI 3

Datoteka METEOROLOSKI PODATKI.xlsx je zelo velika. Koliko zapisov je v 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

Če je hitrost vetra je 0 m/s (brezvetrje), je *Smer vetra* nedoločljiva. To je v spremenljivki *Smer vetra* označeno z oznako »n«. Koliko je takih zapisov? ♦

#### METEOROLOŠKI PODATKI 5

Z ukazom **Filter** izberite podatke treh zaporednih dni (5. - 7. 5. 2004) in jih spravite na nov delovni list z imenom »izbor«. Naloge, ki sledijo, naredite na podlagi teh podatkov.

**Navodilo:** S 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 nov delovni list ter tam uporabimo ukaz **Prilepi**. ♦



## 4 Podatkovna zbirka in vrtilne tabele

	A	B	C	D	E	F	G	H	I	J
1	Datum	Čas	Max. hitrost vetra (m/s)	Min. hitrost vetra (m/s)	Padavine (mm)	Temperatura zraka (°C)	Globalna osočenost [W/m <sup>2</sup> ]	Relativna vlaga (%)	Hitrost vetra (m/s)	Smer vetra
6050	5. 5. 2004		0,0	0,0	0,0	12,8	2	105	0,0	n
6051	5. 5. 2004		0,0	0,0	0,0	12,9	2	106	0,0	n
6052	5. 5. 2004		1,0	0,0	0,0	13,2	2	106	0,3	228
6053	5. 5. 2004		0,0	0,0	0,0	13,2	2	106	0,0	n
6054	5. 5. 2004		0,0	0,0	0,0	13,0	2	106	0,0	n
6055	5. 5. 2004		0,0	0,0	0,0	12,9	2	106	0,0	n
6056	5. 5. 2004		0,0	0,0	0,0	12,7	2	105	0,1	300
6057	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,1	300
6058	5. 5. 2004		0,0	0,0	0,0	12,7	2	106	0,3	300
6059	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,4	300
6060	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,0	n
6061	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,0	n

**Lasten samodejni filter**

Pokaži vrstice, kjer:

Datum

je za ali enako

In  Ali

je pred ali enako

? nadomesti kateri koli znak  
\* nadomesti kateri koli niz znakov

V redu    Prekliči

ali pa

	A	B	C	D	E	F	G	H	I	J
1	Datum	Čas	Max. hitrost vetra (m/s)	Min. hitrost vetra (m/s)	Padavine (mm)	Temperatura zraka (°C)	Globalna osočenost [W/m <sup>2</sup> ]	Relativna vlaga (%)	Hitrost vetra (m/s)	Smer vetra
6050	5. 5. 2004		0,0	0,0	0,0	12,8	2	105	0,0	n
6051	5. 5. 2004		0,0	0,0	0,0	12,9	2	106	0,0	n
6052	5. 5. 2004		1,0	0,0	0,0	13,2	2	106	0,3	228
6053	5. 5. 2004		0,0	0,0	0,0	13,2	2	106	0,0	n
6054	5. 5. 2004		0,0	0,0	0,0	13,0	2	106	0,0	n
6055	5. 5. 2004		0,0	0,0	0,0	12,9	2	106	0,0	n
6056	5. 5. 2004		0,0	0,0	0,0	12,7	2	105	0,1	300
6057	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,1	300
6058	5. 5. 2004		0,0	0,0	0,0	12,7	2	106	0,3	300
6059	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,4	300
6060	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,0	n
6061	5. 5. 2004		0,0	0,0	0,0	12,6	2	106	0,0	n

### METEOROLOŠKI PODATKI 6

Ali so med podatki manjkajoče vrednosti? ♦

### METEOROLOŠKI PODATKI 7

Za vsak dan izračunajte povprečno, minimalno in maksimalno dnevno temperaturo zraka. Vrtilno tabelo parkirajte na delovni list »oblikovane tabele«. Vsebinsko obrazložite vrednosti v stolpcu »Skupna vsota« in ime stolpca ustrezno spremenite.

	A	B	C	D	E	F
1						
2						
3		Datum				
4	<b>Vrednosti</b>	<b>5. 5. 2004</b>	<b>6. 5. 2004</b>	<b>7. 5. 2004</b>	<b>Skupna vsota</b>	
5	Povprečje od Temperatura zraka (°C)	14,1	14,3	10,0	12,8	
6	Min od Temperatura zraka (°C)	11,5	7,8	6,4	6,4	
7	Maks od Temperatura zraka (°C)	17,3	19,1	14,9	19,1	
8						
9						
10						
11						
12						
13						

**Seznam pojmov vrtilne tabele**

Povlecite polja med spodnjimi območji:

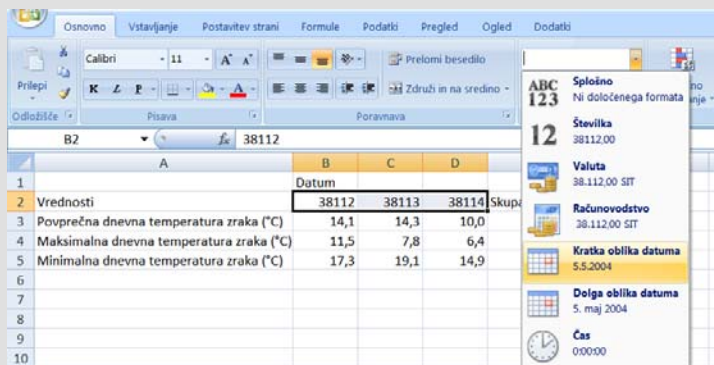
Filter poročila    Oznake stolpcev

Oznake vrstic    Σ Vrednosti

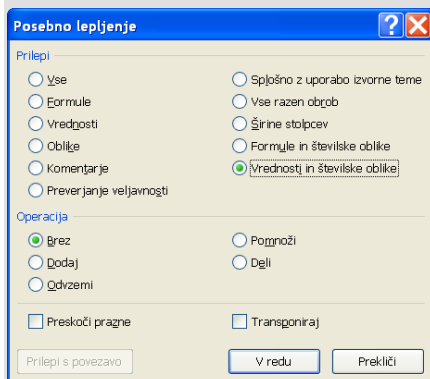
Σ Vrednosti    Povprečje od Temp...  
Min od Temperatur...  
Maks od Temperatur...

Odloži posodobitev postavitev    Posodobiti

**Navodilo:** ob parkiranju vrtilne tabele se vrednosti za datume izpišejo kot zaporedne številke od 1. 1. 1900 dalje. Številke spremenimo nazaj v datum z nastavitvijo oblike v skupini ukazov Število na kartici *Osnovno*:



Če pri parkiranju vrtilne tabele v pogovornem oknu **Posebno lepljenje** namesto možnosti *Vrednosti* izberemo *Vrednosti in številске oblike*, se oblika datuma in število decimalnih mest ne spremeni:



## METEOROLOŠKI PODATKI 8

Za vsak dan in za celotno v tridnevno obdobje izračunajte urno povprečje temperature zraka.

**Navodilo:** podatke razvrstite v skupine po *Datumu* in *Času* ter za vsako skupino izračunajte povprečje temperature zraka. Nato vrednosti spremenljivke *Čas* združite v razrede po urah. V pogovornem oknu **Združevanje v skupine** Excel samodejno napiše v polje *Začni pri* vrednost *0:00:00*, v polje *Končaj pri* pa vrednost v obliki datuma *1.1.1900*, ki številsko pomeni 0, torej računa od polnoči do polnoči. Iste rezultate dobimo, če v polje *Končaj pri* napišemo vrednost *23:55:00*.

## 4 Podatkovna zbirka in vrtilne tabele

The screenshot shows an Excel spreadsheet with a pivot table. The pivot table is based on the following data:

Čas	5.5.04	6.5.04	7.5.04	Skupna vsota
0:00	12,8	13,8	8,5	11,7
0:05	12,9	13,8	8,3	11,7
0:10				
0:15				
0:20				
0:25				
0:30				
0:35				
0:40				
0:45				
0:50				
0:55				
1:00				
1:05				
1:10				
1:15				
1:20				
1:25				

The 'Združevanje v skupine' dialog box is open, showing the following settings:

- Samodejno:  Začni pri: 0:00:00
- Končaj pri: 1.1.1900
- Skupine po: Ura
- Število dni: 1

The 'Seznam polj vrtilne tabele' task pane on the right shows the following fields:

- Filter poročila: Datum
- Oznake stolpcev: Povprečje od Tempera...
- Oznake vrstic: Čas
- Vrednosti: Povprečje od Tempera...

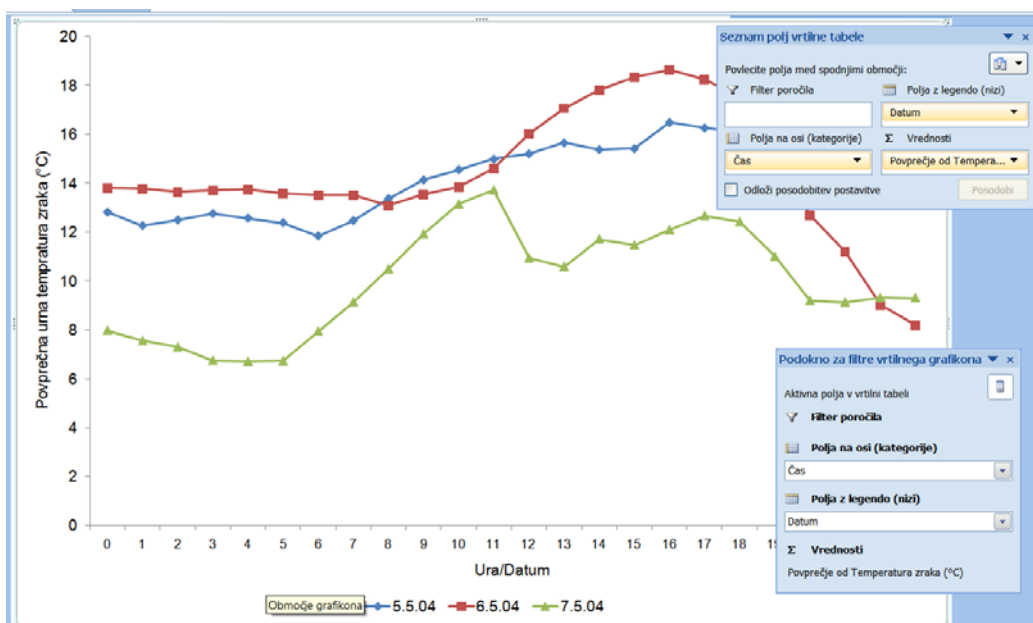
Po izvedenem ukazu **Združi izbor** so v vrtilni tabeli urna povprečja izračunana na podlagi 12 meritev, npr. povprečje ob času 0 je izračunano na podlagi meritev v časih 0:00, 0:05, 0:10,...,0:55, kar pomeni, da je to povprečna temperatura zraka v času od polnoči pa do 1:00. ♦

	A	B	C	D	E
1					
2					
3	Povprečje od Temperatura zraka (°C)	Datum			
4	Čas	5.5.04	6.5.04	7.5.04	Skupna vsota
5	0	12,8	13,8	8,0	11,5
6	1	12,3	13,8	7,6	11,2
7	2	12,5	13,6	7,3	11,1
8	3	12,8	13,7	6,8	11,1
9	4	12,6	13,7	6,7	11,0
10	5	12,4	13,6	6,7	10,9
11	6	11,8	13,5	8,0	11,1
12	7	12,5	13,5	9,1	11,7
13	8	13,4	13,1	10,5	12,3
14	9	14,1	13,6	11,9	13,2
15	10	14,6	13,8	13,2	13,8
16	11	15,0	14,6	13,7	14,4
17	12	15,2	16,0	10,9	14,1
18	13	15,6	17,1	10,6	14,4
19	14	15,4	17,8	11,7	15,0
20	15	15,4	18,3	11,5	15,1
21	16	16,5	18,6	12,1	15,7
22	17	16,3	18,2	12,7	15,7
23	18	16,1	17,5	12,4	15,3
24	19	15,5	15,0	11,0	13,8
25	20	14,9	12,7	9,2	12,3
26	21	14,1	11,2	9,1	11,5
27	22	13,8	9,0	9,3	10,7
28	23	13,7	8,2	9,3	10,4
29	Skupna vsota	14,1	14,3	10,0	12,8

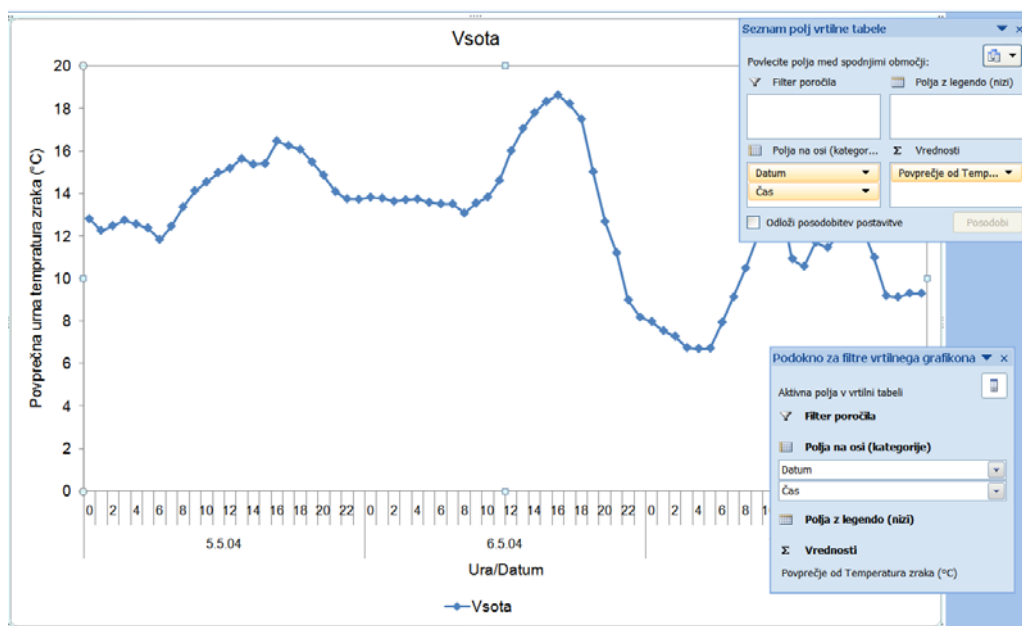
### METEOROLOŠKI PODATKI 9

Povprečne urne temperature zraka za tridnevno obdobje grafično prikažite z uporabo **Vrtilnega grafikona**.

**Navodilo:** ob aktivni vrtilni tabeli uporabimo ukaz **Vrtilni grafikon** na kartici **Možnosti** v skupini Orodja in izberemo vrsto grafikona **Vrstični (Črtni)**. Na grafikonu oblikujemo naslova vodoravne in navpične osi:



Če v **Seznamu polj vrtilne tabele** spremenimo postavitev, se spremeni vrtilna tabela in tudi vrtilni grafikon. Postavite spremenljivki **Datum** in **Čas** v območje **Polja na osi (kategorije)** oz. **Oznake vrstic**. Oblikujte sliko do konca.



## METEOROLOŠKI PODATKI 10

Med podatki imamo tri različne vrednosti za hitrost vetra: maksimalna hitrost vetra, minimalna hitrost vetra in povprečna hitrost vetra za vsak 5-minutni interval. Na podlagi teh podatkov izračunajte minimalno, maksimalno in povprečno dnevno hitrost vetra. ♦

	A	B	C	D	E
4		Datum			
5	<b>Vrednosti</b>	<b>5. 5. 2004</b>	<b>6. 5. 2004</b>	<b>7. 5. 2004</b>	<b>Skupna vsota</b>
6	Min od Min. hitrost vetra (m/s)	0,0	0,0	0,0	0,0
7	Maks od Max. hitrost vetra (m/s)	3,1	7,6	6,2	7,6
8	Povprečje od Hitrost vetra (m/s)	0,2	1,4	0,5	0,7
9					
10					
11					
12					
13					
14					
15					
16					

## METEOROLOŠKI PODATKI 11

Z vrtilno tabelo za vsak dan izračunajte število meritev ter odstotek meritev hitrosti vetra v brezveterju (hitrost vetra 0 m/s).

**Navodilo:** V območju *Oznake vrstic* vrednosti hitrosti vetra, ki so različne od 0 m/s, združimo v »Skupina1« tako, da jih najprej označimo in uporabimo ukaz **Združi izbor**:

	A	B	C	D	E	F
3	<b>Štej od Hitrost vetra (m/s)</b>	Datum				
4	<b>Hitrost vetra (m/s)2</b>	<b>Hitrost vetra (m/s)</b>	<b>5. 5. 2004</b>	<b>6. 5. 2004</b>	<b>7. 5. 2004</b>	<b>Skupna vsota</b>
5		0				
6		0,0	161	84	93	338
7	<b>Skupina1</b>					
8		0,1	19	10	31	60
9		0,3	31	8	23	62
10		0,4	24	11	19	54
11		0,6	15	14	21	50
12		0,7	19	8	20	47
13		0,8	10	16	19	45
14		1,0	2	10	25	37
15		1,1	5	7	13	25

Če odstranimo gumb Hitrost vetra (m/s) iz območja *Oznake vrstic*, vrtilna tabela izgleda takole:

	A	B	C	D	E	F	G
3	Štej od Hitrost vetra (m/s)	Datum					
4	Hitrost vetra (m/s) <sup>2</sup>	5. 5. 2004	6. 5. 2004	7. 5. 2004	Skupna vsota		
5		0	161	84	93	338	
6	Skupina1	127	204	195	526		
7	Skupna vsota	288	288	288	864		
8							
9							
10							
11							
12							

Z ustrezno nastavitvijo polja vrednosti v vrtilni tabeli dobimo strukturo meritev po hitrosti vetra, izraženo v %:

	A	B	C	D	E	F	G
3	Štej od Hitrost vetra (m/s)	Datum					
4	Hitrost vetra (m/s) <sup>2</sup>	5. 5. 2004	6. 5. 2004	7. 5. 2004	Skupna vsota		
5		0	55,9%	29,2%	32,3%	39,1%	
6	Skupina1	44,1%	70,8%	67,7%	60,9%		
7	Skupna vsota	100,0%	100,0%	100,0%	100,0%		
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							

Brezveterje je bilo 5. 5. 2004 izmerjeno v 55,9 % meritev, 6. 5. 2004 v 29,2 % in 7. 5. 2004 v 32,3 % meritev. Vsak dan je bilo izvedenih 288 meritev. ♦

#### METEOROLOŠKI PODATKI 12

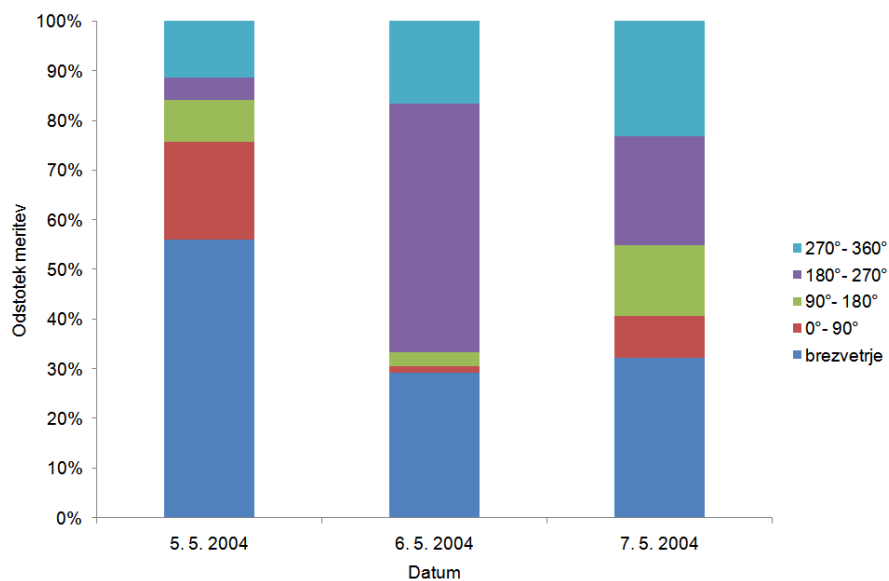
Oblikujte vrtilno tabelo, ki za vsakega od izbranih dni podaja strukturo vetra po njegovi smeri. Smeri vetra naj bodo združene v štiri razrede širine 90°: 0° do pod 90°, 90° do pod 180°, 180° do pod 270° in 270° do pod 360°. Porazdelitev grafično prikažite s tremi strukturnimi stolpci (slika 4.8).

**Navodilo:** Po spremenljivki *Smer vetra* podatkov ne moremo združiti v razrede, dokler je med vrednostmi tudi vrednost »n« (nedoločljivo). Zato na delovnem listu »izbor« najprej vrednosti »n« zamenjajte s praznimi celicami, nato vrtilno tabelo osvežite in uporabite ukaz **Združi izbor**.

## 4 Podatkovna zbirka in vrtilne tabele

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4	Štej od Datum	Datum									
5	Smer vetra (°)	5. 5. 2004	6. 5. 2004	7. 5. 2004	Skupna vsota						
6	<0 ali (prazen)	55,9%	29,2%	32,3%	39,1%						
7	0-90	19,8%	1,4%	8,3%	9,8%						
8	90-180	8,3%	2,8%	14,2%	8,4%						
9	180-270	4,5%	50,0%	21,9%	25,5%						
10	270-360	11,5%	16,7%	23,3%	17,1%						
11	Skupna vsota	100,0%	100,0%	100,0%	100,0%						

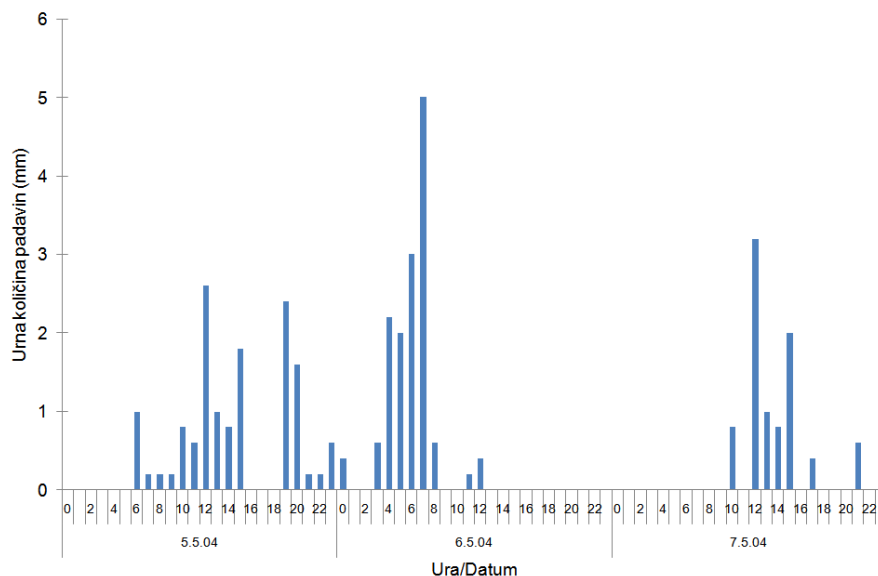
Za oblikovanje grafikona je potrebno v vrtilni tabeli zamenjati vrstice in stolpce: *Smer vetra* gre v območje *Oznake stolpcev* oz *Polja z legendo (Nizi)*, *Datum* pa v območje *Oznake vrstic* oz *Polja na osi (kategorije)*. ♦



Slika 4.8: Struktura smeri vetra v treh dnevih meritev.

### METEOROLOŠKI PODATKI 13

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



Slika 4.9: Urne vsote padavin v danem obdobju.

#### METEOROLOŠKI PODATKI 14

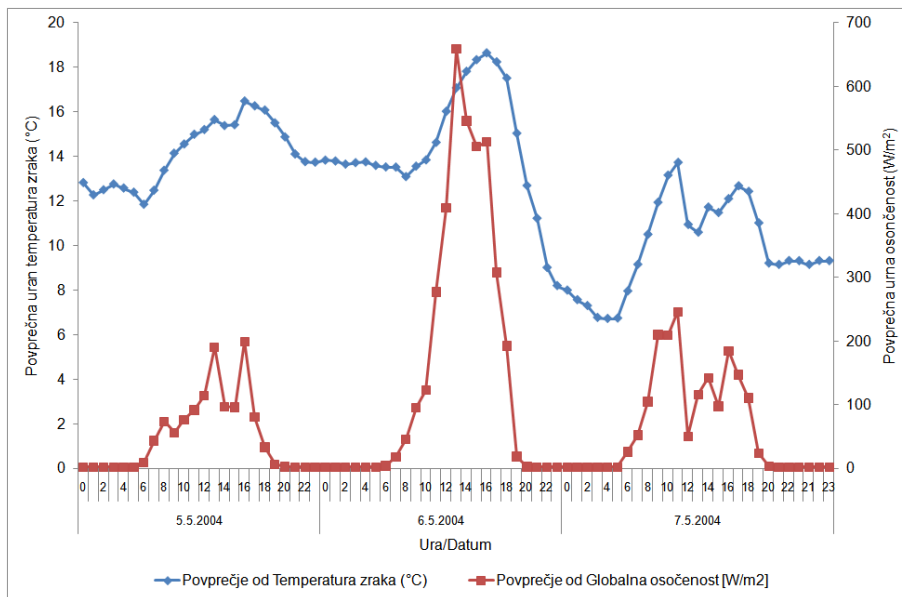
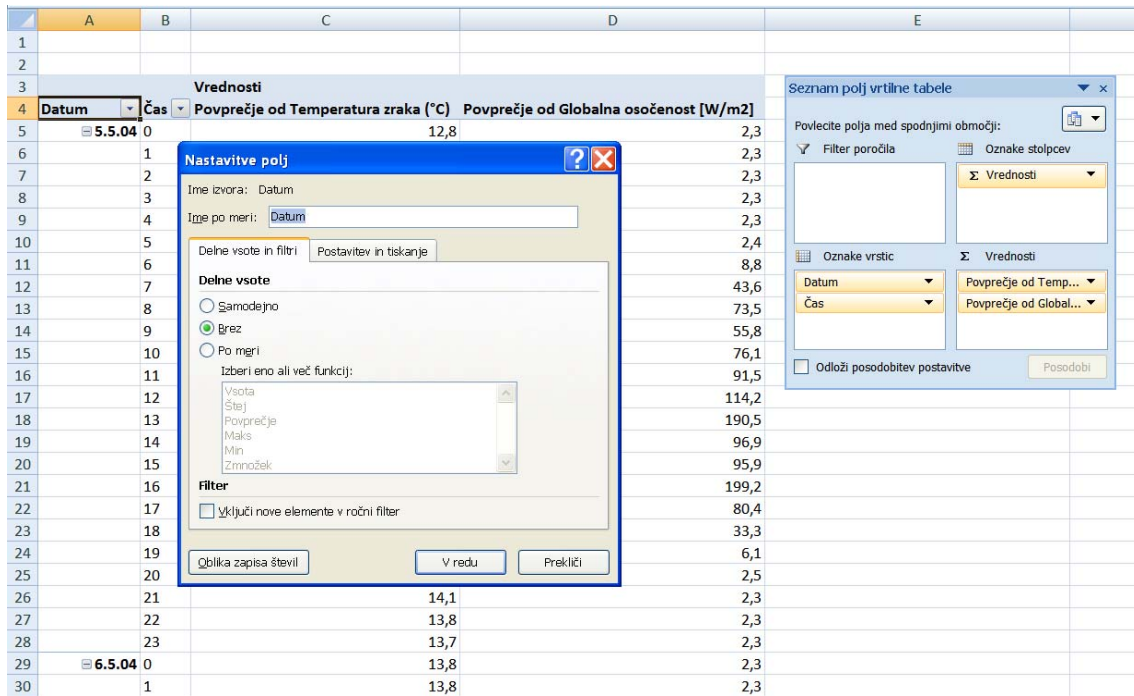
Izračunajte povprečno globalno urno osončenost ( $W/m^2$ ) in povprečno urno temperaturo zraka ( $^{\circ}C$ ) po dnevih v obdobju 5. - 7. 5. 2004 ter ju grafično prikažite na isti sliki (slika 4.10).

**Navodilo:** za grafični prikaz najprej oblikujte vrtilni grafikon. Skalo za povprečno urno osončenost prikažite na sekundarni navpišni osi.

Če grafikon oblikujemo na podlagi parkirane vrtilne tabele, je priročno, da 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 (vrtilni grafikon jih samodejno izpusti). V vrtilni tabeli se tem vrsticam reče *Delne vsote* in se jih lahko znebimo z ukazom **Nastavitve polj** (za datum *Datum*), kjer v razdelku *Delne vsote* izberemo možnost **Brez**:



## 4 Podatkovna zbirka in vrtilne tabele



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

## METEOROLOŠKI PODATKI 15

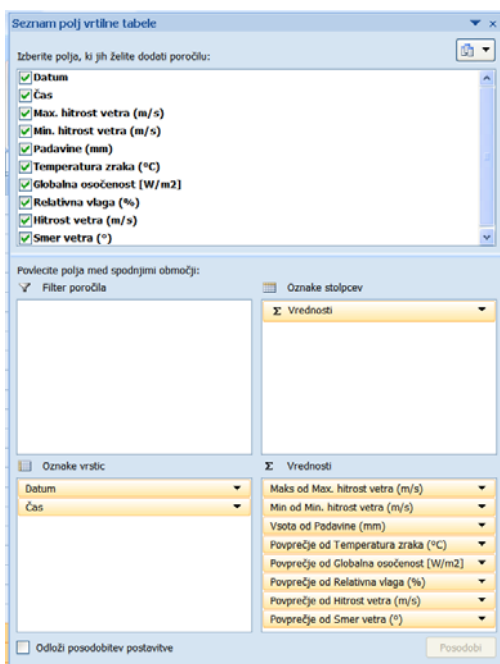
Urne vsote padavin in urno povprečje temperature zraka ustrezno grafično prikažite na istem grafikonu.

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

## METEOROLOŠKI PODATKI 16

Na podlagi vseh podatkov na delovnem listu »podatki« naredite novo podatkovno zbirko, ki bo namesto pet minutnih vrednosti vsebovala urne vrednosti meteoroloških spremenljivk. Delovni list z novo podatkovno zbirko poimenujte »urne vrednosti«.

**Navodilo:** ustrezne urne vrednosti meteoroloških spremenljivk izračunajte z vrtilno tabelo:



V vrtilni tabeli nato z ukazom **Nastavitve polj** odstranite **Delne vsote**, ki predstavljajo dnevne vrednosti meteoroloških spremenljivk:

## 4 Podatkovna zbirka in vrtilne tabele

	A	B	C
1			
2			
3			
4		Vrednosti	
5	Oznake vrstic	Maks od Max. hitrost vetra (m/s)	Min od Min. hitrost vetra (m/s)
6	14. 4. 2004	3,500980392	0
7	0	0,98027451	0
8	1	0,840235294	0
9			
10			
11			
12			
13			
14			
15			0,420117647
16			0,420117647
17			
18			
19			0,420117647
20			
21			
22			
23			
24			
25			
26			
27			
28			
29	23	1,120313725	0
30	15. 4. 2004	2,800784314	0

**Nastavitve polj**

Ime izvora: Datum  
Izpis po meri: Datum

Delne vsote in filtri: Postavitve in tiskanje

Delne vsote

Samodejno  
 Prez  
 Po meri

Izberi eno ali več funkcij:

- Vsota
- Slaj
- Povprečje
- Maks
- Min
- Zmnožnik

Filter

Vključi nove elemente v ročni filter

Oblika zapisa števil: [ ] V redu Prekliči

Z odstranitvijo delnih vsot samo pobrišemo vsebino celic, vrstice, namenjene delnim vsotam pa ostanejo med podatki. Znebiti se jih s spremembo **Postavitve poročila** vrtilne tabele na kartici **Načrt**, izbrati moramo možnost **Pokaži v obliki tabele**:

Datum	Cas	maks od max. hitrost vetra (m/s)	Min od Min. hitrost vetra (m/s)	Vsota od P
14. 4. 2004	0	0,98027451	0	
	1	0,840235294	0	
	2	0,280078431	0	
	3	1,680470588	0	
	4	1,120313725	0	
	5	0	0	
	6	0,700196078	0	
	7	0,98027451	0	
	8	1,680470588	0	
	9	2,520705882	0	
	10	3,220901961	0,420117647	
	11	3,500980392	0,420117647	
	12	2,660745098	0	
	13	2,800784314	0	
	14	3,220901961	0,420117647	
	15	2,380666667	0	
	16	2,100588235	0	
	17	1,680470588	0	
	18	1,540431373	0	
	19	2,660745098	0	
	20	2,240627451	0	
	21	0,840235294	0	
	22	1,120313725	0	
	23	1,120313725	0	
15. 4. 2004	0	0,700196078	0	
	1	0,140039216	0	

V naslednjem koraku vrtilno tabelo parkiramo na nov delovni list »urne vrednosti«. Pri parkiranju podatkovno zbirko primerno uredimo:

Spremenimo imena spremenljivk (imena spremenljivk naj bodo kar enaka kot na delovnem listu »podatki«, v novi podatkovni zbirki gre za urne, v stari pa za pet-minutne vrednosti).

Vrednosti #DEL/0! pri spremenljivki Smer vetra zamenjamo z znakom »n« (pomagajte si s **Filtrom**).

Datum je v novi podatkovni zbirki napisan samo v prvi vrstici podatkov določenega dne, če želimo narediti podatkovno zbirko popolno, je treba vrednosti dopolniti »pešč«.

Pri spremenljivki Čas se pojavijo v levem zgornjem kotu celic zeleni trikotniki, ki opozarjajo, da se je pri kopiranju število shranilo kot besedilo. Čas spremenimo nazaj v številsko spremenljivko tako, da vse vrednosti označimo, kliknemo ikono s klicajem in izberemo ukaz **Pretvori v število**.

Če tega še nismo naredili prej, izbrišemo še zadnjo vrstico »Skupna vsota«.

Izračunane urne vrednosti meteoroloških spremenljivk na koncu oblikujemo na ustrezno število decimalnih mest.

	A	B	C	D	E	F	G	H	I
1	Datum	Čas	Max. hitrost vetra (m/	Min. hitrost vetra (m/	Padavine (mr	Temperatura zraka (°	Globalna osičenost [W/m	Relativna vlaga (%)	Hitrost vetra
2	14. 4.	0	1,0	0,0	0,0	6,2	2,1	95,7	
3	14. 4.		0,8	0,0	0,0	6,1	2,1	96,6	
4	14. 4.		0,3	0,0	0,4	6,1	2,1	95,5	
5	14. 4.		1,7	0,0	0,4	6,1	2,1	95,0	
6	14. 4.		1,1	0,0	0,8	5,9	2,1	96,0	
7	14. 4.		0,0	0,0	0,4	5,9	2,2	96,1	
8	14. 4.		0,7	0,0	0,8	5,9	2,3	97,1	
9	14. 4.		1,0	0,0	0,8	6,0	7,1	96,9	
10	14. 4. 2004	8	1,7	0,0	1,0	6,3	18,5	96,5	
11	14. 4. 2004	9	2,5	0,0	0,4	6,6	50,9	94,2	
12	14. 4. 2004	10	3,2	0,4	0,0	7,2	91,0	89,4	
13	14. 4. 2004	11	3,5	0,4	0,0	8,4	204,5	80,9	
14	14. 4. 2004	12	2,7	0,0	0,0	10,0	237,7	74,2	
15	14. 4. 2004	13	2,8	0,0	0,0	10,5	212,1	71,0	
16	14. 4. 2004	14	3,2	0,4	0,0	10,0	112,0	69,1	
17	14. 4. 2004	15	2,4	0,0	0,0	10,2	119,0	68,5	
18	14. 4. 2004	16	2,1	0,0	0,0	10,2	95,7	69,9	
19	14. 4. 2004	17	1,7	0,0	0,0	10,7	113,3	68,8	
20	14. 4. 2004	18	1,5	0,0	0,0	10,8	73,3	69,0	
21	14. 4. 2004	19	2,7	0,0	0,0	10,1	10,2	75,1	
22	14. 4. 2004	20	2,2	0,0	0,0	9,2	2,2	77,9	
23	14. 4. 2004	21	0,8	0,0	0,0	8,6	2,1	81,6	
24	14. 4. 2004	22	1,1	0,0	0,0	7,8	2,2	89,6	
25	14. 4. 2004	23	1,1	0,0	0,0	7,1	2,2	95,0	
26	15. 4. 2004	0	0,7	0,0	0,0	7,0	2,2	95,2	
27	15. 4. 2004	1	0,1	0,0	0,0	7,0	2,2	95,0	
28	15. 4. 2004	2	1,0	0,0	0,0	6,8	2,2	95,5	
29	15. 4. 2004	3	0,8	0,0	0,0	6,5	2,2	95,4	

## 5 Normalna in binomska porazdelitev

### Novi pojmi:

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

### Datoteke s podatki:

PORAZDELITVE.xlsx

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.

### 5.1 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.

### 5.1.1 Gostota verjetnosti

NORMALNA 1

Odprite datoteko PORAZDELITVE.xlsx. ♦

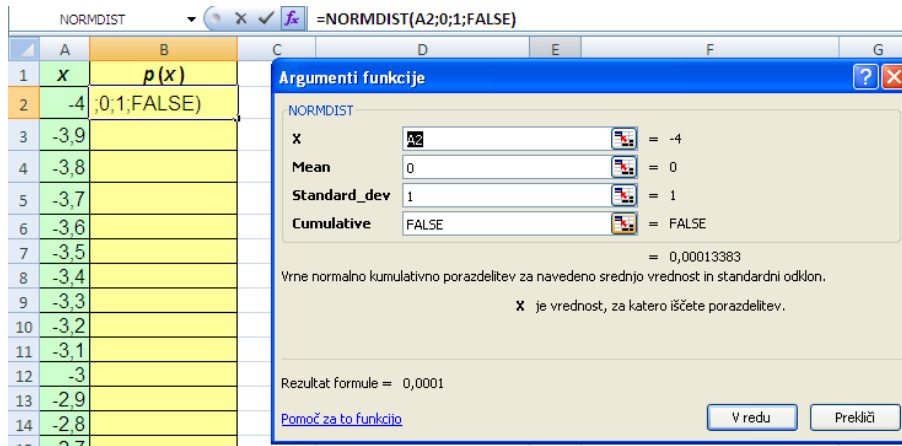
NORMALNA 2

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.

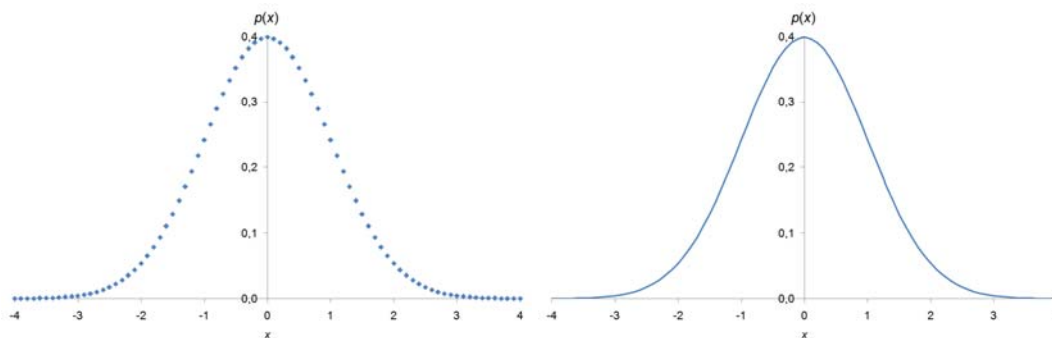
Z uporabo funkcije NORMDIST v sosednji stolpec izračunajmo gostoto verjetnosti  $p(x)$  standardizirane normalne porazdelitve  $N(0, 1)$  za vrednosti  $x$  med -4,0 in 4,0. Za argumente funkcije uporabimo: NORMDIST(referenca celice, v kateri je podatek za  $x$ ; 0; 1; FALSE).

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



NORMALNA 3

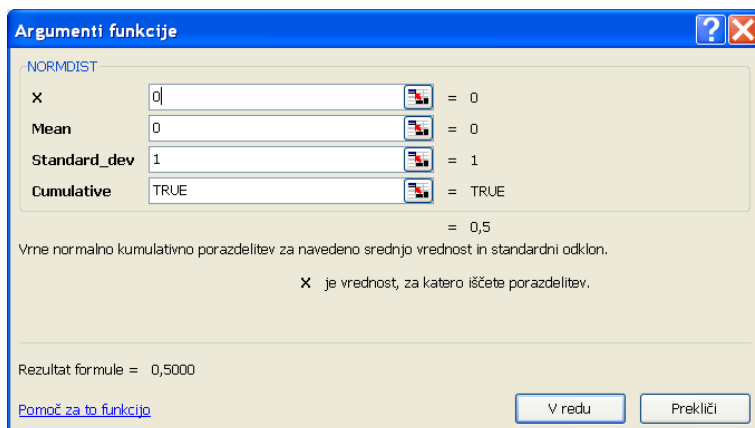
Grafično prikazite izračunane vrednosti gostote verjetnosti  $N(0,1)$  - vrsta grafikona *Raztreseni* (slika 5.1 levo). Točke povežite z daljicami in odstranite predstavitelje točk, da dobite sliko 5.1 (desno). ♦



Slika 5.1: 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).

5.1.2 Izračun verjetnosti

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



#### NORMALNA 4

Za standardizirano normalno porazdelitev z uporabo funkcije  $\text{NORMDIST}(x; 0; 1; \text{TRUE})$  izračunajte verjetnosti:

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

**Navdilo:** na delovnem listu »Tabela 5.1« izračunajte zgornje izraze. Končni rezultati so v celicah vijolične barve. ♦

Tabela 5.1: Izpis formul v tabeli za izračun verjetnosti standardizirane normalne porazdelitve (zgoraj) in rezultati (spodaj)<sup>20</sup>

	A	B	C	D	E
1	Verjetnosti	$x_1$	$P(X < x)$	$P(X > x)$	$P(x_1 < X < x_2)$
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)		=C3-C4
5	$P(-2 < X < 2)$	-2	=NORMDIST(B5;0;1;TRUE)		=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)		=C8-C9

	A	B	C	D	E
1	Verjetnosti	$x_1$	$P(X < x)$	$P(X > x)$	$P(x_1 < X < x_2)$
2	$P(X < 0)$	0	0,5000		
3	$P(X > 1)$	1	0,8413	0,1587	
4	$P(-1 < X < 1)$	-1	0,1587		0,6827
5	$P(-2 < X < 2)$	-2	0,0228		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		0,9153

<sup>20</sup> Pri zveznih porazdelitvah je  $P(X \leq x) = P(X < x)$ . V Excelu zaradi enostavnosti zapisa pišemo znak <.

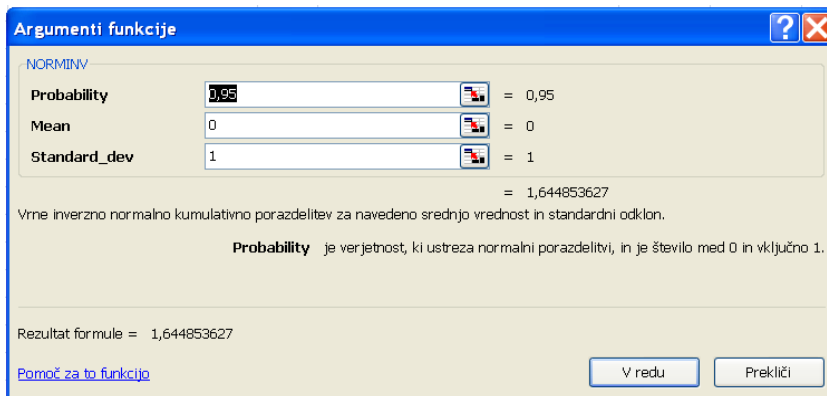


### 5.1.3 Izračun kvantilov

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 5

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 6

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

#### NORMALNA 7

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  $P(-x_4 \leq X \leq +x_4) = 0,99$ . ♦

	D	E	F
1	<b>verjetnost</b>		<b>vrednost spremenljivke X</b>
2	$P(X < x_1) = 0,95$	$x_1 =$	<code>=NORMINV(0,95;0;1)</code>
3	$P(X > x_2) = 0,25$	$x_2 =$	<code>=NORMINV(0,75;0;1)</code>
4	$P(-x_3 < X < x_3) = 0,95$	$x_3 =$	<code>=NORMINV(0,025;0;1)</code>
5	$P(-x_4 < X < x_4) = 0,99$	$x_4 =$	<code>=NORMINV(0,995;0;1)</code>

	D	E	F
1	<b>verjetnost</b>		<b>vrednost spremenljivke X</b>
2	$P(X < x_1) = 0,95$	$x_1 =$	1,6449
3	$P(X > x_2) = 0,25$	$x_2 =$	0,6745
4	$P(-x_3 < X < x_3) = 0,95$	$x_3 =$	1,9600
5	$P(-x_4 < X < x_4) = 0,99$	$x_4 =$	2,5758

**Zgled 5.1.1: 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 5.2, 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

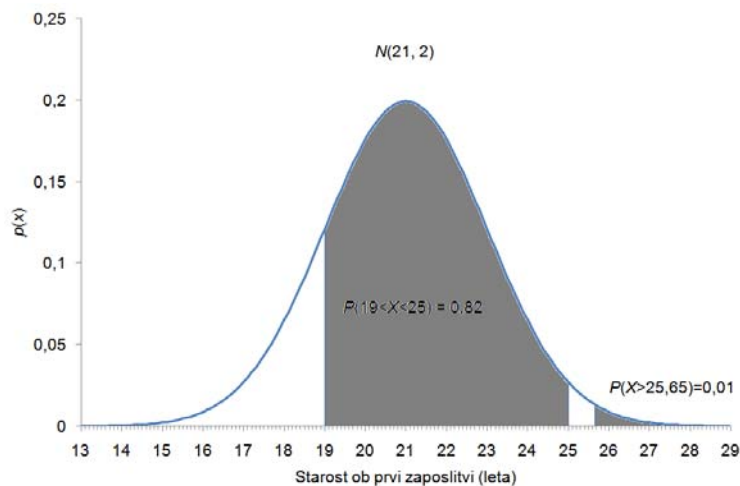
Izračunajte verjetnost, da je starost ob prvi zaposlitvi med 19 in 25 let;  $P(19 \leq X \leq 25)$  (slika 5.2). ♦

## STAROST 3

Izračunajte starost ob prvi zaposlitvi, za katero velja, da je prekoračena pri 1 % oseb (slika 5.2). Kateri centil ste izračunali? ♦

## STAROST 4

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

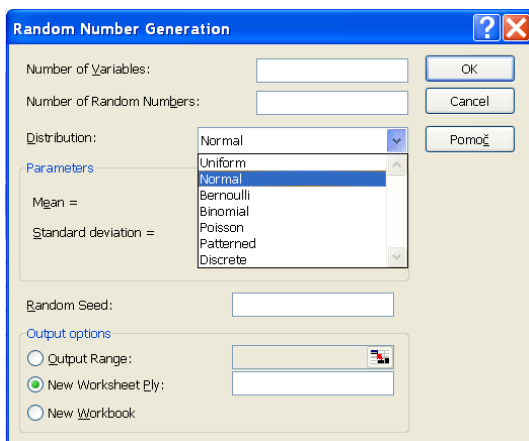


Slika 5.2: Normalna porazdelitev s parametroma  $\mu = 21$  let,  $\sigma = 2$  leti za starost ob prvi zaposlitvi. Svetleje osončena je prikazana verjetnost, da je oseba ob prvi zaposlitvi stara med 19 in 25 let, temneje osončena pa je prikazana verjetnost (njena vrednost je 0,01), da je oseba ob prvi zaposlitvi stara 25,65 let ali več.

## 5.2 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 **Analiza podatkov/Random number generation** na kartici **Podatki**. 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.<sup>21</sup> To polje lahko pustimo prazno in tedaj vedno dobimo druga slučajna števila.



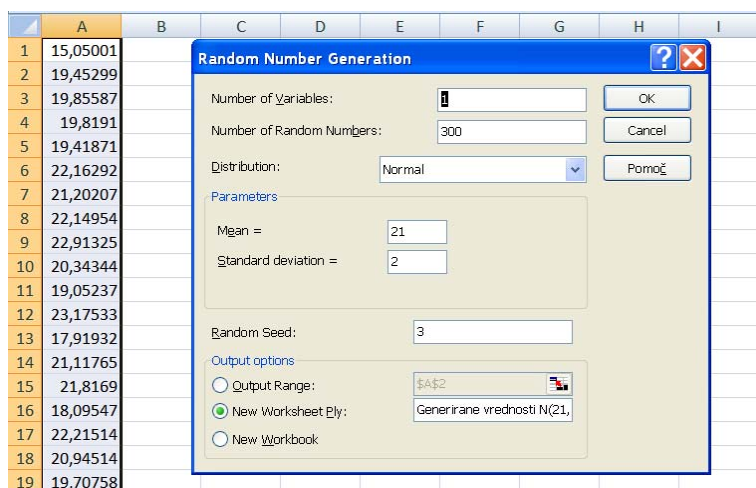
### 5.2.1 Generiranje vrednosti, porazdeljenih po normalni porazdelitvi

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

<sup>21</sup> Taka števila imenujemo psevdoslučajna števila.

## NORMALNA 8

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 N(21, 2)«. ♦



## NORMALNA 9

Izračunajte povprečje in standardni odklon generiranih vrednosti. Primerjete ju s parametroma porazdelitve. ♦

## NORMALNA 10

Generirane vrednosti razporedite v razrede širine 2 s pomočjo vrtlne tabele in narišite histogram (slika 5.3). Kakšna je oblika empirične porazdelitve?

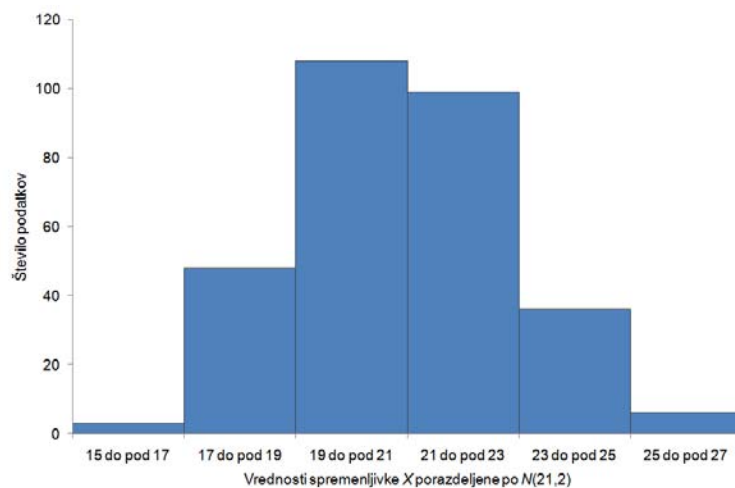
**Navodilo:** vrinite eno vrstico na začetek generiranih vrednosti in v celico A1 napišite ime vrednosti slučajne spremenljivke (npr.  $x$ ) in oblikujte vrtlno tabelo. Vrednosti spremenljivke  $x$  združimo v razrede z ukazom **Združi izbor**, pri tem spremenimo vrednosti v poljih **Začni pri** in **Končaj pri** na cela števila (zaokrožimo navzdol oziroma navzgor):

## 5 Normalna in binomska porazdelitev

The screenshot shows an Excel spreadsheet with data in column A and a pivot table in column B. Two dialog boxes are open:

- Združevanje v skupine (Grouping):** Shows options for grouping. 'Začni pri' (Start at) is 15, 'Končaj pri' (End at) is 27, and 'Po' (By) is 2. Buttons: 'V redu' (OK), 'Prekliči' (Cancel).
- Seznam polj vrtilne tabele (PivotTable Fields):** Shows the list of fields. 'x' is selected. Below, 'Filter poročila' (Filter reports) and 'Oznake stolpcev' (Column labels) are selected. 'Oznake vrstic' (Row labels) is 'x' and 'Vrednosti' (Values) is 'Štej od x'.

	A	B	C
1			
2			
3	x	Štej od x	
4	15-17	3	
5	17-19	48	
6	19-21	108	
7	21-23	99	
8	23-25	36	
9	25-27	6	
10	<b>Skupna vsota</b>	<b>300</b>	
11			



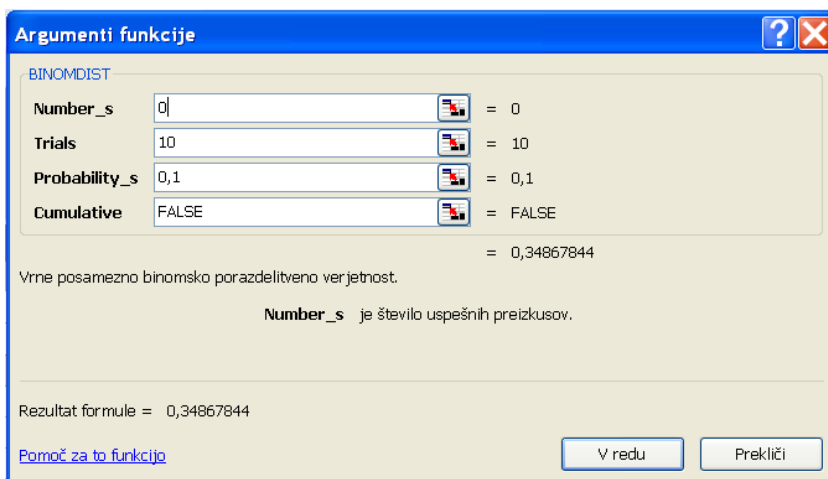
Slika 5.3: Histogram za 300 generiranih vrednosti spremenljivke  $X$  porazdeljene po normalni porazdelitvi  $N(21, 2)$

### 5.3 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  $\bar{A}$ ;  $p$  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.

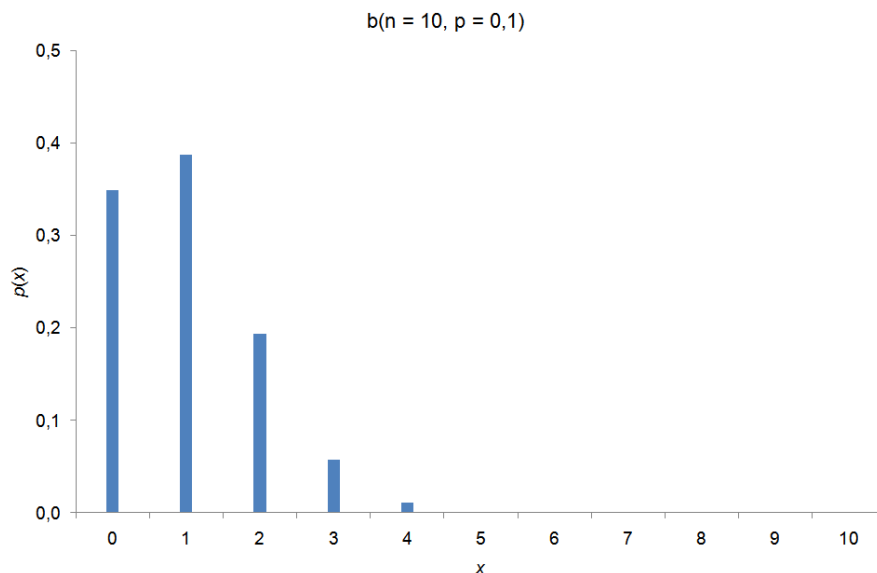


BINOMSKA 1

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 5.4. ♦

	A	B
1	<b>b(n = 10, p)</b>	
2	<b>x</b>	<b>p = 0,1</b>
3	0	=BINOMDIST(A3;10;0,1;FALSE)
4	1	=BINOMDIST(A4;10;0,1;FALSE)
5	2	=BINOMDIST(A5;10;0,1;FALSE)
6	3	=BINOMDIST(A6;10;0,1;FALSE)
7	4	=BINOMDIST(A7;10;0,1;FALSE)
8	5	=BINOMDIST(A8;10;0,1;FALSE)
9	6	=BINOMDIST(A9;10;0,1;FALSE)
10	7	=BINOMDIST(A10;10;0,1;FALSE)
11	8	=BINOMDIST(A11;10;0,1;FALSE)
12	9	=BINOMDIST(A12;10;0,1;FALSE)
13	10	=BINOMDIST(A13;10;0,1;FALSE)
14	<b>Vsota p(x)</b>	=SUM(B3:B13)

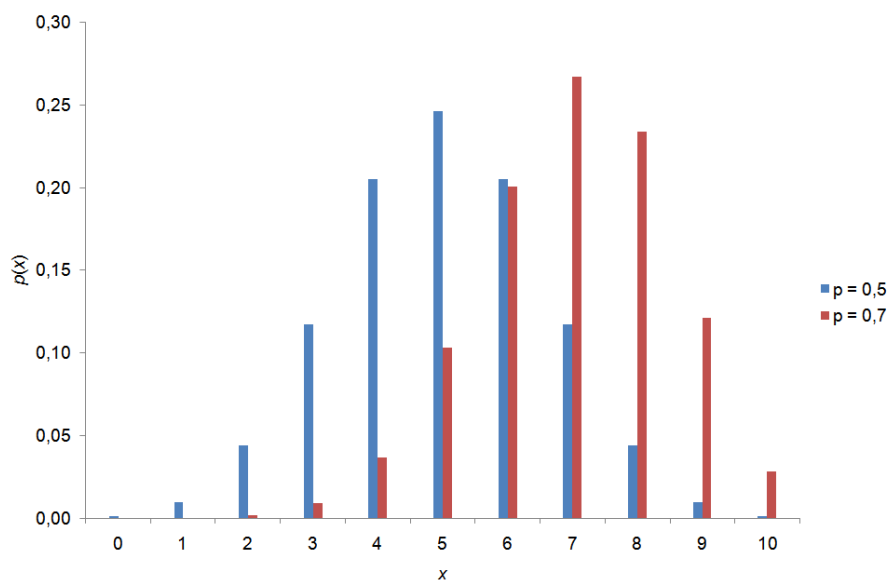
	A	B	C	D	E
1	<b>b(n = 10, p)</b>	<b>p(x)</b>			
2	<b>x</b>	<b>p = 0,1</b>	<b>p = 0,3</b>	<b>p = 0,5</b>	<b>p = 0,7</b>
3	0	0,3487	0,0282	0,0010	0,0000
4	1	0,3874	0,1211	0,0098	0,0001
5	2	0,1937	0,2335	0,0439	0,0014
6	3	0,0574	0,2668	0,1172	0,0090
7	4	0,0112	0,2001	0,2051	0,0368
8	5	0,0015	0,1029	0,2461	0,1029
9	6	0,0001	0,0368	0,2051	0,2001
10	7	0,0000	0,0090	0,1172	0,2668
11	8	0,0000	0,0014	0,0439	0,2335
12	9	0,0000	0,0001	0,0098	0,1211
13	10	0,0000	0,0000	0,0010	0,0282
14	<b>Vsota p(x)</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>



Slika 5.4: Binomska porazdelitev za parametra  $n = 10$  in  $p = 0,1$

## 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 5.5. ♦



Slika 5.5: Binomski porazdelitvi za parametra  $n = 10$  in  $p = 0,5$  ter  $n = 10$  in  $p = 0,7$

## 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. ♦



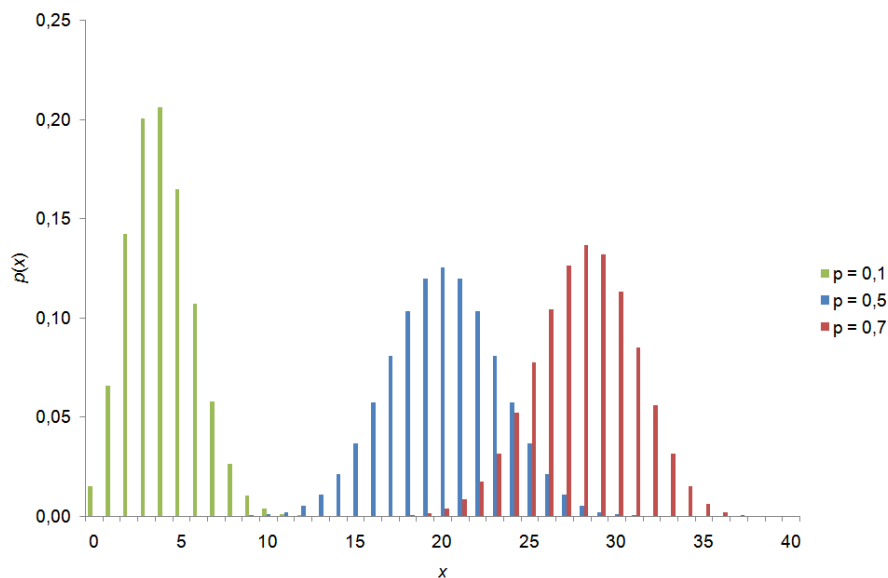
## 5.4 Aproksimacija binomske porazdelitve z normalno porazdelitvijo

Pogoja za aproksimacijo binomske porazdelitve  $b(n, p)$  z normalno porazdelitvijo  $N(np, \sqrt{np(1-p)})$  sta<sup>22</sup>:

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

### 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 5.6. ♦



Slika 5.6: Binomske porazdelitve za parametra  $n=40$  in  $p=0,1$ ;  $n=40$  in  $p=0,5$ ;  $n=40$  in  $p=0,7$

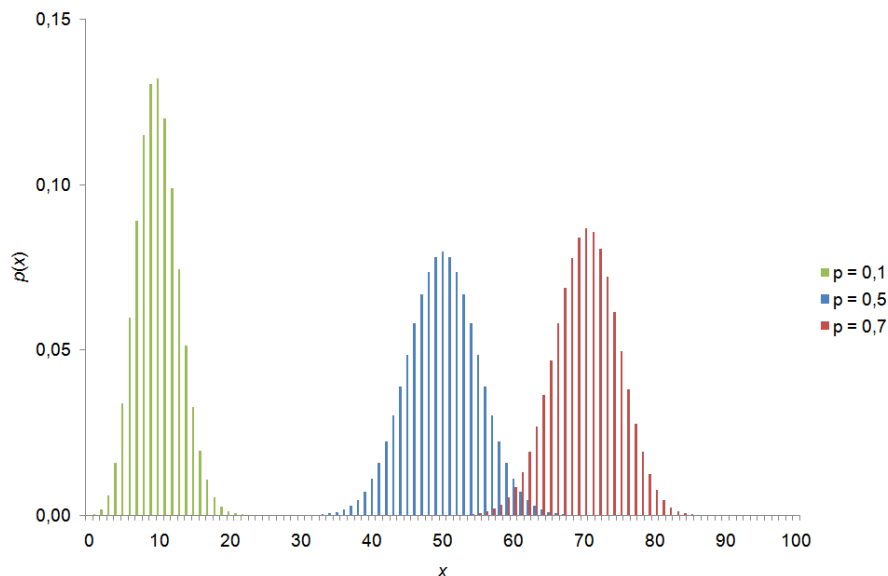
### BINOMSKA 7

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 5.7. ♦

<sup>22</sup> V literaturi za aproksimacijo najdemo tudi drugačne pogoje.



Slika 5.7: 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

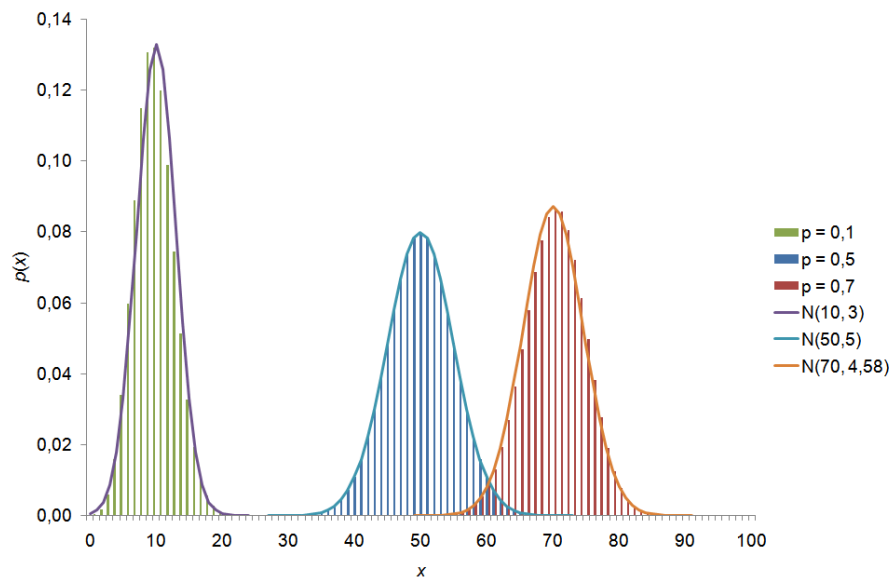
Na sliko 5.7 dodajte še nize za gostote verjetnosti pripadajočih normalnih porazdelitev (slika 5.8).

**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 5.8: Binomske porazdelitve in njihove aproksimacije z normalno porazdelitvijo

## 6 Intervali zaupanja

### Novi pojmi:

- interval zaupanja za povprečje  $\mu$  za normalno porazdelitev  $N(\mu, \sigma)$ ;  $\sigma$  znana,  $\sigma$  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.xlsx

### 6.1 Interval zaupanja za povprečje

#### 6.1.1 $\sigma$ je znana

Pri vnaprej določeni verjetnosti  $\alpha$  takole izračunamo spodnjo mejo ( $l_1$ ) in zgornjo mejo ( $l_2$ ) za  $(1 - \alpha) \cdot 100\%$  interval zaupanja za povprečje v primeru znane  $\sigma$ :

$$l_{1,2} = \bar{x} \pm \Delta,$$

$\Delta = z_{\alpha/2} \cdot \sigma / \sqrt{n}$ ,  $\Delta$  je odklon intervala zaupanja.

#### Zgled 6.1: Interval zaupanja za zgornji krvni tlak pri ženskah

Raziskovalci so na podlagi dolgoletnih podatkov rednih zdravniških pregledov ugotovili, da je za zgornji krvni tlak (spremenljivka  $X$ ) pri ženskah, starih od 25 do 35 let, primerna normalna porazdelitev.

## KRVNI TLAK 1

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  $\sigma$  znana,  $\sigma = 15$  mm.

**Navodilo:** izračune naredimo na delovnem listu »krvni tlak« (stolpec B v tabeli 6.1). ♦

Tabela 6.1: Prikaz dela delovnega lista »krvni tlak«. 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, <b>sigma znana</b>				
3	Zaupanje				
4		0,95	0,99	0,95	0,99
5	n	25	25	250	250
6	$x_{povp}$	123,7			
7	sigma	15			
8	alfa	=1-B4			
9	$z_{alfa/2}$	=NORMINV(1-B8/2;0;1)			
10	delta	=B9*B7/SQRT(B5)			
11	$l_1$	=B6-B10			
12	$l_2$	=B6+B10			
13	$x_{povp}$	=B6			

	A	B	C	D	E
2	Intervali zaupanja za povprečje, <b>sigma znana</b>				
3	Zaupanje				
4		95%	99%	95%	99%
5	n	25	25	250	250
6	$x_{povp}$	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	$z_{alfa/2}$	1,960	2,576	1,960	2,576
10	delta	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	$x_{povp}$	123,7	123,7	123,7	123,7

## KRVNI TLAK 2

Ponovite izračun pri 99 % zaupanju (tabela 6.1, stolpec C). ♦

## KRVNI TLAK 3

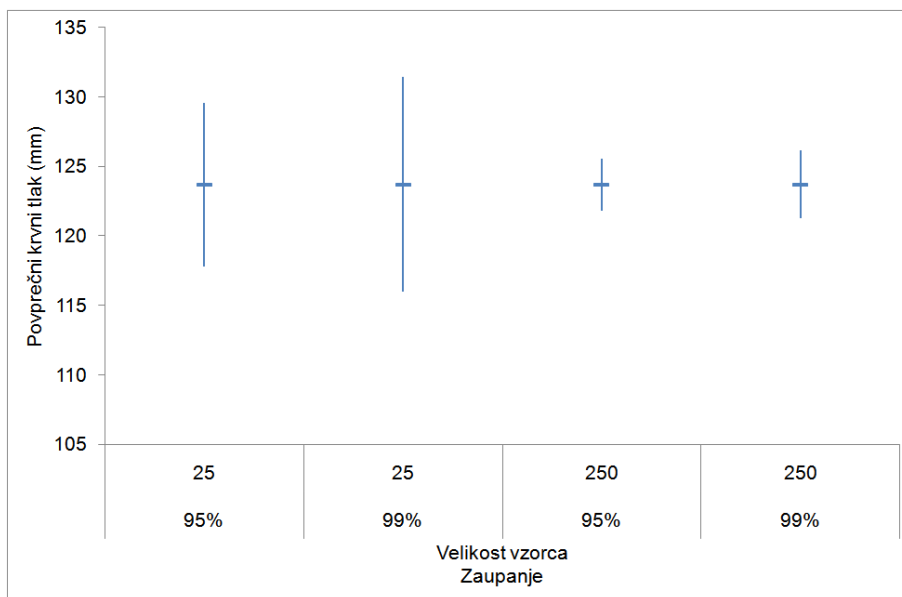
Izračunajte še 95 % in 99 % interval zaupanja za povprečni zgornji krvni tlak ob predpostavki, da je  $\sigma$  znana in da je velikost vzorca  $n = 250$  (tabela 6.1, stolpca D in E). ♦

## KRVNI TLAK 4

Izračunane intervale zaupanja grafično prikažite, kot kaže slika 6.1.

**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  $x_{povp}$ . 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. Spodnji konci črt predstavljajo spodnje meje IZ  $l_1$  in zgornji konci črt predstavljajo zgornje meje IZ  $l_2$ . Vzorčna povprečja so na sredini teh črt po osnovni nastavitvi označena s kratko vodoravno črtico.

Če želimo oznako 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 posamezne oznake, po izbiri niza podatkov z levim gumbom miške ponovno kliknemo predstavitelja in na pomožnem meniju dobimo ukaz **Oblikuj podatkovno točko**. ♦



Slika 6.1: Grafični prikaz različnih intervalov zaupanja za povprečni zgornji krvni tlak žensk, starih od 25 do 35 let, ob predpostavki, da je sigma znana

### 6.1.2 $\sigma$ ni znana

Pri vnaprej določeni verjetnosti  $\alpha$  takole izračunamo spodnjo mejo ( $l_1$ ) in zgornjo mejo ( $l_2$ ) za  $(1 - \alpha) \cdot 100\%$  interval zaupanja za povprečje v primeru neznanega  $\sigma$ :

$$l_{1,2} = \bar{x} \pm \Delta,$$

odklon intervala zaupanja tokrat izračunamo drugače:  $\Delta = t_{\alpha/2}(SP) \cdot s / \sqrt{n}$ ,  $SP = n - 1$ .

V Excelu izračunamo vrednost  $t_{\alpha/2}(SP)$  s funkcijo `TINV(probability; deg_freedom)`. Funkcija ima dva argumenta: »probability« je verjetnost  $\alpha$  in »deg\_freedom« so stopinje prostosti SP. Opozarjamo, da pri funkciji `TINV` verjetnost  $\alpha$  ne razpolovimo, saj to Excel naredi sam.

### KRVNI TLAK 5

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  $\sigma$  ne poznate. Intervale zaupanja obrazložite.

**Navodilo:** izračune najprej naredite z računalom in uporabo tabel, nato pa še v Excelu v datoteki INTERVALI ZAUPANJA.xlsx na delovnem listu »krvni tlak« (tabela 6.2, stolpec B).

#### **Rešitev:**

Z 95 % verjetnostjo trdimo, da je interval (117,4, 130,0) eden tistih, ki vsebujejo  $\mu$ , 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 6

Ponovite izračun pri 99 % zaupanju (tabela 6.2, stolpec C). ♦

### KRVNI TLAK 7

Izračunajte še 95 % in 99 % interval zaupanja za povprečni zgornji krvni tlak ob predpostavki, da  $\sigma$  ne poznate in da je velikost vzorca  $n = 250$  (tabela 6.2, stolpca D in E). ♦

### KRVNI TLAK 8

Izračunane intervale zaupanja grafično prikažite. ♦

Tabela 6.2: Prikaz dela delovnega lista »krvni tlak«. 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, <b>sigma ni znana</b>				
17	Zaupanje				
18		0,95	0,99	0,95	0,99
19	n	25	25	250	250
20	$x_{povp}$	123,7	123,7	123,7	123,7
21	s	15,2	15,2	15,2	15,2
22	alfa	=1-B18			
23	SP=n-1	=B19-1			
24	$t_{\alpha/2}(SP)$	=TINV(B22;B23)			
25	delta	=B24*B21/SQRT(B19)			
26	$l_1$	=B20-B25			
27	$l_2$	=B20+B25			
28	$x_{povp}$	=B20			

	A	B	C	D	E
16	Intervali zaupanja za povprečje, <b>sigma ni znana</b>				
17	Zaupanje				
18		95%	99%	95%	99%
19	n	25	25	250	250
20	$x_{povp}$	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	delta	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	$x_{povp}$	123,7	123,7	123,7	123,7

## Zgled 6.2: Deset intervalov zaupanja

### INTERVALI ZAUPANJA 1

V datoteki INTERVALI ZAUPANJA.xlsx so na delovnem listu »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:** uporabimo ukaz **Analiza podatkov/Random number generation** na kartici **Podatki**. V pogovornem oknu **Random number Generation** izberemo: **Distribution:** *Normal*, za **Random Seed** napišemo 77, če želimo dobiti iste vrednosti slučajne spremenljivke  $X$ , kot so prikazane v nadaljevanju. ♦

	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										
3	sigma (standardni odklon) = 15										
4	n = 25										
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											

**Random Number Generation**

Number of Variables: 10 OK

Number of Random Numbers: 25 Cancel

Distribution: Normal Pomoč

Parameters

Mean = 125

Standard deviation = 15

Random Seed: 77

Output options

Output Range: \$B\$2 ...

New Worksheet Ply:

New Workbook



## 6 Intervali zaupanja

	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	89,42275	124,7275	129,8323	129,5683	125,3001	128,3078	127,6116	106,3279	107,2786	124,5174
3	sigma (standardni odklon) = 15	139,3766	151,452	101,4213	143,4013	136,3367	113,8109	116,7574	105,6675	120,9423	134,0804
4	n = 25	123,6793	119,2678	148,7782	115,9829	142,4861	96,07016	150,1763	136,7783	136,5223	143,1601
5		107,8491	132,2796	135,6699	84,33904	111,2153	126,353	126,9337	126,2758	108,8047	114,3419
6		150,0413	172,9078	109,1538	137,3185	114,6819	113,3261	115,1949	151,7618	138,3392	127,156
7		118,7358	135,6817	115,921	130,7989	125,7571	112,3161	144,5962	114,7848	116,369	109,3034
8		130,7334	119,0189	132,3558	136,1785	123,4418	115,0295	137,4946	121,4035	116,4501	135,8256
9		137,519	116,7841	124,1625	103,0433	142,2349	105,3714	143,4745	128,4208	119,5765	128,0344
10		123,3068	119,2987	111,3686	120,2076	125,6778	106,9516	110,5104	113,3943	133,9046	140,4937
11		111,9312	129,1518	133,4233	135,7202	135,0379	113,9588	122,294	125,2588	147,9781	106,7295
12		117,3529	134,4708	122,1177	142,1156	107,7005	131,3695	120,7468	87,53955	138,6159	136,7144
13		81,08887	127,6244	170,8102	120,7599	123,4083	113,2045	125,0947	104,8942	105,2082	150,9507
14		133,707	147,2973	129,9279	127,8496	168,5106	117,426	101,0878	102,6575	144,7153	117,5976
15		112,8781	102,8982	129,4903	132,2486	127,0703	129,584	113,8533	93,92079	113,0081	112,3832
16		115,2389	119,3504	135,6817	123,1693	108,2554	140,8201	140,9286	123,7276	147,5073	133,1041
17		110,3928	102,668	111,2886	142,3846	132,7413	157,6919	148,1692	146,7183	92,05683	117,998
18		97,00698	124,7125	120,5097	113,8124	130,0746	130,0504	120,608	112,481	97,43413	131,1243
19		96,2235	142,719	140,3419	130,4272	142,0148	136,4869	123,0756	113,2342	113,5147	139,7908
20		146,4629	141,9061	144,5183	114,1476	117,2442	119,1603	131,0582	117,5184	120,2112	119,2382
21		132,2228	101,0057	140,6629	110,0874	120,4076	101,9399	110,4535	120,2197	135,7647	117,9596
22		132,2744	120,6667	131,7268	136,8502	129,9279	135,9361	113,4283	108,2938	123,0976	113,6496
23		127,1966	127,8146	162,5123	116,02	124,4898	110,8953	111,705	116,7561	146,5173	126,7891
24		126,98	104,9448	136,8409	120,0491	118,0377	133,0352	116,3487	128,1071	122,7929	124,864
25		112,2141	133,0378	114,5448	116,6103	108,9966	111,2065	100,7337	130,0637	119,3307	108,0874
26		126,519	145,0524	95,11396	123,1276	102,7923	120,8518	135,0465	130,0588	140,1591	147,2421

### INTERVALI ZAUPANJA 2

Za vzorce velikosti  $n=25$ , generirane po normalni porazdelitvi, izračunajte 95 % intervale zaupanja za povprečje (tabeli 6.3 in 6.4, do vrstice 39). Računajte s predpostavko, da  $\sigma$  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 6.3: Prikaz dela delovnega lista »deset vzorcev«. 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	delta	=B35*B31/SQRT(B32)
37	$l_1$	=B30-B36
38	$l_2$	=B30+B36
39	$\bar{x}_{povp}$	=B30
40	Ali je populacijsko povprečje pod spodnjo mejo, $125 < l_1$ ? DA - 1, NE - 0	=IF(125<B37;1;0)
41	Ali je populacijsko povprečje nad zg. mejo, $125 > l_2$ ? DA - 1, NE - 0	=IF(125>B38;1;0)
42	Ali 95 % IZ vsebuje populacijsko povprečje? DA - 0, NE - 1	=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 6.3 in 6.4, 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 6.3, vrstice 40-41):

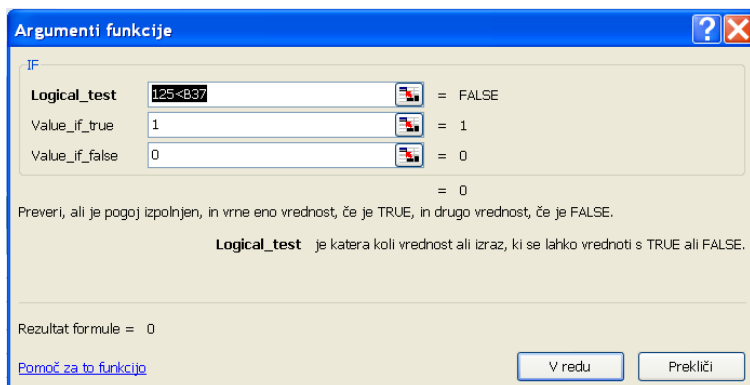


Tabela 6.4: Prikaz izračunov za 95 % interval zaupanja za povprečje na delovnem listu »deset vzorcev«. Brst generatorja slučajnih števil je 77.

	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	$t_{\alpha/2}(SP=n-1)$	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639	2,0639
36	delta	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, $125 < I_1$ ? DA - 1, NE - 0	0	0	0	0	0	0	0	0	0	0
41	Ali je populacijsko povprečje nad zg. mejo, $125 > I_2$ ? 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 6.2 (zgoraj). Grafikon shranite na grafičnem delovnem listu »slika IZ deset vzorcev«. ♦

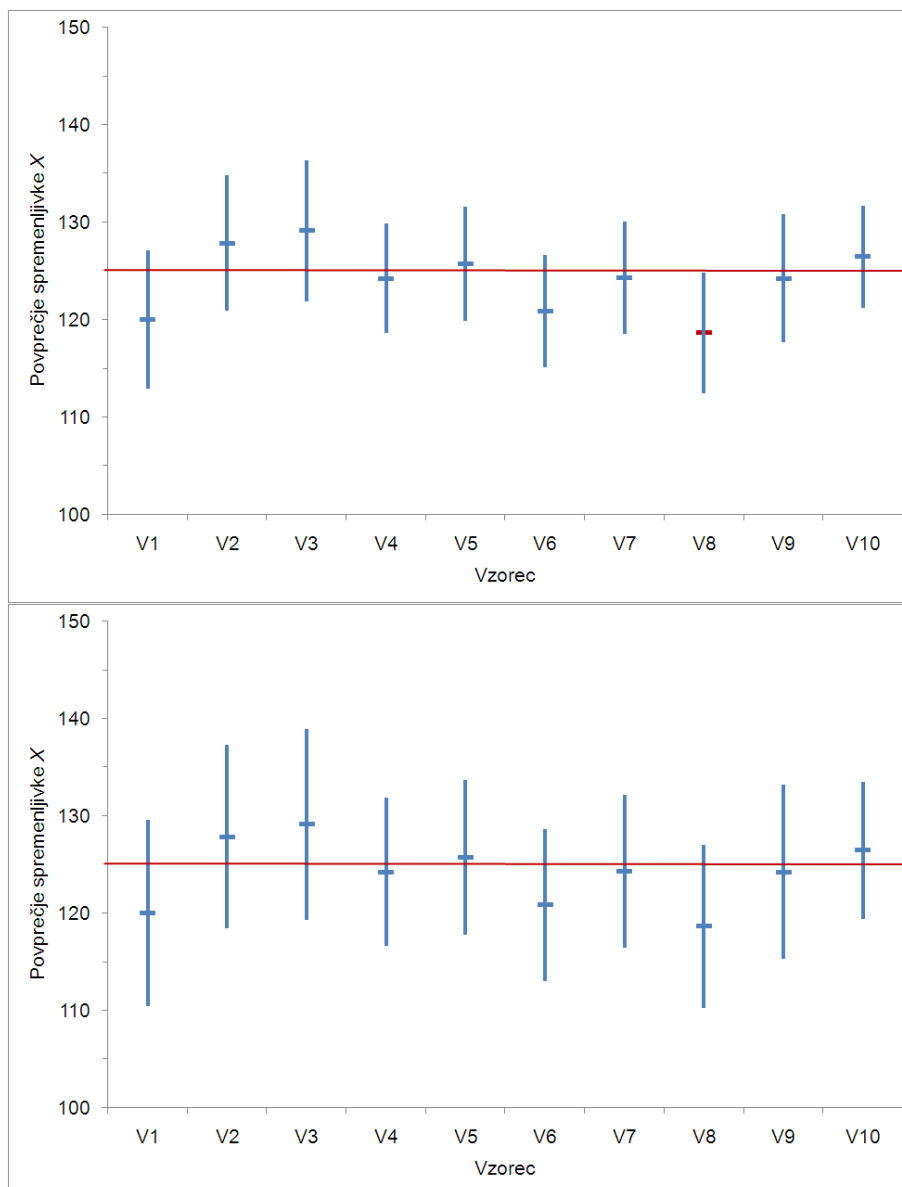
### INTERVALI ZAUPANJA 5

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 6.2 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 6.2: 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 rdečo črtico je označeno vzorčno povprečje, katerega IZ ne vsebuje populacijskega povprečja  $\mu = 125$ . Rdeča črta, ki označuje populacijsko povprečje, 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 uporabite poljuben 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 6.3: Sto intervalov zaupanja za povprečje**

## INTERVALI ZAUPANJA 7

V datoteki INTERVALI ZAUPANJA.xlsx 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  $n = 25$ . ♦

## INTERVALI ZAUPANJA 8

Za vzorce velikosti  $n = 25$ , generirane po normalni porazdelitvi, izračunajte 95 % in 99 % intervale zaupanja za povprečje. Računajte ob predpostavki, da  $\sigma$  ocenimo z vzorčnim standardnim odklonom  $s$ , izračunanim na osnovi vrednosti v vzorcu.

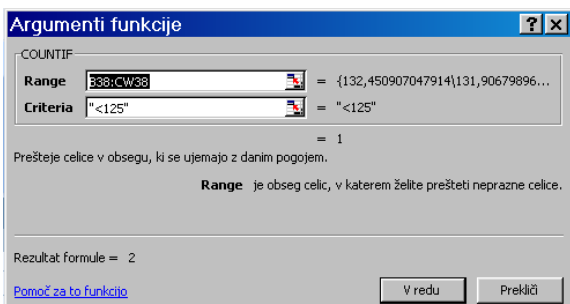
**Navodilo:** izračune uredite enako, kot v tabeli 6.3. Ali lahko formule iz tabele 6.3 na delovnem listu »deset vzorcev« kopiramo na delovni list »sto vzorcev«? Katero formulo moramo popraviti? ♦

## 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 6.3) 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, $125 < l_1$ ? DA - 1, NE - 0	=IF(125<B37;1;0)
41	Ali je populacijsko povprečje nad zg. mejo, $125 > l_2$ ? DA - 1, NE - 0	=IF(125>B38;1;0)
42	Ali 95 % IZ vsebuje populacijsko povprečje? DA - 0, NE - 1	=B40+B41
43	Število intervalov, ki <b>vsebujejo</b> populacijsko povprečje:	=100-SUM(B42:CW42)
44	Število intervalov, ki <b>ne vsebujejo</b> populacijskega povprečja	=COUNTIF(B37:CW37;">125")+COUNTIF(B38:CW38;"<125")

#### INTERVALI ZAUPANJA 10

Intervale zaupanja grafično prikažite v istem stilu, kot na sliki 6.2. 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«. ♦

## 6.2 Velikost vzorca za povprečje pri normalni porazdelitvi

### 6.2.1 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 ( $\mu$ ) največ  $\Delta$ . Pri tem predpostavimo, da je  $\sigma$  znana. V takem primeru velikost vzorca izračunamo z izrazom

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{\Delta} \right)^2. \text{ Da dobimo celo število, rezultat vedno zaokrožimo navzgor.}$$

## 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  $\sigma$  znana, njena vrednost je 15 mm.

**Navodilo:** v datoteki INTERVALI ZAUPANJA.xls na delovnem listu »velikost vzorca, sigma znana« izpolnite tabelo 6.5. Formule napišite tako, da se vam ob spreminjanju vrednosti za  $\sigma$  za zaupanje  $1 - \alpha$  in  $\Delta$  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 želenih decimalnih mest navzgor zaokroženega števila.

**Rešitev:**  $\Delta = 10$  mm,  $n = 9$ ;  $\Delta = 5$  mm,  $n = 35$ . ♦

Tabela 6.5: Prikaz izračuna potrebne velikosti vzorca, če je sigma znana, na delovnem listu »velikost vzorca, sigma znana«

	A	B		B
2	sigma	15	2	15
3	zaupanje	0,95	3	0,95
4	alfa	=1-B3	4	0,05
5	$Z_{\alpha/2}$	=NORMINV(1-B4/2;0;1)	5	1,9600
6	delta	10	6	10
7	Velikost vzorca		7	
8	n	=(B5*B2/B6)^2	8	8,6
9	navzgor zaokroženo število n	=ROUNDUP(B8;0)	9	9

### 6.2.2 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 ( $\mu$ ) največ  $\Delta$ . Pri tem predpostavimo, da  $\sigma$  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:

$$n_0 = \left( \frac{z_{\alpha/2} \cdot s}{\Delta} \right)^2.$$

V drugem koraku velikost vzorca izračunamo z izrazom:

$$n_1 = \left( \frac{t_{\alpha/2}(SP = n_0 - 1) \cdot s}{\Delta} \right)^2.$$

Če je  $n_1$  večji od  $n_0$ , iterativni postopek nadaljujemo:

$$n_2 = \left( \frac{t_{\alpha/2}(SP = n_1 - 1) \cdot s}{\Delta} \right)^2, \text{ sicer je } n_1 \text{ iskana rešitev.}$$

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 na delovnem listu »deset vzorcev« predpostavite, da  $\sigma$  ni znana, za vsak generirani 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 6.6 (velja za generirane vzorce z brstom generatorja slučajnih števil 77). ♦

Tabela 6.6: Prikaz izračuna potrebne velikosti vzorca, če sigme ne poznamo in uporabimo iterativni postopek na delovnem listu »velikost vzorca, sigma ni znana«. Brst generatorja slučajnih števil je 77.

	A	B		B	C	D	E	F
2	Izračun velikosti vzorca če sigma ni znana	V1		V1	V2	V3	V4	V5
3	delta	5		5	5	5	5	5
4	s	17,1348327151802		17,13	16,86	17,57	13,70	14,25
5	1. korak							
6	alfa	0,05		0,05				
7	$Z_{\alpha/2}$	=NORMINV(1-\$B\$6/2;0;1)		1,96	1,96	1,96	1,96	1,96
8	$n_0$	=(B7*B4/B3)^2		45,1	43,7	47,4	28,8	31,2
9	navzgor zaokroženo število $n_0$	=ROUNDUP(B8;0)		46	44	48	29	32
10	2. korak							
11	$SP=n_0-1$	=B9-1		45	43	47	28	31
12	$t_{\alpha/2}(SP)$	=TINV(\$B\$6;B11)		2,0141	2,0167	2,0117	2,0484	2,0395
13	$n_1$	=(B12*B4/B3)^2		47,6	46,3	50,0	31,5	33,8
14	navzgor zaokroženo število $n_1$	=ROUNDUP(B13;0)		48	47	50	32	34
15	3. korak							
16	$SP=n_1-1$	=B14-1		47	46	49	31	33
17	$t_{\alpha/2}(SP)$	=TINV(\$B\$6;B16)		2,0117	2,0129	2,0096	2,0395	2,0345
18	$n_2$	=(B17*B4/B3)^2		47,5	46,1	49,9	31,2	33,6
19	navzgor zaokroženo število $n_2$	=ROUNDUP(B18;0)		48	47	50	32	34
20	itd. dokler se velikost vzorca spreminja							



## 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? ♦

### 6.3 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 (delovni list »IZ za p«, tabela 6.7). ♦

## INTERVAL ZAUPANJA IN VELIKOST VZORCA ZA p 2

Izračunajte 95 % in 99 % interval zaupanja za verjetnost, da oseba zboli za gripo (tabela 6.7). Izračunane vrednosti obrazložite. ♦

Tabela 6.7: Prikaz izračunov za 95 % interval zaupanja za verjetnost<sup>23</sup>, da oseba zboli za gripo na delovnem listu »IZ za p«; izpis formul (levo); izpis rezultatov (desno)

	A	B		A	B
2	Približni interval zaupanja za p		2	Približni interval zaupanja za p	
3	zaupanje	0,95	3	zaupanje	95%
4	n	250	4	n	250
5	x	45	5	x	45
6	p*	=B5/B4	6	p*	0,180
7	1-p*	=1-B6	7	1-p*	0,820
8	alfa	=1-B3	8	alfa	0,05
9	Z <sub>alfa/2</sub>	=NORMINV(1-B8/2;0;1)	9	Z <sub>alfa/2</sub>	1,960
10	delta	=B9*SQRT(B6*B7/B4)	10	delta	0,05
11	l <sub>1</sub>	=B6-B10	11	l <sub>1</sub>	0,132
12	l <sub>2</sub>	=B6+B10	12	l <sub>2</sub>	0,228

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  $\Delta$ . V takem primeru velikost vzorca izračunamo z izrazom

$$n = \hat{p}(1 - \hat{p}) \cdot \left( \frac{z_{\alpha/2}}{\Delta} \right)^2.$$

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 6.8)? ♦

Tabela 6.8: Prikaz izračunov za potrebno velikost vzorca ob vnaprej določeni maksimalni vrednosti za odklon intervala zaupanja  $\Delta = 0,01$ , danem zaupanju in znani oceni za verjetnost, da oseba zboli za gripo  $\hat{p} = 0,18$ ; izpis formul (levo); izpis rezultatov (desno)

	A	B		A	B
2	zaupanje	0,95	2	zaupanje	95%
3	p*	0,18	3	p*	18%
4	1-p*	=1-B3	4	1-p*	0,82
5	alfa	=1-B2	5	alfa	0,05
6	Z <sub>alfa/2</sub>	=NORMINV(1-B5/2;0;1)	6	Z <sub>alfa/2</sub>	1,960
7	delta	0,01	7	delta	1%
8	Velikost vzorca		8	Velikost vzorca	
9	n	=B6^2/B7^2*B3*B4	9	n	5670,0
10	navzgor zaokroženo število n	=ROUNDUP(B9;0)	10	navzgor zaokroženo število n	5670

<sup>23</sup> S p\* v excelovih tabelah označujemo vzorčno oceno verjetnosti  $\hat{p}$ .

### 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  $\Delta$  v tabeli 6.8. Ali se izračuni s tem ustrezno spremenijo? ♦

### INTERVAL ZAUPANJA IN VELIKOST VZORCA ZA p 5

Ponovite izračune potrebne velikosti vzorca še pri 99 % zaupanju. ♦

## 7 Preverjanje statističnih domnev za en vzorec

### Novi pojmi:

- ničelna domneva  $H_0 : \mu = \mu_0$ ;  $\sigma$  je znana,  $\sigma$  ni znana
- ničelna domneva  $H_0 : p = p_0$
- alternativna domneva dvostranska/enostranska

### Datoteke s podatki:

DOMNEVE EN VZOREC.xlsx

### 7.1 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  $\mu_0$  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 poleg 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 7.1: Čakalna doba v zdravstvenem domu**

Podatki o čakalni dobi v zdravstvenem domu za vzorec 20 slučajno izbranih pacientov so v datoteki DOMNEVE EN VZOREC.xlsx na delovnem listu »čakalna doba« in v tabeli 7.1. Na osnovi vzorca bomo preverili ničelno domnevo, da je povprečna čakalna doba v populaciji vseh pacientov tega zdravstvenega doma  $\mu_0$ . Za čakalno dobo v zdravstvenem domu lahko privzamemo normalno porazdelitev.

Tabela 7.1: 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:** 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 7.2: Izračuni za t-preizkus, formule (levo), rezultati (desno)

	A	B		A	B
23	t-preizkus		23	t-preizkus	
24	$x_{povp}$	=AVERAGE(A2:A21)	24	$x_{povp}$	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	$H_0: \mu_i = \mu_{i_0}$ $\mu_{i_0} =$	30	28	$H_0: \mu_i = \mu_{i_0}$ $\mu_{i_0} =$	30
29	test (enostranski, dvostranski)	2	29	test (enostranski, dvostranski)	2
30	SP	=B26-1	30	SP	19
31	t števec	=B24-B28	31	t števec	4,1000
32	t imenovalec	=B25/SQRT(B26)	32	t imenovalec	2,6546
33	t	=B31/B32	33	t	1,5445
34	mejna vrednost $t_{\alpha/2}(SP)$	=TINV(B27;B30)	34	mejna vrednost $t_{\alpha/2}(SP)$	2,0930
35	p-vrednost	=TDIST(ABS(B33);B30;B29)	35	p-vrednost	0,1390

**Rešitev:**

$$H_0 : \mu = \mu_0 = 30 \text{ min in } H_1 : \mu \neq 30 \text{ min}$$

$$t = 1,5445 \text{ in } p = 0,1390$$

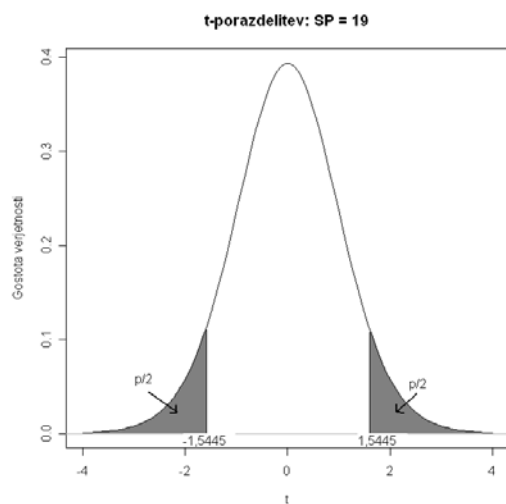
Statistični sklep: ničelno domnevo obdržimo ( $p = 0,1390$ ). Rezultati niso statistično značilni.

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 7.1). 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 7.1: 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):

SP	$\alpha/2$		
	0,1	0,05	0,025
⋮	⋮	⋮	⋮
19	1,328	1,729	2,093
⋮	⋮	⋮	⋮

$$(t_1 = 1,328) < (t = 1,5445) < (t_2 = 1,729) \quad (p_2 = 2 \cdot 0,05) < p < (p_1 = 2 \cdot 0,10)$$

$0,10 < p < 0,20$ . ♦

## ČAKALNA DOBA 3

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 7.2. 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 $t_{\alpha/2}(SP)$	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 7.1: Čakalna doba v zdravstvenem domu - ponovitev snovi**

## ČAKALNA DOBA 4

Za čakalno dobo izračunajte vrednosti, ki določajo okvir z ročaji. Ne pozabite preveriti, ali med podatki obstajajo osamelci. Izračunane vrednosti uredite v tabelo in jih obrazložite. ♦

## ČAKALNA DOBA 5

Izračunajte 2 primerni meri sredine in 2 primerni meri variabilnosti za navedene podatke ter vrednosti obrazložite. ♦

## ČAKALNA DOBA 6

Na osnovi vzorčnih podatkov izračunajte 95 % interval zaupanja za povprečno čakalno dobo v zdravstvenem domu in ga obrazložite.

**Rešitev:**  $l_1 = 28,5$  min in  $l_2 = 39,7$  min. ♦

## ČAKALNA DOBA 7

Kaj bi morali narediti, če bi želeli, da se širina intervala zaupanja 2-krat zmanjša? ♦

## ČAKALNA DOBA 8

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.

**Rešitev:**  $n = 37$ . ♦

**Domača naloga 7.2: 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.xlsx 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  $p = 0,0000$ . ♦

**Zgled 7.2: Premer debla**

V tabeli 3.3 in v datoteki DOMNEVE EN VZOREC.xlsx na delovnem listu »premer debla« (tudi v PREMIER DEBLA.xlsx) so podatki o številu dreves v posameznem razredu glede na premer debla.



## PREMER DEBLA 8

Na podlagi približka za povprečje in ob predpostavki, da je znana  $\sigma = 6$  cm, preverite domnevo, da je povprečen premer debla 22 cm ( $\alpha = 0,01$ ). Izračunajte  $p$ -vrednost. Obrazložite rezultate.

**Navodila:** izračune naredite v datoteki DOMNEVE EN VZOREC.xlsx na delovnem listu »premer debla« (izračun približka za povprečje je prikazan v zgledu 3.3).

**Rešitev:**

$$H_0 : \mu = \mu_0 = 22 \text{ cm in } H_1 : \mu \neq 22 \text{ cm (tabela 7.3)}$$

$$z = 0,4845 \text{ in } p = 0,6280 \blacklozenge$$

Tabela 7.3: Izračuni za z-preizkus za preverjanje ničelne domneve o povprečnem premeru debla dreves; izpis 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	$H_0: \mu = \mu_0, \quad \mu_0 =$	22	17	$H_0: \mu = \mu_0, \quad \mu_0 =$	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

## PREMER DEBLA 9

Isto ničelno domnevo preverite ob predpostavki, da sigma ni znana (izračun približka za standardni odklon je prikazan v tabeli 3.4).

**Rešitev:**  $t = 0,4614$  in  $p = 0,6447$  (tabela 7.4).  $\blacklozenge$

Tabela 7.4: Izračuni za t-preizkus za preverjanje ničelne domneve o povprečnem premeru debela 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	=SQRT(G11/499)	14	približek za standardni odklon	6,30
15	n	500	15	n	500
16	alfa	0,01	16	alfa	0,01
17	$H_0: \mu = \mu_0, \quad \mu_0 =$	22	17	$H_0: \mu = \mu_0, \quad \mu_0 =$	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 7.3: Premer debla - ponovitev snovi**

## PREMER DEBLA 10

Predpostavljamo, da je premer debla slučajna spremenljivka, porazdeljena po normalni porazdelitvi s parametroma  $\mu = 22$  cm in  $\sigma = 6$  cm. Narišite gostoto verjetnosti te normalne porazdelitve. ♦

**Domača naloga 7.4: 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 7.5: Število gledalcev po starosti, ki je izražena v dopoljenih letih

Starost	Število
20-29	20
30-39	75
40-49	45
50-99	10

## STAROST GLEDALCEV 1

Podatke iz preglednice 7.5 vpišite na delovni list »starost gledalcev«. ♦

## STAROST GLEDALCEV 2

Izračunajte približek za povprečje in za standardni odklon za starost gledalcev.

**Rešitev:**  $\bar{x} = 39,3$  let  $s = 11,49$  let. ♦

## 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:**  $t = 4,6175$  in  $p = 0,0000$  ♦

## 7.2 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 7.4: Učinkovitost dveh vrst zdravil

Proučevali so učinkovitost dveh vrst zdravil proti bakterijskemu vnetju. V raziskavi je bilo 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 122 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 (tabela 7.6).

**Rešitev:**  $\hat{p}_A = 0,81$  in  $\hat{p}_B = 0,87$  ♦

#### UČINKOVITOST ZDRAVIL 2

Standardno zdravilo proti bakterijskemu vnetju S ima učinkovitost 0,80. Iz predhodnih raziskav je znano, da je zdravilo A boljše od standardnega zdravila S. Napišite ničelno in alternativno domnevo in na osnovi podatkov preverite to domnevo ( $\alpha = 0,05$ ) (enostranska alternativna domneva) (tabela 7.6).

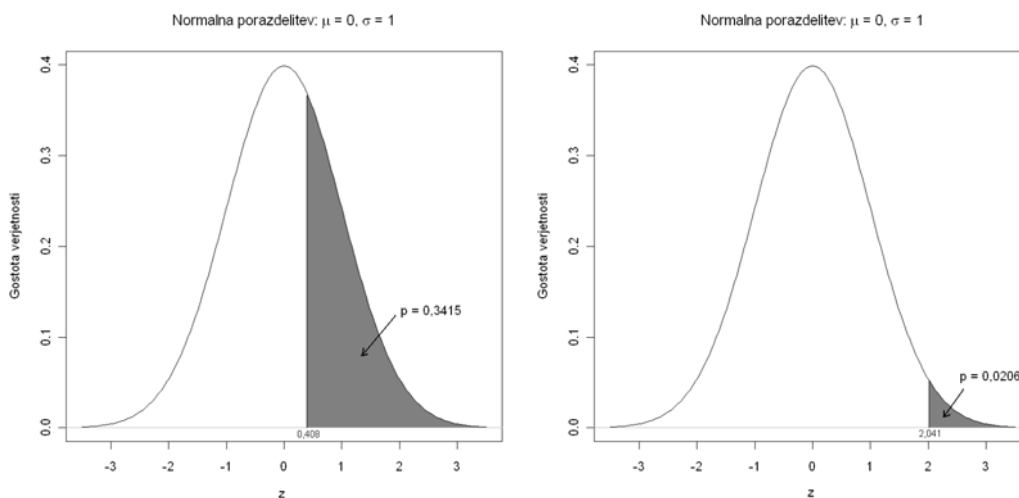
Izračunajte in obrazložite  $p$ -vrednost ter jo shematično grafično prikažite (na papir, slika 7.2).

**Rešitev:**

$H_0: p = p_0 = 0,80$  in  $H_1: p > 0,80$

$z = 0,408$  in  $p = 0,3415$

Ničelno domnevo obdržimo ( $p = 0,3415$ ). Rezultati niso statistično značilni. Podatki nakazujejo, da je učinkovitost zdravila A enaka učinkovitosti standardnega zdravila S. ♦



Slika 7.2: Grafični prikaz  $p$ -vrednosti za vrednost testne statistike  $z = 0,408$  (levo) in  $z = 2,041$  (desno) pri enostranski alternativni domnevi

### UČINKOVITOST ZDRAVIL 3

Standardno zdravilo proti bakterijskemu vnetju S ima učinkovitost 0,80. Tudi za zdravilo B je iz predhodnih raziskav znano, 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 7.6).

**Rešitev:**  $z = 2,041$  in  $p = 0,0206$

Ničelno domnevo zavrnemo ( $p = 0,0206$ , slika 7.2). Rezultati so statistično značilni. S tveganjem, manjšim od  $\alpha = 0,05$ , trdimo, da je zdravilo B učinkovitejše od standardnega zdravila. ♦

Tabela 7.6: Izračun ocene verjetnosti, da zdravila A in B učinkujeta, preverjanje ničelne domneve ob enostranski alternativni domnevi; izpis formul (levo); izpis rezultatov (desno)

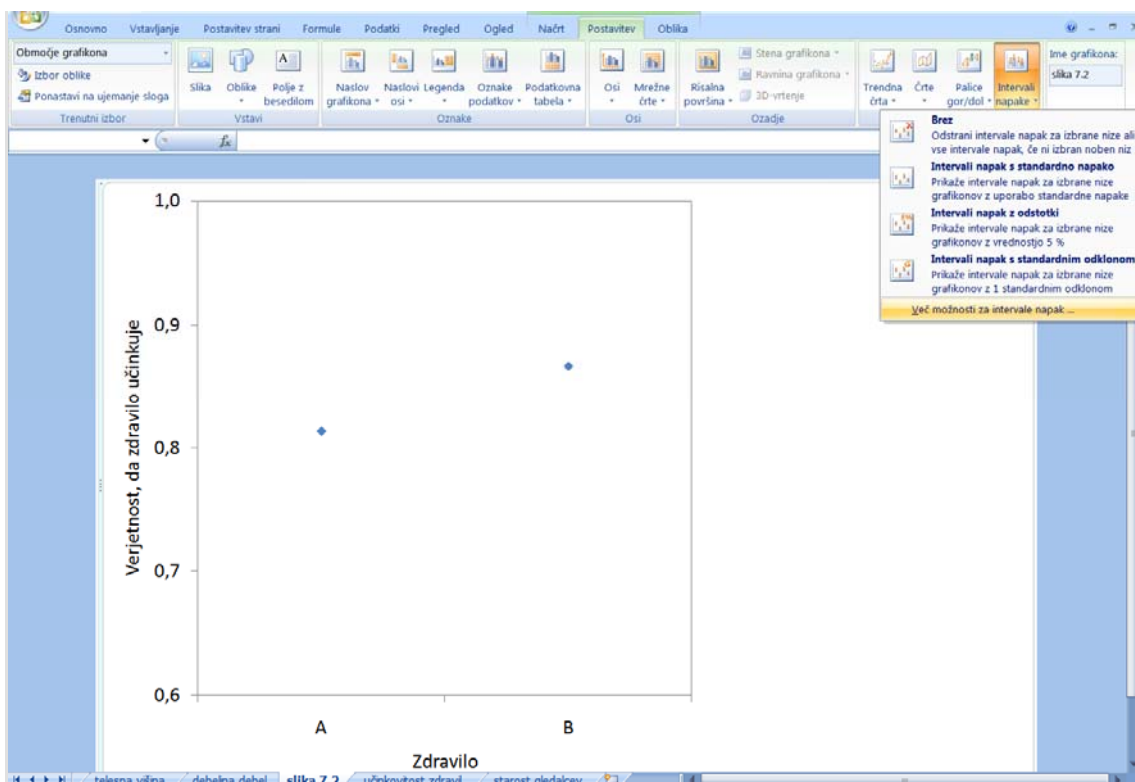
	A	B		A	B	C
2		Zdravilo	2		Zdravilo	
3		A	3		A	B
4	n	150	4	n	150	150
5	x	122	5	x	122	130
6	p*	=B5/B4	6	p*	0,81	0,87
7	1-p*	=1-B6	7	1-p*	0,19	0,13
8	alfa	0,05	8	alfa	0,05	0,05
9	H <sub>0</sub> : p=p <sub>0</sub> , p <sub>0</sub> =	0,8	9	H <sub>0</sub> : p=p <sub>0</sub> , p <sub>0</sub> =	0,8	0,8
10	H <sub>1</sub> : p>p <sub>0</sub>		10	H <sub>1</sub> : p>p <sub>0</sub>		
11	test (enostranski, dvostranski)	1	11	test (enostranski, dvostranski)	1	1
12	z_števlec	=B5-B4*B9	12	z_števlec	2,000	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	0,408	2,041
15	Z <sub>mejna</sub>	=NORMINV((1-B8/B11);0;1)	15	Z <sub>mejna</sub>	1,645	1,645
16	p-vrednost	=B11*(1-NORMDIST(B14;0;1;TRUE))	16	p-vrednost	0,3415	0,0206

### Domača naloga 7.5: Učinkovitost zdravil - ponovitev snovi

#### UČINKOVITOST ZDRAVIL 4

Za vsako zdravilo posebej izračunajte 95 % interval zaupanja za verjetnost, da je zdravilo učinkovito (tabela 7.6). Oba intervala grafično prikažite na isti sliki (slika 7.3). Obrazložite rezultate.

**Navodilo:** oceni za verjetnost, da zdravilo učinkuje (p\*), grafično prikažemo s točkama tako, da uporabimo *Črtni grafikon*, ki mu odstranimo povezovalno črto z ukazom **Oblikuj niz podatkov/Barva črt** možnost *Brez črte*. Nato z ukazom **Intervali napake/Več možnosti** dodamo še črte za interval zaupanja:



V pogovornem oknu **Oblikuj intervale napak** v razdelku *Velikost napake* izberemo možnost *Po meri* in kliknemo gumb **Določanje vrednosti**. V pogovornem okencu **Intervali napak po meri** v polji  *Pozitivna oz. negativna vrednost napake*  izberemo celici z odklonom intervala zaupanja (B19:C19) na delovnem listu »učinkovitost zdravil«.

## 7 Preverjanje statističnih domnev za en vzorec

22			
23			
24	95 % IZ		
25	$p^*$	0,81	0,87
26	Z <sub>alfa/2</sub>	1,960	1,960
27	delta	0,062	0,054
28	$l_1$	0,75	0,81
29	$l_2$	0,88	0,92
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			

**Oblikuj intervale napak**

Navpični intervale napak

Barva črt:

Slog črte:

Senčeno:

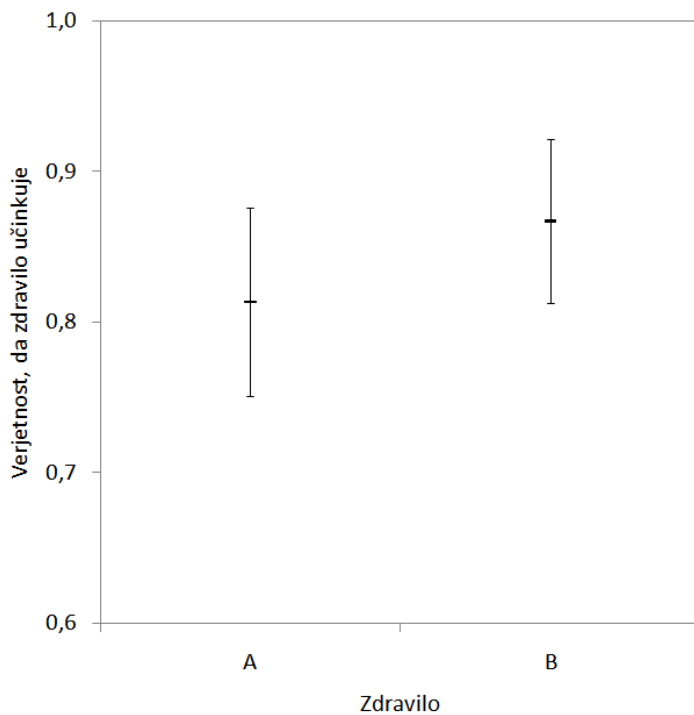
Prikaz

Smer:  Oboje,  Minus,  Plus

Končni slog:  Brez ozaljšane začetnice,  Ozaljšana začetnica

Velikost napake:  Nespremenljiva vrednost: 0,1,  Odstotek: 5,0 %,  Standardni odkloni: 1,0,  Standardna napaka,  Po meri:  Določanje vrednosti

Zapri



Slika 7.3: Ocena za verjetnost, da zdravilo učinkuje, s 95 % intervalom zaupanja, za zdravili A in B

## 8 Preverjanje statističnih domnev za dva vzorca in interval zaupanja za razliko

### Novi pojmi:

- ničelna domneva  $H_0 : \mu_1 - \mu_2 = \delta_0$  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.xlsx

ANKETA.xlsx

### 8.1 Razlika povprečij za neodvisna vzorca

Imamo dva neodvisna vzorca, v katerih analiziramo slučajno spremenljivko  $X$ , ki jo v prvi populaciji označimo z  $X_1$  in je porazdeljena  $X_1 \sim N(\mu_1, \sigma)$ , v drugi populaciji pa z  $X_2$  in je porazdeljena  $X_2 \sim N(\mu_2, \sigma)$ . Preverjamo ničelno domnevo  $H_0 : \mu_2 - \mu_1 = \delta_0$ .

Testna statistika je

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{s_{sk}^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

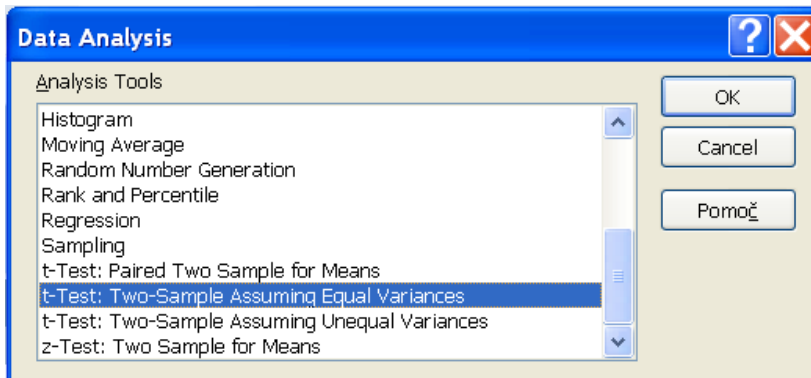
njena ničelna porazdelitev je Studentova porazdelitev:  $t \sim t(SP = n_1 + n_2 - 2)$ . V

$t$ -statistiki je  $s_{sk}^2$  skupna varianca, ki se izračuna  $s_{sk}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ .

Opozarjamo, da sta varianci obeh normalnih porazdelitev enaki.



Za preverjanje ničelne domneve  $H_0 : \mu_1 - \mu_2 = \delta_0$  ob pogoju, da sta vzorca neodvisna in ob pogoju, da lahko predpostavimo, da sta varianci populacij enaki, v Excelu uporabimo ukaz **Analiza podatkov/t-Test: Two-Sample Assuming Equal Variances** na kartici **Podatki** v skupini Analiza. Če pa predpostavke o enakosti varianc ne moremo privzeti, uporabimo **t-Test: Two-Sample Assuming Unequal Variances**.<sup>24</sup>



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 stopnjo značilnosti  $\alpha$ . Pri osnovni nastavitvi je  $\alpha = 0,05$ . V poljih razdelka **Output options** določimo, kam naj se izpiše preglednica z rezultati (tabela 8.2).

### Zgled 8.1: 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.

<sup>24</sup> Za preverjanje predpostavke o enakosti varianc v obeh populacijah lahko uporabimo ohlapno »pravilo palca« na vzorčnih podatkih. Če velja:

$$\frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)} < 2, \text{ je uporaba } t \text{-preizkusa za enake variance vedno dopustna.}$$

Podatki so v tabeli 8.1 (datoteka DOMNEVE DVA VZORCA.xlsx, delovni list »tobak«).

Tabela 8.1: Višina 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 naredimo s formulami v pripravljene tabeli na delovnem listu »tobak«, potem pa uporabimo omenjeno 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 8.2. ♦

## 8 Preverjanje statističnih domnev za dva vzorca, interval zaupanja za razliko

	A	B	C	D	E	F
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	H0: $\mu_{Ne} - \mu_{Ok} = \delta_0$	delta <sub>0</sub>	= 0			
19	SP za ničelno porazdelitev	=B15+C15				
20	skupna varianca	=(B15*B17+C15*C17)/B19				
21	števec_t	=(C16-B16)-B18				
22	imenovalec_t	=SQRT(B20*(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, B19)				
27	p-vrednost	=TDIST(ABS(B23),B19,B25)				

**t-Test: Two-Sample Assuming Equal Variances**

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha:

Output options

Output Range:

New Worksheet Ply:

New Workbook

OK Cancel Pomoc

TOBAK 2

Napišite statistični in vsebinski sklep.

**Rešitev:**

$$t = 3,7780 \text{ in } p = 0,0014$$

Statistični sklep:  $H_0$  zavrnemo pri  $\alpha = 0,05$ . Rezultati so statistično značilni.

Vsebinski sklep: s tveganjem, manjšim od 0,05, trdimo, da povprečna višina okuženih in neokuženih rastlin tobaka ni enaka ( $p = 0,0014$ ). ♦

Tabela 8.2: Rezultat ukaza **Orodja/Analiza podatkov/t-Test: Two-Sample Assuming Equal Variances** za podatke v tabeli 8.1.

	Neokužene	Okužene	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. 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 okužene rastline.

Katero celico v tabeli na delovnem listu »tobak« je potrebno spremeniti?

Kje najdemo rezultat v tabeli 8.2?

**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 neokuženih rastlin večja od povprečne višine okuženih 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:**

$H_0 : \mu_{Ne} - \mu_{Ok} = 7$  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.

Katero celico v tabeli na delovnem listu »tobak« je potrebno spremeniti?

Če pravilno uporabimo excelovo orodje, dobimo kot rezultat spodnjo tabelo:

	E	F	G
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		<i>Neokužene</i>	<i>Okužene</i>
4	Mean	90,70	76,20
5	Variance	58,46	88,84
6	Observations	10	10
7	Pooled Variance	73,65	
8	Hypothesized Mean Difference	7	
9	df	18	
10	t Stat	1,9542	
11	P(T<=t) one-tail	0,0332	
12	t Critical one-tail	1,7341	
13	P(T<=t) two-tail	0,0664	
14	t Critical two-tail	2,1009	

### Rešitev:

$$t = 1,9542 \text{ in } p = 0,0332$$

Statistični sklep:  $H_0$  zavrnemo 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  $\alpha = 0,01$ ?

### Rešitev:

$$t = 1,9542 \text{ in } p = 0,0332$$

Statistični sklep:  $H_0$  obdržimo pri  $\alpha = 0,01$ . ♦

$(1 - \alpha)100\%$  interval zaupanja za  $\mu_1 - \mu_2$ , 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 6

Izračunajte 95 % interval zaupanja za razliko povprečne višine okuženih in neokuženih rastlin tobaka. Obrazložite rezultate.

**Navodila:** v Excelu z vrednostmi in formulami izpolnite pripravljeno tabelo na delovnem listu »tobak« (tabela 8.3).

Tabela 8.3: Izračuni za interval zaupanja za razliko povprečij dveh neodvisnih vzorcev; izpis formul (levo); izpis rezultatov (desno)

	A	B
31	zaupanje	0,95
32	alfa	=1-B31
33	SP	=B19
34	$t_{\alpha/2}(SP=n_1+n_2-2)$	=TINV(B32;B33)
35	delta	=B34*B22
36	$\bar{x}_{Ne}-\bar{x}_{Ok}$	=C16-B16
37	$l_1$	=B36-B35
38	$l_2$	=B36+B35

	A	B
31	zaupanje	0,95
32	alfa	0,05
33	SP	18
34	$t_{\alpha/2}(SP=n_1+n_2-2)$	2,1009
35	delta	8,06
36	$\bar{x}_{Ne}-\bar{x}_{Ok}$	14,50
37	$l_1$	6,44
38	$l_2$	22,56

**Rešitev:**

Pri 95 % zaupanju trdimo, da je povprečna višina neokuženih rastlin od 6,4 cm do 22,6 cm višja od povprečne višine okuženih rastlin tobaka. ♦

## TOBAK 7

Ponovite izračun pri 99 % zaupanju.

**Navodilo:** ustrezno spremenite vrednost za zaupanje v tabeli z izračuni na delovnem listu »tobak«.

**Rešitev:**

Pri 99 % zaupanju trdimo, da je povprečna višina neokuženih rastlin od 3,5 cm do 25,6 cm višja od povprečne višine okuženih rastlin tobaka. ♦

**Zgled 8.2 (nadaljevanje zgleda 7.2): 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. Na delovnem listu »telesna višina« so podatki za študentke.

## TELESNA VIŠINA 2

S pomočjo orodja vrtilne tabele oblikujte tabelo 8.4. Kaj predstavljajo izračuni v stolpcu »Skupaj«? ♦

Tabela 8.4: Število študentk, vzorčno povprečje (cm) in vzorčna varianca (cm<sup>2</sup>) za telesno višino za vsako smer študija posebej

Podatki	Smer študija			Skupaj
	AG	BT	MB	
Število študentk	33	14	32	79
Povprečna telesna višina (cm)	168,8	167,1	168,2	168,3
Varianca telesna višina (cm <sup>2</sup> )	40,6	26,0	33,8	34,8

## TELESNA VIŠINA 3

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 naredimo v Excelu s formulami in pri tem uporabimo oceni povprečja in variance, ki smo jih izračunali z uporabo vrtilne tabele (tabela 8.4). Preden se lotimo izračunov za preverjanje ničelne domneve, parkiramo vrtilno tabelo na nov delovni list z imenom »BT in MB s formulami«. Pomagamo si s tabelo izračunov na delovnem listu »tobak«.

Drugič izračune naredimo 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 **Razvrsti od A do Ž** na kartici **Podatki** v skupini Razvrsti in filtriraj. Preden uporabimo ukaz, mora biti izbrana ena izmed celic v stolpcu z vrednostmi spremenljivke *Smer študija*.

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. Na tem delovnem listu naj bo tudi izpis ukaza **Analiza podatkov/t-Test: Two-Sample Assuming Equal Variances**.

	A	B	C	D	E	F	G	H	I	J
1	BT	MB			Smer študija			t-Test: Two-Sample Assuming Equal Variances		
2	169	165		Vrednosti	BT	MB				
3	174	165		n	14	32			BT	MB
4	167	168		SP	13	31		Mean	167,14	168,20
5	159	163		Povprečje	167,14	168,20		Variance	26,02	33,80
6	167,5	165		Varianca	26,02	33,80		Observations	14	32
7	161	158		$H_0: \mu_1 - \mu_2 = \text{delta}_0$	0			Pooled Variance	31,50	
8	170	169		SP za ničelno porazdelitev	44			Hypothesized Mean Difference	0	
9	164,5	175		skupna varianca	31,50			df	44	
10	160	172		števec_t	1,0603			t Stat	-0,5895	
11	167	165		imenovalec_t	1,7985			P(T<=t) one-tail	0,2793	
12	176	164		t-statistika	0,5895			t Critical one-tail	1,6802	
13	167	159		alfa	0,05			P(T<=t) two-tail	0,5585	
14	173	176		test(eno, dvostranski)	2			t Critical two-tail	2,0154	
15	165	165		mejna vrednost t-statistike	2,0154					
16		168		p-vrednost	0,5585					

**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 4**

Preverite še domnevo, da so v povprečju študentke BT za 3 cm višje od študentk MB ( $\alpha = 0,05$ ).

**Rešitev:**

$$H_0 : \mu_{BT} - \mu_{MB} = 3 \text{ cm}$$

$$H_1 : \mu_{BT} - \mu_{MB} \neq 3 \text{ cm}$$

$$t = -2,2576 \text{ in } p = 0,0290 \blacklozenge$$



## 8.2 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 8.3: 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 8.5.

Tabela 8.5: 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 rešimo najprej v Excelu s formulami (pripravljena tabela v datoteki), nato uporabimo še Excelovo orodje **Analiza podatkov/t-Test: Paired Two Sample for Means** (tabela 8.6).

**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 ( $p = 0,4163$ ). ♦

## 8 Preverjanje statističnih domnev za dva vzorca, interval zaupanja za razliko

	A	B	C	D	E	F	G
1	Plavalec	Na začetku tečaja	Ob koncu tečaja	Razlika, d <sub>i</sub>			
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 <sub>d</sub>	=STDEV(D2:D16)			
20			H <sub>0</sub> : μ <sub>K</sub> - μ <sub>Z</sub> =delta <sub>0</sub> delta <sub>0</sub> =	50			
21			SP	=D17-1			
22			t_števcevec	=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(ABS(D24);D21;D25)			

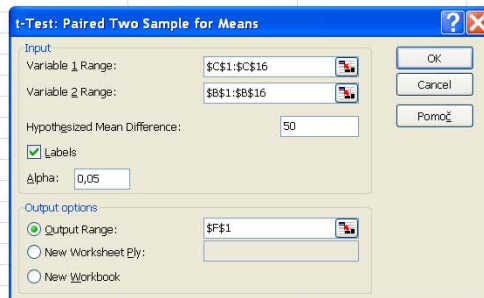


Tabela 8.6 Rezultat ukaza **Analiza podatkov/t-Test: Paired Two Sample for Means** za podatke iz tabele 8.4

	Ob koncu tečaja	Na začetku tečaja	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		$t_{\alpha/2}(SP = 14)$

$(1 - \alpha)$  100% interval zaupanja za  $\mu_D$  izračunamo z izrazom

$$l_{1,2} = \bar{d} \pm t_{\alpha/2}(SP = n - 1) \frac{s_d}{\sqrt{n}}.$$

#### PLAVANJE 2

Izračunajte 95 % interval zaupanja za povprečno razliko preplavane dolžine na začetku in na koncu tečaja. Izračunajte po formulah v pripravljene tabeli na delovnem listu »plavanje«. Obrazložite izračunane vrednosti.

##### **Rešitev:**

Pri 95 % zaupanju trdimo, da plavalci ob koncu tečaja v povprečju preplavajo od 38,0-77,3 m več kot na začetku tečaja. ♦

#### PLAVANJE 3

Ponovite izračune pri 99 % zaupanju.

##### **Rešitev:**

Pri 99 % zaupanju trdimo, da plavalci ob koncu tečaja v povprečju preplavajo od 30,4-84,9 m več kot na začetku tečaja. ♦

### **Zgled 8.4: Kationska izmenjalna kapaciteta tal, določena po dveh različnih metodah**

Na delovnem listu »KIK« so med drugim podatki o količini organske snovi [%] in kationski izmenjalni kapaciteti tal (KIK) [mmol<sub>c</sub>/100 g tal] za 50 talnih vzorcev. Za vsak vzorec (oznaka vzorca) je KIK 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 KIK.

#### KIK 1

Priprava podatkov za analizo: v excelovi datoteki najprej izberite talne vzorce s količino organske snovi pod 20 %.

**Navodilo:** uporabimo ukaz **Podatki/Filter**. Nato tri stolpce podatkovne zbirke: *oznaka vzorcev*, *metoda* in *KIK* izbranih talnih vzorcev kopirajte na delovni list »izbor KIK«.

## 8 Preverjanje statističnih domnev za dva vzorca, interval zaupanja za razliko

	A	B	C	D	E	F	G
	zaporedna številka	oznaka vzorca	metoda	glina (%)	organska snov (%)	pH	Ca (mmol./100 g tal)
2	1					4,58	0,03
3	2					4,73	13,46
4	3					4,53	9,28
5	4					4,21	0,52
6	5					4,3	1,03
7	6					5,26	19,88
8	7						5
9	8						3
10	9						4
11	10						7
12	11						3
13	12						7
14	13						3
15	14						9
16	15						4
17	16						7
18	17	107	EFE		19,1	11,8	1
19	18	112	EFE		23,8	5,0	0
20	19	113	EFE		18,7	6,3	0,90
							24,24

Podatke uredite tako, da bodo v enem stolpcu podatki za KIK po metodi AA, v sosednjem pa podatki za KIK po metodi EFE. Pazite, da bodo v vrsticah podatki za isti talni vzorec (ista oznaka vzorca). Tako ureditev podatkov najlažje naredimo z uporabo vrtljne tabele:

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3	<b>Vsota od KIK</b>	<b>metoda</b>									
4	<b>oznaka vzorca</b>	<b>AA</b>	<b>EFE</b>	<b>Skupna vsota</b>							
5	35	9,71	0,06	9,77							
6	76	28,77	23,30	52,07							
7	78	20,59	19,52	40,11							
8	82	14,29	0,98	15,27							
9	83	13,56	6,15	19,71							
10	85	32,30	23,98	56,28							
11	87	20,70	14,33	35,03							
12	88	20,68	13,75	34,43							
13	89	32,90	15,33	48,23							
14	90	23,30	3,64	26,94							
15	91	20,18	2,61	22,79							
16	93	12,73	5,47	18,20							
17	99	17,60	2,75	20,35							
18	100	17,12	2,71	19,83							
19	101	23,93	14,85	38,78							

Izločite skupno vsoto po vrsticah in stolpcih ter vrtljno tabelo parkirajte na nov delovni list z imenom »KIK za R«:

	A	B	C	D	E	F	G	H	I
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	13,56	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	14,58	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					

Koliko talnih vzorcev ima vsebnost organske snovi pod 20 %? ♦

KIK 2

Kaj je v tem primeru statistični vzorec? Koliko jih je? Kaj je enota proučevanja? Ali sta vzorca odvisna ali neodvisna? Koliko enot je v posameznem statističnem vzorcu? ♦

KIK 3

Preverite ničelno domnevo, ki pravi, da metodi AA in EFE dajeta v povprečju enake rezultate za KIK ( $\alpha = 0,05$ ). Najprej uporabite ukaz **Orodja/Analiza podatkov/t-Test: Paired Two Sample for Means** v Excelu. Izhodno tabelo shranite na delovni list »izbor KIK«. Obrazložite rezultate.

**Rešitev:**

$$H_0 : \mu_{AA} - \mu_{EFE} = 0 \qquad H_1 : \mu_{AA} - \mu_{EFE} \neq 0$$

$$t = 10,3626 \text{ in } p = 0,0000$$

KIK 4

Izračunajte 95 % interval zaupanja za povprečno razliko KIK med metodama AA in EFE.

**Rešitev:**

Pri 95 % zaupanju trdimo, da je povprečna kationska izmenjalna kapaciteta izmerjena po metodi AA od 7,02-10,40 mmol<sub>c</sub>/100 g tal večja od tiste, ki je izmerjena po metodi EFE. ♦

**8.3 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  $X_2 \sim b(n_2, p_2)$ . Če je aproksimacija binomske porazdelitve z normalno utemeljena, velja:  $b(n_1, p_1) \approx N(n_1 p_1, \sqrt{n_1 p_1 (1 - p_1)})$ ,  $b(n_2, p_2) \approx N(n_2 p_2, \sqrt{n_2 p_2 (1 - p_2)})$ . 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 8.5 (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 5

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.xlsx (tabeli 8.7 in 8.8).

Tabela 8.7: 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	=C6-B6	
15	z_imenovalec	=SQRT(B13*(1-B13)*(1/B4+1/C4))	
16	z-statistika	=B14/B15	
17	z_mejna	=NORMINV(1-B12/B11;0;1)	
18	p-vrednost	=B11*(1-NORMDIST(ABS(B16);0;1;TRUE))	

Tabela 8.8: Rezultati formul iz tabele 8.7, 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:

$$H_0: p_A = p_B = p_{sk} \quad H_1: p_A \neq p_B$$

$$z = -2,887 \quad p = 0,0039$$



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 6

Izračunajte 95 % interval zaupanja za razliko učinkovitosti zdravil A in B. Obrazložite rezultate.

**Navodilo:** izračune naredimo s pisanjem formul v pripravljeno tabelo na delovnem listu »učinkovitost zdravil« (tabela 8.9).

Tabela 8.9: Izračun intervala zaupanja za razliko Bernoullijevih verjetnosti za dva neodvisna vzorca; izpis formul (levo); rezultati (desno)

	A	B		A	B
22	zaupanje	0,95	22	zaupanje	0,95
23	alfa	=1-B22	23	alfa	0,05
24	$z_{\alpha/2}$	=NORMINV(1-B23/2;0;1)	24	$z_{\alpha/2}$	1,960
25	$p^*_B - p^*_A$	=C6-B6	25	$p^*_B - p^*_A$	0,133
26	delta	=B24*SQRT(B6*(1-B6)/B4+C6*(1-C6)/C4)	26	delta	0,089
27	$l_1$	=B25-B26	27	$l_1$	0,044
28	$l_2$	=B25+B26	28	$l_2$	0,223

#### Rešitev:

S 95 % zaupanjem trdimo, da je učinkovitost zdravila B od 4,4-22,3 odstotnih točk večja kot učinkovitost zdravila A. ♦

#### UČINKOVITOST ZDRAVIL 7

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?

#### Rešitev:

S 99 % zaupanjem trdimo, da je učinkovitost zdravila B od 1,6-25,1 odstotnih točk večja kot učinkovitost zdravila A. ♦

**Domača naloga 8.1: Športniki med študenti in študentkami na BF**

ANKETA 30

Na podlagi podatkov v datoteki ANKETA.xlsx 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:**

$$\hat{p}_M = 0,853 \quad \hat{p}_Z = 0,720 \quad z = 1,652 \quad p = 0,0985 \blacklozenj$$

**Domača naloga 8.2: Štipendisti med študenti in študentkami BT in AG**

ANKETA 31

Na podlagi podatkov v datoteki ANKETA.xlsx preverite domnevo, da je delež štipendistov in štipendistk (ne glede na spol) pri smereh študija AG in BT enak ( $\alpha = 0,05$ ).

**Rešitev:**

$$\hat{p}_{AG} = 0,25 \quad \hat{p}_{BT} = 0,52 \quad z = -2,413 \quad p = 0,0158 \blacklozenj$$

ANKETA 32

Izračunajte 95 % interval zaupanja za razliko med deležem štipendistov in štipendistk (ne glede na spol) pri smereh študija AG in BT.

**Rešitev:**

S 95 % zaupanjem trdimo, da je delež štipendistov AG od 0,046 do 0,494 manjši od deleža štipendistov BT.  $\blacklozenj$

### 8.4 Razlika Bernoullijevih verjetnosti za odvisna vzorca

Imamo dva odvisna vzorca, v katerih analiziramo slučajno spremenljivko  $X$ , ki je v prvi populaciji porazdeljena  $X_1 \sim b(n, p_1)$  in v drugi populaciji  $X_2 \sim b(n, p_2)$ . 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  $N(0, 1)$ .<sup>25</sup> 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	$f_{11} + f_{21}$	$f_{12} + f_{22}$	$n$

#### Zgled 8.6: 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 8.10 in tudi na delovnem listu »vpliv kajenja«.

Tabela 8.10: 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

<sup>25</sup> McNemarjev preizkus simetrije

## 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« (tabela 8.11).

Tabela 8.11: Izračuni za preverjanje ničelne domneve o razliki Bernoullijevih verjetnosti za dva odvisna vzorca; izpis formul (zgoraj); rezultati (spodaj)

	A	B	C	D
2		Bolnik kadi	Bolnik ne kadi	Skupaj
3	Kontrola kadi	20	29	49
4	Kontrola ne kadi	431	10	441
5	Skupaj	451	39	490
6				
7	$H_0: p_{\text{kadi zdravi}} = p_{\text{kadi bolni}}$			
8	test(enostranski, dvostranski)	2		
9	alfa	0,01		
10	z_števec	=C3-B4		
11	z_imenovalec	=SQRT(C3+B4)		
12	z-statistika	=B10/B11		
13	z_mejna	=NORMINV(1-B9/2,0;1)		
14	p-vrednost	=B8*(1-NORMDIST(ABS(B12);0;1;TRUE))		

	A	B
7	$H_0: p_{\text{kadi zdravi}} = p_{\text{kadi bolni}}$	
8	test(enostranski, dvostranski)	2
9	alfa	0,01
10	z_števec	-402,00
11	z_imenovalec	21,45
12	z-statistika	-18,74
13	z_mejna	2,5758
14	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 99 % interval zaupanja za razliko deležev kadilcev pri bolnikih in zdravih (tabela 8.12).  $\blacklozenge$

## 8 Preverjanje statističnih domnev za dva vzorca, interval zaupanja za razliko

---

Tabela 8.12: Izračuni intervala zaupanja za razliko Bernoullijevih verjetnosti za dva odvisna vzorca; izpis formul (levo); rezultati (desno)

	A	B
18	zaupanje	0,99
19	alfa	=1-B18
20	$z_{\text{alfa}/2}$	=NORMINV(1-B19/2;0;1)
21	$(f_{12}-f_{21})/n$	=(C3-B4)/D5
22	delta	=B20*SQRT((C3+B4)/D5^2)
23	$l_1$	=B21-B22
24	$l_2$	=B21+B22

	A	B
18	zaupanje	0,99
19	alfa	0,01
20	$z_{\text{alfa}/2}$	2,576
21	$(f_{12}-f_{21})/n$	-0,820
22	delta	0,113
23	$l_1$	-0,93
24	$l_2$	-0,71

## 9 Linearna regresija in korelacija

### Novi pojmi:

- linearna regresija
- Pearsonov koeficient korelacije
- Spearmanov koeficient korelacije

### Datoteke s podatki:

REGRESIJSKA ANALIZA.xlsx

ANKETA.xlsx

KORELACIJA.xlsx

### 9.1 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 oceni parametrov regresijskega modela  $a$  in  $b$ .

#### Zgled 9.1: Telesna višina in starost otrok

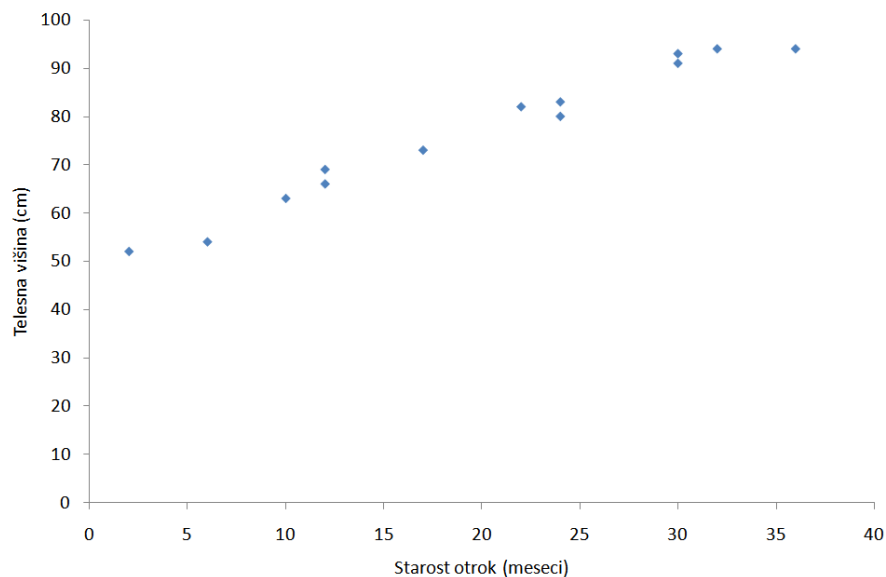
V tabeli 9.1 (datoteka REGRESIJSKA ANALIZA.xlsx, delovni list »otroci«) so podatki o starosti (v mesecih) in telesni višini (cm) za 13 otrok.

OTROCI 1

Narišite sliko, ki prikazuje odvisnost telesne višine od starosti otrok (slika 9.1). ♦

Tabela 9.1: 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 9.1: Odvisnost telesne višine od starosti za 13 otrok

OTROCI 2

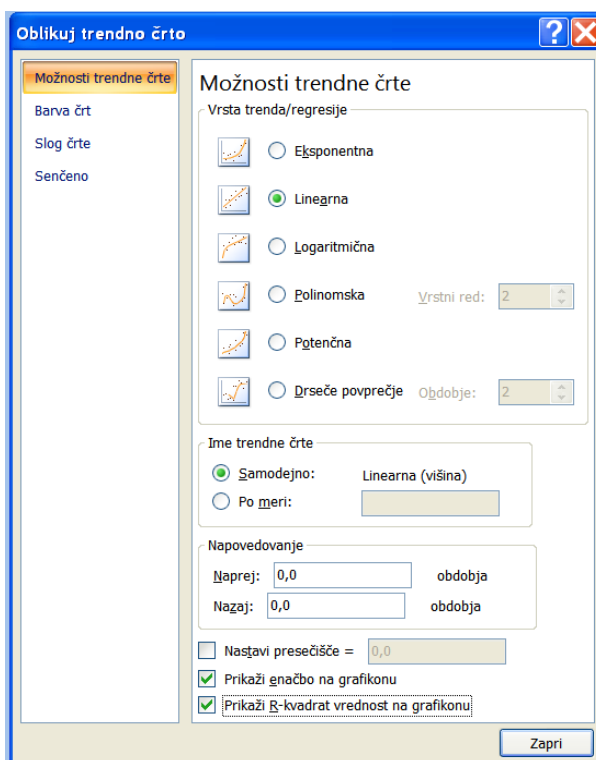
Regresijsko premico dodajte na sliko 9.1. Izračunajte oceni parametrov enostavnega linearnega regresijskega modela in ju obrazložite. Izračunajte in obrazložite koeficient determinacije.

**Navodilo:** če želimo na razsevni grafikon (vrsta grafikona *Raztreseni*) dodati premico za izbrani niz podatkov, ga najprej označimo z levim klikom miške, nato z desnim klikom odpremo seznam ukazov in izberemo ukaz **Dodaj trendno črto**. V pogovornem oknu **Oblikuj trendno črto** izberemo možnost **Linearna**. V spodnjem delu pogovornega okna 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 9.2). Na grafikonu prilagodimo oceni parametrov regresijskega modela na primerno število decimalnih mest.

**Rešitev:**  $a = 49,5$ ,  $b = 1,4$ ,  $r^2 = 98\%$ .

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 pojasnjeno z modelom oz. s starostjo otrok. ♦

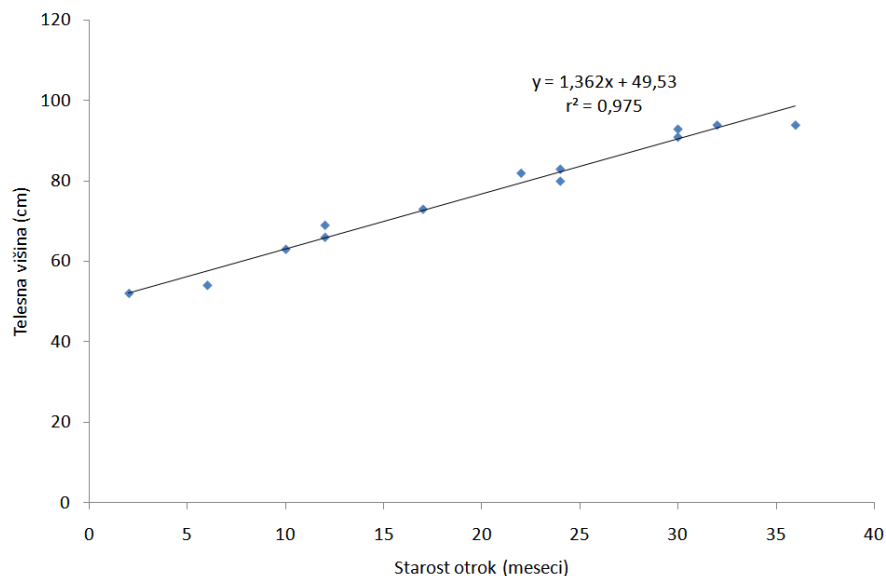


### OTROCI 3

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 9.2: Odvisnost telesne višine od starosti otrok; podatki z dodano regresijsko premico (ukaz **Dodaj trendno črto**)

### 9.1.1 Excelovo orodje za linearno regresijo

Model linearne regresije izračunamo z ukazom **Analiza podatkov/Regression** na kartici **Podatki** v skupini Analiza. V pogovornem oknu **Regression** so naslednja polja za določitev vhodnih podatkov:

	A	B	C	D	E	F	G	H	I	J
	Otrok	Starost (mesece)	Višina (cm)							
1										
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										

**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 drugi interval zaupanja za parametra regresijskega modela, 95 % interval zaupanja se v tabeli izpiše ne glede na to, kaj določimo v tem polju.

Mesto izpisa rezultatov določimo kot pri že opisanih ukazih iz skupine **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 **Analiza podatkov/Regression** (tabela 9.2).

**Rešitev:**  $a = 49,5$  ,  $b = 1,4$  ,  $r^2 = 98\%$  .

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  $\beta$ ).

Ničelno domnevo  $H_0 : \beta = 0$  na podlagi  $F$ -preizkusa iz tabele ANOVA pri  $\alpha = 0,05$  zavrnevo v korist alternativni domnevi  $H_1 : \beta \neq 0$ , ki pravi, da je telesna višina linearno odvisna od starosti otrok ( $p = 0,0000$ ). ♦

Tabela 9.2: Rezultati regresijske analize, narejene z ukazom **Analiza podatkov/Regression** za podatke v datoteki REGRESIJSKA ANALIZA.xlsx na delovnem listu »otroci«. Sivo obarvani stolpci in vrstice s slovenskimi izrazi za izračune v tabelah so dodani.

Razlaga izračunov	Regression Statistics	
koeficient korelacije, r	Multiple R	0,99
koeficient determinacije, r <sup>2</sup>	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			

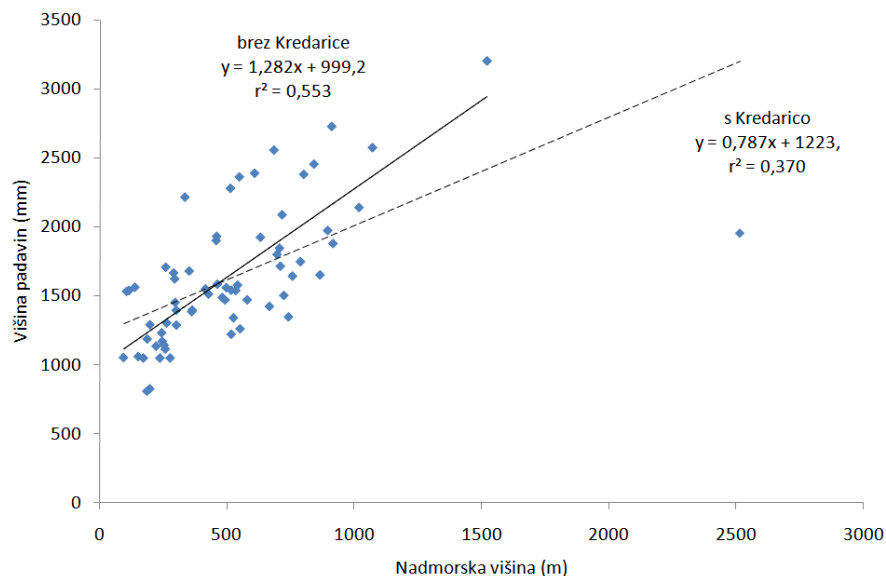
		ocena parametra modela	standardna napaka ocene	t - statistika	p-vrednost	spodnja meja intervala zaupanja	zgornja meja intervala zaupanja
		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
odsek na ordinati	Intercept	49,53	1,45	34,268	1,57E-12	46,35	52,72
	starost	1,36	0,06	21,026	3,12E-10	1,22	1,50

### Zgled 9.2: Padavine v Sloveniji

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 9.4, vsi podatki). Ali je uporaba modela linearne regresije za te podatke upravičena? Naredite linearni regresijski model. ♦



Slika 9.3: 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. Grafično predstavite regresijsko premico, kot kaže slika 9.3. Primerjajte regresijska modela. ♦

**Navodilo:** na sliko 9.3 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 grafikon 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 3

Regresijsko analizo ponovite še enkrat, tako da izločite tudi postajo Dom na Komni. Primerjajte regresijske modele. ♦

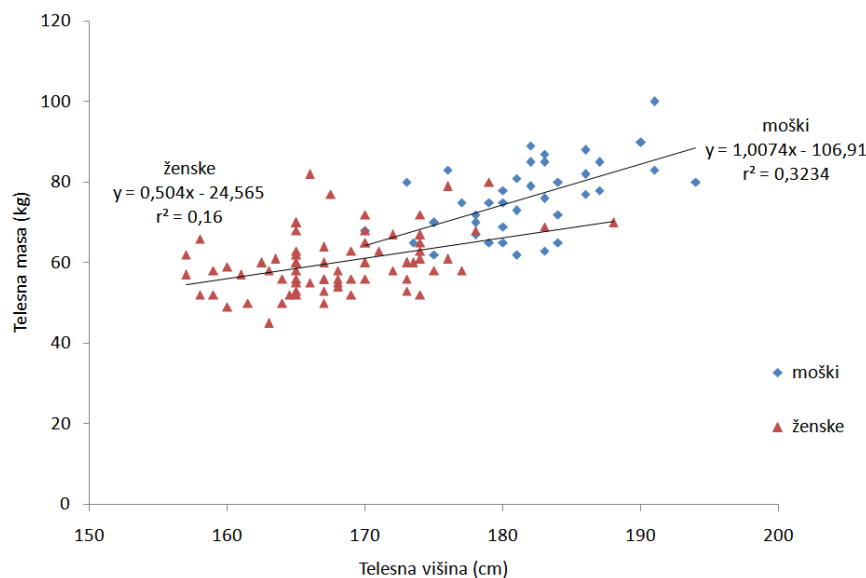
### Domača naloga 9.1: Odvisnost telesne mase od telesne višine

Proučujemo odvisnost telesne mase študentk in študentov AG, BT in MB od telesne višine. Podatki so v datoteki ANKETA.xlsx.

## ANKETA 33

Podatke spremenljivk *spol*, *telesna masa* in *telesna višina* kopirajte v datoteko REGRESIJSKA ANALIZA.xlsx na delovni list »Tmasa\_Tvišina«. Grafično prikažite odvisnost telesne mase od višine posebej za študentke in posebej za študente na istem grafikonu (slika 9.4).

**Navodilo:** kopirane zapise (*spol*, *telesna masa* in *telesna višina*) najprej uredite po spolu (ukaz **Razvrsti**). Vrstni red spremenljivk *telesna masa* in *telesna višina* v datoteki ni ustrezen za direktno uporabo grafikona *Raztreseni*. Zato najprej označite samo celice s *telesno maso* za moške, nato v okviru ukaza **Izberi podatke** določite še vrednosti za vodoravno os (*telesna višina*). Nato na grafikon dodajte še en niz, ki prikazuje podatke za ženske. ♦



Slika 9.4: Odvisnost telesne mase od telesne višine za študente in študentke na BF

## ANKETA 34

Posebej za ženske in posebej za moške izračunajte oceni parametrov regresijskega modela in ju obrazložite. Premici grafično prikažite na sliki, kjer ste prikazali podatke (slika 9.4). Primerjajte rezultate za ženske in moške. ♦

## ANKETA 35

Za oba modela izračunajte koeficienta determinacije in ju obrazložite. ♦

## ANKETA 36

Izračunajte interval zaupanja za naklon premice za vsak model posebej ter vrednosti obrazložite (excelovo orodje).

Navodilo: orodje **Analiza podatkov/Regression** vrne napako v primeru manjkajočih podatkov. Zato je potrebno vrstice z manjkajočimi vrednostmi odstraniti z ukazom **Izbriši**. ♦

**Domača naloga 9.2: Gozdna površina v Sloveniji**

V tabeli 9.3 so podatki za gozdno površino po izbranih letih (Vir: Gozd in gozdarstvo Slovenije, 2004).

Tabela 9.3: Gozdna površina Slovenije po izbranih letih (Vir: Zavod za gozdove Slovenije, 2008)

Leto	Gozdna površina (1000 ha)
1875	737
1947	897
1961	961
1970	1026
1980	1045
1990	1077
2000	1134
2003	1158
2008	1185

## GOZDNA POVRŠINA 1

Podatke iz tabele 9.3 prepisite na delovni list »gozdna površina« v datoteko REGRESIJSKA ANALIZA.xlsx. Izračunajte % gozda v Sloveniji po letih. Za velikost Slovenije upoštevajte vrednost 20 000 km<sup>2</sup>, 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) za obdobje 1961-2008. Obrazložite rezultate. ♦

## 9.2 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 oziroma so podatki ene ali obeh spremenljivk rangi ali so 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, da povezanosti ni ( $H_0 : \rho = 0$ ). Izračunamo testno statistiko:

$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$ , katere ničelna porazdelitev je Studentova porazdelitev  $t(SP = n - 2)$ . Za

Spearmanov koeficient korelacije lahko ta preizkus uporabimo le za velike vzorce, pri majhnih vzorcih je ničelna porazdelitev drugačna.

Pearsonov koeficient korelacije izračunata dve funkciji `CORREL(array1; array2)` ali pa `PEARSON(array1; array2)`. Za argumenta ene ali druge 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 **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, le da za argumente funkcije CORREL uporabimo 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 9.3: Vsebnost žvepla in dušika v lišajih

Dva izmed pokazateljev splošne onesnaženosti zraka sta vsebnost dušika in vsebnost žvepla v drevesnih lišajih (tabela 9.4, datoteka KORELACIJA.xlsx, delovni list »lišaji«).

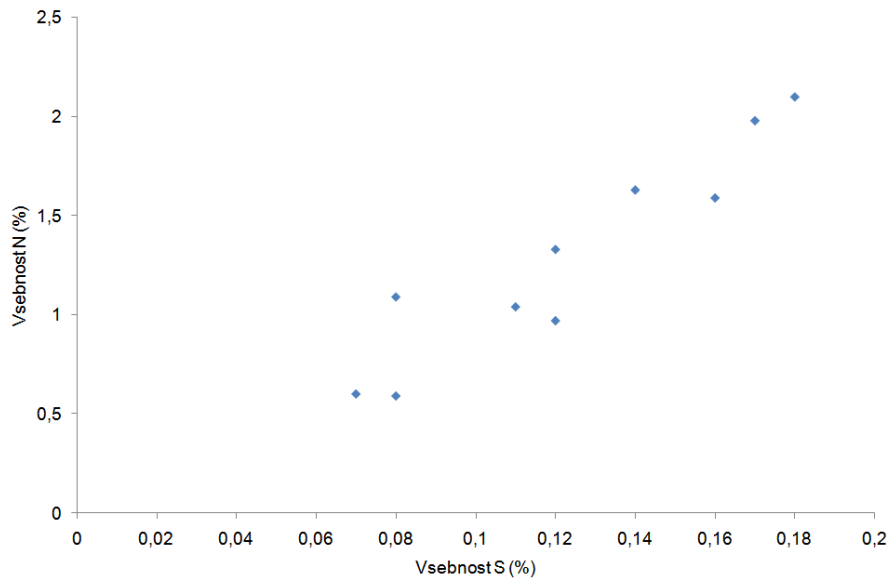
Tabela 9.4: Vsebnost dušika (N) in žvepla (S) v drevesnih lišajih za deset lokacij v zahodni Sloveniji

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

Grafično prikažite povezanost N in S v lišajih (slika 9.5). ♦





Slika 9.5: Razsevni grafikon za povezanost vsebnosti S in N v lišajih

LIŠAJI 2

Katera mera povezanosti je primerna za take podatke? Izračunajte jo. ♦

LIŠAJI 3

Preverite domnevo, da ni povezave med vsebnostjo žvepla in vsebnostjo dušika v lišajih (stopnja značilnosti je 0,01)(tabela 9.5). Obrazložite rezultate.

Tabela 9.5: 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

Rešitev:  $H_0 : \rho = 0$  in  $H_1 : \rho \neq 0$

$r = 0,94$

$t$ -preizkus:  $t = 7,724$  in  $p = 0,0001$ .

Ničelno domnevo pri  $\alpha = 0,01$  zavrnamo. Linearna povezanost med vsebstjo žvepla in dušika v lišajih je statistično značilna ( $p = 0,0001$ ). ♦

#### Zgled 9.4: 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), svineca (Pb) in bakra (Cu). Podatki so v tabeli 9.6 in v datoteki KORELACIJA.xlsx, delovni list »kovine«.

Tabela 9.6: 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 osamelec 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 **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.

**Rešitev:**  $r_s = 0,89$  ◆

## 10 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.xlsx

ANKETA.xlsx

### 10.1 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  $\chi^2 \approx \chi_\alpha^2(SP = k - 1)$ .  $H_0$  obdržimo, če je  $\chi^2 \leq \chi_\alpha^2(SP = k - 1)$  in zavrnemo v korist  $H_1$ , če je  $\chi^2 > \chi_\alpha^2(SP = k - 1)$ . Alternativna domneva je torej enostranska.

Pogoj za uporabo Pearsonove hi-kvadrat statistike je:  $f_i' \geq 5$ ,  $i = 1, \dots, k$ .

### Zgled 10.1: 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.xlsx na delovnem listu »vhod«. Koliko od 200 obiskovalcev trgovine bi vstopilo skozi vsaka od vrat, če predpostavimo, da je uporaba vseh treh vhodov enako verjetna (tabela 10.1)? ♦

#### VHOD 2

Preverite ničelno domnevo, da je verjetnost, da vstopijo skozi katerikoli vhod, enaka ( $\alpha = 0,05$ ). Obrazložite rezultate.

**Navodilo:** nalogo rešite v pripravljeni tabeli na delovnem listu »vhod«. Pri izračunu uporabite funkcije CHIINV, CHIDIST in CHITEST (tabela 10.1). 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  $\alpha$ ); »deg\_freedom« so stopinje prostosti  $SP$  (parameter hi-kvadrat porazdelitve). Funkcija vrne vrednost hi-kvadrat statistike  $\chi_1^2$ , 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.

$$\chi^2 = 6,19 \quad SP = 2 \quad p = 0,0453 \text{ (tabela 10.1)}$$

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 10.1: 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 <sub>mejna</sub> CHIINV	=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 <sub>mejna</sub> CHIINV	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 10.2: Dedovanje mikrosatelitnih lokusov pri hmelju**

Za izdelavo genetske karte hmelja potrebujemo družino križancev, ki ponavadi izhajajo iz križanja dveh staršev z zelo različnimi lastnostmi. Izvedli so kontrolirano križanje med žensko rastlino angleške sorte Wye Target in moško rastlino hmelja 2/1, ki je slovenskega porekla. Navajamo podatke za dva lokusa CO653996-TGAG4 in CO653957-TCCCA4 za 144 rastlin. Za vsako rastlino so dobili genotipske vrednosti (aa, bb, ab, ac, bc) za posamezni lokus. Pričakovan model dedovanja za izbrana lokusa je podan v tabeli 10.2.

Tabela 10.2: Pričakovani modeli dedovanja na posameznih lokusih F1 generacije križancev hmelja.

Lokus	Starševski genotipi	Genotipi potomcev	Predvideno dedovanje
CO653996-TGAG4	ab × ab	aa, ab, bb	1:2:1
CO653957-TCCCA4	ab × ac	aa, ab, ac, bc	1:1:1:1

Rezultati meritev so na delovnem listu »hmelj« v datoteki HI KVADRAT.xlsx. Enota podatkovne zbirke je rastlina (v stolpcu A so zaporedne številke, v stolpcu B pa oznake rastlin), za vsako rastlino je podan genotip na posameznem lokusu (stolpca C in D). Če genotip ni bil določen, je v celici oznaka »--«.

Pred izdelavo genetske karte hmelja je potrebno preveriti, če je dedovanje znotraj posameznih lokusov v skladu s pričakovanji.

HMELJ 1

Z uporabo vrtilne tabele na delovnem listu »hmelj« za vsak lokus posebej ugotovite pogostost pojavljanja posameznega genotipa. ♦

	A	B	C	D	E	F	G
1							
2							
3	CO653996-TGAG4	Štej od Rastlina					
4	--	4					
5	aa	31					
6	ab	69					
7	bb	40					
8	<b>Skupna vsota</b>	<b>144</b>					
9							
10							
11							
12							
13							
14							
15							

**Seznam polj vrtilne tabele**

Izberite polja, ki jih želite dodati poročilu:

- Rastlina
- CO653996-TGAG4
- CO653957-TCCCA4

Povlecite polja med spodnjimi območji:

Filter poročila       Oznake stolpcev

Oznake vrstic       Vrednosti

CO653996-TGAG4      Štej od Rastlina

Odloži posodobitev postavitev      Posodobi

## HMELJ 2

Glede na predvideno dedovanje za vsak lokus izračunajte pričakovane frekvence genotipov. Upoštevajte rastline, na katerih je bil genotip posameznega lokusa določen (tabela 10.3). ♦

## HMELJ 3

Za vsak lokus preverite domnevo, da je dedovanje v skladu s pričakovanji ( $\alpha = 0,05$ ). Obrazložite rezultate.

**Rešitev:** za lokus CO653996-TGAG4

$$H_0 : p_{aa} : p_{ab} : p_{bb} = 1 : 2 : 1$$

$H_1$  : ničelna domneva ne velja.

$$\chi^2 = 1,186 \quad SP = 2 \quad p = 0,5527 \text{ (tabela 10.3)}$$

Ničelno domnevo pri  $\alpha = 0,05$  obdržimo. Rezultati križanja za lokus CO653996-TGAG4 so v skladu s pričakovanji.

Za lokus CO653957-TCCCA4:

$$H_0 : p_{aa} : p_{ab} : p_{ac} : p_{bc} = 1 : 1 : 1 : 1$$

$H_1$  : ničelna domneva ne velja.

$$\chi^2 = 19,118 \quad SP = 3 \quad p = 0,0003$$



Ničelno domnevo pri  $\alpha = 0,05$  zavrnamo. Podatki nakazujejo, da se pri lokusu CO653957-TCCCA4 genotipska vrednost "aa" pojavi manjkrat, genotipska vrednost "bc" pa večkrat, kot bi pričakovali v skladu s predvidenim dedovanjem ( $p = 0,0003$ ). ♦

Tabela 10.3: Izračuni za hi-kvadrat preizkus ničelne domneve o predvidenem dedovanju hmelja za lokus CO653996-TGAG4; izpis formul (zgoraj); izpis rezultatov (spodaj)

	A	B	C	D	E
1	CO653996-TGAG4	Število rastlin	Predvideno dedovanje	Pričakovane frekvence	Prispevki k hi-kvadrat statistiki
2	--	4			
3	aa	31	1	=C3*\$B\$7/\$C\$6	=(B3-D3)^2/D3
4	ab	69	2	=C4*\$B\$7/\$C\$6	=(B4-D4)^2/D4
5	bb	40	1	=C5*\$B\$7/\$C\$6	=(B5-D5)^2/D5
6	Skupaj	144	=SUM(C3:C5)	hi-kvadrat statistika	=SUM(E3:E5)
7	Število določenih rastlin	=(B6-B2)		alfa	0,05
8				SP=( število genotipskih vrednosti - 1)	2
9				hi-kvadrat <sub>mejna</sub> CHIINV	=CHIINV(E7;E8)
10				p-vrednost CHIDIST	=CHIDIST(E6,E8)
11				p-vrednost CHITEST	=CHITEST(B3:B5;D3:D5)

	A	B	C	D	E
1	CO653996-TGAG4	Število rastlin	Predvideno dedovanje	Pričakovane frekvence	Prispevki k hi-kvadrat statistiki
2	--	4			
3	aa	31	1	35	0,457
4	ab	69	2	70	0,014
5	bb	40	1	35	0,714
6	Skupaj	144	4	hi-kvadrat statistika	1,186
7	Število določenih rastlin	140		alfa	0,05
8				SP=( število genotipskih vrednosti - 1)	2
9				hi-kvadrat <sub>mejna</sub> CHIINV	5,991
10				p-vrednost CHIDIST	0,5527
11				p-vrednost CHITEST	0,5527

### Zgled 10.3: Porazdelitev premera debel

V datoteki PREMER DEBLA.xlsx in v datoteki HI KVADRAT.xlsx na delovnem listu »premer debla« je frekvenčna tabela za premer debel dreves. Prvi in zadnji razred sta v tej nalogi odprta navzdol oz. navzgor.

PREMER DEBLA 1

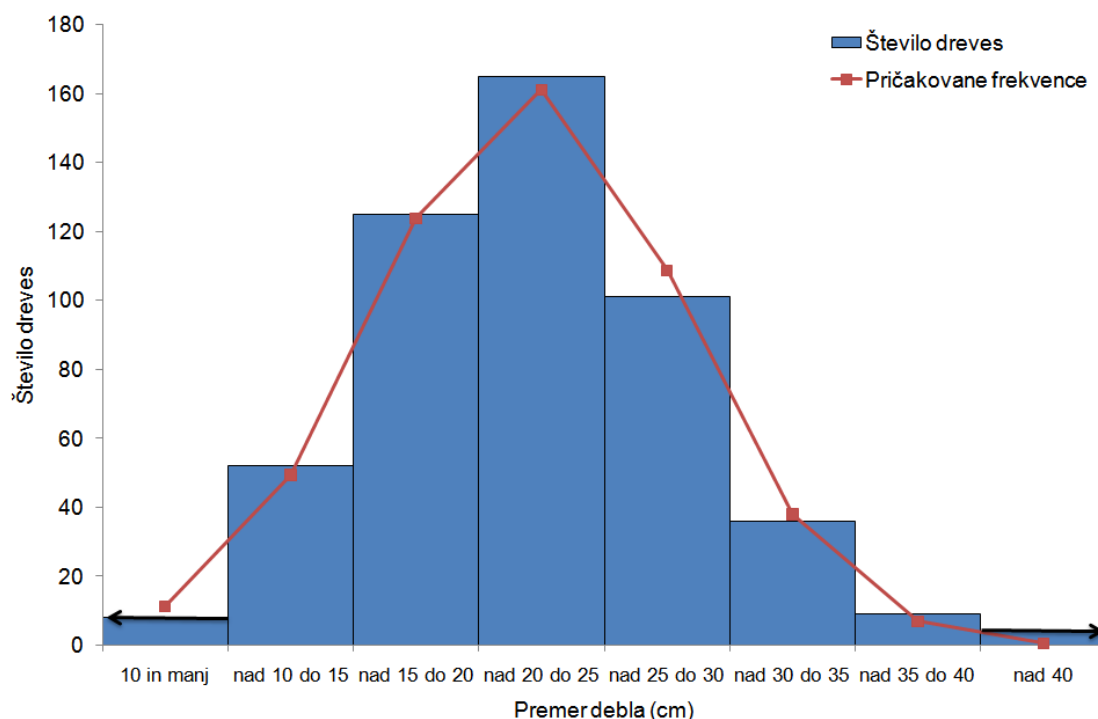
Ob predpostavki, da je premer debla dreves porazdeljen normalno  $N(22, 6)$ , izračunajte verjetnost, da je premer debla v posameznem razredu empirične porazdelitve (tabela 10.5). ♦

## PREMER DEBLA 2

Koliko od 500 dreves bi bilo v posameznem razredu, če bi bil premer debla porazdeljen po normalni porazdelitvi  $N(22, 6)$ ? Dobljene pričakovane frekvence primerjajte z dejanskimi (tabela 10.5). ♦

## PREMER DEBLA 3

Grafično prikažite dejanske in pričakovane frekvence s histogramom in poligonom, kot kaže slika 10.1. ♦



Slika 10.1: Empirična porazdelitev dreves glede na premer debla (histogram, puščici označujeta odprte razrede); poligon predstavlja pričakovane frekvence po normalni porazdelitvi  $N(22, 6)$  za posamezni razred

## PREMER DEBLA 4

Ali empirična porazdelitev premera debla potrjuje predpostavko, da je premer debla porazdeljen normalno s parametroma  $\mu = 22$  cm in  $\sigma = 6$  cm,  $\alpha = 0,05$  (tabeli 10.4 in 10.5)? 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 10.5).

**Rešitev:**

$H_0$ : verjetnostna porazdelitev premera debla 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 premera debla je v skladu s predpostavko, da za premer debla lahko privzamemo normalno porazdelitev  $N(22, 6)$ .

**PREMER DEBLA 5**

Kakšen bi bil statistični sklep, če razredov ne bi združili (kar bi bilo narobe)? Kaj je vzrok za napačen statističen sklep? ♦

Tabela 10.4: Izračuni za hi-kvadrat preizkus ničelne domneve o porazdelitvi premera debla; izpis formul (zgoraj); izpis rezultatov (spodaj)

	A	B	C	D	E	F	G	H	I	J
1	Štev. razreda	Premer debla (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	1	10 in manj	8		10	0	=NORMDIST(E2;\$C\$12;\$C\$13;TRUE)	=G2-F2	=H2*\$C\$10	=(C2-12)^2/12
3	2	nad 10 do 15	52	10	15	=NORMDIST(D3;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E3;\$C\$12;\$C\$13;TRUE)	=G3-F3	=H3*\$C\$10	=(C3-13)^2/13
4	3	nad 15 do 20	125	15	20	=NORMDIST(D4;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E4;\$C\$12;\$C\$13;TRUE)	=G4-F4	=H4*\$C\$10	=(C4-14)^2/14
5	4	nad 20 do 25	165	20	25	=NORMDIST(D5;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E5;\$C\$12;\$C\$13;TRUE)	=G5-F5	=H5*\$C\$10	=(C5-15)^2/15
6	5	nad 25 do 30	101	25	30	=NORMDIST(D6;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E6;\$C\$12;\$C\$13;TRUE)	=G6-F6	=H6*\$C\$10	=(C6-16)^2/16
7	6	nad 30 do 35	36	30	35	=NORMDIST(D7;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E7;\$C\$12;\$C\$13;TRUE)	=G7-F7	=H7*\$C\$10	=(C7-17)^2/17
8	7	nad 35 do 40	9	35	40	=NORMDIST(D8;\$C\$12;\$C\$13;TRUE)	=NORMDIST(E8;\$C\$12;\$C\$13;TRUE)	=G8-F8	=H8*\$C\$10	=(C8-18)^2/18
9	8	nad 40	4	40		=NORMDIST(D9;\$C\$12;\$C\$13;TRUE)	1	=G9-F9	=H9*\$C\$10	=(C9-19)^2/19
10		Skupaj	=SUM(C2:C9)					=SUM(H2:H9)	=SUM(I2:I9)	=SUM(J2:J9)
11							hi-kvadrat statistika		=I10	
12		mi	22				alfa		=0,05	
13		sigma	6				SP		=COUNT(C2:C9)-1	
14							mejna vrednost hi-kvadrat statistike		=CHIINV(I12;J13)	
15							p-vrednost izračunana s funkcijo CHIDIST		=CHIDIST(I11;J13)	
16							p-vrednost izračunana s funkcijo CHITEST		=CHITEST(C2:C9;I2:I9)	

	A	B	C	D	E	F	G	H	I	J
1	Štev. razreda	Premer debla (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	1	10 in manj	8		10	0,0000	0,0228	0,0228	11,38	1,00
3	2	nad 10 do 15	52	10	15	0,0228	0,1217	0,0989	49,46	0,13
4	3	nad 15 do 20	125	15	20	0,1217	0,3694	0,2478	123,88	0,01
5	4	nad 20 do 25	165	20	25	0,3694	0,6915	0,3220	161,01	0,10
6	5	nad 25 do 30	101	25	30	0,6915	0,9088	0,2173	108,66	0,54
7	6	nad 30 do 35	36	30	35	0,9088	0,9849	0,0761	38,04	0,11
8	7	nad 35 do 40	9	35	40	0,9849	0,9987	0,0138	6,89	0,65
9	8	nad 40	4	40		0,9987	1,0000	0,0013	<b>0,67</b>	16,38
10		Skupaj	500					1,0000	500,0000	18,9170
11							hi-kvadrat statistika			18,917
12		mi	22				alfa			0,05
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 10.5: Izpis rezultatov hi-kvadrat preizkusa ničelne domneve o prileganju normalne porazdelitve N(22, 6) empirični porazdelitvi premera debla; zadnja dva razreda sta združena

	A	B	C	D	E	F	G	H	I	J
18	Štev. razreda	Premer debla (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	0,0000	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		0,9849	1,0000	0,0151	7,57	3,90
26		Skupaj	500					1,0000	500,0000	5,7951
27							hi-kvadrat statistika			5,795
28							alfa			0,05
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

PREMER DEBLA 6

Kakšni bi bili rezultati hi-kvadrat preizkusa, če bi predpostavili, da je premer debla dreves porazdeljen normalno s parametroma  $\mu = 23$  cm in  $\sigma = 7$  cm?

**Navodilo:** katere vrednosti v tabelah 10.4 in 10.5 moramo spremeniti, da se formule pravilno preračunajo? Ali je potrebno tudi v tem primeru razrede združiti?

**Rešitev:**  $k = 7$   $\chi^2 = 18,255$  in  $p = 0,0056$  ♦

PREMER DEBLA 7

Po spremembi vrednosti za  $\mu$  in  $\sigma$  v prejšnji nalogi si oglejte ujemanje dejanskih in pričakovanih frekvenc na sliki 10.1. ♦

**Domača naloga 10.1: Porazdelitev mase lešnikov**

V datoteki HI KVADRAT.xlsx 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	K	L
1												
2												
3	Masa (g)	Štej od Masa (g)										
4	1,7-1,95	2										
5	1,95-2,2	7										
6	2,2-2,45	23										
7	2,45-2,7	23										
8	2,7-2,95	14										
9	2,95-3,2	1										
10	Skupna vsota	70										
11												
12												
13												

LEŠNIKI 2

Predpostavimo, da je masa lešnikov porazdeljena normalno  $N(2,5, 0,25)$ . Izračunajte verjetnost za posamezni razred 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  $p = 0,9144$ . ♦

**Zgled 10.4: Kakovost izdelanih zgoščenk (CD-jev)**

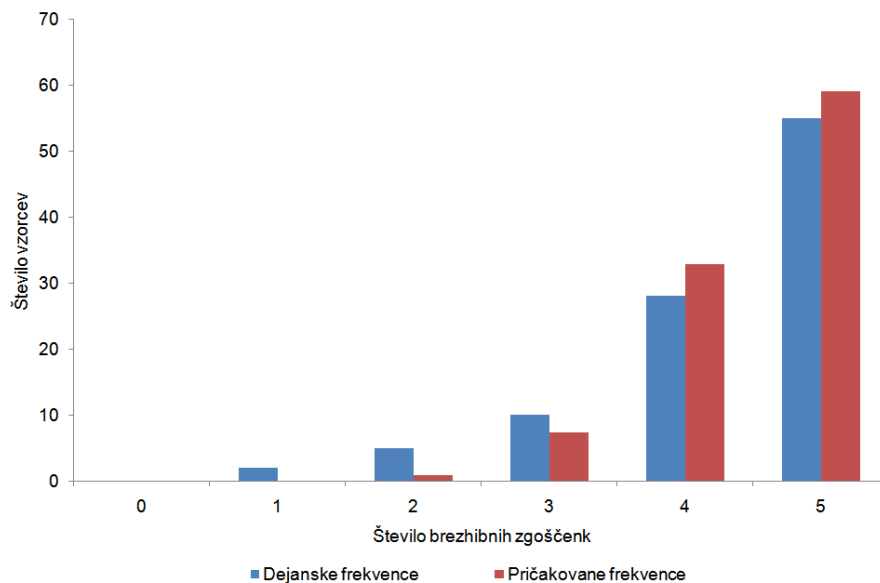
V tovarni zgoščenk trdijo, da je verjetnost, da je izdelana zgoščenska brezhibna, enaka 0,9. V oddelku za kontrolo kakovosti izberejo s slučajno izbiro v kontrolni vzorec 5 zgoščenk. Za vsako zgoščenko ugotovijo, ali je brezhibna ali ne. V kontrolnem vzorcu je lahko 0, 1, 2, 3, 4, 5 brezhibnih zgoščenk. 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 10.2).

♦



Slika 10.2: Število 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 10.6).

**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 10.6). Pričakovane frekvence dodajte na sliko 10.2 (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 10.6).

Tabela 10.6: 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
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	108,5979441
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	0,08	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	10,60497152
19		hi-kvadrat statistika			10,605
20		alfa			0,01
21		SP			2
22		mejna vrednost hi-kvadrat statistike			9,210
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 zavrnemo. 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:** 
$$\hat{p} = \frac{1 \cdot 2 + 2 \cdot 5 + 3 \cdot 10 + 4 \cdot 28 + 5 \cdot 55}{500} = 0,858$$

Ocena tovarne, da je verjetnost brezhibne zgoščenke 0,9, je precenjena. ♦

## 10.2 Hi-kvadrat za preverjanje povezanosti opisnih spremenljivk

Imamo dve opisni spremenljivki,  $X$  ima  $k$  vrednosti,  $Y$  pa  $r$  vrednosti. V vzorcu je  $n$  enot, za vsako enoto imamo vrednost za  $X$  in vrednost za  $Y$ . Podatke uredimo v dvorazsežno kontingenčno tabelo:

$X$	$Y$					Skupaj
	$y_1$	...	$y_j$	...	$y_r$	
$x_1$	$f_{11}$		$f_{1j}$		$f_{1r}$	$f_{1\bullet}$
...						
$x_i$	$f_{i1}$		$f_{ij}$		$f_{ir}$	$f_{i\bullet}$
...						
$x_k$	$f_{k1}$		$f_{kj}$		$f_{kr}$	$f_{k\bullet}$
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:  $n = \sum_{i=1}^k \sum_{j=1}^r f_{ij}$ .

$H_0$ :  $X$  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 10.5: Nastanitev in vožnja z avtom za študente BF

V datoteki ANKETA.xlsx 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 37

Z uporabo vrtilne tabele v datoteki ANKETA.xlsx 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					
2					
3	Štej od ID	avto			
4	nastanitev	da	ne	obcasno	Skupna vsota
5	doma	23	10	18	51
6	lastnik		1		1
7	podnajemnik	3	29	8	40
8	studentski dom	2	28	2	32
9	(prazen)	2			2
10	Skupna vsota	30	68	28	126

## ANKETA 38

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					
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	<b>Skupna vsota</b>	<b>28</b>	<b>68</b>	<b>28</b>	<b>124</b>

Vrednosti tako oblikovane vrtilne tabele parkirajte v datoteko HI KVADRAT.xlsx na delovni list »nastanitev\_avto«. ♦

## ANKETA 39

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 10.7)?

**Rešitev:**

$H_0$  : *nastanitev* in *avto* nista povezani spremenljivki

$H_1$  : *nastanitev* in *avto* sta povezani spremenljivki

$\chi^2 = 45,09$        $p = 0,0000$  (tabela 10.8)

Ničelno domnevo pri  $\alpha = 0,05$  zavrnilo 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 10.7: 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	študentski 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	študentski 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(B13:D13)
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	študentski 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	študentski 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	študentski dom	3,78	6,22	3,78	13,78
20	Skupaj	18,60	19,26	7,23	45,09

Tabela 10.8: Hi-kvadrat preizkus za povezanost spremenljivk *nastanitev* in *avto*; izpis formul (levo); izpis rezultatov (desno)

	A	B		B
23	$H_0$ : <i>nastanitev</i> in <i>avto</i> nista povezani spremenljivki		23	
24	$H_1$ : <i>nastanitev</i> in <i>avto</i> sta povezani spremenljivki		24	
25	alfa	0,05	25	0,05
26	hi-kvadrat statistika	=E20	26	45,09
27	število vrednosti spremenljivke <i>nastanitev</i> ( $k$ )	3	27	3
28	število vrednosti spremenljivke <i>avto</i> ( $r$ )	3	28	3
29	SP	=(B27-1)*(B28-1)	29	4
30	hi_kvadrat <sub>mejna</sub>	=CHIINV(B25;B29)	30	9,49
31	p-vrednost CHIDIST	=CHIDIST(B26;B29)	31	3,8E-09
32	p-vrednost CHITEST	=CHITEST(B3:D5;B10:D12)	32	3,8E-09

### 10.3 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  $\chi^2 \approx \chi_\alpha^2(SP = (k-1)(r-1))$ .  $H_0$  obdržimo, če je  $\chi^2 \leq \chi_\alpha^2(SP = (k-1)(r-1))$  in zavrnilo v korist  $H_1$ , če je  $\chi^2 > \chi_\alpha^2(SP = (k-1)(r-1))$ .

Pogoj za uporabo Pearsonove hi-kvadrat statistike je:  $f'_{ij} \geq 5$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, r$ .

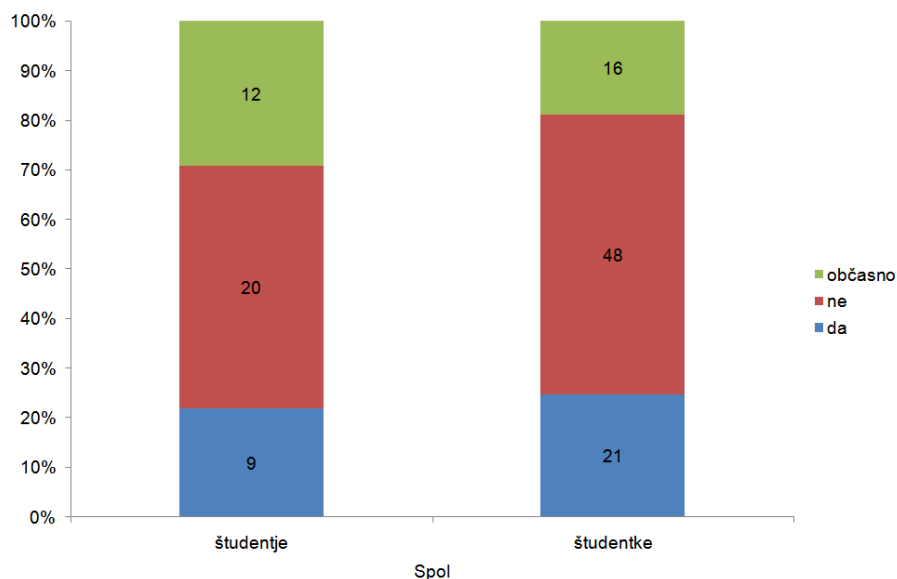
### Zgled 10.6: Struktura študentov in študentk glede na prevoz z avtom

#### ANKETA 40

Na podlagi podatkov v datoteki ANKETA.xlsx, 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.xlsx (tabela 10.9). ♦

#### ANKETA 41

S strukturnima stolpcema prikažite strukturo študentk in študentov glede na prevoz na fakulteto z avtom (slika 10.3). ♦



Slika 10.3: Struktura študentov in študentk glede na to, ali se na fakulteto vozijo z avtom. Številke na stolpcih predstavljajo frekvence.

#### ANKETA 42

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 10.9). 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 10.9: 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				
8	Pričakovane frekvence	spol		
9	avto	študentje	študentke	Skupaj
10	da	9,8	20,2	30,0
11	ne	22,1	45,9	68,0
12	občasno	9,1	18,9	28,0
13	Skupaj	41,0	85,0	126,0
14				
15	Prispevki k hi-kvadrat statistiki	spol		
16	avto	študentje	študentke	Skupaj
17	da	0,06	0,03	0,09
18	ne	0,20	0,10	0,30
19	občasno	0,92	0,44	1,36
20	Skupaj	1,18	0,57	1,75
21				
22	$H_0$ : za študente in študentke je struktura po tem, ali se na fakulteto vozijo z avtom, enaka			
23	$H_1$ : za študente in študentke struktura po tem, ali se na fakulteto vozijo z avtom, ni enaka			
24	alfa	0,05		
25	hi-kvadrat statistika	1,75		
26	število vrednosti spremenljivke <i>avto</i> ( $k$ )	3		
27	število vrednosti spremenljivke <i>spol</i> ( $r$ )	2		
28	SP	2		
29	hi_kvadrat <sub>mejna</sub>	5,99		
30	p-vrednost CHIDIST	0,4171		
31	p-vrednost CHITEST	0,4171		

### Domača naloga 10.3: Struktura študentov AG, BT in MB glede na to, ali prejemajo štipendijo.

ANKETA 43

Na podlagi podatkov v datoteki ANKETA.xlsx 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.xlsx 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 44

Strukturo treh smeri študija glede na to, ali imajo štipendijo ali ne, grafično predstavite. Delovni list v datoteki HI KVADRAT.xlsx poimenujte »Slika 10.4«. ♦



## 11 VIRI

Kastelec D., Košmelj K., 2009. Statistična analiza podatkov s programoma Excel 2003 in R. Ljubljana: Biotehniška fakulteta, URL <http://www.bf.uni-lj.si/agronomija/oddelku/katedre-in-druge-org-enote/za-genetiko-biotehnologijo-statistiko-in-zlahtnjenje-rastlin/statistika/sap-2003.html>.


Košmelj K., 2007. Uporabna statistika, URL <http://www.bf.uni-lj.si/agronomija/oddelku/katedre-in-druge-org-enote/za-genetiko-biotehnologijo-statistiko-in-zlahtnjenje-rastlin/statistika.html>


Microsoft Office Excel 2007, URL <http://office.microsoft.com/en-us/excel/default.aspx>

Microsoft Office Excel 2007, URL <http://office.microsoft.com/sl-si/?CTT=97>

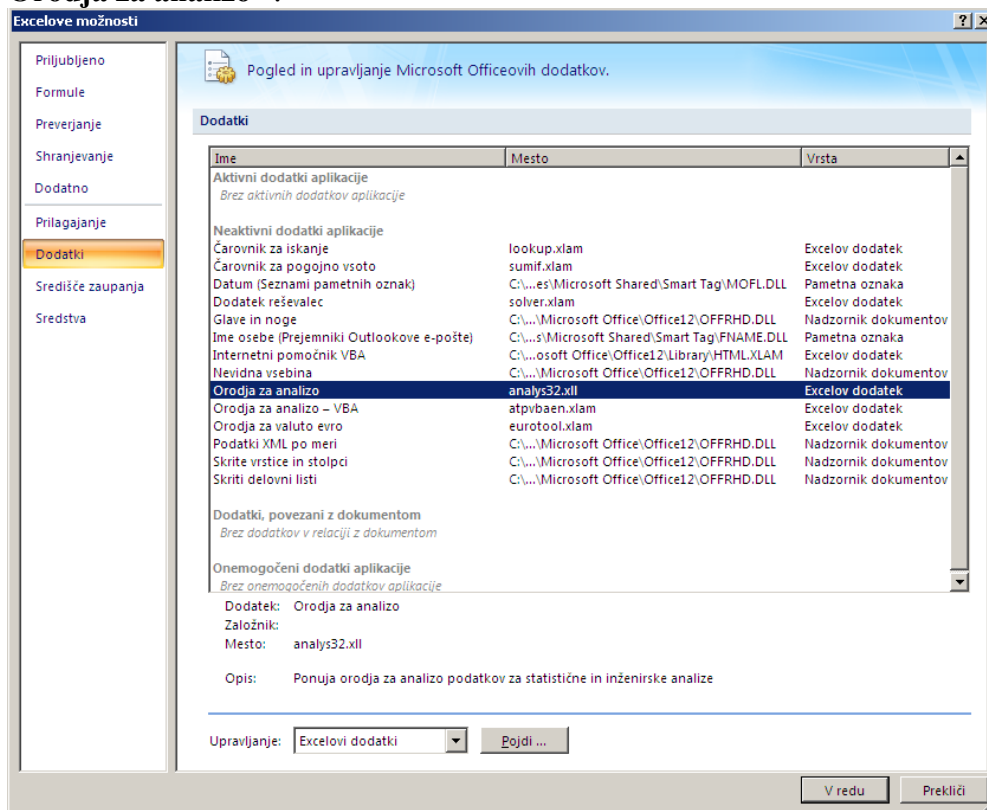
## 12 DODATEK A

### Namestitev orodja **Analiza podatkov**

Kliknemo na gumb Office  in nato v odprtem pogovornem oknu še na gumb

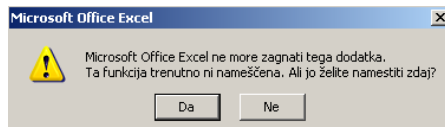
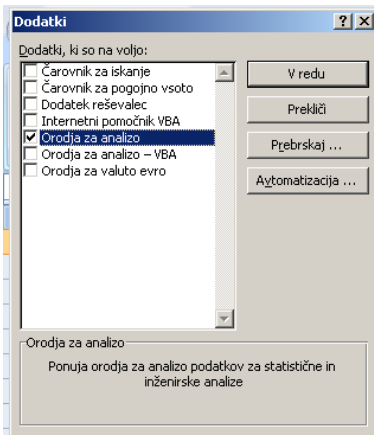
 Excelove možnosti

. Izberemo skupino ukazov *Dodatki*, na desnem seznamu nato izberemo **Orodja za analizo**<sup>26</sup>:

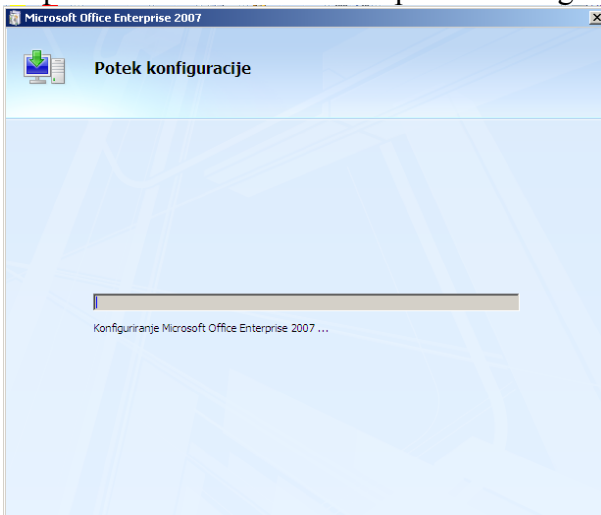


Nadaljujemo s klikom na gumb **Pojdi**. Odpre se pogovorno okno, v katerem ponovno izberemo možnost **Orodja za analizo**. Excel nas opozori, da dodatka ne more zagnati in nas vpraša, ali naj funkcijo namesti. Kliknemo na gumb **Da**.

<sup>26</sup> Če je bilo to orodje pred tem že nameščeno, ga na tem seznamu ni.



Odpre se okno z obvestilom o poteku konfiguracije.



Po končani namestitvi orodja se na kartici **Podatki** pojavi skupina Analiza z ukazom **Analiza podatkov**:

