SVEUČILIŠTE U SPLITU

SVEUČILIŠNI ODJEL ZA STRUČNE STUDIJE

POSLOVNA STATISTIKA

LABORATORIJSKE VJEŽBE U MS EXCELU

Nada Roguljić

Split, 2019.

PREDGOVOR

Poslovna statistika jedan je od temeljnih zajedničkih kolegija koji se izvodi na prvoj godini preddiplomskih studija *Računovodstva i financija* te *Menadžmenta trgovine i turizma* na Sveučilišnom odjelu za stručne studije Sveučilišta u Splitu. Kolegij je zamišljen kao uvodni statistički kolegij u kojem će studenti steći osnove statističke pismenosti, ovladati osnovnim metodama deskriptivne statistike, osnovama regresijske analize, analize vremenskih nizova te trend modela. Izvanredni studenti specijalističkog studija *Menadžmenta trgovine i turizma* u završnom semestru studija slušaju kolegij **Statističke metode** koji se nastavlja na ovaj kolegij.

Obzirom da studenti na prvoj godini prolaze i osnovnu informatičku naobrazbu koja uključuje i upoznavanje s programskim paketom MS Office, posebice MS Excel-om, izbor MS Excela kao alata za statističku analizu posve je očit. Studenti *Računovodstva i financija* u brojnim stručnim kolegijima na višim godinama studija nastavljaju koristiti Excel kao programsku podršku. Isto vrijedi i za studente specijalističkog studija *Menadžmenta trgovine i turizma*.

Ovaj materijal zamišljen je kao praktikum za laboratorijske vježbe kako bi studenti mogli samostalno, uz manju asistenciju nastavnika, kroz riješene primjere proći kroz gradivo i postići željene obrazovne ishode. Na kraju svake laboratorijske vježbe nekoliko je predloženih zadataka za vježbu. Primjeri iz ovog materijala mogu se naći na Moodle stranicama kolegija.

Materijal je nastao kao plod višegodišnjeg iskustva u izvođenju kolegija. Posebno se zahvaljujem umirovljenoj kolegici mr.sc. Karmeli Mikelić uz koju sam stekla dragocjena iskustva.

SADRŽAJ

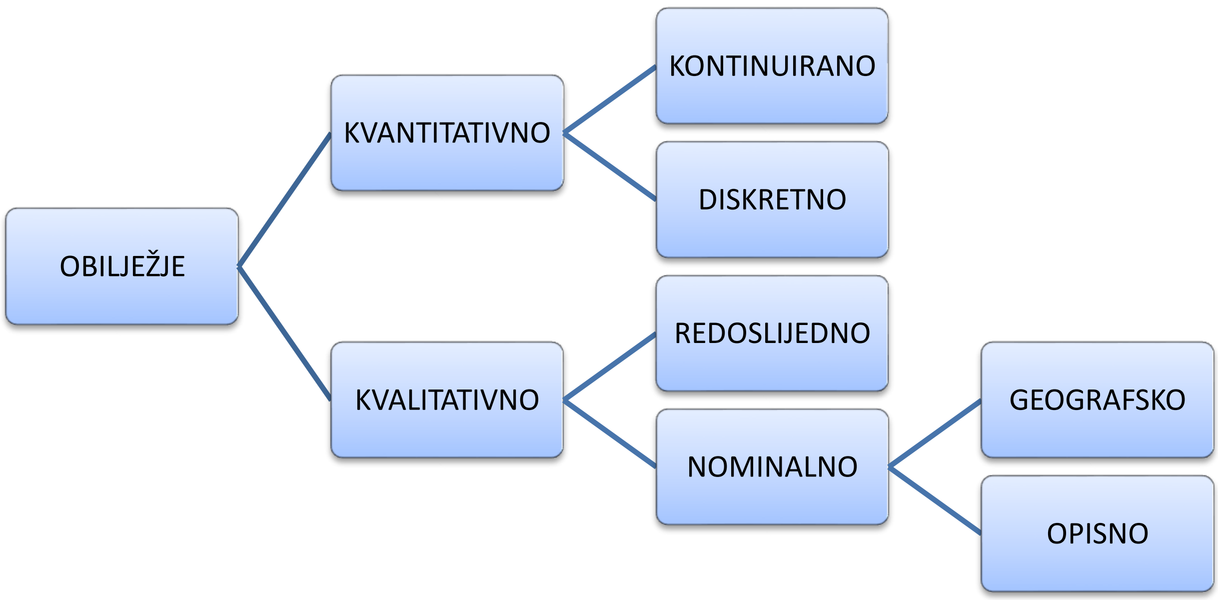
Uvodne vježbe – osnovni statistički pojmovi

Statistički skup tvore statističke jedinice (ljudi, predmeti, pojave) koje se odlikuju sličnim, zajedničkim svojstvima. Upravo ta svojstva, **obilježja**, su predmet statističkog istraživanja. Statistički skup je nužno precizno definirati. Precizno definiran statistički skup podrazumijeva:

- pojmovno
- prostorno
- vremensko određenje

Promotrimo sada skup studenata koji se trenutno nalazi na laboratorijskim vježbama iz Poslovne statistike. Oni čine jedan statistički skup. Određenje skupa je: **pojmovno** - redovni studenti II. semestra smjera RiF (prezimena od A do M), **prostorno** - u Splitu, **vremenski** - dana dd.mm.yy.

Možemo provesti jedno malo statističko istraživanje na ovom skupu. Cilj tog istraživanja neka bude ispitivanje mišljenja studenata o razini njihove motiviranosti za nastavu iz kolegija Poslovna statistika. Dakle, statističkim rječnikom rečeno, obilježje po kojem ćemo promatrati taj skup je razina motiviranosti. Obilježje općenito može biti:



Istraživanje možemo provesti anonimnom anketom.

Na pitanje o razini motiviranosti za kolegij Poslovna statistika ponuđeni su sljedeći odgovori:

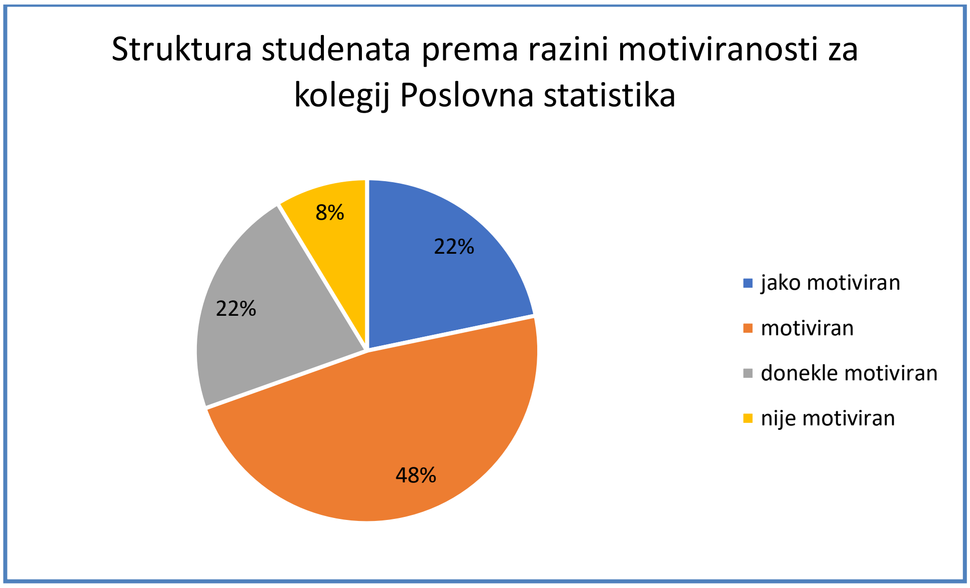
1. Jako sam motiviran(a) - smatram da mi kolegij Poslovna statistika može donijeti korisne spoznaje za budući studij i rad
2. Motiviran(a) sam – želim što lakše položiti kolegij, ali me sadržaj kolegija pretjerano ne zanima
3. Donekle sam motiviran(a) - dolazit ću na nastavu jer je obvezna
4. Nisam motiviran(a) - ne vjerujem da ću pohađati nastavu, Poslovna statistika me uopće ne zanima.

Nakon prikupljenih anketnih listića možemo formirati **statistički niz** kojeg čine **modaliteti** obilježja (pojavni oblici) i pripadne **apsolutne frekvencije** (a_i, f_i). Apsolutna frekvencija je broj jedinica statističkog skupa (u ovom slučaju studenata) koji pripadaju određenom modalitetu. Ukupni broj ispitanika je **opseg** statističkog skupa.

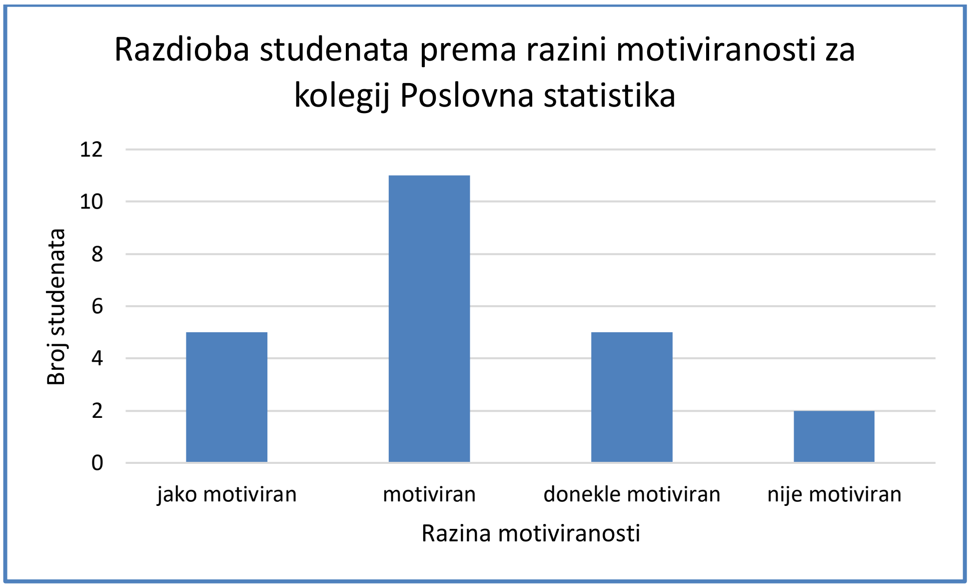
Rezultate istraživanja unijet ćemo u Excelov radni list i formirati tablicu u kojoj ćemo u prvi stupac unijeti moguće odgovore na pitanje (modalitete), a u drugi apsolutne frekvencije (broj odgovora na pojedina pitanja).

Razina motiviranosti	Broj studenata
jako motiviran	5
motiviran	11
donekle motiviran	5
nije motiviran	2

Nakon što smo tablično prikazali distribuciju studentskih odgovora dobro bi bilo napraviti i njihov grafički prikaz. Obzirom da je riječ o jednom statističkom nizu s jednim kvalitativnim obilježjem koristit ćemo neki ravninski graf, npr. grafikon jednostavnih stupaca i/ili strukturni krug. Zato ćemo označiti ćelije u kojima se nalaze modaliteti i pripadne apsolutne frekvencije i aktivirati karticu *Insert,* unutar nje možemo izabrati strukturni krug (*Pie*) i jedan od njenih podtipova. Lijevim klikom miša na bijelu podlogu grafa u vrpci s alatima pojave se opcije *Design* i *Format.* Na kartici *Design* biramo opciju *Quick layout* i unutar nje onaj izgled koji nam omogućava i prikaz **relativnih frekvencija** (% udjela pojedinih modaliteta unutar cijelog skupa). Graf opremamo i naslovom. Neka to bude 'Struktura studenata prema razini motiviranosti za kolegij Poslovna statistika'. Preporuka je da naslov sadržava informaciju o statističkom skupu i promatranom obilježju.



Na sličan način napravit ćemo grafikon jednostavnih stupaca (*Column chart, 1. podtip)* i opremiti ga potrebnim oznakama, naslovom i oznakama na osima.



Odgovorite na pitanja:

Koliki je broj studenata koji su jako motivirani za nastavu iz kolegija PS?

Koliki je udio studenata koji su jako motivirani za nastavu iz kolegija PS?

1. Grafičko prikazivanje statističkih podataka

Vježba 1.1.

VJEŽBA 1.1.

Anketirani su studenti Sveučilišta u Splitu akademske godine 2017/2018.
Na pitanje kojim sportskim utakmicama najrađe prisustvuju, ponudili su slijedeće odgovore:

Sport	Broj studenata
nogomet	1150
košarka	1010
rukomet	720
mali nogomet	110
vaterpolo	270
ragbi	180

a) Odredite statistički skup
b) Odredite opseg statističkog skupa
c) O kojem je obilježju riječ? Koja je vrsta obilježja? Navedite jedan od modaliteta.
d) Podatke iz tablice grafički predočite strukturnim krugom. Navedite sve potrebne oznake.
e) Odredite udio studenata koji najradije prisustvuju košarkaškim utakmicama.

Rješenje:

Iz teksta zadatka razvidno je tko čini statistički skup. To su studenti Sveučilišta u Splitu akademske godine 2017/2018. Opseg skupa iznosi 3440.

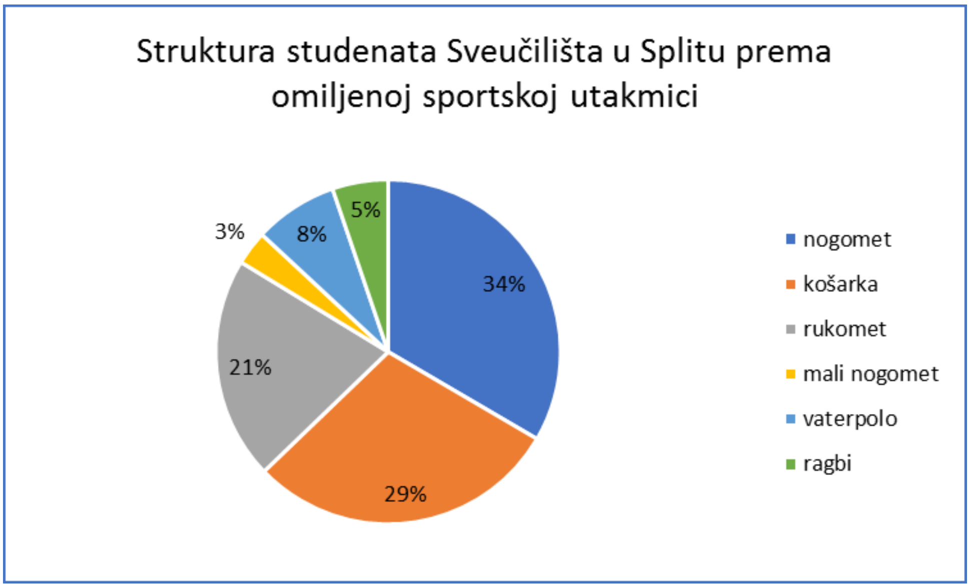
Opseg statističkog skupa je ukupan broj jedinica statističkog skupa kojeg jednostavno dobivamo sumiranjem apsolutnih frekvencija iz dane tablice. Označimo blok ćelija **C7:C12** i na kartici *Home* nađemo ikonicu za sumiranje $\sum$ *Autosum*. U prvoj idućoj ćeliji dobivamo rezultat.

Obilježje koje se promatra na tom skupu je omiljena vrsta sportske utakmice. Riječ je o kvalitativnom, nominalnom, opisnom (atributivnom) obilježju. Ima 6 mogućih modaliteta (nogomet, košarka, rukomet, mali nogomet, vaterpolo, ragbi).

Podatke iz tablice predočit ćemo površinskim grafikonom i to strukturnim krugom. Počinjemo označavanjem bloka ćelija sa podacima (modaliteti i pripadne apsolutne frekvencije) **B7:C12**, naslove i totale ne označavamo kod jednostavnih tablica. Potom na kartici *Insert* izaberemo *Chart*s i unutar njega vrstu *(Pie)* i podvrstu (neka bude 1. podtip u izborniku 2-D Pie).

Dobiveni graf po potrebi izmjestimo s teksta uz pomoć lijeve tipke miša. Graf bez naslova i relativnih frekvencija nije dovoljno informativan, stoga ćemo ih dodati. Najbrži način je da lijevim klikom miša na bijeloj površini grafa u vrpci s alatima aktiviramo opciju *Design* unutar koje izaberemo gotovi predložak *Quick Layout*, u ovom slučaju *Layout* 6 ili bilo koji drugi koji nam daje naslov, legendu i relativne frekvencije. Potrebno je sada napisati naslov grafa 'Struktura studenata Sveučilišta u Splitu prema omiljenoj sportskoj utakmici'. Druga opcija, 'ručnog' opremanja grafa je klikom na ikonicu + koja se pojavi desno od grafa nakon njene aktivacije.

Konačni izgled grafa:



Ukoliko nismo zadovoljni s izgledom moguće je raditi 'kozmetičke' zahvate (povećanje/smanjenje fonta i slično). I sad na kraju jednostavno je odgovoriti na posljednje pitanje.

Udio studenata koji najradije prisustvuju košarkaškim utakmicama je 29%.

Odgovorite na pitanja:

1. Koliko studenata najviše voli rukometne utakmice?
2. Koliki udio studenata preferira ragbi?
3. Koje su sportske utakmice najmanje popularne?
4. Koliki je broj, a koliki udio studenata izabrao upravo te najmanje popularne sportske utakmice?

Vježba 1.2.

VJEŽBA 1.2

Podaci o službenim putovanjima, obzirom na vrstu prijevoznog sredstva, u poduzeću Vihor d.d. na dan 31. prosinca 2018. g. dani su tablicom:

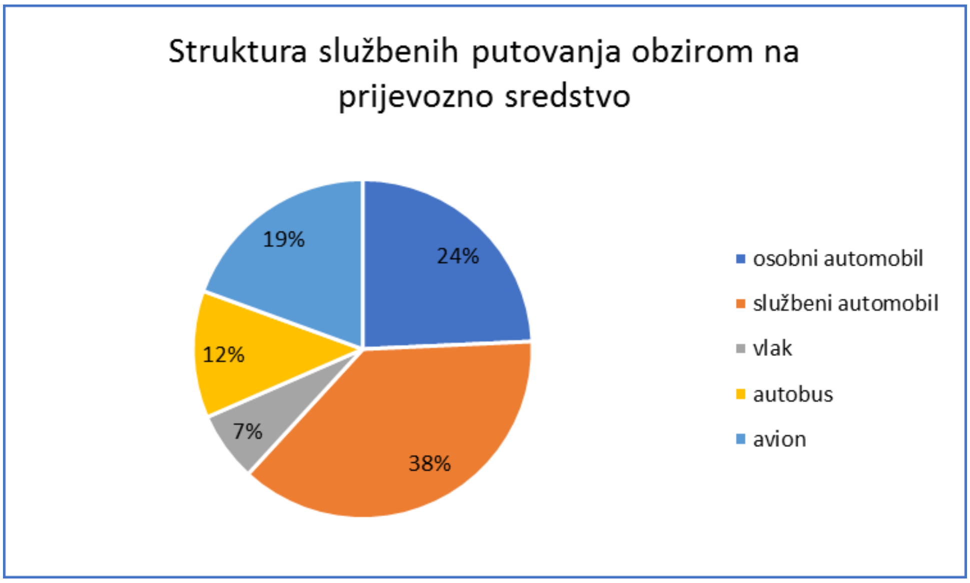
Vrsta prijevoza	Broj putovanja
osobni automobil	110
službeni automobil	170
vlak	30
autobus	55
avion	88

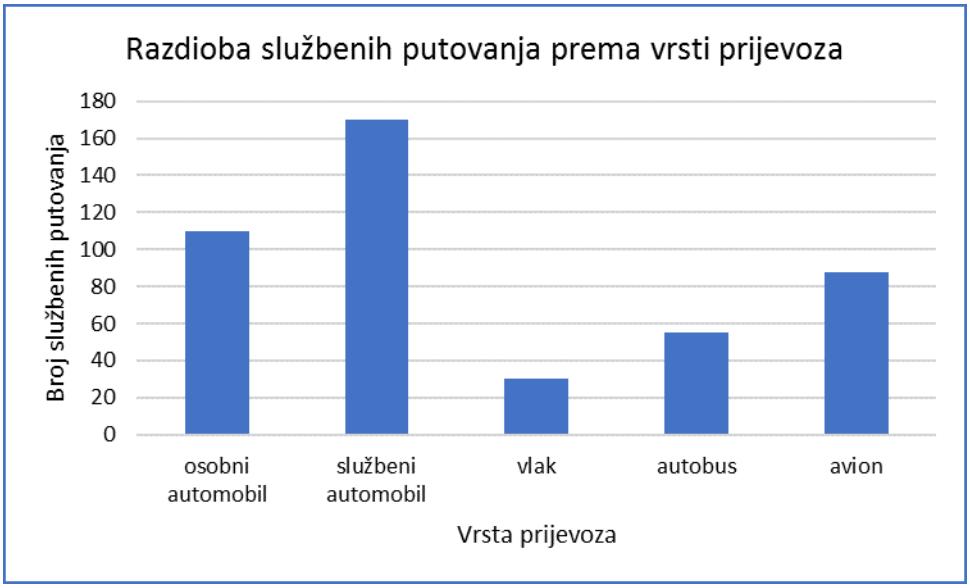
a) Navedite što je statistički skup? Koliki je opseg skupa?
b) Navedite obilježje i vrstu promatranog obilježja jedinica statističkog skupa
c) Navedite jedan od modaliteta. Koliko jedinica statističkog skupa ima tu vrijednost obilježja?
d) Podatke iz tablice grafički predočite strukturnim krugom. Navedite sve potrebne oznake.
e) Podatke iz tablice grafički predočite jednostavnim stupcima. Navedite sve potrebne oznake.
f) Koliko je putovanja u protekloj godini bilo avionom? Koliki je njihov udio?

Rješenje:

Statistički skup: službena putovanja u poduzeću Vihor d.d. na dan 31. prosinca 2018. godine. Opseg skupa je 453 što znači da je u 2018. godini zabilježeno ukupno 453 službena putovanja.

Obilježje: vrsta prijevoznog sredstva, što je kvalitativno, nominalno i opisno obilježje. Ima 5 mogućih modaliteta. Jedan od njih je npr. avion. Tu vrijednost obilježja (službena putovanja avionom) ima 88 jedinica promatranog statističkog skupa.





U proteklој godini zabilježeno je 88 putovanja avionom što predstavlja udio 19 % od ukupnog broja putovanja te godine.

Razmislite, koja bi se još obilježja mogla promatrati na istom skupu službenih putovanja?

Vježba 1.3.

VJEŽBA 1.3.

Anketirani su članovi kluba liječenih alkoholičara u gradu Tuga s juga dana 1. srpnja 2017. godine vezano za neuspjele pokušaje apstinencije. Podaci su dani tablicom:

Broj recidiva	Broj članova kluba
jedan	3
dva	7
tri	6
četiri	9
pet	12
šest	10
sedam	13

a) Navedite tko čini statistički skup

b) Odredite obilježje i vrstu obilježja te navedite barem jedan od modaliteta tog obilježja.

c) Podatke iz tablice prikažite grafički uspravnim stupcima te navedite sve potrebne oznake.

d) Koliki je udio članova kluba koji su načinili najmanje recidiva?

Rješenje:

Iz teksta zadatka vidljivo je da statistički skup čine članovi kluba liječenih alkoholičara u gradu Tuga s juga dana 1. srpnja 2017. godine.

Obilježje po kojem promatramo članove tog skupa je broj recidiva (neuspjelih pokušaja apstinencije). To je kvantitativno, diskontinuirano (diskretno) obilježje. Može poprimiti 7 vrijednosti (modaliteta) npr. broj četiri označava 4 recidiva. Sumiranjem apsolutnih frekvencija dobijemo opseg skupa koji iznosi 60.

Postupak dobivanja grafičkog prikaza je sličan kao u vježbi 1.1. s razlikom odabira tipa grafa. Dakle označimo blok ćelija u kojem su modaliteti i pripadne apsolutne frekvencije **B7:C13** i potom na vrpci alata izaberemo *Insert* /*Charts* /*Column* prvi podtip (2-D).

Dobiveni graf je 'gol' i bez potrebnih oznaka. Kako bi bio informativan moramo dodati naslov i oznake na osima. Najbrži put da 'oživimo' ovaj stupčasti grafikon je da klikom na njegovu bijelu površinu dobijemo karticu *Design* i odaberemo *Quick Layout* u kojem se nalaze gotovi predlošci. Odaberemo onaj koji nam omogućuje upisivanje naslova te oznaka na osima (pr. *Layout 9*). Naslov može glasiti 'Članovi kluba liječenih alkoholičara u gradu Tuga s juga dana 1.7.2017. prema broju recidiva'. Klikom na tekst naslova možemo smanjiti font na prihvatljivu razinu. Legenda je u ovom slučaju suvišna pa je možemo izbrisati. Upišemo oznaku na x osi koju pročitamo iz zaglavne ćelije B6, a to je 'Broj recidiva'. Sličan postupak ponovimo za y os u koju upišemo sadržaj zaglavne ćelije C6 'Broj članova kluba'.

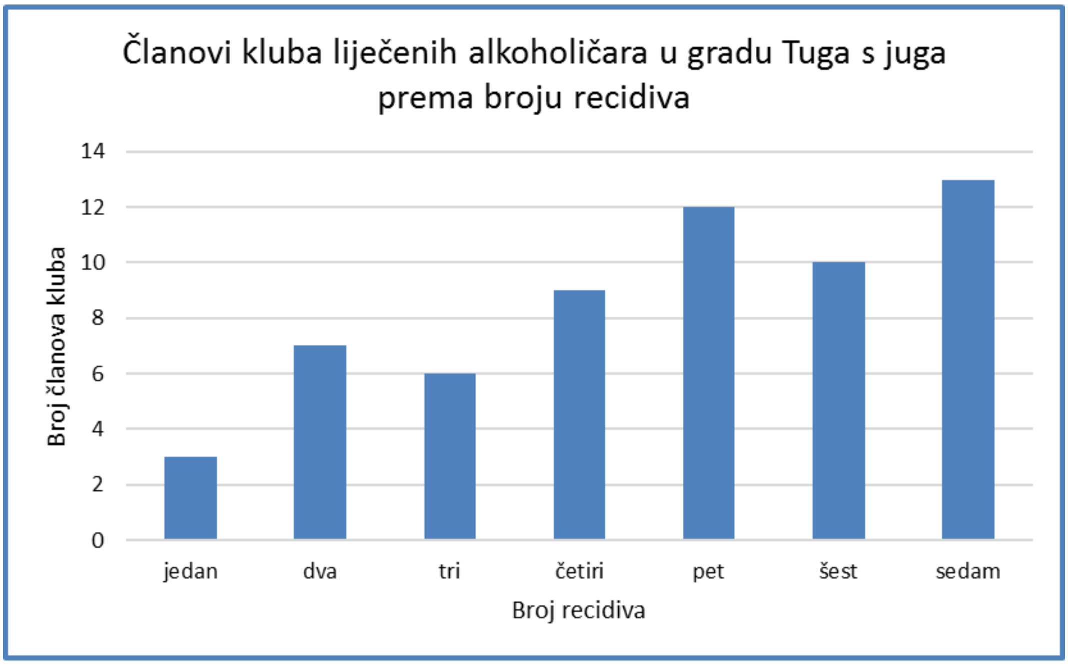
Dodatno se traži udio članova kluba s najmanje recidiva pa taj podatak možemo dobiti i računskim putem. Dakle **broj** članova s najmanje recidiva (oni s jednim) iznosi 3. Da bi dobili njihov **udio** moramo taj broj (aps. frekvenciju) podijeliti s ukupnim brojem članova (opseg skupa). U nekoj slobodnoj ćeliji možemo napraviti račun koristeći relativne adrese.

Broj recidiva	Broj članova kluba		
jedan	3		=C7/C14
dva	7		
tri	6		
četiri	9		
pet	12		
šest	10		
sedam	13		
	60		

Rezultat iznosi 0,05. Budući se udjeli iskazuju u postocima još ćemo taj podatak klikom na ikonicu % pretvoriti u postotak.

Dakle odgovor glasi, udio članova kluba s najmanje recidiva iznosi 5%.

Završni graf izgleda ovako:



Vježba 1.4.

Kao i u prethodnim primjerima ispunimo potrebne podatke:

a) U ovom primjeru :	
Statistički skup :	osuđeni počinitelji kaznenih djela u Hrvatskoj u 2014. godini
Obilježje i vrsta obilježja :	Županija; kvalitativno, nominalno, geografsko (prostorno)
Navedi dva modaliteta:	Splitsko-dalmatinska županija, Grad Zagreb
Opseg statističkog skupa:	61889

Rješenje:

U ovom zadatku dobro je uočiti da ima mnogo modaliteta i stoga bi strukturni krug ili jednostavni stupci bili lošiji izbor za grafičko prikazivanje od položenih stupaca. Postupak izrade grafa je sličan prethodno opisanima s razlikom odabira grafa. Biramo *Bar Chart* i to onaj predložak koji nam omogućava unos naslova te oznaka na koordinatnim osima. Tablica s podacima i već izrađen grafikon izgledaju ovako:

VJEŽBA 1.4.

Dani su podaci o osuđenim počiniteljima kaznenih djela u Hrvatskoj u 2014. godini

ŽUPANIJA	BROJ OSUĐENIH POČINITELJA
Zagrebačka županija	2908
Krapinsko-zagorska ž.	1469
Sisačko-moslavačka ž.	2396
Karlovačka županija	2403
Varaždinska županija	1903
Koprivničko-križevačka ž.	1234
Bjelovarsko-bilogorska ž.	1796
Primorsko-goranska ž.	3739
Ličko-senjska županija	902
Virovitičko-podravska ž.	948
Požeško-slavonska ž.	850
Brodsko-posavska ž.	2103
Zadarska županija	2815
Osječko-baranjska ž.	3585
Šibensko-kninska ž.	2373
Vukovarsko-srijemska ž.	2919
Splitsko-dalmatinska ž.	7352
Istarska županija	5667
Dubrovačko-neretva. ž.	1826
Međimurska županija	1203
Grad Zagreb	11498

a) U ovom primjeru :

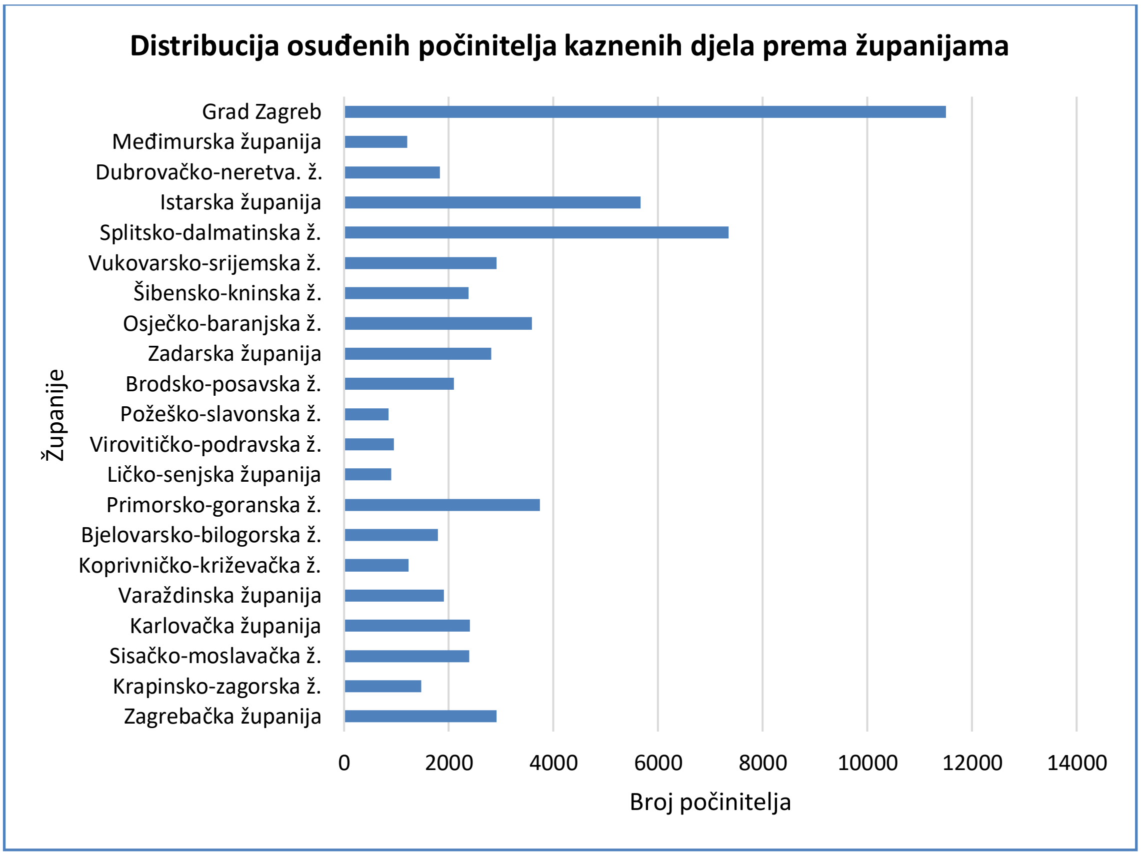
Statistički skup je:

Obilježje i vrsta obilježja

Navedi dva modaliteta

Opseg statističkog skupa:

b) Distribuciju osuđenih počinitelja prema županijama prikažite položenim stupcima (Bar chart) te uz grafikon navedite sve potrebne oznake.



Vježba 1.5.

	A	B	C	D	E	F
1	***Izvoz i uvoz Republike Hrvatske prema ekonomskim grupacijama zemalja***					
2	*Ekonomska grupacija zemalja:*	*Izvoz (u 000 kn)*	*Uvoz (u 000 kn)*			
3	zemlje EU	67.615.063	126.868.502			
4	zemlje EFTA-e	1.700.999	1.446.241			
5	zemlje CEFT-e	18.232.207	10.076.552			
6	zemlje OPEC-a	1.600.184	3.654.241			
7	ostale europske zemlje	3.590.171	4.602.721			
8	ostale azijske zemlje	3.566.956	12.107.920			
9	ostale afričke zemlje	2.216.093	328.608			
10	ostale američke zemlje	5.448.867	1.844.547			
11	oceanijske zemlje	265.055	146.410			
12	neraspoređeno	112.421	1.605.406			
13						
14						
15	*Izvor: Publikacija o robnoj razmjeni za 2017. godinu, www.dzs.hr*					
16						
17	**Zadatak :**					
18	**a)** Odredite statističke skup(ove).					
19	**b)** Odredi opseg skupa(ova)!					
20	**c)** Obilježje i vrsta obilježja?					
21	**d)** Koja je ovo vrsta tablice?					
22	**e)** Distribuciju prikažite grafički dvostrukim stupcima. Uz grafikon navedite sve potrebne oznake.					

Rješenje:

Primijetimo da je u ovom primjeru riječ o dva statistička skupa:

Izvoz RH (u 000 kn) 2017. godine i Uvoz RH (u 000 kn) 2017. godine, ali oba promatramo po istom obilježju, a to je ekonomska grupacija zemalja. Obzirom na modalitete lako je zaključiti da se radi o kvalitativnom, nominalnom, geografskom obilježju. Budući u tablici imamo 2 skupa s istim obilježjem riječ je o složenoj tablici.

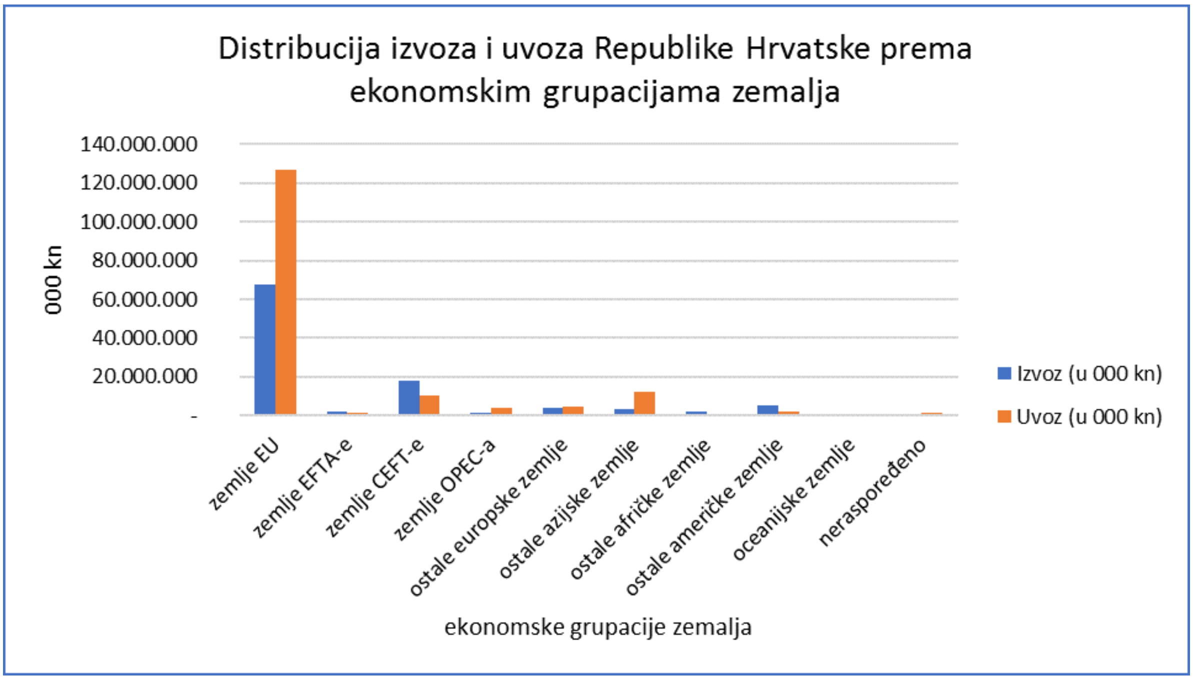
Opseg skupova lako dobijemo sumiranjem apsolutnih frekvencija:

	A	B	C
1	***Izvoz i uvoz Republike Hrvatske prema ekonomskim grupacijama zemalja***		
2	*Ekonomska grupacija zemalja:*	*Izvoz (u 000 kn)*	*Uvoz (u 000 kn)*
3	zemlje EU	67.615.063	126.868.502
4	zemlje EFTA-e	1.700.999	1.446.241
5	zemlje CEFT-e	18.232.207	10.076.552
6	zemlje OPEC-a	1.600.184	3.654.241
7	ostale europske zemlje	3.590.171	4.602.721
8	ostale azijske zemlje	3.566.956	12.107.920
9	ostale afričke zemlje	2.216.093	328.608
10	ostale američke zemlje	5.448.867	1.844.547
11	oceanijske zemlje	265.055	146.410
12	neraspoređeno	112.421	1.605.406
13	**UKUPNO**	104.348.016	162.681.148

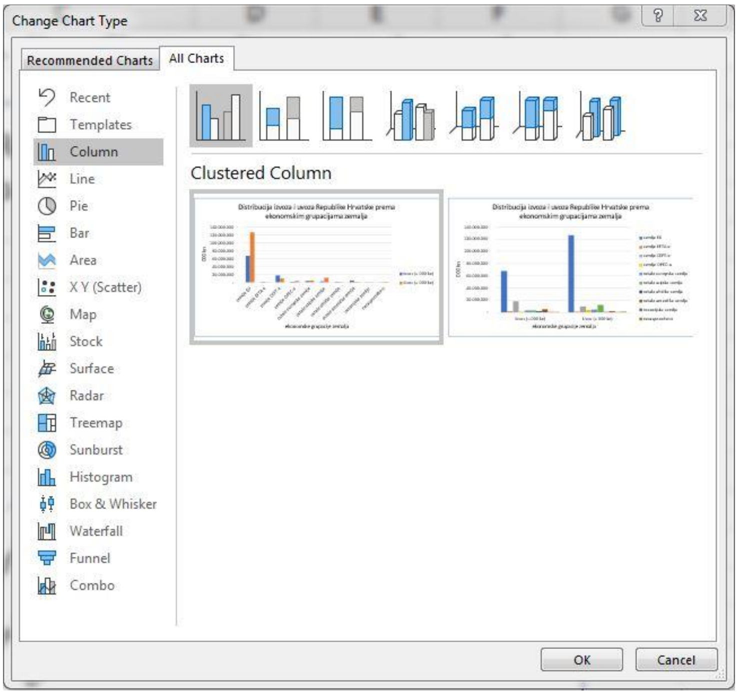
Pažljivo s očitavanjem opsega! Kao što je vidljivo u zaglavnoj ćeliji veličine su izražene u 000 kn.

Opseg skupa Izvoz je 104 348 016 000 kn, a skupa Uvoz 162 681 148 000 kn.

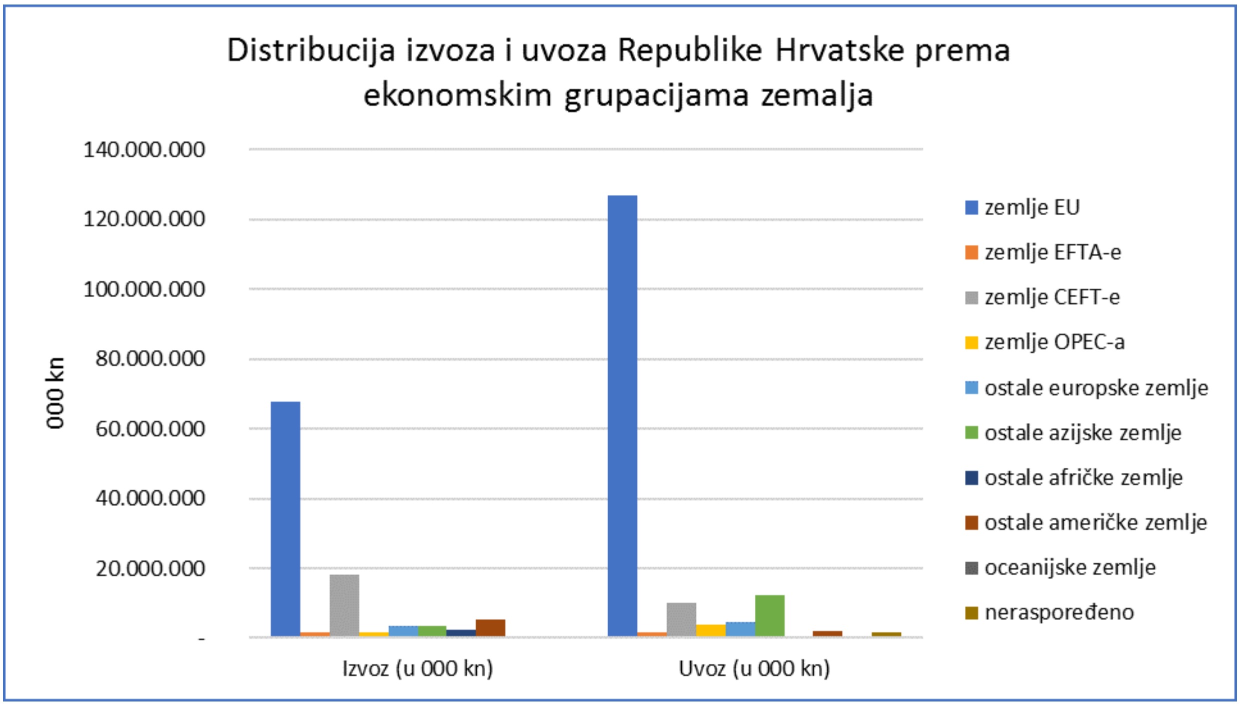
Prilikom grafičkog prikazivanja složenih tablica označavamo i zaglavne ćelije, dakle **A2:C12,** nakon odabira *Insert/Column(1 podtip)/Quick Layout (br 9)* i upisivanja naslova i oznaka na osima dobivamo konačni izgled.



Riječ je o prikazu dvostrukim stupcima. Desnim klikom miša na bijelu površinu grafa dobijemo plutajući izbornik i biramo *Change Chart Type*.



Alternativni prikaz je prikaz višestrukim stupcima, imamo dvije kategorije (Izvoz i Uvoz), a stupci predstavljaju ekonomske grupacije zemalja.



Odgovorite na pitanja:

1. S kojom ekonomskom grupacijom RH ima najživlju razmjenu?
2. S kojom ekonomskom grupacijom RH ima najskromniju razmjenu?
3. Da li, ukupno gledajući, RH više izvozi ili uvozi? Kolika je razlika?
4. S kojim ekonomskim grupacijama RH ima pozitivnu, a s kojim negativnu vanjsko-trgovinsku bilancu (apsolutna razlika izvoza i uvoza)?
5. Ukupnu strukturu izvoza i uvoza RH prikažite strukturnim krugom! (potrebno je označiti ćelije **B2:C2** i **B13:C13** uz pomoć tipke *Ctrl*, odabrati strukturni krug te opremiti graf)

<u>Za vježbu riješite:</u>

U mapi **STATISTIKA.NOVO** možete riješiti:

Primjeri 2 (primjer 2.6. i primjer 2.7)

Zadaci za vježbu 2 (zadatak 2.1, zadatak 2.2, zadatak 2.3.)

2. Grupiranje kvantitativnih podataka

Vježba 2.1.

Prikupljeni su podaci o uspjehu na ispitu iz Poslovne statistike studenata preddiplomskog studija *Trgovinskog poslovanja* OSS-a akademske godine 2017/2018. na ispitnom roku 21.06.2018. Podaci su dani u tablici u stupcu A (**A1:A97**). Zadatak glasi:

	A	B	C	D
1	Ocjene na ispitu			
2	2		Dani su podaci o uspjehu na ispitu iz Poslovne statistike dana 21.06.2018.	
3	2		96 studenata preddiplomskog studija Trgovinskog poslovanja ak. g. 2017/18	
4	2			
5	2			a) Formirajte razdiobu frekvencija te interpretirajte jednu
6	1			b) Razdiobu prikažite grafički odgovarajućim grafikonom
7	2			c) Izračunajte relativne frekvencije i pripadni kumulativni niz
8	2			d) Interpretirajte po jednu frekvenciju u svakom od nizova
9	2			e) Formirajte kumulativne nizove 'veće od' i 'manje od' te
10	3			interpretirajte po jednu frekvenciju u svakom nizu
11	1			
12	3			
13	2			
14	1			
15	1			
16	2			
17	2			
18	1			
19	1			
20	2			
21	1			
22	1			

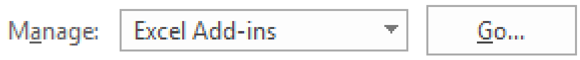
a) Formirajte razdiobu frekvencija te interpretirajte jednu od frekvencija
b) Razdiobu prikažite grafički
c) Izračunajte relativne frekvencije i pripadni kumulativni niz
d) Interpretirajte po jednu frekvenciju u svakom od nizova
e) Formirajte kumulativne nizove 'manje od' i 'veće od' te interpretirajte po jednu kumulativnu frekvenciju u svakom nizu

Rješenje:

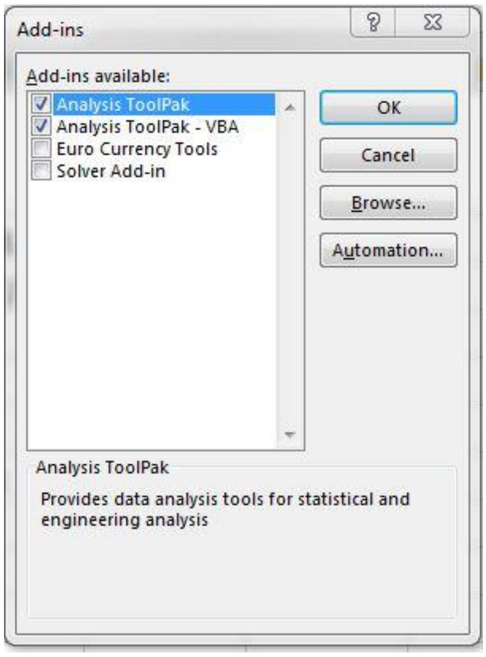
Podaci koje imamo su prikupljeni direktno od ispitivača. U ovakvoj formi kojoj su zadani zovu se sirovi podaci. Tko čini statistički skup?

Obilježje koje promatramo su ocjene na ispitu. Ocjena na ispitu je u pravilu kvalitativno, redoslijedno obilježje (obilježje ranga), ali uobičajeno mu je dati brojčano ruho zbog lakše obrade tako da ga u obradi tretiramo kao diskretno numeričko obilježje. Procedura koju koristimo za grupiranje numeričkih podataka zove se Histogram. Ne nalazi se standardno u izbornicima MS Excela pa je trebamo sami dodati i aktivirati. To radimo na sljedeći način:

Pritiskom na *File* dobijemo padajući izbornik u kojem biramo *Options*, a potom *Add-ins* i na upit Manage Excel Add ins odgovaramo s *Go* :



U dijaloškom okviru označimo prve dvije opcije Analysis ToolPak i Analysis ToolPak – VBA te potvrdimo izbor.



Ako smo to napravili na kartici *Data* pojavi se opcija *Data Analysis* s mnoštvom statističkih alata, među ostalim i nama potrebnim Histogramom.

Da bi grupirali podatke moramo prije svega zadati 'ključ' po kojem ćemo ih grupirati. Budući ocjene na ispitu tretiramo kao kvantitativno, ali diskretno obilježje koje može poprimiti samo 5 vrijednosti(modaliteta) 1, 2, 3, 4, 5 formiramo stupac (bin range) u kojem ih redom navodimo.

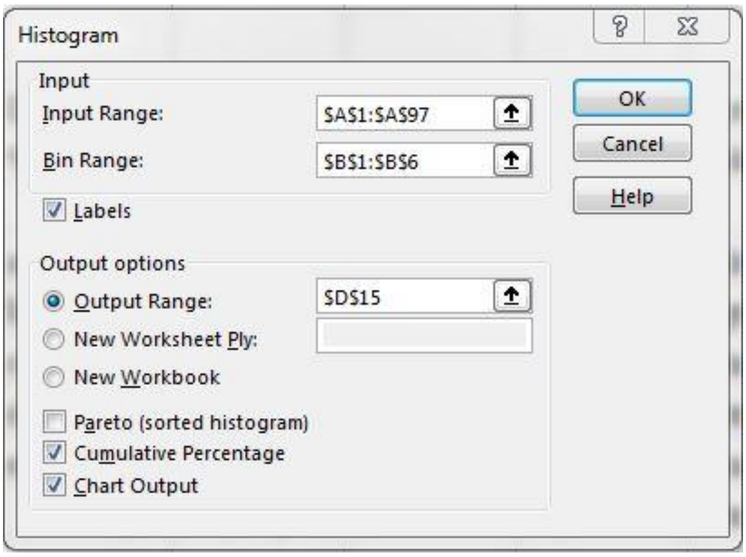
U zaglavnu ćeliju B1 kopiramo naziv stupca iz A1 i navodimo redom sve moguće modalitete obilježja ocjena na ispitu.

	A	B	C
1	**Ocjene na ispitu**	**Ocjene na ispitu**	
2	2	1	
3	2	2	
4	2	3	
5	2	4	
6	1	5	
7	2		
8	2		
9	2		
10	3		
11	1		
12	3		
13	2		
14	1		

Pokrećemo Proceduru Histogram (*Data/Data Analysis/Histogram*) i dobijemo dijaloški okvir. U prvom retku (Input range) unosimo adrese podataka koje želimo obraditi (**A1:A97**), a u drugom (Bin range) modalitete koje obilježje može poprimiti (**B1:B6**). Budući smo obuhvatili i adrese naslovnih ćelija (A1 i B1) moramo aktivirati opciju *Labels*. Time smo dali do znanja da su podaci u te dvije ćelije naslovi. Ako u rasponu podataka ne obuhvatimo naslovne ćelije tada ne stavljamo ni oznaku *Labels*!

U drugom dijelu dijaloškog okvira Output options biramo što želimo da nam Histogram vrati kao rezultat obrade te gdje da smjesti izlazne rezultate. Obzirom da želimo da nam tablica bude na postojećem radnom listu označimo opciju Output range i unesemo adresu koju želimo kao krajnju gornju lijevu ćeliju tablice, npr. **D15**.

Od tri opcije sa dna dijaloškog okvira biramo *Cumulative percentage* (tablici dodaje stupac kumulativnih relativnih frekvencija) i *Chart output* (uz izlaznu tablicu daje i odgovarajući grafički prikaz).



Dobivena tablica izgleda

Ocjene na ispitu	*Frequency*	*Cumulative %*
1	22	22,92%
2	37	61,46%
3	23	85,42%
4	12	97,92%
5	2	100,00%
More	0	100,00%

Malo je preuredimo kako bi je stavili u kontekst našeg zadatka.

Ocjene na ispitu	*Broj studenata*	*Kumulativne frekvencije %*
1	22	22,92%
2	37	61,46%
3	23	85,42%
4	12	97,92%
5	2	100,00%
More	0	100,00%

Kao primjer interpretacije možemo odgovoriti na pitanje:

Koliko je studenata dobilo ocjenu dobar? Odgovor je 23

Dakle, stupac apsolutnih frekvencija (Broj studenata) odgovara na pitanje koliko je studenata dobilo pojedinu ocjenu.

Grafičkim prikazom razdiobe pozabavit ćemo se malo kasnije.

Želimo li odgovoriti na pitanje 'Koliki je udio studenata koji su dobili ocjenu dobar?' moramo formirati sami stupac relativnih frekvencija budući u Excelu ne postoji funkcija za direktni izračun relativnih frekvencija.

Relativne frekvencije računamo tako da podijelimo apsolutne frekvencije s opsegom skupa ($N = \sum f_i$), $p_i = \frac{f_i}{N}$ (%) .

Stoga prvo moramo izračunati opseg skupa kao sumu apsolutnih frekvencija u izlaznoj tablici (stupac *Broj studenata)*. Označimo blok ćelija u kojima se nalaze apsolutne frekvencije, a to je **E16:E21** i odabirom ikonice za sumiranje $\sum$ *Autosum* u idućoj ćeliji dobivamo sumu 96 što smo i znali, ali je bilo korisno provjeriti.

Sada označimo stupac **F** i desnim klikom miša dobijemo plutajući izbornik u kojem biramo *Insert* kako bi ubacili jedan stupac neposredno ispred stupca kumulativnih frekvencija. Zaglavnu ćeliju imenujemo *Relativne frekvencije.*

U ćeliji **F16** računamo prvu relativnu frekvenciju **=E16/$E$22 .** Adresa opsega skupa je kao što vidite apsolutna ('okružena $'). Razlog tomu je da pri kopiranju formule za ostale relativne frekvencije nazivnik uvijek ostane isti. Najbrži način da adresa postane apsloutna je pomoću tipke F4. Kopiranje provedemo tako da označimo ćeliju izračuna, a pokazivač miša dovedemo do donjeg desnog ruba te ćelije do pojave križića (+) te 'razvučemo' njime do dna stupca u kojem želimo izračunati rel. frekvencije. Već označeni podaci su izraženi decimalnim brojevima pa na kartici *Home* prevedemo iste u postotne iznose odabirom ikonice %. Suma relativnih frekvencija trebala bi iznositi 100% što možemo provjeriti sumiranjem. Konačni rezultat izgleda:

Ocjene na ispitu	*Broj studenata*	*Relativne frekvencije*	*Kumulativne rel. frekvencije %*
1	22	22,92%	22,92%
2	37	38,54%	61,46%
3	23	23,96%	85,42%
4	12	12,50%	97,92%
5	2	2,08%	100,00%
More	0	0,00%	100,00%
	96	100,00%	

Sada možemo odgovoriti primjerice na pitanje 'Koliki udio studenata ima ocjenu dobar?' Odgovor čitamo u stupcu relativnih frekvencija i glasi 23,96 %. Interpretirajmo još jednu vrijednost u stupcu relativnih frekvencija. Recimo 22,92% predstavlja udio studenata koji su pali ispit.

Histogramom smo već dobili stupac kumulativnih relativnih frekvencija pa na već ranije postavljeno pitanje 'Koliki je udio studenata koji su dobili ocjenu najviše dobar?' odgovor je 85,42%. Primijetimo da je to zbroj svih relativnih udjela uključivo s tom vrijednošću modaliteta.

22,92%+38,54%+23,96% = 85,42%

Na sličan način, ali 'pješke' formiramo kumulativne nizove frekvencija.

Umetnemo dva potrebna stupca ispred niza relativnih frekvencija i imenujemo ih *Kumulativni niz 'manje od'* i *Kumulativni niz 'veće od'.*

Kumulativni niz 'manje od' formiramo tako da kumulativno zbrajamo apsolutne frekvencije počevši od prve vrijednosti.

U prvu ćeliju stupca **F16** kopiramo(ili upišemo) vrijednost prve apsolutne frekvencije 22. U idućoj ćeliji **F17** unesemo **=F16+E17** i potvrdimo. Ostale članove niza dobivamo kopiranjem pravila iz ćelije **F17.** Posljednji član niza mora biti jednak opsegu skupa (96).

Kumulativni niz 'veće od' formiramo tako da kumulativno zbrajamo apsolutne frekvencije počevši od posljednje vrijednosti.

U posljednju ćeliju stupca **G21** kopiramo(ili upišemo) vrijednost posljednje (iz opcije *More)* apsolutne frekvencije 0. U prethodnoj ćeliji **G20** unesemo pravilo kojim unazad kumulativno zbrajamo **=G21+E20** i potvrdimo. Ostale članove niza dobivamo kopiranjem pravila iz ćelije **G20** u ostale ćelije tog stupca (prema gore). U prvoj ćeliji stupca mora biti vrijednost opsega skupa.

Sada možemo interpretirati dvije frekvencije odgovarajući na dva pitanja:

Koliko studenata je dobilo ocjenu manju ili jednaku 2?

Pitanje se dakle odnosi na studente koji su dobili ili ocjenu 1 ili 2. Odgovor ćemo potražiti u stupcu kumulativnih frekvencija 'manje od' i vidjeti da je odgovor 59 (22+37=59).

Isto pitanje moglo je biti postavljeno na sljedeći način: Koliko studenata je dobilo najviše ocjenu 2?

Iduće pitanje glasi:

Koliko studenata je dobilo ocjenu veću ili jednaku 3 (barem 3)?

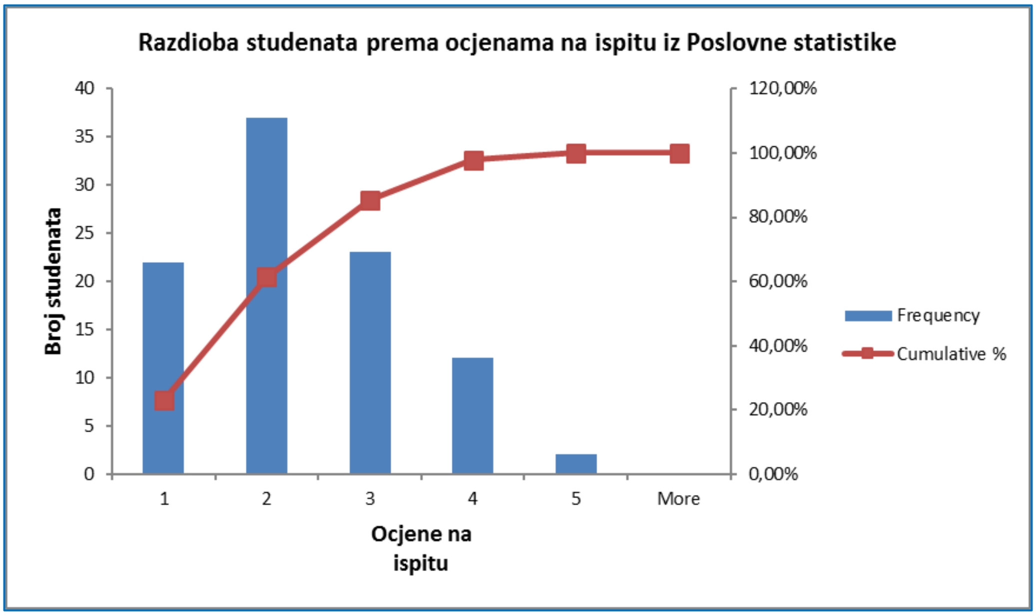
Pitanje se dakle odnosi na studente koji su dobili ocjene 3 ili 4 ili 5. Odgovor ćemo potražiti u stupcu kumulativnih frekvencija 'veće od' i vidjeti da je odgovor 37(2+12+23=37) Isto pitanje moglo je glasiti: Koliko je studenata dobilo barem ocjenu 3?

Konačni izgled tablice je

Ocjene na ispitu	Broj studenata	Kumulativni niz 'manje od'	Kumulativni niz 'veće od'	Relativne frekvencije	Kumulativne frekvencije %
1	22	22	96	22,92%	22,92%
2	37	**59**	74	38,54%	61,46%
3	23	82	**37**	23,96%	85,42%
4	12	94	14	12,50%	97,92%
5	2	96	2	2,08%	100,00%
More	0	96	0	0,00%	100,00%
	96			100,00%	

Sada ćemo se pozabaviti dobivenim grafom i urediti ga na način da bude informativan za konkretni primjer. Izmijenit ćemo naslov i oznaku na ordinati. Stupci se ne spajaju jer je riječ o diskretnom numeričkom obilježju.

Graf nakon uređivanja izgleda ovako:



Za vježbu možete odgovoriti na sljedeća pitanja:

1. Koliko je studenata dobilo ocjenu 4?
2. Koliki je udio studenata dobio ocjenu 4?
3. Koliko je studenata dobilo ocjenu najviše 4?
4. Koliko je studenata dobilo najmanje 4?
5. Koliki je udio studenata dobio ocjenu najviše 4?

Vježba 2.2.

U trgovačkom centru 5 COINS u gradu X planira se broj zaposlenika za sljedeću godinu.
U tu svrhu 1.12.2018.g. je anketirano 60 kupaca.
Na pitanje:'U kojem danu u tjednu obavljate tjednu kupnju?', dali su sljedeće odgovore:

Redni broj	Dan	Redni broj	Dan	Redni broj	Dan	Redni broj	Dan
1	ponedjeljak	16	petak	31	petak	46	petak
2	ponedjeljak	17	subota	32	petak	47	nedjelja
3	utorak	18	subota	33	petak	48	nedjelja
4	srijeda	19	subota	34	subota	49	nedjelja
5	srijeda	20	četvrtak	35	srijeda	50	petak
6	četvrtak	21	petak	36	četvrtak	51	petak
7	četvrtak	22	petak	37	nedjelja	52	petak
8	petak	23	petak	38	utorak	53	subota
9	petak	24	nedjelja	39	utorak	54	subota
10	petak	25	nedjelja	40	utorak	55	subota
11	petak	26	nedjelja	41	ponedjeljak	56	subota
12	subota	27	nedjelja	42	ponedjeljak	57	subota
13	subota	28	srijeda	43	nedjelja	58	subota
14	subota	29	četvrtak	44	nedjelja	59	subota
15	subota	30	četvrtak	45	četvrtak	60	subota

a) Koristeći proceduru Histogram grupirajte zadane podatke. Rezultate predočite tablicom.
b) Dobiveni statistički niz prikažite grafički jednostavnim stupcima.
c) Koliko kupaca kupuje srijedom?
d) Koliki udio kupaca kupuje srijedom?
e) Koliko kupaca kupuje vikendom (petak - nedjelja)?
f) Koliki udio kupaca kupnju obavi od ponedjeljka do srijede?

Rješenje:

U ovom primjeru obilježje je dan u tjednu. To je kvalitativno statističko obilježje. Međutim na jednostavan način te kvalitativne podatke numeriramo i time dobijemo mogućnost da ih grupiramo procedurom Histogram. Danima u tjednu pridružimo numeričke oznake na sljedeći način:

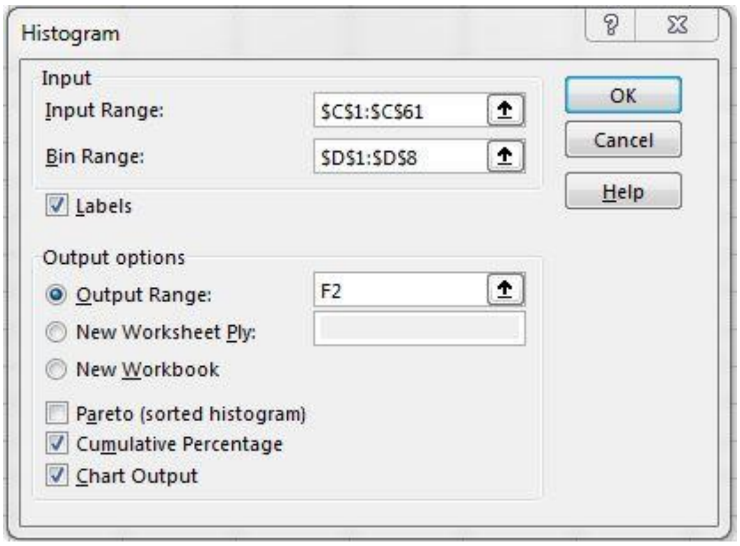
Ponedjeljak – 1, Utorak – 2, Srijeda – 3, Četvrtak – 4, Petak – 5, Subota – 6, Nedjelja – 7

Tako prenumerirani podaci mogu se grupirati na način opisan u prethodnom primjeru. Prvo moramo podatke prikazane na ovaj način kopirati u novi radni list tako da budu smješteni u prva dva stupca jedni ispod drugih. Potom im u susjednom stupcu pridijelimo odgovarajuće numeričke oznake. Zatim formiramo stupac (*bin range* – mogućih modaliteta) po kojem ćemo izvršiti razdiobu.

Ovako izgledaju pripremljeni podaci(dio njih) za obradu:

	A	B	C	D
1	**Redni broj**	**Dan**	**Dan**	**Dan**
2	1	ponedjeljak	1	1
3	2	ponedjeljak	1	2
4	3	utorak	2	3
5	4	srijeda	3	4
6	5	srijeda	3	5
7	6	četvrtak	4	6
8	7	četvrtak	4	7
9	8	petak	5	
10	9	petak	5	
11	10	petak	5	
12	11	petak	5	
13	12	subota	6	
14	13	subota	6	
15	14	subota	6	

Preko kartice *Data/Data Analysis/Histogram* dobijemo dijaloški okvir istoimene procedure koju popunimo na sljedeći način:



Nakon malo uređivanja izlazna tablica izgleda ovako:

Dan	*Broj kupaca*	*Kumulativne relativne frekvencije*
1	4	6,67%
2	4	13,33%
3	4	20,00%
4	7	31,67%
5	16	58,33%
6	15	83,33%
7	10	100,00%
More	0	100,00%

Već sada možemo ponuditi odgovore na pitanja:

c) Koliko kupaca kupuje srijedom (dan 3)? Odgovor je 4

f) Koliki udio kupaca kupnju obavi od ponedjeljka do srijede? Odgovor 20%

Kako bi odgovorili na preostala dva pitanja jasno je da moramo formirati stupac relativnih frekvencija i kumulativni niz 'veće od'. U prethodnom smo primjeru opisali način izrade pa prikazujemo samo gotovu tablicu:

Dan	*Broj kupaca*	*Relativne frekvencije*	*Kumulativni niz 'veće od'*	*Kumulativne relativne frekvencije*
1	4	6,67%	60	6,67%
2	4	6,67%	56	13,33%
3	4	**6,67%**	52	20,00%
4	7	11,67%	48	31,67%
5	16	26,67%	**41**	58,33%
6	15	25,00%	25	83,33%
7	10	16,67%	10	100,00%
More	0	0,00%	0	100,00%
	60	100,00%		

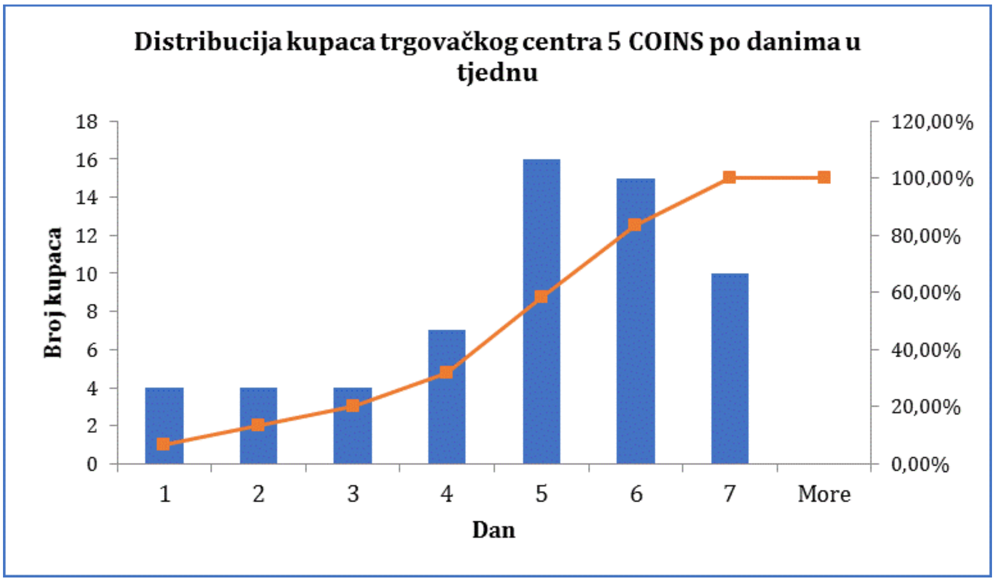
Možemo pronaći i odgovore na preostala dva pitanja:

d) Koliki udio kupaca kupuje srijedom(dan 3)? Odgovor je **6,67%** .

e) Koliko kupaca kupuje vikendom (petak-nedjelja)?

Odgovor pronalazimo u stupcu kumulativni niz 'veće od' (pitanje obuhvaća dane 5, 6 i 7) i glasi **41** (10+15+16).

Grafički prikaz dobivene razdiobe nakon opremanja izgleda ovako:



Vježba 2.3.

	A	B	C	D	E	F	G	H
1	Podaci o godinama starosti korisnika usluga Betanije,doma za starije i nemoćne osobe,dani su donjom tablicom.							
2	a) Grupirajte podatke o životnoj dobi korisnika u razrede širine 10 (godina). Rezultate grupiranja predočite tablicom.							
3	b) Dane podatke predočite grafički odgovarajućim grafikonom.							
4	c) Odredite udio osoba starih između 70 i 80 godina							
5	d) Odredite udio korisnika koji nisu stariji od 70 godina.							
6	e) Koliko je korisnika koji imaju maksimalno 60 godina?							
7	f) Koliko je korisnika starijih od 60 godina?							
8								
9	**Redni broj**	**Godine starosti**	**Redni broj**	**Godine starosti**	**Redni broj**	**Godine starosti**	**Redni broj**	**Godine starosti**
10	1	72,3	21	41,3	41	55,7	61	27,4
11	2	77,8	22	40,6	42	45,8	62	64,9
12	3	81,3	23	43,7	43	33,5	63	76,2
13	4	88,5	24	37,8	44	28,6	64	70,9
14	5	69,8	25	33,5	45	22,7	65	86
15	6	56,8	26	34,8	46	26,8	66	83,6
16	7	66,7	27	55,9	47	37,3	67	27,6
17	8	90,2	28	62,8	48	44,6	68	92,5
18	9	81,7	29	70,5	49	62,7	69	76,9
19	10	30,5	30	84,5	50	66,6	70	90,2
20	11	25,4	31	76,7	51	37,8	71	81,7
21	12	33,1	32	88,9	52	34,8	72	30,5
22	13	45,8	33	51,8	53	29,5	73	25,4
23	14	52,3	34	46,8	54	59,5	74	37,3
24	15	20,6	35	72,6	55	68,4	75	44,6
25	16	48,9	36	77,7	56	78,5	76	62,7
26	17	58,9	37	83,2	57	71,3	77	66,6
27	18	62,7	38	86,4	58	79,4	78	45,8
28	19	74,6	39	92,3	59	85,2	79	52,3
29	20	73,2	40	29,5	60	86,7	80	20,6

Rješenje:

U ovom primjeru na statističkom skupu korisnika doma za stare i nemoćne osobe Betanija promatramo kvantitativno obilježje *Godine starosti*. Međutim riječ je o kontinuiranom numeričkom obilježju. U tom slučaju prilikom upotrebe procedure *Histogram* nužno je podatke grupirati u razrede (nemoguće je i besmisleno navesti sve moguće modalitete tog obilježja).

Prvo moramo podatke predočene na ovaj način smjestiti u novi radni list tako da budu u nizu zapisani u prva dva stupca. Podaci su sada raspoređeni u ćelijama **B1:B81**. Da bismo mogli formirati razrede potrebno je prvo vidjeti koliko je stara najmlađa osoba, a koliko najstarija kako bi formirali dostatan broj razreda. Stoga negdje u slobodnoj ćeliji računamo **=MIN(B2:B81)**, a u drugoj **=MAX(B2:B81)** i vidimo da je najmlađa osoba stara 20,6 godina, a najstarija 92,5. Pri formiranju razreda moramo se voditi principom iscrpnosti i isključivosti. Potrebno je da svi podaci budu raspoređeni u razrede uključivo s minimalnom i maksimalnom vrijednošću. Shodno tome logično je formirati razrede (zadane širine 10) na sljedeći način:

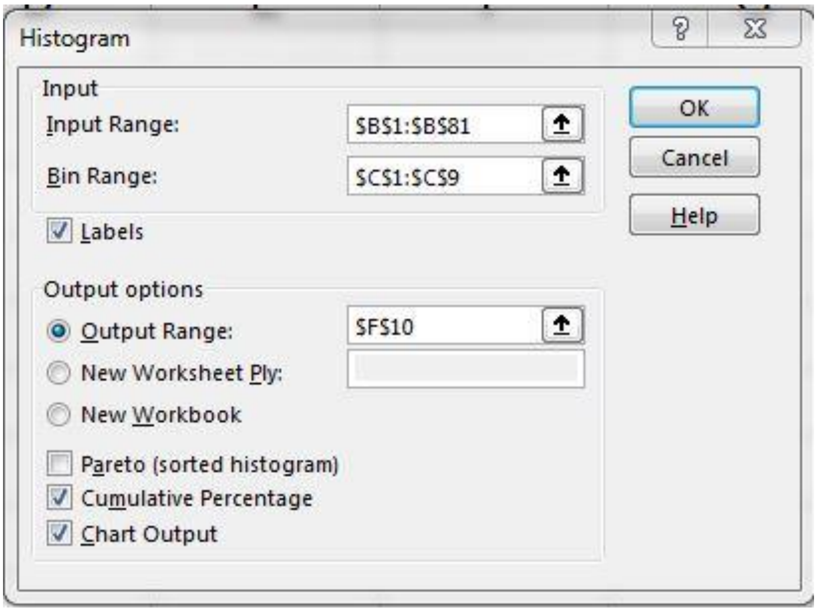
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100

	A	B	C	D	E
1	Redni broj	Godine starosti	Godine starosti		
2	1	72,3	30		20,6
3	2	77,8	40		92,5
4	3	81,3	50		
5	4	88,5	60		
6	5	69,8	70		
7	6	56,8	80		
8	7	66,7	90		
9	8	90,2	100		
10	9	81,7			
11	10	30,5			
12	11	25,4			
13	12	33,1			

Način zapisivanja razreda u proceduri *Histogram* podrazumijeva da u stupcu *bin range* pišemo samo gornje granice razreda. U tablici iznad prikazan je izgled radnog lista.

Pri tome treba imati na umu da razredi ne uključuju svoju donju granicu, a uključuju gornju. To znači da se npr. u razredu označenom sa 50 nalaze osobe starije od 40 godina, a mlađe ili s točno 50 godina. Matematički taj razred možemo zapisati u obliku poluotvorenog intervala (40,50].

Nakon što sve pripremimo, u dijaloškom okviru *Histograma* napravimo sljedeći odabir:



Dobijemo izlaznu tablicu koju malo doradimo.

Godine starosti	*Broj korisnika*	*Kumulativne rel. frekvencije %*
30	11	13,75%
40	11	27,50%
50	10	40,00%
60	8	50,00%
70	10	62,50%
80	14	80,00%
90	12	95,00%
100	4	100,00%
More	0	100,00%
	80	

Kako bi mogli odgovoriti na oba pitanja moramo dodati stupce relativnih frekvencija te kumulativne nizove 'veće od' i 'manje od'. Postupak nećemo ponovno opisati jer je identičan onima u prethodnim primjerima.

Godine starosti	*Broj korisnika*	*Kumulativni niz 'manje od'*	*Kumulativni niz 'veće od'*	*Relativne frekvencije*	*Kumulativne rel. frekvencije %*
30	11	11	80	13,75%	13,75%
40	11	22	69	13,75%	27,50%
50	10	32	58	12,50%	40,00%
60	8	**40**	48	10,00%	50,00%
70	10	50	**40**	12,50%	**62,50%**
80	14	64	30	**17,50%**	80,00%
90	12	76	16	15,00%	95,00%
100	4	80	4	5,00%	100,00%
More	0		0	0,00%	100,00%
	80				

Možemo odgovoriti na pitanja iz zadatka:

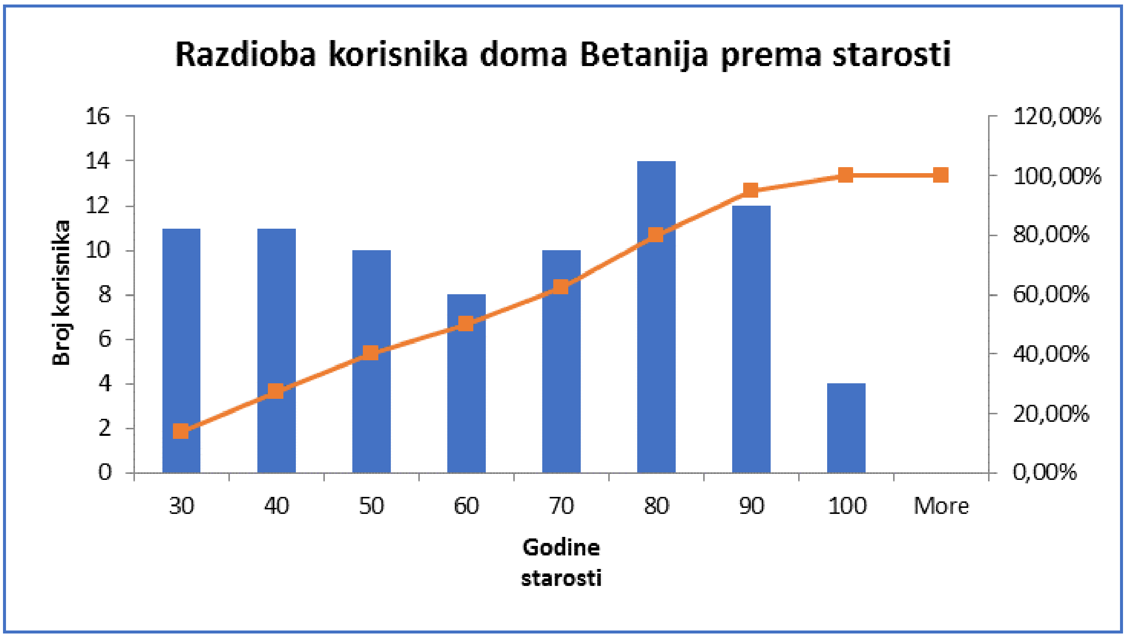
c) Odredite udio osoba starih između 70 i 80 godina? **17,50%**

d) Odredite udio korisnika koji nisu stariji od 70 godina? **62,50%**

e) Koliko je korisnika koji imaju maksimalno 60 godina? **40**

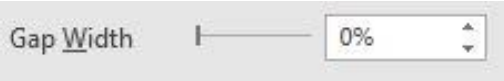
f) Koliko je korisnika starijih od 60 godina? **40**

Grafikon koji dobijemo izgleda (dodali smo naslov, oznake na osima):

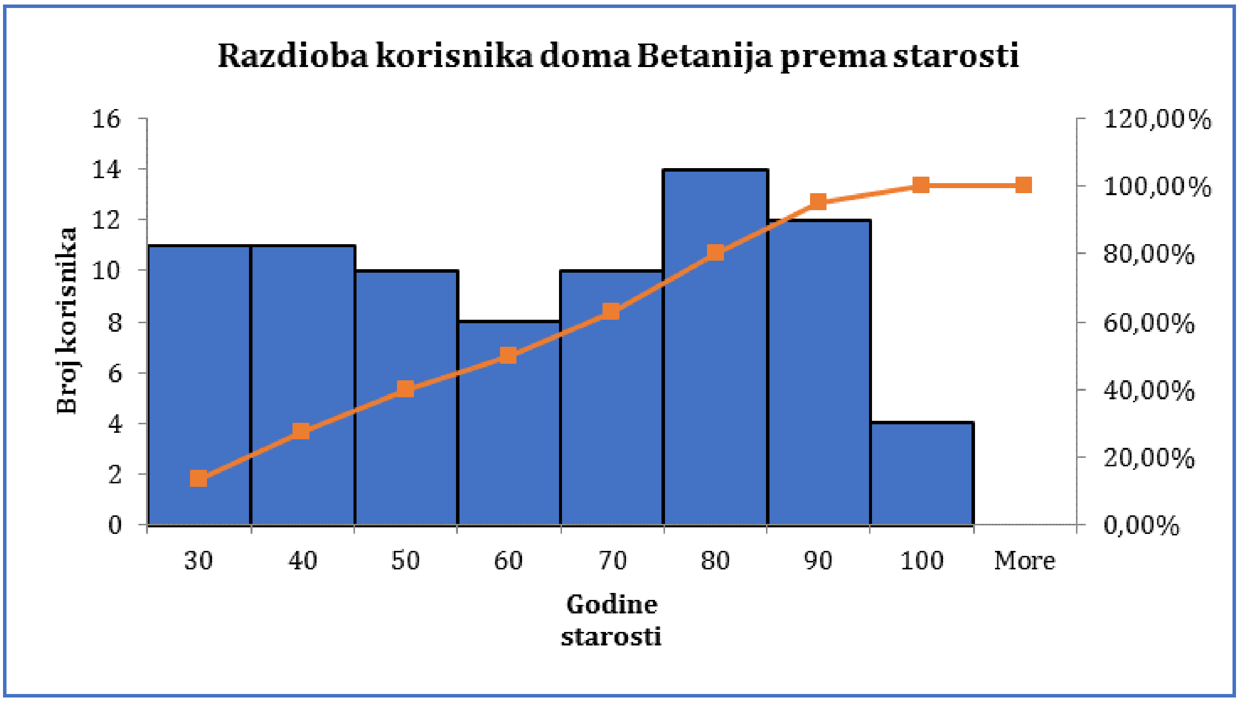


Budući se radi o kontinuiranom obilježju grupiranom u razrede vizualno moramo dočarati kontinuiranost pa radimo grafikon spojenih stupaca.

Postupak spajanja je sljedeći. Desnim klikom miša bilo gdje na površinu stupca dobijemo izbornik u kojem biramo *Format Data Series* u kojem smanjimo razmak među stupcima (*Gap width)* na 0%.



Time dobivamo željeni izgled:



Dodatno možete odgovoriti na pitanja:

1. Koliko je korisnika koje imaju najviše 50 godina?
2. Koliko je korisnika starijih od 40, a imaju maksimalno 50 godina?
3. Koliko je korisnika starijih od 40 godina?
4. Koliko je korisnika mlađih od 80 godina?
5. Koliki je udio korisnika starijih od 80, a s maksimalno 90 godina?
6. Koliki je udio korisnika koji imaju maksimalno 60 godina?

<u>Za vježbu riješite:</u>

U mapi **STATISTIKA.NOVO** možete riješiti:

Primjeri 2 (primjer 2.8 i primjer 2.9)

Zadaci za vježbu 2 (zadatak 2.5)

3. Pivot tablice

3.1. Dvodimenzionalne pivot tablice

U slučaju kada moramo prikazati dva ili više statistička niza koji su nastali grupiranjem istog statističkog skupa prema dva ili više obilježja koristimo proceduru ***Pivot Table.***

Vježba 3.1.

Podaci o spolu i životnoj dobi maloljetnih prekršitelja u gradu Pimat, dani su tablicom:

Redni broj	Spol	Dob	Redni broj	Spol	Dob	Redni broj	Spol	Dob
1	M	17,5	16	M	16,5	31	Ž	13,4
2	M	15,7	17	Ž	12,5	32	Ž	15,3
3	Ž	16,3	18	Ž	17,2	33	M	12,8
4	Ž	17,2	19	M	13,4	34	M	16,4
5	Ž	15,2	20	M	15,3	35	Ž	16,5
6	M	13,1	21	M	12,8	36	M	12,5
7	M	14,2	22	Ž	16,4	37	M	17,2
8	M	12,9	23	Ž	17,3	38	M	13,4
9	Ž	13,7	24	M	12,2	39	M	13,7
10	M	14,8	25	M	12,9	40	M	15,3
11	M	14,3	26	M	13,7	41	M	17,4
12	M	16,5	27	M	14,8	42	Ž	14,4
13	M	12,5	28	M	14,3	43	Ž	12,7
14	M	17,2	29	Ž	16,5	44	M	16,2
15	M	13,4	30	Ž	13,4	45	M	16

a) Formirajte razdiobu prekršitelja prema dobi i spolu (dob grupirajte u razrede širine 2 godine)
b) Odredite udio maloljetnih prekršitelja koji su muškog spola i starosti između 14 i 16 godina.
c) Među maloljetnim prekršiteljima koji su muškog spola odredite udio onih starosti između 14 i 16 godina.
d) Među maloljetnim prekršiteljima starosti između 14 i 16 godina odredi udio onih koji su muškog spola
e) Podatke iz pivot tablice prikažite grafički dvostrukim stupcima.
f) Strukturu prekršitelja prema dobi i spolu prikažite strukturnim stupcima.
g) Strukturu prekršitelja prema spolu prikažite strukturnim krugom

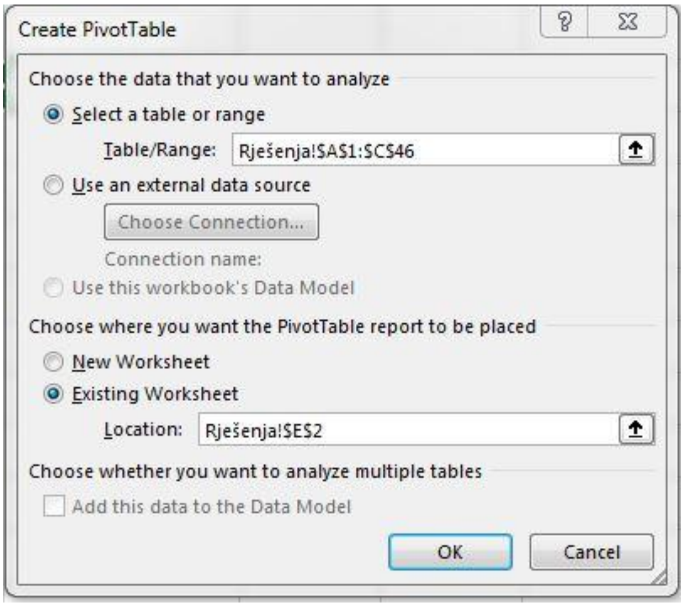
Rješenje:

Statistički skup: Maloljetni prekršitelji u gradu Pimat
Obilježje 1: Spol (kvalitativno, nominalno, atributivno, alternativno – ima samo 2 moguća modaliteta M ili Ž)
Obilježje 2: Dob (kontinuirano kvantitativno obilježje)
Opseg skupa: 45

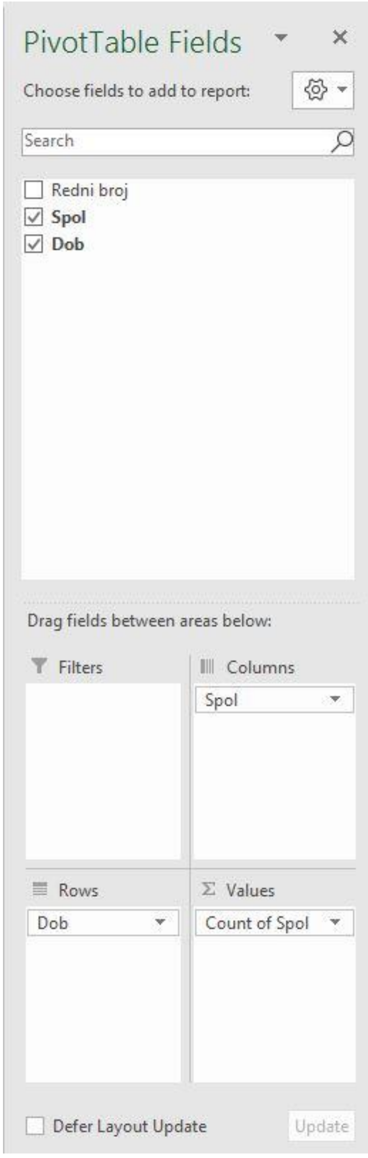
Kao što smo u uvodu napomenuli da bi napravili razdiobu jednog statističkog skupa po dva ili više obilježja, bez obzira da li je riječ o kvalitativnim ili kvantitativnim obilježjima, koristimo proceduru Pivot Table. Prije svega ovako prikazani podaci nisu pogodni za obradu pa ih na drugi radni list kopiramo na način da budu smješteni u prva tri stupca.

	A	B	C	D	E
1	**Redni broj**	**Spol**	**Dob**		
2	1	M	17,5		
3	2	M	15,7		
4	3	Ž	16,3		
5	4	Ž	17,2		
6	5	Ž	15,2		
7	6	M	13,1		
8	7	M	14,2		
9	8	M	12,9		
10	9	Ž	13,7		
11	10	M	14,8		
12	11	M	14,3		
13	12	M	16,5		
14	13	M	12,5		
15	14	M	17,2		
16	15	M	13,4		
17	16	M	16,5		
18	17	Ž	12,5		
19	18	Ž	17,2		
20	19	M	13,4		

Označimo blok podataka koji nam trebaju za formiranje tablice, a to su podaci u stupcima B(obilježje Spol) i C(obilježje Dob) točnije od **B1:C46**, dakle uključujemo i zaglavne ćelije. Najjednostavnije je ulazne podatke označiti samo lijevim klikom miša u bilo koju ćeliju unutar tablice. Zatim na kartici *Insert* odaberemo *Pivot Table* i dobijemo njegov dijaloški okvir u kojem je (ako smo već prethodno označili blok ćelija koje ulaze u obradu) potrebno samo naznačiti adresu ćelije u koju ćemo smjestiti dobivenu tablicu. Neka nama to bude **E2**

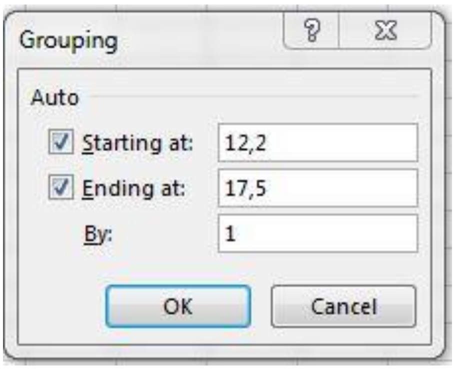


Sada dobijemo praznu tablicu čiji izgled sami definiramo. Na desnoj strani radnog lista imamo ponuđena obilježja (Spol i Dob) koja smjestimo (mišem odvučemo) u stupce odnosno retke. U polje s oznakom Σ Values dovučemo kvalitativno obilježje Spol jer će tada tablica vršiti prebrojavanje (razdioba frekvencija). Dobivena pivot tablica izgleda ovako:

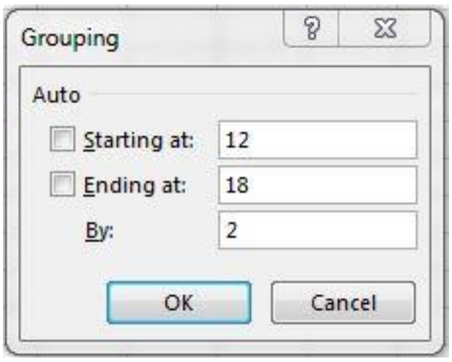


Count of Spol	Spol		
Dob	M	Ž	Grand Total
12,2	1		1
12,5	2	1	3
12,7		1	1
12,8	2		2
12,9	2		2
13,1	1		1
13,4	3	2	5
13,7	2	1	3
14,2	1		1
14,3	2		2
14,4		1	1
14,8	2		2
15,2		1	1
15,3	2	1	3
15,7	1		1
16	1		1
16,2	1		1
16,3		1	1
16,4	1	1	2
16,5	2	2	4
17,2	2	2	4
17,3		1	1
17,4	1		1
17,5	1		1
Grand Total	30	15	45

Ovako dobivena tablica je egzaktna i precizna, ali nepregledna pa ćemo kao što je u tekstu zadatka rečeno grupirati dob prekršitelja u razrede širine 2. Time ćemo izgubiti na preciznosti, ali dobiti na preglednosti podataka. Desnim klikom miša na polje *Dob* dobijemo plutajući izbornik u kojem biramo opciju *Group* koja nas vodi na novi dijaloški okvir:



Kao što je vidljivo najmanja vrijednost obilježja dob je 12,2 godine, a najveća 17,5. Postupak grupiranja je sljedeći. Treba definirati početak prvog razreda (*Starting at*), kraj zadnjeg razreda (*Ending at*) i širinu razreda (*By*). Pri tome je preporuka da granice razreda budu višekratnici širine razreda. Zato kao početak prvog razreda obično uzimamo prvi manji broj od najmanje vrijednosti koji je djeljiv s širinom razreda. U našem primjeru to je 12. Kao kraj zadnjeg razreda uzimamo prvi veći broj od najveće vrijednosti koji je djeljiv s širinom razreda. U našem primjeru to je 18.



Dobivena pivot tablica je sada pregledna i informativna.

Count of Spol	Spol		
Dob	M	Ž	Grand Total
12-14	13	5	18
14-16	8	3	11
16-18	9	7	16
Grand Total	30	15	45

Odgovorite na podpitanja:

Koliko je muških prekršitelja?
Koliko je prekršitelja ženskog spola koje su starosti od 16-18 godina?
Koliko je prekršitelja mlađih od 14 godina?

b) Odredite udio maloljetnih prekršitelja koji su muškog spola i starosti između 14 i 16 godina?

Da bi odgovorili na ovo pitanje potrebno je odrediti relativne frekvencije. Lako je vidljivo da maloljetnih prekršitelja muškog spola starosti od 14-16 godina ima 8. Međutim kako bi odredili relativni broj(frekvenciju) moramo znati na koji se opseg skupa on odnosi. Budući u pitanju nije drugačije naglašeno podrazumijeva se da se traži udio u cijelom skupu prekršitelja (njih 45).

Pivot tablice su jako korisne jer među ostalim mogu mijenjati prikaz podataka pa ćemo koristeći njihova svojstva i riješiti ovaj zadatak. Prvo kopirajmo izvornu pivot tablicu. Označimo je na način da dovedemo pokazivač miša na lijevi dio polja *Count of Spol* dok ne dobijemo crnu strelicu. Zatim je označimo pa na kartici *Home* odaberemo *Copy* i na odabrano (prazno) mjesto uz pomoć opcije *Paste* zalijepimo.

Zatim desnim klikom na polje *Count of Spol* dobijemo plutajući izbornik u kojem biramo *Show Values As.* Dakle biramo vrstu prikaza podataka u tablici. Obzirom da smo u prethodnom razmatranju zaključili da tražimo relativni udio u cijelom skupu biramo *% of Grand Total*

Dobivena tablica izgleda ovako:

Count of Spol	Spol		
Dob	M	Ž	Grand Total
12-14	28,89%	11,11%	40,00%
14-16	**17,78%**	6,67%	24,44%
16-18	20,00%	15,56%	35,56%
Grand Total	66,67%	33,33%	100,00%

Odgovor na pitanje koliki je udio prekršitelja muškog spola koji su stari od 14-16 godina glasi 17,78%.

c) Među maloljetnim prekršiteljima muškog spola odredi udio onih starosti od 14 do 16 godina?

Da bi odgovorili na ovo pitanje moramo odrediti još jednu relativnu strukturu, ali koju? Riječ je o udjelu prekršitelja starosnog raspona od 14 do 16 godina (njih 8) među svim prekršiteljima muškog spola (ima ih 30).

Ponovimo proceduru kopiranja Pivot tablice i opet desnim klikom na polje *Count of Spol* u plutajućem izborniku biramo *Show values as* ovog puta biramo *% of Column Total* jer smo obilježje Spol smjestili u stupce, a ono je skup u kojem tražimo udio. Tu relativnu strukturu nazivamo 'struktura stupca'.

Count of Spol	Spol		
Dob	M	Ž	Grand Total
12-14	43,33%	33,33%	40,00%
14-16	**26,67%**	20,00%	24,44%
16-18	30,00%	46,67%	35,56%
Grand Total	100,00%	100,00%	100,00%

Sada očitamo odgovor: među maloljetnim prekršiteljima muškog spola **26,67%** otpada na prekršitelje starosne dobi 14 – 16 godina.

d) Među maloljetnim prekršiteljima starosti između 14 i 16 godina odredi udio onih koji su muškog spola

Prekršitelji starosne dobi 14 – 16 godina (njih 11) su skup u kojem tražimo udio prekršitelja muškog spola (ima ih 8). Lako je vidjeti da treba naći 'strukturu retka', jer je obilježje Dob smješteno u retke Pivot tablice.

Ponavljamo proceduru kopiranja tablice kao i mijenjanja izlaznih vrijednosti (na ranije opisan način), ali ovog puta biramo prikaz podataka u tablici kao *% of Row Total.*

Count of Spol	Spol		
Dob	M	Ž	Grand Total
12-14	72,22%	27,78%	100,00%
14-16	**72,73%**	27,27%	100,00%
16-18	56,25%	43,75%	100,00%
Grand Total	66,67%	33,33%	100,00%

Odgovor: Među maloljetnim prekršiteljima starosne dobi od 14 do 16 godina **72,73%** je udio prekršitelja muškog spola.

e) Podatke iz pivot tablice prikažite grafički dvostrukim stupcima

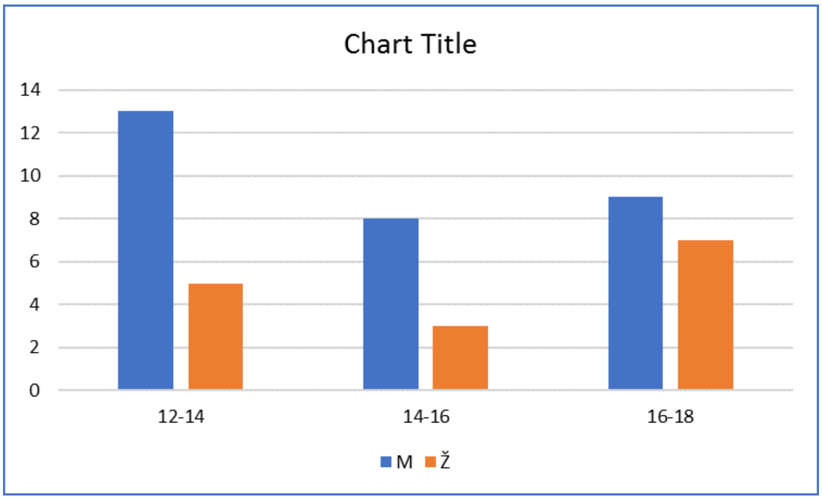
Grafički prikaz iz pivot tablice moguće je dobiti na više načina. Mi ćemo se odlučiti na jedan od njih. Označimo izvornu pivot tablicu s apsolutnim frekvencijama na prethodno opisani način, a zatim na kartici *Home* biramo

Copy i označimo neku slobodnu ćeliju. Zatim na kartici *Home* biramo *Paste(padajući izbornik)/Paste Values*, a potom tablicu uokvirimo .

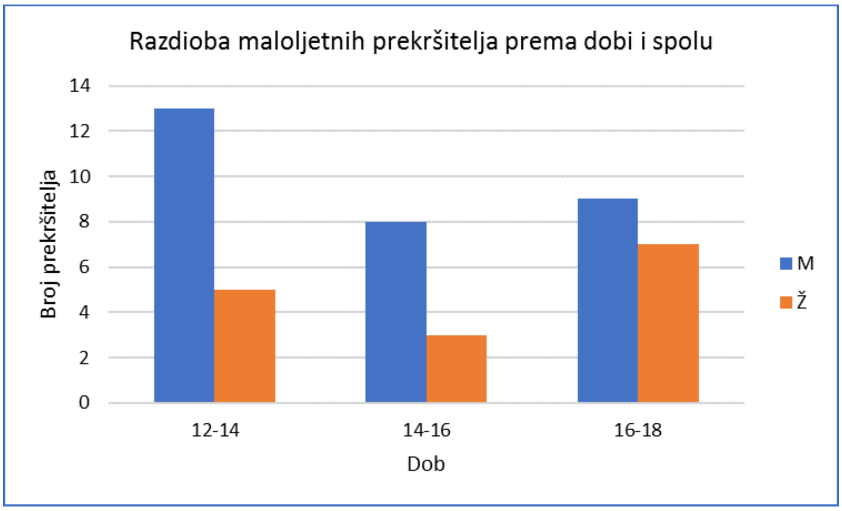
Time smo dobili običnu tablicu s alfanumeričkim podacima koju koristimo za grafičko prikazivanje razdiobe. To je složena statistička tablica iz koje smo već ranije (u poglavlju 1) naučili stvarati grafičke prikaze.

Count of Spol	Spol		
Dob	M	Ž	Grand Total
12-14	13	5	18
14-16	8	3	11
16-18	9	7	16
Grand Total	30	15	45

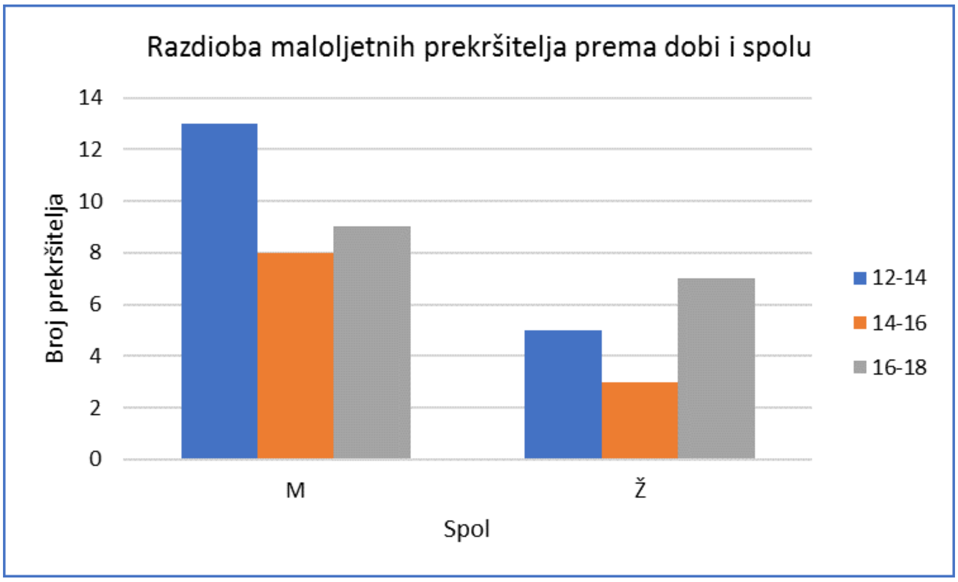
Označimo središnji dio tablice (bez totala po recima i stupcima) i idemo na *Insert/Column/1. Podtip.* Dobijemo graf koji prije uređivanja izgleda



Kao i ranije na kartici *Design* biramo odgovarajući *Quick Layout(npr. br 9)* i nakon uređivanja dobivamo konačni izgled:

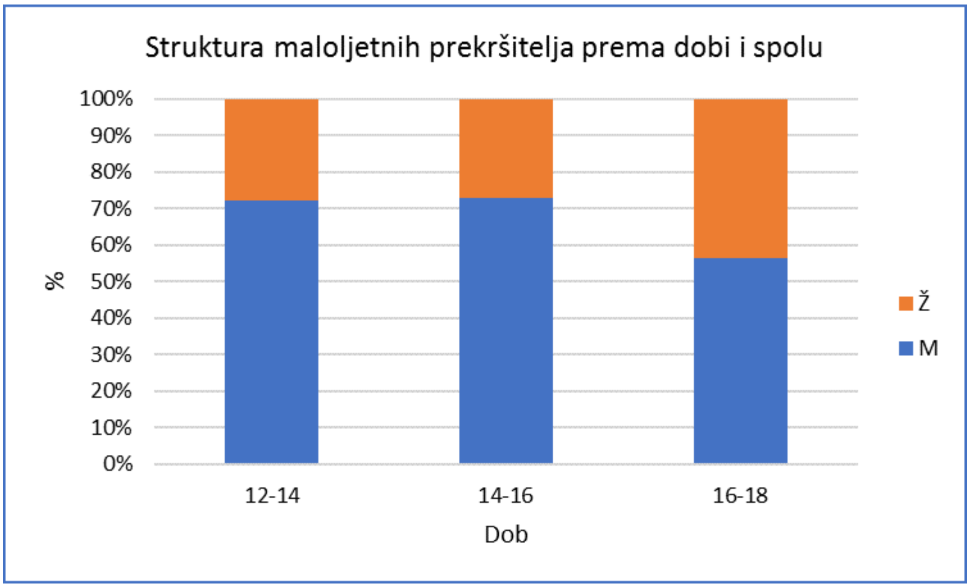


Dobili smo grafikon od tri skupine (dob) po 2 stupca (spol), međutim jednostavnom zamjenom možemo doći i do alternativnog prikaza. Na kartici *Design/Switch row/column* dobijemo dvije skupine po 3 stupca koji uz malo kozmetike izgleda:

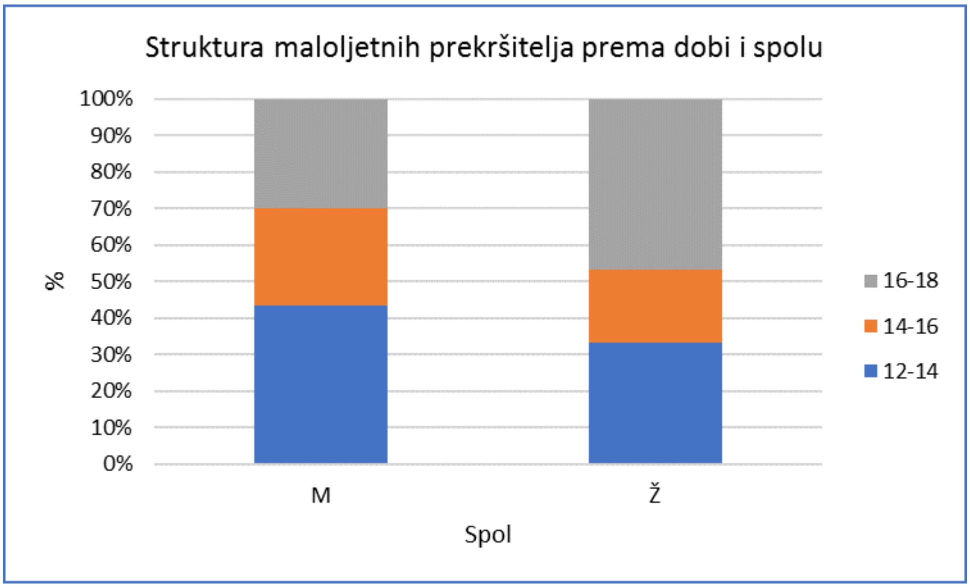


f) Strukturu prekršitelja prema dobi i spolu prikažite strukturnim stupcima

Moguće je i da relativne strukture retka odnosno stupca prikažemo odgovarajućim grafikonima – strukturnim stupcima. Oni predstavljaju grafički prikaz podzadataka c) i d). Postupak je sličan opisanom uz razliku što biramo 3. podtip u izborniku *Column* (stupce koji se protežu do vrha).



Zamjenu redaka/stupaca izvršimo kao ranije *Design/Switch Row/Column* i dobijemo:

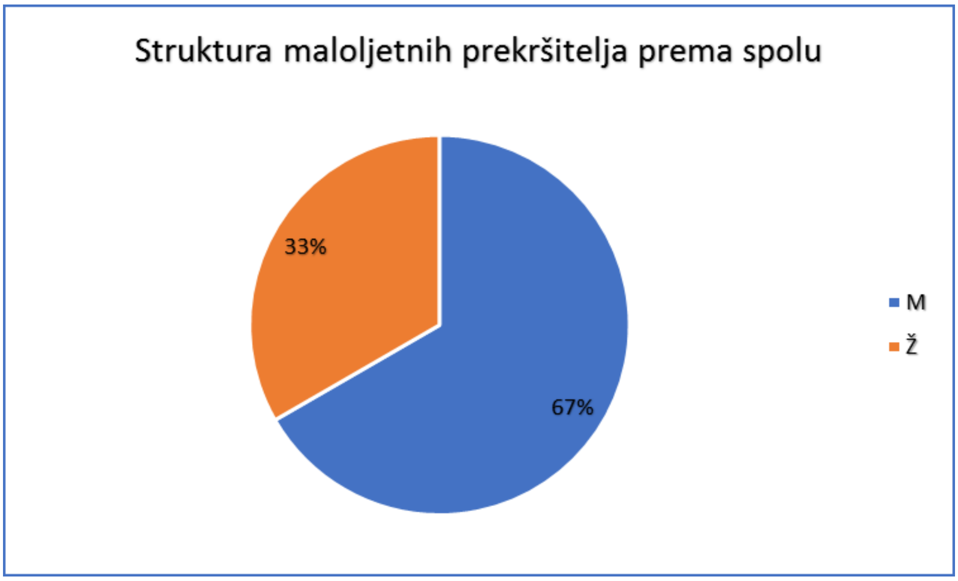


g) Strukturu prekršitelja prema spolu prikažite strukturnim krugom.

Ako želimo grafički prikazati (samo) jedno obilježje iz pivot tablice, u ovom primjeru samo spol tada u tablici za crtanje uz pomoć tipke *Ctrl* označimo samo modalitete tog obilježja i pripadne totale. U ovom slučaju:

Count of Spol	Spol		
Dob	M	Ž	Grand Total
12-14	13	5	18
14-16	8	3	11
16-18	9	7	16
Grand Total	30	15	45

Odaberemo *Insert/Pie* i opremimo graf naslovom, legendom i postotcima:



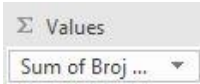
Vježba 3.2.

	A	B	C	D	E	F	G	H	I
1	Rezultati vikend kontrola prometne policije provedenih u Splitu tijekom 2018. godine dani su sljedećom tablicom.								
2									
3	**Redni broj**	**Broj prekršaja**	**Iznos naplaćenih kazni (u kn)**	**Redni broj**	**Broj prekršaja**	**Iznos naplaćenih kazni (u kn)**			
4	1	2	480	16	2	1600			
5	2	1	800	17	3	1800			
6	3	3	900	18	6	2400			
7	4	1	300	19	1	350			
8	5	4	1400	20	1	500			
9	6	2	500	21	5	1500			
10	7	5	1650	22	3	1500			
11	8	6	2000	23	2	800			
12	9	3	2100	24	4	1350			
13	10	4	1600	25	2	600			
14	11	1	600	26	3	1500			
15	12	7	3000	27	3	2400			
16	13	3	2400	28	1	200			
17	14	4	1200	29	1	350			
18	15	2	1000	30	5	1300			
19									
20	a) Formirajte dvodimenzionalnu razdiobu podataka (iznos kazni grupirajte u razrede širine 500, a broj prekršaja u razrede širine 2).								
21	b) Odredite udio prekršitelja koji su načinili između 3 i 4 prekršaja i za to platili između 1000 i 1500 kuna .								
22	c) Odredite udio prekršitelja koji su načinili između 3 i 4 prekršaja među prekršiteljima koji su platili između 1000 i 1500 kuna .								
23	d) Odredite udio prekršitelja koji su platili između 1000 i 1500 kuna među prekršiteljima s 3 do 4 prekršaja.								
24	e) Razdiobu prikažite grafički višestrukim stupcima.								
25	f) Strukturnim krugom prikažite strukturu prekršitelja prema broju prekršaja.								
26	g) Razdiobu prekršitelja prema iznosu naplaćenih kazni prikažite jednostavnim stupcima.								

Rješenje:

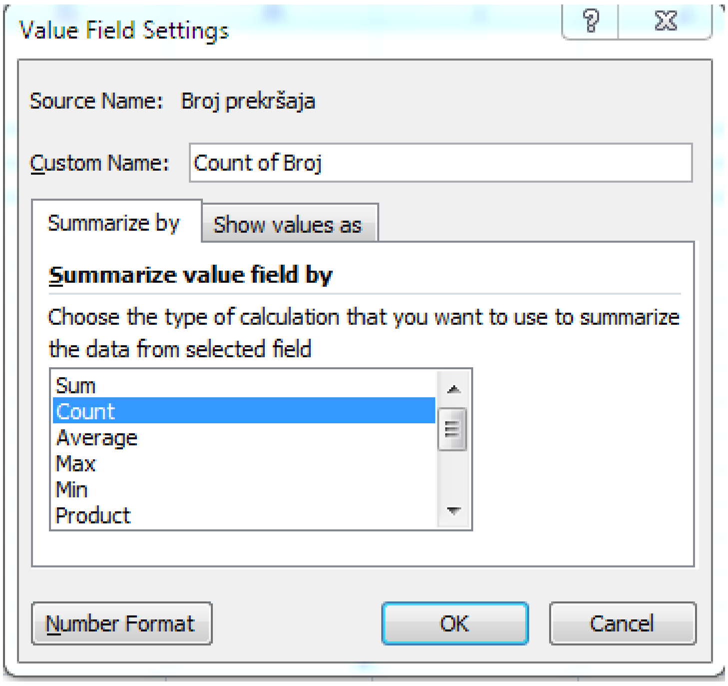
a) Kako bi formirali pivot tablicu potrebne podatke moramo kopirati u novi radni list u stupce, na sljedeći način:

	A	B	C
1	**Redni broj**	**Broj prekršaja**	**Iznos naplaćenih kazni (u kn)**
2	1	2	480
3	2	1	800
4	3	3	900
5	4	1	300
6	5	4	1400
7	6	2	500
8	7	5	1650
9	8	6	2000
10	9	3	2100
11	10	4	1600
12	11	1	600
13	12	7	3000
14	13	3	2400
15	14	4	1200
16	15	2	1000

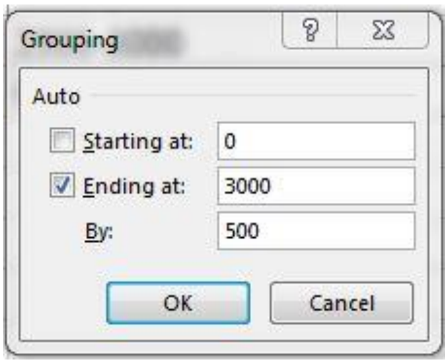
Označimo podatke koji ulaze u izradu pivot tablice (obilježja Broj prekršaja i Iznos kazni), **B1:C31.** Na kartici *Insert* biramo *Pivot Table* zatim u dijaloškom okviru određujemo gdje ćemo ga smjestiti, recimo u **E2**.
Dobili smo praznu tablicu koju sami kreiramo. Odvucimo Iznos kazni u retke, a Broj prekršaja u stupce. Budući su oba obilježja kvantitativna svejedno je koje ćemo staviti u polje ∑ *Values*. U oba slučaja pisat će 

, no kako želimo da tablica vrši prebrojavanje, moramo opciju sumiranja promijeniti u prebrojavanje (count). Lijevim klikom miša na polje *Sum of Broj* dobijemo padajući izbornik u kojem biramo *Value Field Settings* i na kartici *Summarize by* mijenjamo *Sum* u *Count* i potvrdimo izbor.

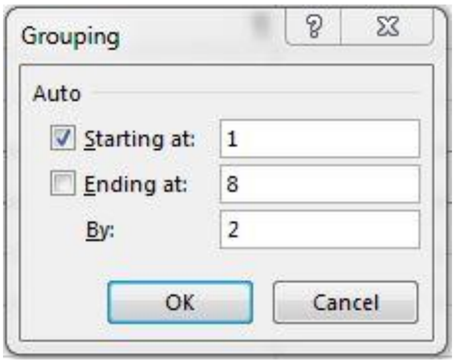
Sada smo dobili preciznu tablicu.



Radi bolje preglednosti potrebno je izvršiti grupiranje oba obilježja. Desnim klikom na obilježje *Iznos naplaćenih kazni* u padajućem izborniku biramo *Group*. U malom dijaloškom okviru vidljivo je da je najmanja kazna 200 pa kao početak prvog razreda biramo prvi manji broj djeljiv s 500 (zadana širina razreda), a to je 0. Gornja granica 3000 već je djeljiva s 500 pa je ostavljamo. U polje *By* upisujemo širinu razreda 500.



Na sličan način grupiramo i obilježje Broj prekršaja:



Dobivena pivot tablica izgleda:

Count of Broj	Broj prekršaja				
Iznos naplaćenih kazni(u kn)	1-2	3-4	5-6	7-8	Grand Total
0-499	5				5
500-999	6	1			7
1000-1499	1	3	1		5
1500-1999	1	4	2		7
2000-2499		3	2		5
2500-3000				1	1
Grand Total	13	11	5	1	30

Postupak računanja relativnih frekvencija nećemo posebno opisivati jer je analogan onom u prethodnom primjeru. Samo ćemo ponuditi rješenja.

b) Odredite udio prekršitelja koji su načinili između 3 do 4 prekršaja i za to platili između 1000 i 1500 kuna?

Očigledno se traži udio prekršitelja s ove dvije osobine u populaciji svih prekršitelja (*% of Grand Total*) pa rješenje izgleda:

Count of Broj	Broj prekršaja				
Iznos naplaćenih kazni(u kn)	1-2	3-4	5-6	7-8	Grand Total
0-499	16,67%	0,00%	0,00%	0,00%	16,67%
500-999	20,00%	3,33%	0,00%	0,00%	23,33%
1000-1499	3,33%	**10,00%**	3,33%	0,00%	16,67%
1500-1999	3,33%	13,33%	6,67%	0,00%	23,33%
2000-2499	0,00%	10,00%	6,67%	0,00%	16,67%
2500-3000	0,00%	0,00%	0,00%	3,33%	3,33%
Grand Total	43,33%	36,67%	16,67%	3,33%	100,00%

c) Odredite udio prekršitelja koji su načinili između 3 i 4 prekršaja među prekršiteljima koji su platili između 1000 i 1500 kuna ?

Sada su populacija u kojoj tražimo udio svi prekršitelji koji su platili kaznu od 1000 do 1500 kn, a kako su nam smješteni u recima onda biramo strukturu retka (*% of Row Total*).

Count of Broj	Broj prekršaja				
Iznos naplaćenih kazni(u kn)	1-2	3-4	5-6	7-8	Grand Total
0-499	100,00%	0,00%	0,00%	0,00%	100,00%
500-999	85,71%	14,29%	0,00%	0,00%	100,00%
1000-1499	20,00%	**60,00%**	20,00%	0,00%	100,00%
1500-1999	14,29%	57,14%	28,57%	0,00%	100,00%
2000-2499	0,00%	60,00%	40,00%	0,00%	100,00%
2500-3000	0,00%	0,00%	0,00%	100,00%	100,00%
Grand Total	43,33%	36,67%	16,67%	3,33%	100,00%

d) Odredite udio prekršitelja koji su platili između 1000 i 1500 kuna među prekršiteljima s 3 do 4 prekršaja?

Populacija u kojoj tražimo udio su svi prekršitelji koji su napravili 3 do 4 prekršaja pa ovaj put radimo strukturu stupca (jer je obilježje Broj prekršaja smješteno u stupcima), dakle *% of Column Total*.

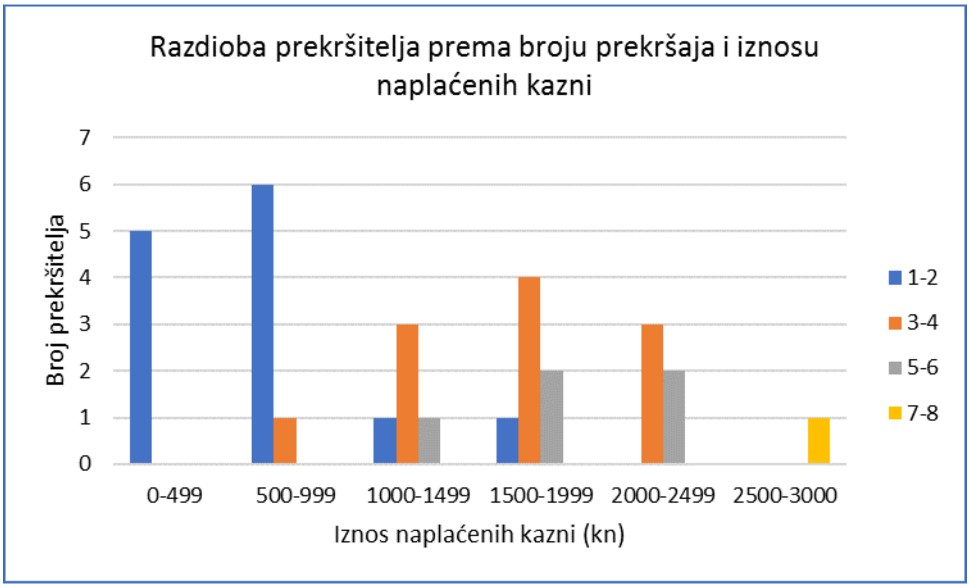
Count of Broj	Broj prekršaja				
Iznos naplaćenih kazni(u kn)	1-2	3-4	5-6	7-8	Grand Total
0-499	38,46%	0,00%	0,00%	0,00%	16,67%
500-999	46,15%	9,09%	0,00%	0,00%	23,33%
1000-1499	7,69%	**27,27%**	20,00%	0,00%	16,67%
1500-1999	7,69%	36,36%	40,00%	0,00%	23,33%
2000-2499	0,00%	27,27%	40,00%	0,00%	16,67%
2500-3000	0,00%	0,00%	0,00%	100,00%	3,33%
Grand Total	100,00%	100,00%	100,00%	100,00%	100,00%

e) Razdiobu prikažite grafički višestrukim stupcima.

Za grafički prikaz kopiramo izvornu tablicu oslobođenu pivot svojstava (s *Paste Values*) i označimo središnji dio, bez totala i zaglavnog retka.

Count of Broj	Broj prekršaja				
Iznos naplaćenih kazni(u kn)	1-2	3-4	5-6	7-8	Grand Total
0-499	5				5
500-999	6	1			7
1000-1499	1	3	1		5
1500-1999	1	4	2		7
2000-2499		3	2		5
2500-3000				1	1
Grand Total	13	11	5	1	30

Biramo *Insert/Column(1. podtip)* i nakon uređivanja grafikon izgleda:

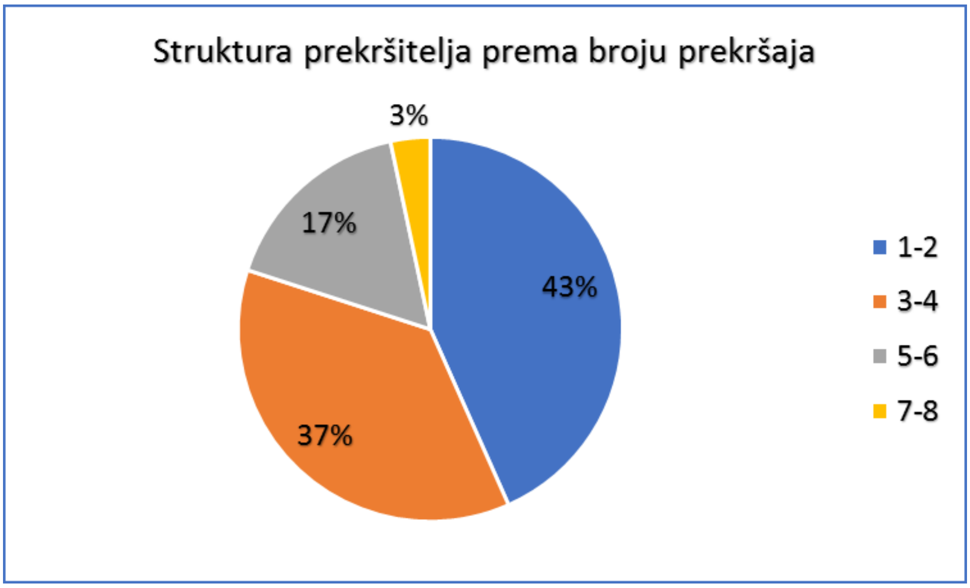


f) Strukturnim krugom prikažite strukturu prekršitelja prema broju prekršaja.

Kako bi prikazali skup po samo jednom obilježju potrebno je u tablici za crtanje označiti samo modalitete i pripadne frekvencije. To za ovaj slučaj izgleda ovako:

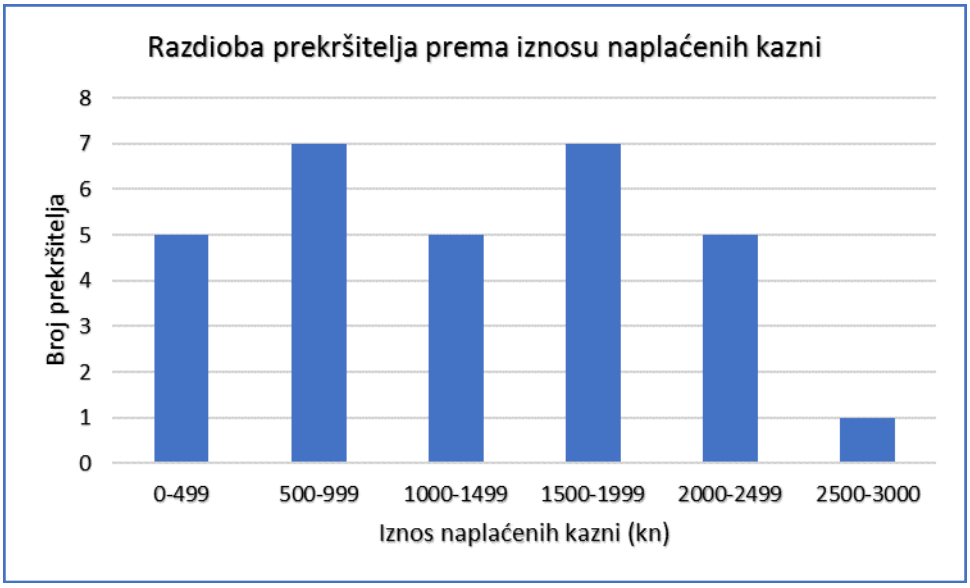
Count of Broj	Broj prekršaja				
Iznos naplaćenih kazni(u kn)	1-2	3-4	5-6	7-8	Grand Total
0-499	5				5
500-999	6	1			7
1000-1499	1	3	1		5
1500-1999	1	4	2		7
2000-2499		3	2		5
2500-3000				1	1
Grand Total	13	11	5	1	30

Biramo *Insert/Pie(1. Podtip).* Nakon opremanja grafa dobijemo:



g) Razdiobu prekršitelja prema iznosu naplaćenih kazni prikažite jednostavnim stupcima.

Na sličan način kao u prethodnom podzadatku potrebno je označiti modalitete obilježja iznos naplaćenih kazni i pripadne totale. Izborom traženog grafičkog prikaza i opremanjem dobivamo:



Za vježbu riješite:

U mapi **STATISTIKA.NOVO** možete riješiti Primjeri 2 (primjer 2.10).

U mapi **Primjeri** Primjer 2c.

3.2. Višedimenzionalne pivot tablice

Vježba 3.3.

	A	B	C	D
1	Masa (u kg)	Spol	Osnovna škola	Godina rođenja
2	18,20	M	Spinut	2012
3	20,21	M	Skalice	2013
4	19,14	Ž	Spinut	2012
5	21,20	M	Manuš	2013
6	22,30	Ž	Spinut	2011
7	24,80	Ž	Skalice	2013
8	24,00	Ž	Spinut	2012
9	20,90	Ž	Manuš	2013
10	19,80	Ž	Skalice	2013
11	22,40	M	Skalice	2012
12	24,30	M	Spinut	2012
13	24,00	Ž	Spinut	2012
14	20,00	M	Spinut	2013
15	22,80	Ž	Manuš	2013
16	22,00	M	Skalice	2013
17	24,10	M	Spinut	2012
18	24,80	Ž	Spinut	2012
19	24,00	M	Manuš	2013
20	19,90	Ž	Skalice	2012
21	25,00	M	Skalice	2013
22	18,80	M	Spinut	2013
23	18,90	Ž	Spinut	2012
24	20,30	M	Manuš	2011
25	20,40	Ž	Skalice	2012
26	22,75	M	Spinut	2013
27	22,70	M	Manuš	2013
28	22,70	Ž	Manuš	2012
29	22,00	M	Skalice	2013
30	24,10	M	Spinut	2012
31	24,30	Ž	Skalice	2012
32	24,00	M	Spinut	2013

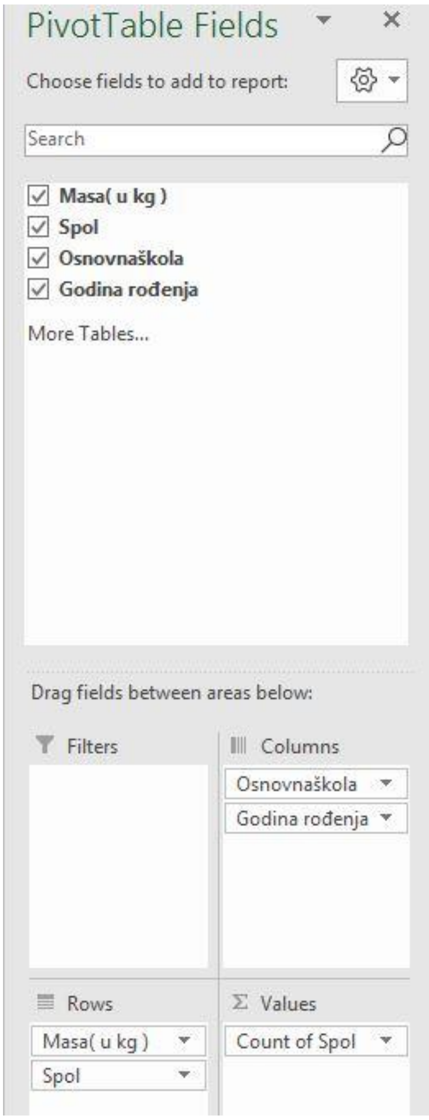
Za 64 predškolca u Splitsko dalmatinskoj županiji dani su podaci
o masi (kg), spolu, osnovnoj školi koju upisuju i godini rođenja.
Napravite višedimenzionalnu razdiobu po svim obilježjima.
Masu grupirajte u razrede širine 2 kg.
Odgovorite na pitanja:
a) Koliko je predškolaca muškog spola prijavljeno u školu Spinut?
b) Koliko je predškolaca ženskog spola koje su rođene 2012. godine?
c) Koliko je predškolaca muškog spola koji pohađaju školu Skalice, a imaju 20-22 kg?
d) Koliko je predškolaca ženskog spola koji imaju masu 18 do 20 kg i rođeni su 2013. godine?
e) Koliko je predškolaca prijavljenih u školu Spinut rođeno 2012. godine?
f) Koliko je predškolaca ženskog spola?

Rješenje:

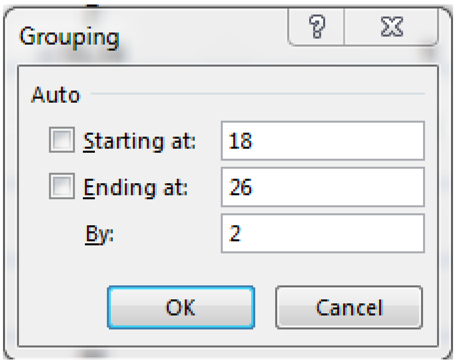
U ovom primjeru na statističkom skupu predškolaca (školskih obveznika) promatramo 4 obilježja (masu, spol, osnovnu školu, godinu rođenja).

Kako bi napravili razdiobu po sva 4 obilježja također ćemo se poslužiti pivot procedurom. Označit ćemo podatke jednostavnim klikom u bilo koju ćeliju unutar tablice, zatim *Insert/Pivot Table* i u dobivenom dijaloškom okviru izaberemo neku ćeliju kao gornji lijevi kut tablice (npr. **F14**). Sada smo dobili prazni predložak tablice i desni prozor za kreiranje tablice. Obzirom da imamo 4 obilježja logično je smjestiti po dva u retke i stupce. Treba voditi računa da je u prozorčiću *Values* opcija *Count of...*

Jedan od načina organizacije tablice ponuđen je u dijaloškom okviru:



Dobivena tablica je prevelika i nepregledna. Kako bi dobili na preglednosti izvršit ćemo grupiranje numeričkog obilježja masa. Desnim klikom na bilo koji modalitet obilježja u padajućem izborniku biramo *Group* i dobijemo dijaloški okvir u kojem postavimo granice razreda u skladu sa zadanom širinom 2 kg.



Count of Spol	Column La											
	Manuš			Manuš Total	Skalice		Skalice Total	Spinut			Spinut Total	Grand Total
Row Labels	2011	2012	2013		2012	2013		2011	2012	2013		
18-20		**2**	**1**	**3**	**1**	**4**	**5**		**5**	**1**	**6**	**14**
M		1	1	2		2	2		3	1	4	8
Ž		1		1	1	2	3		2		2	6
20-22	**1**		**4**	**5**	**4**	**2**	**6**		**1**	**4**	**5**	**16**
M	1		2	3	2	1	3		1	4	5	11
Ž			2	2	2	1	3					5
22-24	**1**	**3**	**3**	**7**	**2**	**2**	**4**	**1**	**2**	**3**	**6**	**17**
M			2	2	1	2	3		1	1	2	7
Ž	1	3	1	5	1		1	1	1	2	4	10
24-26			**2**	**2**	**2**	**4**	**6**		**7**	**2**	**9**	**17**
M			2	2		3	3		3	1	4	9
Ž					2	1	3		4	1	5	8
Grand Total	**2**	**5**	**10**	**17**	**9**	**12**	**21**	**1**	**15**	**10**	**26**	**64**

Odgovorite na pitanja:

a) Koliko je predškolaca muškog spola prijavljeno u školu Spinut?
b) Koliko je predškolaca ženskog spola koje su rođene 2012. godine?
c) Koliko je predškolaca muškog spola koji pohađaju školu Skalice, a imaju 20-22 kg?
d) Koliko je predškolaca ženskog spola koji imaju masu 18 do 20 kg i rođeni su 2013. godine?
e) Koliko je predškolaca prijavljenih u školu Spinut rođeno 2012. godine?
f) Koliko je predškolaca ženskog spola?

I u slučaju višedimenzionalne razdiobe lako je dobiti tablice relativnih frekvencija, ali nećemo ih u ovom razmatranju izrađivati.

Za vježbu riješite:

U mapi **Primjeri** Primjer 2d.

4. Srednje i položajne vrijednosti numeričkih nizova

Zbog izrazite sklonosti ka gomilanju oko neke istaknute vrijednosti, moguće je niz varijabilnih podataka prikazati jednom dominantnom, srednjom vrijednošću. Tu zamjensku, predstavničku ulogu će uspješno obaviti ona vrijednost koja je krajnje tipična varijabilnim podacima iste vrste. Obzirom na način izbora, srednje vrijednosti se dijele na potpune i položajne. U izračun potpunih srednjih vrijednosti su uključeni svi članovi niza. Neke od potpunih srednjih vrijednosti su:

- aritmetička sredina
- geometrijska sredina
- harmonijska sredina

Položajne srednje vrijednosti su određene svojim mjestom u uređenom nizu podataka. To su:

- medijan
- kvartili
- decili
- per(centili)

Odabir srednjih vrijednosti i njihova primjena ovise o prirodi statističkih podataka i vrsti obilježja. Naime, u slučaju kvalitativnog obilježja nisu moguće računske operacije nad modalitetima te je time krajnje sužen izbor srednjih vrijednosti. Kod kvantitativnih varijabli takvih ograničenja nema, ali zato neke od njih pokazuju izuzetnu osjetljivost na ekstremne vrijednosti obilježja. U MS Excelu koji koristimo postoje gotove funkcije za izračunavanje svih srednjih vrijednosti negrupiranih podataka. Nažalost, tako nije i sa grupiranim podacima pa ćemo ih morati sami izračunati pomoću odgovarajućih formula.

Na statističkom skupu prvašića promatramo obilježje masa (u kg). Kako bi mogli odgovoriti na zadana pitanja kopirat ćemo podatke u novi radni list tako da budu smješteni u prva dva stupca.

Vježba 4.1.

Podaci o masi djece upisane u prvi razred dani su slijedećom tablicom:

Redni broj	Masa (u kg)	Redni broj	Masa (u kg)	Redni broj	Masa (u kg)	Redni broj	Masa (u kg)
1	18,20	17	24,80	33	24,00	49	18,80
2	20,21	18	24,00	34	23,80	50	18,90
3	19,14	19	19,90	35	24,30	51	20,30
4	21,20	20	25,00	36	24,00	52	20,40
5	22,30	21	18,80	37	20,00	53	22,75
6	24,80	22	18,90	38	22,80	54	22,70
7	24,00	23	20,30	39	22,00	55	19,14
8	20,90	24	20,40	40	24,10	56	21,20
9	19,80	25	22,75	41	18,80	57	18,20
10	22,40	26	22,70	42	18,90	58	20,21
11	24,30	27	22,70	43	20,30	59	19,14
12	24,00	28	22,00	44	20,40	60	21,20
13	20,00	29	24,10	45	22,75	61	22,30
14	22,80	30	24,30	46	22,70	62	24,80
15	22,00	31	24,00	47	19,14	63	24,00
16	24,10	32	23,80	48	21,20	64	20,90

a) Odredite prosječnu masu prvašića.
b) Kolika je najčešća masa djece upisane u prvi razred?
c) Kolika je maksimalna masa prvih 50% djece razvrstane po masi?
d) Odredite vrijednost koja zadani niz podataka dijeli u omjeru 1 : 3 te onu koja isti niz dijeli u omjeru 3 : 1.
e) Odredite koja je najmanja masa koju ima 30% najteže djece.
f) Odredite najveću masu koju ima prvih 55%, po težini razvrstane, djece.

Redni broj	Masa (u kg)
1	18,20
2	20,21
3	19,14
4	21,20
5	22,30
6	24,80
7	24,00
8	20,90
9	19,80
10	22,40
11	24,30
12	24,00

a) Odredite prosječnu masu prvašića.
b) Kolika je najčešća masa djece upisane u prvi razred?
c) Kolika je maksimalna masa prvih 50% djece razvrstane po masi?
d) Odredite vrijednost koja zadani niz podataka dijeli u omjeru 1 : 3 te onu koja isti niz dijeli u omjeru 3 : 1
e) Odredite koja je najmanja masa koju ima 30% najteže djece.
f) Odredite najveću masu koju ima prvih 55%, po težini razvrstane, djece.

Rješenje:

a) Odredite prosječnu masu prvašića (artimetička sredina)?

Aritmetička sredina (obično je zovemo prosjek) je najvažnija i najčešće korištena potpuna srednja vrijednost. Izračunava se kao omjer zbroja svih vrijednosti obilježja i opsega statističkog skupa. Označava se sa $\bar{x}$. Za negrupirani niz podataka $x_1, x_2, x_3, \dots . x_N$ računamo aritmetičku sredinu kao $\bar{x} = \frac{x_1+x_2+\cdots+x_N}{N}$ gdje N predstavlja broj članova numeričkog niza.

Prosječna vrijednost numeričkog niza računa se u Excelu funkcijom AVERAGE. Sintaksa funkcije je **=AVERAGE(raspon podataka)** pa recimo u ćeliji **D10** računamo **=AVERAGE(B2:B65)** i potvrdimo s Enter, dobivamo rezultat 21,76.

Interpretacija: Prosječna masa prvašića iznosi 21,76 kg.

Prosječnu vrijednost (aritmetičku sredinu) mogli smo dobiti i na način da zbrojimo sve mase i podijelimo s opsegom skupa. Provjerite rezultat na taj način.

b) Kolika je najčešća masa djece upisane u prvi razred?

Najčešća vrijednost numeričkog niza je modalna vrijednost (mod). To je vrijednost koja se u numeričkom nizu pojavljuje najveći broj puta tj. ima najveću frekvenciju. Funkcija kojom izračunavamo mod u Excelu glasi MODE. Sintaksa funkcije je **=MODE(raspon podataka)** pa u ćeliji **D12** računamo **=MODE(B2:B65)** i potvrdimo s Enter, dobivamo rezultat 24.

Interpretacija: Najčešća masa djece upisane u prvi razred je 24 kg ili najveći broj prvašića ima masu 24 kg.

c) Kolika je maksimalna masa prvih 50% djece razvrstane po masi?

Ona položajna vrijednost koja numerički niz dijeli na dva jednaka dijela (u omjeru 1:1) tako da 50% članova niza ima manju ili jednaku vrijednost, a 50% veću ili jednaku naziva se MEDIJAN(srednja položajna vrijednost). Funkcija kojom u Excelu računamo medijan glasi MEDIAN. Sintaksa funkcije je **=MEDIAN(raspon podataka)** pa u ćeliji **D14** računamo **=MEDIAN(B2:B65)** potvrdimo s Enter, dobivamo rezultat 22.

Interpretacija: Maksimalna masa prvih 50% djece razvrstanih po masi iznosi 22 kg ili možemo reći da 50% djece (razvrstanih po masi) ima masu manju ili jednaku 22 kg, a preostalih 50% veću od 22 kg.

d) Odredite vrijednost koja zadani niz podataka dijeli u omjeru 1 : 3 te onu koja isti niz dijeli u omjeru 3 : 1.

Položajna vrijednost koja zadani niz dijeli u omjeru 1:3 naziva se prvi kvartil **Q1**. Ona zadani niz dijeli tako da je 25% vrijednosti manje ili jednako toj vrijednosti, a preostalih 75% veće ili jednako.

Položajna vrijednost koja niz dijeli u omjeru 3:1 naziva se treći kvartil **Q3**. Ona zadani niz dijeli tako da je 75% vrijednosti manje ili jednako toj vrijednosti, a preostalih 25% veće ili jednako. Kvartile u Excelu računamo pomoću funkcije QUARTILE. Sintaksa funkcije glasi **=QUARTILE(raspon podataka;željeni kvartil)** pa u ćeliji **D16** računamo **=QUARTILE(B2:B65;1)** potvrdimo i dobivamo 20,16.

Interpretacija: Vrijednost koja zadani niz dijeli u omjeru 1:3 je 20,16 kg što znači da 25% djece ima masu manju ili jednaku 20,16 kg, a preostalih 75% veću ili jednaku.

Sada računamo gornji kvartil(Q3) u ćeliji **D18** pa pišemo **=QUARTILE(B2:B65;3)** potvrdimo i dobivamo 24.

Interpretacija : Vrijednost koja zadani niz dijeli u omjeru 3:1 je 24 kg što znači da 75% djece ima masu manju ili jednaku 24 kg, a preostalih 25% veću ili jednaku

e) Odredite koja je najmanja masa koju ima 30% najteže djece.

Kako bi izračunali ovu vrijednost moramo izračunati funkciju PERCENTILE. To je ona položajna vrijednost koja izračunata za željeni postotak **P%** dijeli niz tako da P% članova ima manju ili jednaku vrijednost dobivenoj, a preostalih 100% - P% veću ili jednaku vrijednost. Sintaksa funkcije glasi **=PERCENTILE(raspon podataka;željeni postotak)**. Kako se traži najmanja masa 30% najteže djece moramo izračunati P70% jer je to zapravo vrijednost niza od koje 70% djece ima manju ili jednaku masu, a preostalih 30% veću. U ćeliji **D20** pišemo **=PERCENTILE(B2:B65;70%)** i dobivamo 22,9.

Interpretacija: Najmanja masa koju ima 30% najteže djece iznosi 22,90 kg ili 70% djece ima masu manju ili jednaku 22,9 kg, a preostalih 30% veću ili jednaku.

f) Odredite najveću masu koju ima prvih 55%, po masi razvrstane, djece.

Očigledno je da moramo izračunati P55% pa u ćeliji **D22** pišemo **=PERCENTILE(B2:B65;55%)** dobivamo 22,37.

Interpretacija: Najveća masa koju ima prvih 55% po masi razvrstane djece iznosi 22,37 kg, a preostalih 45% ima veću ili jednaku masu.

Primijetite da vrijedi:

Q1=P25%

Q2=MEDIAN=P50%

Q3=P75%

Riješeni primjer na radnom listu Excela izgleda:

	A	B	C	D	E	F	G	H	I	J
1	**Redni broj**	**Masa (u kg)**								
2	1	18,20		a) Odredite prosječnu masu prvašića.						
3	2	20,21		b) Kolika je najčešća masa djece upisane u prvi razred?						
4	3	19,14		c) Kolika je maksimalna masa prvih 50% djece razvrstane po masi?						
5	4	21,20		d) Odredite vrijednost koja zadani niz podataka dijeli u omjeru 1 : 3 te onu						
6	5	22,30		e) Odredite koja je najmanja masa koju ima 30% najteže djece.						
7	6	24,80		f) Odredite najveću masu koju ima prvih 55%, po težini razvrstane, djece.						
8	7	24,00								
9	8	20,90								
10	9	19,80		**21,76**	**ARITMETIČKA SREDINA**					
11	10	22,40								
12	11	24,30		**24**	**MOD**					
13	12	24,00								
14	13	20,00		**22,00**	**MEDIJAN**					
15	14	22,80								
16	15	22,00		**20,16**	**Q1**					
17	16	24,10								
18	17	24,80		**24**	**Q3**					
19	18	24,00								
20	19	19,90		**22,90**	**P70%**					
21	20	25,00								
22	21	18,80		**22,37**	**P55%**					

Vježba 4.2.

1	Rezultati ankete koju se studenti proveli na uzorku od 205 ispitanika, njemačkih državljana.		
2	Pitanje je glasilo:'Koliko ste spremni potrošiti na posjet zabavnom parku(bez ukljucene ulaznice)?'		
3	Podaci su obrađeni histogramom i rezultat je donja tablica:		
4			
5		Potrošnja (eur)	Broj ispitanika
6		10	48
7		20	86
8		30	54
9		40	7
10		50	8
11		60	2
12			
13			
14	Kolika je prosječna planirana potrošnja u zabavnom parku?		

Rješenje:

Promatrano obilježje je novčani iznos koji su ispitanici spremni potrošiti u zabavnom parku. Riječ je o kontinuiranom kvantitativnom obilježju. U ovom primjeru uočavamo da je riječ o grupiranim kvantitativnim podacima.

U slučaju grupiranih numeričkih podataka, odnosno distribucije frekvencija računamo vaganu ili ponderiranu srednju vrijednost koja se računa po formuli:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Apsolutne frekvencije imaju ulogu pondera. Ovaj način računanja srednje vrijednosti ćemo morati provesti sami jer u MS Excelu ne postoji direktna formula kao u slučaju negrupiranih podataka.

U slučaju podataka grupiranih u razrede, suočeni smo s problemom 'usrednjavanja' podataka na nivou svakog razreda s ciljem dobivanja 'predstavničke' vrijednosti, za nama nepoznate, podatke unutar razreda. To znači da ćemo sve vrijednosti unutar razreda zamijeniti sa sredinom razreda x_i koju dobijemo kao aritmetičku sredinu razrednih granica (donje i gornje). Formula za izračun te ponderirane srednje vrijednosti je, formalno, ista kao i u prethodnom slučaju.

Dakle prvi korak u izračunu srednje vrijednosti je određivanje razrednih sredina. Treba imati na umu da kod histogramom dobivenih razdioba zapisana granica predstavlja gornju granicu razreda. Dakle, možemo iščitati da je 48 ispitanika spremno potrošiti od 0 do 10 eura uključivo, njih 86 od 10 do 20 uključivo,... Obzirom da su razredi ekvidistantni dovoljno je odrediti prve dvije razredne sredine, a ostale dobiti 'razvlačenjem' aritmetičkog niza u Excelu. Dakle sredine će biti redom 5, 15, 25, ...

Drugi korak je formirati stupac u kojem množimo frekvencije i pripadne razredne sredine $f_i x_i$, slikovito rečeno uzimamo da 48 ispitanika planira potrošiti po 5 eura uprosječeno, 86 po 15 eura,...

Sumiranjem tog stupca dobit ćemo brojnik, a sumiranjem stupca frekvencija, nazivnik za izračun prosječne vrijednosti. Rezultat u Excelu:

	A	B	C	D	E	F	G
5		Potrošnja (eur)	Broj ispitanika fi	xi	fi*xi		
6		10	48	5	240		
7		20	86	15	1290		
8		30	54	25	1350		
9		40	7	35	245		
10		50	8	45	360		
11		60	2	55	110		
12			205		3595		
13							
14	Kolika je prosječna planirana potrošnja u zabavnom parku?						
15							
16			**17,54**	=E12/C12			
17	Prosječna potrošnja izračunata za razdiobu u razrede iznosi 17,54 eura.						

<u>Za vježbu riješite:</u>

U mapi **STATISTIKA.NOVO** možete riješiti **Primjeri 3** (primjer 3.1, primjer 3.2, primjer 3.6, primjer 3.7).

U mapi **Primjeri** Primjer 2c.

5. Mjere raspršenosti numeričkih nizova

5.1. Negrupirani podaci

Vježba 5.1.

Za 84 učesnika automobilske utrke podaci o maksimalnoj brzini dostignutoj tijekom utrke,dani su tablicom.
Izračunajte i objasnite značenje dobivenih vrijednosti:

Redni broj	Maksimalna brzina (u km/h)	Redni broj	Maksimalna brzina (u km/h)	Redni broj	Maksimalna brzina (u km/h)	Redni broj	Maksimalna brzina (u km/h)
1	179,43	22	260,70	43	260,00	64	280,60
2	201,56	23	300,00	44	198,00	65	230,70
3	240,50	24	280,60	45	220,58	66	287,00
4	199,80	25	230,70	46	188,50	67	250,00
5	200,70	26	287,00	47	198,50	68	240,37
6	250,38	27	250,00	48	199,00	69	239,50
7	250,00	28	240,37	49	200,30	70	244,30
8	260,00	29	239,50	50	290,00	71	199,90
9	198,00	30	244,30	51	287,00	72	245,80
10	220,58	31	199,90	52	280,60	73	310,00
11	188,50	32	245,80	53	230,70	74	223,70
12	198,50	33	310,00	54	287,00	75	220,58
13	199,00	34	223,70	55	250,00	76	188,50
14	200,30	35	201,56	56	240,37	77	198,50
15	290,00	36	240,50	57	239,50	78	199,00
16	287,00	37	199,80	58	244,30	79	200,30
17	250,00	38	200,70	59	199,90	80	290,00
18	240,37	39	250,38	60	245,80	81	287,00
19	239,50	40	250,00	61	310,00	82	280,80
20	244,30	41	260,00	62	223,70	83	276,98
21	260,00	42	270,70	63	265,90	84	256,78

a) raspon varijacije
b) interkvartil
c) koeficijent kvartilne devijacije
d) standardnu devijaciju
e) koeficijent varijacije
f) interpercentil
g) raspon varijacije središnjih 40% podataka

Rješenje:

Potrebno je izračunati:

a) raspon varijacije
b) interkvartil
c) koeficijent kvartilne devijacije
d) standardnu devijaciju
e) koeficijent varijacije
f) interpercentil
g) raspon varijacije središnjih 40% podataka

Mjere raspršenosti dijelimo na

1. Apsolutne mjere raspršenosti:

Raspon varijacije, interkvartil, varijanca i iz nje izvedena standardna devijacija. Apsolutne mjere disperzije se iskazuju u jedinicama mjere promatrane numeričke varijable.

2. Relativne mjere raspršenosti:

To su koeficijent varijacije i koeficijent kvartilne devijacije. One su neimenovani brojevi i obično se iskazuju postotkom ili decimalnim brojem manjim od 1.

Kopirajmo podatke u novi radni list tako da budu smješteni u stupcima A (Redni broj) i B(Maksimalna brzina u km/h). Računamo:

a) **Raspon varijacije R_x** je razlika između najveće i najmanje vrijednosti kvantitativnog obilježja.

Raspon varijacije je po definiciji **=MAX(raspon)-MIN(raspon)** pa npr. u ćeliji **E2** pišemo **=MAX(B2:B85)-MIN(B2:B85)**.

Dobivena vrijednost je 130,57 km/h što znači da razlika najveće i najmanje maksimalno dosegnute brzine tijekom utrke iznosi 130,57 km/h.

b) **Interkvartil I_Q** je apsolutna mjera raspršenosti koja pokazuje raspon varijacije središnjih 50% jedinica uređenog statističkog niza.

Slijedom toga računa se **I_Q =Q3-Q1**. Obzirom da će nam vrijednosti gornjeg Q3 i donjeg Q1 kvartila poslužiti i poslije uputno je prvo njih izračunati negdje na radnom listu, pa u npr. u ćeliji **H2** računamo gornji kvartil, odnosno u **H3** donji kvartil **=QUARTILE(B2:B85;3)**, **=QUARTILE(B2:B85;1)**. Dobili smo sljedeće vrijednosti:

260,175	Q3
200,7	Q1

Sada u ćeliji **E3** računamo interkvartil koristeći relativne adrese izračunatih vrijednosti **=H2-H3**. Dobivena vrijednost iznosi 59,475 km/h.

Možemo reći da raspon varijacije središnjih 50% maksimalno dosegnutih brzina tijekom utrke iznosi 59,48 km/h.

c) **Koeficijent kvartilne devijacije V_q** je relativna mjera raspršenosti središnjih 50% jedinica uređenog statističkog niza. Računa se iz relacije: $V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Koristeći već izračunate kvartile u ćeliji **E4** pišemo **=(H2-H3)/(H2+H3)**

Rezultat iznosi 0,129 što znači da je varijabilitet središnjih 50% maksimalno dosegnutih brzina relativno slab.

Interpretacija koeficijenta kvartilne devijacije prati sljedeću tablicu:

V_q	Varijabilitet
0 – 0,1	vrlo slab
0,1 – 0,2	relativno slab
0,2 – 0,3	umjeren
0,3 – 0,5	relativno jak
0,5 - 1	vrlo jak

d) Standardna devijacija (s ili σ) se definira kao prosječno odstupanje vrijednosti numeričke varijable od njene aritmetičke sredine.

U praksi najčešće korištena potpuna apsolutna mjera disperzije je standardna devijacija. Za razliku od prethodnih mjera disperzije raspona varijacije i interkvartila gdje koristimo samo dvije vrijednosti statističkog skupa, u određivanju standardne devijacije (i varijance) koriste se sve vrijednosti tog statističkog skupa. **Varijanca** (u oznaci s^2 ili σ^2) je prosječno kvadratno odstupanje pojedinih vrijednosti numeričkog obilježja od njihove aritmetičke sredine. Računa se po matematičkoj formuli $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}$. U Excelu se za negrupirane podatke računa funkcijom **VARP(raspon podataka)**, ali se u praksi ne interpretira jer ima kvadratnu dimenziju promatranog obilježja. Standardna devijacija predstavlja drugi korijen iz varijance i matematički se računa formulom $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}}$. U Excelu se računa funkcijom **STDEVP(raspon podataka)**.

Ove relacije za varijancu i standardnu devijaciju odnose se na čitavu populaciju. U situaciji kada ih računamo na uzorku izračun je nešto drugačiji (vidi u predavanjima). Dodatno treba naglasiti da su ovo relacije koje koristimo za negrupirani niz podataka. U sljedećem primjeru bit će prikazano kako se ove vriijednosti računaju za grupirani niz podataka.

Sada možemo u ćeliji **E5** izračunati standardnu devijaciju pa pišemo **=STDEVP(B2:B85)**.

Standardna devijacija iznosi 34,211 km/h što predstavlja prosječno odstupanje maksimalnih brzina od prosječne maksimalne brzine.

e) **Koeficijent varijacije V** je relativna mjera raspršenosti i predstavlja postotni udio standardne devijacije u odnosu na vrijednost aritmetičke sredine. Matematički se računa formulom $V = \frac{\sigma}{\bar{x}}$ (%).

U Excelu se računa **=STDEVP(raspon podataka)/AVERAGE(raspon podataka)** pa u ćeliji **E6** pišemo **=STDEVP(B2:B85)/AVERAGE(B2:B85)**.

Dobivenu vrijednost 0,1423 pritiskom na ikonicu % prikažemo u obliku postotka i dodamo dva decimalna mjesta.

Koeficijent varijacije iznosi 14,23% što predstavlja relativno slabu disperziju maksimalnih brzina.

Interpretaciju vršimo u skladu s tablicom:

V (%)	Varijabilitet
0 – 10	vrlo slab
10 – 30	relativno slab
30 – 50	umjeren
50 – 70	relativno jak
70 - 100	vrlo jak

f) **Interpercentil** se definira kao razlika **P90%-P10%** i predstavlja raspon varijacije središnjih 80% podataka

g) Raspon varijacije središnjih 40% podataka računa se kao **P70%-P30%.**

Kako odrediti raspon središnjih 40% podataka. Obzirom da nam je potrebno 40 % podataka, onda 'višak' od 60% (100%-40) simetrično raspodijelimo na početak uređenog statističkog niza (P30%) i na kraj (P70%) i tako dođemo do gornjeg rezultata.

Rješenje na radnom listu izgleda:

	A	B	C	D	E	F	G	H	I
1	**Redni broj**	**Maksimalna brzina (u km/h)**							
2	1	179,43		**R_x**	**130,57**	**Raspon varijacije**		260,18	**Q3**
3	2	201,56		**I_Q**	**59,48**	**Interkvartil**		200,70	**Q1**
4	3	240,50		**V_Q**	**0,129**	**Koef.kvart.devijacije**			
5	4	199,80		**σ**	**34,21**	**Standardna devijacija**		310,00	**MAX**
6	5	200,70		**V**	**14,23%**	**Koeficijent varijacije**		179,43	**MIN**
7	6	250,38		**P90%-P10%**	**88,35**	**Interpercentil**			
8	7	250,00		**P70%-P30%**	**36,522**	**Raspon varijacije središnjih 40% podataka**			
9	8	260,00							
10	9	198,00							
11	10	220,58							
12	11	188,50							
13	12	198,50							
14	13	199,00							
15	14	200,30							
16	15	290,00							
17	16	287,00							
18	17	250,00							
19	18	240,37							
20	19	239,50							
21	20	244,30							

Za vježbu riješite:

U mapi **STATISTIKA.NOVO** možete riješiti **Primjeri 4** (primjer 4.1, primjer 4.2, primjer 4.3).

U mapi **Primjeri vježbi 1, 2, 3, 4, 5** za vježbu riješite zadatak **Vježba 3.2.**

5.2. Grupirani podaci (razdioba frekvencija)

Vježba 5.2.

	A	B	C	D	E	F
1	*Plaće*			*Plaće*	*fi*	
2	5832			2000	2	
3	2517			4000	111	
4	2121			6000	49	
5	4524			8000	15	
6	4826			10000	11	
7	3741			12000	6	
8	2223			14000	2	
9	4567			16000	2	
10	9012			18000	2	
11	2525			ukupno		
12	3119					
13	4321					
14	3126		*NEGRUPIRANI (IZVORNI) NIZ*			
15	4673			aritmetička sredina		
16	3441			varijanca		
17	2751			standardna devijacija		
18	1987			koeficijent varijacije		
19	4432					
20	3245					
21	11605					
22	2137		*RAZDIOBA FREKVENCIJA*			
23	2422			aritmetička sredina		
24	2615			varijanca		
25	4118			standardna devijacija		
26	2325			koeficijent varijacije		
27	4331					
28	1990					
29	2531					
30	5489					
31	9021					
32	16500					

Rješenje:

U ovom primjeru za isti statistički skup izračunat ćemo aritmetičku sredinu, varijancu, standardnu devijaciju i koeficijent varijacije za negrupirani niz kao i za podatke grupirane u razrede. Riječ je o plaćama 200 zaposlenika neke tvornice (u kunama).

U tablici označenoj **NEGRUPIRANI(IZVORNI) NIZ** računamo koristeći ugrađene Excelove funkcije.

NEGRUPIRANI (IZVORNI) NIZ	
=AVERAGE(A2:A201)	**aritmetička sredina**
=VARP(A2:A201)	**varijanca**
=STDEVP(A2:A201)	**standardna devijacija**
=STDEVP(A2:A201)/AVERAGE(A2:A201)	**koeficijent varijacije**

Rezultat je dan u tablici:

NEGRUPIRANI (IZVORNI) NIZ	
4623,31	**aritmetička sredina**
7685021,17	**varijanca**
2772,19	**standardna devijacija**
59,96%	**koeficijent varijacije**

Kao što smo ranije naveli srednje vrijednosti i mjere raspršenosti drugačije računamo kada su podaci histogramom grupirani u razrede. Tako u slučaju grupiranih numeričkih podataka, odnosno razdiobe frekvencija, imamo vaganu ili ponderiranu **SREDNJU VRIJEDNOST** koja se računa po formuli

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

Gdje su: x_i – sredina i-tog razreda

n – ukupan broj razreda

f_i – frekvencija i-tog razreda

Podaci su prethodno (korištenjem procedure *Histogram)* grupirani u razrede širine 2000 kn. Možete za vježbu sami izvršiti postupak grupiranja. Ovdje je prikazana izlazna tablica.

Plaće	***fi***
2000	2
4000	111
6000	49
8000	15
10000	11
12000	6
14000	2
16000	2
18000	2
ukupno	

U Excelu ne postoje gotove funkcije za izračunavanje srednjih vrijednosti i mjera raspršenosti grupiranih podataka pa ih izračunavamo prateći matematičke izraze. Kako bi izračunali srednju vrijednost moramo prvo odrediti stupac sredina razreda $\boldsymbol{x_i}$. Sredine razreda dobijemo tako da zbrojimo donju i gornju granicu razreda i zbroj podijelimo s 2. Tako je sredina prvog razreda (0+2000)/2=1000, drugog (2000+4000)/2=3000,... prve dvije sredine upišemo u ćelije **F3** i **F4**, obe ćelije označimo i razvučemo do dna tablice. U susjednom stupcu računamo brojnik aritmetičke sredine pa u tu svrhu formiramo stupac umnožaka $x_i{\cdot}f_i$. U ćeliji **G2** računamo prvi umnožak **=E2*F2** .Kopiramo ovu formulu do dna stupca i dobivene umnoške prosumiramo koristeći ikonicu Σ AutoSum. Time smo dobili brojnik našeg izraza. Nazivnik je suma frekvencija koju dobijemo tako da zbrojimo sve frekvencije u stupcu **E**.

Rezultat ovih radnji dan je tablicom:

D	E	F	G
Plaće	***fi***	***xi***	***xi*fi***
2000	2	1000	2000
4000	111	3000	333000
6000	49	5000	245000
8000	15	7000	105000
10000	11	9000	99000
12000	6	11000	66000
14000	2	13000	26000
16000	2	15000	30000
18000	2	17000	34000
ukupno	**200**		**940000**

Na za to predviđeno mjesto (u ćeliju **C23**) sada možemo izračunati aritmetičku sredinu za razdiobu frekvencija. Pišemo **=G11/E11,** rezultat je 4700.

Varijancu za razdiobu frekvencija računamo izrazom $\sigma^2 = \frac{\sum_{i=1}^{n} f_i{\cdot}(x_i-\bar{x})^2}{\sum_{i=1}^{n} f_i}$

Kao što vidimo nazivnik već imamo izračunat u ćeliji **E11**, a kako bi izračunali brojnik moramo formirati novi stupac u tablici, stupac **H**, u kojem ćemo računati traženi umnožak ***f_i*(x_i – x_{pr})^2***. Radi jednostavnosti zapisa umjesto $\bar{x}$ pišemo x_{pr}. U ćeliju **H2** pišemo formulu koristeći relativne adrese **=E2*(F2-$C$23)^2**, a zatim kopiramo formulu u ostale ćelije stupca te zbrojimo sve vrijednosti u stupcu. Zbroj iznosi 1526000000. Tablica sada izgleda:

D	E	F	G	H
Plaće	**fi**	**xi**	**xi*fi**	**fi*(xi-x pr)^2**
2000	2	1000	2000	27380000
4000	111	3000	333000	320790000
6000	49	5000	245000	4410000
8000	15	7000	105000	79350000
10000	11	9000	99000	203390000
12000	6	11000	66000	238140000
14000	2	13000	26000	137780000
16000	2	15000	30000	212180000
18000	2	17000	34000	302580000
ukupno	**200**		**940000**	**1526000000**

U ćeliji **C24** računamo prateći izraz za varijancu **=H11/E11**, i dobivamo 7630000. Standardnu devijaciju možemo dobiti kao drugi korijen iz varijance pa u ćeliji **C25** na taj način i računamo **=SQRT(C24)**.

Standardna devijacija iznosi 2762,25 što znači da je prosječno odstupanje od prosječne plaće 2762,25 kn.

Koeficijent varijacije računamo kao omjer standardne devijacije i aritmetičke sredine, izraženo u postocima. U ćeliji **C26** stoga računamo koeficijent varijacije **=C25/C23**.

Vrijednost koeficijenta varijacije od 58,77% ukazuje na relativno jaku varijabilnost plaći.

Konačna tablica za razdiobu frekvencija:

RAZDIOBA FREKVENCIJA	
4700	**aritmetička sredina**
7630000	**varijanca**
2762,25	**standardna devijacija**
58,77%	**koeficijent varijacije**

Lako je primjetiti da se rezultati grupiranih i izvornih podataka razlikuju. Razlog tome je što se prilikom grupiranja podataka izgubi na preciznosti, ali dobije na preglednosti.

Za vježbu izračunajte aritmetičku sredinu i nađite mjere raspršenosti za grupirani statistički niz u **Vježbi 3.3.** zadan u mapi **Primjeri vježbi 1,2,3,4,5**.

6. Korelacijska i regresijska analiza

6.1. Korelacijska analiza

Korelacijska analiza je dio statistike koji ima za cilj istražiti moguću vezu i utvrditi stupanj povezanosti dviju ili više varijabli. Ako ona pokaže da je veza dovoljno jaka tada se statistička obrada nastavlja pronalaženjem analitičkog zapisa ili algebarskog modela koji će vezu opisati na najbolji mogući način. To je osnovni zadatak ***regresijske analize***.

Vježba 6.1.

Za podatke prikazane u tablici izradite odgovarajući grafikon (dijagram rasipanja) te izračunajte koeficijent korelacije. Koliko je dobiveni koeficijent statističi značajan? Objasnite dobivene vrijednosti.

	A	B
1	Broj kupaca	Tjedna prodaja (u 000 kn)
2	907	112
3	926	110,5
4	506	68,4
5	741	92,1
6	789	94,2
7	889	100,8
8	874	94,5
9	510	67,3
10	529	72,4
11	420	61,2
12	679	76,3
13	872	94,3
14	924	94,6
15	607	76,4
16	452	69,2
17	729	89,5
18	794	93,3
19	844	102,3
20	1010	117,7
21	621	74,1

U shopping centru 'North Gate' praćena je tjedna posjećenost i tjedna prodaja (u 000 kn). Nacrtajte dijagram rasipanja, te izračunajte i protumačite odgovarajuće pokazatelje koji nas upućuju na snagu i smjer ovisnosti promatranih varijabli?

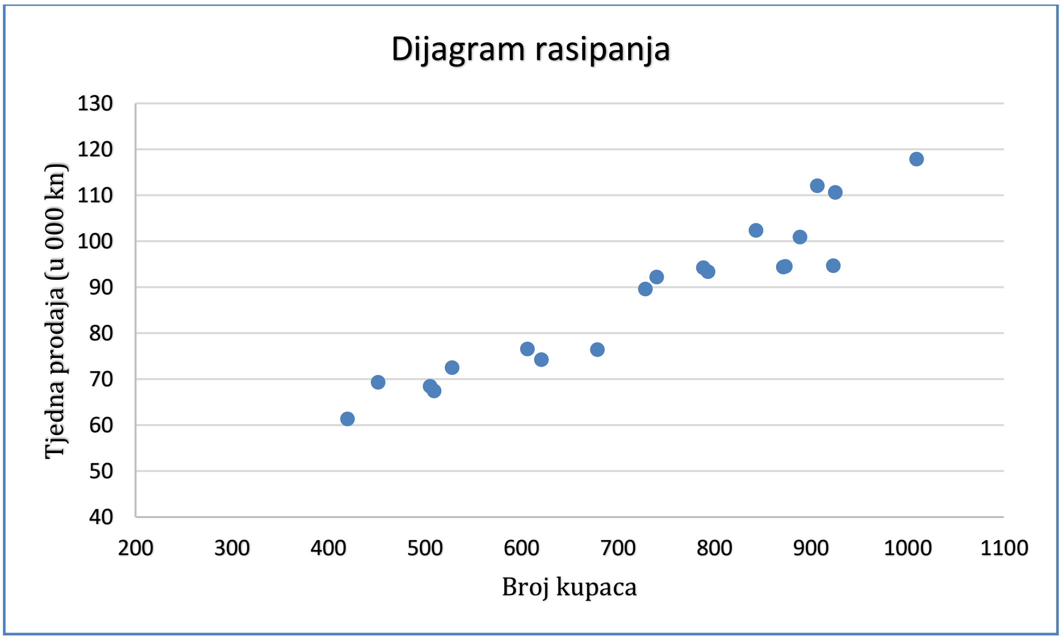
Rješenje:

Mjerenje stupnja jakosti statističkih veza provodi se metodama korelacijske analize kao što su grafički prikaz korelacije varijabli (**dijagram rasipanja**) i izračun odgovarajućeg brojčanog pokazatelja jakosti i smjera veze (**Pearsonov koeficijent korelacije**).

Prvo ćemo utvrditi što je u zadatku nezavisna, a što zavisna varijabla. Nije teško doći do zaključka da *Tjedna prodaja (u 000 kn)* ovisi o *Broju kupaca* pa ćemo u ćelije A1 i B1 kao podsjetnik umetnuti uobičajene oznake za nezavisnu varijablu (X) i zavisnu (Y) varijablu.

Broj kupaca X	Tjedna prodaja (u 000 kn) Y

Potom, da bi izradili dijagram rasipanja označavamo podatke bez zaglavnih ćelija, dakle **A2:B21** i na kartici *Insert* biramo točkasti ravninski grafikon *Scatter* (1. podtip). Dobili smo 20 točaka (uređenih parova) smještenih u ravnini. Uredit ćemo grafikon birajući *Quick Layout* br.1, dodati naslov **Dijagram rasipanja**, a oznake na osima prepisat ćemo iz zaglavnih ćelija.



Način na koji su točke (uređeni parovi) smještene već nam sugerira smjer i tip veze. Ako kao u našem primjeru porastom varijable X raste i varijabla Y riječ je o pozitivnoj vezi. Ako bi se porastom varijable X varijabla Y smanjivala bila bi riječ o negativnoj vezi

Smjer i jakost veze bolje će nam pokazati Pearsonov **koeficijent linearne korelacije r** koji upućuje na intenzitet moguće linearne veze među varijablama.

Računamo ga funkcijom:

=CORREL(Raspon zavisne varijable Y;Raspon nezavisne varijable X)

tj. u našem primjeru

=CORREL(B2:B21;A2:A21)

U interpretaciji nam pomaže tablica:

Apsolutna vrijednost koeficijenta korelacije |r|	Interpretacija
0	odsutnost korelacije
0,00 - 0,50	slaba korelacija
0,50 - 0,80	srednje jaka korelacija
0,80 - 1	čvrsta korelacija
1	potpuna korelacija

Dobiveni iznos koeficijenta linearne korelacije iznosi **0,9549** pa možemo reći da postoji jaka(čvrsta) i pozitivna veza između broja kupaca i tjedne prodaje (u 000 kn).

Koeficijent linearne korelacije može poprimiti vrijednosti iz segmenta $[-1,1]$. Predznak nas upućuje na smjer ovisnosti, a apsolutna vrijednost na intenzitet veze.

Nedostatak prethodne analize i izračuna koeficijenta korelacije jest potpuno ignoriranje veličine uzorka za koji se računa. Posljedica toga je činjenica da, za tako određen koeficijent korelacije, nismo u stanju procijeniti koliko je on **statistički značajan**. Naime, ta činjenica bitno ovisi i o veličini uzorka. Detaljnije o ovoj problematici u materijalima za predavanje.

U praksi se značajnost koeficijenta korelacije procjenjuje na razini značajnosti od 1% i (ili) 5% . U MS Excel-u se ta procjena provodi izborom *Data/Data Analysis/Regression*. U gornjem dijelu dobivenog prozora, u okvir *Input Y Range* unosimo raspon podataka zavisne varijable **(B1:B21),** a u *Input X Range*, analogno, adrese sa podacima nezavisne varijable **(A1:A21).** Zbog obuhvata zaglavnih ćelija, označit ćemo prozorčić uz *Labels.* U *Output options* navedemo odluku o mjestu na kojem želimo izlaznu tablicu (**E5**).

Pearsonov koeficijent korelacije očitat ćemo iz dijela tablice nazvanog *Multiple R*. Međutim, vrijednosti iz te 'rubrike' su uvijek pozitivne pa ćemo predznak koeficijenta odrediti pomoću predznaka koeficijenta smjera regresijskog pravca. Značajnost koeficijenta korelacije može se očitati na dva mjesta u ovoj tablici: *Significante F i P-value* za X varijablu. U oba slučaja iznosi **0,000000000062** što je daleko manje od 0,05 (5%) te zaključujemo da između naših varijabli postoji statistički značajna veza. Provjera značajnosti koeficijenta korelacije je neizostavan dio ozbiljne korelacijske analize. Veći uzorak uz manju vrijednost koeficijenta korelacije može rezultirati jednakom značajnošću kao i osjetno veće vrijednosti koeficijenta korelacije izračunatog na znatno manjem uzorku.

SUMMARY OUTPUT

Regression Statistics		
Multiple R	0,9549132	←koeficijent korelacije
R Square	0,91185922	
Adjusted R Squai	0,90696251	
Standard Error	5,014952145	
Observations	20	

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4683,35	4683,35	186,21875	6,2062E-11
Residual	18	452,695	25,1497		
Total	19	5136,05			

	Coefficients	*ındard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ɔwer 95,0%*	*Ipper 95,0%*
Intercept	24,23044396	4,80965	5,03789	8,554E-05	14,1257525	34,33514	14,1258	34,3351
Broj kupaca X	0,087293382	0,0064	13,6462	6,206E-11	0,07385399	0,100733	0,07385	0,10073

Vježba 6.2.

Poznati su rezultati na ispitima iz Poslovne matematike i Poslovne statistike za 16 studenata preddiplomskog stručnog studija Trgovinskog poslovanja.

	A	B
1	Poslovna matematika	Poslovna statistika
2	45	55
3	80	60
4	100	90
5	15	30
6	30	40
7	67	55
8	75	67
9	85	100
10	48	55
11	77	70
12	82	90
13	36	40
14	10	20
15	90	75
16	40	50
17	95	80

Potrebno je korelacijskom analizom ispitati vezu između rezultata na ispitu iz Poslovne matematike i rezultata na ispitu iz Poslovne statistike.

Rješenje:

Obzirom da rezultate na ispitu možemo shvatiti kao redoslijedno obilježje – obilježje ranga, koristit ćemo **Spearmanov koeficijent korelacije ranga** koji se koristi kada je barem jedna od varijabli redoslijedna.

Računa se $r_S = 1 - \frac{6 \cdot \sum_{i=1}^{n} {d_i}^2}{N^3 - N}$ gdje je $d_i = R(X_i) - R(Y_i)$

razlika rangova odgovarajućeg para vrijednosti varijabli X i Y; N je broj tih parova.

Da bi izračunali sumu kvadrata razlika rangova trebamo najprije rangirati vrijednosti obaju varijabli zasebno. Rangirati broj znači odrediti njegov položaj (po veličini) u odnosu na ostale vrijednosti. Funkcija kojom rangiramo u MS Excelu glasi:

=RANK(Traženi broj;Raspon podataka;Način)

Način podrazumijeva rangiranje ili u padajućem smislu (oznaka 0) ili u uzlaznom(rastućem) nizu (oznaka 1).

Proširit ćemo zadanu tablicu dodatnim stupcima.

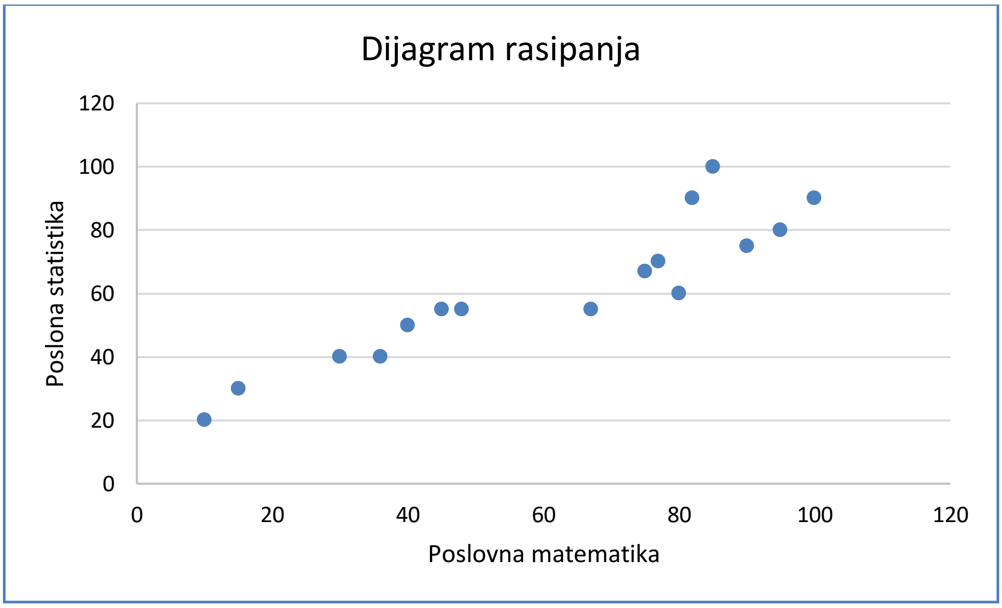
Poslovna matematika	Poslovna statistika	rang X	rang Y	di^2

U ćeliji C2 pišemo **=RANK(A2;$A$2:$A$17;1)**, kopiramo do dna tablice. Prvi podatak je 6-ti po redu tj. 45 bodova na ispitu iz matematike je šesti po veličini u uzlazno sortiranom nizu**.** Na sličan način u ćeliji D2 pišemo **=RANK(B2;$B$2:$B$17;1)**, razvučemo do dna tablice. U ćeliji E2 pišemo **=(C2-D2)^2,** razvučemo do dna tablice i sumiramo. Sada možemo izračunati Spearmanov koeficijent:

=1-6*E12/(16^3-16)

Rezultat iznosi **0,9456** što ukazuje na čvrstu pozitivnu između ostvarenih rezultata na ispitu iz Poslovne matematike i na ispitu iz Poslovne statistike.

	A	B	C	D	E
1	Poslovna matematika	Poslovna statistika	rang X	rang Y	di^2
2	45	55	6	6	0
3	80	60	11	9	4
4	100	90	16	14	4
5	15	30	2	2	0
6	30	40	3	3	0
7	67	55	8	6	4
8	75	67	9	10	1
9	85	100	13	16	9
10	48	55	7	6	1
11	77	70	10	11	1
12	82	90	12	14	4
13	36	40	4	3	1
14	10	20	1	1	0
15	90	75	14	12	4
16	40	50	5	5	0
17	95	80	15	13	4
18					**37**
19					
20		0,9456	=1-6*E18/(16^3-16)		



<u>Za vježbu riješite:</u>

U mapi **Primjeri vježbi 1,2,3,4,5** možete riješiti vježbu **4.1** i vježbu **4.2**.

6.2. Regresijska analiza

Vježba 6.3.

Rješenje:

Na primjeru iz vježbe 6.1. provest ćemo sada regresijsku analizu. Osnovni zadatak regresijske analize jest izgradnja statističkog, analitičkog modela koji će opisati vezu između nezavisne(ih) i zavisne varijable, a uz to i poslužiti u svrhu predviđanja. Promotimo prvo:

Model jednostavne linearne regresije je model koji predstavlja jednadžba pravca

$$\hat{Y} = B \cdot X + A$$

A – parametar regresije (odsječak pravca na y osi)

B – koeficijent regresije (koeficijent smjera pravca)

$\hat{Y}$ – očekivana(regresijska vrijednost) zavisne varijable Y

U MS Excelu možemo ih dobiti pomoću funkcija:

=INTERCEPT(Raspon Y;Raspon X) → =INTERCEPT(B2:B21;A2:A21)

=SLOPE(Raspon Y;Raspon X) → = SLOPE(B2:B21;A2:A21)

Dobijemo A=24,23 i B=0,0873 pa jednadžba regresijskog pravca izgleda

$$\hat{Y} = 0{,}0873 \cdot X + 24{,}23$$

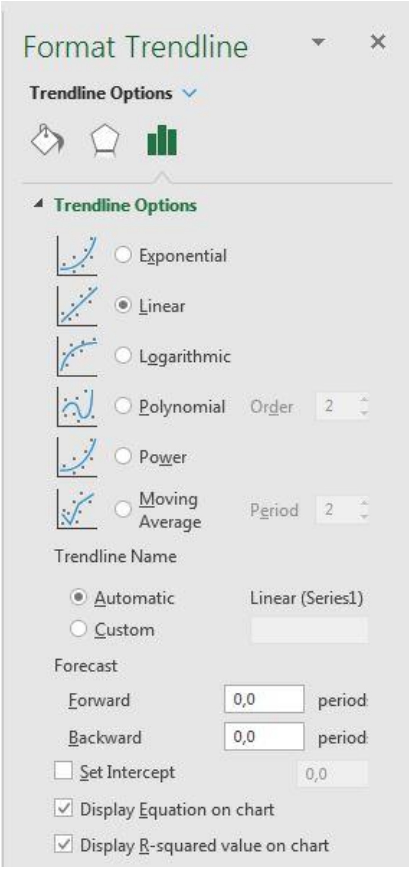
Dobivene vrijednosti interpretiramo:

A – predstavlja očekivanu vrijednost zavisne varijable kada nezavisna iznosi 0. Konkretno, kada ne bi bilo kupaca (X=0) očekivana tjedna prodaja iznosila bi 24,23 (000) kn, tj. 24 230 kuna. Dobiveni rezultat nema uporišta u realnosti, ali ponekad su takve interpretacije.

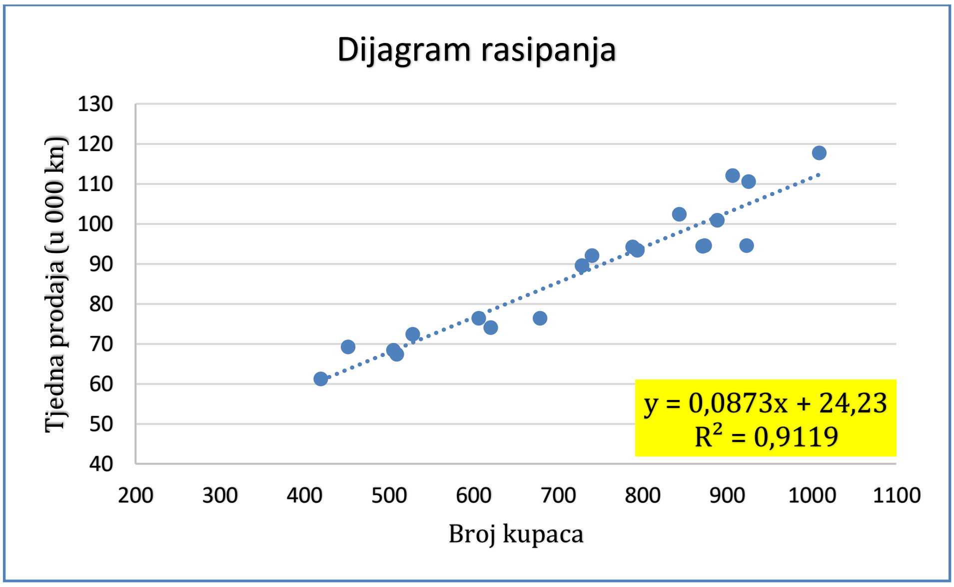
B – predstavlja očekivanu promjenu zavisne varijable Y kada se nezavisna poveća za jedinicu mjere. Konkretno, ako se broj kupaca poveća za 1 očekuje se povećanje tjedne prodaje za 0,0873 (000) kn, tj. 87,3 kune.

Postoji brži način kako ucrtati regresijski model i dobiti njegovu jednadžbu.

Klikom na bijelu površinu dijagrama rasipanja pojavi se kartica *Design* u kojoj biramo *Add Chart Element/Trendline/More trendline options.* Tada dobivamo dijaloški okvir:



u kojem biramo odgovarajući model kao i mogućnost prikazivanja jednadžbe modela na dijagramu i koeficijenta determinacije R^2 (zadnja dva prozorčića). Nakon tih radnji na dijagramu rasipanja dobijemo model:



Već smo spomenuli da nam model služi za prognostičke svrhe pa ćemo postaviti dva pitanja:

b) Uz pomoć dobivenog modela prognozirajte kolika bi bila očekivana tjedna prodaja uz 980 kupaca?

c) Dobivenim modelom prognozirajte koliko bi bilo potrebno kupaca kako bi očekivana tjedna prodaja iznosila 80(000) kn?

Rješenje:

b) Poznato je da ima X=980 pušača. Očekivani broj kupaca dobit ćemo tako da tu vrijednost uvrstimo u model jednostavne linearne regresije:

$$= 0{,}0873 * 980 + 24{,}23$$

Dobiveni broj 109,784 pomnožimo s 1000 jer je riječ o tjednoj prodaji (u 000 kn) i pišemo interpretaciju:

Za 980 kupaca po linearnom modelu može se očekivati tjedna prodaja od 109874 kuna.

c) Poznato je da je tjedna prodaja Y=80(000) kuna. Da bi našli uz koji je to broj kupaca, moramo rješiti jednostavnu linearnu jednadžbu s jednom nepoznanicom. Uz poznati Y dobivamo da je $X = \frac{(Y-A)}{B}$ pa računamo

$$= (80 - 24{,}23)/0{,}0873$$

Dobiveni broj 638,832 zaokružimo na cijeli broj, 639 jer je riječ o broju kupaca.

Dakle uz 639 kupaca može se očekivati tjedna prodaja od 80 000 kuna.

Nije svaki model jednako dobar i pouzdan za prognoziranje zato je potrebno ispitati reprezentativnost regresijskog modela.

Reprezentativnost modela jednostavne linearne regresije

Kako bi procijenili koliko je linearni model uspješan računamo mjere reprezentativnosti, a to su varijanca regresije, standardna devijacija, koeficijent varijacije te koeficijent determinacije R^2.

VARIJANCA REGRESIJE apsolutna je mjera reprezentativnosti modela i predstavlja prosječno kvadratno odstupanje empirijskih od regresijskih vrijednosti. Računa se: $\hat{\sigma}_{\hat{y}}^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N-2}$

Zbog toga ćemo tablicu proširiti sa još dva stupca na sljedeći način:

	A	B	C	D
1	**Broj kupaca X**	**Tjedna prodaja (u 000 kn) Y**	**Yoček**	**(Y-Yoček)^2**

U stupcu C računat ćemo očekivane vrijednosti broja oboljelih (po linearnom modelu) tako da u C2 pišemo **=0,0873*A2+24,23**, a potom razvučemo formulu do dna stupca. Zatim u stupcu D računamo kvadratna odstupanja očekivanih vrijednosti zavisne varijable(Y*oček*) od empirijskih Y. Stoga u D2 pišemo **=(B2-C2)^2** kopiramo do dna stupca i prosumiramo. Time smo dobili brojnik za izračun varijance.

Sada pored tablice u ćeliji po izboru računamo varijancu kao omjer sume kvadratnih odstupanja i N-2, pri čemu je N broj zadanih parova podataka (u našem primjeru N=20). Pišemo **=D22/(20-2)** i dobijemo vrijednost varijance **25,1498**.

STANDARDNA DEVIJACIJA ili standardna pogreška regresije apsolutna je mjera reprezentativnosti modela i predstavlja prosječno odstupanje empirijskih od regresijskih vrijednosti, a računamo je kao i prije kao drugi korijen iz varijance pa pišemo **=SQRT(25,1498)** što iznosi **5,015**.

KOEFICIJENT VARIJACIJE je relativna mjera reprezentativnosti modela i predstavlja postotni udio standardne devijacije regresije u odnosu na aritmetičku sredinu zavisne varijable Y. Formula glasi: $\hat{V}_{\hat{y}} = \frac{\hat{\sigma}_{\hat{y}}}{\bar{Y}}$ (%)

Stoga računamo **=5,015/AVERAGE(B2:B21)** i dobivenu vrijednost izrazimo u postocima. On iznosi **5,70 %**.

KOEFICIJENT DETERMINACIJE R^2 je omjer sume kvadrata odstupanja protumačenih regresijom i sume kvadrata ukupnih odstupanja. Formula je:

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^{N}(y_i - \bar{Y})^2}$$

No, iako u MS Excelu postoji funkcija koja ga računa **=RSQ(Raspon Y;RasponX)** nije ga nužno posebno računati jer ga dobijemo prilikom umetanja regresijskog modela opcijom *Display R-squared value on chart* .

Na našem dijagramu rasipanja, uz jednadžbu modela linearne regresije očitavamo $\boldsymbol{R^2 = 0,9119}$. Koeficijent determinacije može se objasniti kao proporcija varijance zavisne varijable objašnjena nezavisnom varijablom. U našem primjeru to bi značilo da je 91,19 % tjedne prodaje objašnjeno brojem kupaca.

Može se reći što je koeficijent determinacije veći (bliži 1) to je model reprezentativniji.

Kada u sljedećim primjerima budemo ispitivali više modela na istom skupu podataka on će nam poslužiti kao glavni kriterij uspješnosti između više modela.

Radni list riješenog zadatka izgleda ovako:

	A	B	C	D	E	F	G	H	I	J	K
1	**Broj kupaca X**	**Tjedna prodaja (u 000 kn) Y**	**Yoček**	**(Y-Yoček)^2**							
2	907	112	103,411	73,7692		**Yoček**	PREDSTAVLJA OČEKIVANE VRIJEDNOSTI				
3	926	110,5	105,070	29,4871			ZAVISNE VARIJABLE PREMA REGRESIJSKOM				
4	506	68,4	68,404	0,0000			MODELU				
5	741	92,1	88,919	10,1169		**Y**	EMPIRIJSKE VRIJEDNOSTI ZAVISNE VARIJABLE				
6	789	94,2	93,110	1,1888							
7	889	100,8	101,840	1,0810		VARIJANCA REGRESIJE					
8	874	94,5	100,530	36,3633							
9	510	67,3	68,753	2,1112		25,1498					
10	529	72,4	70,412	3,9533							
11	420	61,2	60,896	0,0924		STANDARDNA DEVIJACIJA					
12	679	76,3	83,507	51,9365							
13	872	94,3	100,356	36,6703		5,015					
14	924	94,6	104,895	105,9911							
15	607	76,4	77,221	0,6742		KOEFICIJENT VARIJACIJE					
16	452	69,2	63,690	30,3645							
17	729	89,5	87,872	2,6514		5,70%					
18	794	93,3	93,546	0,0606							
19	844	102,3	97,911	19,2616		KOEFICIJENT DETERMINACIJE					
20	1010	117,7	112,403	28,0582							
21	621	74,1	78,443	18,8643		0,9119					
22				**452,6958**							

Za vježbu riješite:

U mapi **Primjeri vježbi 1,2,3,4,5** vježbu **4.3.**

Vježba 6.4.

Godina	Broj cijepljenih (000)	Broj oboljelih (000)
1998	3,51	10,43
1999	4,83	9,01
2000	5,24	8,30
2001	5,63	8,01
2002	6,01	7,11
2003	6,38	6,56
2004	7,15	6,00
2005	9,20	5,60
2006	11,32	4,72
2007	16,18	3,91

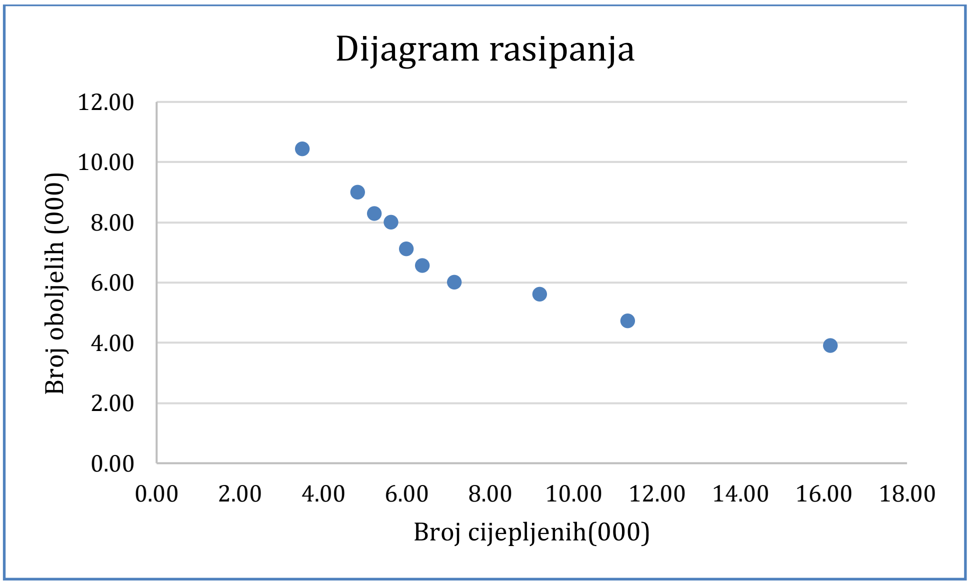
Podaci o broju cijepljenih i broju oboljelih od gripe u gradu X tijekom proteklih deset godina,dani su slijedećom tablicom:

a) Podatke iz tablice predočite dijagramom rasipanja te navedite sve potrebne oznake.
b) Izračunajte Pearsonov koeficijent linearne korelacije za zadane varijable te interpretirajte dobivenu vrijednost.
c) Nađite jednadžbu linearnog regresijskog modela za zadane vrijednosti te obrazložite značenje dobivenih parametara
d) Odredite jednadžbu eksponencijalnog regresijskog modela i interpretirajte dobivene vrijednosti parametara.
e) Odredite jednadžbu dvostruko logaritamskog regresijskog modela i interpretirajte dobivene vrijednosti parametara.
f) Procijenite koji od ova tri modela najbolje opisuje vezu između zadanih varijabli.
g) Uz pomoć sva tri modela procijenite koliko se oboljelih može očekivati u slučaju da se cijepi 18(000) osoba.

Rješenje:

Nije teško uočiti da između ove dvije varijable postoji korelacija pri čemu *Broj oboljelih* (zavisna) ovisi o *Broju cijepljenih* (nezavisna) od gripe.

a) Za dijagram rasipanja označimo podatke, od **B2:C11** i u izborniku *Insert* biramo točkasti (*Scatter*) dijagram te odgovarajući predložak. Navedemo sve potrebne oznake pa dobiveni grafikon izgleda:



Očigledno je da postoji negativna veza između ove dvije varijable.

b) Pearsonov koeficijent linearne korelacije izračunamo funkcijom **=CORREL(C2:C11;B2:B11)**

Pearsonov koeficijent korelacije iznosi **-0,8947** što upućuje na postojanje jake negativne veze između ove dvije veličine.

c) Ucrtat ćemo linearni regresijski model na dijagram rasipanja zajedno s pripadnom jednadžbom i vrijednosti koeficijenta determinacije R^2 na način koji je opisan u prethodnom primjeru i očitati dobivene parametre.

Budući jednadžba modela glasi $\hat{Y} = -0{,}4771 \cdot X + 10{,}565$ iščitavamo:

B=-0,4771 dakle, ako se broj cijepljenih poveća za 1000(jedinicu mjere) očekuje se smanjenje broja oboljelih za 477 slučajeva.

A=10,565 što znači da se u slučaju kada se nitko ne bi cijepio (X=0) očekivani broj oboljelih od gripe bio bi 10565.

d) ***<u>Model jednostavne eksponencijalne regresije</u>*** jedan je od nelinearnih modela koji pretpostavlja eksponencijalnu ovisnost zavisne o nezavisnoj varijabli i ima jednadžbu

$$\hat{Y} = A \cdot B^x$$

Model ćemo dobiti kao i linearni, a u izborniku regresijskih modela nalazi se pod istoimenim nazivom. Jednadžba dobivenog modela glasi

$$\hat{Y} = 11{,}868 \cdot e^{-0{,}076x}$$

Parametar A kod eksponencijalnog modela ima značenje očekivane vrijednosti zavisne varijable Y kada je nezavisna X=0. Direktno ga očitamo i interpretiramo **A = 11,868**.

Dakle možemo reći da se u slučaju izostanka cijepljenja (X=0) može očekivati 11868 oboljelih od gripe.

Parametar B potrebno je izračunati **=exp(-0,076)** što daje iznos **0,9268**. Uz pomoć parametra B računamo stopu S, **S=(B-1)*100** pa sukladno tome računamo **=(0,9268-1)*100** (može se pomnožiti s 100 ili pretvoriti u postotak). Konačno za vrijednost stope dobivamo **S=-7,32(%)**. Stopa opisuje postotnu promjenu zavisne varijable Y kada X naraste za jedinicu mjere.

Dakle u slučaju da se broj cijepljenih poveća za 1000 očekuje se smanjenje broja oboljelih za 7,32%.

e) ***<u>Dvostruko logaritamski model</u>*** je model čija jednadžba glasi

$$\hat{Y} = A \cdot X^{B}$$

Model se u izborniku s regresijskim modelima krije pod imenom ***Power***, zajedno s pripadnom jednadžbom kao i koeficijentom determinacije ucrtamo ga na dijagram rasipanja na već opisani način. Jednadžba modela na konkretnom primjeru glasi

$$\hat{Y} = 24{,}352 \cdot X^{-0{,}671}$$

Očitamo i interpretiramo parametre:

A = 24,352; parametar A kod dvostruko logaritamskog modela predstavlja očekivanu vrijednost zavisne varijable Y kada nezavisna X poprimi jediničnu vrijednost (X=1); dakle u slučaju da se cijepi 1000 osoba očekuje se da će od gripe oboljeti 24352 osoba.

B = -0,671; parametar B se interpretira kao postotna promjena zavisne varijable kada se nezavisna poveća za 1%; dakle u slučaju da se broj cijepljenih poveća za 1% očekuje se smanjenje broja oboljelih za 0,671%.

f) Model koji najbolje opisuje vezu biramo po koeficijentu determinacije R^2.

Očigledno dvostruko logaritamski model ima najveći koeficijent determinacije (najbliži jedinici) pa je on najreprezentativniji tj. najpogodniji za prognoze.

g) Potrebno je odrediti očekivani broj oboljelih u slučaju da se cijepi 18000 osoba prema sva tri modela?

Bitno je uočiti da nam je zadana varijabla X, a obzirom da je (čitamo u tablici) jedinica mjere 1000 možemo zapisati X=18, a Y je nepoznat. Očekivani broj oboljelih dobit ćemo tako da u pojedine jdbe modela na mjesto varijable X uvrštavamo 18.

X=18			
Y=?			
Ako se cijepi 18 000 osoba može se očekivati:			
linearni	1,977	=-0,4771*18+10,565	Očekuje se 1977 oboljelih
eksponencijalni	3,022	=11,868*EXP(-0,076*18)	Očekuje se 3022 oboljelih
dvostruko logaritamski	3,501	=24,352*18^(-0,671)	Očekuje se 3501 oboljelih

Linearni model: **=-0,4771*18+10,565**

Dobivamo vrijednost **1,977** (na tri decimale jer je mjera za broj oboljelih 000). Po linearnom modelu u slučaju 18000 cijepljenih osoba može se očekivati 1977 oboljelih od gripe.

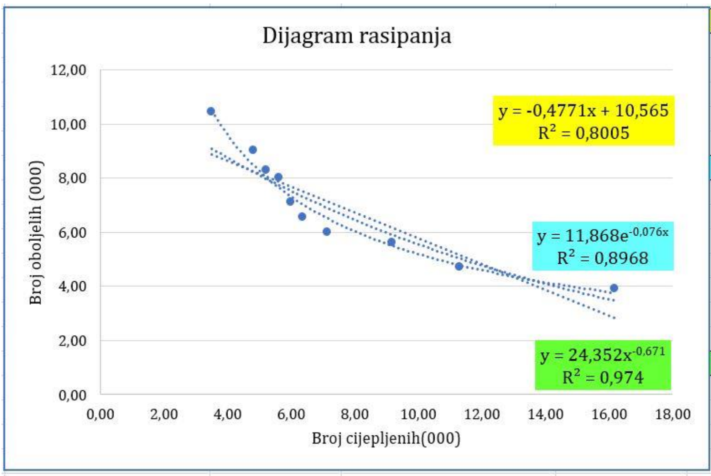
Eksponencijalni model: **=11,868*exp(-0,076*18)**

Dobivamo vrijednost **3,022.** Po eksponencijalnom modelu u slučaju 18000 cijepljenih očekuje se 3022 oboljelih od gripe.

Dvostruko logaritamski model: **=24,352*18^(-0,671)**

Dobivena vrijednost iznosi **3,501**. Po dvostruko logaritamskom modelu u slučaju 18000 cijepljenih očekuje se 3501 oboljelih od gripe.

Radni list (bez interpretacija koje su sastavni dio zadatka) izgleda:



MODEL JEDNOSTAVNE LINEARNE REGRESIJE		
A	10,565	
B	-0,4771	
MODEL JEDNOSTAVNE EKSPONENCIJALNE REGRESIJE		
A	11,868	
B	0,92681621	
S	-7,32%	
DVOSTRUKO LOGARITAMSKI REGRESIJSKI MODEL		
A	24,352	
B	-0,671	

X=18			
Y=?			
Ako se cijepi 18 000 osoba može se očekivati:			
linearni	1,977	=-0,4771*18+10,565	Očekuje se 1977 oboljelih
eksponencijalni	3,022	=11,868*EXP(-0,076*18)	Očekuje se 3022 oboljelih
dvostruko logaritamski	3,501	=24,352*18^(-0,671)	Očekuje se 3501 oboljelih

<u>Za vježbu riješite:</u>

U mapi **Primjeri vježbi 1,2,3,4,5** vježbe **4.5.** i **4.6.**

7. Analiza vremenskih nizova

Relativni pokazatelji dinamike jedne ili skupine pojava su neimenovani brojevi koje zovemo indeksima. Ako se njima iskazuje odnos stanja jedne pojave u različitim razdobljima, riječ je o individualnim indeksima. Obzirom na način praćenja jedne pojave, individualni indeksi se dijele na lančane (verižne) i bazne. Indeksi su relativni brojevi iskazani (ali ne i napisani) u postocima. Ishodišna vrijednost je 100. Ako indeks premašuje tu vrijednost tada je promatrana pojava porasla upravo za iznos tog prekoračenja iskazan u postocima. Analogno, ako je vrijednost indeksa manja od 100 tada se vrijednost promatrane pojave smanjila u navedenom razdoblju i to upravo za, u postocima iskazan, iznos smanjenja od ishodišnih 100. Svaka interpretacija ovih relativnih pokazatelja dinamike nužno mora uključiti podatke o razdoblju na koje se odnosi.

7.1. Individualni indeksi

Vježba 7.1.

Tablicu zadanu u primjeru 7.1. preuredit ćemo na dole prikazan način.

	A	B	C	D	E	F	G	H	I
1	Godina	Prodano pivo (000 l)							
2	2007	18,42							
3	2008	22,67							
4	2009	26,38							
5	2010	29,42							
6	2011	33,62							
7	2012	36,43							
8	2013	37,71							
9	2014	41,92							
10	2015	44,03							
11	2016	51,31							
12									
13									
14	Podaci o godišnjoj prodaji piva u gradu Šibeniku, u razdoblju od 2007. do 2016. godine, dani su tablicom:								
15	a)Izračunajte bazne indekse godišnje prodaje piva u gradu Šibeniku, uzimajući 2009. godinu za baznu.								
16	Obrazložite značenje dobivenog indeksa za 2012. godinu.								
17	b)Izračunajte lančane indekse te obrazložite dobivenu vrijednost tog indeksa za 2014. godinu.								

Rješenje:

a) Potrebno je izračunati *bazne indekse* uzimajući 2009. kao ishodišnu(baznu) godinu, dakle to je referentno razdoblje koje će za ovu pojavu imati indeks 100.

Bazni indeksi su relativni pokazatelji promjene vrijednosti pojave dane vremenskim nizom u odnosu na vrijednost koju pojava ima u nekom odabranom, baznom, razdoblju istog niza. Definiraju se kao omjer vrijednosti pojave u tekućem razdoblju u odnosu na vrijednost pojave u baznom razdoblju.

$$I_t = \frac{Y_t}{Y_b} \cdot 100, \qquad t = 1,2,\dots,n$$

Y_t apsolutna frekvencija promatrane pojave u tekućem razdoblju

Y_b apsolutna frekvencija promatrane pojave u baznom razdoblju

I_t bazni indeks u promatranom razdoblju

Tablicu proširimo s dva stupca:

	A	B	C	D
1	**Godina**	**Prodano pivo (000 l) Y**	**It 2009.=100**	**St**

U ćeliji **C2** računamo prvi bazni indeks, za 2007. godinu pa količinu prodanog piva za 2007. dijelimo s količinom u baznoj 2009. godini, odnosno pišemo **=B2/$B$4*100**. Formulu kopiramo za sva ostala razdoblja i dobivene indekse zaokružimo na dvije decimale. Kako bi jednostavnije interpretirali bazni indeks za 2012. godinu formiramo stupac pripadnih stopa promjena. Pripadna stopa promjene pojave u tekućem, u odnosu na bazno razdoblje računa se:

$$S_t = I_t - 100 \quad t = 1,2,\dots,n$$

U ćeliji **D2** pišemo **=C2-100** i kopiramo za sva razdoblja.

Za 2012. nalazimo **$I_{2012.}$=138,10** odnosno pripadna stopa je **38,10** što znači da je prodaja pive u 2012. godini zabilježila rast od 38,10% u odnosu na baznu 2009. godinu.

Promatrajući dobivenu tablicu popunite sljedeću tvrdnju:

U 2016. godini zabilježena je prodaja____________litara pive što predstavlja __________od_________ % u odnosu na prodaju pive u 2009. godini.

b) *Verižni (lančani) indeksi* predstavljaju, u postocima iskazanu, promjenu vrijednosti pojave tekućeg razdoblja u odnosu na prethodno vremensko razdoblje. Izračun verižnih (lančanih) indeksa prati relaciju:

$$V_t = \frac{Y_t}{Y_{t-1}} \cdot 100 \quad t = 2,3, \dots, n$$

Y_t je apsolutna frekvencije pojave u tekućem razdoblju

Y_{t-1} je apsolutna frekvencije pojave u prethodnom razdoblju

Radi lakše interpretacije može se izračunati i pripadna stopa promjena S^*_t koja se računa tako da od pripadnog verižnog indeksa oduzmemo 100.

$$S_t^* = V_t - 100 \quad t = 2,3, \dots, n$$

Proširit ćemo tablicu s još dva stupca:

	A	B	C	D	E	F
1	Godina	Prodano pivo (000 l) Y	I_t 2009.=100	S_t	V_t	S^*_t

Obzirom da ne znamo vrijednost prodaje piva u 1992. godini ne možemo izračunati prvi verižni indeks pa u ćeliju **E2** za prvo razdoblje pišemo *nepoznat* . U ćeliji **E3** pišemo **=B3/B2*100** i kopiramo formulu do dna stupca (na dvije decimale). U susjednom stupcu izračunamo pripadne stope počevši od **F3** gdje pišemo **=E3-100,** a potom kopiramo do dna.

Za **2014.** godinu očitamo vrijednost **V_t=111,16** i pripadne stope **11,16** što znači da je prodaja pive u 2014. godini porasla za 11,16% <u>u odnosu na prethodnu</u> 2013. godinu.

Promatrajući dobivenu tablicu popunite sljedeću tvrdnju:

U 2012. godini zabilježena je prodaja____________litara pive što predstavlja __________od_________ % u odnosu na prodaju pive u prethodnoj 2011. godini.

Kako se promjenila prodaje pive 2015. godine u apsolutnom iznosu u odnosu na 2009. godinu? Koliko ta promjena iznosi u postotnom iznosu?

Kako se promjenila prodaja pive 2015. godine u apsolutnom iznosu u odnosu na 2014. godinu? Koliko ta promjena iznosi u postotnom iznosu?

Radni list rješenog primjera izgleda:

	A	B	C	D	E	F	G	H
1	**Godina**	**Prodano pivo (000 l) Y**	**I_t 2009.=100**	**S_t**	**V_t**	**S^*_t**		
2	2007	18,42	69,83	-30,17	nepoznat	nepoznat		
3	2008	22,67	85,94	-14,06	123,07	23,07		
4	2009	26,38	100,00	0,00	116,37	16,37		
5	2010	29,42	111,52	11,52	111,52	11,52		
6	2011	33,62	127,45	27,45	114,28	14,28		
7	2012	36,43	**138,10**	**38,10**	108,36	8,36		
8	2013	37,71	142,95	42,95	103,51	3,51		
9	2014	41,92	158,91	58,91	**111,16**	**11,16**		
10	2015	44,03	166,91	66,91	105,03	5,03		
11	2016	51,31	194,50	94,50	116,53	16,53		
12								
13								
14	Podaci o godišnjoj prodaji piva u gradu Šibeniku, u razdoblju od 2007. do 2016. godine, dani su tablicom:							
15	a)Izračunajte bazne indekse godišnje prodaje piva u gradu Šibeniku, uzimajući 2009. godinu za baznu.							
16	Obrazložite značenje dobivenog indeksa za 2012. godinu.							
17	b)Izračunajte lančane indekse te obrazložite dobivenu vrijednost tog indeksa za 2014. godinu.							
18								

Vježba 7.2.

Tablicu zadanu u primjeru 7.2. preuredit ćemo na dole prikazan način.

	A	B	C	D	E	F
1	**Godina**	**I_t 2011.=100**				
2	2005	83,27				
3	2006	87,92				
4	2007	89,29				
5	2008	92,45				
6	2009	97,52				
7	2010	93,87				
8	2011	100				
9	2012	102,33				
10	2013	108,74				
11	2014	110,37				
12	2015	110,29				
13	2016	113,37				
14	2017	118,29				
15	2018	121,42				
16						
17						
18	Kretanje cijena uslužnih djelatnosti u gradu Senju, u razdoblju od 2005. do 2018. godine, dano je tablicom:					
19	a) Izračunajte brojčane pokazatelje dinamike promjene cijena u uzastopnim godinama promatranog razdoblja.					
20	b) Nađite brojčane pokazatelje promjena cijena uslužnih djelatnosti u odnosu na 2008. godinu.					

Rješenje:

a) Potrebno je izračunati verižne indekse iz zadanih baznih indeksa (2011.=100). Bazne indekse preračunavamo u verižne tako da bazne indekse koristimo kao originalne apsolutne frekvencije. Lako se može pokazati da je omjer apsolutnih frekvencija jednak omjeru baznih indeksa pa koristimo formulu:

$$V_t = \frac{I_t}{I_{t-1}} \cdot 100 \quad t = 2,3, \dots , n$$

Proširimo tablicu stupcem verižnih indeksa. U ćeliji **C2** za prvi indeks pišemo da je nepoznat jer za prethodno razdoblje nemamo informaciju. U **C3** pišemo **=B3/B2*100** i isto pravilo kopiramo do dna tablice.

b) Kako pretvoriti bazne indekse u indekse druge baze? Koristeći izraz:

$$I_t^{NB} = \frac{I_t^{SB}}{I_b^{SB}} \cdot 100$$

što znači da se transformacija baza provodi na način da bazni indeksi (po staroj bazi) preuzimaju ulogu originalnih frekvencija. Proširimo tablicu s još jednim stupcem **I_t (2008.=100).** Računamo prvi indeks u ćeliji **D2** formulom **=B2/$B$5*100** i kopiramo je do dna stupca. Konačna tablica izgleda:

	A	B	C	D
1	**Godina**	**It 2011.=100**	**V_t**	**It 2008.=100**
2	2005	83,27	nepoznat	90,07
3	2006	87,92	105,58	95,10
4	2007	89,29	101,56	96,58
5	2008	92,45	103,54	100,00
6	2009	97,52	105,48	105,48
7	2010	93,87	96,26	101,54
8	2011	100	106,53	108,17
9	2012	102,33	102,33	110,69
10	2013	108,74	106,26	117,62
11	2014	110,37	101,50	119,38
12	2015	110,29	99,93	119,30
13	2016	113,37	102,79	122,63
14	2017	118,29	104,34	127,95
15	2018	121,42	102,65	131,34

Za vježbu interpretirajte verižni i bazni indeks za 2007. godinu (iz indeksa na bazi 2008. godine).

Vježba 7.3.

Godina	V_t
2004	----
2005	103,42
2006	110,74
2007	120,17
2008	110,19
2009	128,43
2010	110,71
2011	107,48
2012	112,57
2013	120,61
2014	117,14
2015	119,53
2016	115,53
2017	100,98

Verižni indeksi zapošljavanja u turističkoj djelatnosti, u vremenu od 2004. do 2017. godine u županiji X, su dani tablicom:

a) Izračunajte brojčane pokazatelje dinamike zapošljavanja u turističkoj djelatnosti, u županiji X, u odnosu na 2008. godinu.

b) Znajući da je broj zaposlenih u turističkoj djelatnosti u županiji X, u 2006. godini iznosio 130428 odredite broj zaposlenih, u turizmu u županiji X, tijekom preostalih godina promatranog razdoblja.

Rješenje:

a) Potrebno je iz poznatih verižnih indeksa 'rekonstruirati' bazne indekse. Bazne indekse računamo iz poznatih verižnih dvojako ovisno o tome da li se oni odnose na razdoblja koja prethode baznom razdoblju ili ga slijede. Postupak prati dolje navedene formule, redom od 1 – 3:

$$I_{t-1} = \frac{I_t}{V_t} \cdot 100 \quad , \; t < b \quad (2)$$

$$I_t = 100 \quad , \; t = b \quad (1)$$

$$I_t = \frac{I_{t-1} \cdot V_t}{100} \quad , \; t > b \quad (3)$$

Proširimo tablicu stupcem baznih indeksa **I_t (2008.=100)**. Postupak započinjemo tako da za baznu godinu (2008.) upišemo vrijednost indeksa 100. Za prvo prethodno razdoblje, u ćeliju **C5** primjenimo formulu (2) na prvo prethodno razdoblje $I_{2007} = \frac{I_{2008}}{V_{2008}} \cdot 100$. Stoga pišemo **=C6/B6*100**, i to pravilo primjenimo za sva prethodna razdoblja (kopiramo pravilo prema gore). U prvo razdoblje koje slijedi nakon baznog primjenjujemo pravilo (3) $I_{2009} = \frac{I_{2008} \cdot V_{2009}}{100}$ na način da u **C7** pišemo **=C6*B7/100**. To pravilo kopiramo prema dolje na sva razdoblja koja slijede.

b) Ako znamo vrijednost pojave u jednom od razdoblja moguće je rekonstruirati čitav vremenski niz pomoću verižnih ili baznih indeksa. U slučaju poznatih verižnih indeksa, taj postupak formalno nalikuje preračunavanju baznih indeksa u verižne. I ovdje postoje dva tipa generirajućih formula ovisno o tome da li se odnose na razdoblja koja prethode razdoblju za koje znamo vrijednost pojave opisane nizom, ili ga slijede. Ako je poznata vrijednost $\mathbf{Y_t}$ za dano razdoblje tada za frekvencije koje joj prethode vrijedi:

$$Y_{t-1} = \frac{Y_t}{V_t} \cdot 100$$

dok se članovi niza koji slijede zadanu vrijednost računaju na način:

$$Y_t = \frac{Y_{t-1} \cdot V_t}{100}$$

Obzirom na analogni postupak započinjemo na sličan način. Proširimo niz stupcem $\mathbf{Y_t}$**,** koji predstavlja stvarni broj zaposlenih u turizmu za dani niz godina. Prvo unesemo poznati podatak, za 2006. Znamo da je bilo zaposlenih 130428 osoba. Potom u prethodnu ćeliju **D3** primijenimo pravilo za frekvencije koje prethode poznatom razdoblju **=D4/B4*100** i razvučemo prema gore. U prvoj ćeliji koja slijedi poznatu vrijednost, **D5** pišemo **=D4*B5/100** i kopiramo za sva iduća razdoblja. Broj zaposlenih ćemo zaokružiti na cijeli broj jer je riječ o osobama. Konačna tablica izgleda:

	A	B	C	D
1	**Godina**	$\mathbf{V_t}$	$\mathbf{I_t}$ **2008.=100**	$\mathbf{Y_t}$ **stvarni broj zaposlenih**
2	2004	----	65,94	113884
3	2005	103,42	68,20	117779
4	2006	110,74	75,52	130428
5	2007	120,17	90,75	156735
6	2008	110,19	100	172707
7	2009	128,43	128,43	221807
8	2010	110,71	142,18	245563
9	2011	107,48	152,82	263931
10	2012	112,57	172,03	297107
11	2013	120,61	207,49	358341
12	2014	117,14	243,05	419760
13	2015	119,53	290,52	501739
14	2016	115,53	335,63	579660
15	2017	100,98	338,92	585340

Za vježbu riješite: U mapi **Primjeri vježbi 1,2,3,4,5** vježbu **5.4.**

7.2. Skupni indeksi

Skupni indeksi su brojčani pokazatelji relativne promjene heterogene skupine pojava. Najčešće se relativne promjene prate u dva razdoblja: baznom (nulto razdoblje) i razdoblju u kojem iskazujemo dinamiku promjena (tekuće ili izvještajno razdoblje). Cijene označavamo s p_{ij}, a količine s q_{ij}. Prvi od ova dva indeksa označava mjesto u nizu (i = 1,2,…,k) a drugi razdoblje na koje se odnosi. Za bazno razdoblje je j = 0, a j = 1 je oznaka tekućeg razdoblja. Skupni indeksi se javljaju pod različitim nazivima i u različitim oblicima. Najčešće su u upotrebi skupni indeksi cijena, količina i vrijednosti. Najčešći oblici skupnih indeksa su Laspeyresov i Paascheov oblik. Oznake koje koristimo u relacijama su:

p_0 → cijena u baznom razdoblju

p_1 → cijena u tekućem razdoblju

q_0 → količina u baznom razdoblju

q_1 → količina u tekućem razdoblju

7.2.1. Skupni indeksi cijena

Skupni indeks cijena mjeri prosječnu promjenu cijena skupine pojava u dva različita vremenska razdoblja i to na bazi nepromijenjenih količina. Ovisno iz kojeg razdoblja držimo nepromijenjene količine razlikujemo dva tipa:

Laspeyresov skupni indeks cijena

$$P_{01}(q_0) = \frac{\sum_{i=1}^{k} p_{i1} \cdot q_{i0}}{\sum_{i=1}^{k} p_{i0} \cdot q_{i0}} \cdot 100$$

označava ukupnu promjenu cijena skupine pojava računatu uz nepromijenjene količine baznog razdoblja.

Paascheov skupni indeks cijena

$$P_{01}(q_1) = \frac{\sum_{i=1}^{k} p_{i1} \cdot q_{i1}}{\sum_{i=1}^{k} p_{i0} \cdot q_{i1}} \cdot 100$$

predstavlja relativni brojčani iskaz ukupne promjene cijena tekućeg razdoblja u odnosu na bazno razdoblje uz nepromijenjene količine tekućeg razdoblja.

Vježba 7.4.

	A	B	C	D	E
1	**vrsta povrća**	cijene - 2017.	količine-2017.	cijene - 2018.	količine - 2018.
2	blitva	15	140	18	120
3	kupus	11	100	13	76
4	cvjetača	12	80	15	90
5	brokula	18	220	16	170
6	špinat	16	125	20	110

Na podacima iz Vježbe 7.4. izračunat ćemo skupne indekse cijena, skupne indekse količina i skupni indeks vrijednosti.

Rješenje:

Skupina pojava koju ovdje promatramo su razne vrste povrća. Dva su promatrana razdoblja, 2017. godina i 2018. godina. Starije razdoblje je bazno i označit ćemo ga indeksima 0, a 2018. je izvještajno (tekuće) razdoblje i dobit će indeks 1. Nakon što u zaglavne ćelije ubacimo oznake tablica izgleda:

	A	B	C	D	E
1	**vrsta povrća**	cijene - 2017. p0	količine-2017. q0	cijene - 2018. p1	količine - 2018. q1
2	blitva	15	140	18	120
3	kupus	11	100	13	76
4	cvjetača	12	80	15	90
5	brokula	18	220	16	170
6	špinat	16	125	20	110

Laspeyresov indeks cijena:

$$P_{01}(q_0) = \frac{\sum_{i=1}^{k} p_{i1} \cdot q_{i0}}{\sum_{i=1}^{k} p_{i0} \cdot q_{i0}} \cdot 100$$

Kako bi izračunali Laspeyresov indeks cijena potrebno je formirati umnoške **p_1*q_0** i **p_0*q_0** pa tablicu proširimo s takva dva stupca. Prateći relativne adrese za izračun prvog umnoška pišemo u ćeliji **F2, =C2*D2** kopiramo do dna i sumiramo stupac. U ćeliji **G2** pišemo **=B2*C2** razvučemo do dna i sumiramo. Zatim računamo Laspeyresov indeks cijena tako da podijelimo ove dvije sume **=F7/G7*100** i dobiveni broj zaokružimo na dvije decimale.

Indeks iznosi 109,09 što znači da je došlo do povećanja od 9,09%(109,09-100) i pišemo interpretaciju:

Cijene svih vrsta povrća u 2018. su se povećale za 9,09% u odnosu na 2017. godinu računajući uz neizmjenjene količine iz bazne 2017. godine.

Paascheov skupni indeks cijena

$$P_{01}(q_1) = \frac{\sum_{i=1}^{k} p_{i1} \cdot q_{i1}}{\sum_{i=1}^{k} p_{i0} \cdot q_{i1}} \cdot 100$$

Na analogni način izračunat ćemo i Paascheov indeks cijena. Formiramo još dva stupca potrebnih umnožaka $\mathbf{p_1 * q_1}$ i $\mathbf{p_0 * q_1}$. Prvi računamo u **H2** i pišemo **=D2*E2**, razvučemo i sumiramo. Drugi računamo u **I2** pišući **=B2*E2** razvučemo i sumiramo. Zatim izračunamo Paascheov indeks cijena tako da podijelimo ove dvije sume **=H7/I7*100 i** dobiveni broj zaokružimo na dvije decimale. Indeks iznosi 110,33 što znači da je došlo do povećanja od 10,33% (110,33-100) pa pišemo interpretaciju:

Cijene svih vrsta povrća u 2018. su se povećale za 10,33% u odnosu na 2017. godinu računajući uz neizmjenjene količine iz tekuće 2018. godine.

7.2.2. Skupni indeksi količina

Skupni indeks količina mjeri prosječnu promjenu količina skupine pojava u dva različita vremenska razdoblja i to na bazi nepromijenjenih cijena.

Laspeyresov indeks količina:

$$Q_{01}(p_0) = \frac{\sum_{i=1}^{k} q_{i1} \cdot p_{i0}}{\sum_{i=1}^{k} q_{i0} \cdot p_{i0}} \cdot 100$$

Laspeyresov indeks količina označava ukupnu promjenu količina skupine pojava računatu uz nepromijenjene cijene baznog razdoblja. Za izračun ovog indeksa moramo imati sume umnožaka $\mathbf{p_0 * q_0}$ i $\mathbf{p_0 * q_1}$. No budući ih već imamo odmah ćemo i izračunati indeks **=I7/G7*100**. Dobiveni broj zaokružimo na dvije decimale. Indeks iznosi 84,35 što znači da je došlo do smanjenja od 15,65% (84,35-100) pa pišemo:

Količine svih vrsta povrća zajedno u 2018. godini smanjile su se za 15,65 % u odnosu na 2017. računajući uz fiksne cijene iz bazne 2017. godine.

Paascheov skupni indeks količina

$$Q_{01}(p_1) = \frac{\sum_{i=1}^{k} q_{i1} \cdot p_{i1}}{\sum_{i=1}^{k} q_{i0} \cdot p_{i1}} \cdot 100$$

Paascheov oblik skupnog indeksa količina pokazuje za koliko su se u prosjeku promijenile količine skupine pojava, pri čemu se cijene održavaju nepromijenjene na nivou tekućeg razdoblja. Kako bi ga izračunali moramo imati sume umnožaka $\mathbf{p_1 * q_1}$ i $\mathbf{p_1 * q_0}$ koje već imamo izračunate pa ih koristimo i računamo **=H7/F7*100**. Dobiveni broj zaokružimo na dvije decimale. Indeks iznosi 85,31% što znači da je došlo do smanjenja od 14,69% (85,31-100) pa pišemo:

Količine svih vrsta povrća zajedno u 2018. godini smanjile su se za 14,69% u odnosu na 2017. računajući uz fiksne cijene iz tekuće 2018. godine.

Skupni indeks vrijednosti

$$V_{01} = \frac{\sum_{i=1}^{k} p_{i1} \cdot q_{i1}}{\sum_{i=1}^{k} p_{i0} \cdot q_{i0}} \cdot 100$$

Skupni indeksi vrijednosti mjere promjene uzrokovane istovremenim mijenjanjem cijena i količina promatrane skupine pojava. Budući već imamo sume umnožaka $\mathbf{p_1 * q_1}$ i $\mathbf{p_0 * q_0}$ odmah računamo **=H7/G7*100**. Indeks iznosi 93,06 što znači da je došlo do pada od 6,94% (93,06-100) i interpretiramo:

Ukupna vrijednost povrća u 2018. se smanjila za 6,94% u odnosu na 2017. godinu.

	A	B	C	D	E	F	G	H	I
1	**vrsta povrća**	cijene - 2017. p0	količine-2017. q0	cijene - 2018. p1	količine - 2018. q1	p1*q0	p0*q0	p1*q1	p0*q1
2	blitva	15	140	18	120	2520	2100	2160	1800
3	kupus	11	100	13	76	1300	1100	988	836
4	cvjetača	12	80	15	90	1200	960	1350	1080
5	brokula	18	220	16	170	3520	3960	2720	3060
6	špinat	16	125	20	110	2500	2000	2200	1760
7						11040	10120	9418	8536
8	SKUPNI INDEKSI CIJENA								
9									
10	LASPEYRESOV INDEKS CIJENA P01(q0)			109,09	9,09				
11									
12	PAASCHEOV INDEKS CIJENA P01(q1)			110,33	10,33				
13									
14	SKUPNI INDEKSI KOLIČINA								
15									
16	LASPEYRESOV INDEKS KOLIČINA Q01(p0)			84,35	-15,65				
17									
18	PAASCHEOV INDEKS KOLIČINA Q01(p1)			85,31	-14,69				
19									
20	SKUPNI INDEKS VRIJEDNOSTI V01			93,06	-6,94				

7.2.3. Posebni oblici skupnih indeksa

Deflacioniranje, revalorizacija

Budući se cijene u pravilu mijenjaju s vremenom ne možemo imati pravi uvid u dinamiku mijenjanja pojave s vremenom bez uvažavanja utjecaja promjena cijena na tijek tih promjena. Preciznije rečeno, moramo onemogućiti utjecaj neminovne promjene cijena na vrijednost pojave iskazane vremenskim nizom. Taj se postupak zove **deflacioniranje**. Provodi se dijeljenjem nominalnih vrijednosti (tj. vrijednosti iskazanih u cijenama tekućeg razdoblja) s odgovarajućim indeksom cijena pomnoženim sa 100 (tzv. deflacijskim indeksom ili deflatorom). Pri tome se svi indeksi moraju odnositi na istu baznu godinu tj. godinu uz čije cijene se računaju vrijednosti.

Vježba 7.5.

	A	B	C	D	E	F
1	**Godina**	**Proizvodnja grožđa u tekućim cijenama (000 kn)**	**Lančani indeksi cijena**			
2	2002	4,52	nepoznat			
3	2003	4,80	101,02			
4	2004	5,12	101,82			
5	2005	5,80	102,36			
6	2006	6,10	104,48			
7	2007	6,30	105,52			
8	2008	6,80	105,68			
9	2009	6,80	102,36			
10	2010	7,00	103,78			
11	2011	7,20	108,92			
12	2012	7,25	110,37			
13	2013	7,30	111,48			
14	2014	7,50	111,92			
15	2015	8,00	112,46			
16						
17						
18	Lančani indeksi cijena i proizvodnja grožđa,na poljoprivrednom imanju					
19	Primoštenski vinograd u razdoblju od 2002. do 2015. godine, dani su slijedećom tablicom:					
20	a)Proizvodnju grožđa iskažite u stalnim cijenama iz 2011. godine					
21	b)Izračunajte prosječnu godišnju stopu promjena prosječnih otkupnih cijena grožđa.					

Rješenje:

a) Zadana je proizvodnja grožđa u tekućim cijenama kao i verižni indeksi cijena u tom nizu godina. Kako bi mogli proizvodnju grožđa iskazati u stalnim cijenama (odabrane 2011. godine) moramo prvo verižne indekse pretvoriti u bazne s 2011. kao baznom godinom. Postupak nećemo detaljno obrazlagati jer je objašnjen u **Vježbi 7.3.** Nakon tog postupka tablica izgleda:

	A	B	C	D
1	Godina	Proizvodnja grožđa u tekućim cijenama (000 kn)	Lančani indeksi cijena	It 2011.=100
2	2002	4,52	nepoznat	70,46
3	2003	4,80	101,02	71,17
4	2004	5,12	101,82	72,47
5	2005	5,80	102,36	74,18
6	2006	6,10	104,48	77,50
7	2007	6,30	105,52	81,78
8	2008	6,80	105,68	86,43
9	2009	6,80	102,36	88,47
10	2010	7,00	103,78	91,81
11	2011	7,20	108,92	100
12	2012	7,25	110,37	110,37
13	2013	7,30	111,48	123,04
14	2014	7,50	111,92	137,71
15	2015	8,00	112,46	154,87

Sada ćemo tablicu proširiti stupcem u kojem ćemo izračunati *Proizvodnju grožđa izraženu u stalnim cijenama iz 2011. godine*. Kao što je u uvodu ove cjeline rečeno postupak ćemo provesti tako da proizvodnju u tekućim cijenama podijelimo s deflacijskim faktorom (baznim indeksima na nivou 2011. puta 100) tako da u ćeliji **E2** pišemo **=B2/D2*100** pravilo kopiramo za sve godine i dobijemo:

	A	B	C	D	E
1	Godina	Proizvodnja grožđa u tekućim cijenama (000 kn)	Lančani indeksi cijena	It 2011.=100	Proizvodnja grožđa u stalnim cijenama iz 2011. (000 kn)
2	2002	4,52	nepoznat	70,46	6,42
3	2003	4,80	101,02	71,17	6,74
4	2004	5,12	101,82	72,47	7,07
5	2005	5,80	102,36	74,18	7,82
6	2006	6,10	104,48	77,50	7,87
7	2007	6,30	105,52	81,78	7,70
8	2008	6,80	105,68	86,43	7,87
9	2009	6,80	102,36	88,47	7,69
10	2010	7,00	103,78	91,81	7,62
11	2011	7,20	108,92	100	7,20
12	2012	7,25	110,37	110,37	6,57
13	2013	7,30	111,48	123,04	5,93
14	2014	7,50	111,92	137,71	5,45
15	2015	8,00	112,46	154,87	5,17

b) **Prosječna stopa promjene**

Prosječna stopa promjena se definira kao prosječna promjena vrijednosti pojave (povećanje ili smanjenje, ovisno o predznaku stope) iskazana u postocima. Ona je prikladna zamjenska veličina niza pojedinačnih stopa samo ako one ne očituju velike brojčane varijacije. Računa se pomoću geometrijske sredine lančanih indeksa:

$$\mathrm{G} = \sqrt[N-1]{\mathrm{V}_2 \cdot V_3 \cdot \ldots \cdot V_N}$$

Interpretira se stopa: $\bar{S} = G - 100$

U MS Excelu postoji funkcija koja računa geometrijsku sredinu, a to je **GEOMEAN,** pa prvo izračunamo geometrijsku sredinu verižnih indeksa kao **=GEOMEAN(C3:C15)** a zatim pomoću dobivene vrijednosti (106,25) izračunamo stopu promjene 6,25 (106,25-100). Vrijednost prosječne stope promjene interpretiramo:

Prosječna otkupna cijena grožđa prosječno je godišnje rasla 6,25 %.

<u>Za vježbu rješite</u>:

U mapi **Primjeri vježbi 1,2,3,4,5** vježbu **5.5.**

8. Trend modeli

Model vremenske pojave je analitički zapis njene promjene u vremenu. Veliki broj vremenskih nizova pokazuje osnovnu razvojnu tendenciju (trend). Većina trend modela se ubraja u regresijske modele te se analizira na isti način kao i ti modeli, pri čemu je uvijek nezavisna varijabla vrijeme. Uz određene uvjete, modeli se koriste i u prognostičke svrhe.

8.1. Linearni trend model

Jednadžba linearnog trend modela ima isti oblik kao i linearni regresijski model:

$$\hat{Y} = B \cdot X_t + A$$

Pri čemu su Y – promatrana pojava čiju vremensku dinamiku analiziramo

X_t– vrijeme (dani, mjeseci, kvartali, godine,...)

B - očekivana prosječna promjena (povećanje ili smanjenje, ovisno o predznaku) varijable Y pri povećanju nezavisne varijable X_t za jedinicu vremena

A - očekivana vrijednost (trend vrijednost) varijable Y u ishodištu vremenskog razdoblja ($X_t = 0$)

Reprezentativnost trend modela procijenjuje se na sličan način kao i kod regresijskih modela.

Vježba 8.1.

Poznati su podaci o stanovništvu naselja X u Dalmatinskoj zagori u vremenu od 1930. do 2000. godine i dani su narednom tablicom. Zadatak glasi:

a) Nađite jednadžbu linearnog trenda s ishodištem u 1930. godini. Obrazložite značenje dobivenih vrijednosti parametara.

b) Pomoću dobivenog modela prognozirajte broj stanovnika u 2010. godini te u razdoblju od 2010. do 2050. godine.

c) Ako bi kretanje broja stanovnika zadržalo uočeni trend, odredite desetljeće u kojem će naselje opustjeti.

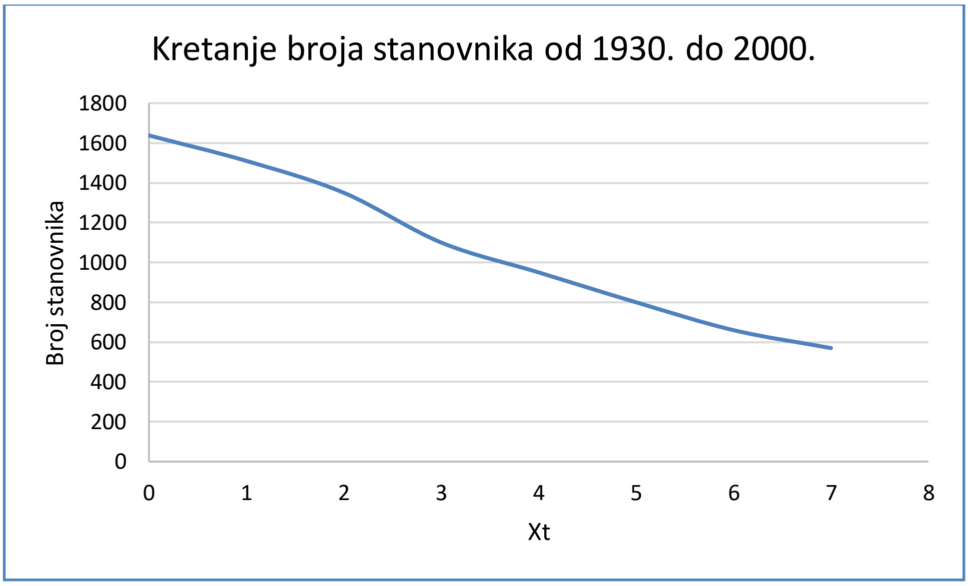
Rješenje:

Rješavanje započinjemo tako da kreiramo novu nezavisnu varijablu X_t koju ubacimo neposredno ispred stupca zavisne varijable Y – broj stanovnika. Ishodišnom desetljeću 1930. pridružimo vrijednost 0, a idućim godinama redom 1, 2, 3,...

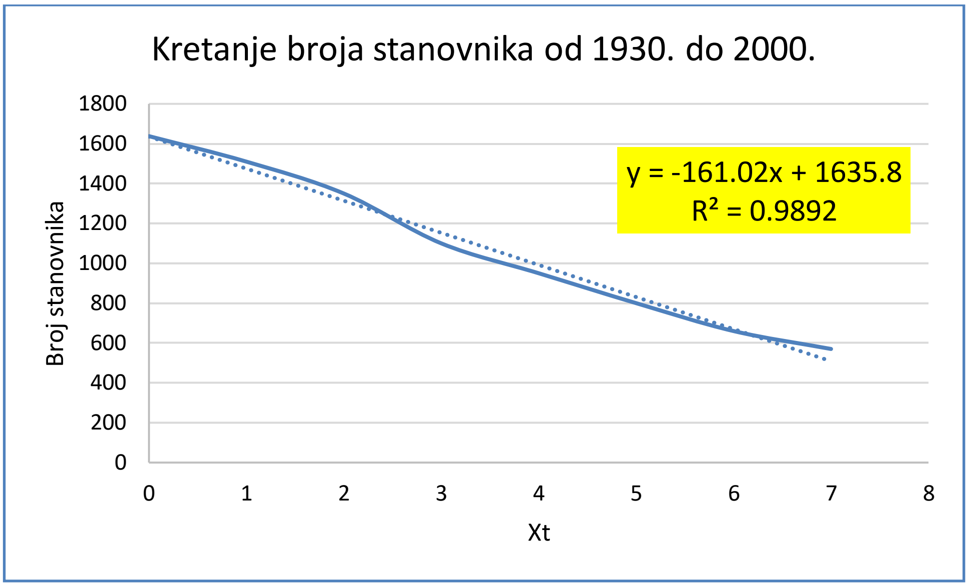
Nakon tog tablica izgleda ovako:

	A	B	C
1	**Godina**	**Xt**	**Broj stanovnika**
2	1930	0	1638
3	1940	1	1510
4	1950	2	1350
5	1960	3	1100
6	1970	4	950
7	1980	5	800
8	1990	6	660
9	2000	7	570

Idući je korak crtanje linijskog dijagrama kretanja pojave u vremenu. Označimo podatke (X,Y) bez zaglavnih ćelija, dakle **B2:C9.** Zatim na kartici *Insert/Scatter/* biramo *Scatter with smooth lines (*treći po redu-linijski*)* te predložak koji nam daje naslove i oznake na osima (br.1).



a) Linearni trend model ubacujemo kao i regresijski: biramo *linearni model* te opciju ubacivanja *jednadžbe modela* i koeficijenta R^2.



B = -161,02 – očekuje se prosječno smanjenje broja stanovnika u naselju X tijekom desetljeća za 161 osobu

A = 1635,8 (≈ 1636) trend broj stanovnika, u ishodišnoj 1930. godini iznosi1636 osoba.

R^2 =0,9892 vrijednost ukazuje na značajnu reprezentativnost modela.

b) Potrebno je napraviti prognozu broja stanovnika u 2010. godini. Kao i kod regresijskog modela prognozu možemo dobiti tako da u jednadžbu modela uvrstimo odgovarajući X (u ovom slučaju to bi bio $X_t = 8$). No možemo koristiti i funkciju koja daje prognozu po linearnom trendu za jedno vremensko razdoblje, **FORECAST(Novi X; Raspon poznate varijable Y; Raspon poznate varijable X)**. Prvo ćemo produžiti godine (i pripadnu varijablu X) do razdoblja za koje tražimo prognozu.

	A	B	C
1	**Godina**	**Xt**	**Broj stanovnika**
2	1930	0	1638
3	1940	1	1510
4	1950	2	1350
5	1960	3	1100
6	1970	4	950
7	1980	5	800
8	1990	6	660
9	2000	7	570
10	2010	8	
11	2020	9	
12	2030	10	
13	2040	11	
14	2050	12	

U ćeliji **C10** pišemo **=FORECAST(B10;C2:C9;B2:B9)** i dobivamo 347,64 te zaokružimo na cijeli broj. To je očekivani broj stanovnika u tom desetljeću.

Prognozu za više vremenskih razdoblja istovremeno možemo dobiti funkcijom **TREND(Raspon poznate varijable Y; Raspon poznate varijable X; Raspon nove varijable X)**

Postupak je sljedeći: Označimo blok ćelija u kojima želimo dobiti prognozu, npr. **C11:C14**, a u vrpci formula pišemo fx =TREND(C2:C9;B2:B9;B11:B14) i držeći tri tipke istovremeno **Ctrl+Shift+Enter** dobijemo prognozu za sva razdoblja istovremeno.

	A	B	C
1	**Godina**	**Xt**	**Broj stanovnika**
2	1930	0	1638
3	1940	1	1510
4	1950	2	1350
5	1960	3	1100
6	1970	4	950
7	1980	5	800
8	1990	6	660
9	2000	7	570
10	2010	8	348
11	2020	9	187
12	2030	10	26
13	2040	11	-135
14	2050	12	-296

c) Lako je vidjeti da se, ako se nastavi trend iseljavanja, između 2030. i 2040. može očekivati da će naselje opustjeti.

8.2. Eksponencijalni trend model

Ako se pojava opisana vremenskim nizom mijenja iz razdoblja u razdoblje za gotovo isti relativni iznos, tada je eksponencijalna funkcija prikladan prikaz trenda koji iskazuje. Model je oblika

$$\hat{Y} = A \cdot B^x$$

Interpretacija parametara je posve analogna onoj koju smo sreli kod eksponencijalnog regresijskog modela.

A - predstavlja očekivanu (trend) vrijednost nezavisne varijable Y u ishodištu promatranog vremenskog razdoblja $(X_t = 0)$

B - prosječan ritam promjene zavisne varijable Y pri jediničnoj promjeni nezavisne varijable. Pomoću parametra B računamo S

$$S = (B - 1) \cdot 100$$

S – predstavlja prosječnu periodičnu stopu promjene za promatrano razdoblje

Vježba 8.2.

U ovom ćemo primjeru pokazati kako je i eksponencijalni model primjenjiv u proučavanju migracija ljudi. Podaci za ovaj primjer preuzeti su s mrežnih stranica Državnog zavoda za statistiku. Na mrežnim stranicama Državnog zavoda za statistiku http://www.dzs.hr/ među bazama podataka nalaze se i podaci o **Naseljima i stanovništvu RH od 1857. do 2001.** gdje odabirom županije pa pojedinog grada/naselja i vremenskog intervala za koji želimo podatke dobivamo Excelov dokument s podacima o broju stanovnika. Ako želimo podatke s posljednjeg popisa stanovništva 2011. godine, potrebno je na istoimenom linku pronaći i taj podatak.

Pogledajmo prvo kretanje stanovništva u gradu Splitu tijekom proteklih 100 godina. Podaci su dobiveni putem popisa stanovništva koja su se obavljala otprilike svakog desetljeća (s iznimkom 2. svj.rata) .

Radi dobivanja jednostavnijeg modela umjesto stvarnih godina kreiramo vremensku varijablu X_t kojoj je ishodište u početku prvog razdoblja, dakle 1857. godine $(X_t = 0)$. Kasnija razdoblja numeriramo dalje 1, 2, 3, ...Pri ovom postupku aproksimiramo jedno desetljeće s jediničnom vrijednošću varijable.

	A	B	C
1	Godina	Xt	Broj stanovnika
2	1857.	0	12417
3	1869.	1	14587
4	1880.	2	16883
5	1890.	3	18438
6	1900.	4	21925
7	1910.	5	25103
8	1921.	6	29155
9	1931.	7	40029
10	1948.	8	54187
11	1953.	9	64874
12	1961.	10	85374
13	1971.	11	129203
14	1981.	12	176303
15	1991.	13	200459
16	2001.	14	188694
17	2011.	15	189388

a) Nađite jednadžbu eksponencijalnog trenda s ishodištem u 1857. godini. Obrazložite značenje dobivenih vrijednosti parametara.

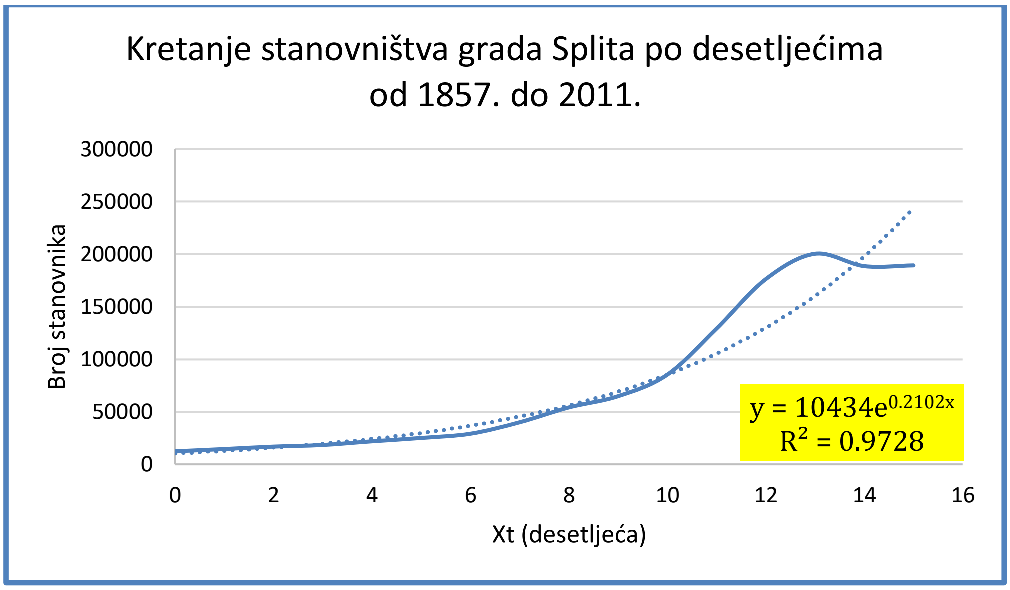
b) Ako rast broja stanovnika nastavi eksponencijalni trend pomoću dobivenog modela prognozirajte broj stanovnika u Splitu po desetljećima u razdoblju od 2021. do 2051. godine.

Rješenje:

a) Izrada eksponencijalnog trend modela identična je kao i linearnog trend modela. Osim jednadžbe modela dobivamo i vrijednost R^2, koeficijenta determinacije koji je pokazatelj reprezentativnosti modela i kazuje u kojem se postotku promjena zavisne varijable s vremenom može objasniti s vrijednostima koje daje izabrani model. U ovom slučaju on iznosi 0,9728.

Koeficijent determinacije za ekponencijalni trend upućuje da se 97,28% periodičnih promjena broja stanovnika u gradu Splitu (u razdoblju od 1857. do 2011.) objašnjava eksponencijalnim trend modelom.

Graf s eksponencijalnim trend modelom nakon uređivanja izgleda:



Očitamo i interpretiramo parametre modela:

A = 10434 predstavlja trend broj stanovnika u Splitu u ishodištu vremenskog razdoblja (1857. godina)

B = exp(0,2102)=1,23392 ne interpretiramo, već računamo stopu S

S = (B-1)*100=(1,23392-1)*100=23,39. Očekuje se prosječni godišnji rast broja stanovnika od 23,39%

b) Funkcija koja u MS Excelu vrši prognozu po eksponencijalnom trend modelu zove se **GROWTH(Raspon poznate varijable Y; Raspon poznate varijable X; Raspon nove varijable X)**.

Prvo ćemo razvući niz godina (i pripadne vrijednosti varijable X_t) do traženog razdoblja. Označimo blok od **C18:C21** u kojem želimo dobiti prognozirane vrijednosti i pišemo u vrpci formula *fx* =GROWTH(C2:C17;B2:B17;B18:B21) i pritiskom na tri tipke istovremeno **Ctrl+Shift+Enter** dobijemo željenu prognozu.

Dobivena prognoza pokazuje da stanovništvo grada Splita ima tendenciju eksponencijalnog rasta tijekom vremena. Naravno taj rast ne može biti beskonačan zbog ograničenih resursa pa je vjerojatno da će i ovaj model nakon nekog vremena ući u fazu stagnacije. Tablica na kraju izgleda:

	A	B	C
1	Godina	Xt	Broj stanovnika
2	1857.	0	12417
3	1869.	1	14587
4	1880.	2	16883
5	1890.	3	18438
6	1900.	4	21925
7	1910.	5	25103
8	1921.	6	29155
9	1931.	7	40029
10	1948.	8	54187
11	1953.	9	64874
12	1961.	10	85374
13	1971.	11	129203
14	1981.	12	176303
15	1991.	13	200459
16	2001.	14	188694
17	2011.	15	189388
18	2021	16	301273
19	2031	17	371743
20	2041	18	458697
21	2051	19	565991

Za vježbu riješite:

Riješite u mapi **Primjeri vježbi 1,2,3,4,5** vježbu **5.8.**

Sa stranica Državnog zavoda za statistiku skinite po uzoru na Vježbu 8.2. povijesne podatke o naseljenosti nekog grada/sela i ispitajte da li slijede neki od ova dva trend modela.

Literatura:

1. K. Mikelić: *Poslovna statistika*, radni materijali za predavanja, Moodle.

2. M. Papić: *Primijenjena Statistika u MS Excelu*, Naklada ZORO, 2005.

3. I. Bacci, I. Plazibat: *Statistika u MS Excelu*, Veleučilište u Splitu, 2002.

4. N. Roguljić, A. Burazin Mišura, I. Baras: *Eksponencijalna funkcija i njezine primjene u realnom životu*, Poučak Vol. 14 No. 53, 2013.