



LLL–Seminari u okviru TEMPUS projekta

Naziv projekta: 511140 – TEMPUS – JPCR
"Master programe in Applied Statistics - MAS"

Broj projekta: 511140

Nosilac projekta: Departman za matematiku i informatiku,
PMF Novi Sad

Rukovodilac: Prof. dr Andreja Tepavčević

Vreme trajanja: 15.10.2010. – 14.10.2013.

Finansiranje: Projekat finansira EU

**PRIPREMA STATISTIČKOG
ISTRAŽIVANJA, PRIKUPLJANJE,
SREĐIVANJE I PRIKAZIVANJE
STATISTIČKOG MATERIJALA**

Naziv *Statistika*:

- *status* (latinska reč, znači stanje) i
- *regione di stato* (ekvivalent italijanskoj reči za državu, državni interes).

Statistika je metod

kvantitativnog istraživanja pojava

STATISTIČKI SKUP I JEDINICE POSMATRANJA

Skup svih pojedinačnih elemenata (jedinica) koji je predmet statističkih istraživanja, naziva se *statistički skup* ili *statistička masa* ili *populacija*. Zavisno od broja statističkih jedinica, populacija može biti konačna ili beskonačna.

Primeri statističkog skupa ili populacije koji se vrlo često koriste za razna istraživanja su:

- skup stanovnika neke države, grada ili ulice;
- mesečna (godišnja) proizvodnja nekog proizvoda u posmatranoj fabrici (grani privrede);
- skup telefonskih poziva u toku jednog dana (meseca);
- skup novorođene dece u jednoj kalendarskoj godini u državi;
- skup upisanih studenata na početku školske godine, i drugi.

Osnovne osobine populacije su:

- *homogenost* (jedinice moraju biti istovrsne, tj. homogene, što znači da imaju bar jedno zajedničko svojstvo);
- *diferenciranost* (jedinice ipak nisu istovetne, tj. imaju i neka različita svojstva);
- *celovitost* (statistički skup obuhvata sve moguće slučajeve posmatrane pojave u vremenu i prostoru).

Populacija je određena:

- *prostorno* (definisana je lokacija na kojoj se posmatraju elementi populacije);
- *vremenski* (definisana je momenat ili vremenski interval u kome se vrši posmatranje jedinica populacije. U nekim slučajevima utvrđuje se tzv. *kritični momenat* – tj. datum i vreme. Naprimer, kod nas se popis stanovnika vrši svake 10-te godine i kritičan momenat je 31. mart godine popisa. Posle toga datuma novorođeni i umrli se ne računaju.);
- *sadržajno* (precizno su utvrđene osobine koje treba da ima pojedini elemenat da bi on bio jedinica posmatranja u populaciji).

Šta je danas najčešće predmet statističkih istraživanja?

Istraživanja se vrše sa ciljem da se procene neki socijalni i ekonomski pokazatelji u društvu, kao što su:

- političko javno mnjenje
- nezaposlenost
- zdravstvena zaštita
- rezultati rada neke radne organizacije
- stanje u obrazovnom sistemu države
- mišljenje javnosti o predloženom zakonu, o načinu regulisanja zarada, o nekoj aktuelnoj situaciji, o gledanosti nakih TV programa ili emisija
- kada treba proceniti poljoprivredne prinose, rod žetve i dr.

STATISTIČKA OBELEŽJA

Statističke jedinice (elementi) koje čine posmatranu populaciju moraju biti homogene, odnosno imati bar jednu zajedničku osobinu. Karakteristika ili osobina koju ispituujemo na nekoj populaciji naziva se *obeležje* ili *karakter*.

Obeležja se dele na:

- *kvantitativna* ili *numerička*, i
- *kvalitativna* ili *atributivna*.

PRIMER 1.

- Broj studenata na nekom fakultetu. (numeričko obeležje)
- Visina i težina svakog posmatranog studenta na nekom fakultetu. (numeričko obeležje)
- Boja kose i pol posmatranog studenata. (atributivno obeležje)

Postoje obeležja koja se mogu prikazati i atributivno i numerički.

PRIMER 2.

Uspeh učenika može se opisati atributivno sa: odličan, vrlo dobar, dobar, dovoljan i nedovoljan ili numerički sa: 5,4,3,2 i 1.

Cilj istraživanja određuje precizno koja će se vrsta obeležja uzeti u razmatranje. Kako je svako obeležje varjabilna karakteristika svakog posmatranog elementa populacije, to se grupisanju elemenata u odnosu na to obeležje posvećuje posebna pažnja jer to bitno utiče na rezultat istraživanja.

Druga podela obeležja vrši se na osnovu primenjene merne skale:

- *nominalna skala* je skup kategorija koje nisu uređene (rangirane). One se označavaju rečima (npr. boja kose: plava, smeđa, crna) ili brojevima (npr. pol: muški (1) , ženski (0)) i te kategorije se mogu navesti u proizvoljnom poretku.
- *ordinalna skala* je skup uređenih (rangiranih) kategorija, ali sa nedefinisanim rastojanjem između njih (npr. za tvrdoću materijala koristimo sledeće oznake: jagoda (1) , drvo (2) , metal (3)).
- *intervalna skala* poseduje jedinicu mere, pa je rastojanje između svake dve susedne kategorije jednako (npr. lenjir, toplomer, težinska vaga,...).
- *skala odnosa* je intervalna skala koja ima pravu (fizičku) nulu. Ovom skalom se postiže najviši stepen merenja. Lenjir i težinska vaga su primeri skale odnosa jer 0 cm i 0 g znače odsustvo dužine i težine. Slično, 10 kg je duplo teže od 5 kg (regulisan odnos). Celzijusova skala za merenje temperature jeste intervalna skala, ali nije skala odnosa, jer 0°C ne znači odsustvo toplote i 4°C ne znači da je duplo toplije od 2°C.

PRIMENA RAČUNARA U STATISTICI

Softverski paketi koji se najčešće koriste:

S-Plus – statistički paket namenjen za korisnike koji se profesionalno bave statistikom; daje mogućnost programiranja i omogućava rad sa više baza podataka.

Statgraphics – namenjen korisnicima koji ne žele da izučavaju statističku teoriju, već da se samo bave njenim primenama. Ovaj paket ima ugrađenu opciju za proveru uslova potrebnih za primenu nekih statističkih testova i u slučaju kada su oni narušeni izveštava korisnika o tome i upućuje ga na naredne moguće opcije. Dobijeni rezultati se daju sa obrazloženjima.

SPSS (*Superior Performance Software Systems*) i *Statistica* – su statistički paketi koji su po mogućnostima između prva dva i oni su najčešće korišćeni programski paketi za statističku analizu. Za njihovu upotrebu potrebno je poznavanje statistike.

SAS (*Statistical Analysis Systems*) – statistički paket koji je u kasnijim sedamdesetim godinama XX veka bio čisto statistički softverski sistem za upravljanje i analizu podataka.

Excel – paket koji se dobija kao sastavni deo *Microsoft*-ovog paketa *Office*.

Modul "**Statistical**" u *Excel*-u sadrži 78 statističkih funkcija koje su raspoređene po abecedi. Pre bilo koje statističke analize uz pomoć računara potrebno je napraviti bazu prikupljenih podataka. Ova baza se može formirati u samom statističkom paketu ili može biti uneta iz nekog drugog paketa, npr. *Microsoft Excel*-a.

PSPP – jedan od novijih statističkih paketa koji se sve više koristi; po mogućnostima moglo bi se reći da je na neki način komplementaran sa *SPSS*, tako da ih je dobro uporedo koristiti za statističke analize.

EduStat – statistički paket pogodan za oblast društvenih nauka, i drugi.

Statistika je zahvaljujući računarima uzela toliko zamaha da je danas nemoguće zamisliti čak i prenos sportskog događaja bez propratne statističke analize.

Na snazi je deviza:

“Statistics all around us”

jer je to zahtev vremena u kome živimo.

ETAPE STATISTIČKOG ISTRAŽIVANJA

Populacije skoro po pravilu sadrže veliki broj elemenata. Ukoliko je taj broj elemenata konačan označićemo ga sa N . Kada se ispituje bilo koja karakteristika populacije, kompletno ispitivanje svih članova populacije obezbeđuje željenu informaciju. Međutim, u mnogim slučajevima to je neekonomično, u drugim je nemoguće dobiti traženu informaciju na celoj populaciji, a nekada je potrebno traženu informaciju dobiti brzo.

PRIMER 3.

- (a) Kod ispitivanja šansi na izborima ne ispituju se svi glasači iz države jer je to neekonomično.
- (b) Fizički je neizvodljivo prebrojavanje svih riba u jednom jezeru ili pak prebrojavanje svih stabljika koprive na nekom staništu.

Prvu etapu u statističkom proučavanju populacije predstavlja *izbor uzorka* nad kojim će se vršiti posmatranje i prikupiti potrebni podaci. To znači da iz populacije izdvajamo n elemenata (statističkih jedinica). Broj n je *obim uzorka*.

Druga etapa je statističko posmatranje izabranog uzorka. U ovoj fazi vrše se anketiranja ili merenja nad elementima iz uzorka sa ciljem prikupljanja podataka.

U **trećoj etapi** grupišemo, sređujemo prikupljene i vršimo njihovo prikazivanje na više načina (tablično, grafički,...).

Četvrta etapa obuhvata statističku obradu podataka, naučnu analizu rezultata i donošenje zaključaka o populaciji.

Prva etapa – **IZBOR UZORKA**

Pre izbora uzorka populacija se mora podeliti na delove koji se zovu *jedinice*. Te jedinice moraju pokrivati celu populaciju i svaki element populacije može pripadati samo jednoj jedinici.

Verovatnoća tačnosti zaključka o populaciji je utoliko veća ukoliko uzorak bolje reprezentuje populaciju. Takav uzorak naziva se *reprezentativni uzorak*.

Da bi uzorak bio reprezentativan moraju biti ispunjeni neki uslovi. Jedan od ovih zahteva je da svaki element populacije ima jednaku verovatnoću da bude izabran u uzorak.

Dalje, kod izbora uzorka obično se postavlja pitanje koju veličinu uzorka koristiti? Odgovor nije uvek jednostavan, jer velika populacija nema direktan uticaj na standardnu grešku i uzoračku sredinu.

Pri izboru uzorka možemo se rukovoditi:

- *subjektivnim kriterijumom* (način izbora se vrši prema sopstvenoj oceni organizatora istraživanja), ili
- *objektivnim kriterijumom* (izbor po strogo objektivnoj proceduri tako da svaki elemenat ima podjednaku mogućnost da bude uključen u uzorak).

Postupak kojim se biraju jedinice iz populacije u uzorak, radi ocene posmatrane karakteristike cele populacije, zove se *plan uzorka*. Skup svih mogućih planova uzoraka deli se na dva osnovna tipa:

- *standardni planovi* i
- *nestandardni planovi*.

Standardni planovi se dele na: *konvencionalne* i *adaptivne planove*.

Konvencionalni plan uzorka je plan kod koga verovatnoće izbora jedinica u uzorak ne zavise ni od jedne vrednosti veličine koja se ispituje niti od vrednosti nekih parametara. Primer konvencionalnog plana su svi poznati verovatnosni planovi: ***prost slučajan uzorak***, *uzorak sa verovatnoćom proporcionalnom veličini*, ***stratifikovani slučajan uzorak***, *sistematski uzorak*, *uzorak skupina*, *višeetapni uzorak*, *višefazni uzorak* i drugi.

PRIMER 4.

Pomoću tablice slučajnih brojeva izdvojiti 10 od prvih 80 prirodnih brojeva.

Rešenje. U tablici slučajnih brojeva izaberemo neko mesto, npr. (7)(1) i idući recimo sa leva na desno beležimo sve dvocifrene brojeve:

1515	(7)(1)93	2522	2263	0487	8444	6186	5410	5927	...
9124	8972	0800	1484	0623	1397	0628	6046	1930	...
...

Dobijamo niz: 71, 93, 25, 22, 22, 63, 04, 87, 84, 44, 61, 86, 54, 10, 59, 27, ...

Izostavimo podvučene brojeve (veće od 80 i ponovljene, već izabrane, brojeve) i izdvojeni brojevi su:

71, 25, 22, 63, 4, 44, 61, 54, 10, 59.

Adaptivni planovi su takođe standardni planovi, ali kod njih postupak izbora elemenata u uzorak može da zavisi od vrednosti veličine koja se ispituje ali samo na elementima koji su uključeni u uzorak.

Planovi kod kojih verovatnoće izbora elemenata u uzorak zavise od vrednosti veličine koju posmatramo i kod jedinica izvan uzorka ili od vrednosti nepoznatih parametara su *nestandardni planovi*.

Naravno, ovo su samo neke mogućnosti izbora uzorka. U praksi, to je često vrlo složen proces. U daljim razmatranjima smatraćemo da je posmatrani uzorak reprezentativan.

Druga etapa – STATISTIČKO POSMATRANJE I PRIKUPLJANJE PODATAKA

Ovu drugu etapu u statističkom istraživanju prati zahtev za tačnošću, istinitošću i kompletnošću prikupljenih podataka. Da bi se to obezbedilo potrebna je saradnja stručnjaka za statistiku i stručnjaka za oblast istraživanja. Kao rezultat te saradnje dobija se *realizovani uzorak*. To je ustvari reprezentativan uzorak obima n izdvojen iz populacije obima N po osnovu obeležja X .

PRIMER 5.

Neka populaciju čine svi studenti završne godine na posmatranom smeru jednog fakulteta. Neka je obeležje X – visina svakog posmatranog studenta uzetog u uzorak. Uređena n – toraka brojeva (x_1, \dots, x_n) je realizacija uzorka obeležja X , gde je x_i ($i = 1, \dots, n$) visina i – tog studenta iz uzorka.

Treća etapa - GRUPISANJE I PRIKAZIVANJE PODATAKA IZ UZORKA

Prikupljeni podaci o posmatranom procesu ili pojavi čine sirov" materijal koji treba obraditi i oplemeniti. Nizovi sređenih statističkih podataka po modalitetima nazivaju se *statističke serije*. One se dele na:

- *serije strukture*
- *vremenske (hronološke) serije, i*
- *geografske serije.*

Serije strukture dele se na *numeričke i atributivne serije*.

Numeričke serije prikazuju raspored ili distribuciju frekvencija (broj pojavljivanja) nekih numeričkih obeležja i ta distribucija može biti *prosta i intervalna*.

Atributivne serije pokazuju distribuciju frekvencija po posmatranim atributivnim obeležjima.

Vremenske serije prikazuju varijacije posmatranih pojava tokom vremena. Mogu biti *momentne i intervalne vremenske serije*.

Geografske serije prikazuju prostorni (teritorijalni) raspored pojave. Ove serije se najčešće predstavljaju grafički pomoću *kartograma* na geografskim kartama.

Statističke serije se prikazuju ***tabelama*** i ***grafikonima***. Takvi prikazi ponekad daju jasnu sliku o posmatranom modalitetu, pa dalja analiza nije ni potrebna. U suprotnom, ova etapa je neophodna za dalju statističku obradu podataka.

STATISTIČKE TABELE

Podaci iz uzorka prikazuju se tabelarno na jedan od sledeća dva načina:

- vrednosti obeležja su dati brojevi;
- vrednosti obeležja pripadaju datim intervalima.

PRIMER 6.

Mesečna potrošnja voća u zimskom periodu u 40 domaćinstava u *kg* data je Tabelom 2.

Tabela 2.

potrošnja u <i>kg</i> (x_i)	9	12	15	7	20	16
broj domaćinstava (f_i)	4	10	6	8	5	7

Grupisati podatke u obliku numeričke serije. Izračunati relativnu frekvenciju u procentima, kumulativ ispod", kumulativ iznad" i kumulativnu frekvenciju u procentima.

Rešenje. Za vredosti obeležja korišćena je skala odnosa. Izračunati podaci su u Tabeli 3.

Tabela 3.

potr. (x_i)	br. dom. (f_i)	rel.fr. %	kum. ispod	kum. iznad	kum.fr. %
7	8	20%	8	40	20%
9	4	10%	12	32	30%
12	10	25%	22	28	55%
15	6	15%	28	18	70%
16	7	17,5%	35	12	87.5%
20	5	12,5%	40	5	100%
Σ	40	100%			

GRAFIČKO PRIKAZIVANJE PODATAKA

Statistički podaci iz uzorka su najpristupačniji korisnicima ako se predstave grafički. Grafičko prikazivanje podataka moguće je na više načina. Bitno je da taj prikaz bude jasan, jednostavan, pregledan i da odgovara prikupljenim statističkim podacima.

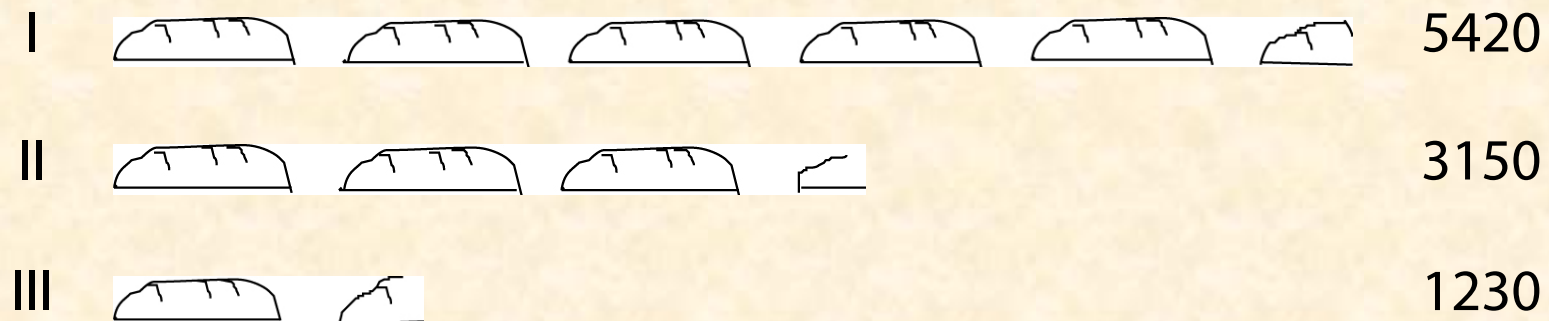
Jedna moguća podela grafikona je na bazi elemenata koje sadrže i to su:

- *dijagrami*
- *kartogrami, i*
- *piktogrami.*

Kartogrami su grafikoni na geografskim kartama i oni na pregledan način ilustruju statističke podatke kod geografskih serija. Ta ilustracija vezana je za teritorijalni raspored i intezitet posmatranih pojava.

Piktogrami na popularan i slikovit način prikazuju pojave. Naprimer, dnevna proizvodnja hleba u tri pekare u jednom gradu može se predstaviti piktogramom na Slici 1.

Ovde se crtežom hleba predstavlja 1000 hlebova.



Slika 1.

Dijagramima se mogu prikazati sve statističke serije. Najčešće korišćeni dijagrami su:

- tačkasti (*stimogrami*)
- linijski (*poligonalni*)
- površinski (*histogrami*)
- prostorni (*sterogrami*) i
- kružni.

PRIMER 7.

Iz godišnjeg izveštaja za 2003/04 godinu Tehničkog fakulteta u Čačku preuzete su prosečne ocene 40 diplomiranih studenata ovoga fakulteta sa različitih smerova:

6.89	6.80	6.77	7.31	7.97	6.53	6.71	7.54
7.57	7.11	8.11	6.87	7.06	6.58	7.53	6.94
8.00	7.34	8.50	7.10	6.86	7.50	7.34	7.90
7.48	6.60	7.20	6.90	7.57	7.33	6.80	7.12
7.53	8.20	7.87	8.60	8.67	7.95	8.48	8.57

Nacrtati histogram i poligon apsolutne i relativne frekvencije prosečne ocene posmatrane grupe studenata Tehničkog fakulteta u Čačku, kao i poligon (apsolutnih) kumulativnih frekvencija.

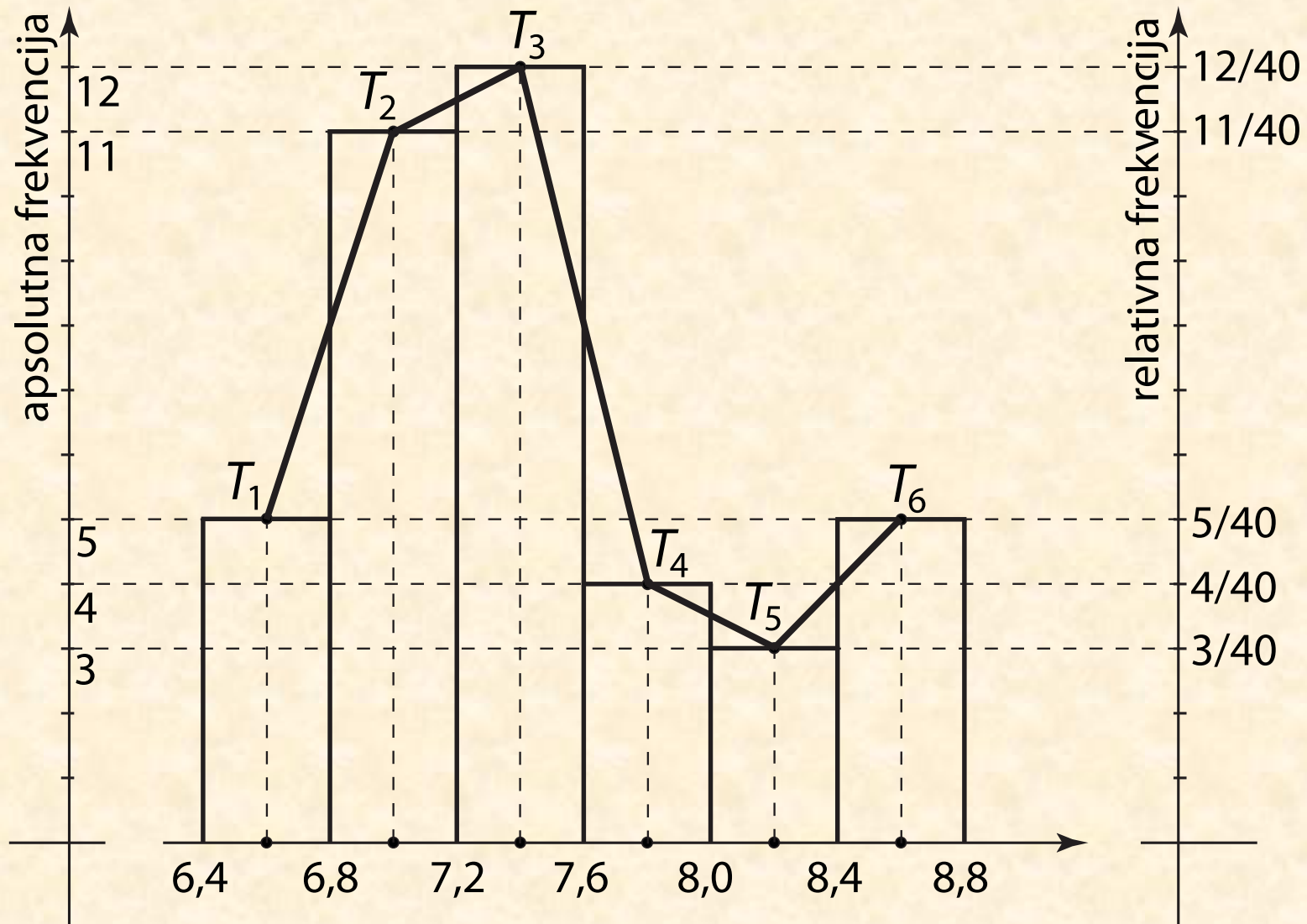
Rešenje. Kako je $x_{\max} = 8.67$ i $x_{\min} = 6.53$, to je raspon varijacije $x_{\max} - x_{\min} = 2.14$. Broj klasa $k = 1 + 3.32 \log 40 = 6.32$ pa ćemo uzeti $k = 6$ i pa je širina klase $d \approx \frac{x_{\max} - x_{\min}}{6} = 0.36$. Zato ako uzmemo $x_M = 8.80$ i $d = 0.40$ (što prikazujemo jediničnom duži na apscisnoj osi), dobijamo $x_m = 6.40$. Praktično, sada nam je interval varijacije $[6.40, 8.80]$ i tabelarni prikaz ove statističke serije dat je u Tabeli 5.

Tabela 5.

pros.ocena (x_i)	[6.4,6.8)	[6.8,7.2)	[7.2,7.6)	[7.6,8.0)	[8.0,8.4)	[8.4,8.8]
frekvencija (f_i)	5	11	12	4	3	5
rel.frekv. (r_i)	5 / 40	11 / 40	12 / 40	4 / 40	3 / 40	5 / 40
kumulat.frekv. (F_i)	5	16	28	32	35	40

Na Slici 2. prikazan je, istovremeno, histogram apsolutne i relativne frekvencije prosečne ocene posmatrane grupe studenata Tehničkog fakulteta u Čačku u obliku niza pravougaonika, sa konstatacijom da se prve frekvencije očitavaju na levoj a druge na desnoj ordinati.

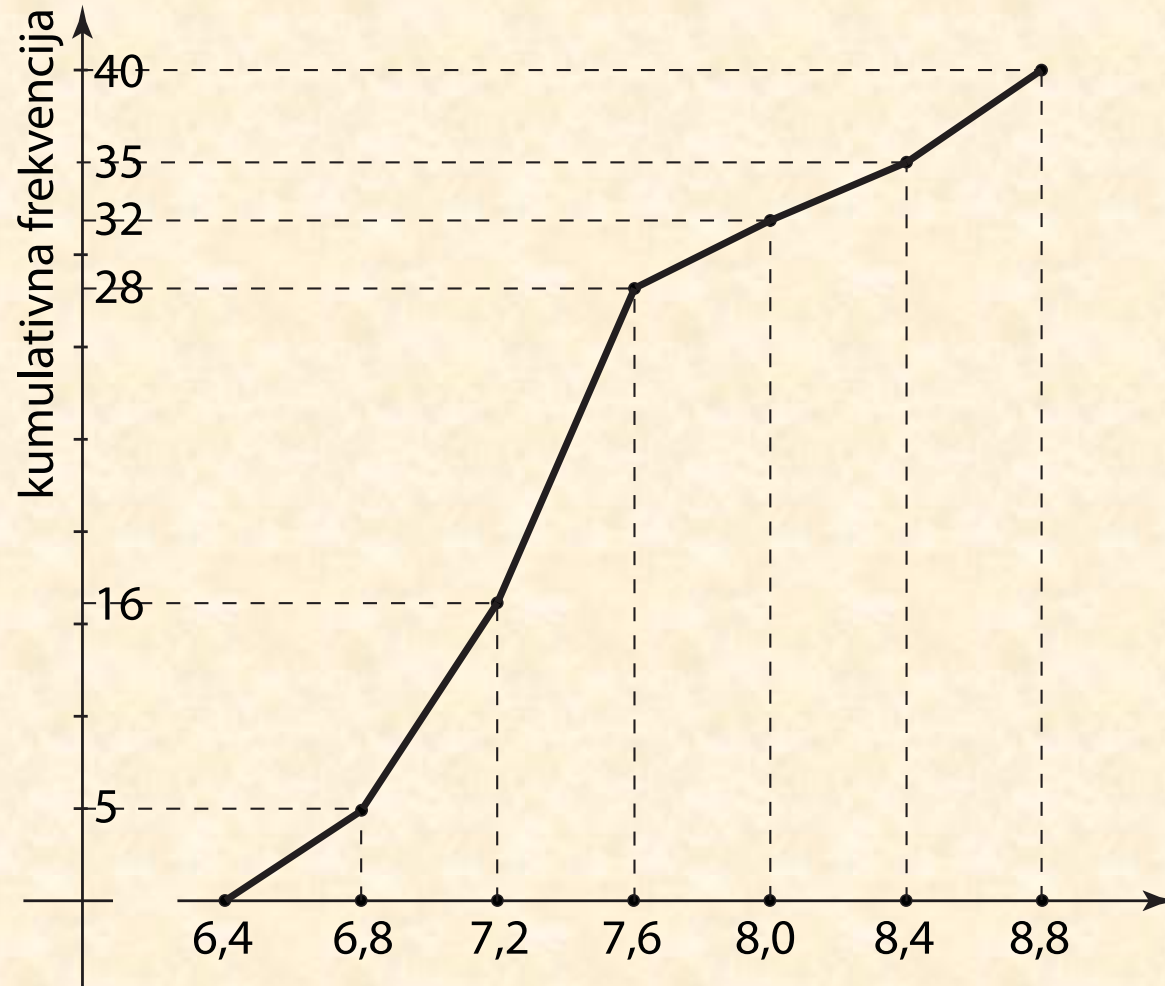
Sredine intervala prezentuju te intervale (klase), pa je izlomljena linija $T_1 T_2 T_3 T_4 T_5 T_6$ poligonalna linija apsolutnih, odnosno relativnih frekvencija.



Slika 2. Histogram i poligon apsolutne i relativne frekvencije prosečne ocene studenata.

Na Slici 3. predstavljen je poligon kumulativnih frekvencija i on je u vezi sa empirijskom funkcijom raspodele obeležja. Preciznije, kumulativnu raspodelu određuju kumulativne relativne frekvencije:

$5 / 40, 16 / 40, 28 / 40, 32 / 40, 35 / 40$ i 1 .



Slika 3. Kumulativna raspodela.

Kao površinski dijagrami, pored histograma (pravougaonici), mogu se koristiti kvadrati, krugovi i slično i taj prikaz ne mora biti u koordinatnom sistemu.

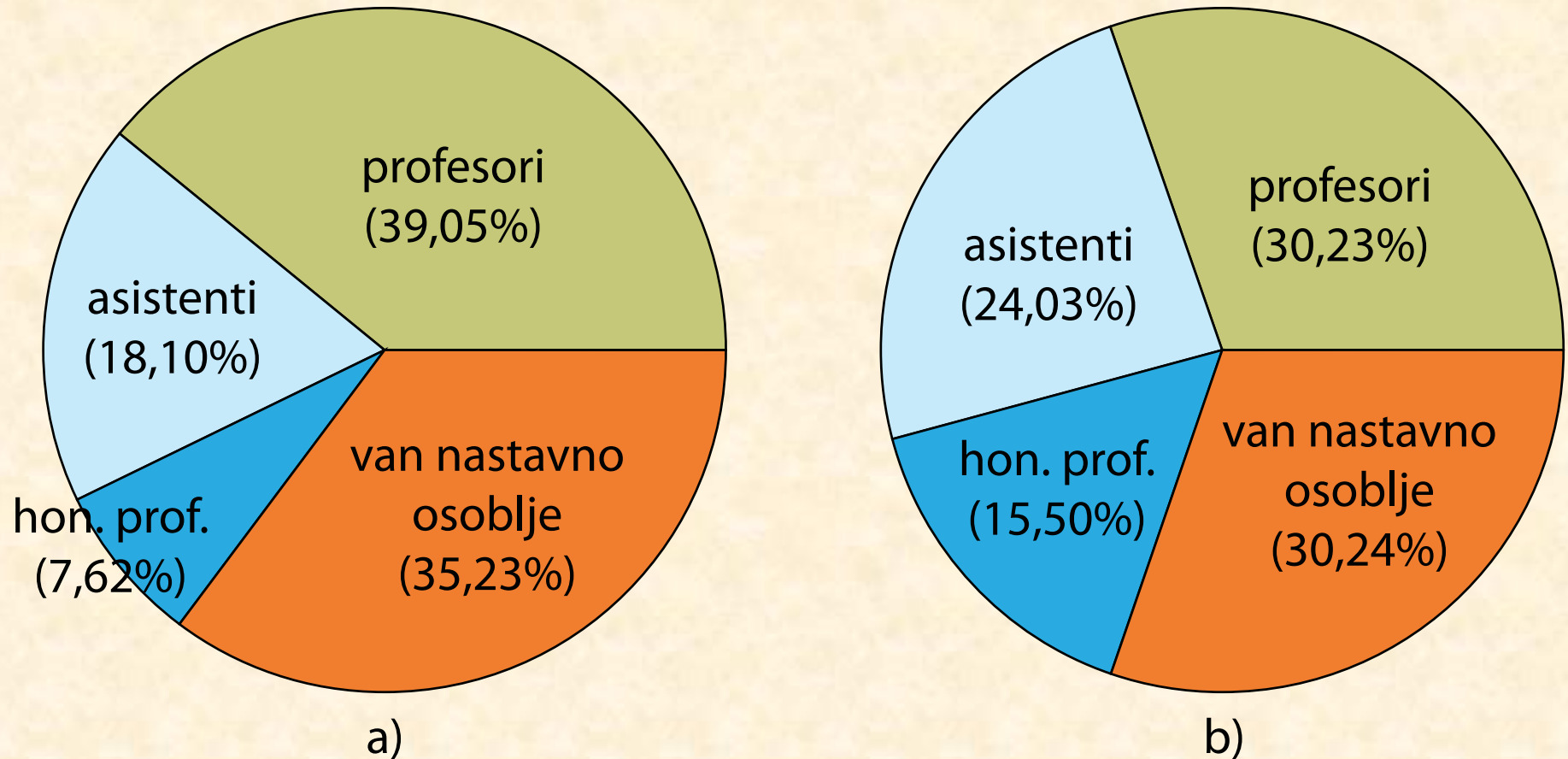
Za grafički prikaz relativne frekvencije (statistička verovatnoća) u procentima pogodno je korišćenje *kružnih dijagrama*. Upotreba kružnih dijagrama dosta je zastupljena u prikazivanju rezultata raznih anketa.

PRIMER 8. Raspored radnika na Tehničkom fakultetu u Čačku, prema opisu poslova koje obavljaju, za 2006/07 i 2011/12 godinu dat je Tabelom 6.

Tabela 6.

opis poslova	profesori	asistenti	honorarni prof.	van nastave	Σ
br.zaposlenih 2006/07	41	19	8	37	105
br.zaposlenih 2011/12	39	31	20	39	129

U strukturi kruga rezultati iz Tabele 6 dati su na Slici 4. pod a) za 2006/07 i pod b) za 2011/12 godinu.



Slika 4.

U narednom delu ilustrovaće se statistička obrada nekih aktuelnih podataka koji prate rad Tehničkog fakulteta u Čačku. Za to će se koristiti raspoloživi podaci iz arhive Fakulteta, tj. „Godišnjaci Fakulteta”. Za obradu tih podataka korišćen je paket *Statistica 8*.

Na Slici 5. prikazan je deo populacije koju čine studenti Tehničkog fakulteta koji su diplomirali u periodu od 2000. do 2004. godine. Obeležja koja ćemo pratiti na toj populaciji (i nekim drugim populacijama) su:

- **numerička:** godina rođenja, godina upisa, godina diplomiranja, prosek i godina na koju se student upisao;
- **atributivna:** ime i prezime, mesto rođenja, smer i pol.

Prikupljene podatke unecemo u tabelu u paketu Statistica, koja ce nam sluziti za sva naredna ispitivanja (Slika 5).

	1 Ime i prezime	2 God. rođenja	3 Mesto rođenja	4 God. upisa	5 Diplomirao/la	6 Prosek	7 Smer	8 Pol	9 God. koju je upisao
1	Remetic Ranko	1958	Nevesinje (BiH)	1999	2002	6.2	Tehnicky obrazovanje	M	3
2	Kostic Sarafin	1951	K. Kamenica	1981	2002	6.26	Tehnicky obrazovanje	M	1
3	Jevremovic Miroslav	1947	Kraljevo	1997	2001	6.29	Tehnika i informatika	M	3
4	Gavrilovic Mila	1961	Cacak	1995	2000	6.33	Tehnicky obrazovanje	Z	1
5	Jonuzi Mehmed	1975	Gora	1994	2003	6.33	Inženjer informatike	M	1
6	Avdic Emir	1969	Olovo (BiH)	1988	2001	6.4	Elektroenergetski sistemi	M	1
7	Colakovic Dušan	1959	Bor	2001	2003	6.4	Tehnicky obrazovanje	M	4
8	Kostic Dragan	1958	Priština	2001	2004	6.4	Tehnicky obrazovanje	M	3
9	Colovic Radojica	1977	Cacak	1996	2004	6.42	Tehnika i informatika	M	1
10	Ivetic Ivan	1969	Kragujevac	1998	2003	6.43	Tehnicky obrazovanje	M	4
11	Novicic Dragica	1956	Andrijevisa (CG)	1995	2001	6.44	Tehnicky obrazovanje	Z	3
12	Nikic Slavica	1959	Kraljevo	1996	2000	6.45	Tehnika i informatika	Z	3
13	Tošic Zoran	1973	Cacak	1991	2000	6.45	Elektroenergetski sistemi	M	1
14	Petriš Denis	1972	Užice	1991	2000	6.45	Elektroenergetski sistemi	M	1
15	Stevanovic Slaviša	1975	Valjevo	1994	2000	6.45	Elektroenergetski sistemi	M	1
				⋮					

Slika 5. Populacija u Statistici 8

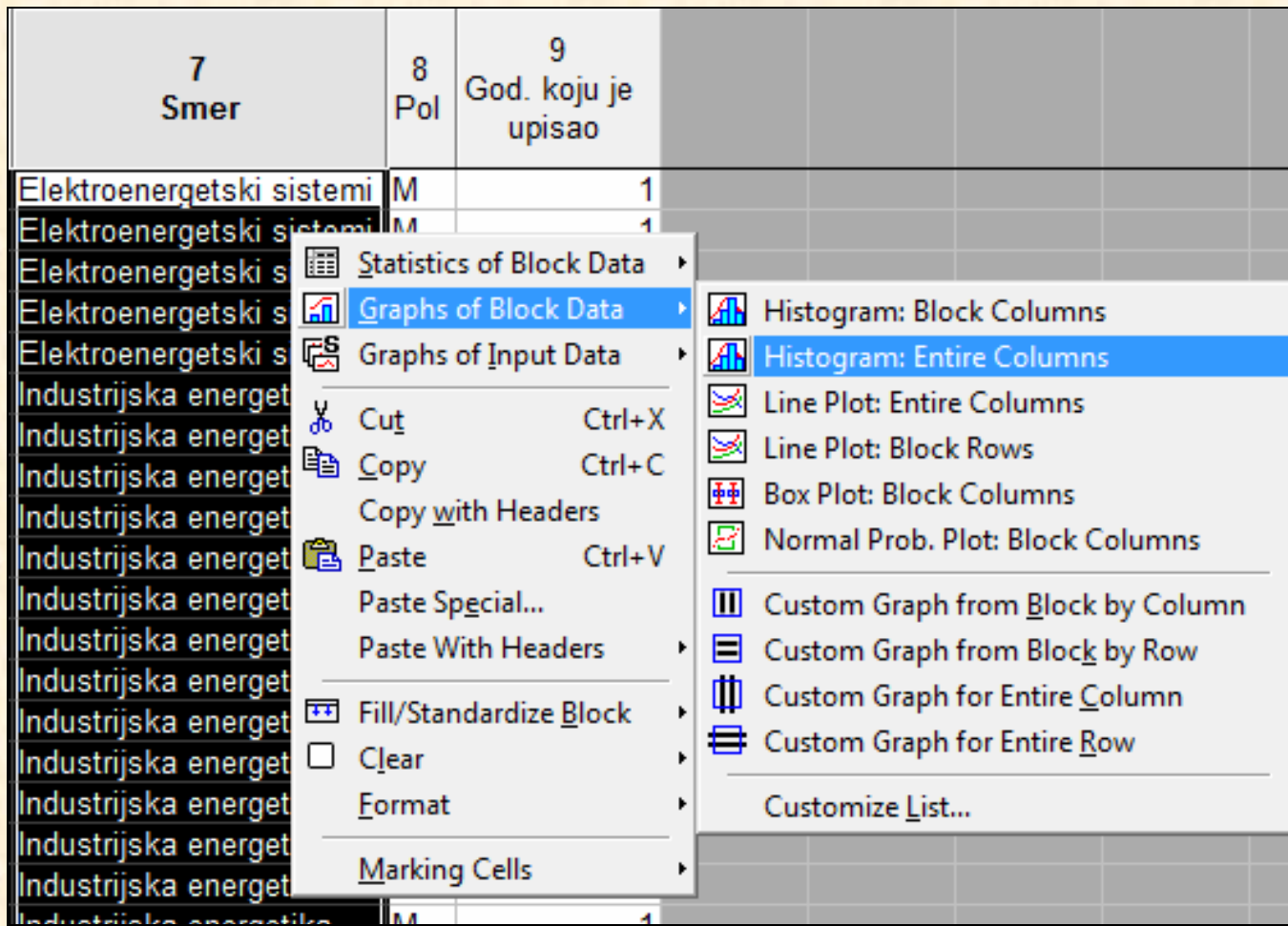
UZORCI IZABRANI IZ POPULACIJE PRIMENOM SUBJEKTIVNOG KRITERIJUMA

Kod subjektivnog kriterijuma organizator istraživanja sam vrši izbor uzorka iz populacije u skladu sa ciljevima istraživanja. Naprimer, ako nas zanimaju studenti koji su diplomirali 2001. godine, onda u tabeli sa Slike 5. uočimo kolonu sa obeležjem – Godina diplomiranja i selektujemo onaj deo te kolone koji ima vrednost 2001 (Slika 6).

	1 Ime i prezime	2 God. rođenja	3 Mesto rođenja	4 God. upisa	5 Diplomirao/la	6 Prosek	7 Smer	8 Pol	9 God. koju je upisao
1	Đikić Vladimir	1975	Surdulica	1994	2001	6.87	Elektroenergetski sistemi	M	1
2	Avdić Emir	1969	Olovo (BiH)	1988	2001	6.4	Elektroenergetski sistemi	M	1
3	Popadić Tanja	1969	Kruševac	1988	2001	6.72	Elektroenergetski sistemi	Z	1
4	Trkulja Aleksandar	1972	Beograd	1991	2001	7.12	Elektroenergetski sistemi	M	1
5	Milošević Radenko	1977	Cacak	1996	2001	8.72	Elektroenergetski sistemi	M	1
6	Stojić Ivan	1975	Trstenik	1992	2001	6.97	Industrijska energetika	M	1
7	Jovanović Srdan	1976	Sarajevo (BiH)	1994	2001	7.03	Industrijska energetika	M	3
8	Ilijević Dragan	1970	Kladovo	1989	2001	6.8	Industrijska energetika	M	1
9	Vasiljević Srdan	1972	Sm. Palanka	1991	2001	6.97	Industrijska energetika	M	1
10	Jović Mišo	1974	Letmathe (FRG)	1993	2001	6.87	Industrijska energetika	M	1
11	Vidić Aleksandar	1972	Valjevo	1996	2001	7.4	Industrijska energetika	M	3
12	Kovacević Rade	1974	Arilje	1993	2001	6.87	Industrijska energetika	M	1
13	Marinković Boban	1974	Prijepolje	1993	2001	6.9	Industrijska energetika	M	1
14	Vujović Milan	1968	Cacak	1987	2001	6.68	Industrijska energetika	M	1
15	Janković Milan	1973	Sarajevo (BiH)	1992	2001	7.09	Industrijska energetika	M	1

Slika 6. Uzorak studenata koji su diplomirali 2001. godine

Pri kreiranju dijagrama, dalje, moramo selektovati još jednu kolonu, odnosno varijablu, naprimer kolonu „Smer“. Zatim desnim klikom na selektovanu kolonu dobijamo prozor u kome biramo tip dijagrama, naprimer histogram. Postupak je prikazan na Slici 7.

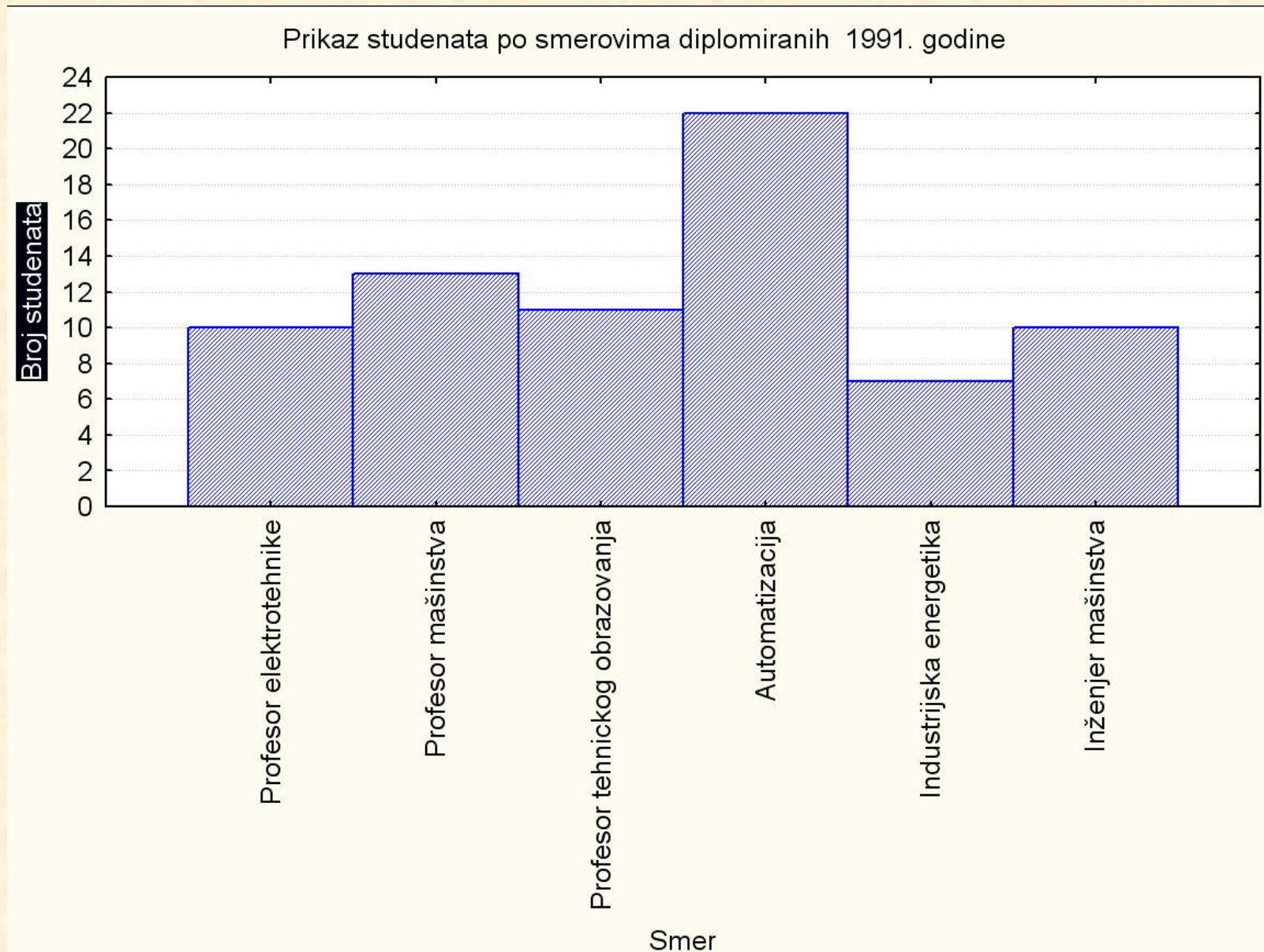


Slika 7. Postupak kreiranja histograma

Korišćenjem navedenih opcija dobijamo niz rezultata koje ćemo prezentovati.

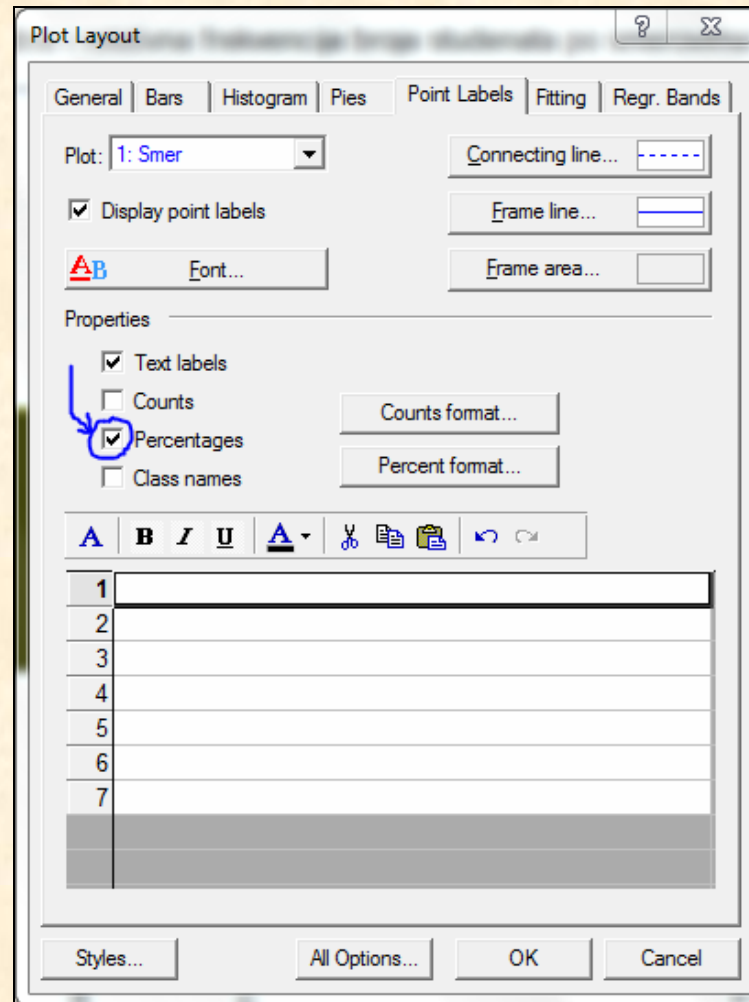
Za populaciju diplomiranih studenata od 1991. do 1994 imamo ukupno 214 unosa, odnosno 214 studenata koji su diplomirali na svim smerovima u tom periodu.

Iz te populacije izdvojimo studente koji su diplomirali 1991. godine (Slika 8).



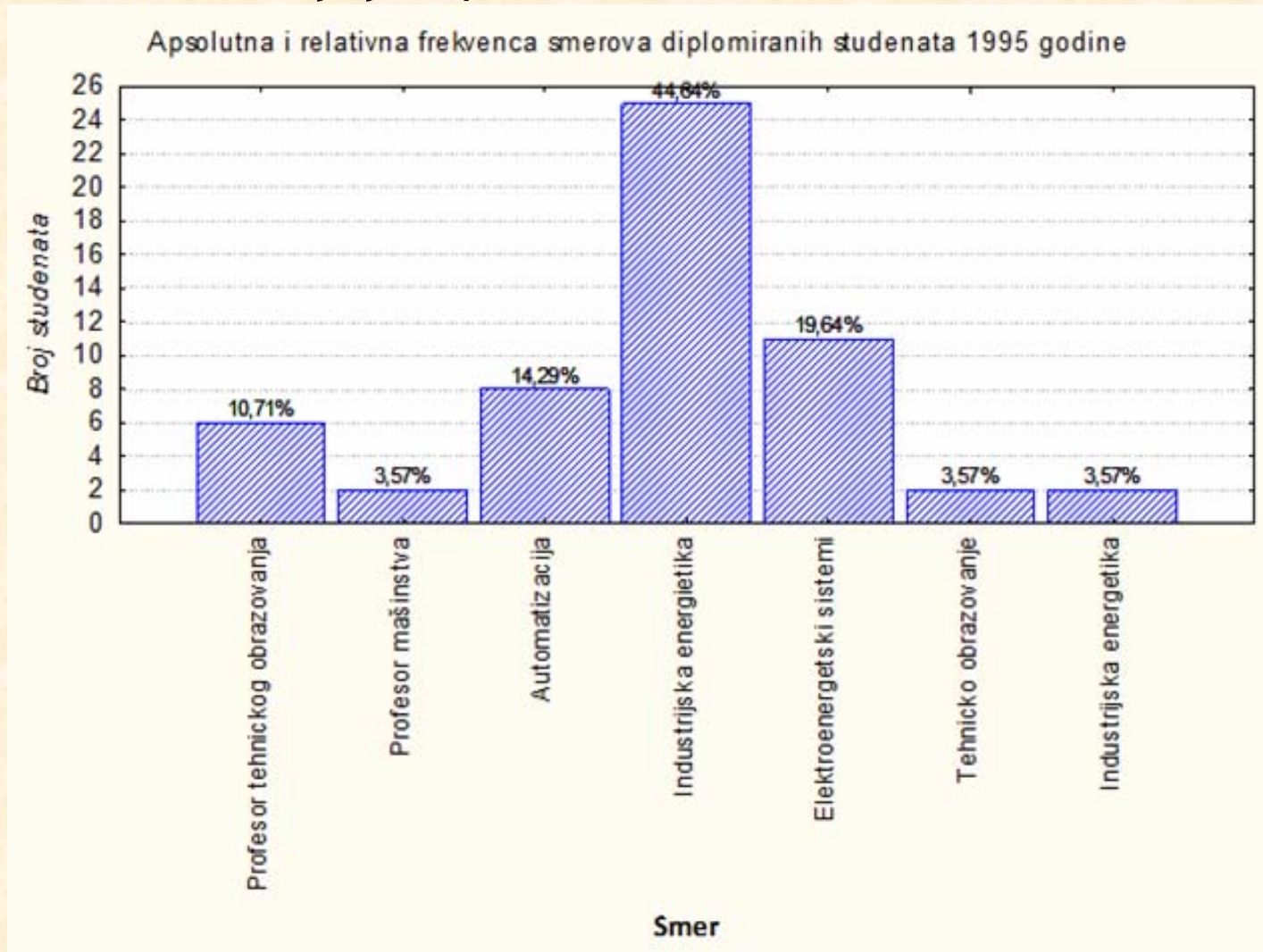
Slika 8. Apsolutna frekvencija smerova 73 studenata koji su diplomirali 1991.

Ako želimo da na dijagramu pokažemo istovremeno apsolutnu i relativnu frekvenciju, dvoklikom na dijagram dobijamo novi prozor, prikazan na Slici 9. U kartici *Point Labels* izaberemo ono što želimo da se vidi na dijagramu, u našem slučaju to su procenti.



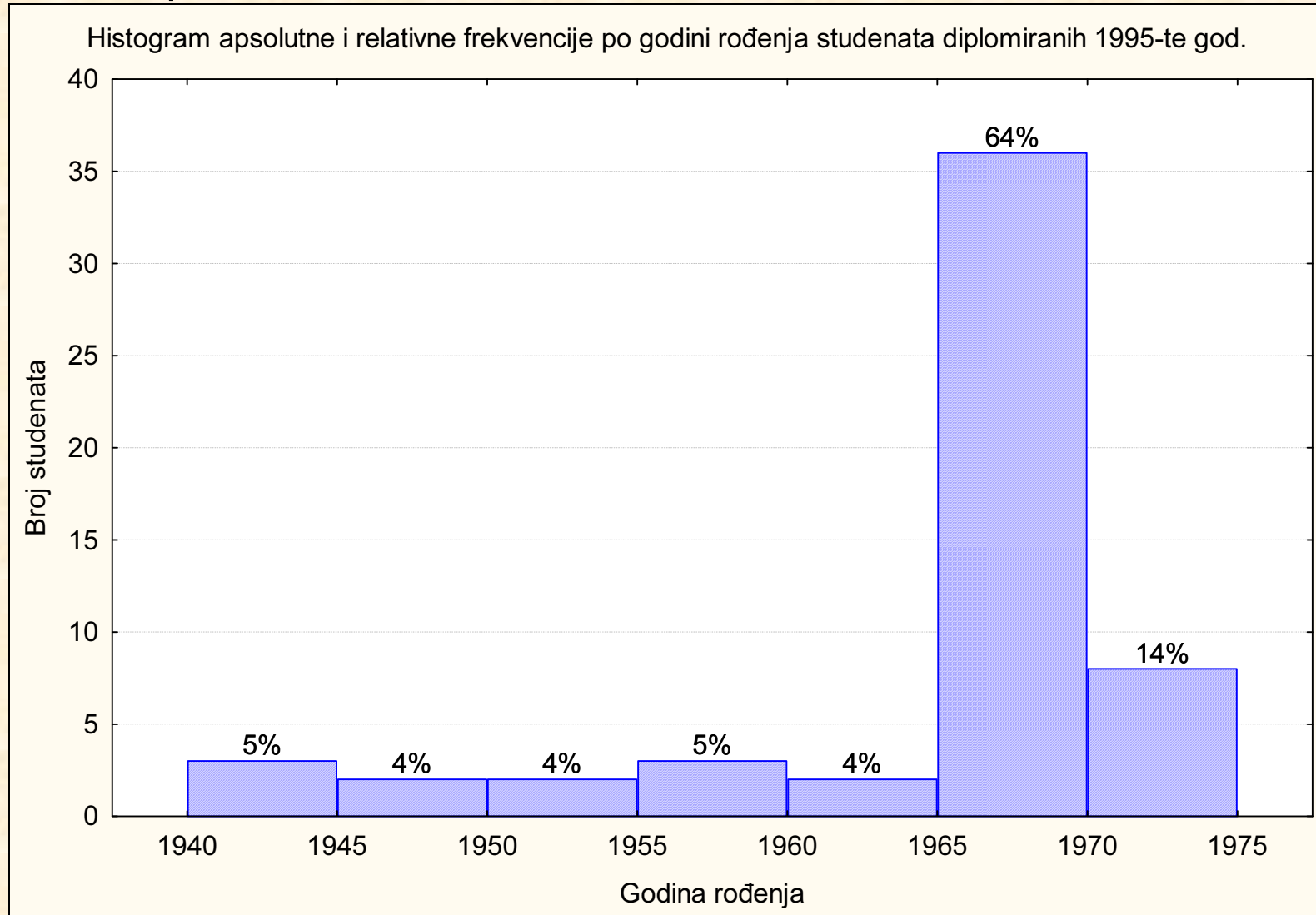
Slika 9. Point Labels

Iz populacije diplomiranih studenata u periodu 1995-1999 izdvajamo generaciju studenata koja je diplomirala 1995. (Slika 10).

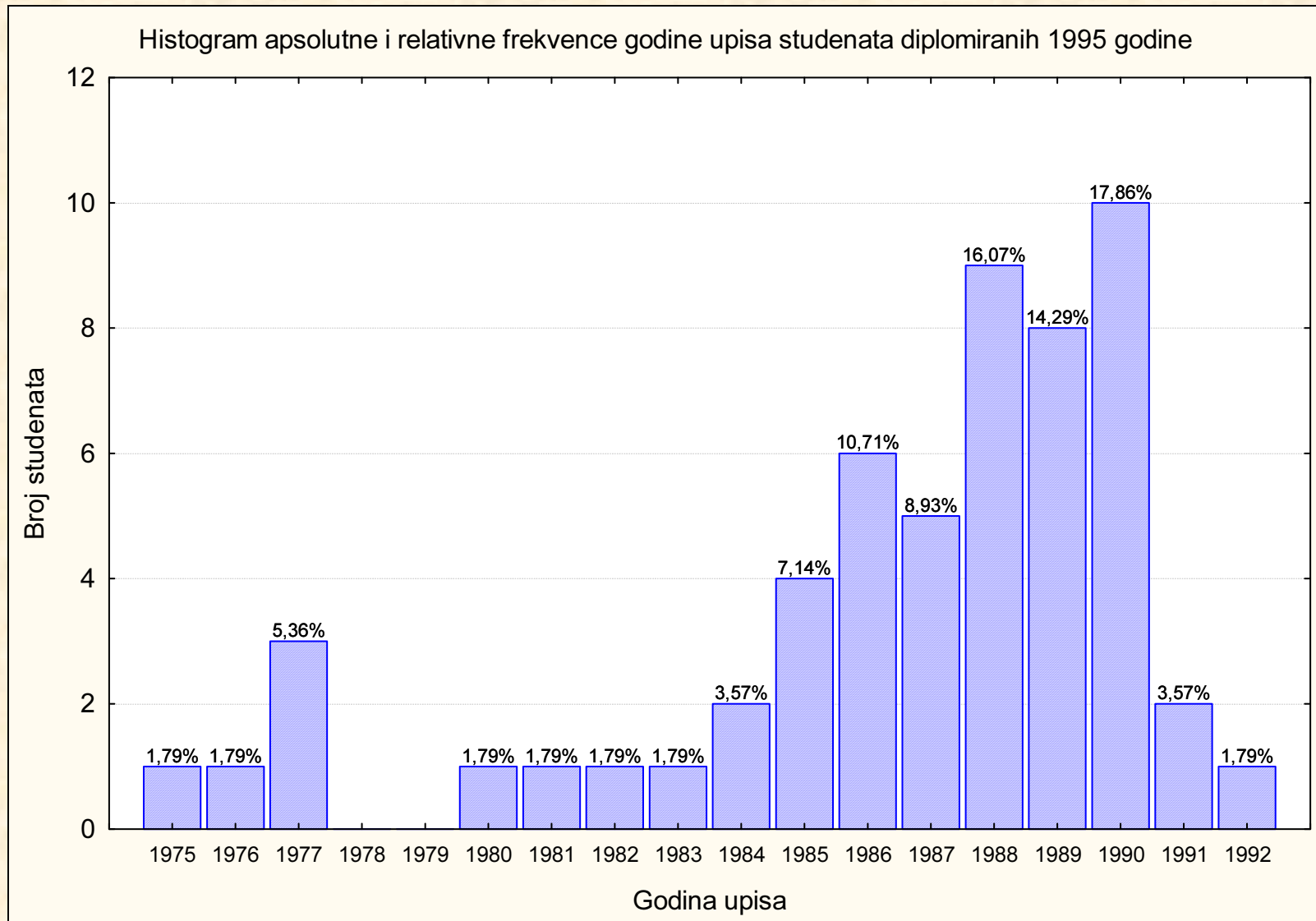


Slika 10. Histogram apsolutne i relativne frekvencija po smerovima 56 diplomiranih studenata 1995. god.

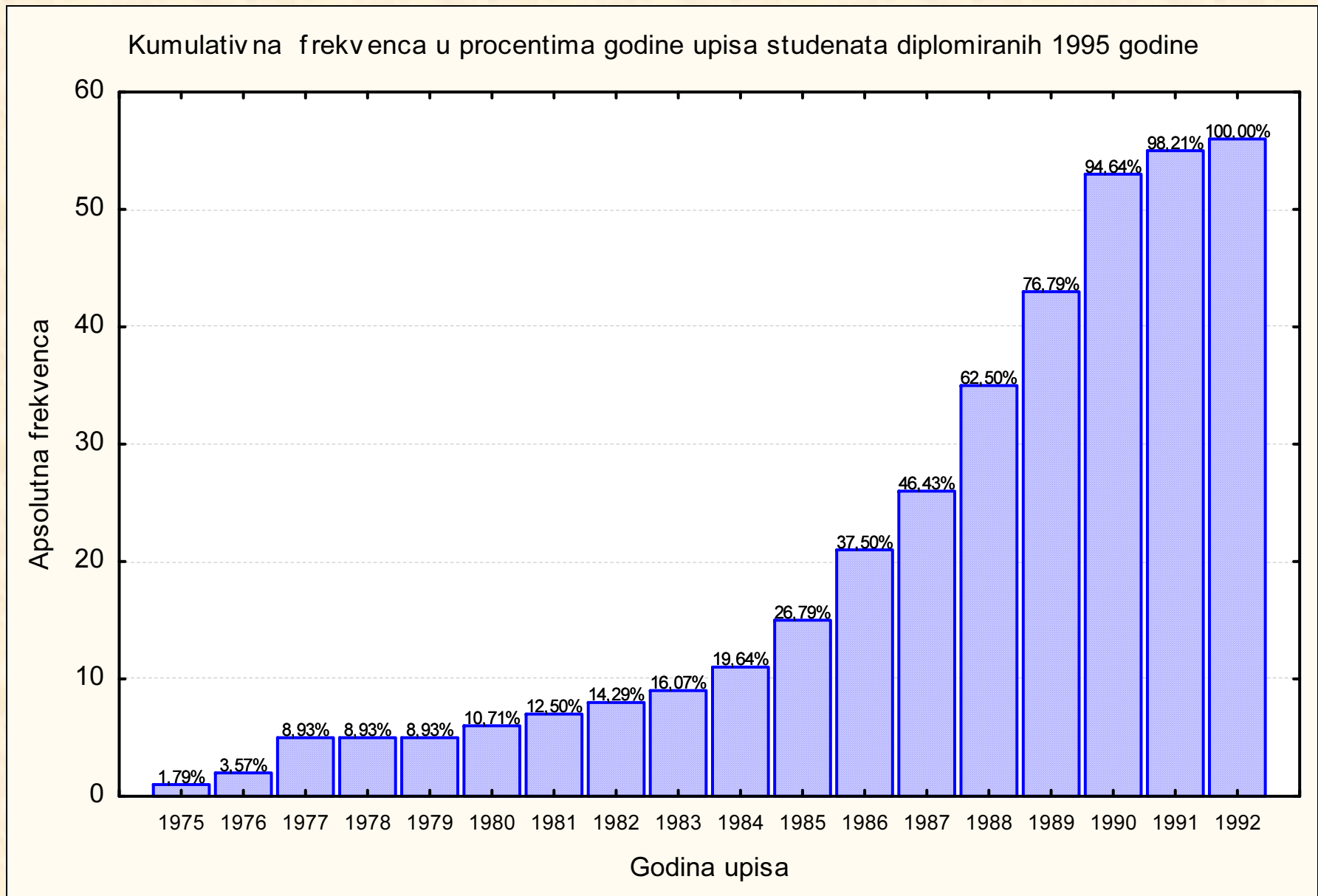
Sve naredne dijagrame dobijamo na sličan način primenom opcija prikazanih na prozorima sa Slika 7. i 9.



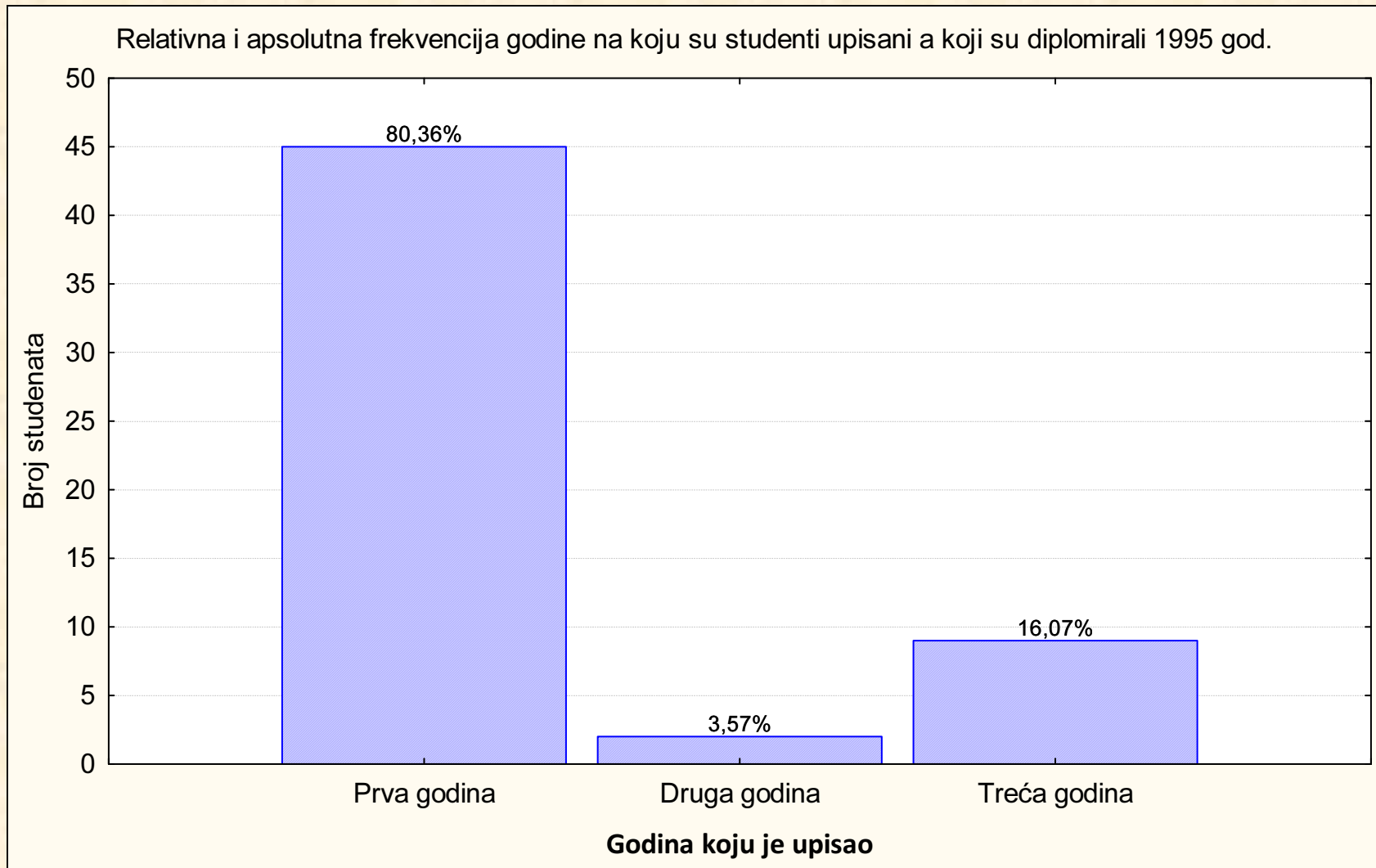
Slika 11. Dijagram godina rođenja studenata koji su diplomirali 1995. god.



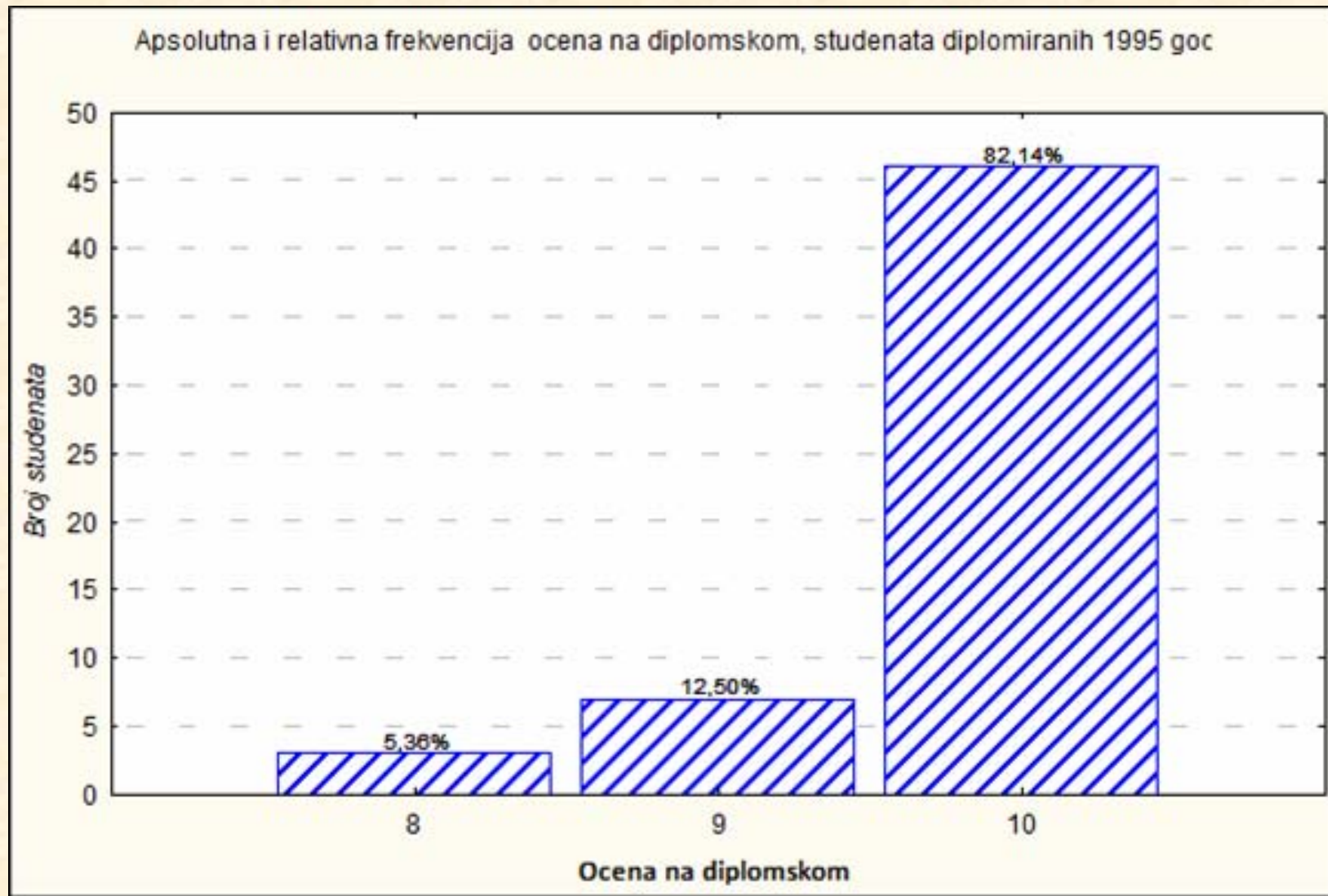
Slika 12. Histogram apsolutne i relativne frekvencije godine upisa studenata diplomiranih 1995. god.



Slika 13. Kumulativna frekvencija godina upisa studenata diplomiranih 1995. god.

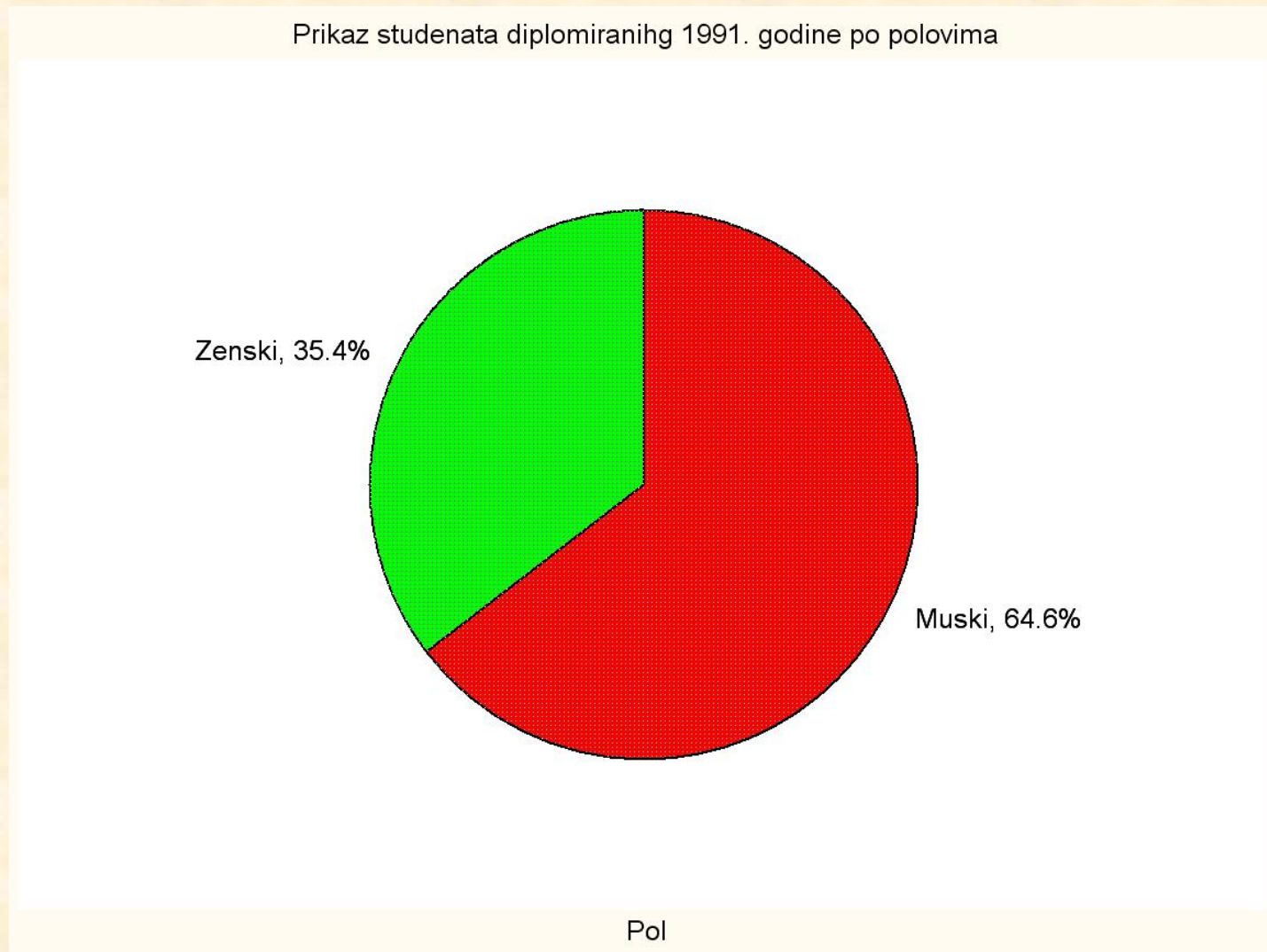


Slika 14. Histogram apsolutne i relativne frekvencije godina na koje su studenti upisani, a koji su diplomirali 1995. god.

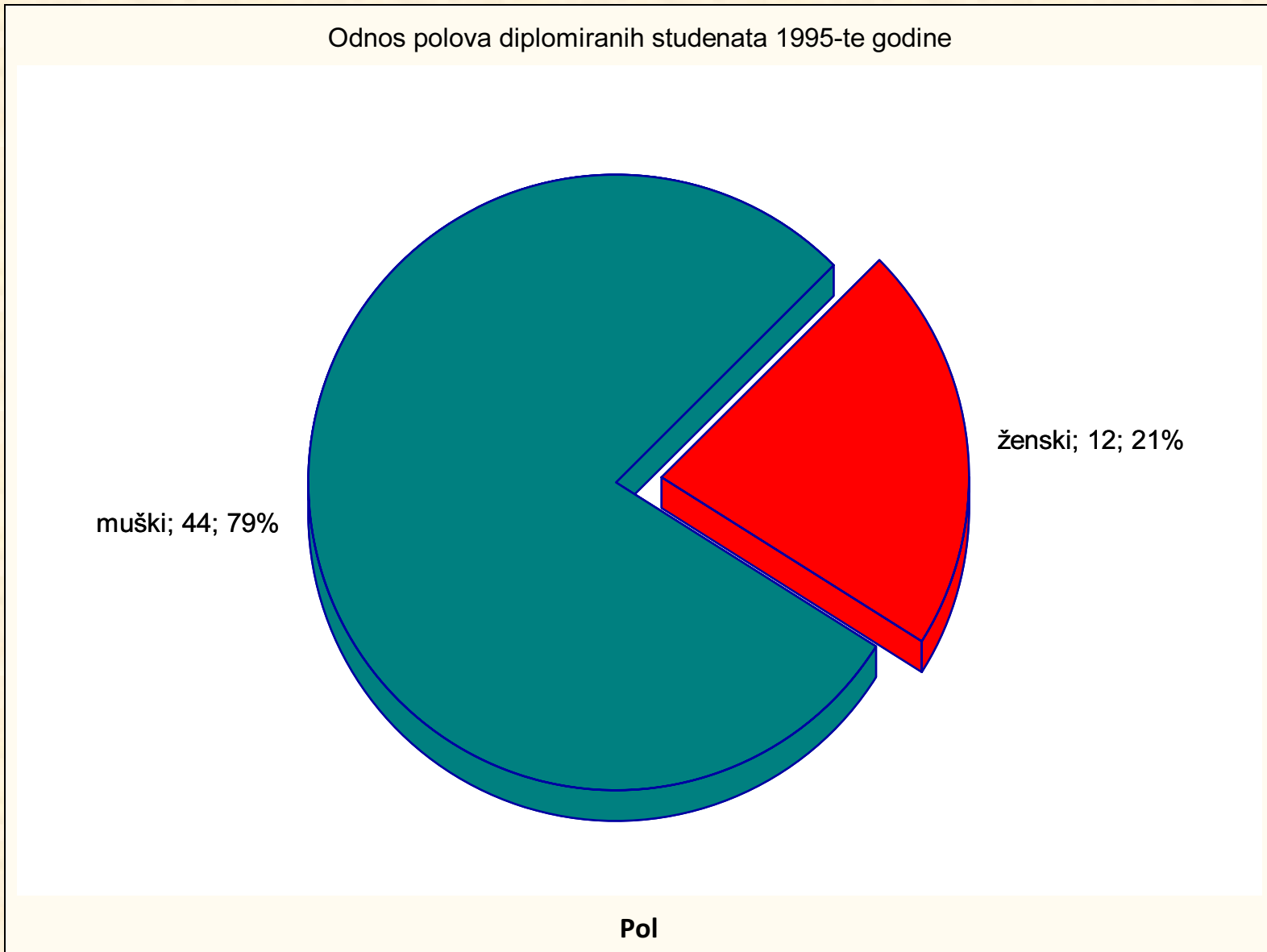


Slika 15. Dijagram ocena na diplomskom, studenata diplomiranih 1995. god.

Iz formirane baze u paketu *Statistika* jednostavnim komandama se dobijaju i kružni dijagrami. Ovi dijagrami pregledno prikazuju odnos polova studenata koji su diplomirali iste godine (Slike 16 i 17).

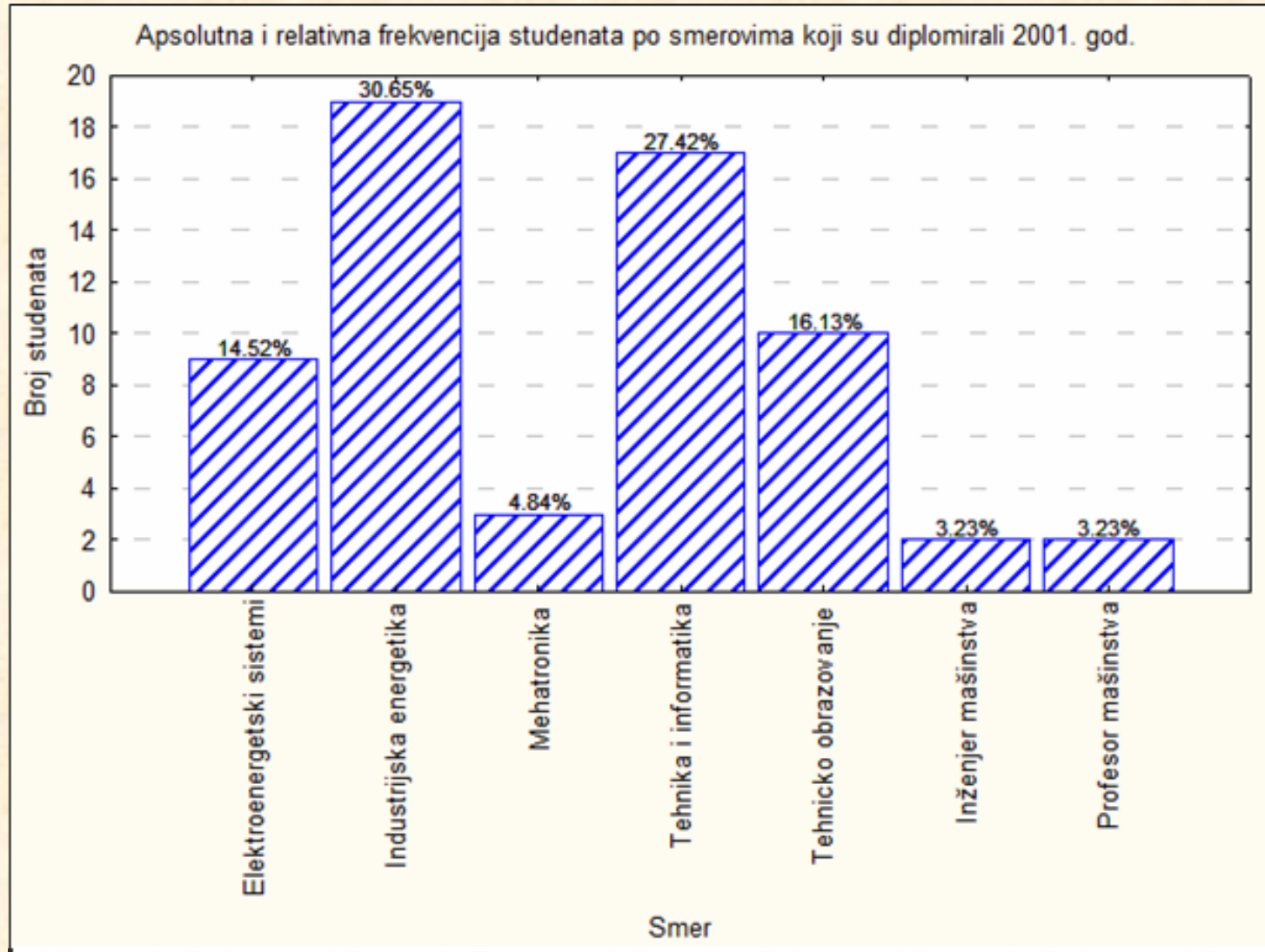


Slika 16.



Slika 17. Prikaz pie chart dijagrama po polovim diplomiranih studenata 1995. god.

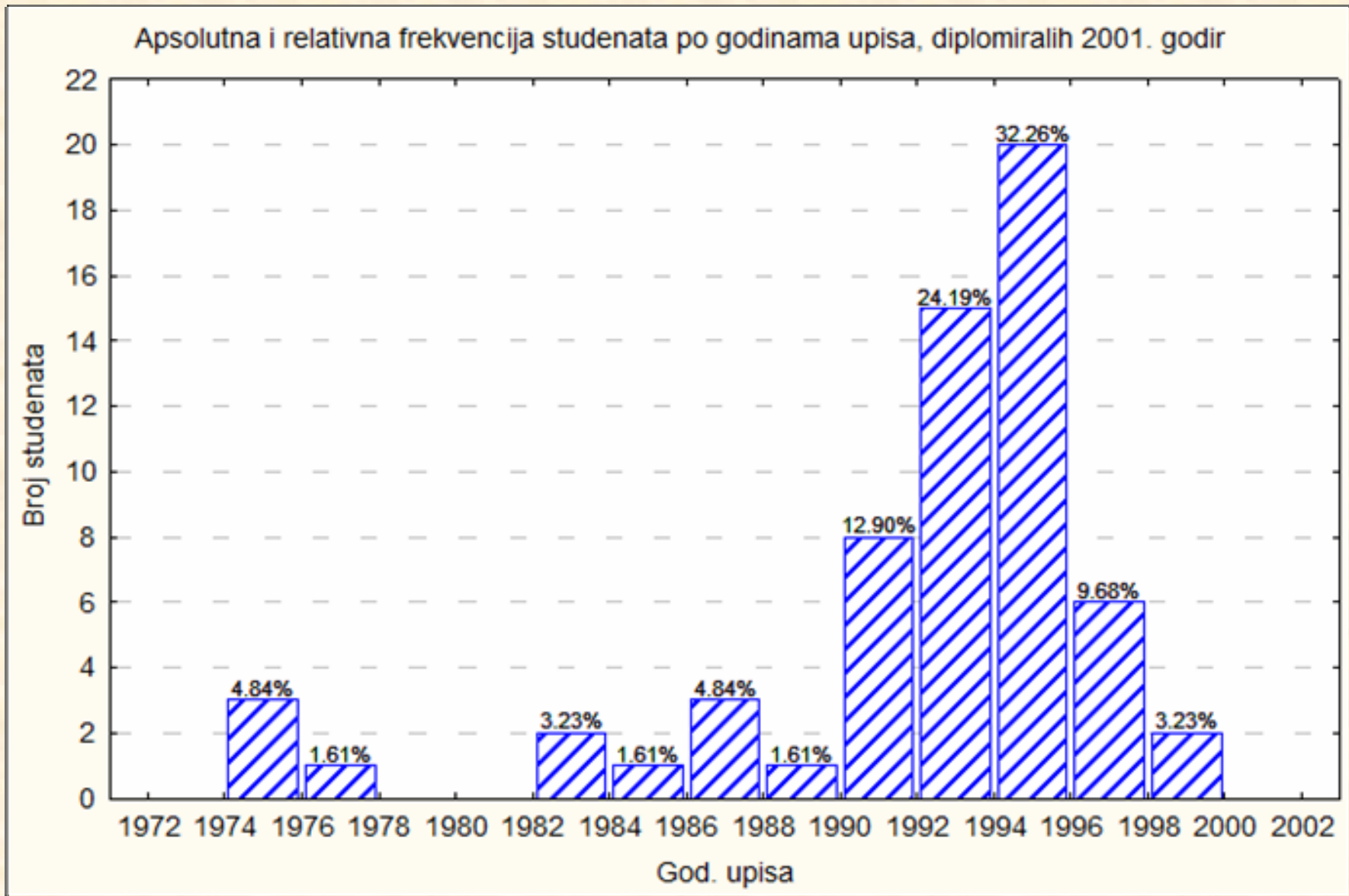
U narednom delu daćemo neke dijagrame vezane za generacije poslednje decenije.



Slika 18. Apsolutna i relativna frekvencija smerova za 62 studenta koji su diplomirali 2001. god.

Na dijagramu možemo videti da je 2001. godine diplomirao najveći broj studenata sa smera Industrijska energetika tj. 30,65 % od ukupnog broja diplomiranih studenata te godine, a najmanji broj studenata je stekao diplomu na smerovima Inženjer mašinstva i Profesor mašinstva.

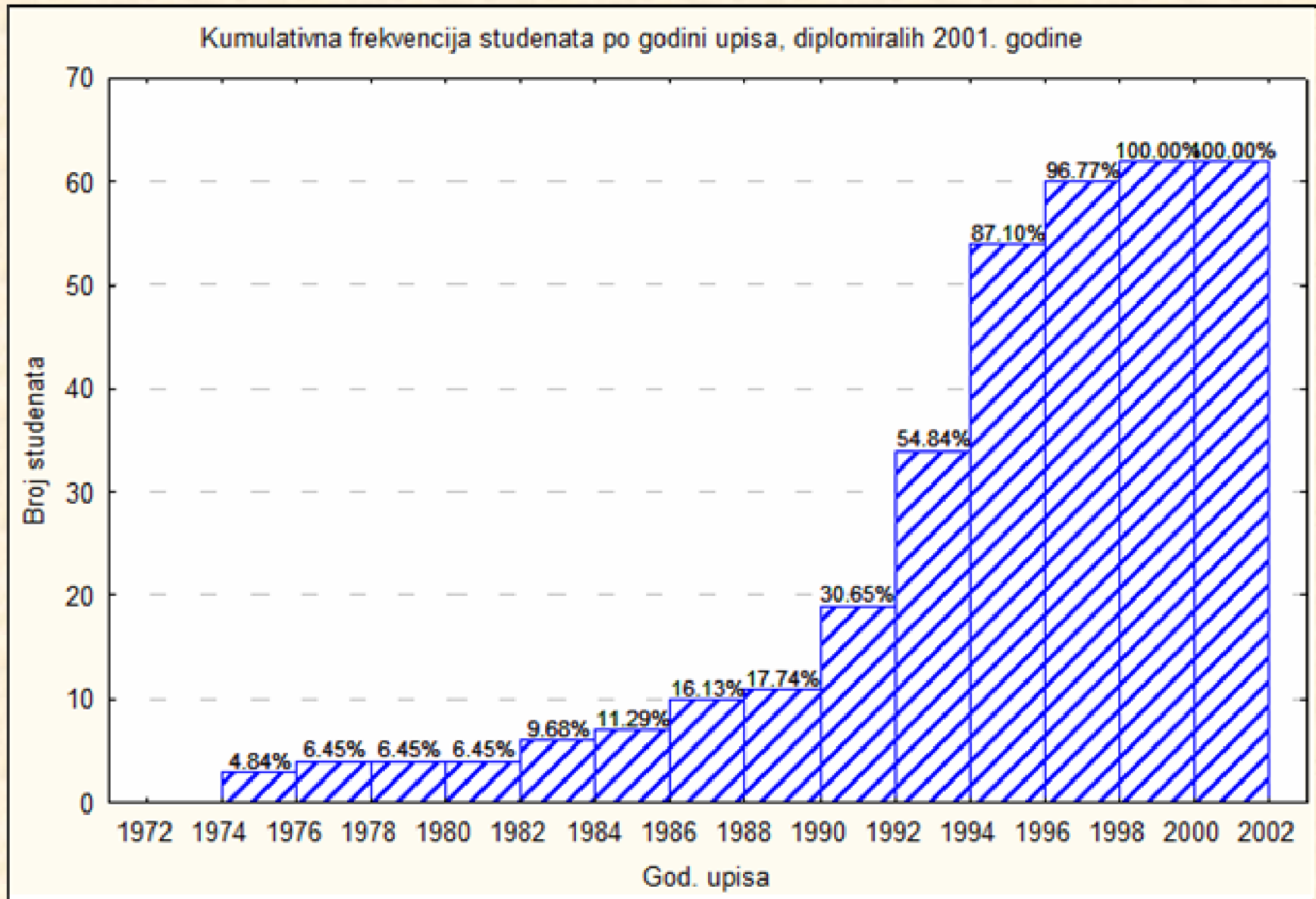
Jedna važna informacija u vezi sa fakultetskim obrazovanjem svakako je i dužina trajanja studija. **Sledeći dijagram** daje informaciju o apsolutnoj i relativnoj frekvenciji godina upisa studenata koji su diplomirali 2001. godine (Slika 19).



Slika 19. Apsolutna i relativna frekvencija godine upisa studenata diplomiranih 2001. god.

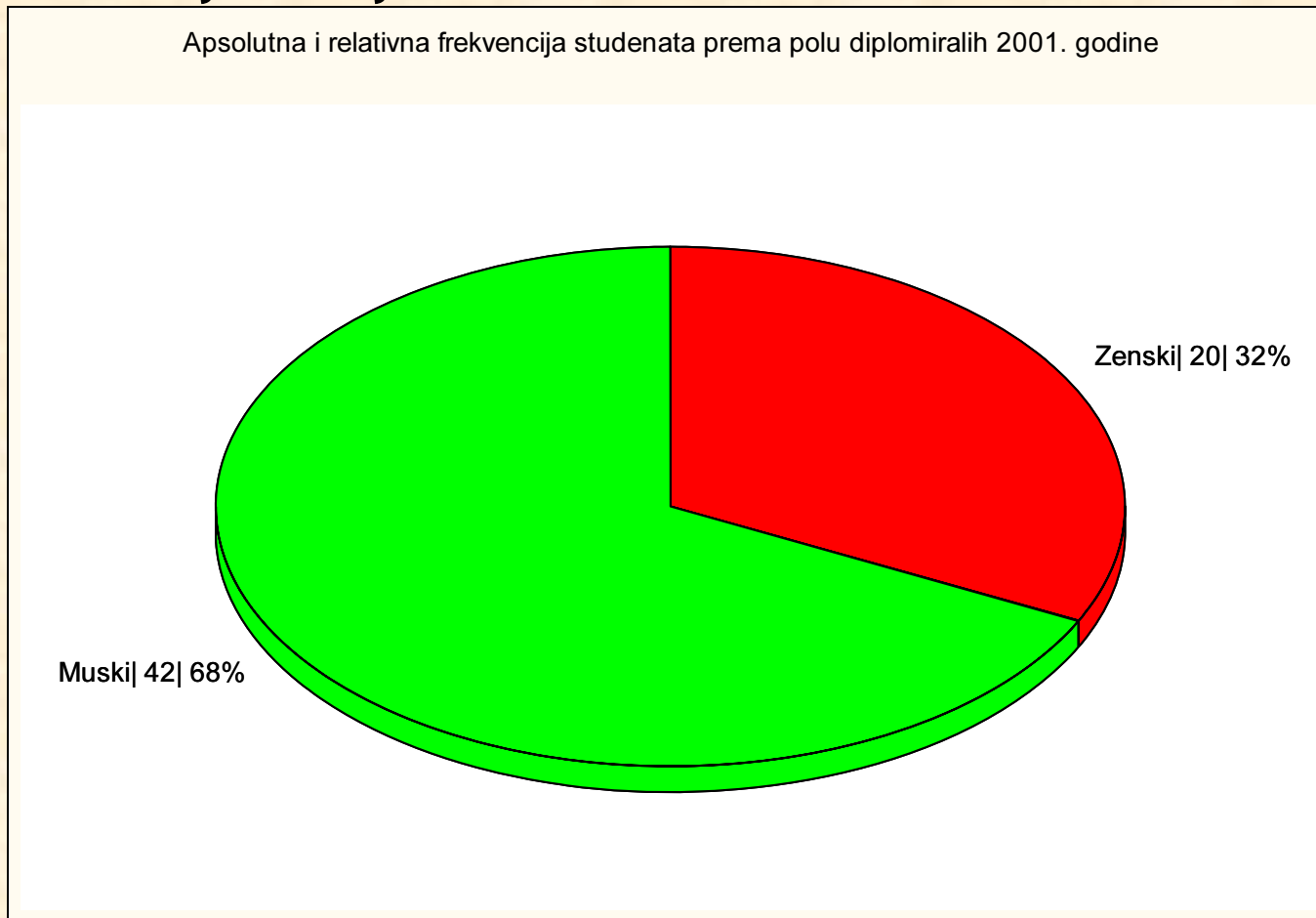
Na ovom dijagramu vidimo da je 2001. godine diplomiralo najviše studenata koji su upisali 1995-1996, i za te studente možemo reći da su diplomirali u roku. Takođe vidimo da postoje i studenti koji su studirali preko 25 godina.

U vezi sa obeležjem – godina upisa na fakultet, iz prethodnog dijagrama može se kreirati i dijagram kumulativne frekvencije koji se dobija sažimanjem relativnih frekvencija (**Slika 20**). Na njemu možemo videti broj/procent studenata koji su upisali studije do, na primer, 1996. godine a diplomirali 2001, što u statističkim istraživanjima može biti veoma koristan podatak.

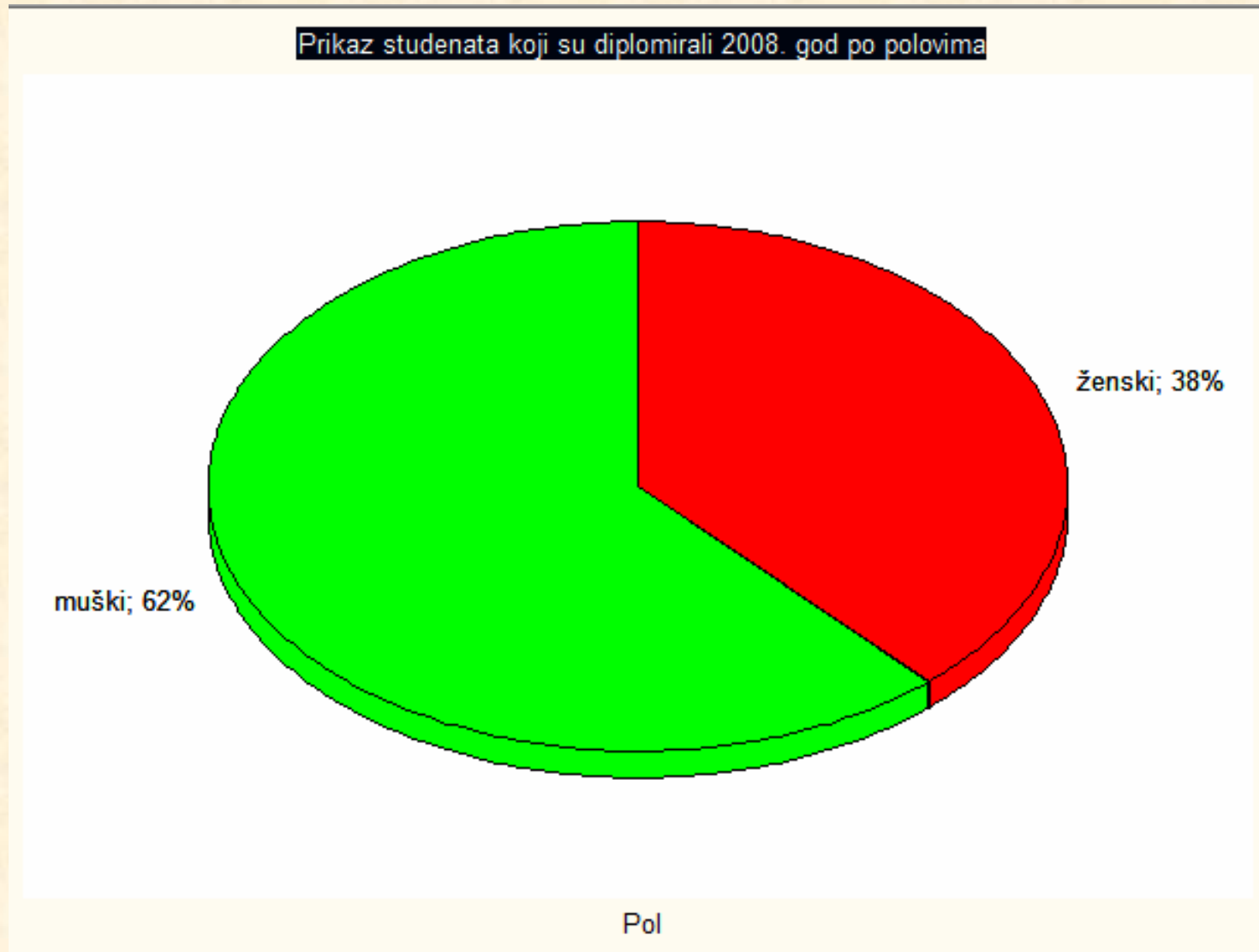


Slika 20. Kumulativna frekvencija godina upisa studenata koji su za diplomirane 2001. god.

Sledi prikaz relativne i apsolutne frekvencije pola studenata diplomiranih 2001. i 2008. godine. Zanimljivo da je i u ovim uzorcima broj muškaraca je mnogo veći od broja devojaka (Slike 21. i 22).

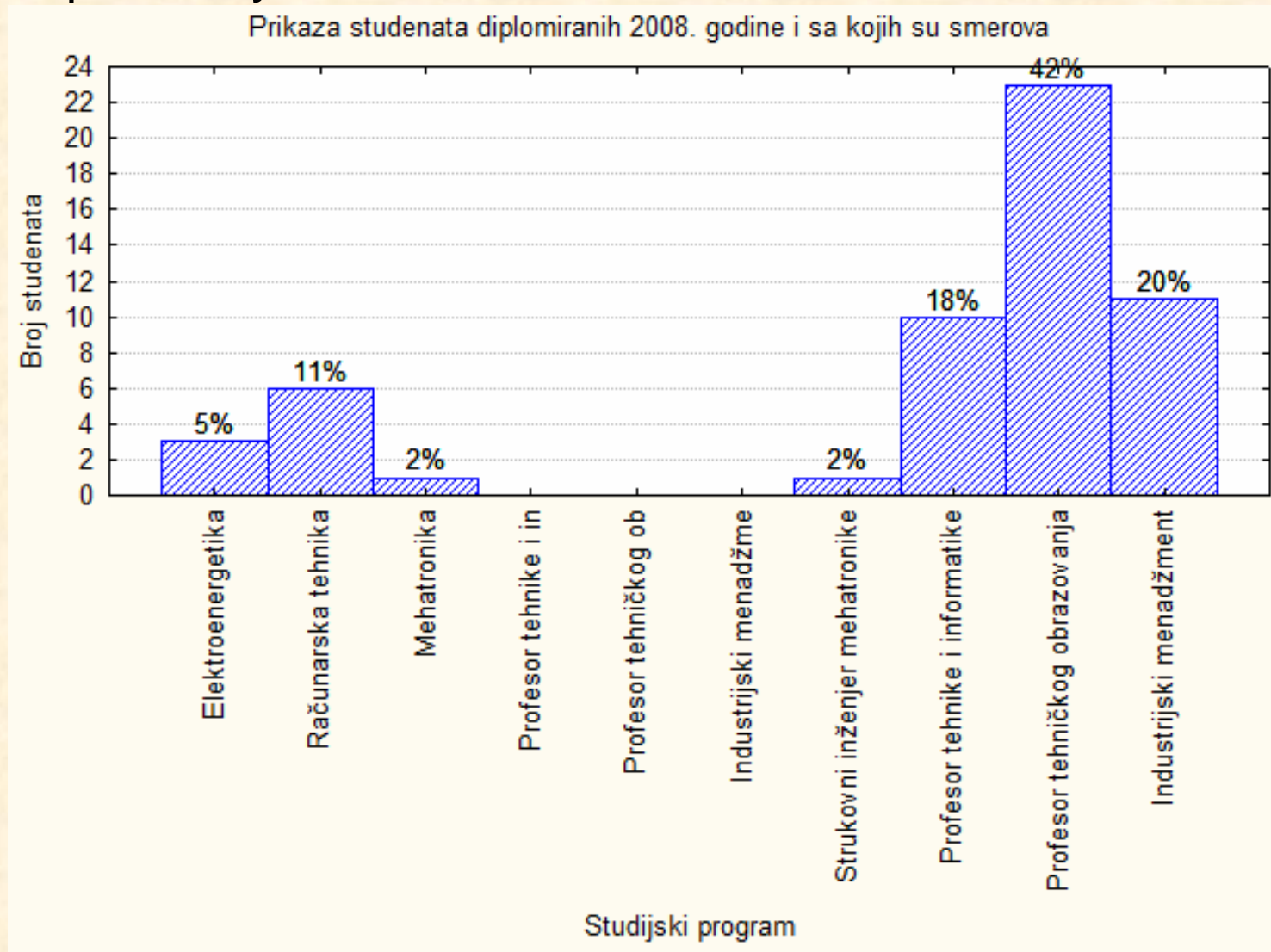


Slika 8. Apsolutna i relativna frekvencija pola studenata.



Slika 22. Odnos polova studenata diplomiranih 2008. god.

Raspored studenata po smerovima (studijskim programima) za diplomirane 2008. god. prikazan je na Slici 23.

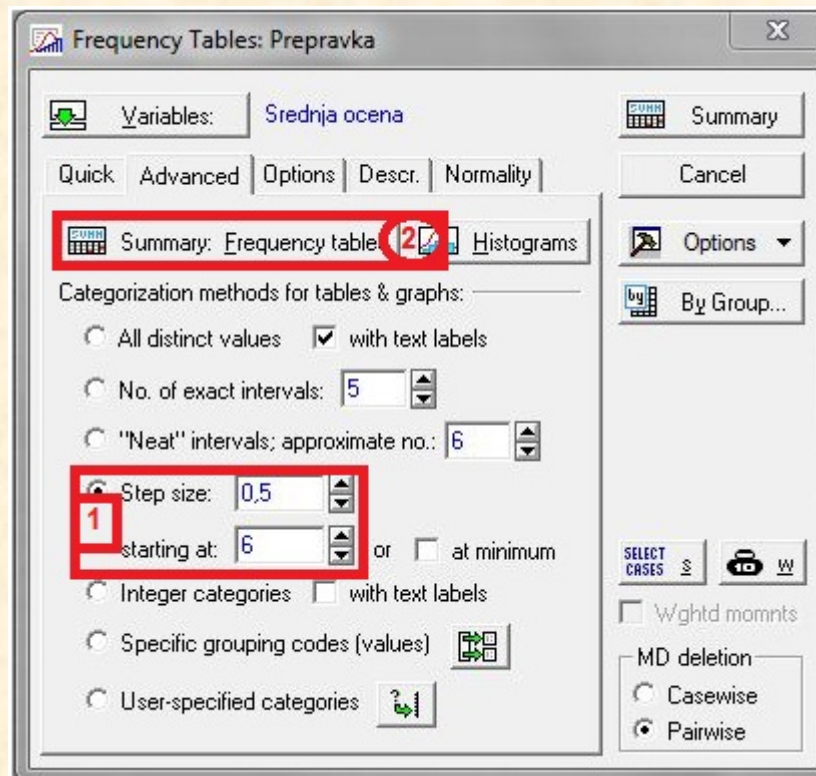


Slika 23.

Za neke statističke serije ponekad je pogodniji prikaz preko poligonalnog dijagrama, tj. poligonalne linije.

Postupak dobijanja poligonalne linije je sledeći:

Selektujemo kolonu "Srednja ocena", zatim idemo na meni "Statistics", pa izabiramo "Basic Statistics/Tabeles". Zatim odabiramo "Frequency Tabeles" i dobijamo sledeći prozor (Slika 24).



Slika 24. Postupak dobijanja tabele pomoću koje crtamo poligonalnu liniju

Broj "1" na slici nam pokazuje da je to prvi korak koji treba da odradimo i služi za određivanje dužine intervala, a zatim rezultate dobijamo klikom na opciju pod brojem "2" posle čega dobijamo sledeću tabelu .

		Frequency table: Srednja ocena			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
6,00	<=x<6,50	2	2	3,57143	3,5714
6,50	<=x<7,00	21	23	37,50000	41,0714
7,00	<=x<7,50	20	43	35,71429	76,7857
7,50	<=x<8,00	5	48	8,92857	85,7143
8,00	<=x<8,50	7	55	12,50000	98,2143
8,50	<=x<9,00	0	55	0,00000	98,2143
9,00	<=x<9,50	1	56	1,78571	100,0000
9,50	<=x<10,00	0	56	0,00000	100,0000

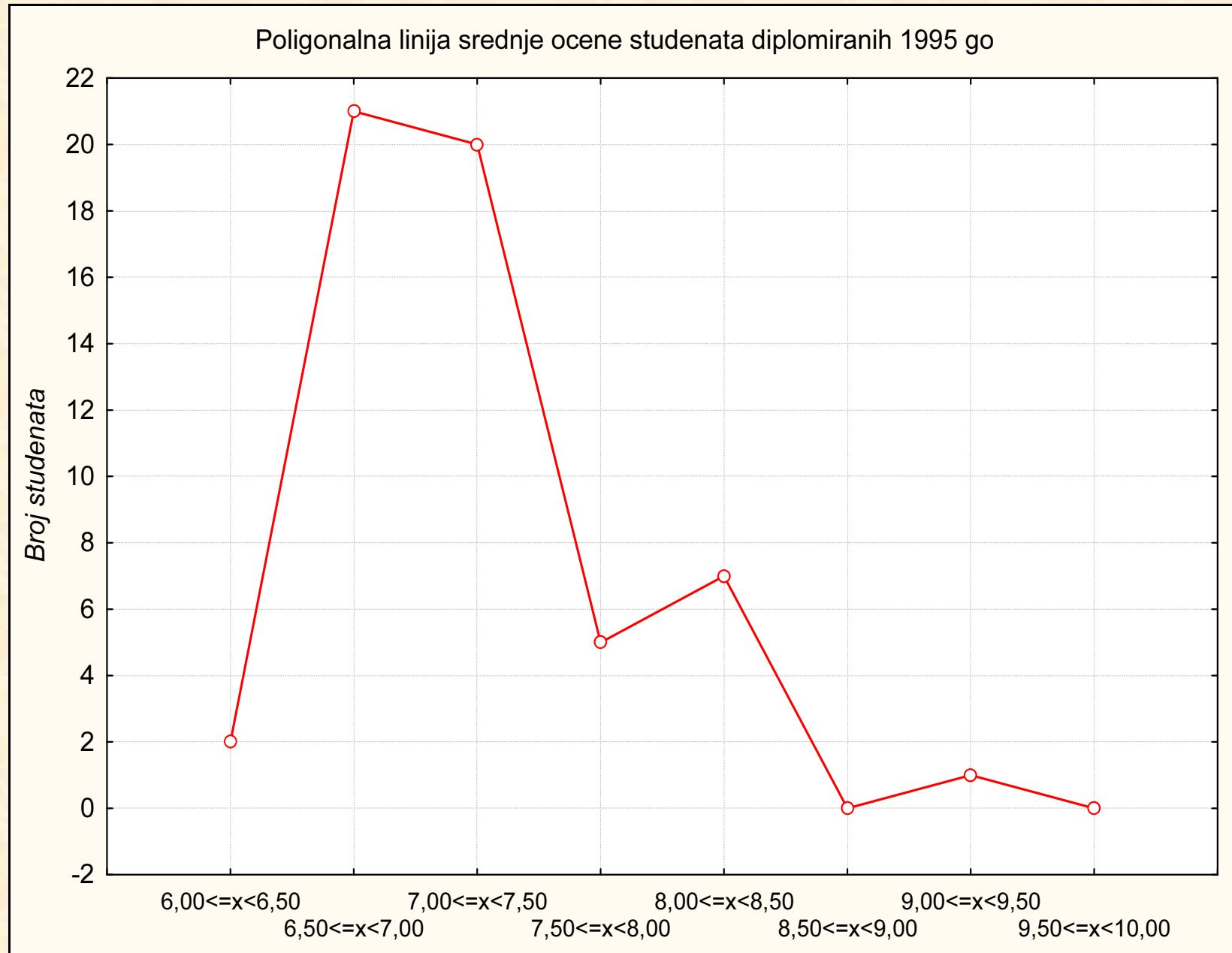
Slika 25. Tablica frekvencije srednje ocene po intervalima.

Konačan izgled Slike 27. dobijamo na način koji je prikazan na Slici 26. Selektujemo kolonu "Count" i kliknemo na desni taster miša.

		Frequency table: Srednja ocena (tabela diplomiranih 1995)			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
6,00	<=x<6,50	2	2	3,57143	3,5714
6,50	<=x<7,00	21	23	37,50000	41,0714
7,00	<=x<7,50	20	43	35,71429	76,7857
7,50	<=x<8,00	5	48	8,92857	85,7143
8,00	<=x<8,50	7	55	12,50000	98,2143
8,50	<=x<9,00	0	55	0,00000	98,2143
9,00	<=x<9,50		55		98,2143
9,50	<=x<10,00		71		100,0000

The image shows a software interface with a data table and an open menu. The menu is titled 'Graphs of Block Data' and contains several options. The option 'Line Plot: Entire Columns' is highlighted with a red oval. Other options include 'Histogram: Block Columns', 'Histogram: Entire Columns', 'Line Plot: Block Rows', 'Box Plot: Block Columns', 'Normal Prob. Plot: Block Columns', and several 'Custom Graph' options.

Slika 26. Postupak za dobijanje poligonalne linije



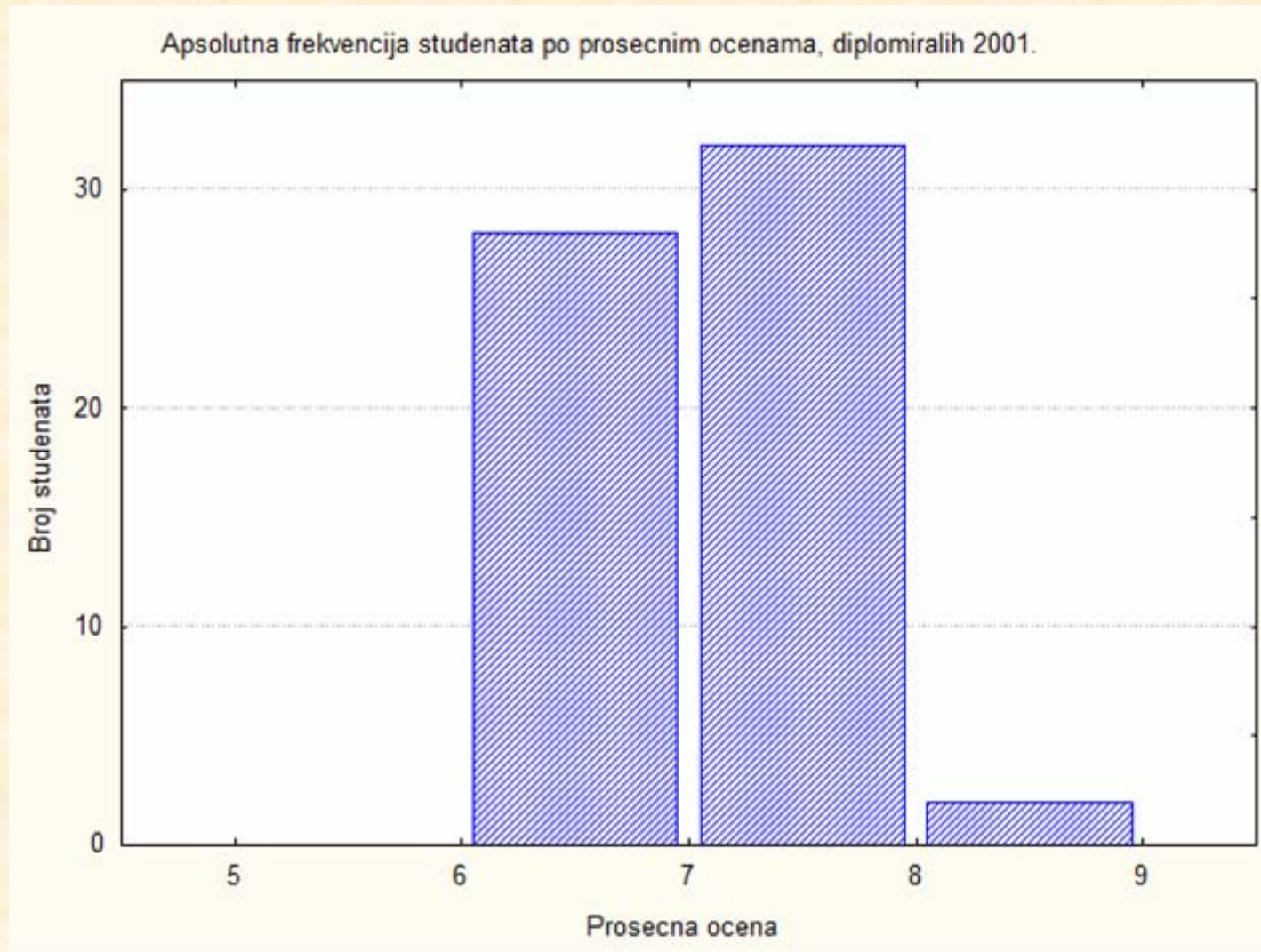
Slika 27. Poligonalna linija srednje ocene studenata diplomiranih 1995. god.

Program nam nudi više opcija za kreiranje intervala. Mi zadajemo korak intervala i pocetak prvog intervala, a zatim klikom na *Summary Frequency tables* dobijamo tablicu sa intervalima (Slika 28).

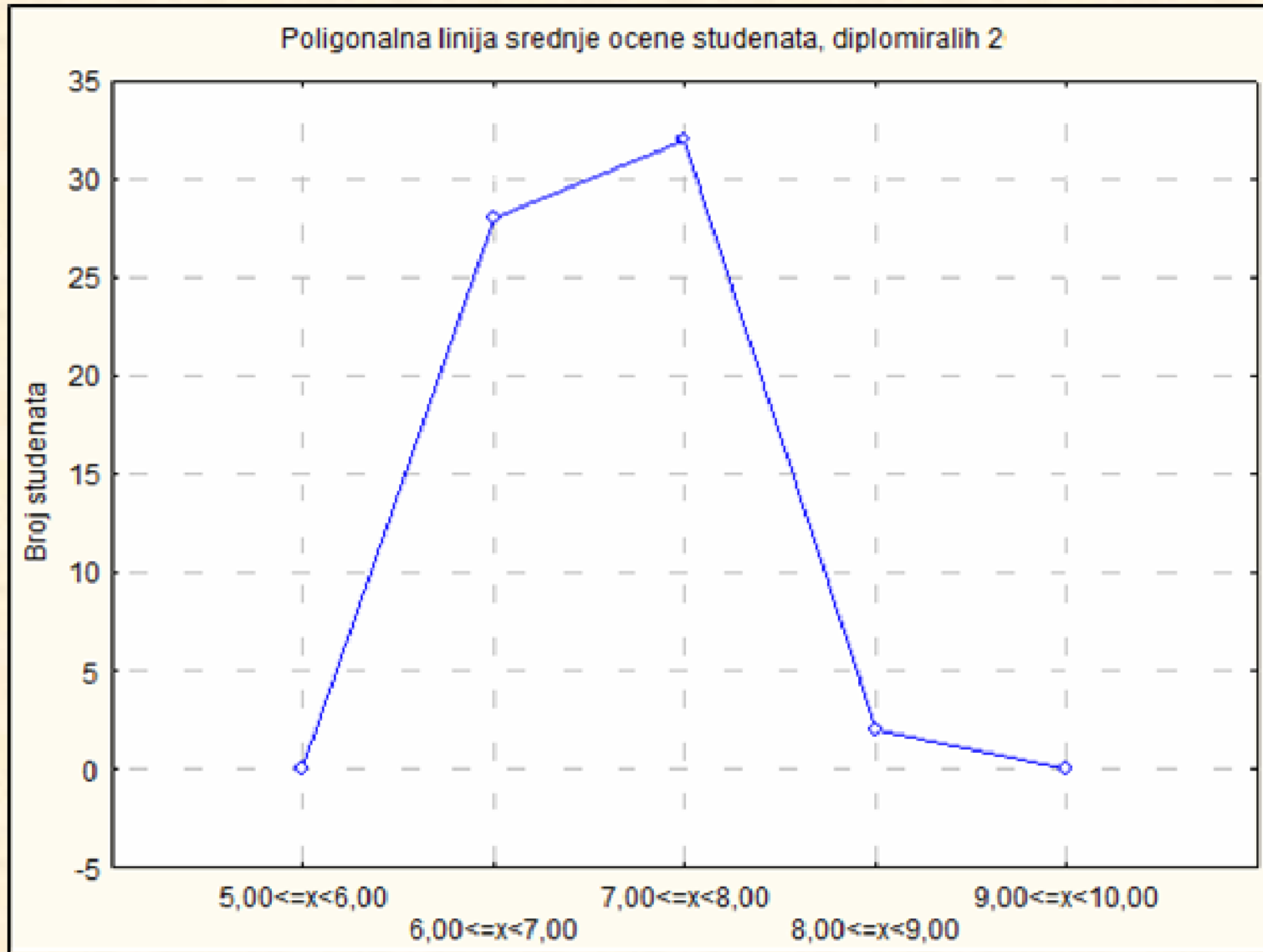
		Frequency table: Prosek (Diplomirali 2001)			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
5,00	<=x<6,00	0	0	0.00000	0.0000
6,00	<=x<7,00	28	28	45.16129	45.1613
7,00	<=x<8,00	32	60	51.61290	96.7742
8,00	<=x<9,00	2	62	3.22581	100.0000
9,00	<=x<10,00	0	62	0.00000	100.0000

Slika 28. Frequency table: Prosek (diplomirali 2001.)

Na osnovu dobijene tabele možemo kreirati histogram apsolutne frekvencije, kao i apsolutnu frekvenciju preko poligonalne linije (Slike 29. i 30).



Slika 29. Apsolutna frekvencija prosečne ocene studenata diplomiranih 2001. god.



Slika 30. Poligonalna linija srednje ocene studenata, diplomiranih 2001. god.

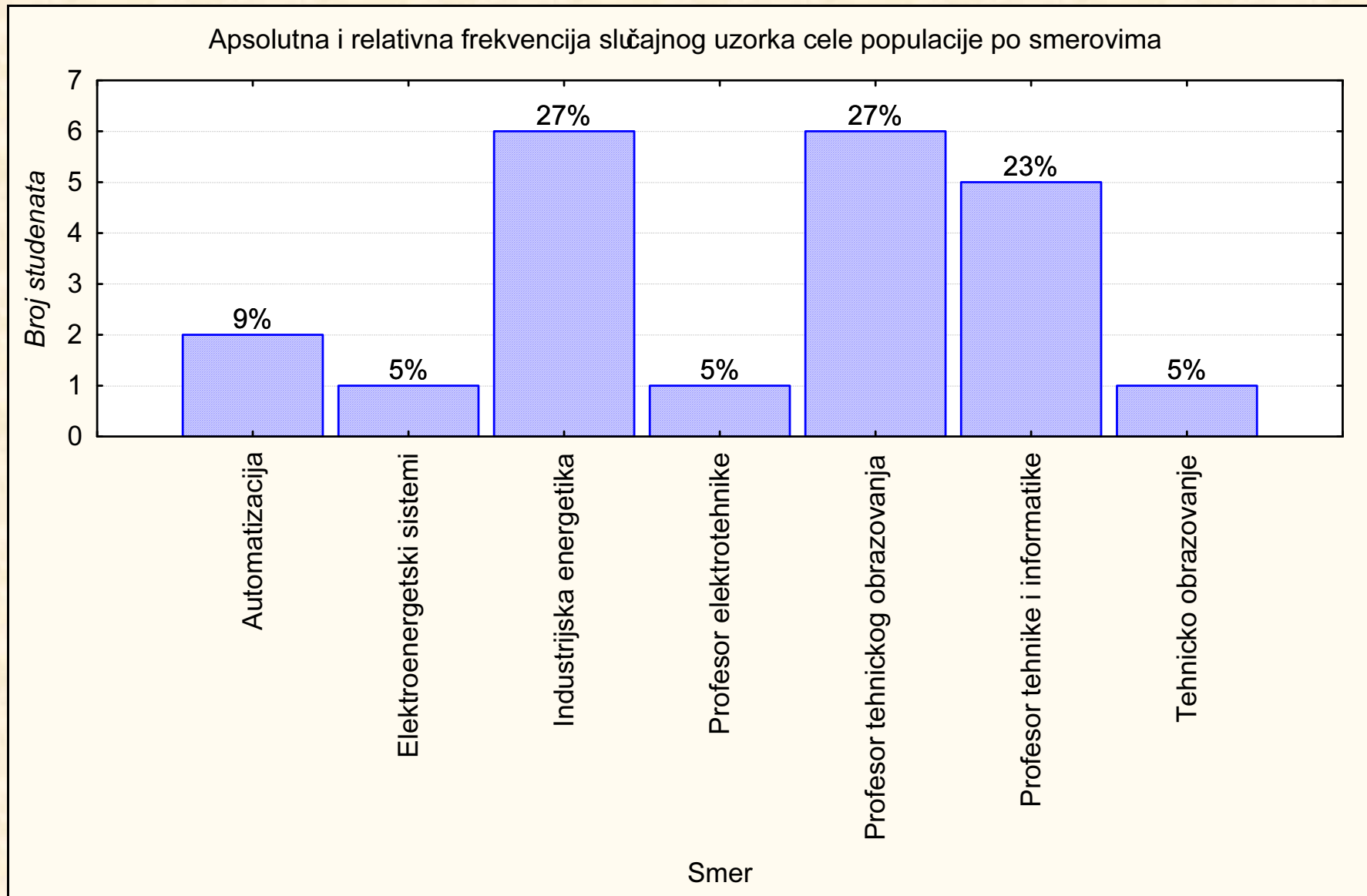
UZORCI IZABRANI IZ POPULACIJE PRIMENOM OBJEKTIVNOG KRITERIJUMA

Izbor uzorka iz populacije po objektivnom kriterijumu znači izbor po strogo utvrđenoj proceduri tako da svaki elemenat populacije ima podjednaku mogućnost da bude uključen u uzorak. Pri primeni paketa *Statistica* taj izbor se vrši zadavanjem neke procentualne vrednosti. Na primer, za populaciju diplomiranih studenata od 1995-1999. god. koja sadrži 504 studenta uzeli smo 4% te populacije i dobili slučajni uzorak od 22 studenta prikazan na Slici 31.

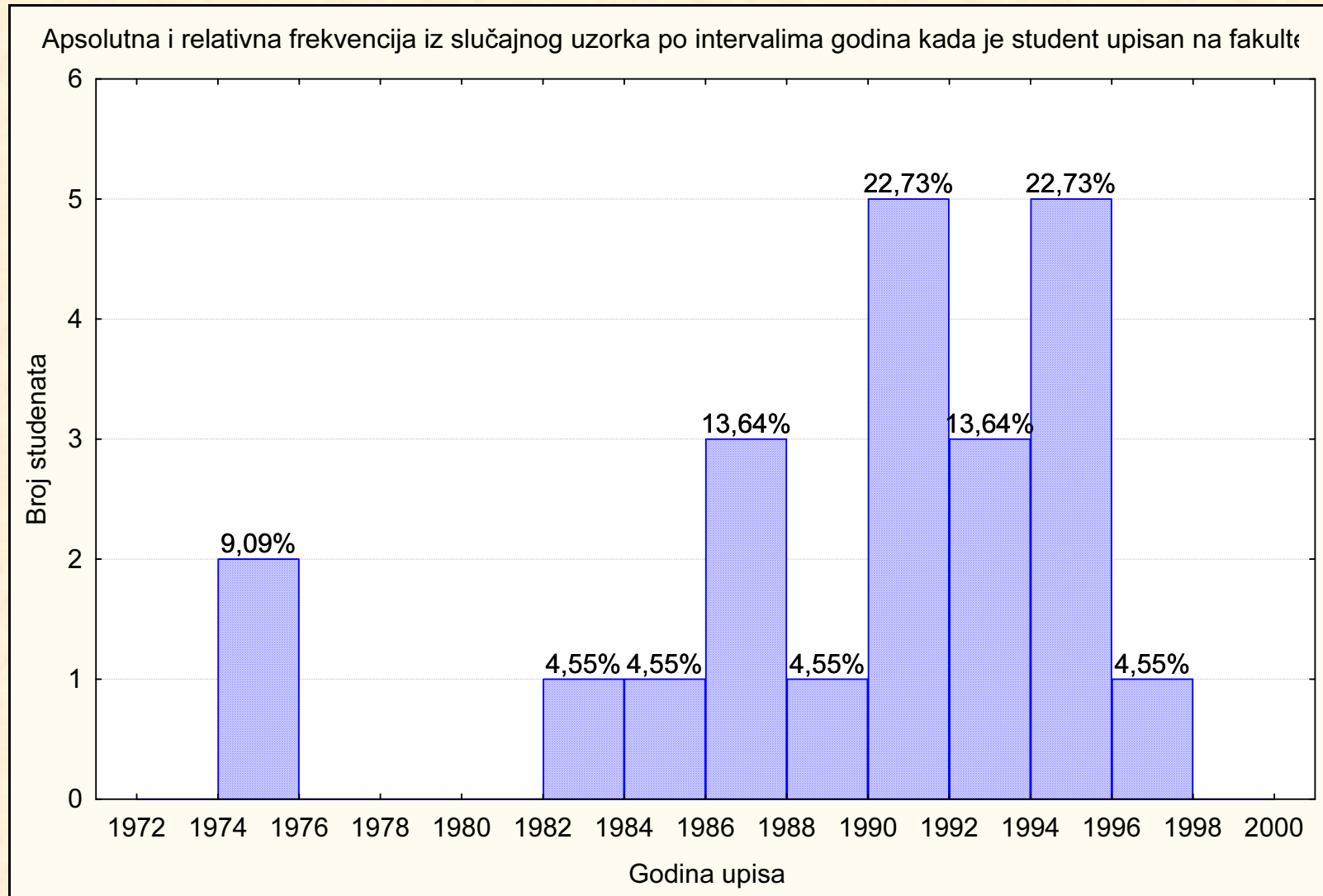
	1 Ime i prezime	2 Opština rođenja	3 Godina upisa	4 Godina diplomiranja	5 Srednja ocena	6 Ocena na diplomskom	7 Pol	8 Godina koju je upisao	9 Godina rođenja	10 Smer
1	Nikolic Dobrinka	Užice	1983	1995	6,69	10	Muški	1	1944	Automatizacija
2	Vojinovic Relja	Sjenica	1985	1996	6,26	10	Muški	1	1955	Automatizacija
3	Zelenovic Dejan	Požega	1990	1998	6,62	10	Muški	1	1971	Elektroenergetski sistemi
4	Dedic Luka	Auterah(Austrija)	1988	1995	6,89	10	Muški	1	1968	Industrijska energetika
5	Racic Dejan	Kruševac	1988	1996	8,05	10	Muški	1	1969	Industrijska energetika
6	Mirosavljaevic Milada	Požega	1991	1997	7,6	10	Ženski	1	1972	Industrijska energetika
7	Jevremovic Goran	Cacak	1992	1997	8,8	10	Muški	1	1973	Industrijska energetika
8	Cvijovic Milomir	Užice	1991	1998	7,09	10	Muški	1	1972	Industrijska energetika
9	Terzic Milosav	Užice	1991	1999	6,97	10	Muški	1	1972	Industrijska energetika
10	Spasojevic Mirjana	Lucani	1975	1997	6,96	9	Ženski	1	1957	Profesor elektrotehnike
11	Tanovic Dragan	Lucani	1976	1995	7,03	10	Muški	1	1958	Profesor tehnickog obrazovanja
12	Balj Snežana	Backi Petrovac	1994	1996	6,61	10	Ženski	3	1961	Profesor tehnickog obrazovanja
13	Bijanac Milovan	Prokuplje	1994	1997	7,05	10	Muški	3	1943	Profesor tehnickog obrazovanja
14	Jovcic Mirjana	Trstenik	1992	1997	6,8	9	Ženski	1	1973	Profesor tehnickog obrazovanja
15	Todorovic Ana	Kragujevac	1995	1997	7,47	10	Ženski	3	1971	Profesor tehnickog obrazovanja
16	Dumic Danijela	Kragujevac	1995	1998	7,66	9	Ženski	3	1967	Profesor tehnickog obrazovanja
17	Todoric Dragan	Kragujevac	1995	1998	7,22	10	Muški	3	1969	Profesor tehnike i informatike
18	Glidžic Sladana	Kruševac	1994	1998	8,05	10	Ženski	1	1975	Profesor tehnike i informatike
19	Jonic Sladana	Aleksinac	1996	1999	6,95	8	Ženski	3	1974	Profesor tehnike i informatike
20	Sreckovic Dragiša	Kragujevac	1997	1999	7,42	10	Muški	3	1968	Profesor tehnike i informatike
21	Novakovic Boban	Savski venac	1996	1999	7,5	10	Muški	3	1970	Profesor tehnike i informatike
22	Kolašinac Leki	Nova Varoš	1988	1996	8,41	9	Muški	1	1969	Tehnicko obrazovanje

Slika 31. Primer dobijenog slučajnog uzorka od 4% iz populacije diplomiranih studenata od 1995–1999. god.

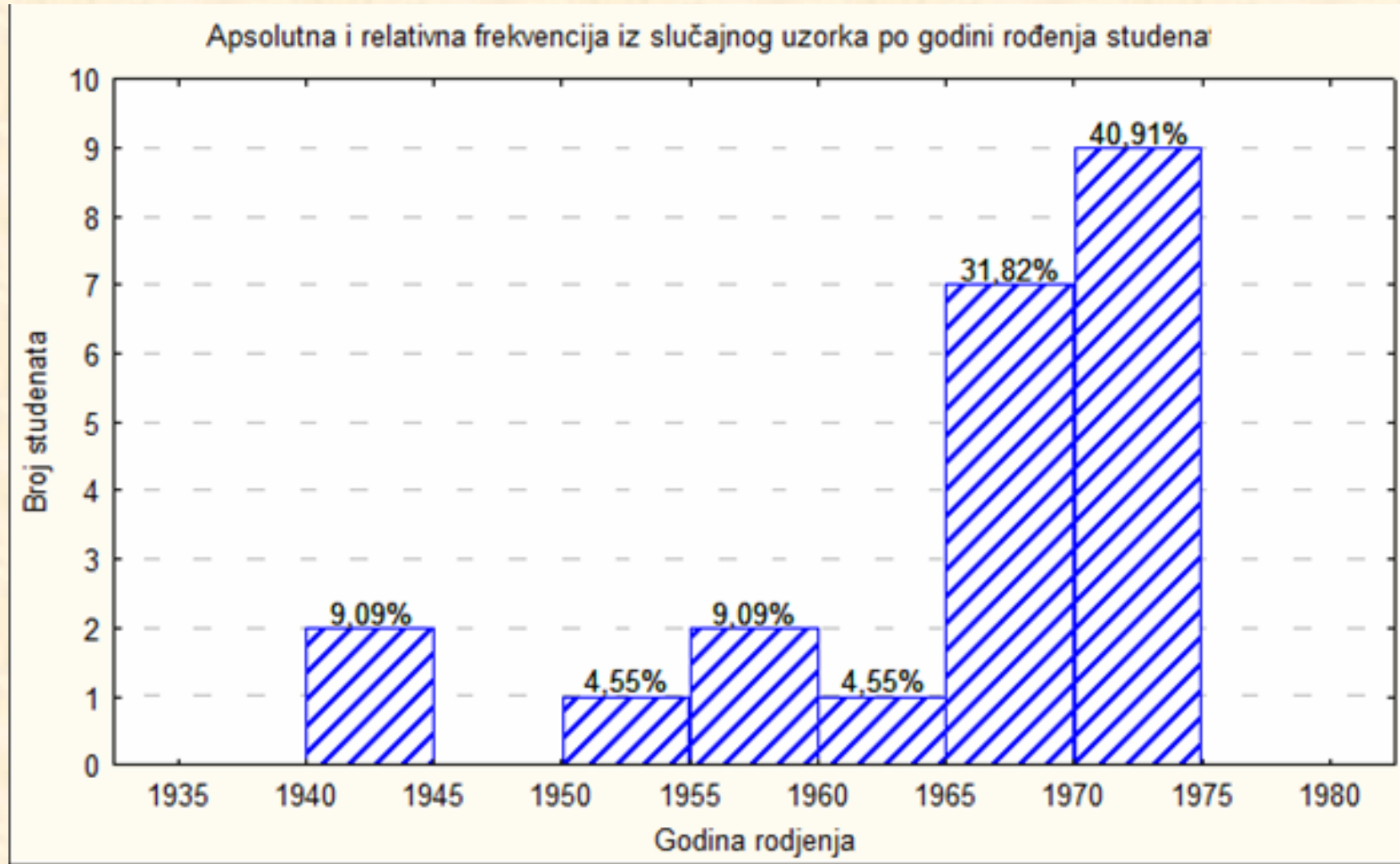
Iz dobijene tabele prikazane na Slici 31. određujemo željene dijagrame za svako od obeležja (varijabli) koje smo pratili na čitavoj populaciji.



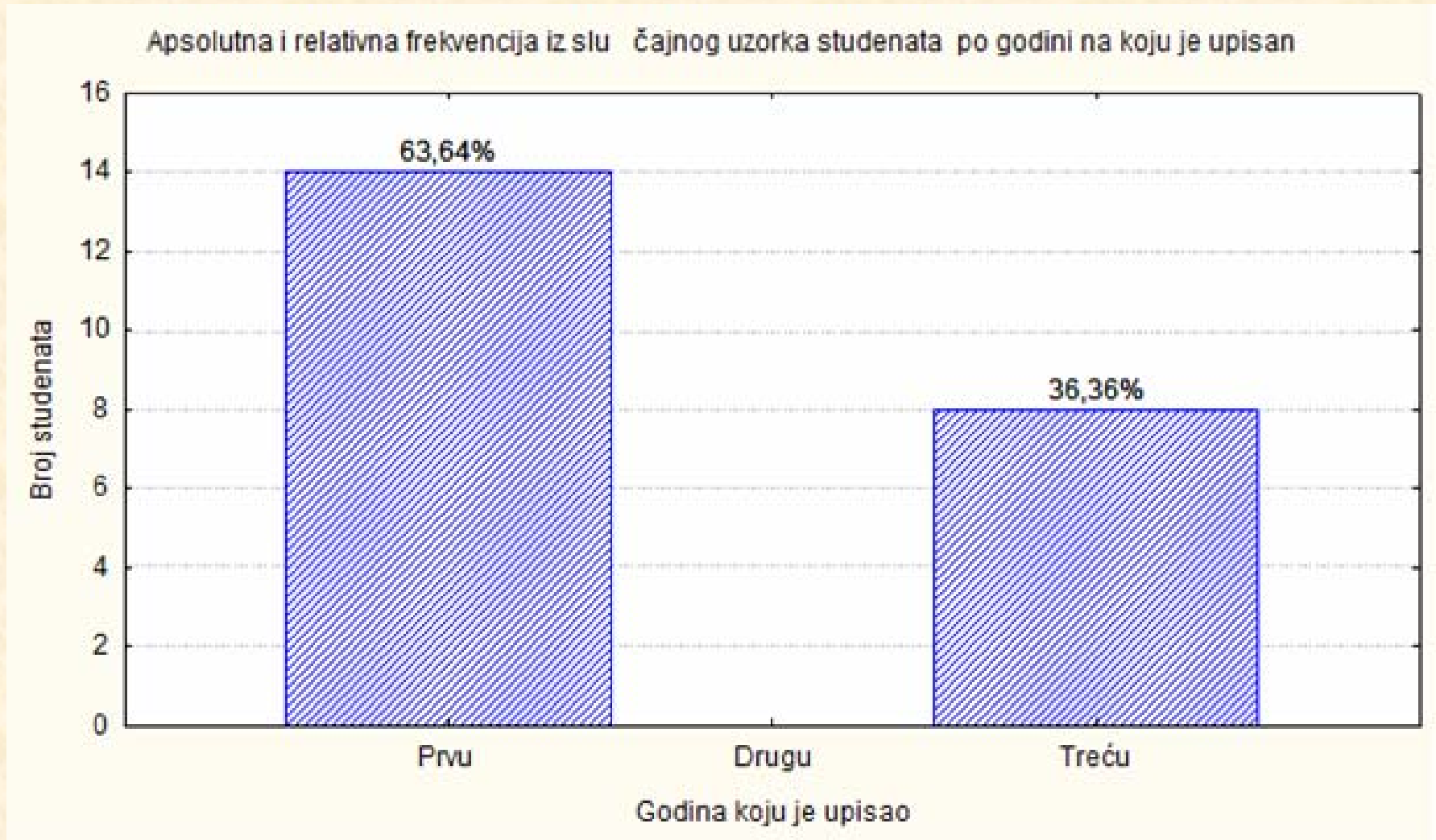
Slika 32. Histogram smerova za slučajni uzorak od 4% populacije diplomiranih studenata od 1995. do 1999. god.



Slika 33. Histogram apsolutne i relativne frekvencije godina upisa studenta na fakultet za diplomirane 1995-1999. god. dobijen na osnovu slučajnog uzorka iz ove populacije.



Slika 34. Histogram apsolutne i relativne frekvencije godine rođenja studenata no osnovu 4% populacije dipl. stud. 1995 – 1999.

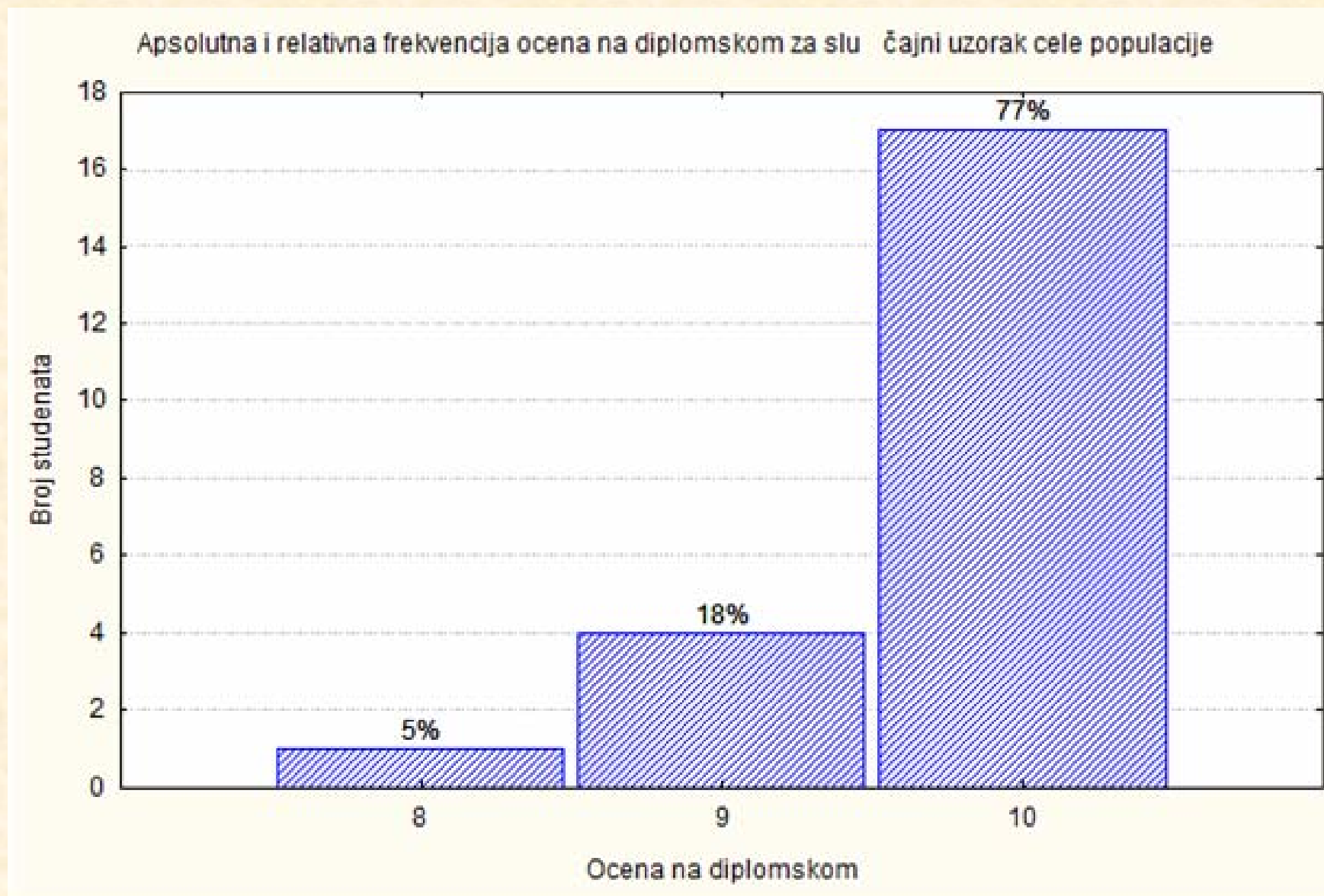


Slika 35. Slučajni uzorak od 4% populacije u odnosu na godinu na koju je upisan za diplomirane od 1995. do 1999. god.

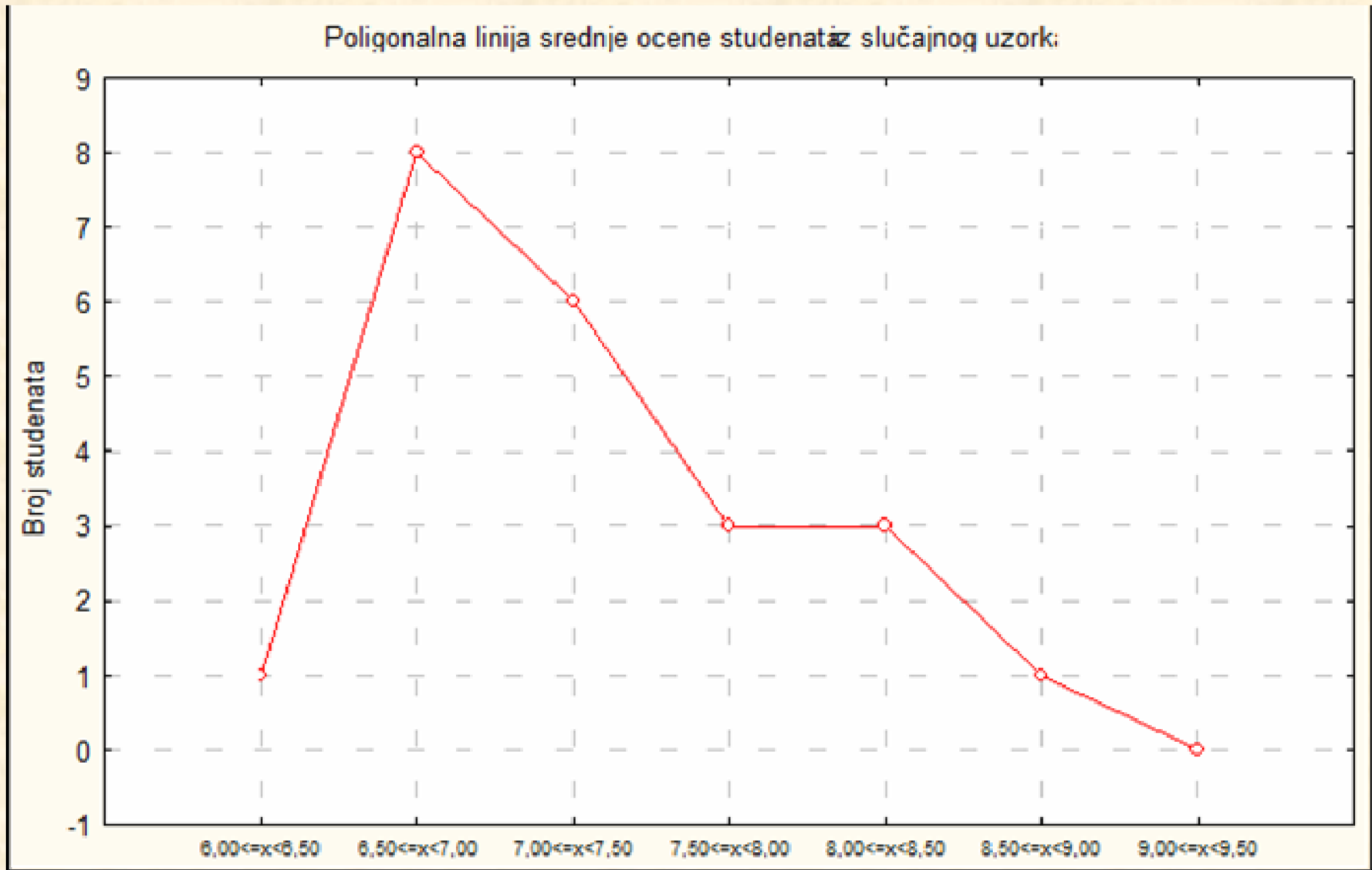
Odnos polova iz slučajnog uzorka svih studenata diplomiranih u periodu 1995-1999 god.



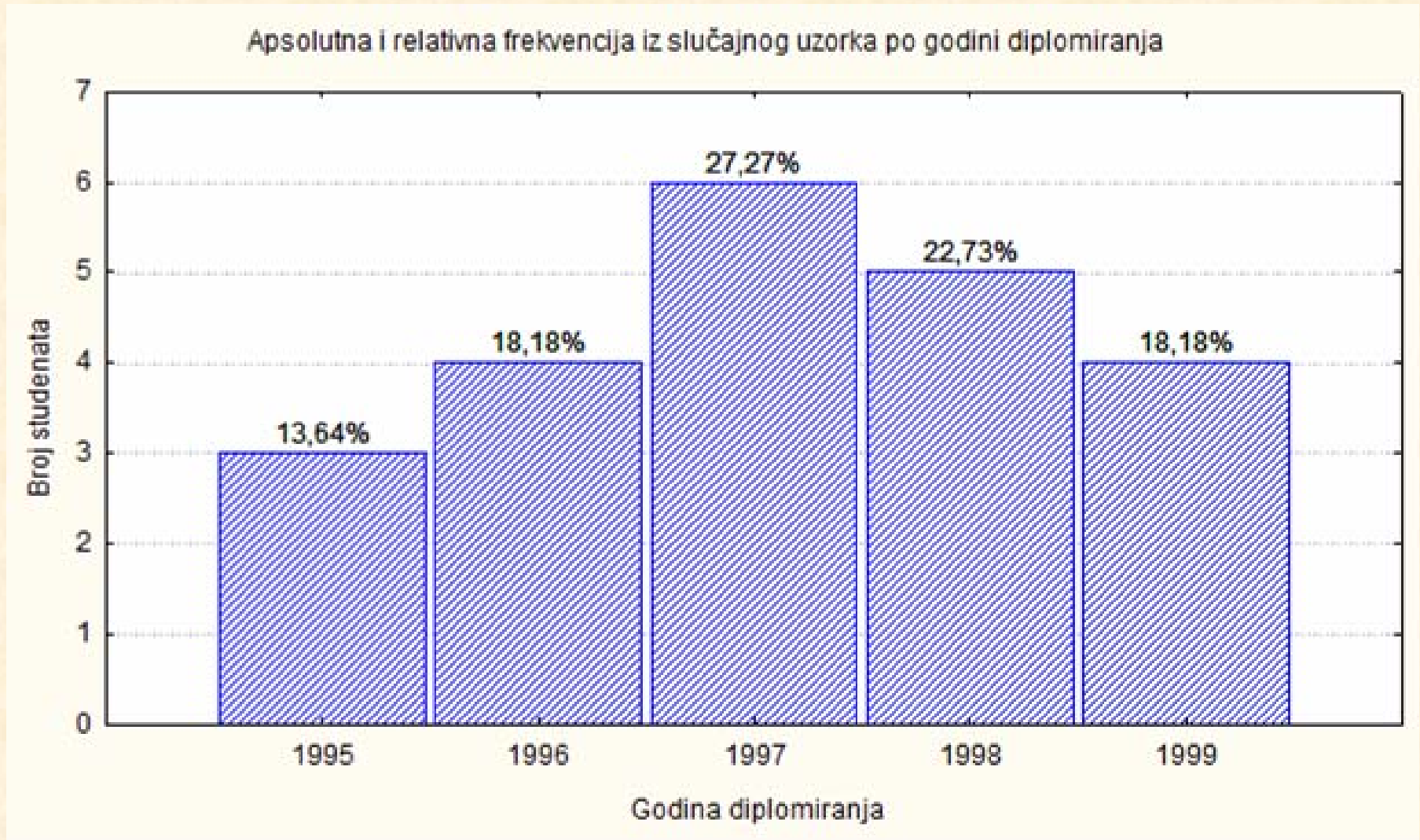
Slika 36. Pol iz slučajnog uzorka od 4% studenata diplomiranih u periodu 1995 – 1999. god.



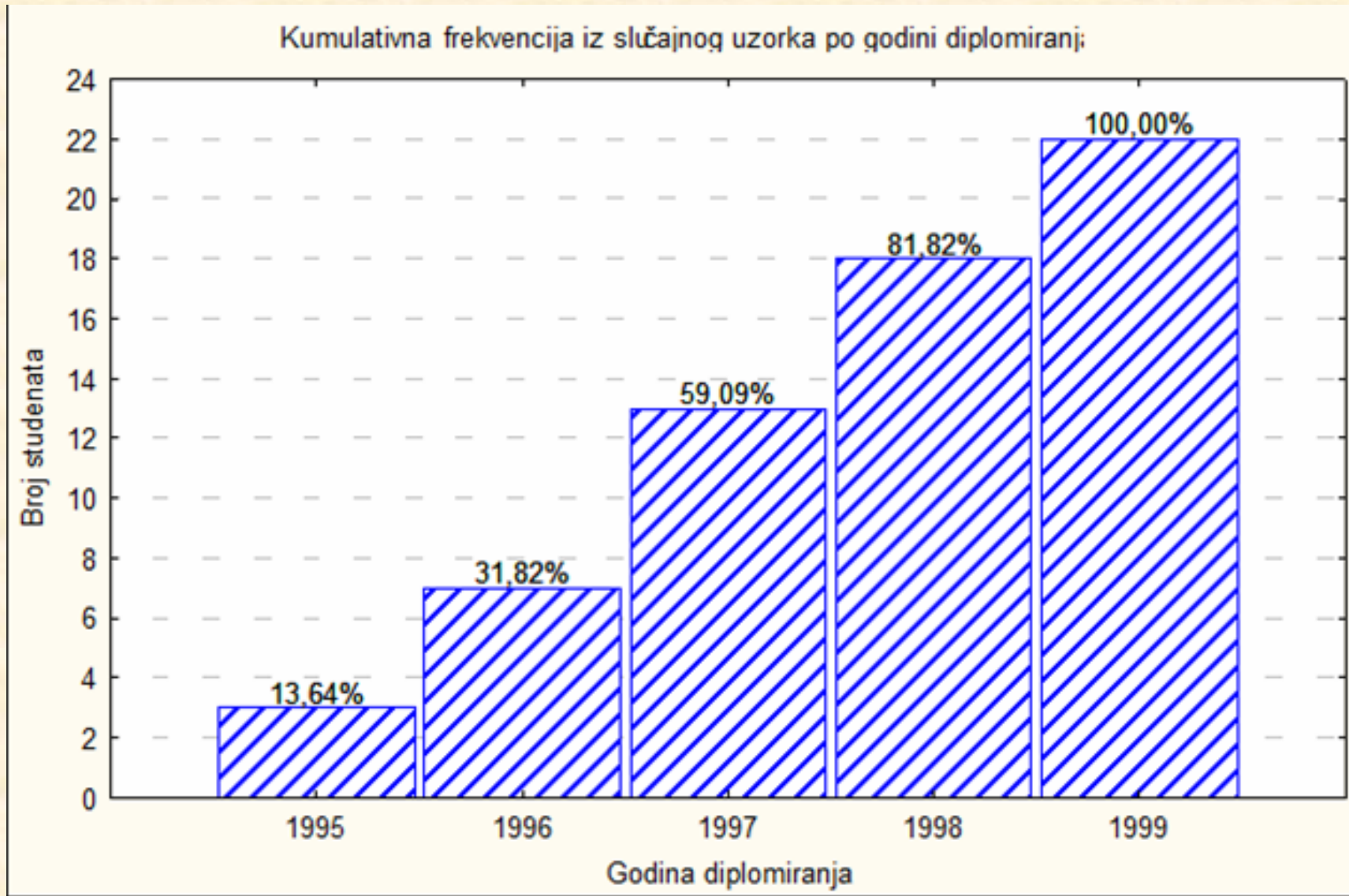
Slika 37. Histogram apsolutne i relativna frekvencija ocena na diplomskom za slučajni uzorak iz cele populacije (1995-1999).



Slika 38. Slučajni uzork od 4% iz cele populacije 1995 – 1999.



Slika 39. Slučajni uzorak od 4% populacije diplomaca od 1995 – 1999.



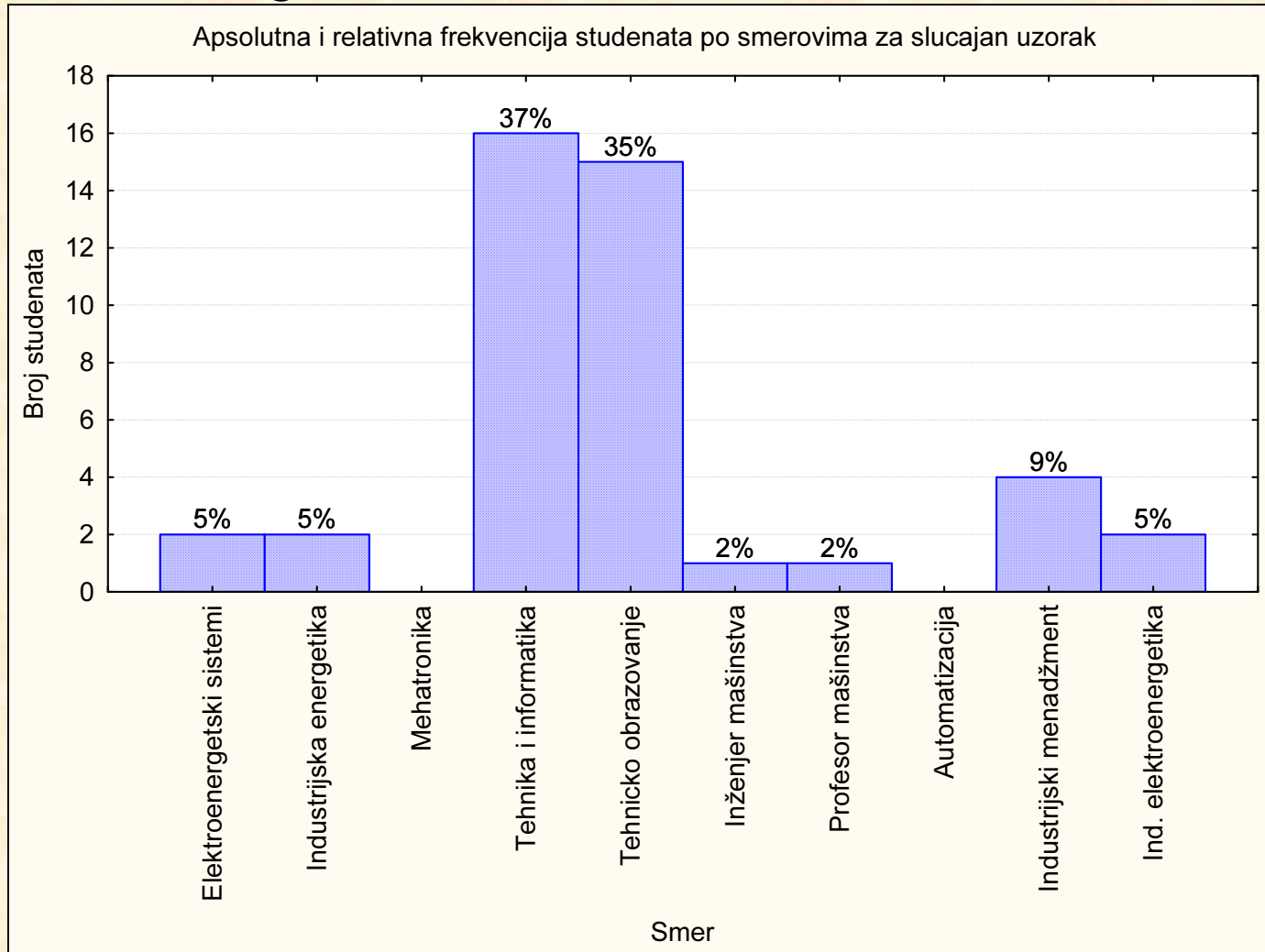
Slika 40. Histogram kumulativne frekvencije za godinu diplomiranja dobijen na osnovu 4% populacije 1995 – 1999. god.

Populacija studenata koji su diplomirali u periodu 2000-2004. ima ukupno 561 elemenat i za obim našeg slučajnog uzorka uzećemo 8% od ukupne populacije. Dakle naš slučajni uzorak ima 43 elementa (Slika 41).

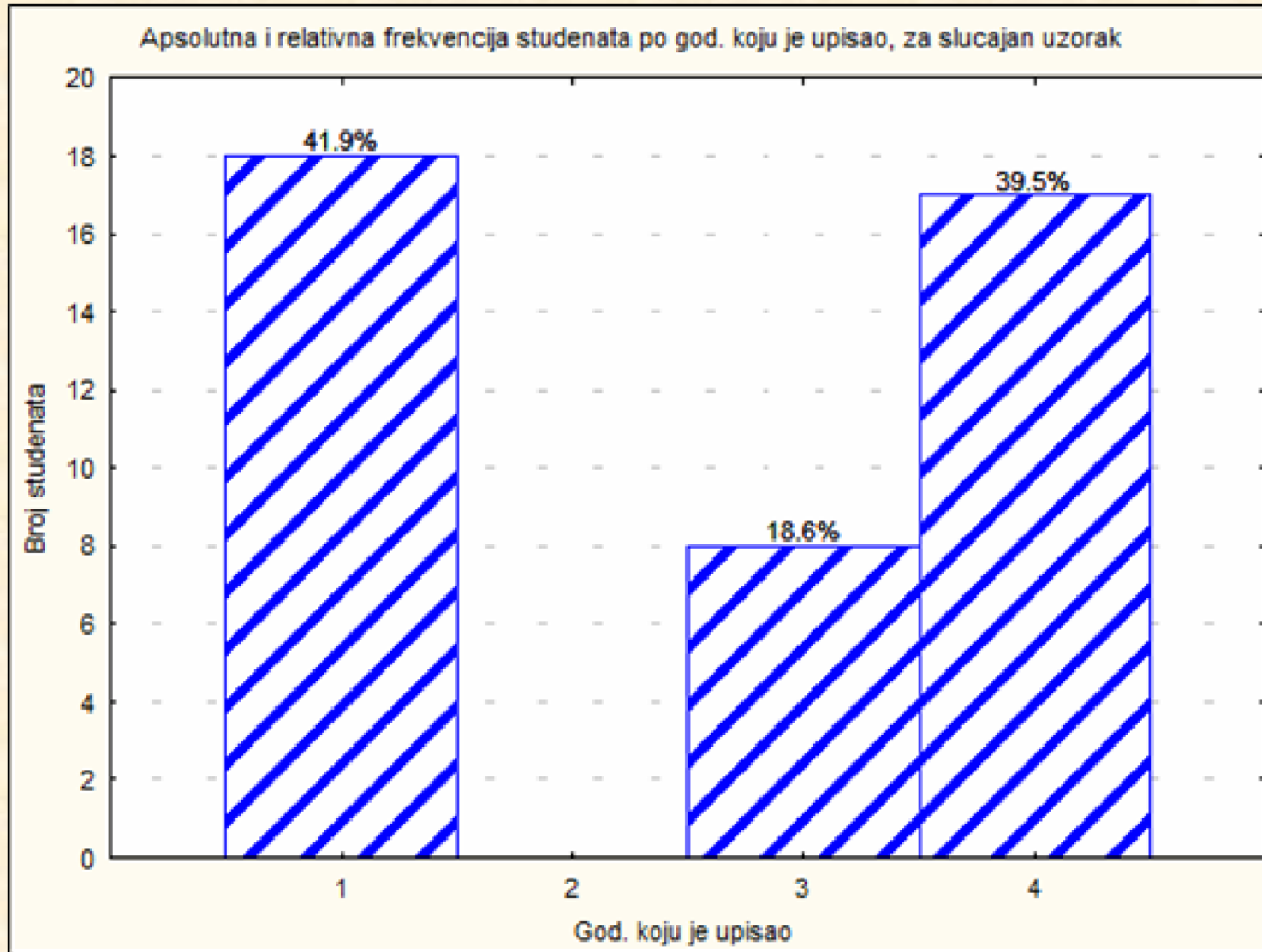
	1 Ime i prezime	2 God. rođenja	3 Mesto rođenja	4 God. upisa	5 Diplomirao/la	6 Prosek	7 Smer	8 Pol	9 God. koju je upisao
1	Maric Miloje	1967	Valjevo	1996	2000	6.5	Tehnika i informatika	M	1
2	Obucina Dragan	1975	Nova Varoš	1997	2002	6.73	Tehnicko obrazovanje	M	1
3	Radivojevic Suzana	1972	Kragujevac	1997	2002	6.77	Tehnika i informatika	Z	3
4	Nikolic Dragan	1970	Kruševac	1996	2000	6.82	Tehnika i informatika	M	1
5	Smiljkovic Dragana	1966	Prokuplje	2003	2004	6.82	Tehnicko obrazovanje	Z	4
6	Nastic Goran	1976	Kruševac	1999	2003	6.83	Ind. elektroenergetika	M	4
7	Zoric Vladimir	1974	Nikšić (CG)	2002	2004	6.87	Tehnicko obrazovanje	M	4
8	Popovic Andrija	1966	Valjevo	1996	2000	6.91	Tehnika i informatika	M	1
9	Cabrilo Zlatan	1956	Kraljevo	2002	2003	6.91	Tehnicko obrazovanje	M	4
10	Vulic Tamara	1979	Smederevo	2001	2003	6.93	Tehnicko obrazovanje	Z	4
11	Gagic Ivica	1975	Valjevo	1994	2002	7	Industrijska energetika	M	1
12	Jekic Dragica	1954	Trstenik	1975	2001	7.03	Profesor mašinstva	Z	1
13	Milenkovic Aleksandar	1973	Split (HR)	1992	2004	7.03	Elektroenergetski sistemi	M	1
14	Tanovic Zeljko	1979	Beograd (BiH)	1999	2004	7.09	Tehnika i informatika	Z	4
				⋮					
41	Rogac Persa	1953	Beograd	2001	2003	8.73	Tehnicko obrazovanje	Z	4
42	Pavicevic Radmila	1957	Kragujevac	2001	2004	8.86	Industrijski menadžment	Z	3
43	Lazic Ana	1974	Kruševac	1997	2000	9.37	Tehnika i informatika	Z	3

Slika 41. Slučajan uzorak diplomiranih studenata u periodu 2000 – 2004. god.

Na osnovu ove tabele vršićemo slične analize kao kod prethodnog uzoraka, tj. diplomiranih 2001. godine.



Slika 42. Apsolutna i relativna frekvencija studenata po smerovima za slučajni uzorak od 8% iz populacije 2000 – 2004. god.



Slika 43. Histogram godine studija na koju su studenti upisani dobijen iz slučajnog uzorka od 8% iz diplomiranih od 2000 – 2004. god.

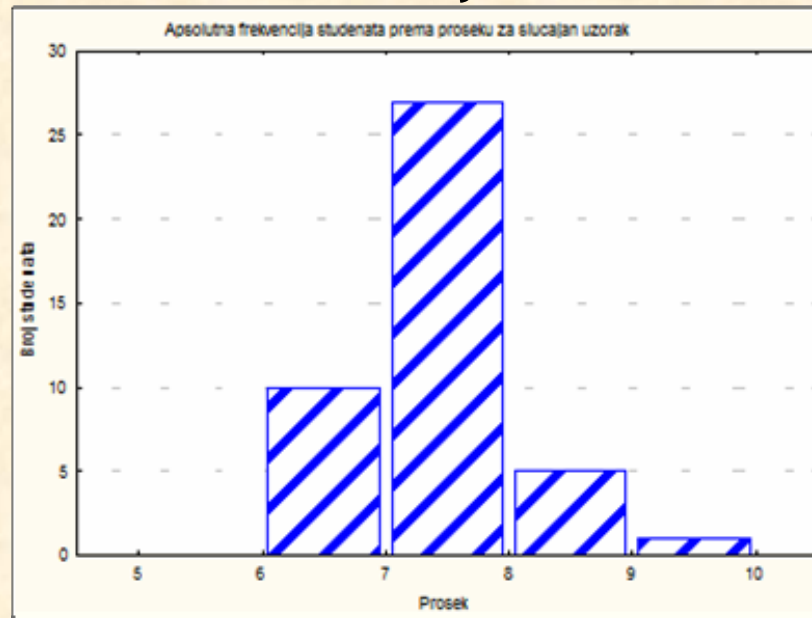


Slika 44. Dijagram relativne frekvencije pola za slučajan uzorak iz 2000 – 2004. god.

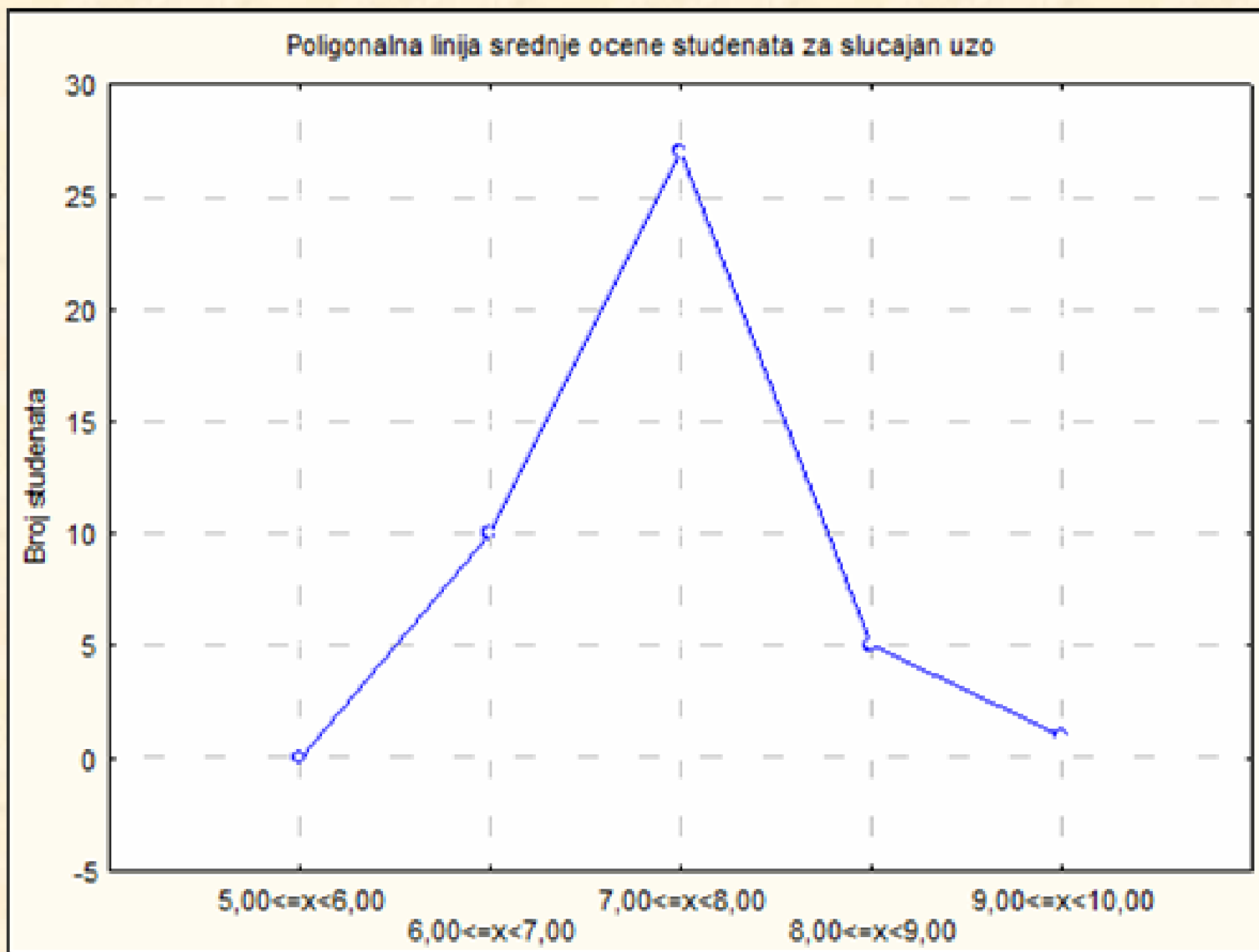
Za analizu prosečne ocene studenata u toku studija kreiramo novu tablicu sa intervalima i primenimo na slučajan uzorak (Slika 45).

		Frequency table: Prosek (Slučajan uzorak stu			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
5,00	$\leq x < 6,00$	0	0	0.00000	0.0000
6,00	$\leq x < 7,00$	10	10	23.25581	23.2558
7,00	$\leq x < 8,00$	27	37	62.79070	86.0465
8,00	$\leq x < 9,00$	5	42	11.62791	97.6744
9,00	$\leq x < 10,00$	1	43	2.32558	100.0000

Slika 45. Frequency Table: Prosek (slučajan uzorak studenata iz 2000 – 2004.)

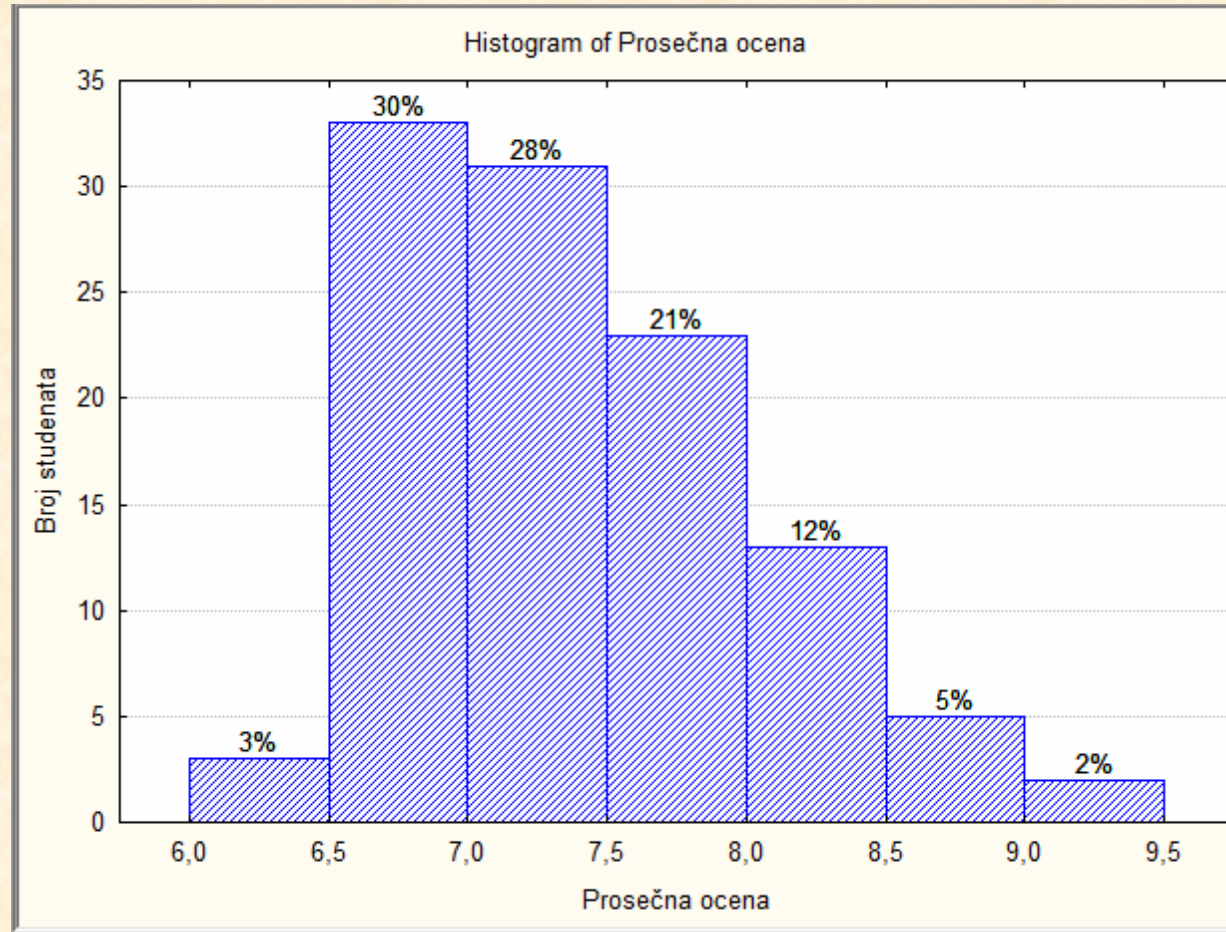


Slika 46. Histogram apsolutne frekvencije na osnovu slučajnog uzorka od 8% iz populacije 2000 – 2004. god.



Slika 47. Poligonalna linija srednje ocene studenata za slučajan uzorak iz 2000 – 2004. god.

Za populaciju diplomiranih studenata od 2008 – 2011. god. (1000 studenata) izabran je 10% slučajni uzorak i frekventnost prosečnih ocena prikazana je na Slici 48.



Slika 48. Prosečne ocene studenata koji su diplomirali 2008 – 2011. god.