

VAJA 02 IZ EXCELA: OPISNA STATISTIKA. UREJANJE IN PRIKAZOVANJE PODATKOV

Datum: oktober 2011

Pripravila Katja L. Manfreda

S to vajo spoznamo načine urejanja podatkov v frekvenčne porazdelitve in njihove grafične predstavitve.

Podatki so že vneseni v datoteko Vaje 2.xlsx.

Literatura: Poglavlja 1 (v celoti), 2.1, 2.2, 4.4 in str. 82-83 v *Kastelec, D., Košmelj, B. (2010): Osnove statistike z Excelom 2007. El. knjiga. Ljubljana: Biotehniška fakulteta. Poglavlje 1.*

NAVODILO

Urejanje in prikazovanje podatkov za nominalne in ordinalne spremenljivke.

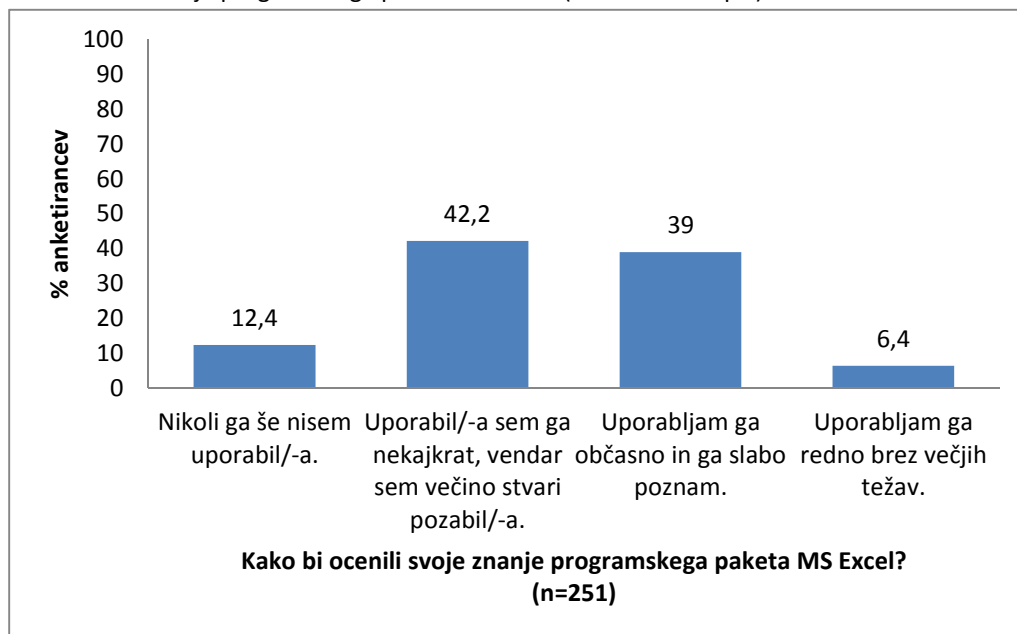
1. Iz ankete o računalniški pismenosti in poznavanju Excela med študenti FDV pri predmetu Statistika v oktobru 2011 smo pridobili naslednje podatke o poznavanju programskega paketa MS Excel:

Kako bi ocenili svoje znanje programskega paketa Microsoft Office Excel (MS Excel)?	
Odgovori	Frekvenca <i>f_i</i>
Nikoli ga še nisem uporabil/-a.	31
Uporabil/-a sem ga nekajkrat, vendar sem večino stvari pozabil/-a.	107
Uporabljam ga občasno in ga slabo poznam.	98
Uporabljam ga redno brez večjih težav.	16
Skupaj	

Tabela 1: Znanje programskega jezika MS Excel

- a) Kaj je v tem primeru enota analize in kaj spremenljivka? Kaj so vrednosti spremenljivke in kakšna je njihova merska lestvica?
- b) Z uporabo formul in funkcij v Excelu izračunajte: skupno število anketirancev ter relativne frekvence/strukturne % (na eno decimalno natančno) za vse odgovore. Kolikšen odstotek anketiranih študentov še nikoli ni uporabljal Excela? Kolikšen odstotek anketiranih študentov uporablja Excel občasno ali redno?
- c) Podatke grafično prikažite s strukturnimi stolpci, pri čemer naj bodo predstavljene relativne frekvence. Grafikon ustrezno uredite tako, kot je prikazano spodaj.

Slika 1: Poznavanje programskega paketa MS Excel (strukturni stolpci)



2. Iz ankete o računalniški pismenosti in poznavanju Excela med študenti FDV pri predmetu Statistika v oktobru 2011 smo pridobili naslednje podatke o računalniški pismenosti:

Katere od spodaj naštetih dejavnosti, povezanih z računalnikom, ste že opravljali? Če ste neko dejavnost že opravljali, izberite odgovor "Da", sicer izberite odgovor "Ne".				
Podvprašanja	Odgovori			
	Da	Ne	Ne vem	Skupaj
Uporaba programov za pisanje in oblikovanje besedila (npr. MS Word ali kak drug urejevalnik).	250	2	1	253
Ustvarjanje predstavitev v programih za njihovo pripravo (npr. MS Power Point ali kak drug program).	243	9	1	253
Računanje v programih za delo s preglednicami (npr. MS Excel ali kak drug program).	123	117	13	253
Kopiranje in lepljenje za podvajanje ali premikanje podatkov/besedila/slik itd.	198	39	16	253
Pisanje računalniških programov s programskimi jeziki.	23	217	13	253

Tabela 2: Dejavnosti, povezanih z računalnikom

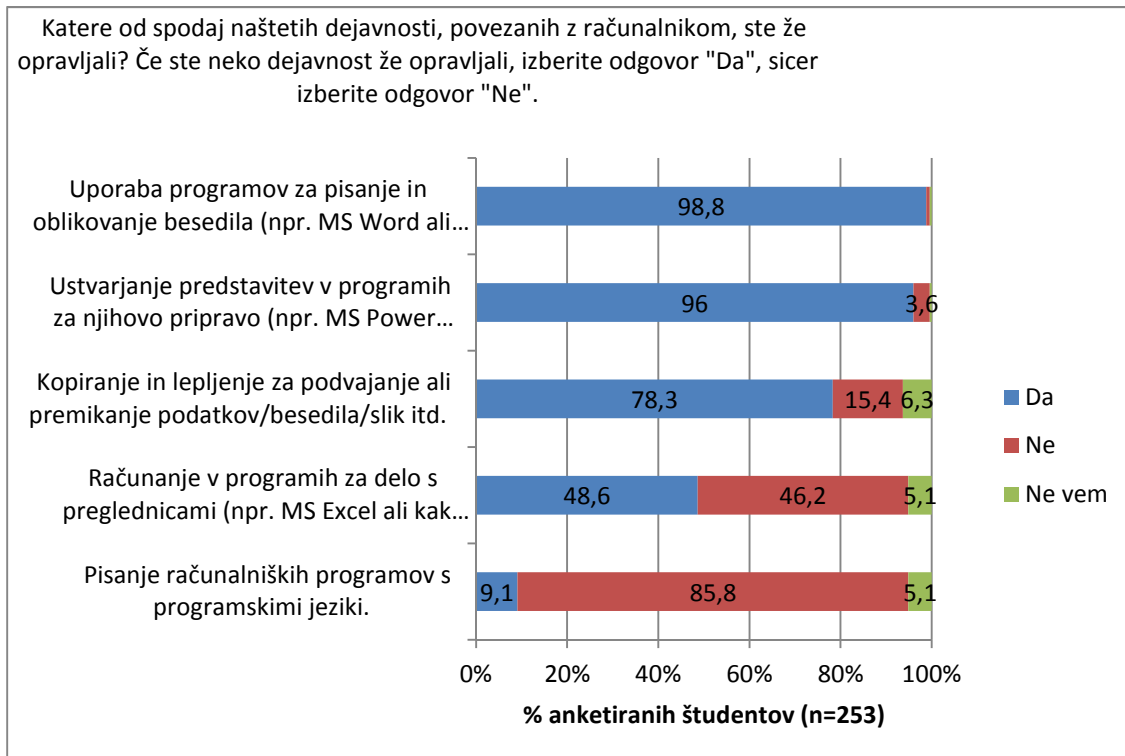
- Koliko in katere spremenljivke obravnavamo v tem primeru? Kakšna je njihova merska lestvica? Kaj je enota analize?
- Naredite novo tabelo, v katero za vsako spremenljivko izračunajte relativne frekvence/strukturne % (na eno decimalko natančno). Kolikšen odstotek študentov še nikoli ni računalno v programih za delo s preglednicami? Katero od naštetih aktivnosti uporablja največji del študentov?
- Podatke predstavite z enim grafikonom s strukturnimi stolpci, kot je prikazano na Sliki 2. Predstavljene naj bodo relativne frekvence, dejavnosti pa naj bodo nanizane od tiste, ki jo uporablja največ študentov, do tiste, ki jo uporablja najmanj študentov.

V ta namen je potrebno pred risanjem grafa tabelo s strukturnimi odstotki kopirati kot vrednosti v nov niz podatkov (ukaz *Posebno lepljenje (Paste Special) – Vrednosti (Values)* znotraj skupine

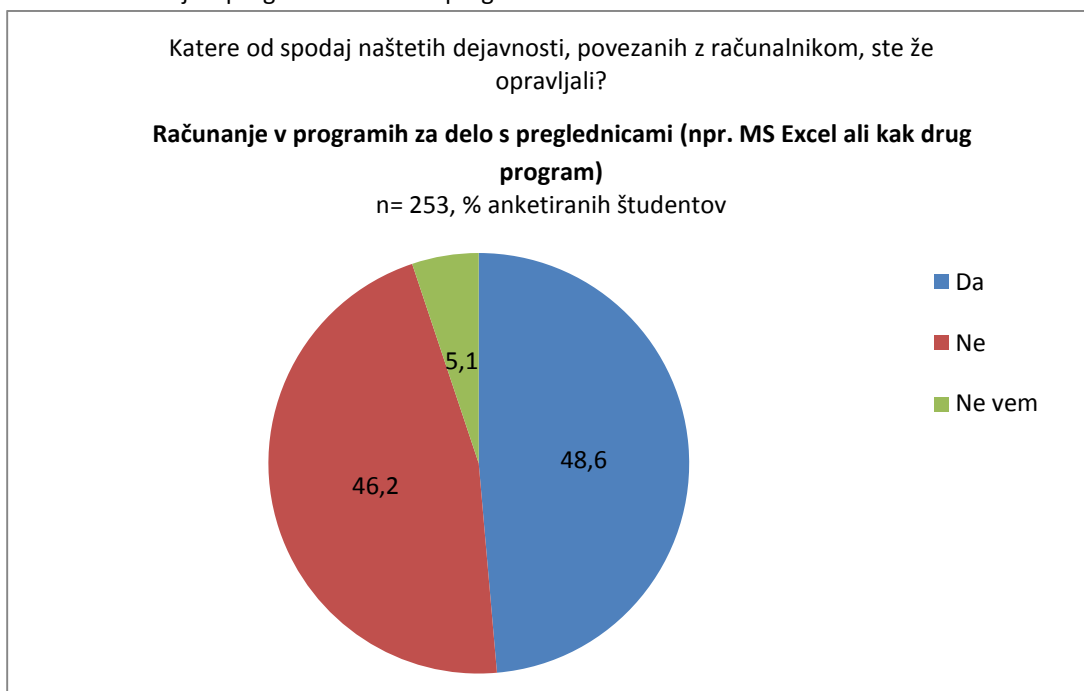
Lepljenje (Paste) v kartici Osnovno (Home) in spremenljivke nato urediti glede na % tistih, ki so rekli Da (skupina Sortiranje (Sort) znotraj kratice Podatki (Data)). Za risanje grafa uporabimo tip grafikona Bar.

- d) Za dejavnost »Računanje v programih s preglednicami« narišite tudi strukturni krog, ki naj bo oblikovan tako, kot je prikazano na Sliki 3.

Slika 2: Računalniška pismenost



Slika 3: Računanje v programih za delo s preglednicami.



Urejanje in prikazovanje podatkov za intervalne in razmernostne spremenljivke

3. 52 staršev šoloobveznih otrok smo vprašali o tem, koliko denarja so namenili za nakup šolskih knjig, delovnih zvezkov in drugih potrebščin za novo študijsko leto. Prosili smo jih, da znesek zaokrožijo na 10 evrov natančno. Podatki so naslednji:

160	150	110	70	270	270	270	270	260	290
180	190	200	180	200	240	200	200	230	200
250	260	270	360	50	70	270	130	520	60
370	400	390	350	360	450	380	260	270	280
450	520	120	480	350	350	200	200	160	150
110	350								

- a) Kaj je v tem primeru enota analize in kaj spremenljivka? Kaj so vrednosti spremenljivke in kakšna je njihova merska lestvica?
- b) Vrednosti spremenljivke uredite v frekvenčno tabelo. V tabelo poleg absolutnih frekvenc vpišite tudi relativne frekvence, kumulativne frekvence in relativne kumulativne frekvence (tako kot je prikazano v Tabeli 3). Pri tem relativne frekvence izračunajte na dve decimalki natančno. Kolikšen % staršev je za šolske potrebščine porabil manj kot 250 evrov?

Frekvenčno tabelo lahko naredimo s pomočjo vrtilnih tabel (angl. *Pivot Tables*), v našem primeru enosmerno vrtilno tabelo. V ta namen zgornje podatke zapišemo v en stolpec, ki naj ima neko naslovno celico (npr. z vrednostjo x_i). Frekvenčno tabelo naredimo z ukazom *Vrtilna tabela* (angl. *Pivot Tables*) na kartici *Vstavljanje (Insert)* v skupini *Tabele (Tables)*. Vrtilno tabelo delamo v novem delovnem listu, kjer se nam pojavi shema prazne vrtilne tabele in pogovorno okno *Seznam polj vrtilne tabele (Pivot Table Field List)*. V slednjem iz seznama spremenljivk izberemo našo spremenljivo (označeno z vrednostjo naslovne celice, v našem primeru x_i) in jo prenesemo v ustrezno območje vrtilne tabele. Najprej jo prenesemo v območje *Oznake vrstic (Row Labels)* in s tem določimo, kaj naj bo zapisano v vrsticah vrtilne tabele (v tem primeru različne vrednosti naše spremenljivke). Ponovno izberemo spremenljivko (x_i) in jo prenesemo v območje *Vrednosti (Values)*. Excel tako v prvi stolpec tabele zapiše izračune funkcije *VSOTA (SUM)*, kar je privzeta funkcija, ki jo izračuna na podlagi spremenljivk(e), ki jih(jo) določimo v območju vrednosti. Ker mi želimo imeti frekvence (kolikokrat se pojavi posamezna vrednost), bomo namesto tega izbrali funkcijo *ŠTEJ (COUNT)*. To lahko naredimo tako, da funkcijo izberemo iz seznama funkcij, ki se nam odpre z desnim klikom na prvo vrstico izpisane vrtilne tabele ter izbiro *Nastavitve polja vrednosti (Value Field Settings)*.

Tabelo nato ustrezno oblikujemo in dopolnimo, kot je zahtevano zgoraj.

Strošek šolskih potrebščin (EUR)	Število staršev			
	f_i	$f_i\%$	F_i	$F_i\%$
x_i				
50	1	1,92	1	1,92
60	1	1,92	2	3,85
70	2	3,85	4	7,69
110	2	3,85	6	11,54
...
	52	100		

Tabela 3: Strošek šolskih potrebščin, frekvenčna porazdelitev, negrupirani podatki

- c) Vrednosti spremenljivke uredite v nekaj smiselnih razredov enake širine (npr. 6 razredov po 100 EUR). Dobljeno tabelo uredite tako, da zapišete ustrezne širine razredov in dodajte absolutnim tudi relativne, kumulativne in relativne kumulativne frekvence. Pripišite tudi pripadajoče meje razredov, širino in sredino razredov (tako kot je prikazano v Tabeli 4). Koliko staršev je za šolske potrebščine porabilo 300 EUR ali manj?

V nadaljevanju sta zapisani dve možnosti razvrščanja vrednosti v razrede.

1. možnost

Vrednosti grupiramo v razrede s pomočjo funkcij *COUNTIF* in *COUNTIFS*. S pomočjo teh funkcij za neko celico določimo, da naj se vanjo zapiše seštevek vseh vrednosti, ki ustrezajo nekemu pogoju (*COUNTIF*) oz. kombinaciji pogojev (*COUNTIFS*). Npr. za vrstico tabele »nad 0 – 100« zapišemo funkcijo *COUNTIF(niz podatkov, "<=100")*, za vrstico tabele »nad 100 – 200« pa funkcijo *COUNTIFS(niz podatkov;">100";niz podatkov;"<=200")*. Podobno naredimo še za ostale razrede. Dobljeno tabelo uredimo in dopolnimo, tako kot je prikazano v Tabeli 4.

2. možnost

Če je potrebno, za rešitev te naloge aktiviramo dodatek *Analysis ToolPak (Datoteka – Možnosti – Dodatki – Analysis ToolPak oz. File – Options – Add-Ins - Analysis ToolPak)*. V okviru kartice *Podatki (angl. Data)* se pojavi nova skupina ukazov *Analiza podatkov (angl. Data Analysis)*. V pogovornem oknu znotraj ukaza *Analiza podatkov* bomo za namene grupiranja podatkov uporabili možnost *Histogram (angl. Histogram)*. Še prej pa je potrebno določiti število in širino razredov. To določimo tako, da v nek niz poleg niza negrupiranih vrednosti v aktivnem delovnem listu vnesemo zgornje meje predvidenih razredov (v našem primeru npr. 100, 200, 300, 400, 500, 600 za 6 razredov po 100 EUR). V pogovornem oknu *Histogram* nato pri določanju obsega podatkov v prvo vrstico (*Input Range*) vpišemo niz podatkov, na katerem se nahajajo negrupirane vrednosti, v drugo vrstico (*Bin Range*) pa niz z zgornjimi mejami razredov. Določimo lahko, da naj se nova tabela izpiše znotraj istega delovnega lista ali v novem delovnem listu. Dobljeno tabelo uredimo in dopolnimo, tako kot je prikazano v Tabeli 4.

Strošek šolskih potrebščin (EUR)								
Razredi	f_i	$f_i\%$	F_i	$F_i\%$	d_i	$x_{i,min}$	$x_{i,max}$	x_i
nad 0 - 100	4	7,69	4	7,69	100	0	100	50
nad 100 - 200	18	34,62	22	42,31	100	100	200	150
nad 200 - 300	15	28,85	37	71,15	100	200	300	250
nad 300 - 400	10	19,23	47	90,38	100	300	400	350
nad 400 - 500	3	5,77	50	96,15	100	400	500	450
nad 500 - 600	2	3,85	52	100	100	500	600	550
	52	100						

Tabela 4: Strošek šolskih potrebščin, frekvenčna porazdelitev za grupirane podatke

- d) Podatke iz Tabele 4 grafično predstavite s histogramom, poligonom in ogivo, kot je prikazano na slikah spodaj. Kakšna je porazdelitev spremenljivke?

Histogram lahko narišemo kot stolpčni grafikon (*Column*) ali s pomočjo ukaza *Podatki – Analiza (Data – Analysis)* istočasno z oblikovanjem tabele, kot je opisano v 2. možnosti pri nalogi c. V tem primeru v pogovornem oknu *Histogram* izberete še možnost *Chart Output*. V obeh primerih je grafikon potrebno še nekoliko oblikovati (vpisati ustrezne oznake za vodoravno os, odstraniti razmik med stolpci itd.), da izgleda kot graf na Sliki 5.

Poligon lahko narišemo s pomočjo črtnega grafikona (angl. *Line*). V ta namen je še prej potrebno v tabelo grupiranih podatkov dodati en razred pred prvim razredom s frekvenco 0 in

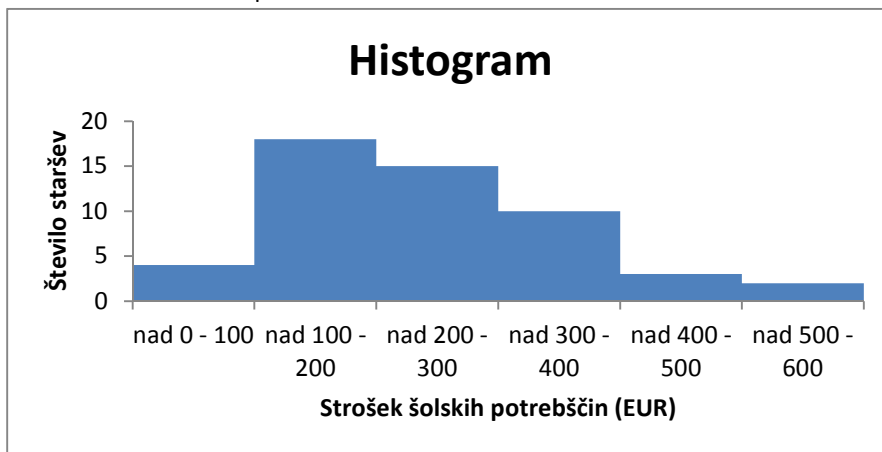
sredino razreda 0 ter en razred za zadnjim razredom s frekvenco 0 in sredino razreda 650. Za izris poligona naj bosta potem uporabljena niza podatkov sredina razreda in frekvenca, ki naj bosta zapisana v dveh sosednjih stolpcih.

Podobno je mogoče narisati ogivo. Potrebno je uporabiti dodatni razred pred prvim razredom, ki naj ima zgornjo mejo razreda 0 in relativno kumulativno frekvenco 0. Za izris naj bosta uporabljena niza podatkov zgornja meja razreda in relativna kumulativna frekvenca, ki naj bosta zapisana v dveh sosednjih stolpcih.

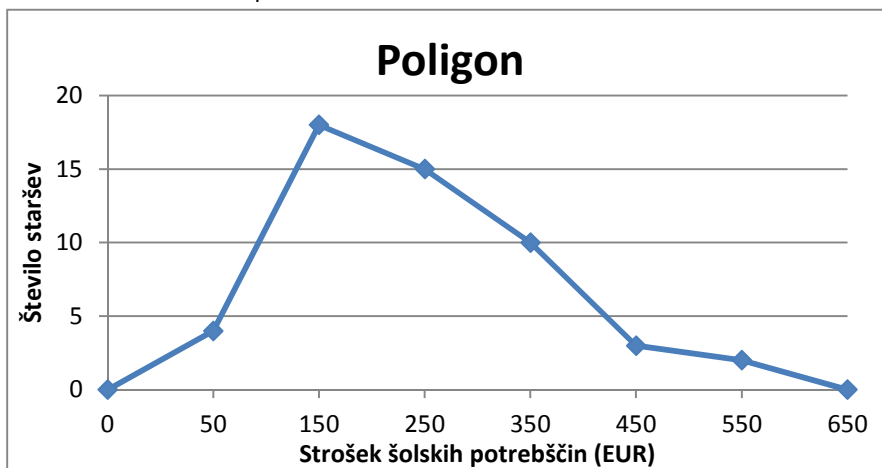
Vse grafe je potrebno še dodatno oblikovati, da so vrednosti na obeh oseh pravilno izpisane.

Poligon in ogivo je mogoče narisati tudi s pomočjo razstresenega (razsevnega) grafikona (angl. *Scatter*), kot je predstavljeno v Kastelec in Košmelj (2010), poglavje str. 82-83.

Slika 4: Strošek šolskih potrebščin



Slika 5: Strošek šolskih potrebščin



Slika 6: Strošek šolskih potrebščin

