

Jedna slika vredi koliko hiljadu reči

Grupisanje podataka – u slučaju velikih setova podataka ponovljenih merenja tj. prilikom posmatranja raspodele neke slučajne promenljive na određenom uzorku (visina učenika u srednjim školama zapadne Srbije, koncentracija tiolnih jedinjenja kod pacijenata koji boluju od reumatoidnog artritisa,...) najpre se podaci moraju grupisati, tj. raspodeliti u grupne intervale.

Formiranje grupnih intervala:

- Određivanje broja grupnih intervala – broj grupnih intervala (n) na koje se skup deli određuje se približno kao $\sqrt{N}$, gde N predstavlja ukupan broj podataka, dok se kod jako velikog broja podataka broj grupnih intervala može odrediti prema izrazu: $n = 1 + 3,22 \log N$.
- Utvrdjivanje širine grupnog intervala – grupni intervali moraju biti jednake širine da bi bili međusobno uporedivi. Širina grupnog intervala se izračunava tako što se nađe razlika između najviše i najniže vrednosti u skupu, a zatim se ona podeli prethodno određenim brojem grupnih intervala. Dobijeni rezultat se zaokruži na najbliži ceo broj ili na manji broj decimalnih mesta, ali uvek sa istom tačnošću sa kojom su izraženi i sami podaci.
- Određivanje granica intervala –
 - donja granica prvog intervala, koji obavezno mora da sadrži najnižu vrednost seta podataka, treba da bude broj koji je deljiv sa širinom intervala;
 - donja granica prvog intervala bez obzira na širinu intervala može da počne nulom;
 - preporuka je da donja granica grupnog intervala bude prva dekadna jedinica manja od najniže vrednosti seta podataka;
 - donja granica intervala mora biti za jedinicu mere veća od gornje granice prethodnog intervala;
 - najviša vrednost seta podataka mora biti obuhvaćena poslednjim grupnim intervalom.

Najefikasniji način kojim se istovremeno podaci mogu grupisati u vidu tablice frekvencija i grafički prikazati u obliku histograma, dobija se korišćenjem alatke histogram, u okviru Data Analysis ToolPack-a.

Primer 1. Bilirubin

Kod 120 osoba vršeno je određivanje koncentracije bilirubina ($\mu\text{mol/dm}^3$). Dobijene vrednosti grupisati u grupne intervale.

4,1	5,5	5,9	6,7	7,0	7,4	7,6	8,0	8,1	8,1	8,7
9,0	9,2	9,4	9,4	9,6	10,2	10,4	10,8	10,6	11,1	11,2
11,2	11,4	11,5	11,6	11,7	11,7	12,0	11,9	12,0	12,0	12,1
12,2	12,4	12,5	12,6	12,8	13,0	12,9	13,0	13,1	13,4	13,5
13,5	13,6	13,8	13,8	13,9	13,9	13,9	14,0	14,0	14,0	14,2
14,2	14,3	14,4	14,5	14,4	14,6	14,7	14,8	14,8	14,9	15,0
15,0	15,1	15,4	15,3	15,5	15,5	15,6	15,6	15,6	15,8	15,9
16,0	16,0	16,0	16,1	16,2	16,3	16,4	16,4	16,5	16,5	16,7
17,0	16,9	17,1	17,2	17,5	17,5	17,8	18,0	18,2	18,4	18,6
18,4	18,8	18,9	19,0	19,2	19,5	19,7	19,9	20,2	20,6	20,3
20,8	21,0	21,2	21,3	21,6	22,0	22,1	22,4	22,5	23,0	

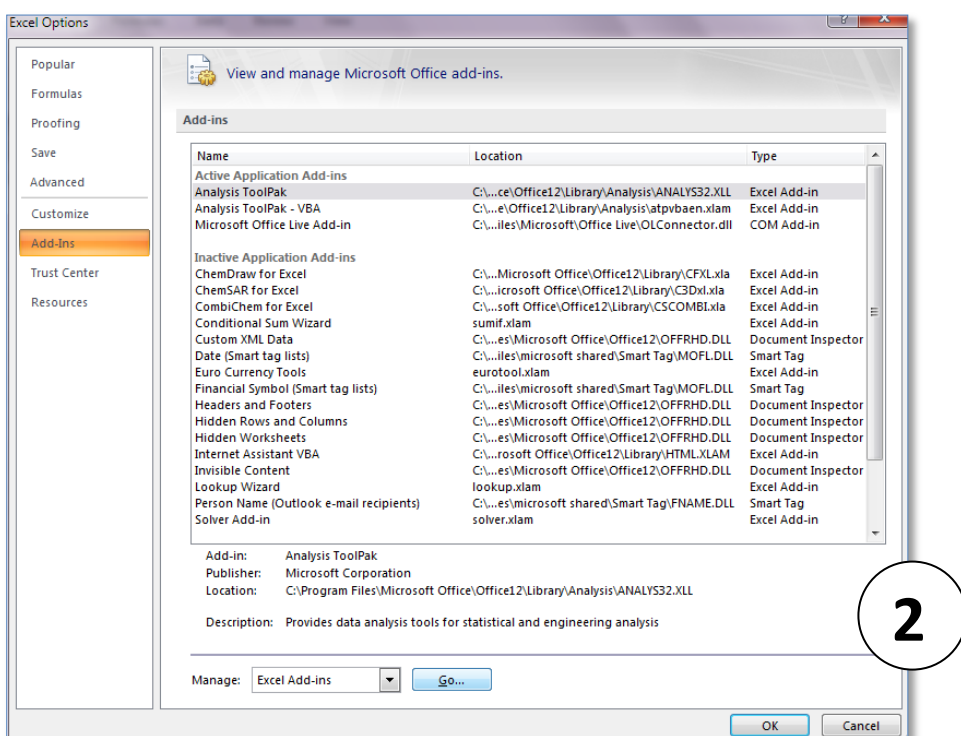
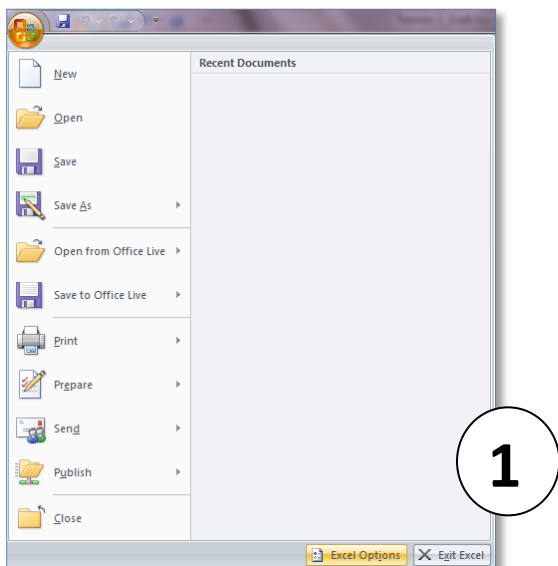
Proverite da li je u okviru programskog okruženja Excel instaliran Analysis toolpack (Tools/Data Analysis) – paket alata za statističku obradu podataka. Ukoliko nije instaliran pratite sledeću proceduru (koraci 1-4). Nakon instaliranja paketa za statističku obradu podataka, odaberite opciju sa padajućeg menija Tools/Data Analysis; trebalo bi da vam se otvori dijalog kao na slici 1, korak 4. Tada odaberite opciju histogram.

Šta ćete raditi danas?

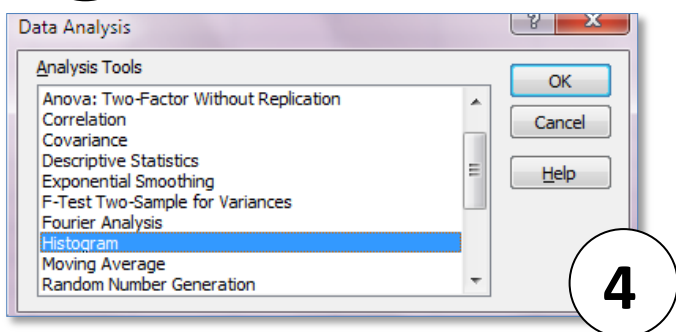
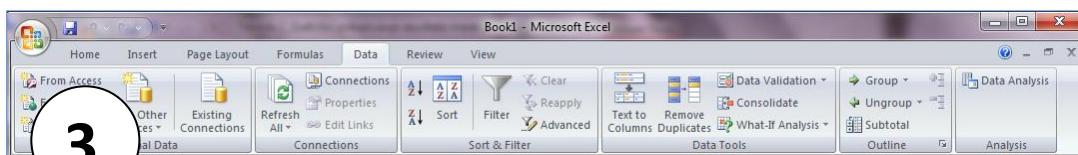
Grupisanje podataka
Tabelarno prikazivanje podataka
Kreiranje dijagrama pomoću Chart Wizard-a
Izbor tipa i podtipa dijagrama
Izbor opsega podataka i serije
Paleta alata chart
Postavljanje labela
Promena tipa dijagrama
Promena serije podataka
Izmena boja, linija, tekstura i stilova na dijagramu

Pravilno izabran broj i širina intervala omogućavaju i dobru preglednost i dobru informaciju o setu podataka

Obavezno pre početka uređivanja, prikazivanja i statističke obrade podataka proveriti prisustvo rezultata koji su posledica grubih grešaka!!!



Slika 1. Procedura za instaliranje paketa alata za statističku obradu podataka (Analysis Toolpack) i alatke za izradu histograma i grupisanje podataka



Pre nego što pristupite grupisanju podataka i crtanju histograma, neophodno je izračunati nekoliko parametara:

- Širina intervala unutar koga su smešteni svi podaci - **R** (range, opseg):

$$R = \max(Ax : Xz) - \min(Ax : Xz)$$
- Širina grupnog intervala **w**: $w = \frac{R}{\sqrt{N}}$
- Zaokružite širinu grupnog intervala na tačnost sa kojom su dati rezultati merenja. Ne zaboravite da kada za zaokruživanje koristite alatku koja smanjuje broj decimalnih mesta, excel i dalje čuva u svojoj memoriji broj sa svim raspoloživim decimalama. Najbolje je da pošto smanjite prikaz decimala na odgovarajući nivo, jednostavno broj obrišete i ručno unesete u ćeliju zaokruženu vrednost.
- Izračunajte donje granice grupnih intervala. Pri ovome treba zadovoljiti uslov da je donja granica svakog intervala deljiva sa širinom **w**. Za donju granicu prvog intervala, d_0 , ovaj uslov je ispunjen kada važi $d_0 = k \cdot w$ gde je **k** ceo broj. Kako na početku ne poznajemo d_0 , niti **k**, najpre ćemo izračunati vrednost k' kao, $k' = \frac{\min()}{w}$, a zaokruživanjem ove vrednosti na prvi manji celi broj dobija se **k**. Tada se d_0 može izračunati prema pređašnjoj formuli. Donja granica sledećeg intervala dobija se kada se na donju granicu prethodnog doda širina intervala, pa bi za drugi interval $d_1 = d_0 + w$.

Ukoliko ste sve parametre izračunali pažljivo trebalo bi da dobijete sledeće rezultate (slika 2). Ukoliko je minimalna vrednost seta podataka smeštena između prve i druge donje granice (d_0 i d_1) znači da ste sve uradili kako valja.

Starujte komandu histogram, iz paketa alata za analizu podataka (Tools/Data Analysis/Histogram). U polje **Input Range** unesite opseg ćelija između kojih su smešteni vaši podaci. U polje **Bin Range** uneti opseg ćelija u kojima su smeštene vaše granice intervala. U polje **Output Range** unesite ćeliju ispod koje i desno do koje nema nikakvih podataka na radnom listu, u suprotnom excel će vam saopštiti da će rezultate prepisati preko već postojećih podataka. Odaberite još opcije **Chart Output** (prikazuje dijagram) i **cumulative percentage** (daje kumulativnu raspodelu u procentima).

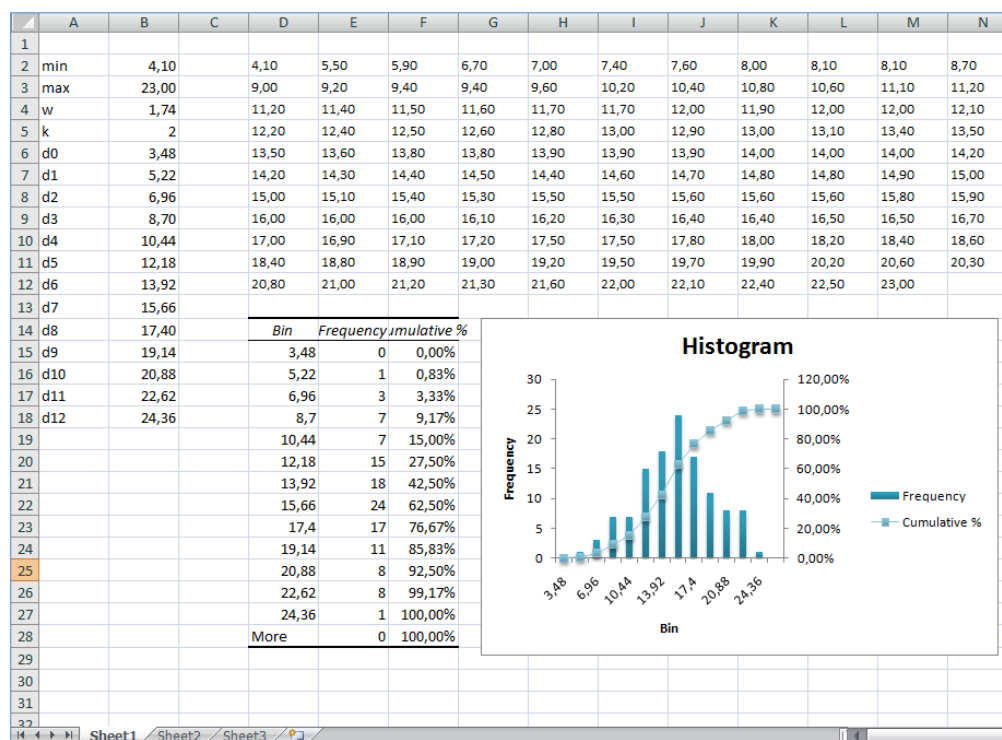
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1															
2	min	4,10			4,10	5,50	5,90	6,70	7,00	7,40	7,60	8,00	8,10	8,10	8,70
3	max	23,00			9,00	9,20	9,40	9,40	9,60	10,20	10,40	10,80	10,60	11,10	11,20
4	w	1,74			11,20	11,40	11,50	11,60	11,70	11,70	12,00	11,90	12,00	12,00	12,10
5	k	2			12,20	12,40	12,50	12,60	12,80	13,00	12,90	13,00	13,10	13,40	13,50
6	d0	3,48			13,50	13,60	13,80	13,8	13,90	13,90	13,90	14,00	14,00	14,00	14,20
7	d1	5,22			14,20	14,30	14,40	14,50	14,40	14,60	14,70	14,80	14,80	14,90	15,00
8	d2	6,96			15,00	15,10	15,40	15,30	15,50	15,50	15,60	15,60	15,60	15,80	15,90
9	d3	8,70			16,00	16,00	16,0	16,10	16,20	16,30	16,40	16,40	16,50	16,50	16,70
10	d4	10,44			17,00	16,90	17,10	17,20	17,50	17,50	17,80	18,00	18,20	18,40	18,60
11	d5	12,18			18,40	18,80	18,90	19,00	19,20	19,50	19,70	19,90	20,20	20,60	20,30
12	d6	13,92			20,80	21,00	21,20	21,30	21,60	22,00	22,10	22,40	22,50	23,00	
13	d7	15,66													
14	d8	17,40													
15	d9	19,14													
16	d10	20,88													
17	d11	22,62													
18	d12	24,36													
19															
20															
21															
22															
23															
24															
25															
26															
27															

Dogovorićemo se da oznakom Ax:Xz predstavljamo bilo koji set podataka, na primer C2:D4, označava konkretno set podataka koji se nalaze u pravougaoniku između ćelija C2 i D4.

N-broj merenja u nekom setu podataka možete dobiti pomoću funkcije prebrojavanja **count**, na primer unošenjem **=count (C2:D4)** dobijate broj podataka između polja C2 i D4.

Slika 2. Izračunati parametri neophodni za grupisanje podataka pomoću alatke histogram, u slučaju primera 1. Izgled dijaloga alatke histogram.

Ukoliko ste sve ispravno uradili trebalo bi da konačan rezultat izgleda ovako:



Tabeliranje – pregledno prikazivanje prethodno grupisanih podataka tabelom.

Svaka tabela ima tekstualni i brojčani deo.

Tekstualni deo čine:

- redni broj tabele;
- naslov, odnosno naziv tabele, koji objašnjava sadržaj tabele;
- zaglavlje tabele, koje obuhvata prvi red tabele, i u njoj se upisuje tekst kojim se tumače podaci o kolonama;
- pretkolona, koja obuhvata prvu kolonu, i u njoj upisani tekst objašnjava podatke u redovima;
- izvor podataka.

Brojčani deo tabele su kolone i redovi, odnosno polja koja se dobijaju njihovim ukrštanjem.

Svaka tabela mora da ima kolonu i red za zbir. U tabeli nijedno polje nesme da ostane prazno; ukoliko za neko polje nema podataka u njega se upisuje crtica.

Tabela broj ... Naziv tabele ...

	Zbir po redovima
ZAGLAVLJE	
Redni broj kolone	
PREDKOLONA	
Redni broj	
Zbir po kolonama	Ukupan zbir

Primer 2. Krvne grupe

Dvadeset slučajno izabranih bolesnika hematološkog odeljenja Interne A klinike Medicinskog fakulteta Univerziteta u Beogradu pregledom krvi koji je izvršen 24.10.2005. godine pokazalo je sledeće pripadništvo krvnim grupama A, B, 0 i Rh sistema: od 14 Rh-pozitivnih pet bolesnika imalo je 0, troje A, četvoro B i dvoje AB krvnu grupu, a od 6 Rh-negativnih krvnu grupu 0 imalo je dvoje, A dvoje, B jedan i AB jedan bolesnik. Dobijene rezultate prikazati tabelom.

Tabela 1. Zastupljenost faktora krvnih grupa kod pacijenata hematološkog odeljenja interne A klinike Medicinskog fakulteta Univerziteta u Beogradu

Rh faktor	Krvne grupe			
	A	B	AB	0
+	3	4	2	5
-	2	1	1	2

Grafičko prikazivanje – metoda prikazivanja grupisanih i tabelarnih podataka u vizuelnom obliku.

Histogram – način grafičkog predstavljanja statističkih podataka u obliku serije pravougaonika.

Pripada grupi površinskih dijagrama (dve dimenzije) koji se konstruišu u pravougaonom koordinatnom sistemu tako što se na apscisu nanose grupni intervali, a na ordinatu broj slučajeva u svakom intervalu, tj. frekvencija.

Površine ili visine pravougaonika, ako su grupni intervali jednake širine, predstavljaju frekvencije grupnih intervala.

Poligon – pripada grupi linijskih dijagrama i konstruiše se u pravougaonom koordinatnom sistemu na taj način što se grupni intervali nanose na apscisu, a frekvencije na ordinatu. Dobija se tako što se iz sredine grupnih intervala dižu ordinate na koje se nanose odgovarajuće frekvencije. Spajajući ove tačke pravom linijom, dobija se izlomljena, poligonalna linija.

Primer 3. Krvne grupe - nastavak

Podatke date u primeru 1, prikažite pomoću poligona frekvencija.

Štapićasti dijagram – koristi se za prikazivanje stanja jednog prekidnog numeričkog obeležja. Pripada grupi linijskih dijagrama i konstruiše se u pravougaonom koordinatnom sistemu na taj način što se grupe nanose na apscisu, a frekvencije na ordinatu. Dobija se tako što se nad utvrđenim grupama dižu štapići čije visine predstavljaju vrednosti odgovarajućih frekvenci.

Kružni dijagram – upotrebljava se kada treba grafički prikazati strukturu jedne pojave (odnos delova prema celini). Pripada grupi površinskih dijagrama. Površina celog kruga predstavlja pojavu u celini, a površine pojedinih isečaka delove te celine. Konstruiše se u ugaonom sistemu. Veličina isečka određena je veličinom ugla, preko odnosa $100\% = 360^\circ$, tj. $1\% = 3,6^\circ$.

Primer 4. Raspodela studenata po godinama studiranja

Prikazati u obliku tabele broj studenata jednog fakulteta u odnosu na godinu studiranja i podatke prikazati u obliku štapičastog dijagrama frekvencija.

I godina – 505 studenata
II godina – 430 studenata
III godina – 410 studenata
IV godina – 380 studenata
V godina – 260 studenata

Primer 5. Prolaznost studenata na ispitu iz organske hemije

U septembarskom ispitnom roku za ispit iz organske hemije prijavilo se 342 studenta. Na ispitu se nije pojavio 121 student. 12 studenata je odustalo. Ocenu deset dobilo je 11 studenata, ocenu devet 35, ocenu osam 51, ocenu sedam 39, ocenu šest 32 i ocenu pet 41 student. Prikazati tabelarno i grafički strukturu izlaska studenata na ispit iz organske hemije

Zadaci

- 1) Spektrofotometrijskim merenjem absorbance u rastvoru nekog uzorka dobijene su sledeće vrednosti:

Grupisati date podatke.

0,3410	0,3350	0,3470	0,3590	0,3530	0,3460	0,3470	0,3460
0,3430	0,3420	0,3560	0,3500	0,3630	0,3530	0,3480	

- 2) Vrednosti dobijene kod 90 ponovljenih određivanja fosfora (vrednosti su date u %) u uzorku prirodnih fosfatida treba grupisati i prikazati grafički kao histogram frekvencija.

3,92	4,14	4,41	4,27	3,46	4,15	4,51	4,03	4,27	3,46	4,12	3,61
4,07	3,32	3,84	4,02	3,85	3,46	3,75	3,87	3,65	3,85	4,05	4,10
3,85	3,82	4,47	4,29	4,07	3,97	4,03	3,51	3,92	4,11	3,99	3,87
4,12	3,98	4,09	3,46	4,07	3,86	3,65	3,56	3,65	3,75	3,92	3,87
3,92	4,32	3,98	4,42	3,79	4,02	3,94	3,90	4,33	3,77	4,12	3,61
3,65	4,35	4,25	4,03	4,42	3,79	4,02	3,94	4,11	3,90	4,33	3,77
4,12	4,33	3,63	3,52	3,82	3,65	4,23	4,02	3,91	3,88	4,21	4,61
3,71	3,93	3,95	3,63								

- 3) Kontrolom kvaliteta težine tableta tokom 10 dana (svakog dana 10 tableta) dobijeno je 100 vrednosti (vrednosti su date u mg). Grupisati dobijene podatke i prikazati ih grafički histogramom.

303	308	298	304	305	300	297	301	302	300	301	301
301	292	300	299	301	298	299	298	302	294	296	302
302	303	298	300	297	300	301	302	299	300	302	302
304	301	295	305	303	303	298	296	297	298	301	297
299	303	303	300	302	302	304	302	303	304	306	307
298	299	296	301	300	295	301	301	295	300	298	304
307	301	304	302	299	300	305	302	306	303	304	296
295	297	298	301	300	300	302	296	298	301	306	305
300	296	299	300								

- 4) U okviru kontrole kvaliteta određivan je kalijum u jednom uzorku kontrolnog seruma u toku 92 dana (vrednosti su date u mmol/dm³). Dobijene podatke grupisati i prikazati ih pomoću poligona.

4,62	4,65	4,67	4,70	4,72	4,72	4,72	4,72	4,72	4,73	4,73	4,73
4,74	4,74	4,75	4,75	4,75	4,76	4,76	4,76	4,76	4,77	4,77	4,77
4,78	4,78	4,78	4,79	4,79	4,80	4,80	4,80	4,80	4,80	4,81	4,81
4,81	4,82	4,82	4,82	4,83	4,83	4,83	4,83	4,83	4,83	4,84	4,84
4,84	4,85	4,85	4,85	4,86	4,86	4,87	4,87	4,87	4,87	4,87	4,88
4,88	4,88	4,89	4,89	4,89	4,89	4,90	4,90	4,91	4,91	4,91	4,92
4,92	4,92	4,93	4,93	4,94	4,95	4,96	4,97	4,98	4,98	4,99	5,00
5,02	5,04	5,05									

- 5) Gravimetrijskom analizom određivan je maseni udeo sumpora u uglju i dobijeni su sledeći rezultati:

1,125	1,250	1,580	1,345	1,680	1,080	1,460	1,290	1,830	1,950
1,620	1,390	1,690	1,425	1,680	1,750	1,790	1,520	1,940%	2,050

Grupisati date podatke.