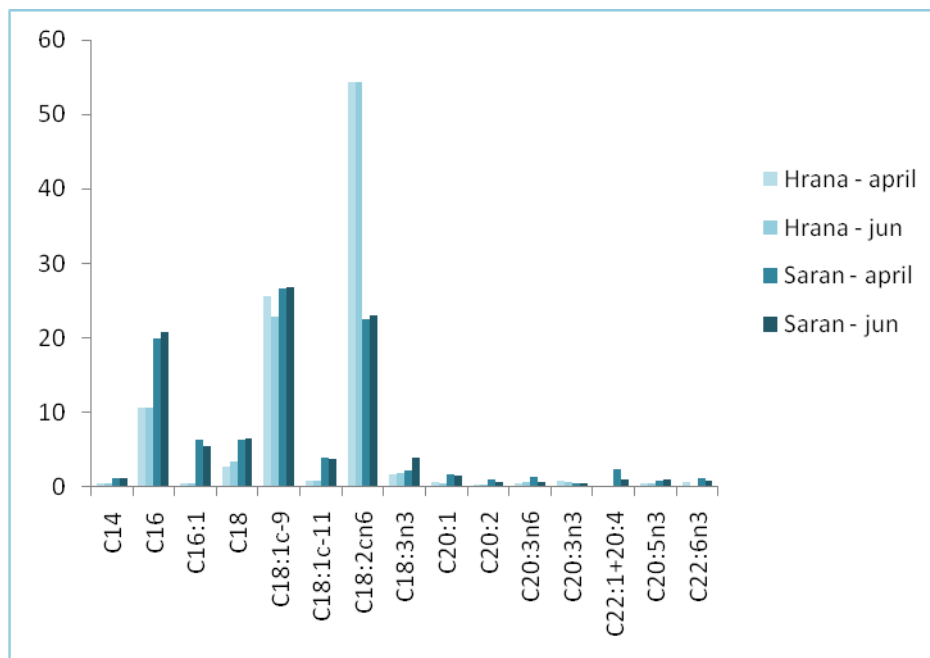


**Увод у статистичку обраду  
резултата у хемији**

**I део**

# Grafičko prikazivanje rezultata merenja

*Jedna slika vredi koliko hiljadu reči*



**Slika 1.** Distribucija sadržaja masnih kiselina u šaranu i hrani korišćenoj za ishranu ribe u aprilu i junu mesecu

## 1. 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:  $n = 1 + 3,22 \log N$ .
- Utvrđivanje 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 zaokrugli na najbliži ceo broj ili na manji broj decimalnih mesta, ali uvek sa istom tačnošću kojom su izraženi i sami podaci.

c) Određivanje granica intervala:

- donja granica prvog intervala, koji obavezno mora da sadrži najnižu vrednost seta podataka, treba da bude broj koji je deljiv 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.

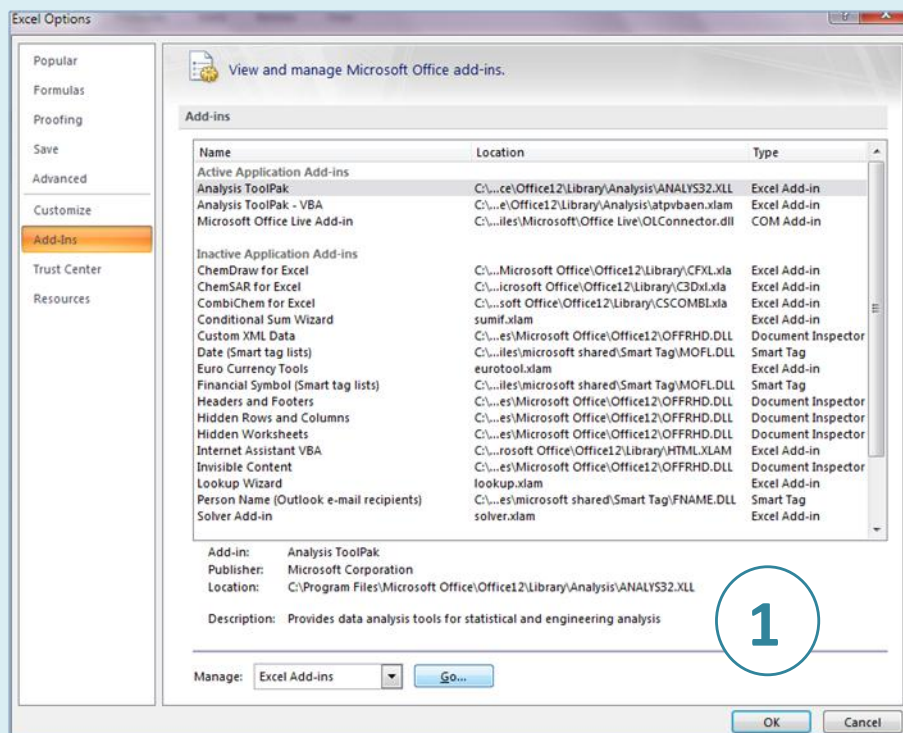
### Primer 1. Grupisanje rezultata

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

**Komentar:**

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 1-3). Nakon instaliranja paketa za statističku obradu podataka, odaberite opciju sa padajućeg menija Tools/Data Analysis; trebalo bi da vam se otvori dijalog kao na. Tada odaberite opciju histogram.



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

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

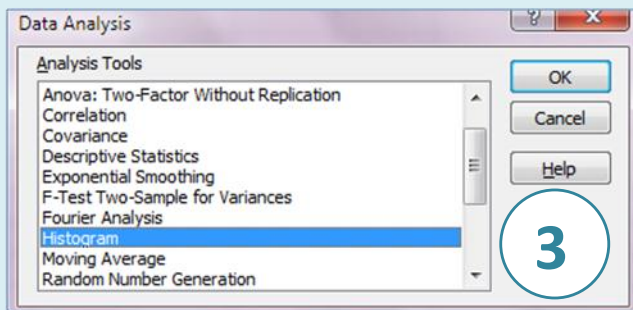
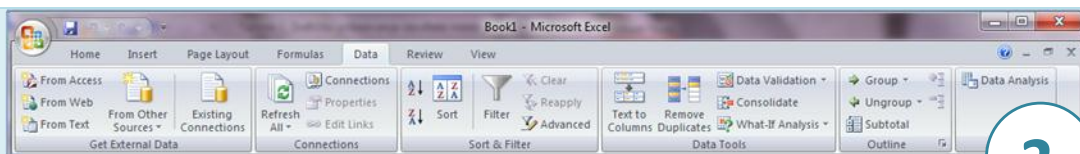
$$R = \max(Ax : Xz) - \min(Ax : Xz)$$
- **Širina grupnog intervala  $w$ :**  $w = \frac{R}{\sqrt{N}}$
- **Zaokruglite širinu grupnog intervala na tačnost kojom su dati rezultati merenja.** Ne zaboravite da kada za zaokrugljivanje koristite alatku koja smanjuje broj decimalnih mesta, Excel i dalje čuva 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 zaokrugljenu vrednost.
- **Izračunajte donje granice grupnih intervala.** Pri ovome treba zadovoljiti uslov da je donja granica svakog intervala deljiva 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 zaokrugljivanjem ove

vrednosti na prvi manji celi broj dobija se  $k$ . Tada se  $d_0$  može izračunati prema predašnjoj formuli. Donja granica sledećeg intervala dobija se kada se na donju granicu prethodnog doda širina intervala, pa je za drugi interval  $d_1 = d_0 + w$ .

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).

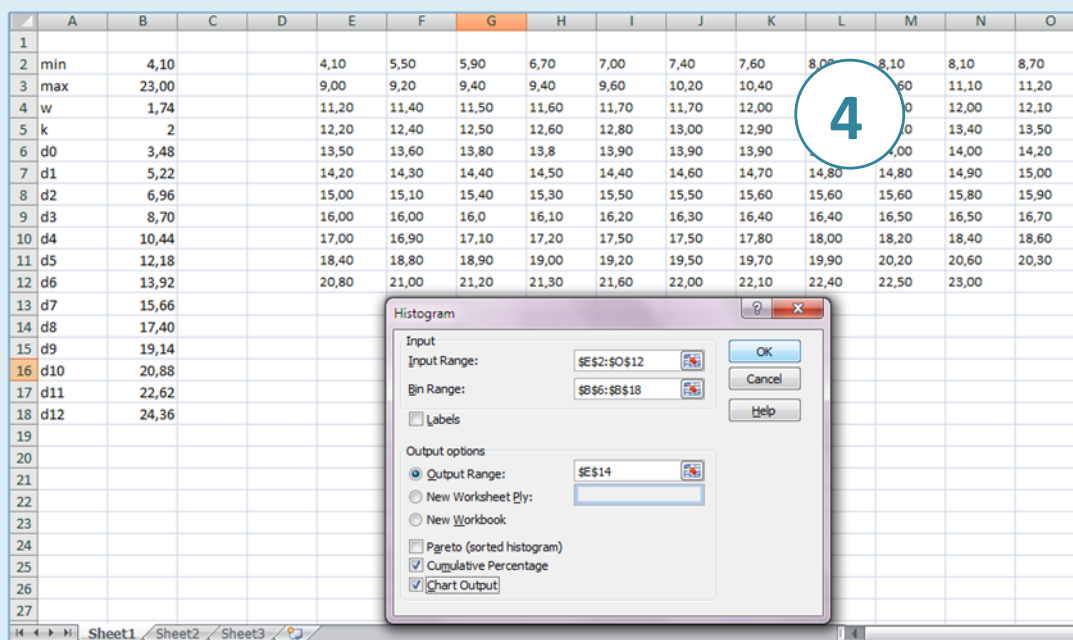


U polje **Input Range** unesite opseg ćelija između kojih su smešteni vaši podaci.

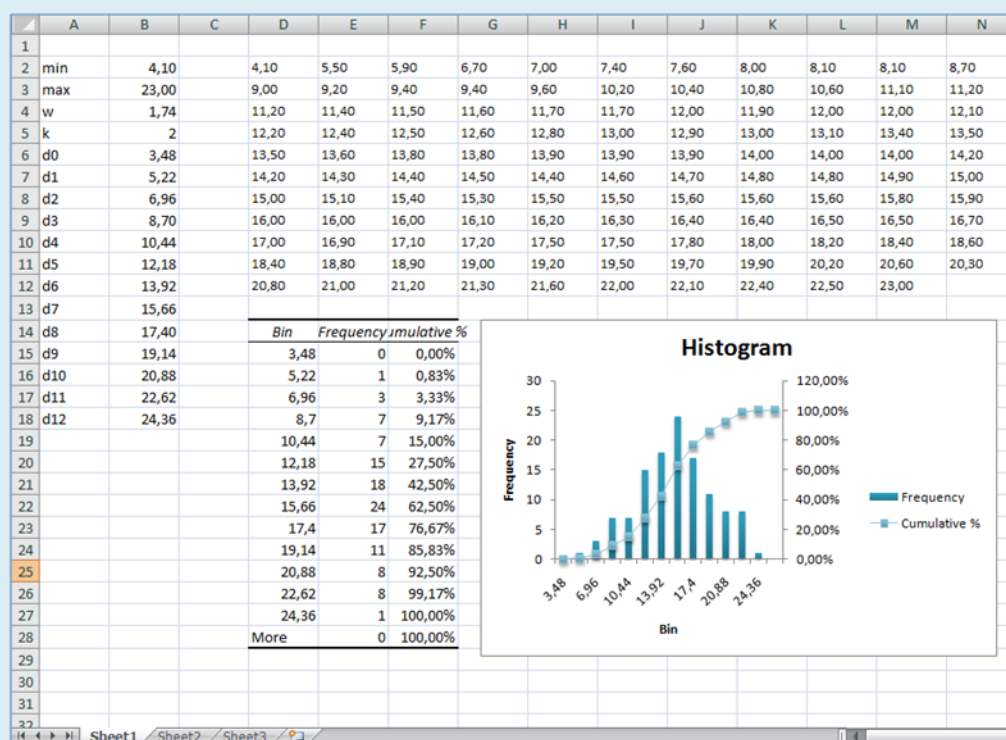
U polje **Bin Range** uneti opseg ćelija u kojima su smeštene vaše granice intervala.

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).



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



## 2. Tabeliranje

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 nju se upisuje tekst kojim se tumače podaci u 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. Statistička tabela obično ima kolonu i red za zbir. U tabeli nijedno polje ne sme da ostane prazno; ukoliko za neko polje nema podataka u njega se upisuje crtica ili neki drugi pogodan znak.

Tabela broj ... Naziv tabele ...

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

### Primer 2. Tabelarno prikazivanje – raspodela krvnih grupa

Dvadeset slučajno izabranih pacijenata hematološkog odeljenja Interne A klinike Kliničkog centra u Beogradu pregledom krvi koji je izvršen 24.10.2005. godine pokazalo je sledeće pripadništvo krvnim grupama A, B, AB, 0 i Rh sistema: od 14 Rh-pozitivnih pet bolesnika imalo je 0, troje A, četvero 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 Kliničkog centra u Beogradu

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

### 3. 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. Spajanjem ovih tačaka linijom, dobija se izlomljena, poligonalna linija.

**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 frekvencija.

**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 3. Grafičko prikazivanje – kružni dijagram

Od ukupno 342 uzorka suhomesnatih proizvoda 121 uzorak nije zadovoljavalo osnovne parametre kvaliteta. 12 uzoraka nije čuvano pod adekvatnim uslovima, 11 proizvoda je ocenjeno I klasom, u drugu klasu je svrstano njih 35, III klasi je pripalo 51, dok se u četvrtoj klasi našlo 112. Prikazati tabelarno i grafički strukturu ispitivanih uzoraka.

Komentar:



# Statistika ponovljenih merenja

## 4. Greške koje prate hemijsku analizu

*Nemoguće je izvesti hemijsku analizu potpuno oslobođenu grešaka, odnosno nesigurnosti! (Možemo samo da se nadamo da ćemo greške svesti na minimum i da ćemo njihovu veličinu odrediti sa prihvatljivom tačnošću)*

- Greška – razlika između izmerene vrednosti i „prave” ili „tačne” vrednosti.
- Podatak bez poznatog kvaliteta je bezvredan.

### Neki značajni termini

Pojedinačni rezultati nekog seta merenja prikazuju se preko jedne centralne vrednosti koja je opšta i reprezentativna za celu distribuciju. Njena uloga je da, zanemarujući individualne razlike između podataka skupa, istaknu onu veličinu koja je za sve njih karakteristična i koja može da služi kao sredstvo za upoređivanje raznih serija.

Najčešće korišćene mere centralne tendencije su srednja vrednost i medijana.

#### 1. Aritmetička sredina – Srednja vrednost

Vrednost dobijena deljenjem sume eksperimentalno dobijenih vrednosti sa brojem merenja:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

#### 2. Medijana

Prosečna vrednost centralnog para seta rezultata.

Medijana se uvek upotrebljava kada niz dobijenih podataka sadrži vrednost koja značajno odstupa od niza. Ova vrednost može da ima veliki uticaj na srednju vrednost, a da pritom uopšte ne utiče na medijanu.

#### 3. Preciznost

Veličina koja opisuje reproduktivnost merenja, odnosno bliskost određenog rezultata sa drugim rezultatima dobijenim na potpuno isti način. Preciznost možemo odrediti jednostavnim ponavljanjem merenja.

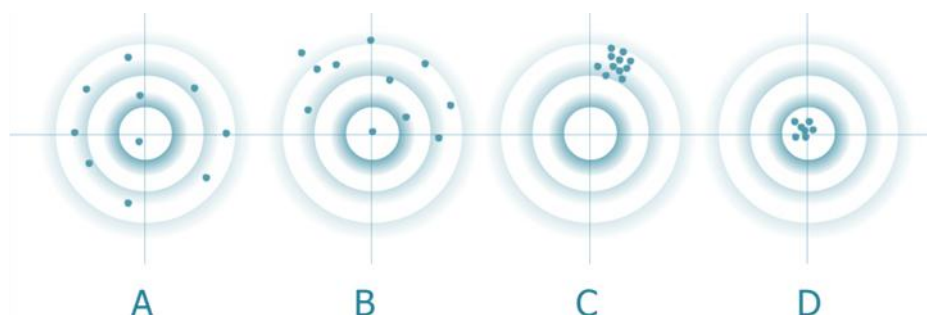
Preciznost seta ponovljenih merenja se izražava pomoću tri veličine: standardne devijacije, varijanse i koeficijenta varijacije (relativne standardne devijacije). Sva tri parametra su funkcija odstupanja pojedinačnih rezultata od srednje vrednosti, tj. devijacija od **srednje vrednosti**

$$d_i = |\bar{x} - x_i|$$

#### 4. Tačnost

Tačnost ukazuje na bliskost određenog rezultata sa stvarnom ili prihvatljivom vrednošću i izražava se preko *greške*.

Tačnost određuje slaganje između rezultata i njihove stvarne vrednosti, dok preciznost opisuje slaganje između više rezultata koji su dobijeni na isti način. Preciznost možemo odrediti jednostavnim ponavljanjem merenja, dok tačnost nikada ne možemo sigurno da odredimo jer nikada ne možemo da znamo stvarnu vrednost merenja.



**Slika 2.** Vežba u streljaštvu: A – tačni i neprecizni rezultati, B – netačni i neprecizni rezultati, C – netačni i precizni rezultati, D – tačni i precizni rezultati

Tačnost se izražava pomoću termina apsolutne i relativne greške.

**Apsolutna greška** – razlika između eksperimentalnog rezultata i prave vrednosti:

$$\Delta x = |\mu - x|$$

$\Delta x$  – apsolutna greška merenja,  $\mu$  – prava vrednost,  $x$  – merena vrednost.

$\Delta x$  ima dimenzije merene veličine.

**Relativna greška** – količnik apsolutne greške i stvarne vrednosti:

$$\delta = \frac{\Delta x}{\mu} = \frac{|\mu - x|}{\mu}$$

Bezdimenzionalna veličina; izražava se u procentima ili promilima kako se ne bi izazvala zabuna ukoliko se rezultat merenja izražava u procentima.

*Apsolutnu grešku uvek treba procenjivati u odnosu na pravu vrednost. Obazrivo sa korišćenjem apsolutne greške kao merila tačnosti!!!*

#### Primer 4. Statistika ponovljenih merenja – apsolutna i relativna greška

Dve različite analitičke metode upotrebljene su za analizu bakra i cinka u mesingu.

Dobijeni su sledeći rezultati:

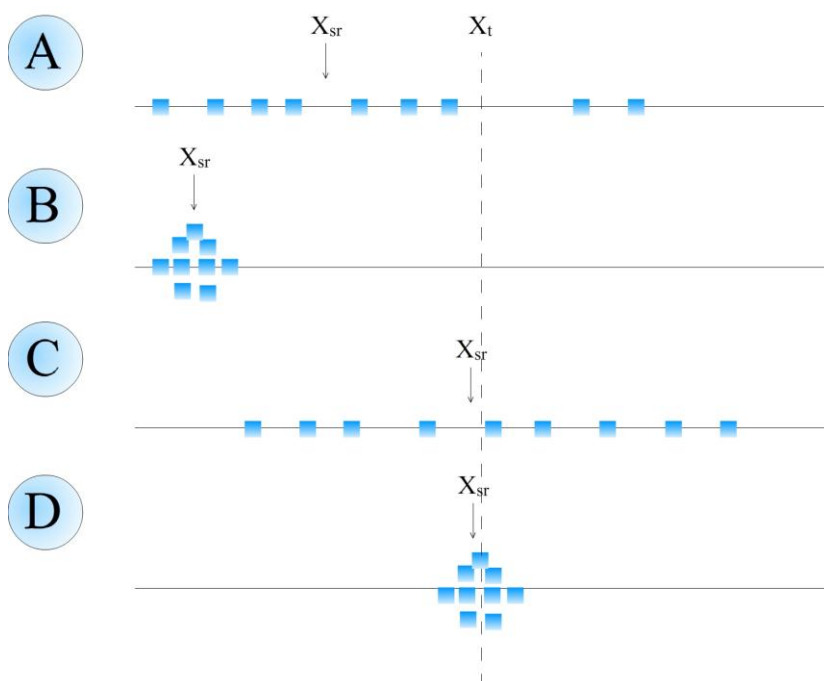
$\mu_{Zn} = 32,6 \%$ ,  $d_{Zn} = 0,3 \%$ , i  $\mu_{Cu} = 7,9 \times 10^{-3} \%$ ,  $d_{Cu} = 0,6 \times 10^{-3} \%$ . Koji je metal određen tačnije?

Komentar:

$$\delta_{Zn} = \frac{0,3}{32,6} \times 100 = 0,92\%$$

$$\delta_{Cu} = \frac{0,6 \times 10^{-3}}{7,9 \times 10^{-3}} \times 100 = 8,00\%$$

### 5. Vrste grešaka u eksperimentalnim podacima



Slika 3. Šematski prikaz tačnosti i preciznosti

#### Slučajne greške – neodređene

Javljaju se pri svakom merenju; nikada ne mogu biti potpuno eliminisane i najčešće su glavni izvor nesigurnosti datog određivanja. Prouzrokuju manje ili više simetričnu raspodelu rezultata oko srednje vrednosti.

#### Sistematske greške – određene

Greške koje utiču na tačnost rezultata.

Prouzrokuju veliku razliku srednje vrednosti seta podataka od stvarne vrednosti.

Mogu se grubo podeliti u sledeće kategorije:

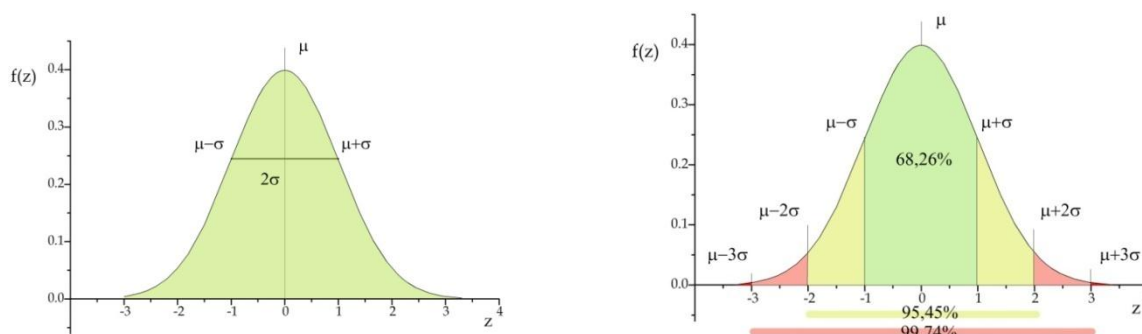
1. Greške instrumenta – uslovljene neispravnim instrumentima, pogrešnom kalibracijom ili korišćenjem instrumenta u neprilagođenim uslovima. Mogu se korigovati kalibracijom instrumenta.
2. Greške metode – uslovljene neidealnim hemijskim ili fizičkim ponašanjem analitičkog sistema. Najteže ih je detektovati i eliminisati. Neki od načina za njihovu detekciju su analiza standardnog referentnog materijala, paralelna analiza sa drugom nezavisnom i pouzdanom metodom, korišćenje slepe probe.
3. Lične greške – uslovljene nepažnjom ili nestručnošću eksperimentatora.

### **Grube greške**

Greške koje dovode do pojave rezultata koji se u velikoj meri razlikuju od ostalih rezultata. Grube greške u rezultatima analiza mogu biti otkrivene statističkim metodama i potom eliminisane.

## 6. Gauss-ova kriva (kriva normalne raspodele)

Ukoliko se broj ponovljenih merenja iz Primera 1. beskonačno poveća i širina intervala se istovremeno redukuje, tada frekvencija merenja dobija zvonast oblik koji se označava kao Gauss-ova ili kriva normalne raspodele.



Slika 4. Gauss-ova raspodela, parametri lokalizacije i rasipanja, osnovna svojstva Gauss-ove krive.

### Svojstva raspodele

- Kriva zvonastog oblika, simetrična oko vrednosti  $\mu$ , proteže se u beskonačnost u oba pravca asimptotski težeći nuli.
- Sve normalne krive imaju istu unutrašnju distribuciju (građu):
 

$\mu \pm 1\sigma$	- $P = 0,6826$ (68,26% podataka)
$\mu \pm 2\sigma$	- $P = 0,9544$ (95,44% podataka)
$\mu \pm 3\sigma$	- $P = 0,9974$ (99,74% podataka)
- Teorijska raspodela određena dvema veličinama:  $\mu$  i  $\sigma$

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Za slučaj standardne promenljive kada su vrednosti korigovane za srednju vrednost i podešene na jediničnu standardnu devijaciju raspodela ima oblik:

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right] \quad z = \frac{x-\mu}{\sigma}$$

Za ovakvu raspodelu kažemo da je standardna normalna raspodela, a promenljiva  $z$  standardna promenljiva.

## Momenti Gauss-ove krive

Ukupna površina ispod Gauss-ove krive data je jednačinom  $\int_a^b f(x)dx = P$ , površina ispod krive za interval  $a < x < b$  odgovara verovatnoći da se veličina  $x$  nađe u datom intervalu.

$n$ -ti moment Gauss-ove krive definisan je relacijom  $m_n = \int_{-\infty}^{+\infty} (x - \mu)^n f(x)dx = 1$

$$n = 1 \quad m_1 = \int_{-\infty}^{+\infty} xf(x)dx = \mu$$

$$n = 2 \quad m_2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx = \sigma^2$$

$n = 3 \quad s = \frac{m_3}{\sigma^3}$   $s$  - iskrivljenje krive  $s = 0$  simetrična,  $s < 0$  rep na levoj strani,  $s > 0$  rep na desnoj strani

$n = 4 \quad k = \frac{m_4}{\sigma^4}$   $k$  - izduženje krive,  $k = 3$  normalna kriva,  $k > 3$ , kriva izduženija,  $k < 3$  kriva spljoštenija

*Set koji se sastoji iz podataka koji prikazuju rezultate velikog broja merenja naziva se **populacija**. Za posmatranja seta podataka se, međutim, često uzima njen reprezentativni deo koji se označava kao **uzorak**.*

*Ako ne postoje sistematske greške, srednja vrednost populacije, označena sa  $\mu$ , predstavlja zapravo stvarnu vrednost merene veličine. Odstupanje rezultata merenja od prave vrednosti označava se sa  $\sigma$ .*

U analitičkoj praksi se obično rade 2-3 paralelne analize istog uzorka na osnovu čega nije moguće tačno odrediti parametre normalne raspodele, tj. mora se izvršiti procena odgovarajućih parametara. Procena predstavlja tačno izračunatu veličinu zasnovanu na preciznim merenjima.

Kada  $n \rightarrow \infty$ , procena mora asimptotski da se približava pravoj vrednosti parametra (**konzistencija**), a rasipanje procene u odnosu na pravu vrednost parametra da bude što je moguće manje (**efikasnost**).

Aritmetička sredina (srednja vrednost) rezultata je najčešće konzistentna i efikasna procena parametra  $\mu$  Gausove raspodele. Medijana je manje efikasna procena prave vrednosti od aritmetičke sredine.

## 7. Mere rasipanja

**Standardna devijacija** - mera odstupanja vrednosti obeležja posmatranja od aritmetičke sredine:

$$s = \sqrt{\frac{\sum (\bar{x} - x_i)^2}{n-1}}$$

**Apsolutna standardna devijacija (ukupna standardna devijacija)**

$$s_{aps} = \sqrt{\frac{\sum_{i=1}^{N_1} (x_i - \bar{x}_1)^2 + \sum_{j=1}^{N_2} (x_j - \bar{x}_2)^2 + \sum_{k=1}^{N_3} (x_k - \bar{x}_3)^2 + \dots}{N_1 + N_2 + N_3 + \dots - N_t}}$$

$N_1$ -broj podataka u setu 1,  $N_2$ -broj podataka u setu 2, itd....  $t$ -ukupan broj setova podataka.

**Varijansa** - prosečno kvadratno odstupanje od aritmetičke sredine :

$$s^2 = \frac{\sum (\bar{x} - x_i)^2}{n-1}$$

**Koeficijent varijacije** – relativna standardna devijacija

$$v = \frac{s}{\bar{x}} \cdot 100$$

**Interval varijacije**, raspon - razlika najveće i najmanje vrednosti obeležja posmatranja.

Najnetačnija mera grupisanja (rasipanja) rezultata oko neke srednje vrednosti.

$$R = x_n - x_1 \quad x_1 < x_2 < \dots < x_n$$

**Kvartili i kvartilno odstupanje** – kvartilno odstupanje označava razmak u kome se oko medijane nalazi polovina svih vrednosti. Gornja kvartilna vrednost ( $Q_3$ ) deli gornju polovinu raspodele vrednosti u dva jednaka dela, a donja kvartilna vrednost ( $Q_1$ ) deli donju polovinu raspodele u dva jednaka dela.

**Interval pouzdanosti** – daje informaciju kolika je bliskost izračunate srednje vrednosti  $\bar{x}$  sa populacionom srednjom vrednošću  $\mu$ , a izražava se kao verovatnoća.

Verovatnoća da se nepoznata populaciona srednja vrednost  $\mu$  nalazi unutar nekog intervala vrednosti označava se kao  $(1-\alpha)\%$ , gde je  $\alpha$  verovatnoća da  $\mu$  **nije** unutar tog intervala. Tipične vrednosti verovatnoće za koje se izračunava interval pouzdanosti su 99%, 95% ili 90%.

#### Interval pouzdanosti za veliki broj podataka

$$L_{1,2} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}}$$

$z$  – standardna promenljiva

$$z = \frac{x - \bar{x}}{s}$$

$(1-\alpha) = 90\%$	$z = 1,65$
$(1-\alpha) = 95\%$	$z = 1,96$
$(1-\alpha) = 99\%$	$z = 2,58$

#### Interval pouzdanosti za mali broj podataka

- $z$  se zamenjuje parametrom  $t$  iz Studentove raspodele (za odgovarajuću verovatnoću i broj stepenai slobode);
- populaciona standardna devijacije  $\sigma$  zamenjuje se standardnom devijacijom uzorka  $s$ .

$$L_{1,2} = \bar{x} \pm t \frac{s}{\sqrt{N}}$$

#### Relativna širina intervala pouzdanosti (%)

$$i = 100 \times \frac{L_2 - L_1}{\bar{x}}$$



## Primer 5. Deskriptivna statistika

Pri određivanju sadržaja olova u uzorku krvi (ppm Pb) dobijeni su sledeći rezultati:

0,752      0,756      0,752      0,751      0,760

Izračunati srednju vrednost, medijanu, standardnu devijaciju, koeficijent varijacije, raspon i 95% interval pouzdanosti.

Komentar:

Najefikasniji način za određivanje mera centralne tendencije i mera varijabilnosti je korišćenje alatke Descriptive Statistics, u okviru Data Analysis ToolPack-a. Odaberite opciju sa padajućeg menija Tools/Data Analysis; starujte komandu Descriptive Statistics.

U polje **Input Range** unesite opseg ćelija između kojih su smešteni vaši podaci.

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 opciju **Summary Statistics**.

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

	A	B	C	D	E	F
1						
2		0,752		Column1		
3		0,756				
4		0,752		Mean	0,7542	
5		0,751		Standard Error	0,00168523	
6		0,760		Median	0,752	
7				Mode	0,752	
8				Standard Deviation	0,003768289	
9				Sample Variance	0,0000142	
10				Kurtosis	0,020829201	
11				Skewness	1,143717867	
12				Range	0,009	
13				Minimum	0,751	
14				Maximum	0,76	
15				Sum	3,771	
16				Count	5	
17				Confidence Level(95,0%)	0,004678948	
18						

Excel ne racuna koeficijent varijacije; pomenuti parametar morate sami da izračunate:

	A	B	C	D	E	F
1						
2		0,752		Column1		
3		0,756				
4		0,752		Mean	0,7542	
5		0,751		Standard Error	0,00168523	
6		0,760		Median	0,752	
7				Mode	0,752	
8				Standard Deviation	0,003768289	
9				Sample Variance	0,0000142	
10				Kurtosis	0,020829201	
11				Skewness	1,143717867	
12	koef					
13	varijanse	=(E8/E4)*100		Range	0,009	
14				Minimum	0,751	
15				Maximum	0,76	
16				Sum	3,771	
17				Count	5	
18				Confidence Level(95,0%)	0,004678948	

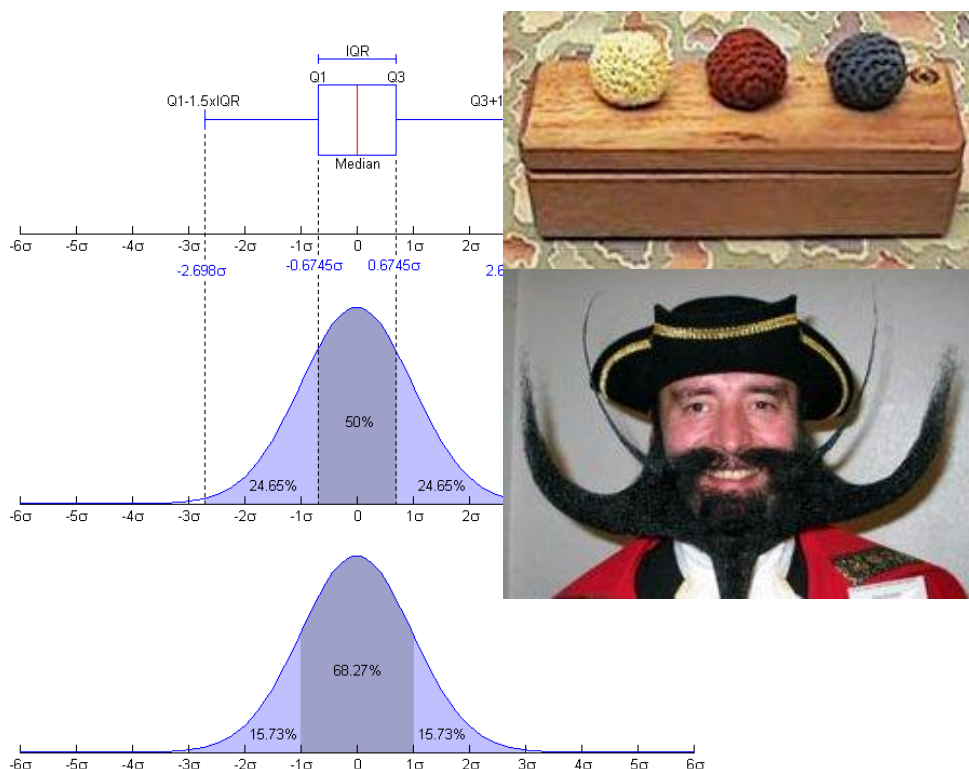
## 8. Grafički prikaz deskriptivne statistike: Box-and-whisker plots

- Grafičko prikazivanje najznačajnijih detalja deskriptivne statistike;
- Duž ose promenljive, u ovom slučaju ordinate, crta se kutija (box) sa gornjom i donjom kvartilnom vrednošću koje definišu vrh i dno kutije. Širina kutije nema nikakvog značaja.
- Gornja i donja vrednost brkova (whiskers) se dobijaju na sledeći način:

$$\text{gornja vrednost} = Q_3 + 1,5 (Q_3 - Q_1)$$

$$\text{donja vrednost} = Q_1 - 1,5 (Q_3 - Q_1)$$

- Vrednosti koje se nalaze van gornje i donje vrednosti brkova se smatraju grubim greškama (spoljašnjim rezultatima).



Slika 5. “Box and whiskers” grafik i kriva normalne raspodele

Najefikasniji način za crtanje „box-and-whisker” grafika u Excel-u je sledeći:

1. Selektujte tabelu sa izračunatim vrednostima  $Q_1$ , donja vrednost brkova, medijana, gornja vrednost brkova,  $Q_3$ , zajedno sa oznakama pojedinih parametara.
2. Odaberite opciju **Insert** → **Line Chart** → **Line with Markers**.
3. U Chart Tools-u odaberite opciju **Design** → **Switch Row/Column**.
4. Desni klik na podatke iz prve serije i odaberite **Format Data Series** → **Line Colour** → **No line** da biste uklonili liniju koja povezuje podatke. Isto ponoviti na podacima iz drugih serija.
5. Selektujte bilo koju seriju i u opciji Chart Tools selektujte **Layout** → **Analysis** → **Lines** → **High-Low Lines**, zatim **Layout** → **Analysis** → **Lines** → **Up/Down Bars** → **Up/Down Bars**.

#### Primer 6. Deskriptivna statistika – spektrofotometrijsko određivanje Fe(III)

Sadržaj Fe(III) u uzorku magnetita određivan je spektrofotometrijski od strane tri analitičara i za 15 ponovljenih merenja svaki od analitičara je dobio sledeće vrednosti apsorbancija:

Analitičar 1	Analitičar 2	Analitičar 3	Analitičar 4
0,3410	0,3334	0,3448	0,3556
0,3350	0,3214	0,3418	0,3496
0,3470	0,3454	0,3478	0,3616
0,3590	0,3694	0,3538	0,3736
0,3530	0,3574	0,3508	0,3676
0,3460	0,3434	0,3473	0,3606
0,3470	0,3454	0,3478	0,3616
0,3460	0,3434	0,3473	0,3606
0,3430	0,3374	0,3458	0,3576
0,3420	0,3354	0,3453	0,3566
0,3560	0,3634	0,3523	0,3706
0,3500	0,3514	0,3493	0,3646
0,3630	0,3774	0,3558	0,3776
0,3530	0,3574	0,3508	0,3676
0,3480	0,3474	0,3483	0,3626

- a) Distribuciju podataka svakog od analitičara predstaviti grafički pomoću histograma frekvencija.
- b) Odrediti parametre deskriptivne statistike i uporediti tačnost i preciznost određivanja sadržaja gvožđa od strane tri analitičara.
- c) Parametre deskriptivne statistike predstaviti grafički pomoću „box-and-whisker” grafika i prokomentarisati ga.

### Primer 7. Deskriptivna statistika – određivanje molarne mase litijuma

Richards i Willard su početkom XX veka određivali atomsku masu litijuma i dobili sledeće rezultate:

Molarna masa,	
Eksperiment	g/mol
1	6,9391
2	6,9407
3	6,9409
4	6,9399
5	6,9407
6	6,9391
7	6,9406

- Odrediti srednju vrednost atomske mase litijuma određenu od strane pomenutih istraživača;
- Odrediti medijanu atomske mase;
- Pretpostavljajući da je kasnije prihvaćena vrednost atomske mase litijuma koja iznosi 6.941 prava vrednost, utvrditi koji je od dva prethodno određena parametra bolja procena prave vrednosti;
- Izračunati apsolutnu i relativnu grešku srednje vrednosti određene od strane Richards-a i Willard-a.

Komentar:

### Primer 8. Deskriptivna statistika i grafičko prikazivanje - histogram

Vršena je kalibracija pipete od 10ml i odgovarajući postupak kalibracije ponavljen 50 puta. Za dobijene vrednosti odrediti srednju vrednost, medijanu, standarnu devijaciju i raspon. Grupisati dobijene vrednosti u intervale i distribuciju podataka predstaviti grafički pomoću histograma.

---

9,988 9,973 9,986 9,980 9,975 9,982 9,986 9,982 9,981 9,990 9,980  
9,989 9,978 9,971 9,982 9,983 9,988 9,975 9,980 9,994 9,992 9,984 9,981 9,987  
9,978 9,983 9,982 9,991 9,981 9,969 9,985 9,977 9,976 9,983 9,976 9,990 9,988  
9,971 9,986 9,978 9,986 9,982 9,977 9,977 9,986 9,978 9,983 9,980 9,984 9,979

---

Komentar:

## 9. Devijacije od Gauss-ovog zakona raspodele

U slučajevima kada dolazi do nesimetrične distribucije podataka (kada su greške određenog znaka verovatnije od grešaka suprotnog znaka), rezultate je potrebno predstaviti log-normalnom distribucijom, jer logaritam promenljive ima normalnu raspodelu, tj. vrednosti  $x_i = \log x_i$  podležu Gaussovoj raspodeli.

$$\log \bar{x}_g = \sum \frac{\log x_i}{N} \quad \bar{x}_g = \sqrt[N]{x_1 + x_2 + \dots + x_n}$$

$$\log s_g = \sqrt{\frac{\sum (\log x_i - \log \bar{x}_g)^2}{N-1}}$$

$$\log L_1 = \log \bar{x}_g + \log t \frac{s_g(A)}{\sqrt{N}} \quad \log L_2 = \log \bar{x}_g - \log t \frac{s_1(B)}{\sqrt{N}}$$

### Primer 10. Log-normalna raspodela

**Primer 9.** Određivani su tragovi nikla u kalijum-hidroksidu metodom emisione spektrofografije i objeni sledeći rezultati:

Broj analize	Ni (%)
1	$1,3 \times 10^{-4}$
2	$2,0 \times 10^{-4}$
3	$5,8 \times 10^{-4}$
4	$8,9 \times 10^{-4}$

Odrediti parametre deskriptivne statistike.

## 10. Standardna devijacija izvedenog rezultata

$$\begin{aligned}
 y = x_1 \pm x_2 & \quad s_y = \sqrt{(s_{x_1})^2 + (s_{x_2})^2} \\
 y = x_1 \times x_2 & \quad \frac{s_y}{y} = \sqrt{\left(\frac{s_{x_1}}{x_1}\right)^2 + \left(\frac{s_{x_2}}{x_2}\right)^2} \\
 y = x^n & \quad \frac{s_y}{y} = \left| \frac{ns_x}{x} \right| \\
 y = f(x) & \quad s_y = \left| s_x \frac{d_y}{d_x} \right|
 \end{aligned}$$

### Primer 10. Standardna devijacija zbira i razlike

Pri nekoj titraciji utrošeno je  $V \text{ cm}^3$  titracionog sredstva. Kolika je standardna devijacija zapremine  $V$ , ako su početna i krajnja zapremina titracionog sredstva očitane sa birete sa standardnim devijacijama od po  $0,02 \text{ cm}^3$ :

- a)  $0,02 \text{ cm}^3$                       b)  $0,03 \text{ cm}^3$                       c)  $0,04 \text{ cm}^3$ ?

Komentar:

### Primer 11. Standardna devijacija količnika i proizvoda

3,4842 g uzorka koji sadrži benzoevu kiselinu ( $122,123 \text{ g/mol}$ ), je rastvoreno i rastvor titrovan rastvorom NaOH. Za titraciju je utrošeno  $41,36 \text{ ml}$   $0,2328 \text{ mol/dm}^3$  NaOH. Izračunati udeo benzoeve kiseline u uzorku, kao i nesigurnost rezultata.

Komentar:

$$\% \text{ HBz} = \frac{41,36 \text{ ml} \times 0,2328 \frac{\text{mmol NaOH}}{\text{ml NaOH}} \times \frac{1 \text{ mmol HBz}}{\text{mmol NaOH}} \times \frac{122,123 \text{ g HBz}}{1000 \text{ mmol HBz}}}{3,842 \text{ g}} \times 100\% = 33,749\%$$

Rezultat je sam po sebi beznačajan ukoliko ne postoji i podatak o njegovom kvalitetu. Zbog toga je neophodno da se naglasi pouzdanost podataka.

- Najbolji način za izražavanje pouzdanosti je prikazivanje intervala pouzdanosti najčešće na 95% nivou pouzdanosti.
- Drugi način je prikazivanje apsolutne standardne devijacije ili koeficijenta varijacije podataka (ovde je dobro da se naglasi broj podataka na osnovu kojih je dobijena vrednost standardne devijacije).
- Treći način je prikazivanje rezultata preko značajnih cifara (najmanje pouzdan).

## 11. Značajne cifre

Po definiciji, značajnim ciframa se nazivaju sve cifre od prve sleva, različite od nule, do poslednje zapisane cifre sdesna, pri čemu se nule koje slede iz množioca  $10^n$  ne uzimaju u obzir. Obuhvataju sve sigurne cifre plus prvu nesigurnu cifru..

### Primer 12. Značajne cifre

Odrediti broj značajnih cifara u sledećim brojevima:

15,13	4	
1513	4	
15013	5	
0,001513	4	
0,15130	5	
15000	2	$(1,5 \times 10^4)$

Komentar:

## 12. Pravila zaokrugljivanja brojeva

Određeni broj možemo zapisati kao:  $n, d_1 d_2 d_3 d_4 d_5 d_6 d_k d_{k+1} \dots d_m$  pri čemu je  $n$  ceo broj a  $d_1, d_2, \dots, d_n$  odgovarajuće decimale. Ako želimo da dati broj zaokruglimo na  $k$  decimala onda postupamo na jedan od sledećih načina:

- Ako je  $d_{k+1} < 5$ , tada zaokrugljeni broj ima oblik  $n, d_1 d_2 d_3 d_4 d_5 d_6 d_k$
- Ako je  $d_{k+1} > 5$ , tada zaokrugljeni broj ima oblik  $n, d_1 d_2 d_3 d_4 d_5 d_6 d_{k+1}$
- Ako je  $d_{k+1} = 5$  i bar jedna od decimala iza njega je različita od nule tada se postupa kao pod b)
- Ako je  $d_{k+1} = 5$  i svi ostale decimale, ako ih uopšte ima, su nule onda razlikujemo dva slučaja: ako je  $d_k$  parno postupa se kao u slučaju a) ako je  $d_k$  neparno postupa se kao u slučaju b).



### 13. Značajne cifre u rezultatu nekih aritmetičkih operacija

#### Sabiranje i oduzimanje –

- apsolutna nesigurnost rezultata ne sme da bude manja od najnesigurnije vrednosti koja podleže sabiranju, odnosno oduzimanju.

#### Množenje i deljenje –

- relativna nesigurnost rezultata treba da bude jednaka onoj koju ima komponenta sa najvećom relativnom nesigurnošću;
- proizvod i količnik treba da sadrže onoliko značajnih cifara koliko ima i član sa najmanjim brojem značajnih cifara.

#### Logaritmovanje –

- rezultat treba da sadrži onoliko cifara desno od decimalnog zareza koliko ima značajnih cifara broj čiji se logaritam traži.

#### Stepenovanje i korenovanje –

- rezultat treba da sadrži onoliko značajnih cifara koliko ima broj koji se stepenuje ili korenuje.

#### Primer 13. Značajne cifre i aritmetičke operacije – brza procena

Prikazati pravilno krajnje rezultate:

- a)  $12,2 + 0,365 + 1,04 =$
- b)  $4,16 \cdot 10^{-3} + 1,724 \cdot 10^{-1} + 3,86 \cdot 10^{-4} =$
- c)  $128,6 / 0,069 =$
- d)  $\log 3,67 \cdot 10^{-3} =$
- e)  $23,15^2 =$
- f)  $1,13 \times 10^4 + 11\,000 =$
- g)  $\frac{0,00480 \times 2,300}{0,2084} =$
- h)  $\frac{416,00 \times 0,000187}{\sqrt{73,84}} =$
- i)  $\log (783,6 \times 1654) =$

Komentar: