Merenje i greške merenja

Značajne cifre

U civilizaciji u kojoj živimo koristimo se dekadnim sistemom brojeva (0,1,2,3,4,5,6,7,8,9). U ovom sistemu brojevi nose sa sobom neku informaciju. Jedan sto – 1; dva polaznika - 2; tri jabuke – 3; ništa novca na računu – 0. Dakle, date cifre su nam od nekog značaja (značajne cifre) jer nose informaciju o merenoj veličini. Broj nula je jedinstven jer pored informativne vrednosti njime se služimo da ostale cifre dovedemo na određenu poziciju. Tako jedna nula iza broja označava desetice, dve nule stotine, tri hiljade itd. U ovom slučaju broj nula ne predstavlja značajnu cifru.
Pitanje koliko neki broj ima značajnih cifara se upravo svodi na rešavanje ove dileme, kada je nula značajna cifra, a kada ne.

Šta ćete raditi danas?

- Merenje i greške merenja.
- Statistika ponovljenih merenja.
- Statistički testovi.

Primer 1. Značajne cifre

Odrediti broj značajnih cifara u sledećim brojevima:

15,13	4
1513	4
151300	4
15013	5
0,001513	4
0,15130	5

U prvom broju sve četiri cifre su značajne jer tako naveden rezultat pretpostavlja da je zaista izmereno 15,13 jedinica merene vrednosti, sa tačnošću na stotim delovima. Isto važi i za drugi broj. Kod trećeg broja očigledno je da je uloga dve poslednje nule da dovedu cifre 1513 na određenu poziciju, te osim toga nemaju nikakv drugi značaj – nisu značajne cifre. Kod broja 15013 jasno je da se nule izmedju drugih cifara koje nisu nule moraju uzeti kao značajne cifre te stoga ovaj broj ima pet značajnih cifara. U broju 0,001513 tri nule služe da cifre 1513 dovedu na određeni položaj te stoga nemaju značaj i ovaj broj zato ima 4 značajne cifre. Kod poslednejg slučaja treba obratiti naročitu pažnju. Nula na kraju decimalnog zapisa mora biti smatrana značajnom cifrom, u ovom slučaju ona služi da demonstrira preciznost sa kojom je izvršeno merenje (na 10^{-5}). Klasičan primer je navođenje koncentracije nekog standardnog rastvora C=0,1000mol/dm^3. Nule iza zareza ukazuju na analitičku preciznost sa kojom je ova koncentracija određena.

Značajne cifre obuhvataju tzv. sigurne cifre i nesigurnu cifru (tj. cifru na kojoj se nalazi greška).

Primer 2. Značajne cifre

Navesti sigurne cifre i nesigurnu cifru za svaki od brojeva iz primera 1.

15,13	sigurne: 1,5,1	nesigurna:3
1513	sigurne: 1,5,1	nesigurna:3
151300	sigurne: 1,5,1	nesigurna:3
15013	sigurne: 1,5,0,1	nesigurna:3
0,001513	sigurne: 1,5,1	nesigurna:3
0,15130	sigurne: 1,5,1,3	nesigurna:0

Primer 3. Značajne cifre

U sledećem broju odrediti značajne cifre (sigurne i sumnjivu cifru).

12,400483217±0,001

Ovakav zapis nosi informaciju i o grešci sa kojom je izmerena neka vrednost. Vidi se da je greška (±0,001) na hiljaditim delovima, pa je jedino smisleno zadržati cifre sve do ovog mesta iza decimalnog zareza. Zbog toga su 1,2,4,0,0 značajne cifre. Poslednja cifra - nula je sumnjiva jer je opterećena greškom dok su sve preostale cifre nepotrebne i kao takve treba ih izbaciti iz konačnog zapisa rezultata (12,400±0,001).

Pravila zaokruživanja brojeva

Nekakav broj možemo zapisati kao: $n,d_1d_2d_3d_4d_5d_6d_kd_{k+1}...d_m$ pri čemu je n ceo broj a $d_1,d_2,...,d_n$ odgovarajuće decimale. Ako želimo da dati broj zaokružimo na k decimala onda postupamo na jedan od sledećih načina:

a) Ako je $d_{k+1}<5$, tada zaokruženi broj ima oblik $n,d_1d_2d_3d_4d_5d_6d_k$
b) Ako je $d_{k+1}>5$, tada zaokruženi broj ima oblik $n,d_1d_2d_3d_4d_5d_6d_k+1$
c) Ako je $d_{k+1}=5$ i bar jedan od decimala iza njega je različit od nule tada se postupa kao pod b)
d) Ako je $d_{k+1}=5$ i svi ostali decimali, 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).

Primer 4. Zaokruživanje brojeva

Zaokružiti na tri decimale sledeće brojeve:
1,45549; 2,45977; 3,455501; 4,9985; 5,999500

Rešenje: 1,445; 2,460; 3,456; 4,998; 6,000

Primer 5. Zaokruživanje brojeva

Zaokružiti na tačnost do na stotinu sledećebrojeve:
145549,88; 245977; 34550,1; 49850; 599950,0

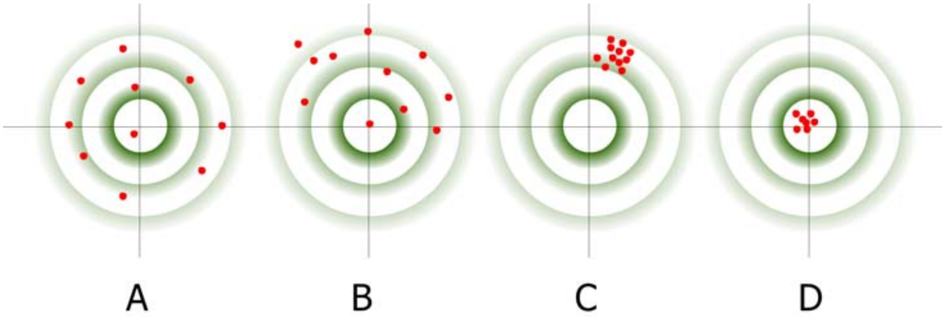
Rešenje: 145500; 345600; 49800; 60000

Greške merenja

Preciznost – bliskost određenog rezultata sa drugim rezultatima dobijenim na potpuno isti način; veličina koja opisuje reproduktivnost merenja.

Tačnost – bliskost određenog rezultata stvarnoj ili prihvatljivoj vrednosti.

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 1. Vežba u streljaštvu
Diskutujte rezultate strelaca A, B, C i D.

Tačnost se izražava pomoću termina apsolutne i relativne greške.

<u>*Apsolutna greška*</u> – razlika između eksperimentalnog rezultata i prave vrednosti:

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

Δ – apsolutna greška merenja, *μ* - prava vrednost, *x* – merena vrednost.
d ima dimenzije merene veličine.

Δ uvek treba procenjivati u odnosu na pravu vrednost. Voditi računa prilikom korišćenja apsolutne greške kao merila tačnosti!!!

<u>*Relativna greška*</u> – količnik apsolutne greške i stvarne vrednosti:

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

Bezdimenzionalna veličina; izražava se u procentima ili promilima kkao se ne bi izazvala zabuna ukoliko se izražava greška merenja datih u procentima

Primer 6. Tačnost dve analitičke metode

Dve različite analitičke metode upotrebljene su za analizu dva metala u leguri. Dobijeni su sledeći rezultati:
μ_1 = 32,6 %, d_1 = 0,3 %,
μ_1 = 7,91 x 10^{-3} %, d_1 = 0,63 x 10^{-3} %.
Koji je metal određen tačnije?

$$\delta_1 = \frac{0{,}3}{32{,}6} \times 100 = 0{,}92\% \qquad \delta_2 = \frac{0{,}63 \times 10^{-3}}{7{,}91 \times 10^{-3}} \times 100 = 8{,}00\%$$

Funkcije rasipanja rezultata oko srednje vrednosti poput **standardne devijacije, varijanse i koeficijenta varijanse**, koriste se za opisivanje preciznosti rezultata Sa njima ćete se upoznati na nekom od sledećih termina

<u>Kada je *μ* nepoznato:</u>

$$\delta x = \frac{\Delta x}{x}$$

$$\Delta x = |\bar{x} - x|$$

Δ – devijacija od srednje vrednosti – mera preciznosti pojedinačnog određivanja

Izračunavanje apsolutnih i relativnih grešaka izvedenog rezultata

Zbir i razlika: apsolutne greške se sabiraju, relativne se izračunavaju iz odnosa $\delta_y = \frac{dy}{y}$

$$y = x_1 + x_2 - x_3 \qquad dy = dx_1 + dx_2 + dx_3$$

Proizvod i količnik: relativne greške se sabiraju, apsolutne izračunavaju iz odnosa $d_y = e_y \times y$

$$y = \frac{x_1 \times x_2}{x_3}$$

$$dy = \frac{\partial y}{\partial x_1} dx_1 + \frac{\partial y}{\partial x_2} dx_2 + \frac{\partial y}{\partial x_2} dx_3 \Rightarrow \delta_y = \delta x_1 + \delta x_2 + \delta x_3$$

Greške koje prate hemijsku analizu

Slučajne greške – neodređene
Greške koje utiču na preciznost merenja.
Prouzrokuju manje ili više simetričnu raspodelu rezultata oko srednje vrednosti.
Slučajne greške u rezultatima analiza mogu biti eliminisane statističkim metodama.
Sistematske greške – određene
Greške koje utiču na tačnost rezultata.
Prouzrokuju veliku razliku srednje vrednosti seta podataka od stvarne vrednosti.
Grube greške
Greške koje dovode do pojave rezultata koji se u velikoj meri razlikuju od ostalih rezultata.
Grube greške u rezultatima analiza mogu biti eliminisane statističkim metodama.

Grešku uvek treba zaokruživati na jednu značajnu cifru. Nema smisla greške prikazivati sa više cifara jer neodređenost (nepreciznost) demonstrirana prvom cifrom isključuje preostale. Opravdano je grešku prikazati sa više cifara ukoliko će se dati podatak koristiti u daljim statističkim proračunima pa bi se na taj način sprečio nepotreban gubitak informacija.

Sve međurezultate ne treba zaokruživati. Konačan rezultat se zaokružuje na tačnost iskazanu prethodno izračunatom greškom.

Zadaci

1. Izračunati apsolutnu grešku krajnjeg rezultata:

   a) y = 6,75 (±0,03) + 0,843 (±0,001) – 7,021(±0,001)

   b) y = 67,1 (±0,3) · 1,03 (±0,02)

   c) % Cr

$$y = \frac{40{,}64\ (\pm 0{,}04)\,\mathrm{cm}^3 \times 0{,}1027\ (\pm 0{,}001)\,\mathrm{mmol/cm}^3 \times \left(\frac{51{,}996\ (\pm 0{,}001)\,\mathrm{mg/mmol}}{3} \right)}{346{,}4\ (\pm 0{,}2)\,\mathrm{mg}} \times 100$$

   d) $y = \frac{[(B-A)-(D-C)] \times E}{(G-F)} \times 100$

   A = 1,38 (±0,02) B = 29,82 (±0,02)
   C = 0,89 (±0,02) D = 10,35 (±0,02)
   E = 0,00589 (±0,00006) F = 25,8943 (±0,0001)
   G = 26,7854 (±0,0001)

2. Izračunati maksimalnu vrednost relativnih grešaka merenja analitičkom vagom sledećih masa: 100g, 1g, 0,1g. Pri merenju koje mase je greška najveća?

Statistika ponovljenih merenja

Mere centralne tendencije

- *Centralna tendencija je težnja ka okupljanju podataka skupa oko jedne centralne vrednosti, koja je opšta i reprezentativna za celu distribuciju.*
- *Njihova 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.*

ARITMETIČKA SREDINA – *Srednja vrednost*
(procena parametra μ)
Vrednost dobijena deljenjem sume eksperimentalno dobijenih vrednosti sa brojem merenja :

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

MEDIJANA
(procena parametra μ)
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.

MODA
(procena parametra μ)
Vrednost koja je u nizu rezultata najčešće postignuta.

GEOMETRIJSKA SREDINA
(procena parametra μ)
Prosečna mera brzine nekih promena :

$$G = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}$$

HARMONIJSKA SREDINA
(procena parametra μ)
Koristi se kada želimo da dobijemo prosek nekih odnosa :

$$H = \frac{n}{\sum \frac{1}{x}}$$

Mere varijabilnosti

- Daju informacije o različitim odstupanjima u statističkom skupu.

INTERVAL VARIJACIJE – RASPON
Razmak od najmanje do najveće vrednosti obeležja posmatranja.
Najnetačnija mera grupisanja rezultata oko neke srednje vrednosti.

$$R = x_n - x_1 \qquad x_1 < x_2 < \ldots < x_n$$

STANDARDNA DEVIJACIJA
Mera odstupanja vrednosti obeležja posmatranja od aritmetičke sredine :

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

Apsolutna standardna devijacija (ukupna standardna devijacija)

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

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

VARIJANSA
Prosečno kvadratno odstupanje od aritmetičke sredine :

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

KOEFICIJENT VARIJANSE – RELATIVNA STANDARDNA DEVIJACIJA
Količnik standardne devijacije i aritmetičke sredine :

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

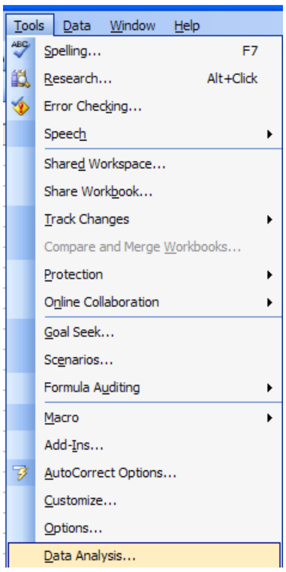
Primer 1. Statistika ponovljenih merenja

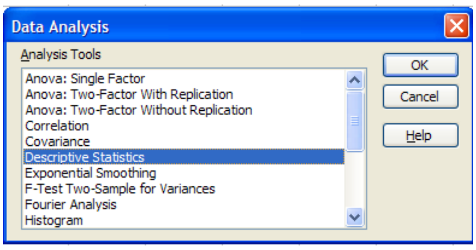
Pri određivanju sadržaja olova u uzorku krvi dobijeni su sledeći rezultati:

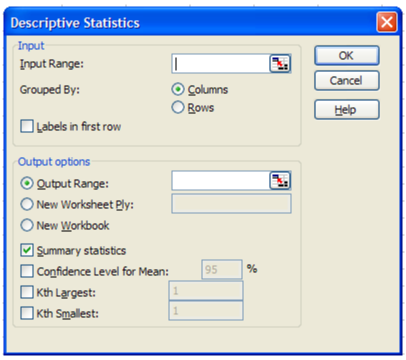
0,752; 0,756; 0,752; 0,751; 0,760 ppm Pb.

Izračunati srednju vrednost, medijanu, standardnu devijaciju, koeficijent varijacije, raspon.

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.**

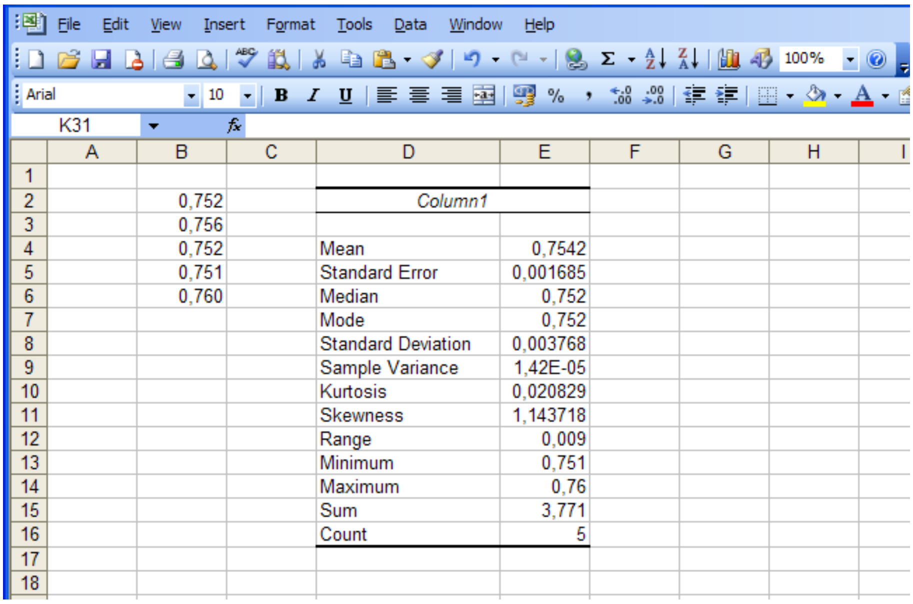






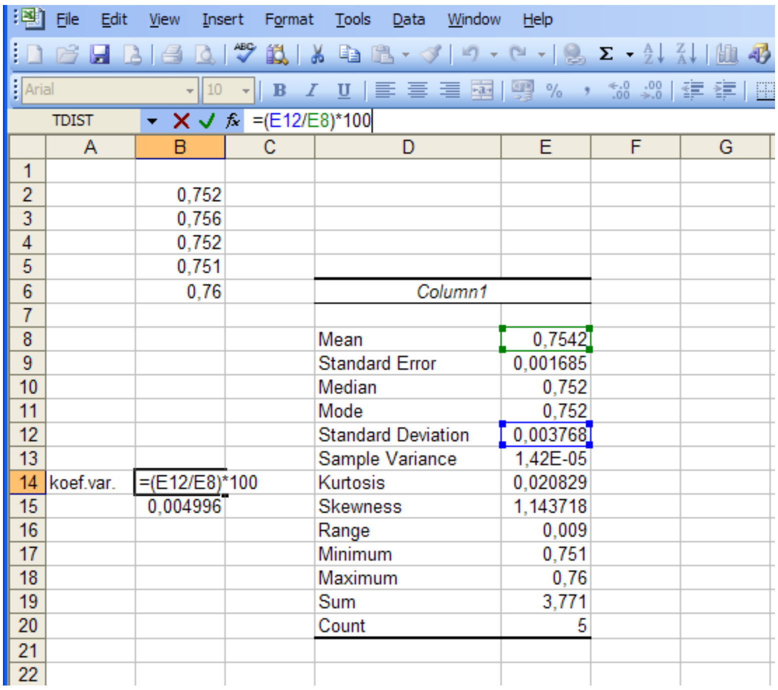
*Set koji se sastoji iz podataka koji prikazuju rezultate velikog broja merenja naziva se populacija. Ako ne postoje sistematske greške, srednja vrednost populacije, označena sa **μ**, predstavlja zapravo stvarnu vrednost merene veličine. Odstupanje rezultata merenja od prave vrednosti označava se sa **σ**. Za posmatranja seta podataka se, međutim, često uzima njen deo koji se označava kao uzorak.*

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



	A	B	C	D	E
1					
2		0,752		*Column1*	
3		0,756			
4		0,752		Mean	0,7542
5		0,751		Standard Error	0,001685
6		0,760		Median	0,752
7				Mode	0,752
8				Standard Deviation	0,003768
9				Sample Variance	1,42E-05
10				Kurtosis	0,020829
11				Skewness	1,143718
12				Range	0,009
13				Minimum	0,751
14				Maximum	0,76
15				Sum	3,771
16				Count	5

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



TDIST =(E12/E8)*100

	A	B	C	D	E
1					
2		0,752			
3		0,756			
4		0,752			
5		0,751			
6		0,76		*Column1*	
7					
8				Mean	0,7542
9				Standard Error	0,001685
10				Median	0,752
11				Mode	0,752
12				Standard Deviation	0,003768
13				Sample Variance	1,42E-05
14	koef.var.	=(E12/E8)*100		Kurtosis	0,020829
15		0,004996		Skewness	1,143718
16				Range	0,009
17				Minimum	0,751
18				Maximum	0,76
19				Sum	3,771
20				Count	5

Zadaci

3. Richards i Willard su početkom dvadesetog veka određivali atomsku masu litijuma i dobili sledeće rezultate:

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

a) Odrediti srednju vrednost atomske mase litijuma određenu od strane pomenutih istraživača;

b) Odrediti medijanu atomske mase;

c) 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;
d) Izračunati apsolutnu i relativnu grešku srednje vrednosti određene od strane Richards-a i Willard-a.

4. Za svaki set rezultata merenja izračunati srednju vrednost, medijanu, standardnu devijaciju, koeficijent varijacije, raspon:

A	B	C	D	E	F
3,5	70,24	0,812	2,7	70,65	0,514
3,1	70,22	0,792	3,0	70,63	0,503
3,1	70,10	0,794	2,6	70,64	0,486
3,3		0,900	2,8	70,21	0,497
2,5			3,2		0,472

5. Prihvaćene vrednosti za setove podataka iz prethodnog zadatka su sledeće:
set A-3,0; set B-70,05; set C-0,830; set D-3,4; set E-70,05; set F-0,525.
Za srednju vrednost svakog seta, izračunati:
a) apsolutnu grešku,
b) relativnu grešku u ppt-u.

6. Data metoda ima koeficijent varijacije ≤0,5%. Analizom uzorka tom metodom dobijeni su sledeći rezultati: 40,12; 40,15 i 40,55. Kako se poslednji rezulatat učinio sumnjivim, urađena su dva dodatna određivanja i dobijeni su rezultati 40,20 i 40,39. Uporedite reproduktivnost oba seta rezultata sa poznatim koeficijentom varijacije.

7. Radi utvrđivanja efikasnosti dijete koja je prepisana pacijentu koji boluje od dijabetesa, vršeno je određivanje koncentracije glukoze spektrofotometrijskom analitičkom metodom. Dobijeni su sledeći rezultati. Izračunati ukupnu standardnu devijaciju metode.

Vreme	Konc.glukoze, mg/L
Mesec I	1108, 1122, 1075, 1099, 1115, 1083, 1100
Mesec II	992, 975, 1022, 1001, 991
Mesec III	788, 805, 779, 822, 800
Mesec IV	799, 745, 750, 774, 777, 800, 758

8. Analizom K^+ jona u nekoliko uzoraka hrane dobijeni su sledeći rezultati:

Uzorak	Procenat K^+
1	5,15; 5,03; 5,04; 5,18; 5,20
2	7,18; 7,17; 6,97
3	4,00; 3,93; 4,15; 3,86
4	4,68; 4,85; 4,79; 4,62
5	6,04; 6,02; 5,82; 6,06; 5,88

Uzorci su nasumično izabrani iz iste populacije.
a) Odrediti srednju vrednost i standardnu devijaciju za svaki uzorak.
b) Odrediti ukupnu standardnu devijaciju.
c) Zašto je ovo bolja procena σ od standardne devijacije pojedinih uzoraka?

9. Analiziran je sadržaj zaostalog šećera u šest boca vina iz iste serije i dobijeni sledeći rezultati:

Boca	Procenat (w/v) šećera
1	0,99; 0,84; 1,02
2	1,02; 1,13; 1,17; 1,02
3	1,25; 1,32; 1,13; 1,20; 1,12
4	0,72; 0,77; 0,61; 0,58
5	0,90; 0,92; 0,73

6	0,70; 0,88; 0,72; 0,73

a) Izračunati standardnu devijaciju za svaki set podataka.
b) Izračunati apsolutnu standardnu devijaciju metode.

Svojstva Gauss-ove raspodele

Svojstva raspodele

- Kriva zvonastog oblika, simetrična oko vednosti μ, proteže se u beskonačnost u oba pravca asimptotski težeci 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: μ i σ

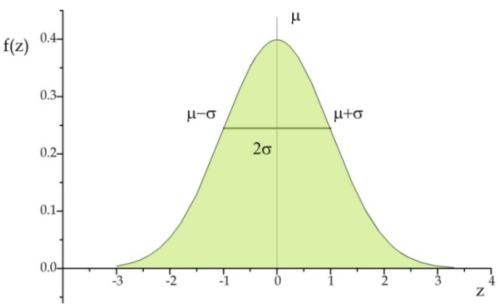
$$y = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

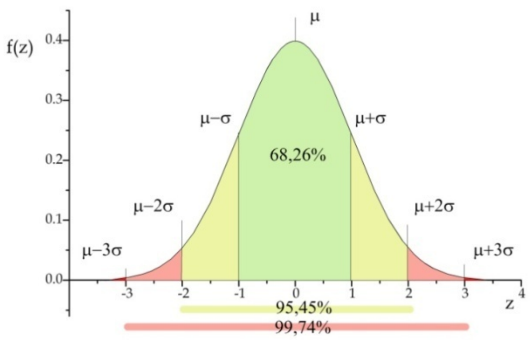
Za slučaj standardne promenljive kada su vrednosti korigovane za srednju vrednost i podešene na jediničnu standardnu devijaciju raspodela ima oblik:

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

$$z = \frac{x-\mu}{\sigma}$$

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





Momenti Gauss-ove krive

Ukupna površina ispod Gauss-ove krive data je jednačinom $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, površina ispod krive za interval *a<x<b* odgovara verovatnoći da se veličina *x* nadje 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 $m_1 = \int_{-\infty}^{+\infty} x f(x)\,dx = \mu$

n=2 $m_1 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx = \sigma^2$

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

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

Zadaci

10. Naći površinu za oblast ispod normalne krive koja leži između datih Z vrednosti:
    a) Z=0 i Z=2,37
    b) Z=0 i Z=-1,94
    c) Z=-1,85 i Z=1,85
    d) Z=-0,76 i Z=1,13
    e) Z=0 i Z=3,09
    f) Z=-2,77 i Z=-0,96
11. Naći oblast ispod normalne krive koja pada ispod –Z ili iznad +Z.
    a) Z=1,73
    b) Z=-2,41 i Z=2,41
    c) Z=2,55
    d) Z=-3 i Z=3
12. Naći Z vrednosti koje odgovaraju sledećoj verovatnoći: 95%, 80%, 50%, 30%, 20%.
13. Visina učenika u predadolescentnoj fazi normalno raspoređena oko 170 cm sa standardnom devjacijom od 15cm. Koji procenat populacije se očekuje:
    a) 100-120cm
    b) 90-130cm
    c) 150-170cm
    d) 170-190cm
    e) >200cm
14. Nivo holesterola koji je uzet iz populacije srednjoškolaca ima srednju vrednost od 195 sa standardnom devijacijom 10 za muskarce i 185 sa standardnom devijacijom od 12 za devojke.
    a) Koji je nivo holestorla kod najviših 5% muškaraca, a koji kod žena?
    b) Koji je nivo holeseterola kod najnižih 5% muškaraca a koji kod žena?
    c) Koji procenat muškaraca a koji žena će imati nivo holesterola veći od 180?
15. Dužina života se pokorava normalnoj raspodeli i za muškarce u severnoj Americi iznosi 55±10 godina. Kolika je verovatnoća da ako ste mlad i zdrav 25-to godišnjak umrete za dve godine. Kolika je verovatnoća da ćete ako preživite svoj 27. rođendan dočekati osamdeseti?

Centralna granična teorema

A. Srednja vrednost raspodele srednjih vrednosti uzoraka μ_{xsr} skoro je identična srednjoj vrednosti populacije μ.
B. Standardna devijacija srednjih vrednosti uzoraka izračunata po formuli $\sqrt{\Sigma(\bar{x}-\mu)^2/(n-1)}$ je veoma bliska standardnoj grešci srednje vrednosti $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.
C. Bez obzira kakva je raspodela populacije, raspodela srednjih vrednosti uzoraka je uvek približna normalnoj (sličnost raste sa porastom veličine uzorka – *n*).

Greška srednje vrednosti

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$

Razlikuje se od drugih grešaka po tome što ilustruje grešku uzorkovanja. Njena veličina određena je voljom eksperimentatora, tj. veličinom uzorka. Ona takođe ilustruje i činjenicu da je srednja vrednost tačnija od bilo kog pojedinačnog rezultata i to za $\sqrt{n}$ puta.

<u>**Standardna devijacija krajnjeg rezultata:**</u>

$y = x_1 \pm x_2 \qquad \sigma_y = \sqrt{(\sigma_{x_1})^2 + (\sigma_{x_2})^2}$

$y = x_1 \times x_2 \qquad \frac{\sigma_y}{y} = \sqrt{\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2}$

$y = x^n \qquad \frac{\sigma_y}{y} = \left|\frac{n\sigma_x}{x}\right|$

$y = f(x) \qquad \sigma_y = \left|\sigma_x \frac{d_y}{d_x}\right|$

Zadaci

16. Data je funkcija $y = k\frac{a \times b}{c \times d}$. Čemu je jednaka relativna standardna devijacija ove funkcije?

17. Pri nekoj titraciji utrošeno je V 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 cm^3:
    a) 0,02 cm^3 b) 0,03 cm^3 c) 0,04 cm^3?

18. Proizvod rastvorljivosti AgCl iznosi $1{,}8 \times 10^{-10}$ sa standardnom devijacijom $0{,}1 \times 10^{-10}$. Kolika je standardna devijacija izračunate rastvorljivosti ove soli u vodi?

19. Standardna devijacija prečnika kruga iznosi ±0,02cm. Kolika je standardna devijacija izračunate zapremine kruga prečnika 2,15cm?

20. Odrediti apsolutnu standardnu devijaciju i koeficijent varijacije za rezultate sledećih izračunavanja. Zaokružiti svaki rezultat tako da on sadrži samo značajne cifre. Brojevi u zagradama su standardne devijacije.
    a) y = 5,75(±0,03) + 0,833(±0,001) - 8,02(±0,001) = -1,438
    b) y = 18,97(± 0,04) + 0,0025(±0,0001) + 2,29(±0,08) = 21,2625
    c) y = 66,2(±0,3) x 1,13(±0,02) x 10^{-17} = 7,4806 x 10^{-16}
    d) y = 251(±1) x $\frac{860(\pm 2)}{1.673(\pm 0{,}006)}$ = 129 025,70
    e) y = $\frac{157(\pm 6) - 59(\pm 3)}{1220(\pm 1) + 77(\pm 8)}$ = 7,5559 x 10^{-2}
    f) y = $\frac{1.97(\pm 0{,}01)}{243(\pm 3)}$ = 8,106996 x 10^{-3}
    g) y = $[4.73(\pm 0{,}03) \times 10^{-4}]^3$
    h) y = $[2.145(\pm 0.002)]^{1/4}$

21. Pri volumetrijskom određivanju analita A, dobijeni podaci i njihove standardne devijacije su sledeće:

Početno očitavanje birete	0,23 ml	0,02 ml
Konačno očitavanje birete	8,76 ml	0,03 ml
Masa uzorka	50,0 mg	0,2 mg

    Iz navedenih podataka izračunati koeficijent varijacije konačnog rezultata za % A koji je dobijen korišćenjem sledeće jednačine (ekvivalentna masa analita A iznosi 63,54 g/mol i može da se tretira kao da nema nesigurnost):

$$\%A = \frac{zapremina\ titranta \times ekvivalentna\ masa}{masa\ uzorka} \times 100$$

22. 3,4842 g uzorka koji sadrži benzoevu kiselinu (122,123 g/mol), je rastvoreno i rastvor titrovan rastvorom NaOH. Za titraciju je utrošeno 41,36ml 0,2328mol/dm^3 NaOH. Izračunati maseni udeo benzoeve kiseline u uzorku, kao i nesigurnost rezultata.
23. U čemu je značaj centralne granične teoreme?
24. Ukoliko se greška srednje vrednosti smanji dva puta, koliko puta treba povećati veličinu uzorka?
25. Ukoliko je uzet uzorak 5 puta veći od prethodnog, koliko puta se promenila standardna greška srednje vrednosti?

Interval pouzdanosti

- Interval pouzdanosti daje informaciju kolika je bliskost izračunate srednje vrednosti $\overline{x}$ sa populacionom srednjom vrednošću μ, a izražava se kao verovatnoća.
- Verovatnoća da se nepoznata populaciona srednja vrednost μ nalazi unutar nekog intervala vrednosti označava se kao $(1-\alpha)\%$, gde je α verovatnoća da μ **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

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

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

$(1-\alpha) = 90\%$ *z* = 1,65.

Interval pouzdanosti za mali broj podataka

- *z* se zamenjuje sa *t* iz Studentove raspodele (za odgovarajuću verovatnoću i broj stepena slobode);
- populaciona standardna devijacije *σ* zamenjuje se sa standardnom devijacijom uzorka *s*.

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

Relativna širina intervala pouzdanosti

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

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 na 90% ili 95% nivou pouzdanosti.***
- ***Drugi način je prikazivanje apsolutne standardne devijacije ili koeficijenta varijacije podataka (ovde je dobro da se naglasi sa koliko podataka je rađeno).***
- ***Treći način je prikazivanje rezultata preko značajnih cifara.***

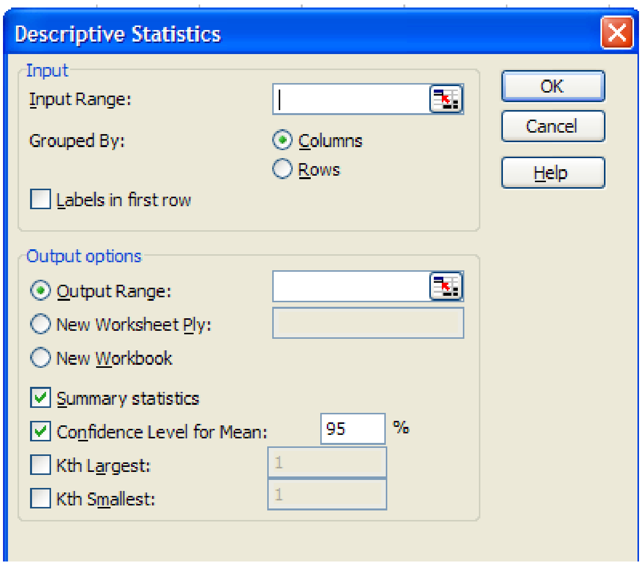
Primer 1. Statistika ponovljenih merenja

Merenjem pH vrednosti nekog pufera dobijeni su sledeći rezultati:

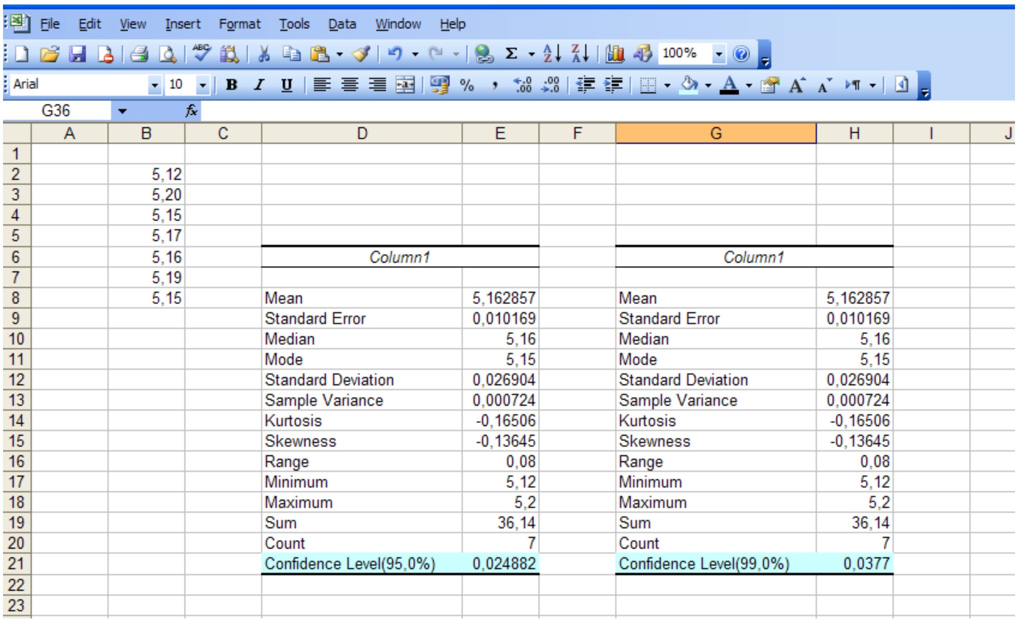
5,12 5,20 5,15 5,17 5,16 5,19 5,15

Izračunati 95% i 99% interval pouzdanosti prave vrednosti pH.

Otvorite alatku **Descriptive Statistics**, u okviru Data Analysis-a i odaberite opciju **Confidence Level for Means,** kao i nivo pouzdanosti na kome želite da izračunate interval pouzdanosti. U polje **Input Range** unesite opseg ćelija između kojih su smešteni vaši podaci.



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



	A	B	C	D	E	F	G	H	I	J
1										
2		5,12								
3		5,20								
4		5,15								
5		5,17								
6		5,16		*Column1*			*Column1*			
7		5,19								
8		5,15		Mean	5,162857		Mean	5,162857		
9				Standard Error	0,010169		Standard Error	0,010169		
10				Median	5,16		Median	5,16		
11				Mode	5,15		Mode	5,15		
12				Standard Deviation	0,026904		Standard Deviation	0,026904		
13				Sample Variance	0,000724		Sample Variance	0,000724		
14				Kurtosis	-0,16506		Kurtosis	-0,16506		
15				Skewness	-0,13645		Skewness	-0,13645		
16				Range	0,08		Range	0,08		
17				Minimum	5,12		Minimum	5,12		
18				Maximum	5,2		Maximum	5,2		
19				Sum	36,14		Sum	36,14		
20				Count	7		Count	7		
21				Confidence Level(95,0%)	0,024882		Confidence Level(99,0%)	0,0377		
22										
23										

Obratite pažnju na činjenicu da vam excel izračunava vrednost izraza $t\frac{s}{\sqrt{N}}$, interval pouzdanosti je, međutim, $\bar{x} \pm t\frac{s}{\sqrt{N}}$; ne zaboravite da ga izrazite na ovaj način, sa donjom i gornjom granicom.

Za pomenuti primer interval pouzdanosti je

$L_{1,2}$ = 5,16 ± 0,03 za 95% nivo pouzdanosti i

$L_{1,2}$ = 5,16 ± 0,04 za 99% nivo pouzdanosti

Zadaci

26. Odrediti 95% interval pouzdanosti za srednju vrednost koncentracije glukoze za prvi mesec u zadatku 7. Pretpostaviti da je vrednost ukupne standardne devijacije dobra aproksimacija σ.
27. Koliko je merenja potrebno izvršiti tokom prvog meseca, u zadatku 7, da bi povećali 95% interval pouzdanosti na 1100,3 ± 10,0 mg/L glukoze?
28. Prilikom određivanja sadržaja alkohola u krvi jednog pacijenta dobijeni su sledeći podaci:

    % C_2H_5OH: 0,084; 0,089; 0,079.

    Izračunati 95% interval pouzdanosti pretpostavljajući

    a) da su tri dobijena rezultata jedini indikatori preciznosti metode;
    b) na osnovu prethodnog iskustva, da je standardna devijacija metode 0.005% C_2H_5OH.
29. Za setove podataka iz zadatka 8 odrediti:
    a) 95% interval pouzdanosti;
    b) 95% interval pouzdanosti pretpostavljajući da je s dobra aproksimacija σ i ima sledeće vrednosti: set A-0,20; set B-0,070; set C-0,0090; set D-0,30; set E-0,15; set F-0,015.
30. Određivanjem sadržaja bakra u gorivima atomskom apsorpcionom metodom dobijena je zajednička standardna devijacija od 0,30 μg Cu/ml. Analizom nekog motornog ulja dobijena vrednost za sadržaj bakra iznosi 8,53 μg Cu/ml. Izračunati 90% i 99% interval pouzdanosti rezultata zasnovanom na
    a) jednom merenju;
    b) srednjoj vrednosti četiri merenja;
    c) srednjoj vrednosti šesnaest merenja.

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 Gausovoj raspodeli.

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

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

$$\log L_1 = \log \bar{x}_g + \log t \frac{s_{g(A)}}{\sqrt{N}} \qquad \log L_2 = \log \bar{x}_g - \log t \frac{s_{1(B)}}{\sqrt{N}}$$

Statistički testovi

- Prilikom izvođenja statističkih testova postoje određeni koraci kojih se treba pridržavati da bi zaključak bio pouzdan:
  1. Postaviti nultu hipotezu
  2. Izabrati nivo pouzdanosti
  3. Odrediti veličinu uzorka
  4. Izabrati statistički test za testiranje hipoteze
  5. Utvrditi kritičnu vrednost za odabrani statistički test
  6. Prikupiti podatke
  7. Izračunati statističku veličinu za odabrani statistički test
  8. Doneti statistički zaključak
  9. Izraziti statistički zaključak.

Za testiranje hipoteze koriste se parametrijski i neparametrijski testovi. Parametrijske metode koriste se za upoređivanje dve ili više grupa podataka i zasnivaju se na pretpostavci da su podaci normalno raspodeljeni. Ove metode se uvek zasnivaju na teoriji verovatnoće i uvek se u njima pojavljuje potreba za ocenjivanjem pojedinih parametara (srednje vrednosti, standardne devijacije ili varijanse). Međutim, kada ne može sa sigurnošću da se utvrdi da li je raspodela jedne grupe podataka normalna, izračunavanje pojedinih parametara i primena parametrijskih metoda daju vrlo nepouzdane zaključke. U tim slučajevima se primenjuju neparametrijske metode, koje se zasnivaju na pretpostavci da postoji bilo koja verovatnoća raspodele.

Q- i G-test

- Eliminisanje „spoljnih” rezultata, vrednosti koje se izdvajaju u odnosu na ostale.

Dixon-ov test (Q-test) – za male uzorke (3-7)

$$Q = \frac{|sumnjiva\ vrednost - najbliža\ vrednost|}{(\max vrednost - \min vrednost)}$$

Grubb-ov test (G-test)

$$G = \frac{|sumnjiva\ vrednost - \bar{x}|}{s}$$

Nulta hipoteza, H_0 - *posmatrana vrednost nije posledica grube greške*.

Primer 1. *Q*- i *G*-test

Određivan je sadržaj cinka u nekom uzorku i dobijene sledeće vrednosti:

16,84 16,86 16,91 16,93 17,61%.

Da li je poslednji rezultat posledica grube greške?

$$Q = \frac{|17{,}61 - 16{,}93|}{(17{,}61 - 16{,}84)} = 0{,}883$$

Izračunata vrednost Q upoređuje se sa kritičnom vrednošću; ukoliko je $Q_{izr} > Q_{krt}$ nulta hipoteza se odbacuje, tj. posmatrana vrednost jeste posledica grube greške i treba je odbaciti; ukoliko je $Q_{izr} < Q_{krt}$ nulta hipoteza se zadržava, tj. posmatrana vrednost nije posledica grube greške.

Kritična vrednost Q za veličinu uzorka *n*=5 iznosi 0,717. Pošto je izračunata vrednost Q veća od kritične vrednosti, nulta hipoteza se odbacuje, tj. posmatrana vrednost jeste posledica grube greške.

Zadaci

31. Na koliko merenja se može primeniti Q, a na koliko G test? Da li se Q i G test može primeniti na merenja koja ne pripadaju istoj populaciji?

32. Na osnovu Q i G testa proveriti da li neki od rezultata predstavljaju spoljne vrednosti pri nivou pouzdanosti od 95%?

25.34	25.64	25.55	25.44	25.98	25.52		
4.13	4.12	4.12	4.14	4.13	4.15	4.13	4.25
183.12	183.56	183.33	183.99	183.55	184.01	183.25	183.15
15.3	15.4	15.6	15.4	15.3	15.2	15.9	

33. U sledećem setu podataka koristeći se Q testom za nivou poverenja od 95% izračunati kritične vrednosti ispod i iznad kojih bi rezultat bio odbačen kao spoljašnja vrednost:

11.2	11.3	11.2	11.4	11.3	11.2	11.1	11.3
150.18	150.22	150.19	150.16	150.21	150.19	150.17	150.19
0.5558	0.5552	0.5554	0.5554	0.5553	0.5556	0.5555	0.553

34. Pri određivanju himozinskog broja spektrofotometrijski su izmerene sledeće vrednosti apsorbancije plavog jod - skrobnog inkluzionog kompleksa:

    0,341; 0,335; 0,347; 0,359; 0,353; 0,346; 0,347; 0,346; 0,343; 0,342; 0,356; 0,350; 0,363; 0,353; 0,348.

    Proveriti dati set rezultata na eventualno prisustvo spoljnih vrednosti (Q ili G test?). Izračunati: srednju vrednost, medijanu, geometrijsku sredinu, varijansu, standardnu devijaciju, relativnu standardnu devijaciju, interval i interval pouzdanosti za α = 0,05 ovih merenja.

F-test

F-test se koristi za utvrđivanje postojanja slučajnih grešaka.

- Utvrđivanje da li je razlika između varijansi dva uzorka značajna.

$$F = \frac{s_1^2}{s_2^2} \qquad s_1 \rangle s_2$$

Nulta hipoteza, H_0 - *razlika koja se javlja između standardnih devijacija je posledica slučajnih grešaka*

Primer 2. *F*-test

Sadržaj titana u čeliku određivan je atomsko-apsorpcionom spektrometrijom u dve laboratorije. Dobijeni su sledeći rezultati:

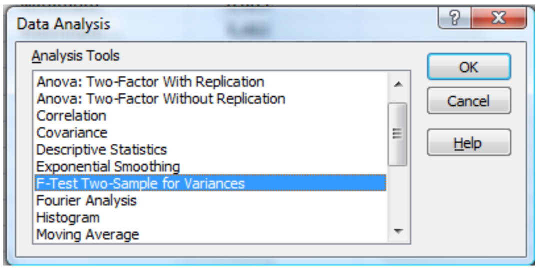
Lab 1: 0,529; 0,490; 0,489; 0,521; 0,486; 0,502

Lab 2: 0,470; 0,448; 0,463; 0,449; 0,482; 0,454; 0,477; 0,409.

Da li postoji statistički značajna razlika u preciznosti u radu između ove dve laboratorije?

Nakon provere prisustva „spoljnih” rezultata, postojanje razlike u preciznosti utvrđuje se primenom F-testa. U okviru Data Analysis ToolPack-a postoji alatka **F-Test Two-Sample for Variances**. Odaberite opciju sa padajućeg menija Tools/Data Analysis; starujte komandu F-Test Two-Sample for Variances.
U polje **Input Range,** kao **Variable 1** unesite opseg ćelija između kojih su smešteni vaši podaci sa većom standardnom devijacijom, odnosno varijansom; **Variable 2** su vam podaci sa manjom varijansom.
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.

*Kada ne znamo da li će ishod određenog posmatranja biti pozitivan ili negativan, test mora da pokrije obe mogućnosti ⇨ **<u>dvosmerni test</u>** (two-tailed). Kada nas interesuje samo jedan ishod određene analize ⇨ **<u>jednosmerni test</u>** (one-tailed).*



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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2		0,529	0,470		*Column1*			*Column1*			F-Test Two-Sample for Variances			
3		0,490	0,448											
4		0,485	0,463		Mean	0,502167		Mean	0,4565			*Variable 1*	*Variable 2*	
5		0,521	0,449		Standard Error	0,0077		Standard Error	0,008113		Mean	0,4565	0,502167	
6		0,486	0,482		Median	0,496		Median	0,4585		Variance	0,000527	0,000356	
7		0,502	0,454		Mode	#N/A		Mode	#N/A		Observations	8	6	
8			0,477		Standard Deviation	0,018862		Standard Deviation	0,022947		df	7	5	
9			0,409		Sample Variance	0,000356		Sample Variance	0,000527		F	1,480103		
10					Kurtosis	-1,70769		Kurtosis	2,189574		P(F<=f) one-tail	0,344173		
11					Skewness	0,666826		Skewness	-1,26569		F Critical one-tail	4,875872		
12					Range	0,044		Range	0,073					
13					Minimum	0,485		Minimum	0,409					
14					Maximum	0,529		Maximum	0,482					
15					Sum	3,013		Sum	3,652					
16					Count	6		Count	8					
17														
18														

Izračunata F-vrednost upoređuje se sa kritičnom vrednošću; ukoliko je $F_{izr} > F_{krt}$ nulta hipoteza se odbacuje, tj. razlika koja se javlja između standardnih devijacija ne može biti objašnjena samo uticajem slučajnih grešaka; ukoliko je $F_{izr} < F_{krt}$ nulta hipoteza se zadržava, tj. razlika koja se javlja između standardnih devijacija je posledica slučajnih grešaka. Kada ovaj test radite u excel-u morate voditi računa o tome da excel daje kritičnu vrednost samo za jednosmerni test; ukoliko vi radite dvosmerni test kritičnu vrednost očitavate iz tablica.
U Primeru 2. primenjujemo dvosmerno testiranje. Kritična vrednost parametra F za dvosmerni test iznosi 6,853. Pošto je izračunata vrednost parametra F manja od kritične, nulta hipoteza se zadržava, tj. ne postoji statistički značajna razlika u preciznosti u radu između ove dve laboratorije.

Zadaci

35. U rudarsko-topioničarskom basenu Bor ispitivan je sastav jalovine koja zaostaje posle prerade rude bakra na sadrzaj olova. U cilju formiranja standardne metode testirane su dve metode: prva, zasnovana na spektrofotometrijskom određivanju kompleksa olova sa ditizonom i druga zasnovana na polarografskom određivanju. Dobijeni su sledeći rezultati:

Spektrofotometrijski	0.153	0.162	0.158	0.154	0.157	0.157	0.160	0.152
Polarografski	0.160	0.158	0.159	0.161	0.160	0.158	0.159	0.159

Da li postoji razlika u preciznosti između ove dve metode?

36. Pri određivanju sadržaja vitamina E u uzorku nekog ulja standardnom voltametrijskom i novom FIA metodom dobijeni su sledeći rezultati:

Standardna metoda: 32,1; 32,3; 31,9; 32,1; 32,0; 32,1; 31,8 %
FIA metoda: 31,9; 31,8; 31,7; 31,8; 31,6; 31,9; 31,8 %.

a) Da li se razlika među ovim rezultatima može pripisati isključivo slučajnim greškama na nivou značajnosti od P = 0,05?
b) Da li je FIA metoda preciznija od standardne voltametrijske metode?

<u>t - test</u>

<u>t-test se koristi za utvrđivanje postojanja sistematskih grešaka</u>

Koristi se u sledećim slučajevima:
a) Kada se upoređuje srednja vrednost grupe podataka sa pravom vrednošću (određivanje tačnosti)
b) Kada se upoređuju srednje vrednosti dve grupe podataka
c) Kod paralelnih određivanja.

<u>Upoređivanje eksperimentalno određene srednje vrednosti sa pravom vrednošću</u>

<u>Nulta hipoteza, H_0</u> – između posmatrane i poznate, prave vrednosti, ne postoji druga razlika od one koja može da se pripiše slučajnim greškama.

$$t = \frac{(\bar{x} - \mu) \times \sqrt{N}}{s}$$

Dobijena vrednost se upoređuje sa kritičnom *t*-vrednošću, koja se za dati nivo pouzdanosti i broj stepeni slobode, očitava u tablici. Ako vrednost *t* prelazi određenu kritičnu vrednost nulta hipoteza se odbacije. U suprotnom ne postoje dokazi za postojanje sistematske greške (ovo ne znači da sistematska greška ne postoji već samo da ona nije izražena).

Primer 3. ***t*-test**

U standardnom uzorku seruma, u kome je sadržaj natrijuma 135,4 mmol/dm^3, određivan je natrijum plamenom fotometrijom i dobijeni sledeći rezultati:

134,6 137,5 135,6 135,9 135,8 136,2 135,8 134,2 136,7
137,6 135,7 134,9 135,8 136,5 136,0 mmol/dm^3.

Pokazati kakva je tačnost metode.

$N = 15$, $\bar{x} = 135{,}9 \text{mmol/dm}^3$, $s = 0{,}9367 \text{mmol/dm}^3$, $\mu = 135{,}4 \text{mmol/dm}^3$

$$t = \frac{(135{,}9 - 135{,}4) \times \sqrt{15}}{0{,}9367} = 2{,}067$$

$t_{krit} = 2{,}145$ $t \langle t_{krit}$

H_0 se zadržava, tj. između izračunate srednje vrednosti i deklarisanog sadržaja seruma nema statistički značajne razlike, odnosno metoda daje tačne vrednosti.

<u>Upoređivanje dve eksperimentalno određene srednje vrednosti</u>

<u>Nulta hipoteza, H_0</u> – dve metode daju jednake rezultate, tj. $\bar{x}_1 - \bar{x}_2$ se ne razlikuje mnogo od nule.

1) Standardne devijacije dve metode su bliske

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \qquad s = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{(N_1 + N_2 - 2)} \qquad \nu = N_1 + N_2 - 2$$

2) Standardne devijacije dve metode se značajno razlikuju

$$t=\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{s_1^{\,2}}{N_1}+\frac{s_2^{\,2}}{N_2}}} \qquad \nu=\frac{\left(\frac{s_1^{\,2}}{N_1}+\frac{s_2^{\,2}}{N_2}\right)^2}{\left(\frac{s_1^{\,4}}{N_1^{\,2}(N_1-1)}+\frac{s_2^{\,4}}{N_2^{\,2}(N_2-1)}\right)}$$

Primer 4. ***t*-test**

Pri određivanju sadržaja kalaja u hrani, uzorci sa hlorovodoničnom kiselinom su refluktovani različito vreme. Neki od dobijenih rezultata su sledeći:

Vreme refluktovanja (min)	Sadržaj kalaja (mg/kg)
30	55, 57, 59, 56, 56, 59
75	57, 55, 58, 59, 59, 59

Da li dužina refluktovanja ima uticaja na ishod analize?

Nakon provere prisustva „spoljnih" rezultata, bliskost standardnih devijacija utvrđuje se dvosmernim F-testom.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2		55	57		*Column1*			*Column1*			F-Test Two-Sample for Variances			
3		57	55											
4		59	58		Mean	57		Mean	57,83333			*Variable 1*	*Variable 2*	
5		56	59		Standard Error	0,68313		Standard Error	0,654047		Mean	57	57,83333	
6		56	59		Median	56,5		Median	58,5		Variance	2,8	2,566667	
7		59	59		Mode	59		Mode	59		Observations	6	6	
8					Standard Deviation	1,67332		Standard Deviation	1,602082		df	5	5	
9					Sample Variance	2,8		Sample Variance	2,566667		F	1,090909		
10					Kurtosis	-1,78571		Kurtosis	1,239669		P(F<=f) one-tail	0,463129		
11					Skewness	0,384181		Skewness	-1,35376		F Critical one-tail	5,050329		
12					Range	4		Range	4					
13					Minimum	55		Minimum	55					
14					Maximum	59		Maximum	59					
15					Sum	342		Sum	347					
16					Count	6		Count	6					
17														
18														

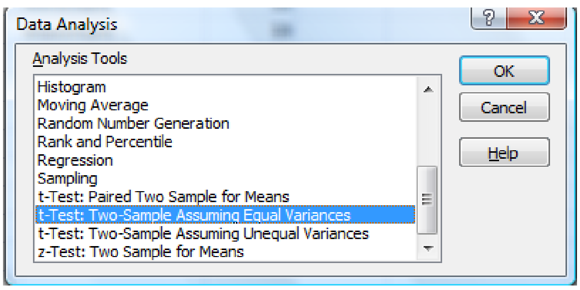
Kritična vrednost parametra F za dvosmerni test iznosi 7,146. Pošto je izračunata vrednost parametra F manja od kritične, nulta hipoteza se zadržava, tj. standardne devijacije su bliske. Za upoređivanje srednjih vrednosti ova dva seta podataka se zbog toga koristi t-test koji pretpostavlja da su standardne devijacije bliske.

U okviru Data Analysis ToolPack-a postoji alatka **t-Test: Two-Sample Assuming Equal Variances**. Odaberite opciju sa padajućeg menija Tools/Data Analysis; starujte komandu t-Test: Two-Sample Assuming Equal Variances.

U polje **Input Range,** kao **Variable 1** unesite opseg ćelija između kojih je smešten jedan set vaših podataka; **Variable 2** vam je drugi set podataka.

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.

Za upoređivanje dva seta podataka čije se standardne devijacije statistički značajno razlikuju koristi se alatka **t-Test: Two-Sample Assuming Unequal Variances**, koja se nalazi u okviru Data Analysis ToolPack-a.



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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2		55	57		*Column1*			*Column1*			F-Test Two-Sample for Variances			
3		57	55											
4		59	58		Mean	57		Mean	57,83333			*Variable 1*	*Variable 2*	
5		56	59		Standard Error	0,68313		Standard Error	0,654047		Mean	57	57,83333	
6		56	59		Median	56,5		Median	58,5		Variance	2,8	2,566667	
7		59	59		Mode	59		Mode	59		Observations	6	6	
8					Standard Deviation	1,67332		Standard Deviation	1,602082		df	5	5	
9					Sample Variance	2,8		Sample Variance	2,566667		F	1,090909		
10					Kurtosis	-1,78571		Kurtosis	1,239669		P(F<=f) one-tail	0,463129		
11					Skewness	0,384181		Skewness	-1,35376		F Critical one-tail	5,050329		
12					Range	4		Range	4					
13					Minimum	55		Minimum	55					
14					Maximum	59		Maximum	59		t-Test: Two-Sample Assuming Equal Variances			
15					Sum	342		Sum	347					
16					Count	6		Count	6			*Variable 1*	*Variable 2*	
17											Mean	57	57,83333	
18											Variance	2,8	2,566667	
19											Observations	6	6	
20											Pooled Variance	2,683333		
21											Hypothesized Mean	0		
22											df	10		
23											t Stat	-0,88113		
24											P(T<=t) one-tail	0,199464		
25											t Critical one-tail	1,812461		
26											P(T<=t) two-tail	0,398928		
27											t Critical two-tail	2,228139		
28														

$$t = -0{,}88 \qquad t_{krit} = 2{,}23 \qquad t \langle t_{krit}$$

H_0 se zadržava, tj. vreme trajanja refluktovanja nema uticaja na količinu pronađenog kalaja.

Uporedni t-test (Paired t-test)

- Upoređivanje dve metode ispitivanjem uzoraka koji sadrže različite količine analita.
- U ovom slučaju ne može da se upotrebi test za upoređivanje dve srednje vrednosti jer on ne razdvaja varijaciju prouzrokovanu određenom metodom od varijacije prouzrokovane razlikama između uzoraka.
- Ne može da se koristi kada je širok opseg koncentracija, jer se zasniva na pretpostavci da bilo koja greška, slušajna ili sistematska, je nezavisna od koncentracije. *Alternativa – linearna regresiona analiza.*

$$t = \frac{\bar{d}\sqrt{N}}{s_d}$$

$\bar{d}$ - srednja vrednost razlike parova

s_d – standardna devijacija razlike

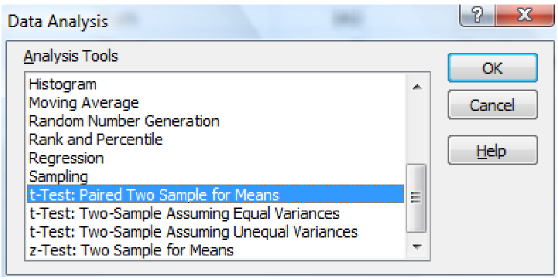
Primer 5. ***t*-test**

Podaci u tabeli pokazuju koncentraciju gvožđa (μg/dm^3) određenu dvema različitim metoda u svakom od četiri uzorka.

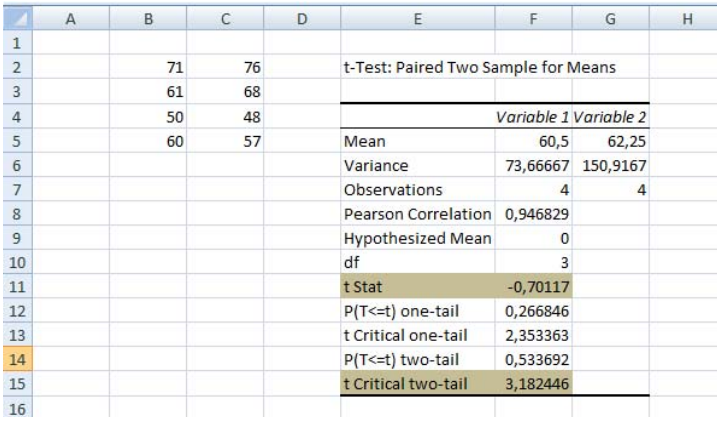
Uzorak	Oksidacija	Ekstrakcija
1	71	76
2	61	68
3	50	48
4	60	57

Utvrditi da li se srednje vrednosti dobijene različitim metodama značajno razlikuju.

U okviru Data Analysis ToolPack-a postoji alatka **t-Test: Paired Two-Sample for Means**.
Odaberite opciju sa padajućeg menija Tools/Data Analysis; starujte komandu t-Test: Paired Two-Sample for Means.
U polje **Input Range,** kao **Variable 1** unesite opseg ćelija između kojih je smešten jedan set vaših podataka; **Variable 2** vam je drugi set podataka.
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.



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



	A	B	C	D	E	F	G	H
1								
2		71	76		t-Test: Paired Two Sample for Means			
3		61	68					
4		50	48			*Variable 1*	*Variable 2*	
5		60	57		Mean	60,5	62,25	
6					Variance	73,66667	150,9167	
7					Observations	4	4	
8					Pearson Correlation	0,946829		
9					Hypothesized Mean	0		
10					df	3		
11					t Stat	-0,70117		
12					P(T<=t) one-tail	0,266846		
13					t Critical one-tail	2,353363		
14					P(T<=t) two-tail	0,533692		
15					t Critical two-tail	3,182446		
16								

$t = -0{,}70 \qquad t_{krit} = 3{,}18 \qquad t \langle t_{krit}$

H_0 se zadržava, tj. dve metode ne daju značajno velike razlike srednjih vrednosti.

Zadaci

37. U cilju poređenja dva metoda za određivanje hroma u zelenom čaju urađeno je pet merenja:
    Metod 1: $\bar{x} = 1{,}48;\ s = 0{,}28;\ n = 5$
    Metod 1: $\bar{x} = 2{,}33;\ s = 0{,}31;\ n = 5$
    Da li postoji statistička razlika između rezultata dobijenih pomoću ove dve metode?

38. 10,00 cm^3 rastvora NaOH, C = 0,1 M titrovano je rastvorom HCl iste koncentracije. Dobijeni su sledeći rezultati:
    9,88; 10,18; 10,23; 10,39; 10,25 cm^3. Ovakvi rezultati ukazuju na prisustvo:
    a) pozitivne sistematske greške   b) negativne sistematske greške
    c) samo slučajnih grešaka u ovim merenjima.

39. Koncentracija tiola u krvnom lizatu kod grupe obolelih od reumatoidnog artritisa i zdravih pacijenata iznosi:

Zdravi	1.84	1.92	1.94	1.92	1.85	1.91	2.07
Reumatoidna grupa	2.81	4.06	3.62	3.27	3.27	3.76	

Da li se koncentracija tiola statiistički razlikuje u ove dve grupe?
Da li je koncentracija tiola veća u grupi sa reumatoidnim artritisom?

40. U sledećoj tabeli data je koncentracija norepinefrina (μmol/g kreatinina) u urinu zdravih doborovoljaca u njihovim ranim dvadesetim godinama

Momci	0.48	0.36	0.20	0.55	0.45	0.46	0.47	0.23
Devojke	0.35	0.37	0.27	0.29				

Da li postoji razlika izmežu polova u nivou norepinefrina?

41. Sledeći rezultati se odonose na određivanje hroma u različitim biljnim uzorcima koristeći se dvema metodama. Za svaki biljni materijal odredi da li se rezultati dobijeni različitim metodama značajno razlikuju?

Materijal	Metod	$\bar{x}$	s	*n*

Iglice bora	I	2.15	0.26	5
	II	2.45	0.14	5
Lišće	I	5.12	0.80	5
	II	7.27	0.44	5
Vodene biljke	I	23.08	2.63	5
	II	32.01	4.66	5

42. Pošto je otklonjen kvar na Pašen – Rungeovom cilindru ICP uredjaja u valjaonici bakra – Sevojno, pristupilo se određivanju Hg u standardnim uzorcima legura. Hemičar koji je analizu ponovio nekoliko puta posmatra rezultate koje je imao pre i nakon popravke:

Pre popravke	100.12	100.15	100.22	100.14
Posle popravke	101.12	100.22	100.55	101.23

Da li postoji razlika između ova dva seta rezultata na nivou poverenja od 95%? Da li je preciznost nakon popravke difrakcione rešetke smanjena?

43. Nova enzimska metoda za određivanje peroksida u vodi poređena je sa konvencionalnom, permanganometrijskom. Obe metode primenjene su na na uzorke peroksida za farmaceutske namene:

Broj uzorka	Enzimska metoda	Permanganometrijska
1	31.1	32.6
2	29.6	31.0
3	31.0	30.3

Da li postoji značajna razlika između rezultata dobijenih jednom i drugom metodom?

44. Šest analitičara je napravilo po šest određivanja paracetamola u tabletama iz istog pakovanja. Rezultati su prikazani ispod:

Analitičar	Sadržaj paracetamola (% m/m)					
A	84.32	84.51	84.63	84.61	84.64	84.51
B	84.24	84.25	84.41	84.13	84.00	84.30
C	84.29	84.40	84.68	84.28	84.40	84.36
D	84.14	84.22	84.02	84.48	84.27	84.33
E	84.50	83.88	84.49	83.91	84.11	84.06
F	84.70	84.17	84.11	84.36	84.61	83.81

Da li postoji značajna razlika između srednjih vrednosti dobijenih od strane različitih analitičara?

45. Sledeći rezultati se donose na koncentraciju albumina (g/dm^3) u krvnom serumu šesnaest zdravih odraslih jedinki:

37, 39, 37, 42, 39, 45, 42, 39, 44, 44, 40, 39, 45, 47, 47, 43, 41

Prvih osam rezultata se odnosi na muškarce a preostalih osam na žene. Da li se srednje vrednosti nivoa albumina razlikuju značajno za muškarce i žene?