

**Увод у статистичку обраду
резултата у хемији**

III део

Analiza varijanse - ANOVA

Analiza varijanse (ANalysis Of VAriance) predstavlja statističko upoređivanje srednjih vrednosti više setova rezultata; npr. upoređivanje srednjih vrednosti koncentracija proteina u rastvorima čuvanim pod različitim uslovima, upoređivanje srednjih vrednosti rezultata dobijenih za koncentraciju analita primenom nekoliko različitih metoda; upoređivanje srednjih vrednosti rezultata dobijenih od strane nekoliko eksperimentatora i sl.

Nulta hipoteza:

Svi uzorci potiču iz iste populacije sa srednjom vrednošću μ i varijansom σ_0^2 .

Analiza varijanse razdvaja ukupan varijabilitet u podacima na dva dela: na varijabilitet unutar grupa (u ovom slučaju to je varijabilitet ponovljenih merenja – repetabilnost) i na varijabilitet između grupa (npr. među laboratorijama, učesnicama međulaboratorijskog ispitivanja).

Ukoliko postoji h grupa i n ponovljenih merenja u svakoj od njih, onda su ova dva varijabiliteta data kao sume kvadrata odstupanja sledećim formulama:

$$SS_{\text{unutar}} = \sum \sum (x_{ij} - \bar{x}_j)^2$$

$$SS_{\text{izmedju}} = \sum \sum (\bar{x}_j - \bar{x})^2$$

Kako je u svakoj grupi po n merenja, to je broj stepeni slobode ponovljenih merenja za svaku pojedinačnu grupu (set rezultata) $n - 1$, odnosno za ukupan varijabilitet svih grupa $h(n - 1)$. Broj stepeni slobode za varijabilitet među grupama je $h - 1$.

Shodno tome F vrednost se definiše kao:

$$F = \frac{SS_{\text{izmedju}} / h - 1}{SS_{\text{unutar}} / h(n - 1)} = \frac{\sum \sum (x_{ij} - \bar{x}_j)^2 / h - 1}{\sum \sum (\bar{x}_j - \bar{x})^2 / h(n - 1)}$$

Ukoliko varijabilitet izađu grupa premašuje vrednost prosečnog varijabiliteta unutar grupa kaže se da među grupama rezultata postoji statistički značajna razlika. U tu svrhu se koristi jednosmerni Fisher-ov F test.

Analiza varijanse potvrđuje ili opovrgava postojanja razlike među rezultatima, ali ne upućuje na njen uzrok. Da bi se ustanovilo koji od setova rezultata doprinose ukupnom odstupanju neophodno je uraditi sledeće:

a) Srednje vrednosti poređati u rastući niz vrednosti.

b) Odrediti tzv. najmanju značajnu razliku (least significant difference) $s\sqrt{\frac{2}{n}} \times t_{h(n-1)}$

s - odstupanje unutar uzorka, $h(n-1)$ – broj stepeni slobode pomenutog određivanja.

c) Uporediti razlike srednjih vrednosti sa najmanjom značajnom razlikom.

Primer 1. Analiza varijanse – međulaboratorijsko ispitivanje

Četiri laboratorije (A - D) učestvuju u međulaboratorijskom ispitivanju. Laboratorijama je prosleđen na analizu isti uzorak pijaće vode uz zahtev da se odredi sadržaj neorganskih anjona. Rezultat tri ponovljena merenja sadržaja hlorida ($\mu\text{g/l}$) u svakoj od laboratorija dat je u sledećoj tabeli.

Lab. A	Lab. B	Lab.C	Lab.D
102	101	97	90
100	101	95	92
101	104	99	94

Da li na nivou značajnosti od $P = 0,05$ postoji statsitički značajna razlika među rezultatima ovih laboratorija?

Ukoliko je razlika u rezultatima prisutna odredite koje laboratorije doprinose datoj razlici.

Komentar:

U okviru Data Analysis ToolPack-a nalazi se alatka “Anova: Single Factor“. Odaberite opciju sa padajućeg menija Data/Data Analysis; starujte komandu Anova: Single Factor.

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.

Ukoliko ste sve uradili ispravno vaši rezultati bi trebalo da izgledaju ovako:

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	186	3	62	20,66667	0,0004	4,066181
Within Groups	24	8	3			
Total	210	11				

Izračunata F -vrednost upoređuje se sa kritičnom vrednošću; ukoliko je $F_{izr} > F_{krt}$ nulta hipoteza se odbacuje, tj. svi uzorci ne potiču iz iste populacije; ukoliko je $F_{izr} < F_{krt}$ nulta hipoteza se zadržava, tj. uzorci potiču iz iste populacije.

Pošto je izračunata vrednost parametra F veća od kritične, nulta hipoteza se odbacuje, tj. srednje vrednosti uzoraka se značajno razlikuju.

Utvrđivanje seta rezultata koji uslovljava odstupanja vrši se tako što se srednje vrednosti poređaju u rastući niz i upoređuju razlike susednih vrednosti sa najmanjom značajnom razlikom. Srednje vrednosti uzoraka, poređane u rastući niz, su sledeće:

$$\bar{x}_D = 92 \quad \bar{x}_C = 97 \quad \bar{x}_A = 101 \quad \bar{x}_B = 102$$

Najmanja značajna razlika iznosi:

$$d = \sqrt{3} \cdot \sqrt{2/3} \cdot 2,306 = 3,26$$

Upoređivanjem ove vrednosti sa razlikama između srednjih vrednosti uočava se da laboratorije D i C, kao i C i A daju rezultate koji se značajno razlikuju jedni od drugih i od rezultata dobijenih u laboratorijama A i B. Sa druge strane, rezultati dobijeni u laboratorijama A i B se ne razlikuju značajno jedni od drugih.

Primer 2. Analiza varijanse

Četiri analitičara vršili su određivanje sadržaja Hg (ppb) u istom uzorku. Dobijeni su sledeći rezultati:

Analitičar 1	Analitičar 2	Analitičar 3	Analitičar 4
10,38	10,14	10,20	10,19
10,26	10,25	10,11	10,15
10,29	10,04	10,02	10,16
10,42	10,28	10,15	10,28
10,31	10,16	10,50	10,10
10,19	10,09	10,12	10,32

- Odrediti srednju vrednost, standardnu devijaciju i 95%-interval pouzdanosti svakog seta podataka.
- Utvrđiti da li postoji statistički značajna razlika ($P = 0,05$) između srednjih vrednosti prvog i trećeg seta rezultata.
- Uporediti srednje vrednosti sva četiri seta rezultata ($P = 0,05$).
- Odrediti zajedničku srednju vrednost i ukupnu standardnu devijaciju rezultata.
- Uporediti zajedničku srednju vrednost sa pravom vrednošću sadržaja Hg od 10,18 ppb. Odrediti apsolutnu i relativnu grešku određivanja.

Komentar:

Primer 3. Analiza varijanse – homogenost uzorka

Pri određivanju sadržaja antimona u rečnoj vodi u blizini industrijske zone dobijeni su sledeći rezultati za uzorke uzete sa površine, dna i središnjeg dela toka:

Površina	Sredina	Dno
49.0	50.8	51.3
52.0	50.6	49.3
51.0	51.0	49.3
51.2	54.2	50.7
52.0	49.4	48.3
55.8	52.8	52.3
50.6	53.4	49.7
49.6	54.2	51.5

- Za svaki set rezultata odrediti srednju vrednost, medijanu, opseg (interval), standardnu devijaciju, 95% interval pouzdanosti;
- Utvrditi da li postoji statistički značajna razlika na nivou pouzdanosti $P = 0,05$ između srednjih vrednosti prvog i trećeg seta rezultata;
- Uporediti srednje vrednosti sva tri seta rezultata na nivou pouzdanosti $P = 0,05$. Da li postoje razlike između mesta uzorkovanja?

Komentar: