

Analiza variance (ANOVA)

Analiza variance je ena najpomembnejših kemometrijskih metod. Pri ANOVI gre za analizo oziroma za primerjavo povprečij večjega števila vzorcev (k). Ničelna hipoteza ANOVE je vedno **statistična** enakost vseh povprečij. Testiramo predpostavko, da so vsi vzorci vzeti iz iste populacije. ANOVA je uporabna predvsem pri vrednotenju laboratorijev za delo z neko predpisano metodo, za primerjanje rezultatov iste metode, ki jih dajejo **različni** laboratoriji ali en laboratorij pod **različnimi** pogoji (**posledično** je torej uporabna za odkrivanje slabosti v laboratorijih), za testiranje ustreznosti kalibracije analitskih instrumentov (glej poglavje o kalibracijski premici!), za testiranje modelov dobljenih z MLR (ali drugimi modelnimi metodami) s katerimi napovedujemo lastnosti kompleksnih vzorcev in podobno.

ANOVA je osnovana na dveh predpostavkah:

- a) vse meritve x_{ij} so normalno porazdeljene in
- b) variance v vzorcih so homogene (Hartleyjev, Cochranov, Bartlettov test)

K sreči je ANOVA zelo robustna metoda, kar pomeni, da daje uporabne rezultate tudi, če podatki malo odstopajo od obeh gornjih predpostavk.

Pri ANOVI vedno najprej sestavimo več skupini vzorcev, ki smo jih dobili z meritvami pri obdelavi ali ovrednotenju nekega analitskega procesa. Ničelna hipoteza H_0 pri ANOVA testu je, da so povprečja v skupinah vzorcev statistično enaka. Z drugimi besedami, da se ob upoštevanem tveganju α , ne razlikujejo med seboj. Ničelno hipotezo pri ANOVI preverjamo z F -testom, to je s primerjavo variance povprečij med skupinami vzorcev, ki smo jih izbrali, in variance povprečij znotraj vzorcev.

$$F = \frac{\sigma_{med\ skupinami}^2}{\sigma_{znotraj\ skupin}^2} < F_{\alpha, ps_1, ps_2}^{tabele} \rightarrow H_0 \text{ sprejeta}$$

Obe oznaki ps_1 in ps_2 v zgornji enačbi pomenita število prostostnih stopenj v števcu (ps_1) in v imenovalcu (ps_2)

Večkratni t-test

Če želimo primerjati več povprečnih vrednosti bi lahko za to uporabili večkratni t-test. Vendar pri taki primerjavi uporabimo isto povprečno vrednost večkrat, zato posamezni t-testi niso več med seboj neodvisni. Kot posledica se verjetnost da bo vsaj en test spodletel pri primerjavi povprečnih vrednosti iz iste populacije poveča. En način da se znebimo tega problema je da prilagodimo tveganje (Bonferronijeva prilagoditev) posameznega testa (α') tako, da bo skupno tveganje (α) take določitve doseglo predpisano vrednost.

$$\alpha' = 1 - (1 - \alpha)^{\frac{1}{k}}$$

Primer:

Primerjamo 5 povprečij. Za tako primerjavo moramo narediti 10 t-testov. Če želimo, da bo skupno tveganje napovedi $\alpha=0.05$, potem mora biti izbrano tveganje posameznega testa α' :

$$\alpha' = 1 - (1 - \alpha)^{\frac{1}{k}} = 1 - (1 - 0.05)^{\frac{1}{10}} = 0.005$$

Testiranje enakosti varianc

Ker ANOVA zahteva enakost varianc moramo najprej preveriti ali je ta predpostavka izpolnjena. Na voljo imamo več testov. Pogledali si bomo tri najpogostejše.

Hartleyev test

Pri njem primerjamo največjo in najmanjšo varianco posameznih serij. Dobljen faktor primerjamo z kritičnimi F vrednostmi.

$$F_{izr} = \frac{S_{\max}^2}{S_{\min}^2}$$

Cochranov test

Pri tem testu primerjamo največjo varianco z vsoto vseh varian. Uporabimo naslednjo zvezo

$$C_{izr} = \frac{S_{\max}^2}{\sum_{j=1}^k S_j^2}$$

Osnovna predpostavka pri Cochranovem testu je, da je število elementov v posamezni seriji enako. Pri anjših odstopanjih od te zahteve test še vedno lahko izvajamo. Namesto varianc lahko test izvedemo tudi s kvadrati območij. Kritične vrednosti dobimo iz Cochranovih C tabel. Kritične vrednosti so pri izbrani stopji zaupanja odvisne od števila serij k in števila elementov v posamezni seriji.

Bartlettov test varianc večjega števila vzorcev

Testiranje varianc, če so statistično enake ali ne, je v bistvu test porazdelitev, iz katerih smo variance izračunali. Ker porazdelitve primerjamo z uporabo χ^2 testa, je logično, da χ^2 tabele uporablja tudi za Bartlettov test.

Najprej moramo izračunati skupno varianco (pooled variance) vseh k vzorcev, katerih povprečja prizkušamo. Če ima vsak od k vzorcev n_j meritev in že določeno ali izračunano varianco s_j^2 , potem je skupna varianca enaka:

$$s_{pooled}^2 = \frac{\sum_{j=1}^k (n_j - 1) s_j^2}{\sum_{j=1}^k (n_j - 1)}$$

Pri Bartlettovem testu moramo vse izračunane variance logaritmirati z naravnim logaritmom in potem izračunati izraz M . Število prostostnih stopenj pri testiranju je število vzorcev k , oziroma število varianc, ki jih primerjamo, zmanjšano za eno ($k-1$):

$$M = \left[\sum_{j=1}^k (n_j - 1) \right] \ln(s_{pooled}^2) - \sum_{j=1}^k (n_j - 1) \ln(s_j^2) < \chi^2(\alpha, k-1) \rightarrow H_0 \text{ sprejeta}$$

Podobno kot moramo pred primerjavo dveh povprečij s t -testom, preveriti ali imata vzorca primerljivi varianci (to naredimo z F -testom; $F = s_1^2/s_2^2$), moramo tudi preden pričnemo primerjati več povprečij z ANOVO, narediti Bartlettov test. Ta nam zagotov, bodisi da so variance, ki jih obravnavamo homogene, bodisi da to niso.

Bartlettov test je močno občutljiv na odstopanja od normalne porazdelitve znotraj serij podatkov, tako odstopanja od kritičnih vrednosti pogosto nakazujejo odstopanje od normalne porazdelitve in ne neenakost varianc.

Eno-faktorska ANOVA (one-way ANOVA), medlaboratorijski test

Najenostavnejši primer ANOVE je primerjava povprečij k vzorcev. Vsak j -ti vzorec ima n_j meritev. Testiramo ali so vsa povprečja iz iste populacije. Če gre pri tem za medlaboratorijsko primerjavo k laboratorijev pri kateri je bilo v vsakem laboratoriju narejenih n_j meritev, preverjamo ali dobijo vsi laboratoriji statistično enake rezultate (seveda ob naprej predpisani stopnji tveganja α).

Oznaka x_{ij} pomeni i -to meritev v j -tem laboratoriju, \bar{x}_j je povprečje meritev j -tega laboratorija. Skupno število meritev v vseh laboratorijih ($n_1 + n_2 + \dots + n_j + \dots + n_k$) je enako N . Z $\bar{\bar{x}}$ pa označimo celotno povprečje vseh N meritev.

Izvor variance	Vsota kvadratov napak (SS)	SS/(pr.s t)	F-test
Med k laboratoriji	$SS_{med} = \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 = \sum_{j=1}^k n_j \bar{x}_j^2 - N \bar{\bar{x}}^2$	$SS_{med} / (k - 1)$	
Znotraj laboratorija	$SS_{znotraj} = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}^2 - \sum_{j=1}^k n_j \bar{x}_j^2$	$SS_{znotraj} / (N - k)$	$F = \frac{SS_{med} / (k - 1)}{SS_{znotraj} / (N - k)} \rightarrow F_{\alpha, k-1, N-k}^{tabele}$
Skupna varianca	$SS_{Tot} = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}^2 - N \bar{\bar{x}}^2$	$SS_{Tot} / (N - 1)$	

Ker je $SS_{Tot} = SS_{med\ lab.} + SS_{znotraj\ lab.}$ ni treba računati vseh treh vrednosti, ampak moramo le dve. Najlažje je izračunati SS_{tot} , in $SS_{med\ lab.}$, zato navadno izračunamo $SS_{znotraj\ lab.}$ posredno, iz razlike med prvima dvema:

$$SS_{znotraj\ lab.} = SS_{Tot} - SS_{med\ lab.}$$

V večini primerov je $SS_{znotraj\ lab.}$, ki mu pravimo tudi ostanek ("residual", SS_{res}), manjši kot $SS_{med\ lab.}$. Zato bo količnik F večji od ena (medlaboratorijske razlike so večje kot znotraj laboratorijske). Če temu ni tako, potem razlike med laboratoriji z gotovostjo niso signifikantne. Take postanejo šele, ko iz podatkov izračunani F preseže $F_{tabelaričen}$.

Kalibracijska premica

Izračun vseh parametrov kalibracijske premice je osnovna naloga večine kvantitativnih analiznih postopkov. Izhajali bomo iz 4 ponovitev vsake meritve pri 5 različnih standardnih koncentracijah, ki so podane v spodnji tabeli.

Tabela 1. Ponovitve za umeritev kalibracijske premice pri petih različnih standardnih koncentracijah. Skupno število meritev $N=20$. Za račun premice ni nujno, da je pri vsaki koncentraciji narejeno enako število ponovitev. Primerno pa je, da sta v vsaki točki narejeni najmanj dve ponovitvi.

	j	1	2	3	4	5
Standardne koncentracije	x_j	1.00	2.00	3.00	5.00	10.00
Ponovitve meritev	y_{j1}	10.60	24.80	31.00	52.30	102.40
	y_{j2}	8.70	22.20	32.30	51.40	85.90
	y_{j3}	12.80	23.80	35.20	59.30	95.20
	y_{j4}	9.50	21.80	32.30	54.60	98.40
Število ponovitev	n_j	4	4	4	4	4

Za začetni izračun parametrov same kalibracijske premice (naklona in odsek na ordinatni osi), ne potrebujemo posameznih vrednosti vseh ponovitev, ampak povprečja meritev za vsako koncentracijo.

Posamezne vrednosti vseh meritev različnih koncentracij potrebujemo šele pri ovrednotenju kvalitete modela, t.j. pri ovrednotenju linearnosti kalibracijske premice z F -testom (glej poglavje ANOVA).

Metoda najmanjših kvadratov

Model premice

$$y_i = b_0 + b_1 x_i + e_i$$

Napaka modela

$$E = \sum_{i=1}^n e_i^2$$

$$\frac{\delta E}{\delta b_0} = \sum 2(y_i - b_0 - b_1 x_i)(-1) = 0$$

$$\frac{\delta E}{\delta b_1} = \sum 2(y_i - b_0 - b_1 x_i)(-x_i) = 0$$

$$\sum x_i y_i - b_0 \sum x_i - b_1 \sum x_i^2 = 0 \Rightarrow b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\sum y_i - n b_0 - b_1 \sum x_i = 0 \Rightarrow b_0 = \bar{y} - b_1 \bar{x}$$

Tabela 2

Standardne koncentracije	x_j	1.00	2.00	3.00	5.00	10.00	
Povprečja meritev	$\bar{y}_j = (1/n_j) \sum_{i=1}^{n_j} y_{ij}$	\bar{y}_j	10.40	23.15	32.70	54.40	95.48
Število različnih koncentracij na x-osi k = 5							
Vsota koncentracij	$S_x = \sum_{j=1}^k x_j$		21.00				
Vsota povprečij meritev \bar{y}_j	$S_y = \sum_{j=1}^k \bar{y}_j$		216.13				
Vsota kvadratov koncentracij	$SS_x = \sum_{j=1}^k x_j^2$		139.00				
Vsota kvadratov povprečij \bar{y}_j	$SS_y = \sum_{j=1}^k \bar{y}_j^2$	13788.21		ne potrebujemo za izračun kalibracijske premice			
Vsota produktov	$SP_{xy} = \sum_{j=1}^k x_j \bar{y}_j$	1381.55					
Totalna vsota vseh meritev y_{ij}	$TS_y = \sum_{j=1}^k \sum_{i=1}^{n_j} y_{ij}$	864.50		ne potrebujemo za izračun kalibracijske premice			

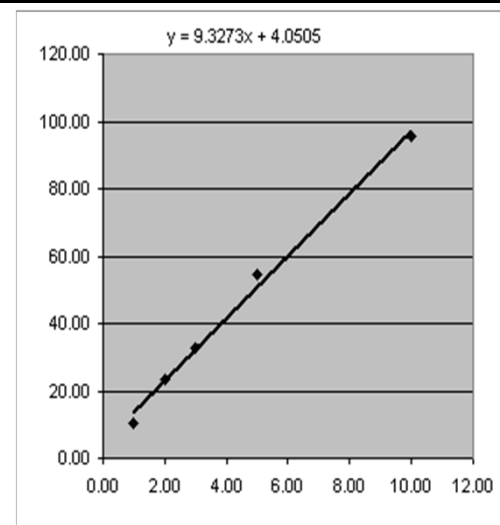
Enačba kalibracijske premice $y = a + bx$

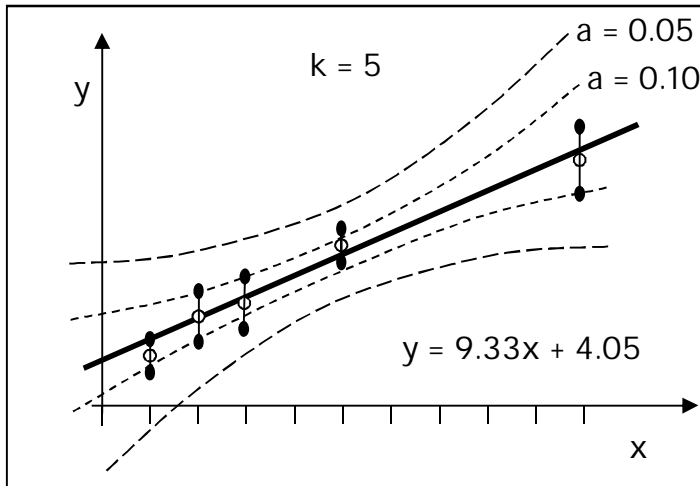
$$b = \frac{\sum_{j=1}^k (x_j - \bar{x})(\bar{y}_j - \bar{\bar{y}})}{\sum_{j=1}^k (x_j - \bar{x})^2} = \frac{k \sum_{j=1}^k x_j \bar{y}_j - \sum_{j=1}^k x_j \sum_{j=1}^k \bar{y}_j}{k \sum_{j=1}^k x_j^2 - (\sum_{j=1}^k x_j)^2}$$

$$a = \bar{y} - b\bar{x} = \frac{\sum_{j=1}^k \bar{y}_j \sum_{j=1}^k x_j^2 - \sum_{j=1}^k x_j \sum_{j=1}^k x_j \bar{y}_j}{k \sum_{j=1}^k x_j^2 - (\sum_{j=1}^k x_j)^2}$$

Naklon premice
b = 9.33

Odsek na
ordinatni osi
a = 4.05





Interval zaupanja (Confidence interval)

$$y_0 = a + bx_0 \pm t(\alpha, k-2) s_e \sqrt{\frac{1}{k} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

napaka modela $s_{\hat{y}_0} = s_e \sqrt{\frac{1}{k} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$

celotna napaka s_e
$$s_e = \sqrt{\frac{\sum_{j=1}^k (\bar{y}_j - \hat{y}_j)^2}{k-2}} = \sqrt{\frac{\sum_{j=1}^k \bar{y}_j^2 - (1/k)(\sum_{j=1}^k \bar{y}_j)^2 - b \left[\sum_{j=1}^k x_j \bar{y}_j - (1/k)(\sum_{j=1}^k x_j)(\sum_{j=1}^k \bar{y}_j) \right]}{k-2}}$$

napaka naklona s_b
$$s_b = \frac{s_e}{\sqrt{\sum_{j=1}^k x_j^2 - \frac{1}{k}(\sum_{j=1}^k x_j)^2}}$$

napaka odseka s_a
$$s_a = s_b \sqrt{\frac{\sum_{j=1}^k x_j^2}{k}}$$

Za izračun hiperbolične oblike intervala zaupanja pri določeni stopnji zaupanja α in za določitev napak naklona s_b in odseka s_a , potrebujemo samo vrednosti, ki so že podane v prvi tabeli.

To so S_x , S_y , S_{x^2} , S_{y^2} in S_{xy}

Določitev koncentracije analita in napake določitve s pomočjo kalibracijske premice

Pri analitskem delu je, potem ko določimo vse lastnosti kalibracijske premice, t.j., njene parametre in njihove napake, osnovna naloga določiti koncentracijo iskane substance v neznanem vzorcu in določiti napako napovedi.

Enačbe za izračun napake in intervala zaupanja, ki smo jih uporabili na prejšnji strani, veljajo le za regresijsko premico, ne pa tudi za napako, ki jo storimo pri določitvi koncentracije analita na osnovi te premice.

Meritev neznanega vzorca \bar{y}_s dobimo na podlagi n_s ponovljenih meritev:

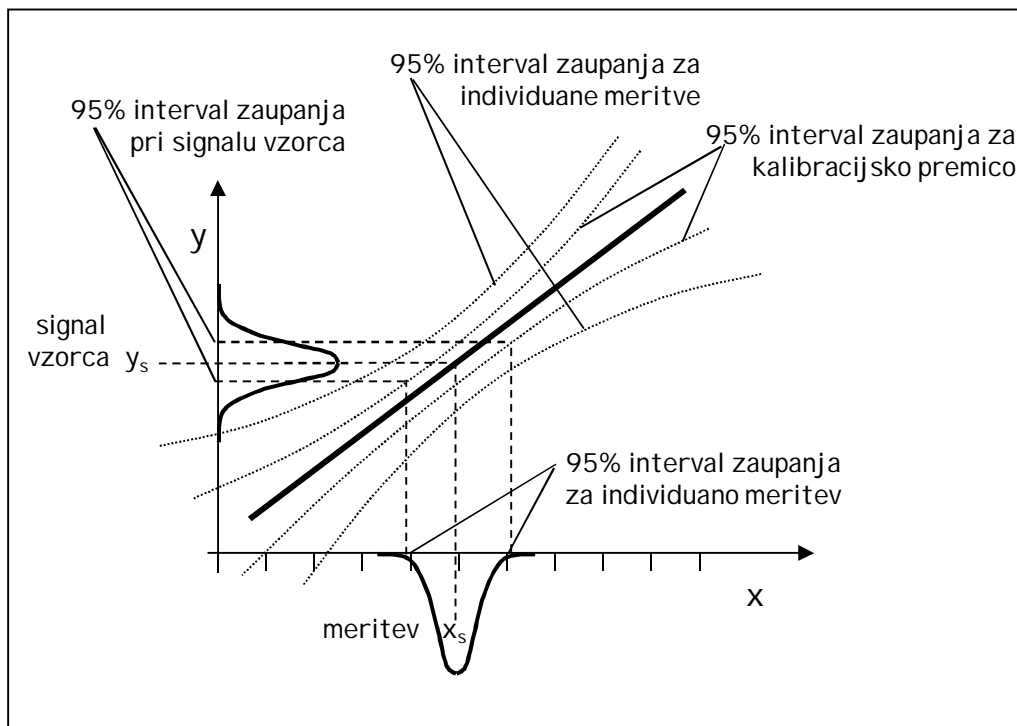
$$\bar{y}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} y_{is}, \text{ pri neznanem } x_s, \quad \hat{x}_s = \frac{\bar{y}_s - a}{b}$$

Koncentracijo v neznanem vzorcu določimo iz kalibracijske premice, $y = a + bx$.

Napaka pri določitvi koncentracije neznanega vzorca, je odvisna tako od napake same meritve, \bar{y}_s , kot tudi od vseh napak s katerimi smo določili parametre kalibracijske premice. Skupni izraz za določitev koncentracije in napake te določitve je:

$$\hat{x}_s = \frac{\bar{y}_s - a}{b} \pm t(\text{obojestranski}, 0.05, k - 2) \frac{s_e}{b} \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(\bar{y}_s - \bar{y})^2}{b^2 \sum_{j=1}^k (x_j - \bar{x})^2}}$$

V zgornjem izrazu sta le n_s in \bar{y}_s dobljena iz meritev neznanega vzorca, vsi ostali parametri pripadajo kalibracijski premici: k je število točk na njej, a in b sta naklon in odsek, x_j , \bar{x} in \bar{y} pa so vrednosti dobljene s kalibracijo na prvotnih podatkih. Niz vrednosti $\{x_j\}$ so koncentracije pri katerih je bila narejena kalibracija, \bar{x} je njihovo povprečje, \bar{y} pa povprečje vseh odzivov pri kalibraciji.



$$y_0 = bx_0 + a$$

$$\text{napaka kalibracijske premice} \Rightarrow s_{y_0} = s_e \sqrt{\frac{1}{k} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

$$\text{napaka enega odziva} \Rightarrow s_{y_s} = s_e \sqrt{1 + \frac{1}{k} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

$$\text{napaka odziva za } n_s \text{ replikacij} \Rightarrow s_{y_s} = s_e \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

$$\text{napaka meritve koncentracije} \Rightarrow s_{x_s} = \frac{s_e}{b} \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(y_s - \bar{y})^2}{b^2 \sum_{j=1}^k (x_j - \bar{x})^2}}$$

$$\text{napaka meritve koncentracije} \Rightarrow s_{x_s} = \frac{1}{b} \sqrt{\frac{s_s^2}{n_s} + \frac{s_e^2}{k} + \frac{(y_s - \bar{y})^2}{b^2 \sum_{j=1}^k (x_j - \bar{x})^2}}$$

Pri računanju standardnih napak je treba vedno paziti katero napako računamo in kakšni podatki so nam na voljo. Če se standardna napaka meritve na vzorcu s_s razlikuje od standardne napaka meritev pri kalibraciji s_e , potem moramo to tudi upoštevati (glej razliko med levim zgornjim in levim spodnjim izrazom). Napaka s_e je podana na prejšnji strani, napako s_s pa izračunamo z znanim obrazcem za standardno napako vzorca v katerem smo naredili n_s ponovitev (ne smemo pozabiti, da merimo odzive y_{is}):

$$s_s = \sqrt{\frac{\sum_{i=1}^{n_s} (y_{is} - \bar{y})^2}{n_s - 1}}$$

Testiranje ustreznosti modelov z ANOVA testom

V analizi kemiji pogosto pridemo do problema modeliranja, t.j., do iskanja takega **matematičnega** modela, ki bo dani niz meritev opisal najbolje. Z modelom, ki je lahko polinomska ali kakšna **drugačna analitična** funkcija, **matrično-vektorski** izraz, ali nevronska mreža, želimo napovedati lastnosti neznanih vzorcev. Najpreprostejša uporaba modelov je kalibracijska premica. Modelna funkcija je premica z dvema parametroma: $f(x) = ax + b$. Predno model (v našem primeru premico, lahko pa je model tudi bolj kompliciran) **začnemo** uporabljati, moramo z ANOVA preveriti ali je, glede na **izračunano** ali izmerjeno **natančnost** meritev, ustrezen ali ne.

V skladu z idejo ANOVA testa preverjamo ali sta dve skupini **povprečij** enaki ali ne. Prva skupina so variance meritev v posameznih **točkah** (ponovitve ali replikacije meritev) na katerih smo model naredili (razpršenost merskih povprečij), druga skupina pa predstavlja razpršenost ali odstopanje točk s katerimi smo model izdelali, od dejanskih napovedi modela. Z drugimi besedami: primerjamo variance posameznih meritev glede na **povprečje** ponovitev z varianco, ki jo dajo **povprečne** vrednosti ponovitev glede na napovedi modela,

$$F = \frac{\sigma_{modela}^2}{\sigma_{meritev}^2} < F(\alpha, ps_1, ps_2) \rightarrow model\ ustrezen$$

V F -testu sta ps_1 in ps_2 števili prostostnih stopenj modela ($ps_1 = k - p$) in meritev povprečij ($ps_2 = N - k$). Modelna funkcija $f(x_j)$ ima p parametrov (v primeru premice sta to a in b , $p = 2$) in je narejena na osnovi k **merskih točk** x_j , $j = 1 \dots k$. V primeru kalibracijske premice je x_j j -ta koncentracija. Pri vsaki točki x_j je bilo izmerjeno n_j ponovitev y_{ij} . ($i = 1 \dots n_j$). Vrednost modelne funkcije $f(x_j)$ v točki x_j velikokrat označimo s strešico \hat{y}_j .

$$\left. \begin{aligned} \sigma_{modela}^2 &= \frac{SS_{modela}}{k - p} = \frac{\sum_{j=1}^k n_j [\bar{y}_j - f(x_j)]^2}{k - p} \\ \sigma_{meritev}^2 &= \frac{SS_{meritev}}{N - k} = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{N - k} \end{aligned} \right\} F = \frac{SS_{modela}}{SS_{meritev}} = \frac{\frac{\sum_{j=1}^k n_j (\bar{y}_j - \hat{y}_j)^2}{k - p}}{\frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{N - k}} < F_{\alpha, k-p, N-k}^{tabele} \rightarrow Model\ je\ ustrezen$$

Ovrednotenje modela premice (ANOVA)

Na osnovi 20 meritev ($N = 20$) smo določili model kalibracijske premice z dvema parametroma ($p = 2$): $a = 4.050$ in $b = 9.327$. V vsaki od petih točk ($k=5$), t.j. pri petih različnih koncentracijah, so bile narjane po 4 ponovitve ($n_1 = n_2 = n_3 = n_4 = n_5 = 4$).

		$j \Rightarrow$	1	2	3	4	5	
			1.00	2.00	3.00	5.00	10.00	
Ponovitve meritev	$i \downarrow$	{	10.60	24.80	31.00	52.30	102.40	
			8.70	22.20	32.30	51.40	85.90	
			12.80	23.80	35.20	59.30	95.20	
			9.50	21.80	32.30	54.60	98.40	
Povprečje replikacij	A	\bar{y}_j	10.40	23.15	32.70	54.40	95.48	
Modelne vrednosti	B	$\hat{y}_j = a + bx_j$	13.38	22.71	32.03	50.69	97.32	
Račun za SS_{lof}	C	$n_j(\bar{y}_j - \hat{y}_j)^2$	35.47	0.79	1.78	55.15	13.66	
Račun za SS_{pe}	{	D	$\sum_{i=1}^{n_j} y_{ij}^2$	442.14	2149.56	4286.62	11874.90	36610.17
		E	$\frac{1}{n_j}(\sum_{i=1}^{n_j} y_{ij})^2$	432.64	2143.69	4277.16	11837.44	36461.90
		D-E	$\sum_{i=1}^{n_j} y_{ij}^2 - \frac{1}{n_j}(\sum_{i=1}^{n_j} y_{ij})^2$	9.50	5.87	9.46	37.46	148.27

SS_{lof} = vsota vrstice C = 106.86

SS_{pe} = vsota vrstice D-E = 210.56

Število prostostnih stopenj:

$N - k = 15$ pri napaki (PE - pure error)

$K - p = 3$ pri modelu (LOF - lack of fit)

$$F = \frac{\frac{SS_{\text{LOF}}}{k-p}}{\frac{SS_{\text{PE}}}{N-k}} = \frac{\sum_{j=1}^k n_j (\bar{y}_j - \hat{y}_j)^2}{k-p} = \frac{210.56}{3} = 2.537 \quad F_{\text{tabelarična}} = F(0.05, 3, 15) = 3.287$$

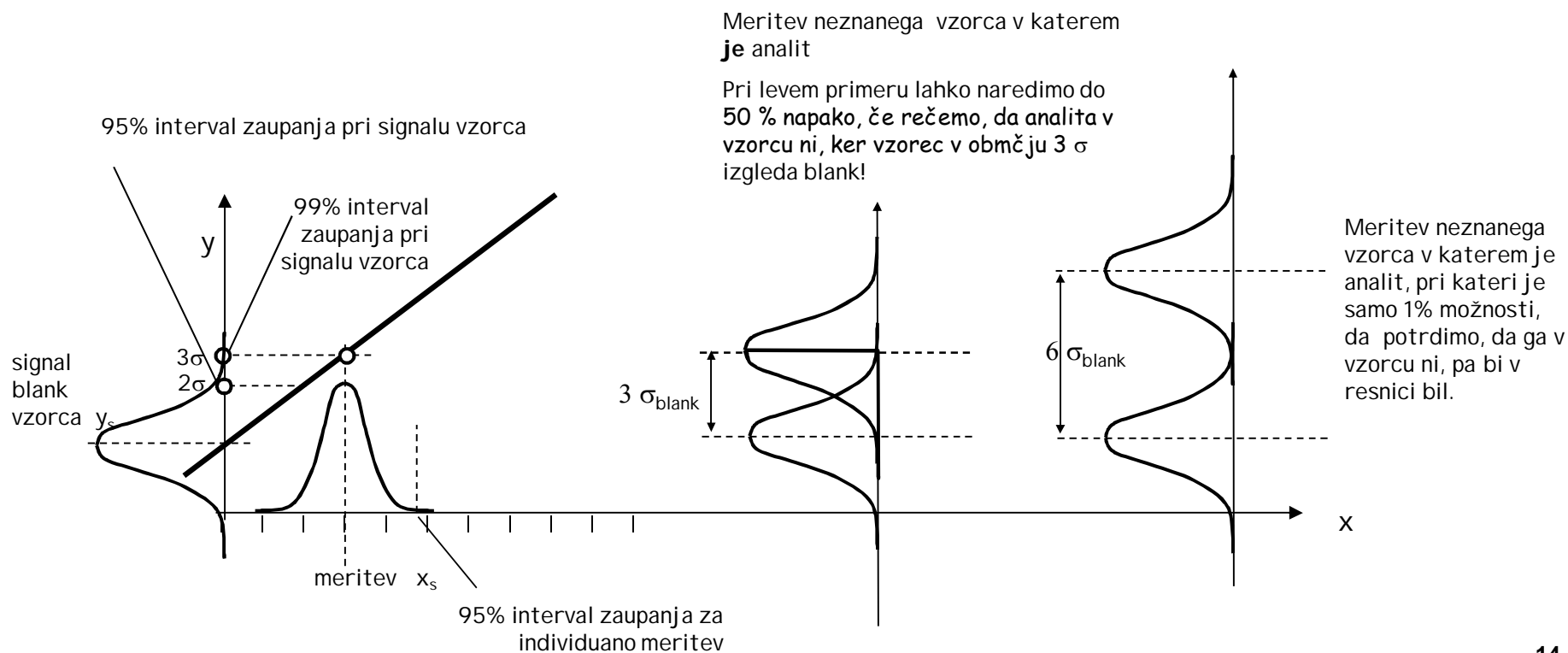
$$= \frac{106.86}{15}$$

Meja detekcije

Detekcijska limita je najmanjša količina substance, ki jo je ob predpisanem tveganju (stopnji zanesljivosti α) možno razlikovati od meritve praznega vzorca (blank sample). Stara IUPAC definicija limite detekcije, ki podaja enačbo:

$$y_{ld} = \bar{y}_{blank} + k\sigma_{blank}$$

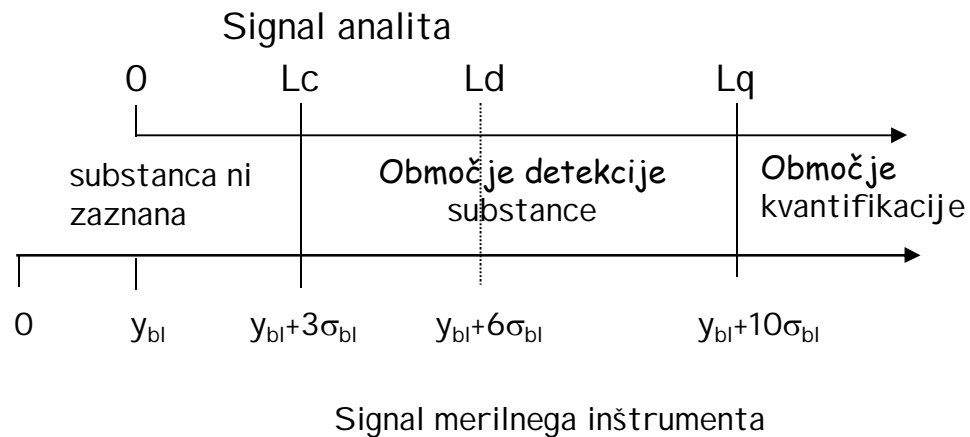
pri kateri je veljalo priporočilo, da je $k = 3$, danes splošno ni več sprejemljiva (čeprav se še precej uporablja). Stara definicija upošteva samo napake α , ne pa tudi napak β . Z drugimi besedami povedano, stara definicija zagotovi, da je napaka potrditve navzočnosti substance v analitu, če te v resnici ni v njem, z zanesljivostjo $3\sigma_{blank}$ (t.j. s tveganjem, ki je manjše od 0.13 %), a je hkrati pri tem lahko napaka pri zanikanju navzočnosti substance v analitu, če ta v njem v resnici je, velika do 50% (glej spodnjo sliko!).



Detekcijska limita (Detection limit)

Sedanji izračun in definicije detekcijskih območij

$$y_{ld} = \bar{y}_{blank} + k\sigma_{blank}$$



Lc odločitveni ali posteriorni nivo (decision level) $1/3 = 33.3\%$

Ld limita detekcije (detection limit) $1/6 = 16.7\%$

Lq kvantifikacijska limita (quantification limit) $1/10 = 10\%$

Relativna standardna napaka na nivoju $L_i = k_i\sigma_{bl}$ je $1/k_i$

Določitev limite detekcije s pomočjo intervala zaupanja kalibracijske premice

Če limita detekcije ni predpisana z določitvijo faktorja k_i (n.pr. $k_i = 3$ ali 10), jo lahko določimo tudi na podlagi intervala zaupanja. V vsakem primeru je koristno, da ti dve določitvi med seboj primerjamo. Tveganje β (da naredimo β -napako) in število meritev v vzorcu n_s s katerim določamo neznane koncentracije, moramo določiti posebej v predpisu (protokolu, ki opisuje analizo meritev). Število meritev v vzorcih, n , s katerimi smo določili kalibracijsko premico, je lahko različno od števila meritev n_s s katerimi smo določili koncentracije v neznanih vzorcih. To ni priporočljivo, primerneje je vedno delati z istim številom meritev.

$$y_0 = bx_0 + a \pm t_{\alpha, k-2} s_e \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

Če je $x_0 = 0$

$$y_c = a + t_{\alpha, k-2} s_e \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(\bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

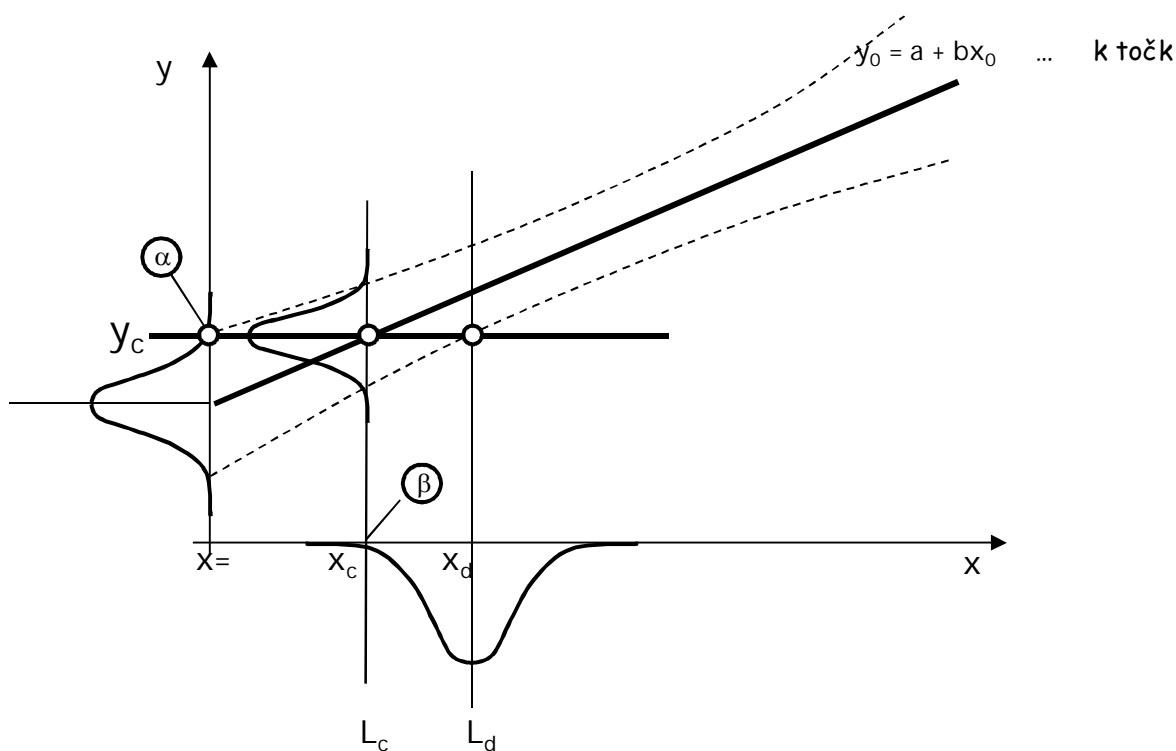
$$x_c = (y_c - a) / b$$

$$x_c = x_d - t_{\beta, k-2} \frac{s_e}{b} \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(x_d - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

gornji izraz obrnemo, da izračunamo x_d in aproksimiramo x_d pod korenem z vrednostjo $2x_c$

$$x_d = x_c + t_{\beta, k-2} \frac{s_e}{b} \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(2x_c - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

$$x_d = x_c + t_{\beta, k-2} \frac{1}{b} \sqrt{\frac{s_s}{n_s} + \frac{s_e}{k} + \frac{(2x_c - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$



1. Najprej s pomočjo kalibracijske premice izračunamo za "blank" vzorec ($x_0 = 0$) točko y_c , ki leži na zgornji meji intervala zaupanja. Zato je med "a" in "t" znak "+":

$$y_c = bx_0 + a + t_{\alpha, k-2} s_e \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}} \Rightarrow y_c = a + t_{\alpha, k-2} s_e \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(\bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

2. Izračunamo koncentracijo x_c , ki jo da kalibracijska premica za vrednost odziva y_c (to je meja področja pod katerim koncentracija analita ni zaznavna:

$$x_c = \frac{y_c - a}{b}$$

3. Predpostavimo, da leži koncentracija x_c na skrajnem levem robu detekcijske limite. Zato imamo znak "-":

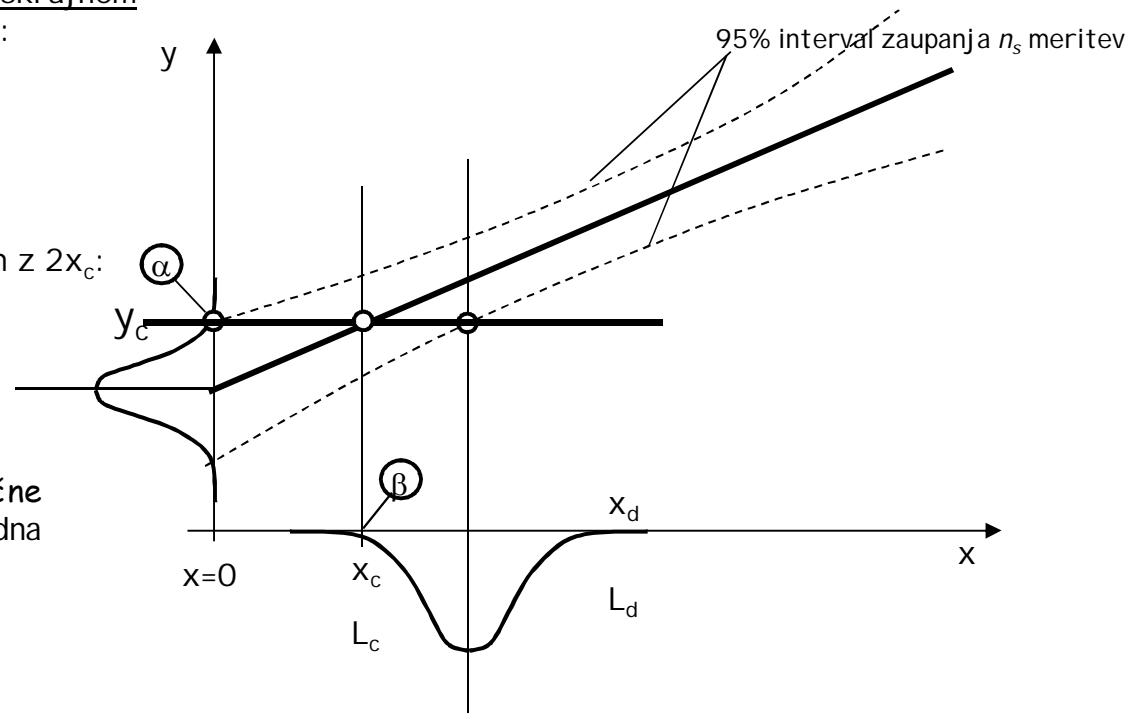
$$x_c = x_d - t_{\beta, k-2} \frac{s_e}{b} \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(x_d - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

4. Izraz obrnemo in aproksimiramo x_d pod korenem z $2x_c$:

$$x_D = x_c + t_{\beta, k-2} \frac{s_e}{b} \sqrt{\frac{1}{n_s} + \frac{1}{k} + \frac{(2x_c - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$

5. Če imamo pri meritvi in kalibraciji različne standardne napake, vnesemo dva različna standardna odmika:

$$x_D = x_c + t_{\beta, k-2} \frac{1}{b} \sqrt{\frac{s_s}{n_s} + \frac{s_e}{k} + \frac{(2x_c - \bar{x})^2}{\sum_{j=1}^k (x_j - \bar{x})^2}}$$



$$s_e = \sqrt{\frac{\sum_{j=1}^k (\bar{y}_j - \hat{y}_j)^2}{k-2}}$$

Pri premici (s_e) imamo dve prostostni stopnji manj, ker računamo 2 parametra, a in b,

$$s_s = \sqrt{\frac{\sum_{i=1}^{n_s} (y_{is} - \bar{y}_s)^2}{n_s - 1}}$$

pri meritvah pa imamo eno samo prostostno stopnjo manj, ker smo izračunali samo povprečje.