HD 589 Arkivets

WORK RESEARCH INSTITUTES
GYDAS VEI 8, OSLO 3 - P. O. BOX 5387
TELEPHONE 46 68 50

NØYAKTIGHET AV DOBBELTANALYSER

I. INNLEDNING

Resultatene av en kjemisk analyse må angis med feilgrenser. Ved å bruke en bestemt fremgangsmåte kan man bestemme om analysemetoden er beheftet med feil og hvor store disse i så fall er. Her er angitt hvordan man kan bestemme:

1. en eventuell konstant feil A
2. en eventuell proporsjonal feil B
3. tilfeldige feil.

det er gjort følgende forutsetninger:

1. Prøven er homogen
2. Det søkte stoff kan tilsettes prøvene i kjent mengde.
3. Reproduserbarheten av analysen er tilnærmet konstant for det mengdeområde som er aktuelt.
4. Det foreligger et forholdsvis stort antall, M, prøver av samme type.

II. PREPARERING AV PRØVENE

De M prøvene deles vilkårlig i to grupper X ogY med henholdsvis N_x og N_y prøver med omtrent samme antall i hver gruppe.

Fra hver av prøvene i gruppe X tas to delprøver med forskjellig prøvemengde e_1 og e_2 hvor disse velges slik at $e_1=2e_2$. Den analyserte mengde av det søkte stoff i disse to prøvene betegnes med henholdsvis x'_1 og x'_2, mens det virkelige innhold (ukjent) er x_1 og x_2.

Fra hver av prøvene i gruppe Y tas to delprøver med samme prøvemengde $e = e_2$, begge med virkelig innhold av det søkte stoff = y. Til den ene av delprøvene settes en kjent mengde z av det søkte stoff. Analysert mengde av det søkte stoff uten tilsats betegnes med y'_1, og med tilsats med y'_2. Tilsatsen z bør være av samme størrelsesorden som y.

III. BEREGNING AV GAUSSFORDELING

Det forutsettes at man har mange målinger, helst 100 - 200. Inntil man får samlet så mange, vil man som regel anta at fordelingen er normal. For kjmiske analyser må man imidlertid regne med at fordelingen er normal for logaritmene til analyseresultatene (log-normal fordeling) hvis spredningen er stor.

Verdiene som er funnet inndeles i 6 klasser eller mer og antallet som finnes i hver klasse noteres og omregnes til kumulativ prosent, regnet fra laveste klasse.

På to typer sannsynlighetspapir, ett med liniær abcisse og ett med logaritmisk abcisse, avsettes så de enkelte klassers øvre verdi. Ordinaten er delt i kumulative prosenter. Sett den beregnede kumulative prosent for hver klasse inn i begge diagrammer. Ligger punktene tilnærmet på en rett linje ved liniær abcisse, så er Gauss-fordelingen normal. Ligger de tilnærmet på en rett linje ved logaritmisk abcisse, så er fordelingen log-normal.

Man kan støte på tilfelle som ligger utenfor begge kategorier, men disse blir ikke behandlet her.

IV. BEREGNING AV KONSTANT OG PROPORSJONAL FEIL

En eventuell konstant feil beregnes for hver av dobbeltmålingene i gruppe X efter ligningen

$$A = 2x'_2 - x'_1 \qquad (1)$$

En eventuell proporsjonal feil beregnes for hver av dobbeltmålingene i gruppe Y efter ligningen

$$B = \frac{y'_1 - y'_2 + z}{z} \qquad (2)$$

(Nedenfor er middelverdier betegnet med en strek over, f. eks. betyr $\overline{\log A}$ middelverdien av alle log $A = \Sigma \log A : N$)

Middelverdiene $\bar{A}$ og $\bar{B}$ beregnes forskjellig eftersom A og B har normal eller log-normal Gaussfordeling.

1. Normal Gaussfordeling.
Her beregnes middelverdiene på vanlig måte:

$$\bar{A} = \frac{\Sigma A}{N_x} \quad \text{og} \quad \bar{B} = \frac{\Sigma B}{N_y} \tag{3}$$

Standardavvikene for A og B beregnes efter formlene:

$$s_A = \sqrt{\frac{\Sigma d_A^2}{N_x - 1}} = \sqrt{\frac{N_x \cdot \Sigma A^2 - (\Sigma A)^2}{N_x^2 - N_x}}$$

$$\text{og} \quad s_B = \sqrt{\frac{\Sigma d_B^2}{N_y - 1}} = \sqrt{\frac{N_y \cdot \Sigma B^2 - (\Sigma B)^2}{N_y^2 - N_y}} \tag{4}$$

hvor $d_A = \bar{A} - A$ og $d_B = \bar{B} - B$ finnes for de enkelte prøver.

For å kunne avgjøre om eventuelle avvikelser av A og B fra 0 er signifikante, brukes Student's t-test hvor t beregnes slik:

$$t_A = \frac{\bar{A}\sqrt{N_x}}{s_A} = \frac{\Sigma A}{s_A \cdot \sqrt{N_x}} = \sqrt{\frac{(\Sigma A)^2}{s_A^2 \cdot N_x}}$$

$$t_B = \frac{\bar{B}\sqrt{N_y}}{s_B} = \frac{\Sigma B}{s_B \cdot \sqrt{N_y}} = \sqrt{\frac{(\Sigma B)^2}{s_B^2 \cdot N_y}} \tag{5}$$

Antall frihetsgrader for $\bar{A}$ og $\bar{B}$ er henholdsvis $N_x - 1$ og $N_y - 1$.

I vedlagte tabell er angitt de kritiske verdier av Student's $t = t(P,n)$ for forskjellige sannsynlighetsnivåer P og frihetsgrader n. Man velger vanligst $P = 95\%$.

Hvis en t-verdi beregnet efter en av formlene (5) er større enn t(P,n)- verdien svarende til det valgte sannsynlighetsnivå og antall frihetsgrader, er vedkommende feil signifikant ved det valgte sannsynlighetsnivå og analysen inneholder da med tilsvarende sannsynlighet feil som må rettes.

Er det ikke mulig å forandre analysemetoden slik at feilen(e) forsvinner, må man korrigere resultatene. Man får da følgende virkelige innhold i delprøvene, når man foreløpig ser bort fra de tilfeldige feil:

$$x_1 = \frac{x'_1 - \bar{A}}{1 - \bar{B}} \quad \text{og} \quad x_2 = \frac{x'_2 - \bar{A}}{1 - \bar{B}}$$

$$y_1 = \frac{y'_1 - \bar{A}}{1 - \bar{B}} \quad \text{og} \quad y_2 = \frac{y'_2 - \bar{A}}{1 - \bar{B}} - z \qquad (6)$$

Uttrykt i mengde av det søkte stoff pr. mengdeenhet prøve blir resultatene:

$$q_{x_1} = \frac{x_1}{e_1} = \frac{x'_1 - \bar{A}}{e_1(1 - \bar{B})} = \frac{x_1}{2\, e_1} = \frac{x'_1 - \bar{A}}{2\, e_2\, (1 - \bar{B})}$$

$$q_{x_2} = \frac{x_2}{e_2} = \frac{x'_2 - \bar{A}}{e_2(1 - \bar{B})}$$

$$q_{y_1} = \frac{y_1}{e_2} = \frac{y'_1 - \bar{A}}{e_2(1 - \bar{B})} \qquad (7)$$

$$q_{y_2} = \frac{y_2}{e_2} = \frac{y'_2 - \bar{A}}{e_2(1 - \bar{B})} - \frac{z}{e_2}$$

Standardavvikene og konfidensgrensene for de respektive dobbeltanalyser finnes på forskjellig måte eftersom resultatene har normal eller log-normal Gaussfordeling, se V.

2. Log-normal fordeling.

Middelverdiene $\bar{A}$ og $\bar{B}$ finnes ut fra formlene (3).

For å avgjøre om eventuelle avvikelser av $\bar{A}$ og $\bar{B}$ fra 0 er signifikante, brukes formlene (5) hvor man istedenfor A, B, s_A og s_B setter inn henholdsvis log A, log B, $s_{\log A}$ og $s_{\log B}$.

V. BEREGNING AV TILFELDIGE FEIL, MIDDELVERDIER OG KONFIDENSGRENSER

Finn først differansene $q_{x_1} - q_{x_2}$ for alle målingene i X-gruppen og $q_{y_1} - q_{y_2}$ for alle målingene i Y-gruppen. Differansene betegnes med d_q. Undersøk om disse har en normal eller log-normal Gaussfordeling.

1. Normal Gaussfordeling.

a. Tilfeldige feil.

Disse finnes ut fra alle dobbeltbestemmelsene av

$$s_q = \sqrt{\frac{\Sigma d_q^2}{2\ M}} \qquad (8)$$

som gir standardavviket for en enkelt analyse.

b. Middelverdier av dobbeltanalysene .

Disse finnes på vanlig måte for hver av prøvene i både X- og Y-serien av

$$m_q = \frac{q_1 + q_2}{2} \qquad (9)$$

idet q_1 betegner både q_{x_1} og q_{y_1}

og q_2 betegner både q_{x_2} og q_{y_2}.

c. Konfidensgrenser for middelverdiene.

Konfidensgrensene angir den høyeste og laveste verdi det virkelige innhold med en viss, på forhånd valgt sannsynlighet ligger.

Ofte st brukes et sannsynlighetsnivå P på 95 %, hvilket vil si at for 95 % av analysene ligger den virkelige verdi innen-enfor de angitte grenser, mens man må vente at den virkelige verdi Q for 5 av 100 analyser ligger utenfor grensene. Er man ikke tilfreds med dette, må man velge et høyere sannsynlighetsnivå.

Konfidensgrensene blir:

$$Q = \frac{q_1 + q_2}{2} \pm \frac{t(P,n) \cdot s_q}{\sqrt{2}} = m_q \pm o,707 \cdot t(P,n) \cdot s_q \quad (10)$$

hvor $t(P,n)$ finnes av vedlagte tabell for $n = M$ frihetsgrader og det valgte sannsynlighetsnivå P, og hvor s_q beregnes efter formel (8).

2. Log-normal fordeling for d_q.

a. Tilfeldige feil.

Standardavviket $s_{\log q}$ beregnes ut fra alle dobbeltanalysene av

$$s_{\log q} = \sqrt{\frac{\Sigma(\log q_1 - \log q_2)^2}{2\ M}} = \sqrt{\frac{\Sigma \log^2(q_1/q_2)}{2\ M}} \quad (11)$$

b. Middelverdier av dobbeltanalysene.

Middelverdiene m_q finnes av (9).

c. Konfidensgrenser for middelverdiene.

For logaritmene til det virkelige innhold blir konfidensgrensene:

$$\log Q = \log m_q \pm 0{,}7071 \cdot t(P,n) \cdot s_{\log q} \qquad (12)$$

De to konfidensgrenser man finner her ved å ta antilogaritmen til de to verdiene av (12), ligger ikke symmetrisk om m_q.

VI. KONTROLL AV RESULTATER SOM MISTENKES FOR Å VÆRE GALE.

Hvis et resultat i en måleserie skiller sig meget ut fra middelverdien, kan man for et valgt sannsynlighetsnivå regne ut om avvikelsen er tilfeldig eller skyldes en feil.

For en serie dobbeltmålinger vurderes A, B, og d for den mistenkte prøve, og dermed også de enkelte standardavvik.

Beregn først middelverdien m og standardavviket s for vedkommende størrelse uten den mistenkelige verdi. Hvis avvikelsen bare er tilfeldig, vil den mistenkelige verdien ligge innenfor grensene

$$m \pm s \cdot g(P,n) \qquad (13)$$

hvor n er antall prøver uten den mistenkelige verdi og g(P,n) finnes i siste kolonne i vedlagte tabell.

Ligger den mistenkelige verdi utenfor grensene, må man regne med at det er gjort en feil og målingen(e) tas om igjen. Ligger verdien innenfor grensene, skal den tas med både for beregning av middelverdien og standardavviket.

Ved logaritmisk normalfordeling gås frem på samme måte, men man regner da med logaritmene til de aktuelle verdier istedenfor disse selv.

VII. LITTERATUR

1. W. J. Youden: Statistical Methods for Chemists, Chapman & Hall, Ltd, London 1951.

2. Klaus Doerffel: Beurteilung von Analysenverfahren und -ergebnissen. Z. f. Analyt. Chemie, 185 (1961) p.1-98. Rettelsetil side 41 senere.

3. U. Graf u. H.-J. Henning: Zum Ausreisserproblem, Mitteilungsbl. math. Stat. 4, 1 (1952)

YRKESHYGIENISK INSTITUTT

2. mai 1969.

Jørgen Jahr

Jørgen Jahr.

TABLE I. CRITICAL VALUES OF *t*

n	50	60	70	80	90	95	98	99	99,9	g(P,n) 95%	99%
					P %						
1	1.000	1.376	1.963	3.078	6.314	12.706	31,821	63.657	636,619		
2	.816	1.061	1.386	1.886	2.920	4.303	6,965	9.925	31,598		
3	.765	.978	1.250	1.638	2.353	3.182	4,541	5.841	12,941		
4	.741	.941	1.190	1.533	2.132	2.776	3,747	4.604	8,610	7,3	60
5	.727	.920	1.156	1.476	2.015	2.571	3,365	4.032	6,859	5,9	18
6	.718	.906	1.134	1.440	1.943	2.447	3,143	3.707	5,959	5,3	12
7	.711	.896	1.119	1.415	1.895	2.365	2,998	3.499	5,405	4,8	9,8
8	.706	.889	1.108	1.397	1.860	2.306	2,896	3.355	5,041	4,5	8,5
9	.703	.883	1.100	1.383	1.833	2.262	2,821	3.250	4,781	4,3	7,8
10	.700	.879	1.093	1.372	1.812	2.228	2,764	3.169	4,587	4,2	7,1
11	.697	.876	1.088	1.363	1.796	2.201	2,718	3.106	4,437		
12	.695	.873	1.083	1.356	1.782	2.179	2,681	3.055	4,318		
13	.694	.870	1.079	1.350	1.771	2.160	2,650	3.012	4,221		
14	.692	.868	1.076	1.345	1.761	2.145	2,624	2.977	4,140 ?		
15	.691	.866	1.074	1.341	1.753	2.131	2,602	2.947	4,073	3,9	5,8
16	.690	.865	1.071	1.337	1.746	2.120	2,583	2.921	4,015		
17	.689	.863	1.069	1.333	1.740	2.110	2,567	2.898	3,965 ?		
18	.688	.862	1.067	1.330	1.734	2.101	2,552	2.878	3,922		
19	.688	.861	1.066	1.328	1.729	2.093	2,539	2.861	3,883		
20	.687	.860	1.064	1.325	1.725	2.086	2,528	2.845	3,850	3,85	5,2
21	.686	.859	1.063	1.323	1.721	2.080	2,518	2.831	3,819		
22	.686	.858	1.061	1.321	1.717	2.074	2,508	2.819	3,792		
23	.685	.858	1.060	1.319	1.714	2.069	2,500	2.807	3,767		
24	.685	.857	1.059	1.318	1.711	2.064	2,492	2.797	3,745		
25	.684	.856	1.058	1.316	1.708	2.060	2,485	2.787	3,725	3,8	5,0
26	.684	.856	1.058	1.315	1.706	2.056	2,479	2.779	3,707		
27	.684	.855	1.057	1.314	1.703	2.052	2,473	2.771	3,690		
28	.683	.855	1.056	1.313	1.701	2.048	2,467	2.763	3,674		
29	.683	.854	1.055	1.311	1.699	2.045	2,462	2.756	3,659		
30	.683	.854	1.055	1.310	1.697	2.042	2,457	2.750	3,646	3,8	4,8
40	.681	.851	1.050	1.303	1.684	2.021		2.704		3,8	4,7
50	.680	.849	1.048	1.299	1.676	2.008		2.678			
60	.679	.848	1.046	1.296	1.671	2.000		2.660			
120	.677	.845	1.041	1.289	1.658	1.980		2.617		3,8	4,3
∞	.674	.842	1.036	1.282	1.645	1.960		2.576			
	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001		

P

TABLE 2.7.1
Student's t

P % →	50	60	80	90	95	97,5	99	99,5	99,9
n ↓	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001 P
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657		
2	.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	1.330	1.734	2.101	2.445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850
21	.686	.859	1.323	1.721	2.080	2.414	2.831	3.135	3.819
22	.686	.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	.685	.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	.685	.857	1.318	1.711	2.064	2.391	2.797	3.090	3.745
25	.684	.856	1.316	1.708	2.060	2.385	2.787	3.078	3.725
26	.684	.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	.684	.855	1.314	1.703	2.052	2.373	2.771	3.056	3.690
28	.683	.855	1.313	1.701	2.048	2.363	2.763	3.047	3.674
29	.683	.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659
30	.683	.854	1.310	1.697	2.042	2.360	2.750	3.030	3.646
35	.682	.852	1.306	1.690	2.030	2.342	2.724	2.996	3.591
40	.681	.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	.680	.850	1.301	1.680	2.014	2.319	2.690	2.952	3.520
50	.680	.849	1.299	1.676	2.008	2.310	2.678	2.937	3.496
55	.679	.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	.679	.848	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	.678	.847	1.294	1.667	1.994	2.290	2.648	2.899	3.435
80	.678	.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	.678	.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	.677	.846	1.290	1.661	1.982	2.276	2.625	2.871	3.390
120	.677	.845	1.289	1.658	1.980	2.270	2.617	2.860	3.373
∞	.6745	.8416	1.2816	1.6448	1.9600	2.2414	2.5758	2.8070	3.2905

* Parts of this table are reprinted by permission from R. A. Fisher's *Statistical Methods for Research Workers*, published by Oliver and Boyd, Edinburgh (1925–1950); from Maxine Merrington's "Table of Percentage Points of the *t*-Distribution," Biometrika, 32:300 (1942); and from Bernard Ostle's *Statistics in Research*, Iowa State Uni-