Om Variationsanalyse. IV.

Af R. K. Kristensen.

En Bestemmelse af Middelfejlen eller Middelafvigelsen er som bekendt præget af en vis Usikkerhed, navnlig hvis Bestemmelsen hviler paa et for ringe Antal Gentagelsesresultater. *R. A. Fisher* har opstillet Sandsynlighedstal (paa Grundlag af »Student«s Undersøgelser), hvori denne Usikkerhed er indbefattet. Jo mere usikkert, Middelfejlen er bestemt, desto større skal Udslaget være i Forhold til Middelfejlen, for at man — med et i og for sig uheldigt Udtryk — kan »være paa den sikre Side«. Sætter man 50 pCt. Sandsynlighed som Grænse[1]), skal Udslaget være, eftersom Middelfejlen m er bestemt ved Hjælp af 5, 10, 15, 20 eller uendelig mange Gentagelsesresultater:

5	10	15	20	∞
0.741 m	0.703 m	0.692 m	0.688 m	0.675 m

Man ser, at naar Antallet af Gentagelsesresultater ikke er under 15—20 (og færre bør man i Almindelighed ikke have til en Middelfejlsberegning), er Forskellen mellem den foreliggende og den helt sikre Middelfejlsbestemmelse temmelig betydningsløs.

Det vil dog ikke være uden Interesse at kunne vurdere en Middelfejlsbestemmelse paa samme Maade, som man vurderer selve Forsøgsresultaterne ved Hjælp af deres Middelfejl. Man vil derigennem kunne faa et Indblik i, hvad Gentagelsesresul-

[1]) Sandsynligheden for Fejl, der er større, og Fejl, der er mindre, er da lige stor. Hvis Udslaget ligger over denne Fejlgrænse, den saakaldte »Sandsynlige Fejl« (0.675 m), er Sandsynligheden for, at det virkelige (fejlfri) Udslag overstiger 0, større end $\frac{1}{2}$. Der er med andre Ord over 50 pCt. Sandsynlighed for, at ikke hele Udslaget skyldes Forsøgsfejl. Denne tidligere meget brugte Fejlgrænse er for saa vidt mere naturlig og mindre vilkaarlig end Fejlgrænserne 2 m eller 3 m.

taternes Antal betyder for Middelfejlsbestemmelsen, og man kan se, hvorledes specielle Middelfejlsformler virker i denne Henseende. Ved disse Vurderinger kan man da — med Hensyn til Sandsynlighederne — støtte sig til den almindelige Fejllovstabel, fordi Virkningen af Gentagelsesresultaternes Antal *er* indgaaet i Beregningerne. Ifølge Fejlteorien er Middelfejlen paa Middelfejlen:

$$(m) = \frac{m}{\sqrt{2\,(n \div 1)}}, \text{ eller } (m)^2 = \frac{m^2}{2\,(n \div 1)};$$

(m) er saaledes givet ved m og Elementernes Antal, n. For samme Værdi af m er (m) alene afhængig af Antallet; $(m)^2$ er dog ikke omvendt proportional med dette, men med $n \div 1$. Dette vil sige — da $m^2 = \frac{[v^2]}{n \div 1}$ — at $(m)^2$ er proportional med Afvigelseskvadratsummen, $[v^2]$, men ikke med Fejlkvadratsummen, som jo heller ikke kendes. Dette er meget naturligt, thi den Del af de sande Fejl, som ikke kommer til Udtryk i $[v^2]$, kan selvfølgelig ikke give noget Bidrag til Bestemmelse af m og altsaa ikke komme i Betragtning ved Fastsættelsen af den Sikkerhed, hvormed m er bestemt. Ud fra denne Betragtning kan Middelfejlens Middelfejl betegnes saaledes:

$$(m)^2 = \frac{m^2}{2D},$$

hvor D er Divisor i Formlen til Bestemmelse af m. Bestemmes Middelfejlen f. Eks. af 100 Observationsresultater, der falder i 10 Rækker med 10 Elementer (Gentagelser) i hver, bliver

$$m^2 = \frac{[[v^2]]}{10\,(10 \div 1)}, \text{ og } (m)^2 = \frac{m^2}{2 \cdot 10\,(10 \div 1)}.$$

Dette stemmer med den almindelige Lov for Gentagelsernes Betydning ved Fastsættelsen af enhver anden Observationsstørrelses Gennemsnitsværdi. Bestemmes m ved Hjælp af een Gentagelsesrække à 10 Elementer, har man

$$(m)^2 = \frac{m^2}{2\,(10 \div 1)},$$

har man 10 Gentagelsesrækker, bliver

$$(m)^2 = \frac{m^2}{10 \cdot 2\,(10 \div 1)}.$$

Man ser altsaa, hvor meget Middelfejlsbestemmelsen taber i Sikkerhed, naar Middelfejlen skal bestemmes af 10 Gentagelsesrækker med 10 Elementer i hver i Stedet for af een Gentagelsesrække à 100 Elementer. I første Tilfælde bliver

$$(m)^2 = \frac{m^2}{10 \cdot 2\,(10 \div 1)} = \frac{m^2}{180},$$

i sidste Tilfælde

$$(m)^2 = \frac{m^2}{2\,(100 \div 1)} = \frac{m^2}{198}.$$

For de specielle Middelfejlsformler, der kommer til Anvendelse ved Behandling af Forsøgsresultater, har man da:

$$m^2 = \frac{[[v^2]]}{p\,(r \div 1)},\quad (m)^2 = \frac{m^2}{2p\,(r \div 1)}.$$

$$m^2 = \frac{[[v^2]]}{(p \div 1)\,(r \div 1)},\quad (m)^2 = \frac{m^2}{2\,(p \div 1)\,(r \div 1)}.$$

$$m^2 = \frac{[[v^2]]}{(p \div s)\,(r \div 1)},\quad (m)^2 = \frac{m^2}{2\,(p \div s)\,(r \div 1)}.$$

Spørgsmaalet om, hvad Middelfejlen paa Middelfejlen bliver, naar man danner forskellige Funktioner af Forsøgsresultaterne, har Oberst *N. P. Johansen* besvaret saaledes:

»Lad m være Middelfejlen paa en Størrelse og (m) Middelfejlen paa m. Lad endvidere m^2 være det tilsvarende Middelfejlkvadrat og μ Middelfejlen paa m^2. Saa er der, naar bortses fra Størrelser af højere Orden, følgende almindelige Forbindelse mellem μ og (m):

$$\mu = 2\,m\,(m). \qquad (1)$$

At Forbindelsen er omtrent af denne Art, ses let. $m^2 + \mu$ og $m + (m)$ er at betragte som mere fuldkomne Former af Middelfejlkvadrat og Middelfejl end m^2 og m alene, og de maa derfor tilfredsstille de almindelige Love. Følgelig maa man have:

$$m^2 + \mu = (m + (m))^2 = m^2 + 2\,m\,(m) + (m)^2,$$

hvis sidste Led $(m)^2$ altsaa er af den Størrelsesorden, som der ses bort fra. Hvor nøje denne Tilnærmelse er, beror paa Omstændighederne; men saaledes regner man. Al anden Fejlberegning er jo ogsaa Tilnærmelse.

Lad dernæst A og B være to af hinanden uafhængige Størrelser med Fejlrepræsentanterne m_a, (m_a) o. s. v., m_b, (m_b) o. s. v. Vi danner Funktionen

$$P = A \pm B$$

og vil nu beregne Middelfejl m_p og Middelfejls-Middelfejl (m_p) for P.

Idet som ovenfor bemærket $m^2 + \mu$ kun er en fuldkomnere Form for Middelfejlkvadratet, maa vi nu have[1]):

$$m_p^2 + \mu_p = m_a^2 + \mu_a + m_b^2 + \mu_b,$$

og tillige er $$m_p^2 = m_a^2 + m_b^2,$$

hvoraf ved Subtraktion: $\mu_p = \mu_a + \mu_b$,

eller ved Indførelse af (1):

$$2\, m_p\, (m_p) = 2\, m_a\, (m_a) + 2\, m_b\, (m_b).$$

Følgelig er $$(m_p) = \frac{m_a\, (m_a) + m_b\, (m_b)}{m_p}$$

eller idet $$m_p = \sqrt{m_a^2 + m_b^2}, \tag{2}$$

$$(m_p) = \frac{m_a\, (m_a) + m_b\, (m_b)}{\sqrt{m_a^2 + m_b^2}}. \tag{3}$$

Man kan ogsaa danne en Formel til Beregning af (m_p), naar Funktionen er

$$P = a\,A \pm b\,B \pm c\,C \pm \ldots,$$

hvor a, b, c...., er konstante Faktorer. Man har (jvf. den almindelige Formel for Middelfejlen paa en Funktion af Observationsstørrelser, sammenholdt med ovenstaaende):

$$m_p^2 + \mu_p = a^2\, (m_a^2 + \mu_a) + b^2\, (m_b^2 + \mu_b) + c^2\, (m_c^2 + \mu_c) \ldots,$$

$$m_p^2 = a^2\, m_a^2 + b^2\, m_b^2 + c^2\, m_c^2 \ldots,$$

altsaa $$\mu_p = a^2 \mu_a + b^2 \mu_b + c^2 \mu_c + \ldots$$

og deraf i Henhold til (1):

$$m_p\, (m_p) = a^2\, m_a\, (m_a) + b^2\, m_b\, (m_b) + c^2\, m_c\, (m_c) + \ldots$$

eller $$(m_p) = \frac{a^2\, m_a\, (m_a) + b^2\, m_b\, (m_b) + c^2\, m_c\, (m_c) + \ldots}{\sqrt{a^2\, m_a^2 + b^2\, m_b^2 + c^2\, m_c^2 + \ldots}} \tag{4}$$

Ligesom man kan finde Middelfejlen, kan man ogsaa finde Middelfejls-Middelfejlen paa en hvilken som helst Funktion. Grundidéen, der maa gaas frem efter, er den almindelige Fejlophobningslov. Vi sætter

$$u = f\,(x, y, z),$$

hvor x, y, z er indbyrdes uafhængige. Det forudsættes, at vi kender Middelfejl m_x, m_y, m_z og Middelfejls-Middelfejl (m_x), (m_y), (m_z) for x,

[1]) Som Udgangspunkt kan ogsaa sættes: Hvis Fejlen v_p paa en Størrelse er sammensat af flere, af hverandre uafhængige Fejlelementer

$$v_p = v_a \pm v_b \pm v_c \pm \ldots,$$

saa er $$m_p^2 + \mu_p = m_a^2 + \mu_a + m_b^2 + \mu_b + m_c^2 + \mu_c + \ldots$$

o. s. v.

y, z og vil finde de tilsvarende Størrelser m_u og (m_u) for u. Vi har da Fejlligningen
$$v_u = \frac{df}{dx} v_x + \frac{df}{dy} v_y + \frac{df}{dz} v_z,$$
hvor de partielle Differentialkvotienter har ganske bestemte Værdier. Sætter vi for at afkorte
$$\frac{df}{dx} = a, \quad \frac{df}{dy} = b, \quad \frac{df}{dz} = c,$$
bliver Fejlligningen
$$v_u = a\,v_x + b\,v_y + c\,v_z,$$
og heraf faas umiddelbart:
$$m_u^2 + \mu_u = a^2(m_x^2 + \mu_x) + b^2(m_y^2 + \mu_y) + c^2(m_z^2 + \mu_z),$$
$$m_u^2 \qquad = a^2 m_x^2 \qquad + b^2 m_y^2 \qquad + c^2 m_z^2,$$
og hermed er Sagen ført tilbage til Formel 4.

Er x, y, z indbyrdes afhængige, bliver Sagen langt mere indviklet. Nu beror Behandlingen alene paa Arten af Afhængighedsforholdet.

Formlen
$$m_p^2 + \mu_p = m_a^2 + \mu_a + m_b^2 + \mu_b + \ldots\ldots$$
er en almengyldig algebraisk Relation og kan altsaa behandles rent algebraisk. Kender man saaledes m_p og m_a — samt (m_p) og (m_a) — og vil finde m_b og (m_b), saa er
$$m_b^2 + \mu_b = m_p^2 + \mu_p \div m_a^2 \div \mu_a,$$
$$m_b^2 \qquad = m_p^2 \qquad \div m_a^2,$$
$$\overline{\mu_b = \qquad \mu_p \qquad \div \mu_a,}$$
og deraf (jvf. Udviklingen af Formel 3):
$$(m_b) = \frac{m_p(m_p) \div m_a(m_a)}{\sqrt{m_p^2 \div m_a^2}}. \tag{5}$$

(m) er betydelig mere usikker end m, og kan altsaa lettere komme ud for Urimeligheder, naar Materialet er fattigt.«

Eksempel. Et Forsøgsresultat er fremskaffet gennem Forsøg, der er udført paa 20 forskellige Steder. Forsøgene er anlagte som Dobbeltforsøg, saaledes at der er fremkommet to sideordnede Forsøgsresultater, a og b, paa hvert Forsøgssted.

Forskellen mellem a og b betragtes som Forsøgsfejl (Arbejdsfejl, tilfældig Fejl). Der spørges: Kan Forskellen fra Forsøgssted til Forsøgssted forklares som Arbejdsfejl, eller er der en særlig Variation, »Stedvariation«, til Stede? I saa Fald regner vi med, at den er uafhængig af Arbejdsfejlen, og at den i Hovedsagen følger den almindelige For-

Forsøg Nr.	a	b	Sum	Gennemsnit	Forskel, d
1	525	519	1044	522	6
2	454	394	848	424	60
3	474	481	955	478	÷ 7
4	512	475	987	494	37
5	488	419	907	454	69
6	477	486	963	482	÷ 9
7	436	484	920	460	÷48
8	436	435	871	436	1
9	447	504	951	476	÷57
10	536	540	1076	538	÷ 4
11	447	413	860	430	34
12	470	452	922	461	18
13	427	468	895	448	÷41
14	461	481	942	471	÷20
15	366	397	763	382	÷31
16	469	466	935	468	3
17	404	482	886	443	÷78
18	488	516	1004	502	÷28
19	410	401	811	406	9
20	438	489	927	464	÷51
Gennemsnit:	458	465	923	462	

delingslov (»Fejlloven«). En fælles Værdi for Arbejdsfejlen beregnes efter Formlen

$$m^2 = \frac{[[v^2]]}{20\,(2 \div 1)},$$

hvor v er Enkeltresultaternes Afvigelser fra Middeltallet af to sammenhørende Forsøgsresultater[1]). Man finder

$$m^2 = 739, \quad m = 27.2.$$

Middelfejlen paa denne Størrelse, 27.2, bliver

$$(m) = \frac{27.2}{\sqrt{2 \cdot 20\,(2 \div 1)}} = 4.3.$$

For Gennemsnitstallene (Gennemsnit af a og b) faar man:

$$m^2 = \frac{739}{2} = 370, \; m = \frac{27.2}{\sqrt{2}} = 19.2, \; (m) = \frac{4.3}{\sqrt{2}} = 3.0^{2)}.$$

Dernæst beregnes den totale Middelafvigelse af selve Gennemsnitstallene (522, 424 o. s. v., Divisor 20 ÷ 1), og man faar[3]):

[1]) I øvrigt er det simplest at beregne Arbejdsfejlen efter Formlen $m^2 = \frac{[d^2]}{2 \cdot 20}$, hvor d er Forskellen mellem a og b. Arbejdsfejlen kan naturligvis ogsaa være bestemt ad anden Vej, ved Markforsøg f. Eks. paa Grundlag af Fællesparcellerne.

[2]) (m) kan ogsaa beregnes i Overensstemmelse med Formel 4, idet Middeltallet af a og b er $\frac{1}{2}a + \frac{1}{2}b$:

$$(m) = \frac{(\frac{1}{2})^2 \cdot 27.2 \cdot 4.3 + (\frac{1}{2})^2 \cdot 27.2 \cdot 4.3}{\sqrt{(\frac{1}{2})^2 \cdot 27.2^2 + (\frac{1}{2})^2 \cdot 27.2^2}} = \frac{\frac{1}{2} \cdot 27.2 \cdot 4.3}{\sqrt{\frac{1}{2} \cdot 27.2^2}} = 3.0.$$

[3]) Afrundingsfejl kan undgaas ved at regne med Sumtallene i Stedet for med Gennemsnitstallene. Kvadratsummen maa da yderligere divideres med $2^2 = 4$.

$$m^2 = 1383,\ m = 37.2,\ (m) = \frac{37.2}{\sqrt{2\,(20 \div 1)}} = 6.0.$$

Heraf skal Arbejdsfejlen udskilles, og Stedvariationen bliver da:

$$m^2 = 1383 \div 370 = 1013,\ m = 31.8.$$

Middelfejlen paa Stedvariationen bliver (jvf. Formel 5):

$$(m) = \frac{37.2 \cdot 6.0 \div 19.2 \cdot 3.0}{\sqrt{37.2^2 \div 19.2^2}} = 5.2.$$

Der er altsaa *en udpræget Stedvariation, og den er bestemt med* ret stor Sikkerhed. Sammenstilles Resultaterne, har man:

	m^2	m
Totalafvigelse	1383	37.2 ± 6.0
Arbejdsfejl	370	19.2 ± 3.0
Stedvariation.............	1013	31.8 ± 5.2

Det bør bemærkes, at selv om vi kan betragte Arbejdsfejl og Stedvariation som uafhængige af hinanden, er der dog visse Baand til Stede, fordi Arbejdsfejl og Totalafvigelse er beregnet af de samme Elementer (i forskellige Kombinationer). Forbindelsen er udtrykt i Ligningen nederst paa Side 536, Tidsskrift for Planteavl, 39. Bind (»Om Variationsanalyse«). Betydningen af disse Baand er dog underordnet, da de kun gælder den Del af Totalafvigelsen, som repræsenterer Arbejdsfejlen, og vi kan derfor se bort fra dem i denne Forbindelse.

Beregnes Totalafvigelsen uden at sammendrage a- og b-Rækken til Gennemsnitstal, faar man:

$$m^2 = \frac{[[v^2]]}{2\,(20 \div 1)} = 1765,\ m = 42.0,$$

idet v er Afvigelserne fra de respektive Middeltal (458 og 465), og Stedvariationen bliver:

$$m = \sqrt{42.0^2 \div 27.2^2} = 32.0,$$

der stemmer særdeles tilfredsstillende med den Værdi (31.8), der blev beregnet ved Benyttelse af Gennemsnitstallene. Men Bestemmelsen af Middelfejlen paa Størrelsen 42.0 støder paa den Vanskelighed, at Stedvariationen i Virkeligheden kun hviler paa 20 reelle Gentagelser (hvis der ingen Arbejdsfejl fandtes, vilde a- og b-Rækken være ens), medens Arbejdsfejlen, der ogsaa indgaar i Totalafvigelsen, er bestemt af 20 Gange 2 Gentagelser. Det gør dog ingen nævneværdig Forskel, da Divisor i Middelfejlsformlen bliver omtrent uforandret; i første Tilfælde har man $D = 20 \div 1 = 19$, i sidste Tilfælde $D = 20\,(2 \div 1) = 20$. Regner vi med den for Stedvariationen gældende Divisor, bliver Middelfejlen paa Totalafvigelsen:

$$(m) = \frac{42.0}{\sqrt{2\,(20 \div 1)}} = 6.8,$$

medens Stedvariationen faar Middelfejlen

$$(m) = \frac{42.0 \cdot 6.8 \div 27.2 \cdot 4.3}{\sqrt{42.0^2 \div 27.2^2}} = 5.3,$$

eller praktisk talt det samme som før (5.2).

En Sammenstilling af disse Resultater, der altsaa gælder Enkelttallene, medens Sammenstillingen paa Side 366 gjaldt Gennemsnitstallene, giver:

	m^2	m
Totalafvigelse	1765	42.0 ± 6.8
Arbejdsfejl	739	27.2 ± 4.3
Stedvariation	1026	32.0 ± 5.3

At Stedvariationen er ens (med Tilnærmelse) i begge Sammenstillingerne, er et Udtryk for, at Enkelttal og Gennemsnitstal vilde have samme Variation, hvis der ingen Arbejdsfejl var til Stede.

Medens Stedvariationen her er vurderet paa Grundlag af den Usikkerhed, hvormed den er bestemt, idet det foreliggende Materiale opfattes som en Prøve, der er udtaget af et større Materiale, maa man ved en Vurdering efter R. A. Fishers z-Kriterium eller G. W. Snedecors F-Kriterium stille Spørgsmaalet saaledes: Er Forskellen mellem Arbejdsfejl og Totalafvigelse saa stor, at den ikke kan tilskrives den tilfældige Fordeling eller Gruppering af de Afvigelser, der repræsenterer Arbejdsfejlen? Med andre Ord: Kunde Forskellen fremkomme i et Materiale, hvor der ikke fandtes nogen reel Stedvariation? (Se f. Eks. *G. W. Snedecor:* Calculation and Interpretation of Analysis of Variance and Covariance, Side 15). Som en Fortsættelse af denne Tankegang er det naturligt at ræsonnere saaledes: Naar der ud over Arbejdsfejlen er en selvstændig Variation til Stede (i dette Tilfælde Stedvariationen), vil vi fremskaffe et direkte, talmæssigt Udtryk for denne Variation og samtidig vurdere dette Udtryk ved Hjælp af den Sikkerhed (eller Usikkerhed), hvormed Variationen er bestemt. Hertil tjener de Fremgangsmaader, som er beskrevet i dette og det foregaaende Arbejde om Variationsanalyse (Tidsskrift for Planteavl, 40. Bind, Side 835). Da (m) er en mere usikker Størrelse end m, og da Uregelmæssigheder i selve Resultaterne ogsaa kan gøre sig gældende og forøge Usikkerheden, maa Materialet til saadanne Beregninger ikke være for lille, men dette gælder selvfølgelig enhver Form af Variationsanalyse. Med passende Forbehold vil man i Reglen kunne benytte Middelfejlen paa Middelfejlen til en løselig Vurdering af Variationsberegningens Paalidelighed.