Yrkeshygienisk Institutt

ENKLE STATISTISKE BEREGNINGSMETODER, 3. utgave,

av

Jørgen Jahr, Overing.

Januar 1972, revidert september 1977.

Januar 1972.
Revidert mars 1973.

ENKLE STATISTISKE BEREGNINGSMETODER

Av Jørgen Jahr, Yrkeshygienisk Institutt, Oslo.

A. GAUSS-- FORDELING.

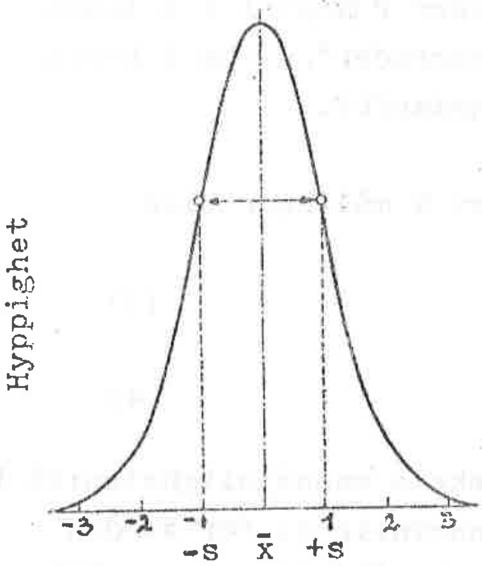

Fig.1. Normal fordeling.

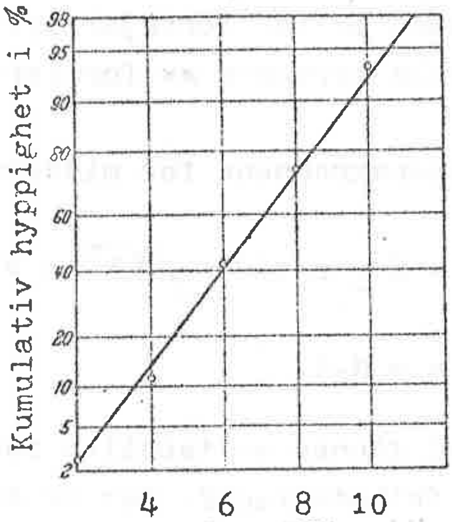

Fig. 2. Normal fordeling på sannsynlighetspapir.

Langs abcissen på sannsynlighetspapiret avsettes de øvre grenser i hver klasse. Hvis punktene tilnærmet ligger på en rett linje på sannsynlighetspapir med liniær abcisse, er Gauss - fordelingen normal. Fåes en rett linje med logaritmisk abcisse, er fordelingen log - normal,): at logaritmene til måleverdiene har en normal Gaussfordeling.

B. BEREGNING AV MIDDELVERDIER OG KONFIDENSGRENSER.*

B. 1. Normal fordeling.

Det aritmetiske middel $\bar{x}$ finnes ut fra N målinger av

$$\bar{x} = (\Sigma x)/N \qquad (1)$$

og standardavviket s av

$$s = \sqrt{\left[N\Sigma x^2 - (\Sigma x)^2\right]/N\cdot(N-1)} \qquad (2)$$

For M prøveserier, hver med N_j målinger, finnes s av (13), side 4. Det areal av kurven i fig. 1 som ligger mellom -s og +s omfatter 68,4 % av alle målingene. Tar man bare en måling x_1 er det 68,4 % sannsynlighet for at verdien ligger innenfor konfidensgrensene $x_1 \pm s$. De 68,4 % kalles sannsynlighets-

*Liste over de brukte symboler finnes i tabell. 1.

nivået og betegnes med P. Som regel vil man forlange at det skal være minst 95 % sannsynlighet for at målingen eller middelverdien skal ligge innenfor de angitte konfidensgrenser. Da må s multipliseres med en faktor t(P,n) = Student's t som finnes i vedlagte tab. 2 for en rekke sannsynlighetsnivåer P (angitt i % langs abcissen) og for forskjellige "frihetsgrader" n som i hvert tilfelle beregnes av formler angitt nedenfor.

Konfidensgrensene for middelverdien av N målinger blir

$$m = \bar{x} \pm s \cdot t(P,n) / \sqrt{N} = \bar{x} \pm K_i \qquad (3)$$

hvor $n = N-1$. (4)

t(P,n) finnes av tabellen for det ønskede sannsynlighetsnivå P % og n frihetsgrader. Det er da P % sannsynlighet for at den virkelige middelverdi ligger innenfor konfidensgrensene hvis målingene ikke har andre feil enn rent tilfeldige.

For en måleperiode over tidsrommet T med N-enkeltmålinger, hver tatt i løpet av tiden τ, blir konfidensgrensene for den aritmetiske middelverdi, $\bar{x}$,

$$m_T = (\Sigma x)/N \pm s \cdot t(P,n)\sqrt{(T-N\tau)/(T-\tau)(N-1)} \qquad (5)$$ hvor

$$n = (T - \tau)(N - 1)/(T - N \cdot \tau) \qquad (6)$$

og s finnes av (2)

B. 2. Log - normal fordeling.

Ved log-normal Gaussfordeling finnes den aritmetiske middelverdi $\bar{x}$ for N målinger av formlen (1) og konfidensgrensene av

$$\log m = \log \bar{x} \pm s_1 \cdot t(P,n)/\sqrt{N} \qquad (7)$$

hvor s_1 finnes av (2) ved å sette inn log x istedenfor x. Antall frihetsgrader n = N - 1 .

For en måleserie over tidsrommet T og med N-enkeltmålinger, hver tatt i løpet av tiden τ , blir konfidensgrensene for det aritm. $\bar{x}$ ved log - normal Gauss-fordeling

$$\log m_T = \log \bar{x} \pm s_1 \cdot t(P,n) \cdot \sqrt{(T-N\tau)/(T-\tau)(N-1)} \qquad (8)$$

med antall frihetsgrader som (6) og hvor s_1 beregnes etter (2) med innsetting av log x istedenfor x.

C. SAMMENLIGNING AV MIDDELVERDIER.

For å undersøke om det er noen statistisk signifikant forskjell (grensen settes ofte til 95% eller p = 0,05, dvs. at man er villig til å ta feil i 1 av 20 tilfelle) mellom to middelverdier $\overline{x_1}$ og $\overline{x_2}$ med henholdsvis N_1 og N_2 prøver, kan man bruke Student's t-test.

Ved normal fordeling er brukt formelen:

$$t = (\overline{x_1} - \overline{x_2}) \cdot \sqrt{N_1 \cdot N_2/(N_1+N_2)}/s \qquad (10)$$

hvor s finnes av (13) for $N = N_1 + N_2$ og M = 2.

Ved log - normal Gauss-fordeling brukes den tilsvarende formel:

$$t = (\log \overline{x_1} - \log \overline{x_2}) \cdot \sqrt{N_1 \cdot N_2/(N_1+N_2)}/s_1 \qquad (11)$$

Av tabell 2 kan man for en utregnet tallverdi av t og antall frihetsgrader = N_1+N_2-2 finne hvor stor sannsynlighet det er for at en forskjell mellom to middelverdier er reell.

D. MÅLEMETODENS PRESISJON (REPRODUSERBARHET).

Ved å ta flere målinger samtidig og på samme sted, kan man finne ut usikkerheten i selve metoden, ofte kalt presisjonen og betegnet ved s_p. Ut fra dobbeltmålinger x' og x" finner man ved normal fordeling

$$s_p = \sqrt{(\Sigma d^2)/2M} \qquad (12)$$

hvor d = x' - x'' og M er antall dobbeltmålinger.

For M prøveserier med N_j paralellbestemmelser i hver serie er

$$s_p = \sqrt{\frac{\left[\Sigma x_1^2 - (\Sigma x_1)^2/N_1\right] + \left[\Sigma x_2^2 - (\Sigma x_2)^2/N_2\right] + \cdots \left[\Sigma x_M^2 - (\Sigma x_M)^2/N_M\right]}{N - M}} \qquad (13)$$

Ved log - normal fordeling brukes de samme formler (12) og (13) men med innsetting av log x istedenfor x til beregning av s_1.

E. LINEÆR KORRELASJON.

Om det er noen sammenheng mellem f. eks. fluoridkonsentrasjonen i en arbeidsatmosfære i mg F^-/m^3 og konsentrasjonen av utskilt F^- i arbeidernes urin i mg F^-/l, kan man avgjøre ved en såkalt korrelasjonstest hvor korrelasjonskoeffisienten r beregnes etter formelen:

$$r = (N\cdot\Sigma xy - \Sigma x\cdot\Sigma y) \Big/ \sqrt{\left[N\Sigma x^2 - (\Sigma x)^2\right]\cdot\left[N\Sigma y^2 - (\Sigma y)^2\right]} \qquad (14)$$

Hvis r = 0 er det ingen sammenheng. Er tallverdien av r omkring 0,2, er det bare en svak korrelasjon, mens verdier omkring 0,5 viser en rimelig sammenheng. Hvis tallverdien av r er nær 1, er det en høy grad av korrelasjon mellom de to målesett. Positive verdier av r viser at høye verdier av x tilsvarer høye verdier av y og at lave verdier av x tilsvarer lave verdier av y. Er r negativ, svarer høye verdier av x til lave verdier av y og omvendt. Måleverdiene av x må være de sikreste.

Ved hjelp av t-testen:

$$t = \sqrt{r^2(N-2)/(1-r^2)} \qquad (15)$$

hvor antall frihetsgrader er N-2, kan man så av tabeller finne hvor stor sannsynlighet det er for at sammenhengen mellom målingene er reell og ikke bare skyldes tilfeldige målefeil. (C. Mack har angitt en annen test: $C = \sqrt{r^2(N-1)}$ hvor C > enn henholdsvis 1,96, 2,58 og 3,29 viser at det er henholdsvis 95, 99 og 99,9% sannsynlighet for at korrelasjonen er reell.)

F. LINEÆR REGRESJON

1. Liten eller ingen usikkerhet i x.

Hvis det er en signifikant korrelasjon mellom to sett av måleresultater x og y, beregnes sammenhengen mellom x og y av

$$y = a + bx \tag{16}$$

hvor det forutsettes at x-verdiene bare har små feil. I ligning 16 er:

$$b = (N\Sigma xy - \Sigma x \cdot \Sigma y)/[N\Sigma x^2 - (\Sigma x)^2] \text{ og} \tag{17}$$

$$a = \bar{y} - b \cdot \bar{x} \tag{18}$$

Standardavvikene for a og b finnes av formlene:

$$s_b = \sqrt{\{N \cdot \Sigma y^2 - (\Sigma y)^2 - b^2 [N \cdot \Sigma x^2 - (\Sigma x)^2]\}/\{[N\Sigma x^2 - (\Sigma x)^2](N-2)\}} \tag{19}$$

$$s_a = s_b \cdot \sqrt{\Sigma x^2 / N}$$

Hvis man på forhånd forutsetter (vet) at regresjonslinjen skal gå gjennom origo, blir ligningen forenklet til

$$y = b' \cdot x, \text{ hvor} \tag{20}$$

$$b' = \frac{\Sigma(xy)}{\Sigma x^2} \tag{21}$$

Standardavviket for b' finnes av

$$s_{b'} = \sqrt{\frac{N\Sigma y^2 - (\Sigma y)^2 - b'^2[N\Sigma x^2 - (\Sigma x)^2]}{N(N-1)\Sigma x^2}} \tag{22}$$

Konfidensgrensene for a, b og b' finnes på vanlig måte av formel (3), bortsett fra at antall frihetsgrader n = N - 2, ved å sette inn a,b,eller b' og de tilhørende verdier av s istedenfor x og s i formlen.

Standardavviket for punktenes spredning omkring regresjonslinjen (i y-retningen) er gitt av formlen

$$s_{xy} = \sqrt{\frac{\Sigma(y - Y)^2}{N - 2}} \tag{23}$$

hvor y er den målte verdi og Y regresjonslinjens verdi for en gitt verdi av x.

2. Hvis det er omtrent like stor usikkerhet i både x og y, kan man ifølge Bøckman (1972) bruke nedenstående verdier av a og b for ligningen

$$y = a + b \cdot x \tag{16}$$

$$b = \frac{s_y}{s_x} = \sqrt{\frac{N\sum y^2 - (\sum y)^2}{N\sum x^2 - (\sum x)^2}} \tag{24}$$

hvor man velger + når $r > 0$ og - når $r < 0$. Man får da:

For $r > 0$:

$$y = \bar{y} - b \cdot \bar{x} + b \cdot x \qquad \text{hvor } \bar{y} - b \cdot \bar{x} = a_1 \tag{25}$$

For $r < 0$:

$$y = \bar{y} + b \cdot \bar{x} - b \cdot x \qquad \text{hvor } \bar{y} + b \cdot \bar{x} = a_2 \tag{26}$$

Konfidensgrensene beregner Bøckman slik at konfidenslinjenes avstand fra regresjonslinjen i y-retningen blir

$$\pm\, t(P,n) \cdot s_y \cdot \sqrt{2\,(1 - |r|)}, \tag{27}$$

det vil si at man får to linjer parallellt med regresjonslinjen. (Beregningen av r forutsetter ifølge Doerffel at der bare er usikkerhet i y, men det er mulig man kan se bort fra dette krav.)

Det ville vært fordelaktig om man kunne beregnet standardavviket for a_1, a_2 og b også i dette tilfelle, men jeg har foreløpig ikke funnet noen formler for dette når både x og y er usikre.

G. CHI-KVADRAT-TEST.

Denne testen brukes for å avgjøre om hyppigheten av en eller flere foreteelser i en gruppe avviker fra det normale. Her redegjøres bare for testens enkleste form, nemlig hvor det bare er én foreteelse i gruppen som skal vurderes.

Det finnes to litt forskjellige formler for chi-test = χ^2. Den opprinnelige gir litt for høye verdier, og Yates har innført en korreksjon som imidlertid overkompenserer noe. Det er en fordel å beregne begge to; hvis begge ligger over det valgte kritiske nivå, eller begge ligger under er man sikker på resultatet. Hvis de to beregnede verdier ligger på hver sin side av den kritiske verdi, må man gjøre flere undersøkelser. De kritiske verdier av χ^2 for forskjellige sannsynlighetsnivåer er gjengitt i tabell 3.

Den generelle formel er

$$\chi^2 = \sum (o_i - e_i)^2 / e_i$$

hvor i er én av M-klasser og hvor antall frihetsgrader n = M - antall relasjoner mellom de observerte (o) og forventede (e) frekvenser av foreteelser.

NB! o og e må være dimensjonsløse, her brukes antallet foreteelser.

Med Yates' korreksjon er formelene

$$\chi_Y^2 = \sum (|o_i - e_i| - 0,5)^2 / e_i.$$

Eksempel:

I en bedrift hadde man 8 tilfeller av hjerteinfarkt blant 250 arbeidere i løpet av 1 år. Hvis det normale i en større, tilsvarende befolkningsgruppe med samme aldersfordeling og røkevaner var 1 %, er det da sannsynlig at bedriftens høye antall hjerteinfarkt skyldes tilfeldigheter?

Det er 2 klasser blant de 250 arbeidere, nemlig de som har fått hjerteinfarkt og de som ikke har fått hjerteinfarkt i løpet av året. Her er antall observerte o = 8, og det antall man skulle vente e = 2,5 for dem som har fått hjerteinfarkt, mens det for klassen som ikke har fått hjerteinfarkt er 242, mens man skulle vente 247,5. Vi har to klasser, og det er én sammenheng mellom de observerte tilfeller. Antall frihetsgrader er derfor 2 - 1 = 1. Her blir

$$\chi^2 = 12,10 + 0,12 = 12,22 \text{ og } \chi_Y^2 = 10,00 + 0,10 = 10,10.$$

Av tabellen over kritiske verdier for χ^2 i bilag 3, ser vi at begge de verdier vi har funnet er større enn den kritiske verdi ved 99,5% sannsynlighetsnivå, mens de ligger på hver sin side av 99,9%-verdien. Det er således over 99,5% sannsynlighet for at bedriften har flere tilfelle av hjerteinfarkt enn det normale. Det er mulig at sannsynligheten er over 99,9%.

Når det bare er hyppigheten av én foreteelse i en gruppe man skal undersøke, kan man bruke andre formler hvor antallet N i hele gruppen inngår:

$$\chi^2 = N (o - e)^2 / e (N - e)$$

$$\chi_Y^2 = N (|o - e| - 0,5)^2 / e (N - e).$$

For annen bruk av testen vises til faglitteraturen.

H. ENKEL VARIANSANALYSE

Ved t-testen kan man vurdere om to måleserier har middelverdier som for et valgt sannsynlighetsnivå (oftest 95%) er statistisk signifikant forskjellige. Ofte har man behov for å sammenligne fler enn 2 middelverdier. Til dette kan man bruke variansanalyse. Her er bare beskrevet den enkleste form, de mer avanserte må studeres i litteraturen.

For en gruppe forsøksresultater er variansen = kvadratet av standardavviket = s^2. Ved å dele opp en gruppe forsøksresultater i mindre grupper, kan man sammenligne to og to varianser ved en såkalt F-test, hvor

$$F_{2-1} = \frac{s_2^2}{s_1^2}, \quad F_{3-1} = \frac{s_3^2}{s_1^2}, \quad F_{3-2} = \frac{s_3^2}{s_2^2}, \text{ osv.}$$

hvor telleren i hver av brøkene er større enn nevneren. Til hver av variansene er knyttet et bestemt antall frihetsgrader, og av tabell 4 kan man finne ut om den beregnede F-verdi er større enn den tilsvarende kritiske verdi ved 95 eller 99% sannsynlighetsnivå (P). Hvis det er tilfelle, er gruppene signifikant forskjellige.

For å kunne utføre en variansanalyse, må forsøkene være lagt opp etter et bestemt mønster ("experimental design"), slik at resultatene danner et rektangel. Et eksempel illustrerer prinsippet:

I et gassrenseanlegg ønsket man å bestemme følgende:

1. optimal gassmengde pr. time
2. fordelingen av forurensningen over tverrsnittet av rense-enheten
3. om ett målepunkt kan gi et representativt resultat
4. målemetodens reproduserbarhet.

For å gjøre det enkelt, antas at fire målepunkter var tilstrekkelig. Måleresultatene er vist i skjemaet på neste side:

Gassmengde m^3/h	Måleresultater i punktene A	B	C	D	middel
2000	110	120	107	108	111,25
2500	105	110	107	106	107,0
3000	103	112	101	106	105,5
3500	108	118	105	105	109,0
4000	115	125	117	112	117,25
Middel	108,2	117,0	107,4	107,4	110,0

Man kan forandre enheten hvis dette gjør beregningene lettere, her er dette ikke aktuelt, men man kan forenkle regnearbeidet ved å trekke en konstant, f. eks. 100, fra hvert tall. Man får da følgende nye tabell:

	A	B	C	D	Sum_H	$\bar{x}_H$
2000	10	20	7	8	45	11,25
2500	5	10	7	6	28	7,00
3000	3	12	1	6	22	5,50
3500	8	18	5	5	36	9,00
4000	15	25	17	12	69	17,25
Sum_v	41	85	37	37	200	
X_v	8,2	17,0	7,4	7,4		10

Ved variansanalysen benytter man seg av "summen av kvadrater" = SS (efter engelsk "sum of squares") definert ved

$$SS = \sum_{1}^{N} (x_i - \bar{x})^2 = \sum_{1}^{N} x^2 - \frac{(\sum_{1}^{N} x)^2}{N}$$

Ved å planlegge forsøket på den måten som er vist foran, kan serien deles opp i mindre serier, hvor man for hver enkelt kan beregne en SS-verdi som refererer seg til enkeltmålingenes varians. Man får da

$$SS = SS_1 + SS_2 + \cdot \cdot + SS_j + \cdot \cdot + SS_M$$

hvor

$$SS_j = \sum_1^{N_j} x_{ij}^2 - \frac{(\sum_i^{N_j} x_{ij})^2}{N_j}$$

Også for frihetsgradene har man:

$$n = n_1 + n_2 + \cdots\cdots + n_j + \cdots\cdot + n_M$$

SS-verdiene for de enkelte delserier kan således adderes. Det samme gjelder frihetsgradene for delseriene. Den generelle formel for variansen kan da skrives:

$$s^2 = \frac{\sum_1^M SS}{\sum_1^M n}$$

For den enkelte, j-te gruppe blir

$$s_j^2 = \frac{SS_j}{n_j}$$

hvorav $SS_j = n_j \cdot s_j^2$

I eksemplet er det tre varianser som interesserer oss. Disse er forårsaket av:

a. forskjellig "gassflux"
b. målepunktenes forskjellige plassering
c. målemetodens usikkerhet (presisjonen)

For å beregne variansene må vi først finne de tilsvarende SS.
Dette gjøres slik:

1. Beregn $\sum x^2$ og $(\sum x)^2$ for hele gruppen.

$$SS_{totalt} = \sum x^2 - \frac{(\sum x)^2}{N_{totalt}} = 2734 - \frac{40\ 000}{20} = 734$$

2. For middelverdiene $\overline{x}_H$ av målingene i A, B, C og D for hver "gassflux" kan man beregne

$$SS_{Vm} = \sum \overline{x}_{Hi}^2 - \frac{(\sum \overline{x}_{Hi})^2}{N_V}$$

(V = vertikalt, H = horisontalt)

hvor N_v her er 5. Da det for hver middelverdi er N_H = 4 målinger, blir

$$SS_V = N_H \cdot SS_m = 4 \cdot SS_m$$

hvor SS_V er "sum of squares" redusert til enkeltmåling ved varierende"gassflux".

Samtidig er:

$$\overline{x}_{Hi} = \frac{\sum x_{Hi}}{N_H} \quad \text{slik at}$$

$$SS_V = N_H \left[\sum \left(\frac{\sum x_{Hi}}{N_H} \right)^2 - \frac{\left(\sum \frac{\sum x_{Hi}}{N_H} \right)^2}{N_V} \right]$$

$$SS_V = \sum \frac{(\sum x_{Hi})^2}{N_H} - \frac{(\sum x)^2}{N_H \cdot N_V}$$

$$SS_V = \frac{\sum (\sum x_{Hi})^2}{N_H} - \frac{(\sum x)^2}{N_{total}} = \frac{9350}{4} - 2000 = 337{,}5.$$

3. Analogt til pkt. 2 beregnes SS_H for middelverdiene $\overline{x}_{Vm}$ for de forskjellige målepunkter A, B, C og D,

$$SS_H = \frac{\sum(\sum x_{Vi})^2}{N_V} - \frac{(\sum x)^2}{N_{total}}$$

4. Vi kan nå sette opp variansanalysen slik:

Kilde til variansen	n	SS	s^2	s	F
Forskjellig "gassflux"	4	337,5	84,375	9,186	14,96
Forskjellige målepunkter	3	328,8	109,600	10,469	19,43
Presisjon*	12	67,7	5,642	2,375	
Totalt	19	734,0			

Kritiske verdier av F for	P = 95%	P = 99%
Forskjellig "gassflux"	3,26	5,41
Forskjellige målepunkter	3,49	5,95

Av tabellen over kritiske verdier av F ovenfor ser man at det er over 99% sannsynlighet for at det var en reell forskjell på resultatene både ved variasjon av "gassfluxen" og for de forskjellige målepunkter.

Man kan trekke følgende konklusjoner:

1. Laveste middelkonsentrasjon, 105,5, ble funnet for "gassflux" 3000 som derfor antas å være nær den optimale verdi.
2. Fordelingen av forurensningen over tverrsnittet av renseenheten var ujevn.
3. Målepunkt A ligger med 108,2 nærmest midlet av alle måleresultatene som var 110. Med t-testen får man:

*hvor n og SS finnes som differans mellem "Totalt" og de tilsv. verdiene for de andre kilder til variansen.

$$t = \frac{110 - 108,2}{2,375} \cdot \sqrt{\frac{5 \cdot 20}{25}} = 1,515$$

(Den verdi som er brukt for s er presisjonen, dette gir sannsynligvis en noe for høy t-verdi, men er bedre enn s for "gassfluxen" som her ikke kan brukes.)

Med n = 18 frihetsgrader er det mellom 80 og 90% sannsynlighet for at avvikelsen fra totalmidlet var reell. Forskjellen er ikke signifikant på 95 %-nivået, men det bør gjøres flere målinger hvis kravet til nøyaktighet er stort. Eventuelt kunne man flytte punkt A nærmere punkt B.

4. Målemetodens reproduserbarhet er tilfredsstillende med s = 2,375 eller 2,16% av middelverdien.

LITTERATUR

Klaus Doerffel, Z. Analyt. Chemie, Bd. 195, Heft 1, 1962.

C. Mack: Essentials of Statistics for Scientists and Technologists. Plenum Press, New York, 1967.

Juda u. Budzinski, Staub - Reinh. Luft, 27, April 1967, p. 176-179.

Oluf Chr. Bøckman: En kort orientering om statistisk kontroll av forsøksresultater. Fiskaa Verk, Kristiansand S. Stensil. 19.5.1953.

W. Duckworth: Statistical techniques in technological research; an aid to research productivity. London, Methuen, c. 1968.

Oluf Chr. Bøckman: Kort kurs i statistisk analyse, des. 1972.

Symbol-liste

d = differans mellom to (måle-)verdier

G = Geometrisk middel $= \sqrt[N]{x_1 \cdot x_2 \cdot x_3 \ \ldots \cdot x_N} = 10^{\frac{\sum \log x}{N}}$

): $\log G = \overline{\log x}$ = middel av alle log x.

F = s_2^2/s_1^2

K_i = Konfidensintervall

M = Antall måleserier

m = $\bar{x} \pm K_i$ = Konfidensgrenser mellom hvilke den virkelige middelverdi antas å ligge med sannsynlighet P%

N = Antall målinger

N_j = Antall målinger i den j-te måleserie

n = Antall frihetsgrader

P = Sannsynlighetsnivå i %

p = $1 - \frac{P}{100}$ en annen måte å angi sannsynlighetsnivå på

r = korrelasjonskoeffisient

s = standard avvik

s_l = standard avvik for logaritmene til måleverdiene

s_p = presisjonens standardavvik

Σ = summetegn; $\Sigma x = x_1 + x_2 + \ \ldots \ + x_N$

SS = "Sum of Squares", Sum av kvadrater.

T = tidsrom for en måleperiode, f. eks. en uke, men oftest angitt i timer.

t = Student's t

$t(P,n)$ = Student's t for sannsynlighetsnivået P % og n frihetsgrader

τ = den tid den enkelte prøve er tatt i løpet av, med samme tidsenhet som for T (oftest timer)

x = enkelt måleverdi

$\bar{x}$ = Aritmetisk middel $= \frac{x_1 + x_2 + \ \ldots\ldots + x_N}{N} = \frac{\Sigma x}{N}$

χ^2 = test for kontroll av om funnet hyppighet er normal

χ_Y^2 = Yates' korrigerte χ^2 - test.

/ er brukt istedenfor brøkstrek i teksten.

Tabell 2.

Student's t

n \ P%	50	60	80	90	95	97,5	99	99,5	99,9
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	2.368	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

Tabell 3.

Critical Values of χ^2

Note: For $\nu > 100$, use the formula $\chi^2 = 0{\cdot}5\{\beta + \sqrt{(2\nu - 1)}\}^2$, the values of β are in the last column below

P% \ n	1	2	3	4	5	6	7	8	9	10	12
0,5	0·000	0·010	0·072	0·207	0·412	0·676	0·989	1·34	1·73	2·16	3·07
1	0·000	0·020	0·115	0·297	0·554	0·872	1·24	1·65	2·09	2·56	3·57
2,5	0·000	0·051	0·216	0·484	0·831	1·24	1·69	2·18	2·70	3·25	4·40
5,0	0·000	0·103	0·352	0·711	1·15	1·64	2·17	2·73	3·33	3·94	5·23
95	3·84	5·99	7·81	9·49	11·1	12·6	14·1	15·5	16·9	18·3	21·0
97,5	5·02	7·38	9·35	11·1	12·8	14·4	16·0	17·5	19·0	20·5	23·3
99	6·63	9·21	11·3	13·3	15·1	16·8	18·5	20·1	21·7	23·2	26·2
99,5	7·88	10·6	12·8	14·9	16·7	18·5	20·3	22·0	23·6	25·2	28·3
99,9	10·8	13·8	16·3	18·5	20·5	22·5	24·3	26·1	27·9	29·6	32·9

P% \ n	14	16	18	20	22	24	26	28	30	40	60	100	β
0,5	4·07	5·14	6·26	7·43	8·64	9·89	11·2	12·5	13·8	20·7	35·5	67·3	−2·58
1	4·66	5·81	7·01	8·26	9·54	10·9	12·2	13·6	15·0	22·2	37·5	70·1	−2·33
2,5	5·63	6·91	8·23	9·59	11·0	12·4	13·8	15·3	16·8	24·4	40·5	74·2	−1·96
5	6·57	7·96	9·39	10·9	12·3	13·8	15·4	16·9	18·5	26·5	43·2	77·9	−1·64
95	23·7	26·3	28·9	31·4	33·9	36·4	38·9	41·3	43·8	55·8	79·1	124·3	+1·64
97,5	26·1	28·8	31·5	34·2	36·8	39·4	41·9	44·5	47·0	59·3	83·3	129·6	+1·96
99	29·1	32·0	34·8	37·6	40·3	43·0	45·6	48·3	50·9	63·7	88·4	135·8	+2·33
99,5	31·3	34·3	37·2	40·0	42·8	45·6	48·3	51·0	53·7	66·8	92·0	140·2	+2·58
99,9	36·1	39·3	42·3	45·3	48·3	51·2	54·1	56·9	59·7	73·4	99·6	149·4	+3·09

Tabell 4.

(Fra Duckworth: Statistical Techniques in Technological Research.)

Probability points of the variance ratio (*F-Distribution*)

Probability Point p*	ϕ_d	ϕ_n (Corresponding to greater mean square)																		
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
0.100	1	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.7	61.2	61.7	62.0	62.3	62.5	62.8	63.1	63.3
0.050		161	199	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
0.025		648	800	864	900	922	937	948	957	963	969	977	985	993	997	1001	1006	1010	1014	1018
0.010		4052	4999	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
0.100	2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
0.050		18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
0.025		38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5	39.5
0.010		98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
0.100	3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
0.050		10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
0.025		17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.3	14.2	14.1	14.1	14.0	14.0	13.9	13.9
0.010		34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
0.100	4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
0.050		7.71	6.49	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
0.025		12.2	10.6	10.0	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
0.010		21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
0.100	5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
0.050		6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
0.025		10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
0.010		16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
0.100	6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
0.050		5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
0.025		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
0.010		13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
0.100	7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
0.050		5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
0.025		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
0.010		12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
0.100	8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
0.050		5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
0.025		7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
0.010		11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
0.100	9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
0.050		5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
0.025		7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
0.010		10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
0.100	10	3.28	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
0.050		4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.74	2.70	2.66	2.62	2.58	2.54
0.025		6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.83	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
0.010		10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
0.100	12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
0.050		4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
0.025		6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
0.010		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
0.100	15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.92	1.87	1.85	1.82	1.79	1.76
0.050		4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
0.025		6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
0.010		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
0.100	20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
0.050		4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
0.025		5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
0.010		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42

* Probability point p	=	Sannsynlighetsnivå P %
0.100		90.0
0.050		95.0
0.025		97.5
0.010		99.0