

Jdf

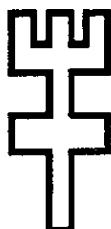
**Onderafdeling
der Wiskunde**

**Toegepaste
Statistiek**

SYLLABUS VAN HET COLLEGE
VAN

PROF. DR. H.C. HAMAKER

VOORJAARSSEMESTER 1967



**TECHNISCHE HOGESCHOOL
EINDHOVEN**

TECHNISCHE HOGESCHOOL EINDHOVEN

Afdeling Algemene Wetenschappen

Onderafdeling der Wiskunde

TOEGEPASTE STATISTIEK

Syllabus naar het college van

Prof. Dr. H.C. Hamaker

Voorjaarssemester 1967

Onderafdeling der Wiskunde

TOEGEPASTE STATISTIEK

Syllabus van het college

van

Prof. dr. H. C. Hamaker

voorjaarssemester 1967

T e c h n i s c h e H o g e s c h o o l E i n d h o v e n

Inhoudsbeschrijving

TOEGEPASTE STATISTIEK

H.C. Hamaker

voorjaarssemester 1967

Gewenste voorkennis. Definities en symbolen	1
Bewerking van een uitgebreide waarnemingsreeks in een frequentietabel	4
Twee-dimensionale waarnemingsreeksen	6
Formules voor het toevoegen, weglaten of wijzigen van waarnemingen	9
Uitzetten van een kleine reeks waarnemingen	12
De normale verdeling	14
Populatie en steekproef	15
Praktijkmoeilijkheden bij aselekt trekken	17
De centrale limietstelling	22
Parameters, schatters en schattingen	23
Betrouwbaarheidsintervallen voor μ	24
Betrouwbaarheidsintervallen	28
Voorbeelden van betrouwbaarheidsintervallen aan de hand van Tabel 27	32
Discrete verdelingen $x = 0, 1, 2, \dots$	37
Toepassingen van betrouwbaarheidsintervallen bij tellingen	41
Het toetsen van hypothesen	42
Enige gangbare toetsen	44
Steekproefkeuringen	47
Practische aspecten	56
De χ^2 -verdeling en haar toepassingen	57
De verdeling van s^2	58
Het vergelijken van fracties of percentages	59
Grafische methoden	67
Kruisjeskaarten	69
De \bar{x} - en R -kaarten	73
Variantie analyse	79
De F -toets	83
Regressie analyse	84
Foutendiscussies en de wet van de voortplanting van fouten	89
Voorbeelden	92

Toegepaste statistiek 1965

Enquete

Naam en voorletters :

Geboorte datum :

Studie-richting :

Lengte :

Gewicht :

Grootte van het gezin waartoe U behoort, Uzelf meegerekend :

Totaal: Jongens: Meisjes:

Rookt U : ja, neen

Zo ja hoeveel per dag : sigaretten :

 sigaren :

 pijpen :

Waarnemingen van de secundewijzer :

t1 :

t2 :

t3 :

t4 :

t5 :

t6 :

Welke middagen heeft U vrij voor practicum:

Ma, Di, Wo, Do, Vr,

Gewenste voorkennis: Aanvullingsdiktaat Wiskunde IV; kansberekening.

Bij dit kollege behoren:

1. Een statistisch compendium, bewerkt door A.J. Bosch en J.H.L. Kamps.
2. Vraagstukken met oplossingen, bewerkt door A.J. Bosch, beide verkrijgbaar bij de Administratie.

Aanbevolen boekwerken:

1. H.W. Alexander: Elements of Mathematical Statistics.
 2. P.G. Hoel: Introduction to Mathematical Statistics, beide ca. f 35,-- uitgegeven door Wiley and Sons, New York, eventueel gezamenlijk bestellen door tussenkomst van Sunya.
-

Definities en symbolen

1. Een populatie: Een verzameling gelijksoortige objecten of elementen.
2. Een statistische populatie: Een verzameling metingen die men aan de elementen van een populatie zou kunnen verrichten.
3. Een steekproef: Een deel van een populatie waaraan daadwerkelijk metingen worden verricht.
4. Een reeks waarnemingen: De waarnemingen aan een steekproef uitgevoerd.
5. a) De steekproefgrootte
b) Het aantal waarnemingen : } n
6. De individuele waarnemingen : $x_i, i = 1, \dots, n,$
7. Hun som : $x_{\cdot} = \sum_i x_i,$
8. Het gemiddelde : $\bar{x} = \bar{x} = \frac{x_{\cdot}}{n},$
9. De kwadratensom : $KS = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - x_{\cdot}^2/n = \sum_i x_i^2 - nx_{\cdot}^2/n.$
10. De variantie : $s^2 = KS/(n-1),$
11. De standaardafwijking : $s = \sqrt{s^2},$
12. Aantal vrijheidsgraden : $v = (n-1),$
13. De spreidingsbreedte : $R = x_{\max} - x_{\min},$
14. De mediaan : M = middelste waarnemingen als n even is;
M = het gemiddelde van het middelste paar als n oneven is.
15. Het afrondingsinterval : a = kleinste positieve verschil tussen de afgeronde waarnemingen.

16. Toelaatbare afronding : voor x_i : $a < \frac{R}{2\sqrt{n}}$ of $a < \frac{s}{2}$,
 voor \bar{x} en s : $a < \frac{R}{2n}$ of $a < \frac{s}{2\sqrt{n}}$,
 voor s^2 : $a < \frac{2s^2}{n}$.

17. Een getal met een 5 als laatste cijfer afronden op een even cijfer.

18. Koderen: Het vervangen van x_i door $y_i = \frac{(x_i - c)}{a}$ ter vereenvoudiging
 van de waarnemingen en de berekeningen $\bar{x} = c + a\bar{y}$, $s_x^2 = a^2 s_y^2$.

Oefenmateriaal

Dikten van 75 hardpapieren schijfjes in mikrons

3085	2988	3041	3118	3069
3080	3059	3032	3091	3098
3096	3045	3041	3107	3041
3068	3030	3080	3083	3053
3068	3073	3115	3088	3040
3077	3065	3087	3005	3087
3050	3050	3041	3016	3093
2948	2999	3071	3039	3078
3089	3041	3033	3088	3067
3065	3131	3030	3083	3058
3058	3050	3059	3097	3061
3080	3081	3005	3028	3059
3115	3073	3048	3072	3023
3058	3082	3085	3057	3095
3078	3093	2923	3090	3110

Gewichten van dezelfde 75 schijfjes in dezelfde volgorde in
miligrammen

1141	1110	1107	1146	1130
1131	1124	1102	1141	1130
1126	1116	1112	1139	1131
1129	1094	1128	1145	1103
1138	1119	1130	1144	1116
1127	1126	1141	1092	1155
1105	1139	1097	1114	1145
1141	1109	1125	1122	1131
1136	1135	1092	1142	1109
1144	1136	1131	1150	1129
1154	1116	1128	1154	1135
1134	1130	1100	1133	1134
1132	1164	1132	1139	1085
1124	1134	1148	1130	1140
1144	1146	1154	1134	1140

1. Bewerking van een uitgebreide waarnemingsreeks in een frequentie tabel

- Regels:
- a) Waarnemingen vooraf niet afronden,
 - b) Aantal klassen bij voorkeur >5 en <16 ,
 - c) Ondubbelzinnige klassegrenzen,
 - d) Bij voorkeur een oneven aantal malen de laatste decimaal per klasse.

Tabel 1.4. Frequentietabel van de dikten van 30 hardpapieren schijfjes uit tabel 1.3.

Klasse grenzen μ	Turfstaat	frequen- tie f	y	fy	fy ²
2938 - 2962	/	1	-5	-5	25
2963 - 2987		0	-4	0	0
2988 - 3012	//	2	-3	-6	18
3013 - 3037	////	4	-2	-8	16
3038 - 3062	//// ///	8	-1	-8	8
3063 - 3087	//// //// /	11	0	0	0
3088 - 3112	//	2	1	2	2
3111 - 3137	//	2	2	4	8
	50 m. =	30		-21	77

Berekening: $\bar{y} = -21/30 = -0.70$

$$KS_y = 77 - \frac{21^2}{30} = 62.3$$

$$s_y^2 = \frac{62.3}{29} = 2.14, \quad s_y = 1.46.$$

Hieruit

$$n = 30$$

$$\bar{x} = 3075 + 25(-0.70) = 3057.5\mu$$

$$s_x = 25 * 1.46 = 36.6\mu.$$

Toelaatbare afronding voor \bar{x} en s: $a < \frac{36.6}{2\sqrt{30}} = 3.3\mu.$

We kunnen dus met $n = 30$, $\bar{x} = 3058\mu$, $s = 37\mu$ volstaan.

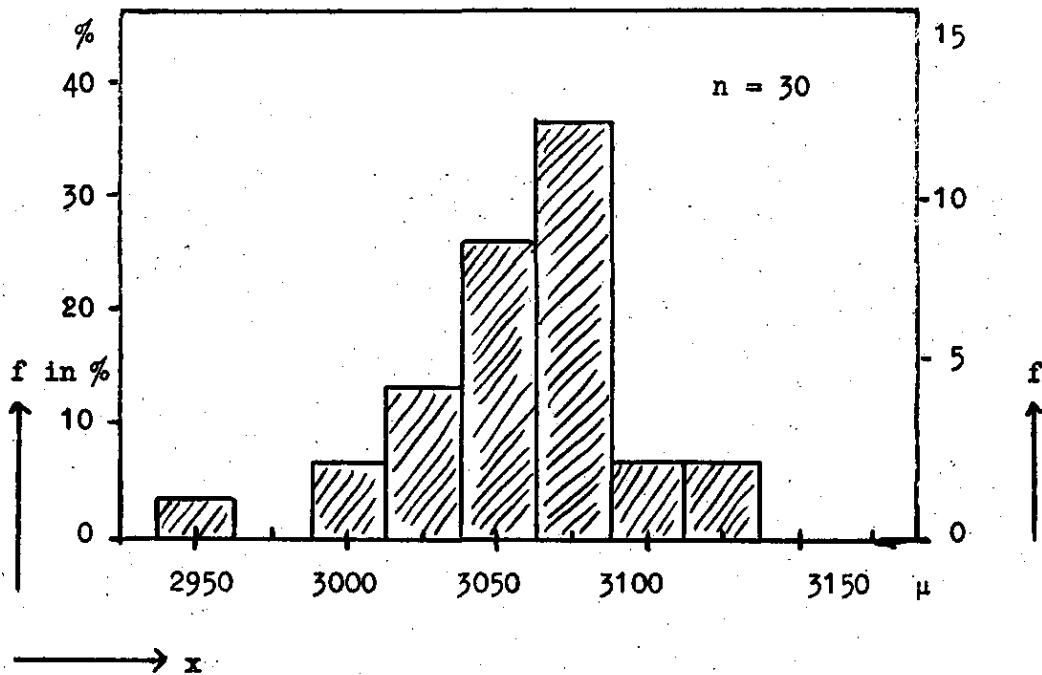
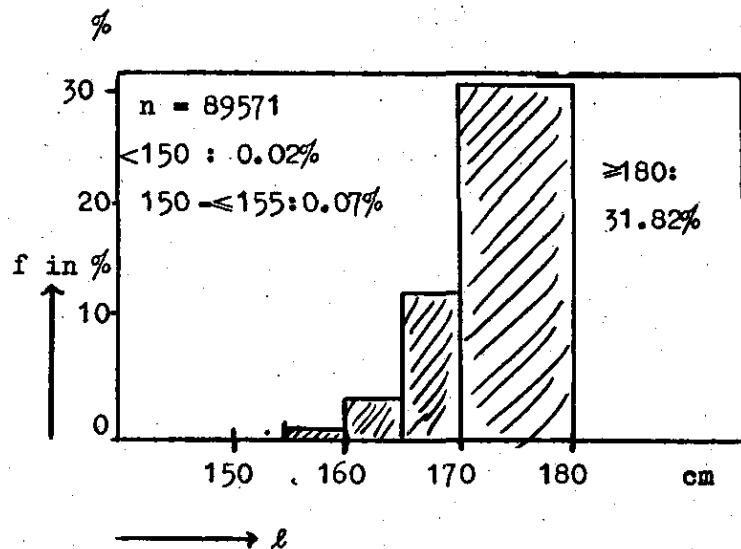


Fig. I.5.A. Histogram van de dikten van 30 hardpapier schijfjes uit Tabel I.4

Voorbeeld van een histogram met ongelijke klassen.

klassen cm	f in %
< 150	0.02
150 - <155	0.07
155 - <160	0.46
160 - <165	2.70
165 - <170	10.68
170 - <180	54.25
≥ 180	31.82
100.00	
n = 89 571	



Tabel en Fig. I.5.B. Lengten van rekruten in 1963.

Bron: Statistisch zakboek 1964.

Regels bij het construeren van Histogrammen

- Steeds het totaal aantal waarnemingen vermelden.
- Frequentie f bij voorkeur in % voor onderlinge vergelijkbaarheid.
- Het oppervlak van de kolom boven een klasse interval is evenredig met de frequentie! (van belang bij verschillende klasse breedten)
- Onvolledig gedefinieerde klassen (\geq of $<$) niet tekenen doch frequenties apart vermelden.

Algemene regels voor Tabellen en Figuren

- a) Goede tabellen en figuren moeten gemakkelijk leesbaar en begrijpelijk zijn. Zij moeten steeds van een duidelijk onderschrift zijn voorzien.
- b) Men moet zich vaak de moeite getroosten figuren en tabellen enige malen over te maken ten einde een zo sprekend mogelijk resultaat te verkrijgen.
- c) Dimensies van de schalen moeten steeds worden vermeld.

Voorbeelden van het nut van histogrammen

- 1. Kopshulzen (AKU). Tweetoppige verdeling.
- 2. Staalstaaf 6 mm ϕ . Twee partijen met verschil in spreiding. Verband met toleranties.
- 3. Afgeknotte verdelingen bewijzen sortering.

Twee-dimensionale waarnemingsreeksen

$$x_{1i}, x_{2i}, \quad i = 1, \dots, n$$

Naast $\bar{x}_1, \bar{x}_2, s_{x_1}^2, s_{x_2}^2$ enz. zijn nu nog van belang:

$$1. \text{ de Productsom} = PS = \sum_1 (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$$= \sum_1 x_{1i} x_{2i} - n\bar{x}_1 \bar{x}_2 = \sum_1 x_{1i} x_{2i} - \frac{x_1 \cdot x_2}{n}$$

$$2. \text{ cov}(x_1, x_2) = PS / (n-1)$$

$$3. \text{ de correlatie coëfficiënt } r = \frac{\text{cov}(x_1, x_2)}{\sqrt{\text{var } x_1 \cdot \text{var } x_2}} = \frac{PS}{\sqrt{KS_{x_1} \cdot KS_{x_2}}}$$

Voorbeeld. Dikten en gewicht van 6 hardpapieren schijfjes.

x_{1i}	= 3085	3080	3096	3068	3068	3077 μ .
x_{2i}	= 1141	1131	1126	1129	1138	1127 mgr.

Zet eerst x_1 en x_2 tegen elkaar uit.

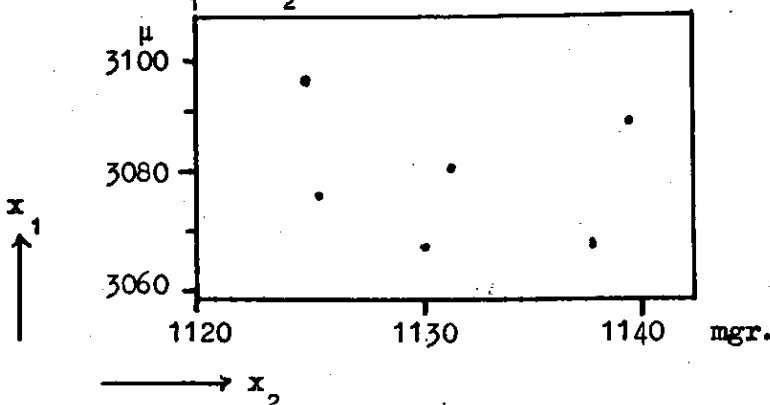


Fig. I.6. De dikte x_1 , van 6 hardpapieren schijfjes uitgezet tegen het gewicht. x_2 .

Berekening met gecodeerde waarnemingen

$$\begin{array}{rcccccc}
 y_0 = & & 1 & 1 & 1 & 1 & 1 & 1 \\
 y_1 = x_1 - 3000 = & 85 & 80 & 96 & 68 & 68 & 77 & \mu. \\
 y_2 = x_2 - 1100 = & 41 & 31 & 26 & 29 & 38 & 27 & \text{mgr.}
 \end{array}$$

Met afronden is enige voorzichtigheid geboden. Berekening met een rekenmachine.

$$(850000041)^2 = 7225\ 0006790\ 0001681$$

$$(800000031)^2 = 6400\ 0004960\ 0000961$$

$$\begin{array}{r}
 \dots\dots\dots \\
 \hline
 4740000192 \quad 38018\ 0030192\ 0006332
 \end{array}$$

optellen

$$\Sigma \quad y_1 \quad y_2 \quad y_1^2 \quad 2y_1y_2 \quad y_2^2$$

Matrix van ruwe productsommen

	y_0	y_1	y_2
y_0	6	474	192
y_1		38018	15096
y_2			6332

Matrix van gecorrigeerde productsommen

	y_1	y_2
y_1	572	-72
y_2		188

Hieruit volgt: $\bar{x}_1 = 3000 + 474/6 = 3079 \mu$

$$\bar{x}_2 = 1100 + 192/6 = 1132 \text{ mgr.}$$

$$\text{var } x_1 = 572/5 = 114.4 \mu^2,$$

$$\text{var } x_2 = 188/5 = 37.6 \text{ mgr}^2.$$

$$\text{cov}(x_1, x_2) = -72/6 = -12.0 \mu \cdot \text{mgr.}$$

$$r = \frac{-12.0}{\sqrt{37.6 \cdot 114.4}} = -0.18$$

Tabel I.8. Een twee-dimensionale frequentietabel.

Dikten en gewichten van 30 hardpapieren schijfjes uit Tabel.

Dikten in microns	Gewicht in mgr.					$n_{i.}$	$\sum_j n_{ij} y_{2j}$
	x_2	1090- 1104	1105- 1119	1120- 1134	1135- 1169		
x_1	$y_{1i} \backslash y_{2j}$	0	1	2	3	$n_{i.}$	$\sum_j n_{ij} y_{2j}$
2930 - 2964	0		n_{ij}		1	1	3
2965 - 2999	1		2			2	2
3000 - 3034	2	3	1	1		5	3
3035 - 3069	3	2	2	3	4	11	20
3070 - 3104	4		1	5	3	9	20
3105 - 3139	5			1	1	2	5
	$n_{.j}$	5	6	10	9	30	
	$\sum_i n_{ij} y_{1i}$	12	14	36	29		$\sum_{ij} n_{ij} y_{1i} y_{2j} = 173$

$$\sum n_{i.} y_{1i} = 91$$

$$\sum n_{.j} y_{2j} = 53,$$

$$\bar{y}_1 = 3.03$$

$$\bar{y}_2 = 1.77,$$

$$KS(y_1) = 39.0$$

$$KS(y_2) = 36.6,$$

$$PS(y_1, y_2) = 173 - \frac{91 * 53}{30} = 12.2.$$

$$r(x_1, x_2) = r(y_1, y_2) = \frac{12.2}{\sqrt{39.0 * 36.6}} = 0.32$$

$$PS(x_1, x_2) = 12.2 * 35 * 15 = 6405 \mu \text{ mgr.}$$

Formules voor het toevoegen, weglaten of wijzigen van waarnemingen

Samenvoegen van twee reeksen:

$$n = n_1 + n_2, \quad \bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n}, \quad KS = KS_1 + KS_2 + \frac{n_1 n_2}{n} (\bar{x}_1 - \bar{x}_2)^2,$$

Weglaten van n waarnemingen:

$$n_1 = n - n_2, \quad \bar{x}_1 = \frac{n\bar{x} - n_2 \bar{x}_2}{n_1}, \quad KS_1 = KS - KS_2 - \frac{n n_2}{n_1} (\bar{x}_2 - \bar{x})^2.$$

Toevoegen van één waarneming, x_{n+1}

$$n = n_1 + 1, \quad \bar{x} = \frac{n\bar{x}_1 + x_{n+1}}{n+1}, \quad KS = KS_1 + \frac{n}{n+1} (x_{n+1} - \bar{x}_1)^2,$$

Weglaten van één waarneming, x_n

$$n_1 = n - 1, \quad \bar{x}_1 = \frac{n\bar{x} - x_n}{n-1}, \quad KS_1 = KS - \frac{n}{n-1} (x_n - \bar{x})^2.$$

Wijzigen van één waarneming x_k in x'_k

$$\bar{x}' = \frac{n\bar{x} - x_k + x'_k}{n}, \quad KS' = KS - \frac{n}{n-1} (x_k - \bar{x})^2 + \frac{n}{n-1} (x'_k - \bar{x}')^2$$

De KS voor een serie van twee waarnemingen is

$$n = 2, \quad KS = \frac{(x_1 - x_2)^2}{2}.$$

Voor twee-dimensionale reeksen gelden de formules:

$$n = n_1 + n_2, \quad PS = PS_1 + PS_2 + \frac{n_1 n_2}{n} (\bar{x}_{1,1} - \bar{x}_{1,2})(\bar{x}_{2,1} - \bar{x}_{2,2}),$$

$$n_1 = n - n_2, \quad PS_1 = PS - PS_2 - \frac{n_1 n_2}{n_1} (\bar{x}_{1,2} - \bar{x}_1)(\bar{x}_{2,2} - \bar{x}_2),$$

$$n = n_1 + 1, \quad PS = PS_1 + \frac{n_1}{n+1} (x_{1,n+1} - \bar{x}_{1,1})(x_{2,n+1} - \bar{x}_{2,1}),$$

$$n_1 = n - 1, \quad PS_1 = PS - \frac{n}{n-1} (x_{1,n} - \bar{x}_1)(x_{2,n} - \bar{x}_2).$$

De normale verdeling: $x \cong \mu + u\sigma$; x is $N(\mu, \sigma)$.

$$\varphi(x)dx = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Compendium Tabel 1.1.

Aanpassing bij een histogram

$\mu = \bar{x} = 3058$ micron, $s = 37$ micron, $\Delta x = 25$ micron.

\bar{x}		u	$\varphi(u)$	$\varphi(x)\Delta x = \frac{25}{37}\varphi(u)$
$\bar{x} =$	3058 micron		0.399	0.270 = 27.0 %
$\bar{x} \pm 0.5s$	3040 , 3076 "	± 0.5	0.352	0.238 = 23.8 "
$\bar{x} \pm 1.0s$	3021 , 3095 "	± 1.0	0.242	0.163 = 16.3 "
$\bar{x} \pm 1.5s$	3003 , 3113 "	± 1.5	0.130	0.088 = 8.8 "
$\bar{x} \pm 2.0s$	2984 , 3132 "	± 2.0	0.054	0.036 = 3.6 "
$\bar{x} \pm 2.5s$	2966 , 3150 "	± 2.5	0.018	0.012 = 1.2 "
$\bar{x} \pm 3.0s$	2947 , 3169 "	± 3.0	0.004	0.003 = 0.3 "

Met deze gegevens kan in Fig. 1.5.A de kromme worden ingetekend.

Uitzetten van een kumulatieve verdeling op lineaire waarschijnlijkheids-schaal

Tabel 10.

Gegevens van Tabel I.4.

Klasse midden f	F	F%	u	klasse grens	2F'	F' %	u	Klasse midden
2950 1	1	3.3	-1.84	2962.5	1	1.7	-2.12	2950
2975 0	1	3.3	-1.84	2987.5	2	3.3	-1.84	2975
3000 2	3	10.0	-1.28	3012.5	4	6.6	-1.51	3000
3025 4	7	23.1	-0.74	3037.5	10	16.7	-0.95	3025
3050 8	15	50.0	0.00	3062.5	22	36.5	-0.35	3050
3075 11	26	86.5	1.10	3087.5	41	67.0	+0.44	3075
3100 2	28	93.2	1.49	3112.5	54	90.0	1.28	3100
3125 2	30	100.0			58	96.2	1.77	3125
micron					60			micron
	Methode A met klassegrenzen				Methode B met klassemiddens			

Methode B verdient de voorkeur; zij geeft een meer symmetrische behandeling en één punt meer op de grafiek van Fig. 11).

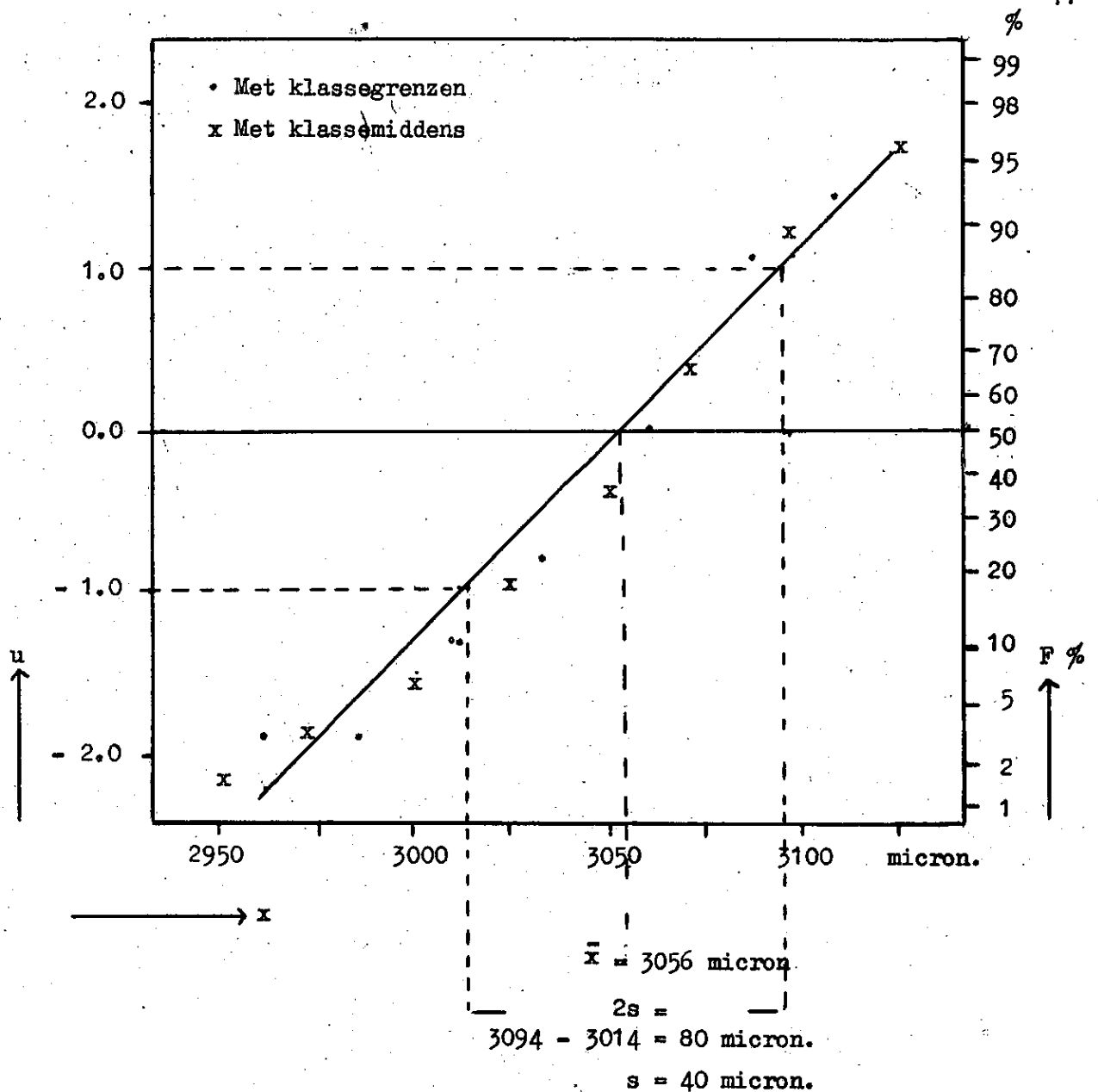


Fig. 11. De gegevens van Tabel 10 uitgezet op lineaire waarschijnlijkheids-schaal.

Bij het trekken van een rechte moeten de punten aan de uiteinden niet te zwaar in rekening worden gebracht.

\bar{x} en s kunnen uit de figuur worden afgelezen als aangegeven.

Grafisch papier met een schaal voor F in % als rechts aangegeven, worden uitgegeven door de drukkerij Mercurius te Wormerveer.

Uitzetten op waarschijnlijkheidsschaal kan ook geschieden met onvolledige gegevens.

Voorbeeld: de lengten van rekruten gegeven in Tabel 1.5.B. Uit een grafiek
Uit een grafiek vond ik

$$\bar{x} = 177.5 \text{ cm}, \quad s = 7.0 \text{ cm.}$$

Deze gegevens zijn uit Tabel 1.5.B niet te berekenen.

Uitzetten van een kleine reeks waarnemingen

$n = 7.$	$x_i =$	1141	1131	1126	1129	1138	1127	1105 mgr.
Gerangschikt naar grootte	$x'_i =$	1105	1126	1127	1129	1131	1138	1141 mgr.
$F = 100 \frac{i-\frac{1}{2}}{n} \%$		7.1	21.4	35.7	50.0	64.3	78.6	92.8
$u =$		-1.47	-0.79	-0.34	0.00	0.34	0.79	1.47.

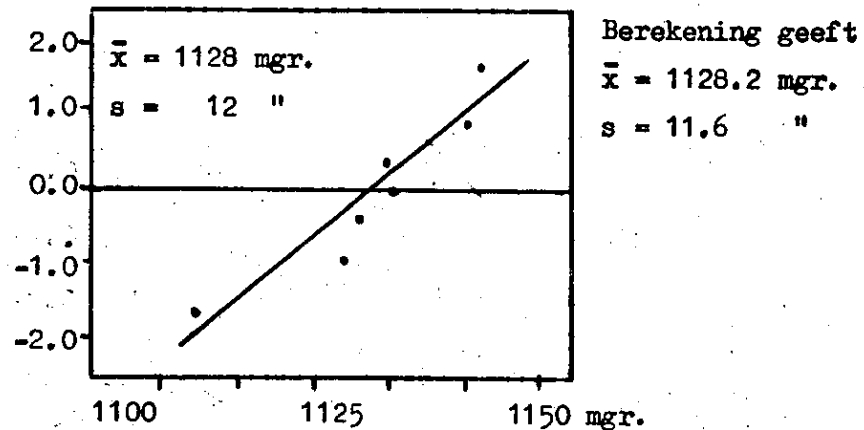


Fig. 12. Een reeks van 7 waarnemingen uitgezet op normale waarschijnlijkheidsschaal.

Voor de berekening van de kumulatieve frequentie F bestaan verschillende formules:

$$F = 100 \frac{i-\frac{1}{2}}{n}, \quad F = 100 \frac{i}{n+1}, \quad F = 100 \frac{i-\frac{3}{4}}{n+\frac{1}{2}} \%$$

De laatste twee berusten op theoretische overwegingen waarop we hier niet in kunnen gaan. Bij grote aantallen waarnemingen is het verschil gering; bij kleine aantallen zijn de statistische fluctuaties zó groot dat het er weinig toe doet welke formule men gebruikt.

Voorbeeld van een logaritmisch normale verdeling

Tabel 12. Aantallen auto's geparkeerd op 60 parkeerterreinen rondom de Philipsfabrieken op een weekday in oktober 1963.

22	126	5	42	2	12	24	21	20	22
109	14	4	39	18	41	12	14	19	77
47	97	23	38	4	24	12	187	44	63
28	44	49	7	4	27	13	16	17	18
77	93	82	25	5	35	32	59	13	12
47	27	8	28	22	70	109	80	137	21

Deze gegevens afgeturft.

Tabel 13.

A. op lineaire schaal

Klassen	
0 - 24	
25 - 49	
50 - 74	
75 - 99	
100 - 124	
125 - 149	
150 - 174	
175 - 199	

B. op geometrische schaal

Klassen	
2	
3 - 4	
5 - 8	
9 - 16	
17 - 32	
33 - 64	
65 - 128	
129 - 256	

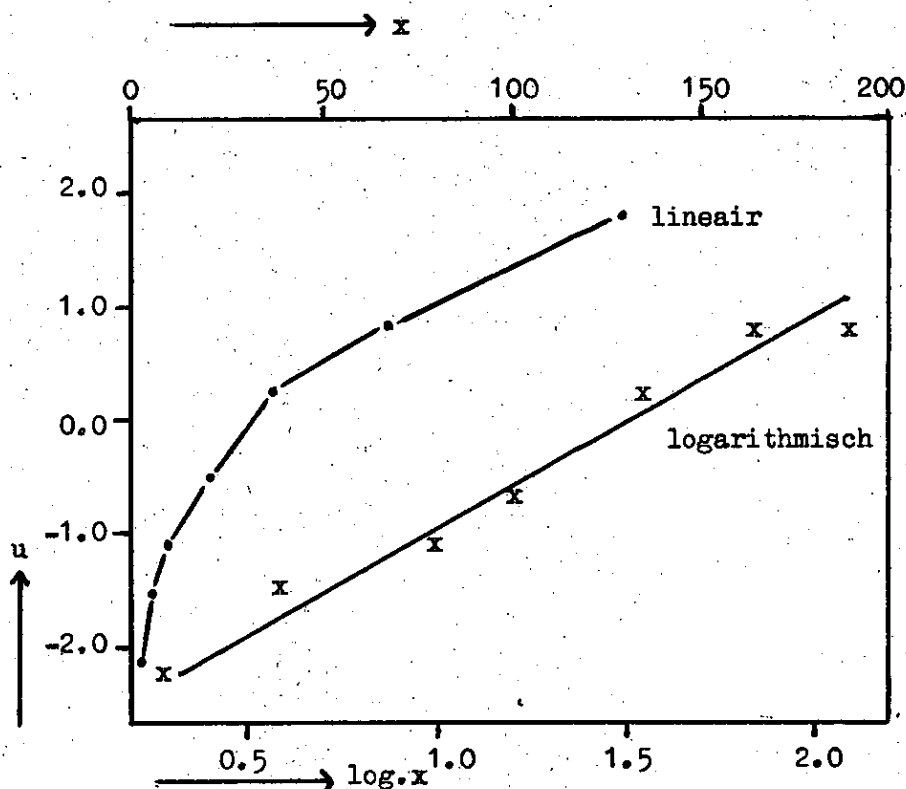


Fig. 13. Gegevens van Tabel 12 uitgezet met lineaire en logaritmische schaal.

De log. normale verdeling komt veel voor bij economische problemen bijv. salarissen, en het electriciteitsverbruik per abonnee. Ook in de techniek komt men haar soms tegen: deeltjesgrootte van gemalen poeders, elektronen in radiobuizen of transistoren.

Er is een apart boek aan deze verdeling gewijd:

J. Aitchison en J.A.C. Brown, The log. normal distribution.
Cambridge University Press 1957.

De normale verdeling: (aanvulling)

De normale verdeling speelt in de techniek een grote rol omdat producten onder homogene omstandigheden geproduceerd in hun eigenschappen veelal een normale verdeling bezitten. Daar geldt:

68 % van de producten ligt tussen $\mu \pm \sigma$,

95 % van de producten ligt tussen $\mu \pm 2\sigma$,

99,7 % van de producten ligt tussen $\mu \pm 3\sigma$,

De waarde van σ is in hoge mate bepalend voor de tolerantie eisen die men kan stellen. Kennis van σ is voor goed kwaliteitsbeleid essentieel.

Bij een normale verdeling kan men σ of S op eenvoudige wijze schatten uit de gemiddelde spreidingsbreedte \bar{R} :

$$\text{benaderd } s = \frac{\bar{R}}{\sqrt{n}}$$

exacter $s = A(n)\bar{R}$, $A(n)$ uit compendium tabel 8.3

Hierbij is n de serie grootte; de formules zijn toepasbaar voor $n \leq 12$. Een grote serie kan men splitsen in een aantal kleinere; mits dit op aselechte wijze geschiedt.

Een loterij-probleem

Een loterij van een Buurt-vereniging leverde 70 prijzen op. De winnende nummers waren.

26 52 63 101 189 285 356 468 476 4621

De onderlinge afstanden zijn tussen deze nummers:

26 26 11 38 88 96 71 112 8

Tabel 14: Frekwentietabel van de onderlinge afstanden tussen de prijswinnende nummers:

Klasse		f	$\frac{2(1-F)}{140}$	$(1-F)\%$	$e^{-\log(1-F)}$
0 - 19		15	125	89.2	4.50
20 - 39		17	93	66.3	4.19
40 - 59		7	69	49.3	3.88
60 - 79		10	52	37.2	3.61
80 - 99		5	37	26.5	3.26
100 - 119		6	26	18.6	2.92
120 - 139		2	18	12.8	2.55
140 - 159		6	10	7.1	1.95
211		1	3	2.1	0.74
477		1	1	0.7	0.37

Stel de kans op een prijs p dan is $p \ll 1$ nl. ongeveer $70/4621 = 0.015$ of 1,5%. De kans op een rij van x nieten gevolgd door een prijs is dan bij benadering.

$$(1-p)^x p \approx e^{-xp} p \quad x = 1, 2, \dots$$

en de afstanden tussen de prijswinnende nummers volgt bij benadering een exponentiële verdeling.

$$P(x < x+dx) = e^{-px} p dx$$

Dan is de cumulatieve verdeling

$$F = P(x \leq x) = 1 - e^{-px} \quad \text{of} \quad 1 - F = e^{-px}$$

$$\log(1-F) = -px$$

In bovenstaande frekwentietabel zijn de waarden van $(1-F)$ in % en behorende bij de klasse-middens uitgerekend en daaruit $e \log(1-F)$. Uitgezet tegen x geeft Fig. 15:

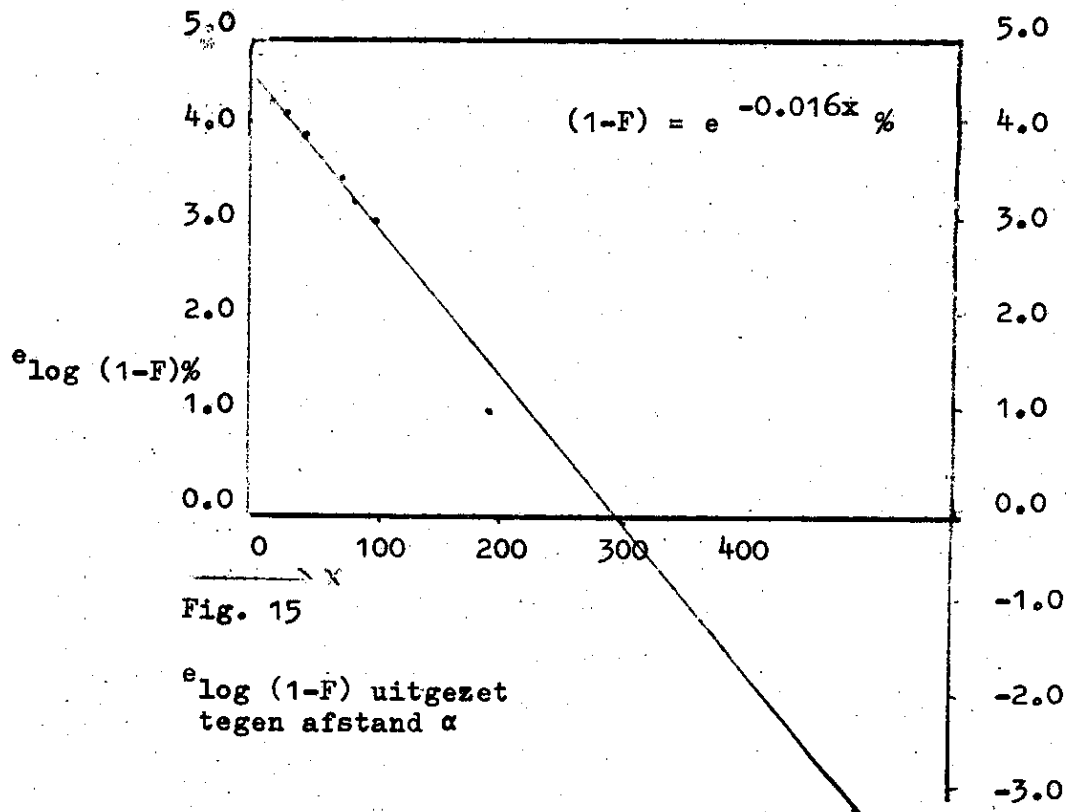


Fig. 15

$e \log(1-F)$ uitgezet tegen afstand x

Aan een exponentiële verdeling wordt goed voldaan. Alleen de hoogste waarde $x = 477$ wijkt sterk af. De kans op een waarneming $x > 477$ is ongeveer $e^{-3} = 1/20\%$. De kans op minstens één dergelijke waarneming in een serie van 70 is dan ongeveer $70 \times 1/20 = 3,5\%$.

Populatie en steekproef.

1. Een eindige Populatie = een verzameling gelijksoortige objecten of elementen.
2. Een eindige statistische populatie = een verzameling metingen die aan de elementen van een populatie kunnen worden uitgevoerd.

3. Een oneindige populatie = een meetprocédé dat in principe onbeperkt kan worden herhaald.
4. Een steekproef = een deel van een populatie waaraan metingen worden verricht.
5. Een steekproef } de waarnemingen aan een steekproef uitge-
Een serie waarnemingen } voerd.
6. Omvang van de populatie : N ,
7. Steekproef grootte } n .
Serie grootte } :
8. De steekproef-populatie = de verzameling van alle mogelijke steekproeven.
9. De steekproef-populatie omvat:
bij trekken zonder terugleggen:
 $\binom{N}{n}$ elementen de volgorde niet meegeteld
of $\frac{N!}{(N-n)!}$ elementen de volgorde wel meegeteld
en bij trekken met teruglegging:
 N^n elementen de volgorde telt steeds mee.
10. Trekken met terugleggen = trekken uit een oneindige populatie. Bij trekken met teruglegging zijn waarnemingen aan verschillende elementen van de steekproef steeds onderling onafhankelijk.

Een aselecte trekking = een trekking zodanig dat ieder element van de steekproefpopulatie dezelfde kans bezit te worden getrokken.

Uit een genummerde populatie wordt aselect getrokken met behulp van aselecte getallen;

Compendium Tabel 8.4

Voorbeeld: trekking van een steekproef uit de 365 dagen van een jaar.

Tabel 8,4, pag. 3. Kolom 13, rij 16 van boven naar beneden! geeft de getallen rij:

311	137	343	101	500	828	931	329	123	443	853
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

De steekproef bestaat uit de dagen met rangnummers.

311	137	343	101	135	x	x	329	123	78	x
-----	-----	-----	-----	-----	---	---	-----	-----	----	---

Het getal 000 en getallen > 730 worden weggelaten; getallen < 730 maar > 365 worden met 365 verminderd.

Dit is trekking met teruglegging; trekking zonder teruglegging krijgt men door nummers die meerdere malen voorkomen slechts één maal mee te rekenen. Bij populaties met minder dan 100 elementen kan trekken zonder terugleggen ook worden uitgevoerd met behulp van de aselecte permutaties van Tabel 8.5.

Met behulp van aselecte getallen rijen kan men uit iedere willekeurige verdeling trekkingen verrichten. Het principe ligt Fig. 17 toe.

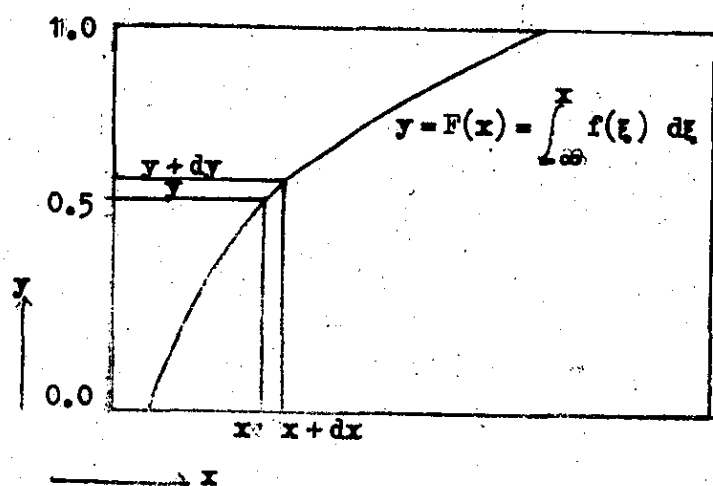


Fig. 17 Trekkingen uit een verdeling met kansdichtheid $f(x) dx$ met behulp van aselechte getallen.

We vatten groepen van bijv. 4 aselechte getallen op als decimale breuken. Tabel 8,4 pag. 2 kolom 11 en 12 van onder naar boven gelezen geven de breuken:

0.8406 0.0143 0.4165 0.5599 0.0556 0.4499 enz.

Beschouwen we dit als waarnemingen van een statistische variabele y dan bezit y praktische een standaard rechthoekige verdeling d.w.z.

$$P(y < \underline{y} < y + d y) = d y, \quad 0 \leq y \leq 1$$

Wij zetten nu de kumulative verdeling $F(x)$ als in Fig. 17 uit en zoeken met behulp van deze kromme bij een \underline{y} de bijbehorende waarde van \underline{x} . Dan geldt:

$$P(x < \underline{x} < x + d x) = P(y < \underline{y} < y + d y) = d y$$

$$F(x) dx = f(x) dx$$

Bij een discrete verdeling is $F(x)$ een discontinue trapfunctie doch is de methode even goed toepasbaar.

Voorbeelden: naspelen, van een dobbelspel, trekkingen uit een Poisson verdeling, uit een normale verdeling (Tabel 8.6) en uit een exponentiële verdeling (Tabel 8.7)

Praktijk moeilijkheden bij aselechte trekken.

De relatie tussen populatie en steekproef.

toegelicht aan een enquête betreffende:

geboorte datum, lengte, gewicht, gezinssamenstelling, rookgewoonten en tijdwaarnemingen.

Trekken uit een kaartsysteem met ongelijke aantallen elementen per kaart (gezinnen).

Gegeven: \underline{x}_k , $k = 1, 2, \dots$, $E(\underline{x}_k) = \mu_k$, $\text{var } \underline{x}_k = \sigma_k^2$

$$\underline{y} = \sum_i a_i \underline{x}_i, \quad \underline{z} = \sum_j b_j \underline{x}_j,$$

dan is

$$E(\underline{y}) = \sum a_i E(\underline{x}_i) = \sum a_i \mu_i, \quad (18,1)$$

$$\text{var } \underline{y} = \sum a_i^2 \text{var } \underline{x}_i + \sum_{i \neq i'} \sum a_i a_{i'} \text{cov}(\underline{x}_i, \underline{x}_{i'}), \quad (18,2)$$

$$= \sum a_i^2 \text{var } \underline{x}_i + 2 \sum_{i < i'} \sum a_i a_{i'} \text{cov}(\underline{x}_i, \underline{x}_{i'}), \quad (18,3)$$

$$\text{cov}(\underline{y}, \underline{z}) = \sum a_i b_j \text{cov}(\underline{x}_i, \underline{x}_j). \quad (18,4)$$

$E(\underline{z})$ en $\text{var } \underline{z}$ volgen uit (18,1) en (18,2).

In (18,4) is voor $\underline{x}_i = \underline{x}_j$, $\text{cov}(\underline{x}_i, \underline{x}_j) = \text{var } \underline{x}_i$.

Zijn \underline{x}_i , $\underline{x}_{i'}$ onderling onafhankelijk voor iedere $i \neq i'$ dan geldt

$$\text{var } \underline{y} = \sum a_i^2 \text{var } \underline{x}_i = \sum a_i^2 \sigma_i^2. \quad (18,5)$$

Deze formules vormen de basis van vele statistische analysemethoden.

Voorbeelden

A. Standen van de secondewijzer

$$\underline{t}_0, \underline{t}_1, \underline{t}_2, \underline{t}_3, \underline{t}_4, \underline{t}_5.$$

\underline{t}_0 , \underline{t}_1 en \underline{t}_2 kort na elkaar, \underline{t}_3 en \underline{t}_4 met ± 15 min. tussenpauze, \underline{t}_5 ongeveer 90 min. na \underline{t}_1 .

Model:

$$\begin{aligned} \underline{t}_0 &= \underline{A} + \underline{e}_0, \\ \underline{t}_1 &= \underline{A} + T_{10} + \underline{e}_1, \\ \underline{t}_2 &= \underline{A} + T_{20} + \underline{e}_2. \end{aligned} \quad (18,6)$$

\underline{A} = de "ware" aanvangsstand

T_{10} , T_{20} = de tijden verlopen tussen de waarnemingen \underline{t}_0 , \underline{t}_1 , en \underline{t}_0 , \underline{t}_2 .

\underline{e}_0 , \underline{e}_1 , \underline{e}_2 = toevallige waarnemingsfouten waarvoor we onderstellen:

$$E(\underline{e}_1) = 0, \text{ var } \underline{e}_1 = \sigma_0^2, \text{ cov } (\underline{e}_1, \underline{e}_j) = 0. \quad (19, 1)$$

Dan is

$$\underline{d}_{10} = \underline{t}_1 - \underline{t}_0 = T_{10} + \underline{e}_1 - \underline{e}_0 \quad (19, 2)$$

$$\underline{d}_{21} = \underline{t}_2 - \underline{t}_1 = T_{20} - T_{10} + \underline{e}_2 - \underline{e}_1$$

en

$$\text{var } \underline{d}_{10} = \text{var } \underline{e}_1 + \text{var } \underline{e}_0 = 2\sigma_0^2 \quad (19, 3)$$

$$\text{var } \underline{d}_{21} = \text{var } \underline{e}_2 + \text{var } \underline{e}_1 = 2\sigma_0^2 ; \text{ cov } (\underline{d}_{10}, \underline{d}_{21}) =$$

$$\text{cov } (\underline{e}_1, \underline{e}_2) - \text{var } \underline{e}_1 - \text{cov } (\underline{e}_0, \underline{e}_2) + \text{cov } (\underline{e}_0, \underline{e}_1) = -\sigma_0^2.$$

$$r = \frac{-\sigma_0^2}{\sqrt{2\sigma_0^2 * 2\sigma_0^2}} = -\frac{1}{2} \quad (19, 4)$$

B. Bij grotere intervallen tussen de aflezingen moeten we rekening houden met verschil in loopsnelheid der horloges. Voor \underline{t}_3 , \underline{t}_4 en \underline{t}_5 wordt het model

$$\begin{aligned} \underline{t}_3 &= \underline{A} + \underline{b} T_{30} + \underline{e}_3 \\ \underline{t}_4 &= \underline{A} + \underline{b} T_{40} + \underline{e}_4 \\ \underline{t}_5 &= \underline{A} + \underline{b} T_{50} + \underline{e}_5 \end{aligned} \quad (19, 5)$$

waarin \underline{b} nu een factor is die ongeveer 1.00 zal bedragen, maar tussen de horloges aan toevallige variaties onderhevig is. We vinden nu

$$\underline{d}_{30} = (\underline{t}_3 - \underline{t}_0) = \underline{b} T_{30} + \underline{e}_3 - \underline{e}_0$$

en

$$\text{var } \underline{d}_{30} = T_{30}^2 \text{var } \underline{b} + 2\sigma_0^2 = T_{30}^2 \sigma_b^2 + 2\sigma_0^2, \quad (19, 6)$$

en evenzo voor \underline{t}_4 en \underline{t}_5 .

Verder bijv.

$$\text{cov } (\underline{d}_{30}, \underline{d}_{40}) = T_{30} T_{40} \sigma_b^2 - \sigma_0^2. \quad (19, 7)$$

Enzovoort.

C. Afrondingsfouten met een afrondingsinterval a hebben een variantie:

$$\text{var } \underline{x}_a = \int_{-\frac{1}{2}a}^{+\frac{1}{2}a} \frac{1}{a} \xi^2 d\xi = \frac{a^2}{12} \quad (20, 1)$$

Een variantie σ^2 vóór afronden wordt

$$\sigma'^2 = \sigma^2 + \frac{a^2}{12}$$

na afronden.

Is $a < \frac{1}{2}\sigma$ dan is $\sigma'^2 < 1.02 \sigma^2$ en

$$\sigma' < 1.01 \sigma$$

Hierop is de keuze van een afrondingsinterval gebaseerd; vergelijk pag. 2 no 16.

D. Met een eindige statistische populatie

$$\begin{aligned} X_k, k = 1, \dots, N, \\ \mu = \Sigma X_k / N, \\ \sigma^{*2} = \Sigma (X_k - \mu)^2 / (N-1), \end{aligned} \quad (20, 2)$$

korrespondeert een oneindige statistische populatie gedefinieerd door

$$\begin{aligned} P(\underline{x} = X_k) &= 1/N, \\ E(\underline{x}) &= \mu = \Sigma X_k / N, \\ \sigma^2 &= \Sigma (X_k - \mu)^2 / N. \end{aligned} \quad (20, 3)$$

Zij

$$\underline{x}_i, \quad i = 1, \dots, n$$

een steekproef getrokken zonder teruglegging en waarin de volgorde der elementen wordt meegerekend dan bestaat de steekproefpopulatie uit $N! / (N-n)!$ elementen.

Vatten we de steekproef op als een trekking met teruglegging van één element uit deze steekproefpopulatie dan geldt :

$$\begin{aligned}
 P(\underline{x}_i = X_k) &= 1/N, \\
 P(\underline{x}_i = X_k, \underline{x}_j = X_l) &= \frac{1}{N(N-1)} \\
 \text{var } \underline{x}_i &= \sigma^2, \\
 \text{cov}(\underline{x}_i, \underline{x}_j) &= \frac{-\sigma^2}{N-1}.
 \end{aligned}
 \tag{21, 1}$$

Hieruit volgt

$$\begin{aligned}
 E(\underline{x}_i) &= n\mu, \\
 E(\bar{\underline{x}}_i) &= \mu, \\
 \text{var } \underline{x}_i &= n\sigma^{*2} \left(\frac{1}{n} - \frac{1}{N} \right), \\
 \text{var } \bar{\underline{x}}_i &= \frac{\sigma^{*2}}{n} \left(\frac{1}{n} - \frac{1}{N} \right), \\
 E \underline{s}^2 &= \sigma^{*2}.
 \end{aligned}
 \tag{21, 2}$$

Deze formules gelden bij trekken zonder teruglegging. Formules voor trekken met teruglegging volgen hieruit door $N \rightarrow \infty$ en $\sigma^{*2} = \sigma^2$ te stellen; dus dan geldt:

$$\begin{aligned}
 \text{var } \underline{x}_i &= n\sigma^2, \\
 \text{var } \bar{\underline{x}}_i &= \frac{\sigma^2}{n}, \\
 E \underline{s}^2 &= \sigma^2.
 \end{aligned}
 \tag{21, 3}$$

De centrale limietstelling

Deze zegt: Een steekproefsom of een steekproef gemiddelde heeft in goede benadering een normale verdeling, ook al is de verdeling binnen de populatie niet normaal.

Voorbeelden:

- A. De som van de worpen met drie of meer dobbelstenen,
- B. Steekproeven van 10 elementen uit de zeer scheve log. normale verdeling van Tabel 12.
- C. Een binomiale verdeling wanneer $np > 5$, $n(1-p) > 5$.

De centrale limietstelling geldt strikt voor de limiet $n \rightarrow \infty$, doch de convergentie is zeer snel, zoals de voorbeelden tonen.

Theoretische toelichting

Door de transformatie $y = (\underline{x} - \mu) / \sigma$ kan iedere variabele \underline{x} worden omgezet in een variabele y waarvoor $E(y) = 0$, $\text{var } y = E(y^2) = 1$.
Dan definiëren we

$$\beta_1 = E(y^3) = E\left(\frac{\underline{x} - \mu}{\sigma}\right)^3 = \frac{\mu_3}{\sigma^3}, \quad (22.1)$$

$$\beta_2 = E(y^4) = E\left(\frac{\underline{x} - \mu}{\sigma}\right)^4 = \frac{\mu_4}{\sigma^4}. \quad (22.2)$$

en

$$\gamma_1 = \beta_1^2, \quad \gamma_2 = \beta_2 - 3. \quad (22.3)$$

β_1 en γ_1 zijn een maat voor de scheefheid van een verdeling; voor een symmetrische verdeling is $\beta_1 = 0$, $\gamma_1 = 0$.

β_2 en γ_2 zijn maten voor de z.g. scherpstoppigheid of platvoetigheid. Fig. 23 geeft enige voorbeelden, die de betekenis toelichten.

Waarden van deze coëfficiënten kan men voor gangbare verdelingen afleiden uit de tabel in het Compendium pag. 11.

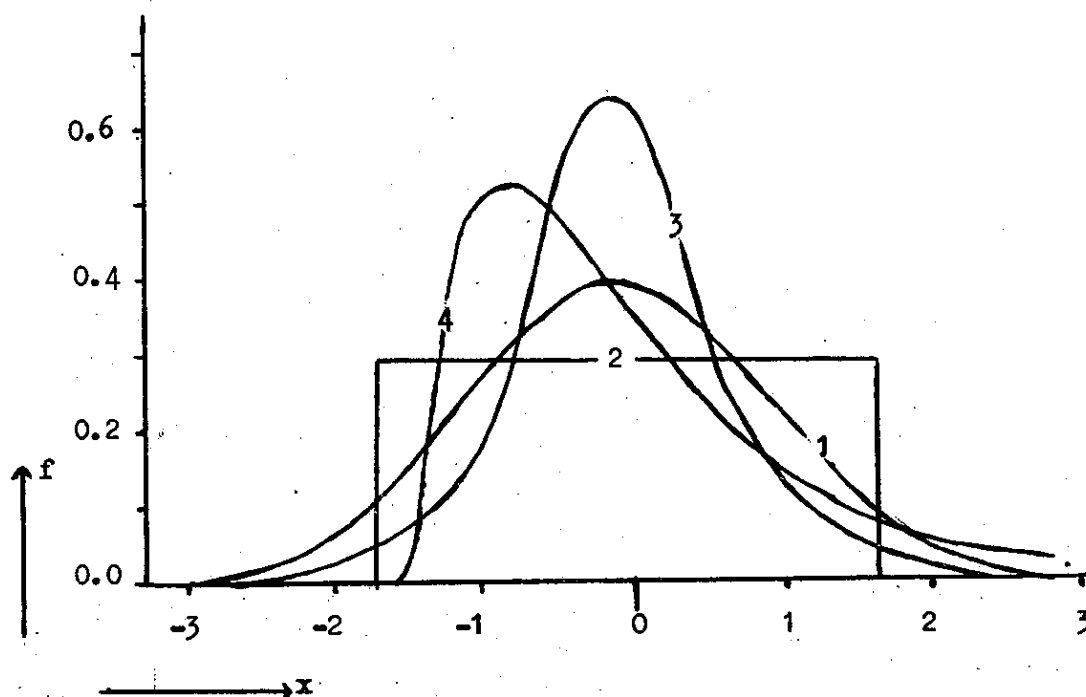


Fig. 23. 4 verdelingen met $\mu = 0$, $\sigma = 1$,

1. De normale verdeling: $\beta_1 = 0$, $\beta_2 = 3$,
2. De rechthoekige verdeling: $\beta_1 = 0$, $\beta_2 = 1.8$,
3. Een scherptoppige verd.: $\beta_1 = 0$, $\beta_2 = \infty$,
4. Een scheve verdeling: $\beta_1 = 4$, $\beta_2 = 12$.

Algemeen geldt nu bij onafhankelijke waarnemingen

$$\beta_1(\underline{x}_n) = \beta_1(\bar{x}_n) = \beta_1(\underline{x})/\sqrt{n}, \quad (23.1)$$

$$\beta_2(\underline{x}_n) = \beta_2(\bar{x}_n) = \frac{\beta_2(\underline{x})}{n} + 3 \frac{n-1}{n}. \quad (23.2)$$

Ook hieruit zien we dat voor $n \rightarrow \infty$, $\beta_1(\underline{x}_n) \rightarrow 0$ en $\beta_2(\underline{x}_n) \rightarrow 3$.

Parameters, schatters, en schattingen

μ en σ^2 of σ zijn twee belangrijke parameters van een populatie. In de praktijk is men meestal niet in de volledige populatie geïnteresseerd, maar alleen in deze parameters. Doel van een steekproef is dan deze parameters numeriek te bepalen. Uit een steekproef kan men echter de exacte waarde niet vinden; men kan de parameters alleen schatten. De formules

$$\bar{x} = \frac{\sum x_i}{n}, \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)} \quad (23.3)$$

noemt men daarom schatters voor de parameters μ en σ^2 ; de waarden

$$\bar{x} = \frac{\sum x_i}{n}, \text{ en } s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)} \quad (24.1)$$

uit een gegeven steekproef becijferd heten schattingen van die parameters. Schatten zijn stochastische variabelen met een bepaalde steekproefpopulatie; schattingen zijn numerieke getallen.

Daar $E(\bar{x}) = \mu$, en $E(s^2) = \sigma^2$ (24.2)

noemt men \bar{x} en s^2 zuivere schattingen van μ en σ^2 . Daar $E(s) < \sigma$ is s een onzuivere schatting van σ .

Naast μ en σ kent men nog andere parameters, bijv. β_1 en β_2 of de mediaan maar die spelen een ondergeschikte rol en we zullen er geen aandacht aan besteden.

De begrippen schatter, schatting, en zuiverheid zijn ook bij deze van toepassing.

Betrouwbaarheidsintervallen voor μ

Op grond van de centrale limietstelling kan men in goede benadering aannemen dat

$$\bar{x} \approx \mu + u \sigma / \sqrt{n^*}, \quad n^* = \frac{nN}{N-n}. \quad (24.3)$$

in woorden: het steekproefgemiddelde, ofwel de schatter \bar{x} heeft een normale verdeling met verwachting = μ en standaardafwijking $\sigma / \sqrt{n^*}$.

Dan is

$$P\left\{ \mu - u_\alpha \sigma / \sqrt{n^*} < \bar{x} < \mu + u_\alpha \sigma / \sqrt{n^*} \right\} = 1 - \alpha, \quad (24.4)$$

of ook

$$P\left\{ \bar{x} - u_\alpha \sigma / \sqrt{n^*} < \mu < \bar{x} + u_\alpha \sigma / \sqrt{n^*} \right\} = 1 - \alpha. \quad (24.4.A)$$

De kans $1 - \alpha$ noemt men de betrouwbaarheid, de kans α de onbetrouwbaarheid, en de grenzen $\bar{x} \pm u_\alpha \sigma / \sqrt{n^*}$ betrouwbaarheidsgrenzen voor de onbekende parameter μ .

Bij een gegeven α kan men de bijbehorende waarde van u_α aflezen uit Compendium Tabel 1.1, of uit de tabel op de omslag afgedrukt.

Het door (24.4.A) gegeven interval heet een tweezijdig betrouwbaarheidsinterval met betrouwbaarheid $(1-\alpha)$. Daarnaast onderscheid men éénzijdige intervallen gegeven door

$$P\left\{ \bar{x} - u_\alpha \sigma / \sqrt{n^*} < \mu \right\} = 1 - \alpha, \quad (24.5)$$

en

$$P\left\{ \mu < \bar{x} + u_\alpha \sigma / \sqrt{n^*} \right\} = 1 - \alpha.$$

Bij gegeven α behoort dan echter een andere waarde van u_{α_i} bijv.

voor $\alpha = 5\%$ tweezijdig is $u_{\alpha} = 1.96$,
en voor $\alpha = 5\%$ eenzijdig is $u_{\alpha} = 1.64$.

De waarde $u_{\alpha} = 1.96$ behoort bij $\alpha = 2.5\%$ éénzijdig, enz.

Doorgaans kent men σ niet, doch neemt men aan dat bij benadering

$$\sigma = s \quad (25.1)$$

mag worden gesteld. Wanneer s bepaald is uit ten minste 30 vrijheidsgraden is dit in ieder geval toelaatbaar.

Toepassingen

Tabel 27

Resultaat gewichten en wegingen bij de enquête.

A Opgegeven gewicht zonder kleren				B Opgegeven gewicht met kleren			
	Eigen opgave	Controle	Vershil		Eigen opgave	Controle	Vershil
	69	74	-5		76	77	-1
	82	84	-2		79	80	-1
	57	59	-2		85	87	-2
	72	72	0		59	60	-1
	75	76	-1		70	71	-1
	65	70	-5		79	79	0
	70	73	-3		64	65	-1
	78	80	-2		75	76	-1
	69	71	-2		77	76	+1
	70	80	-10		63	62	+1
	80	81	-1		70	71	-1
	80	83	-3		85	86	-1
Σx	867	903	-36		882	890	-8
Σx^2	63213	68493	186		65608	66838	14
KS	572	542	78		781	830	8.67
s^2	52.0	49.3	7.1		71.0	75.5	0.79
n	12	12	12		12	12	12
\bar{x}	72.2	75.2	-3.0		73.5	74.2	-0.7
s	7.4	7.0	2.7		8.4	8.7	0.9
	x_1	x_2	d_{12}		x_3	x_4	d_{34}

kg
"

Resultaat Lengten:

	n	\bar{x}	s	KS	s^2
zonder schoenen	18	180.6	5.9	582.3	34.2
met schoenen	4	184.2	4.9	72.7	24.2
		cm	cm	cm ²	cm ²

Betrouwbaarheidsintervallen

Parameter	Interval	Tabel Comp.	Voorwaarden
Algemeen: Centrale Limietstelling			
1 μ	$\bar{x} \pm u_{\alpha} s / \sqrt{n}$	Omslag	$\sigma \approx s$
2 $\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2 \pm u_{\alpha} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	"	$\sigma_1 = \sigma_2 \approx s$
3 "	$\bar{x}_1 - \bar{x}_2 \pm u_{\alpha} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	"	$s_1 \approx \sigma_1 \neq \sigma_2 \approx s_2$
Normale populaties			
4 μ	$\bar{x} \pm t_{v, \alpha} s / \sqrt{n}$	2.1	
5 $\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2 \pm t_{v, \alpha} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	"	$\sigma_1 = \sigma_2$
6 σ	$a_1 s, a_2 s$	3.2	
7 $\frac{\sigma_1^2}{\sigma_2^2}$	$\frac{1}{F_{v_1}^2} \frac{s_1^2}{s_2^2}, F_{v_1}^2 \frac{s_1^2}{s_2^2}$ α verdubbelen!	4	
Binomiale populaties, $\hat{p} = x/n$, $\hat{q} = (1-\hat{p})$.			
8 p	$\hat{p} \pm u_{\alpha} \sqrt{\frac{\hat{p}\hat{q}}{n}}$	Omslag	$n\hat{p} > 50, n\hat{q} > 50$
9 "	$\hat{p} + \frac{u_{\alpha}^2}{n} (\frac{1}{2} - \hat{p}) \pm u_{\alpha} \sqrt{\frac{\hat{p}\hat{q}}{n}}$	"	$n\hat{p} > 5, n\hat{q} > 5$
10 "	$\frac{a_1}{n}, \frac{a_2}{n}$	6.3	$\hat{p} < 0.10$
11 "	a_1, a_2	5.3	
12 $p_1 - p_2$	$\hat{p}_1 - \hat{p}_2 \pm u_{\alpha} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$	Omslag	$n_1 \hat{p}_1 > 5, n_2 \hat{p}_2 > 5$ $n_1 \hat{q}_1 > 5, n_2 \hat{q}_2 > 5$

Parameter	Interval	Tabel Comp.	Voorwaarden	
Poisson verdeling; n waarnemingen, $x. = \sum x_i$, $\bar{x}. = x./n$.				
13	μ	$\bar{x}. \pm u_{\alpha} \sqrt{\frac{x.}{n}}$	Omslag	$x. > 50$
14	"	$\bar{x}. \pm \frac{1}{2}u_{\alpha}^2 \pm u_{\alpha} \sqrt{\frac{x.}{n}}$	"	$x. > 5$
15	"	$\frac{a_1}{n}, \frac{a_2}{n}, a_1$ en a_2 opzoeken bij $x.$	6.3	
16	$\mu_1 - \mu_2$	$\bar{x}_{1.} - \bar{x}_{2.} \pm u_{\alpha} \sqrt{\frac{x_{1.}}{n_1} + \frac{x_{2.}}{n_2}}$	Omslag	
Poisson proces; geteld x_1, \dots, x_n , in T_1, \dots, T_n , $x. = \sum x_i, T. = \sum T_i$				
17	λ	$\frac{x.}{T.} \pm u_{\alpha} \sqrt{\frac{x.}{T.}}$	Omslag	$x. > 50$
18	"	$\frac{x.}{T.} \pm \frac{1}{2}u_{\alpha}^2 \pm u_{\alpha} \sqrt{\frac{x.}{T.}}$	"	$x. > 5$
19	$\lambda_1 - \lambda_2$	$\frac{x_{1.}}{T_{1.}} - \frac{x_{2.}}{T_{2.}} \pm u_{\alpha} \sqrt{\frac{x_{1.}}{T_{1.}^2} + \frac{x_{2.}}{T_{2.}^2}}$	"	

Noot: Hebben we in de formules 13-19 slechts één waarneming, x , dan is

$$n = 1, \quad x. = \bar{x}. = x$$

De formules blijven toepasbaar.

Toelichting

1,2,3. Aan de voorwaarde $s \approx \sigma$ is redelijk voldaan wanneer $v > 30$. Men gebruikt de formules echter ook vaak als eerste benadering bij kleinere aantallen waarnemingen.

Heeft men een aantal reeksen waarnemingen met eenzelfde σ dan kunnen de schattingen s_i worden samengevoegd volgens:

$$s^2 = \frac{\sum v_i s_i^2}{\sum v_i} = \frac{\sum K s_i^2}{\sum v_i}, \quad v = \sum v_i.$$

Men gebruikt dan verder alleen deze eindwaarde en vergeet de afzonderlijke s_i 's.

Bij trekking zonder teruglegging moet n door n^* worden vervangen:

$$n^* = \frac{nN}{N-n}.$$

4 - 7. Altijd geldt $t_{v,\alpha} > u_\alpha$, doch $\lim_{v \rightarrow \infty} t_{v,\alpha} = u_\alpha$.

v is het aantal vrijheidsgraden van s , n het aantal waarnemingen in \bar{x} . Bij een normale verdeling zijn \bar{x} en s onafhankelijk. Het doet er dus niet toe of \bar{x} en s uit dezelfde of uit verschillende waarnemingsreeksen zijn geschat.

De α -waarden bij Compendium Tabel 4 zijn éénzijdig. Voor een tweezijdig interval moeten zij worden verdubbeld.

In $F_{v_a}^{v_b}$ geeft v_a het aantal vrijheidsgraden in de teller, v_b dat in de noemer. Bij het gebruik van Tabel 4 is dit belangrijk.

8-12. Heeft men een aantal steekproeven uit één binomiale verdeling dan kunnen de formules 8 en 9 worden toegepast op $x = \sum x_i$ en $n = \sum n_i$.

9 is een vereenvoudigde vorm van de formule

$$\frac{\hat{p} + u_\alpha^2/2n \pm u_\alpha \sqrt{\hat{p}\hat{q}/n + u_\alpha^2/4n}}{1 + u_\alpha^2/n}$$

gegeven op pag. 17 van het Compendium. Deze laatste formule is voor de praktijk te gecompliceerd.

13-16. Deze formules onderstellen n waarnemingen x_i uit één Poisson verdeling. Hebben we slechts één waarneming dan is $\bar{x} = x$ en blijven de formules van toepassing.

14 is een vereenvoudiging van

$$\bar{x} + u_\alpha^2/2n \pm u_\alpha \sqrt{\bar{x}/n + u_\alpha^2/4n}.$$

De Compendium Tabel pag. 17 geeft deze laatste formule voor $n = 1$.

17-19. Deze formules zijn o.m. van toepassing bij Geiger tellingen! Men lette erop dat zij gebaseerd zijn op de sommen x, T .

Voorbeelden van betrouwbaarheidsintervallen aan de hand van Tabel 27

A. $s \approx \sigma$. Centrale limietstelling

Voor gemiddelde lengte zonder schoenen.

$$180.6 \pm 1.96 * 5.9/\sqrt{18} = 177.8 \leftrightarrow 183.4\text{cm}; \alpha = 5\% \text{ (tweezijdig)}, \quad (32.1)$$

$$180.6 \pm 2.58 * 5.9/\sqrt{18} = 177.0 \leftrightarrow 184.2\text{cm}; \alpha = 1\% \text{ (tweezijdig)}, \quad (32.2)$$

$$180.6 - 1.64 * 5.9/\sqrt{18} = 178.3\text{cm} < \mu; \quad \alpha = 5\% \text{ (eenzijdig)}. \quad (32.3)$$

B. Normale verdelingen; $\sigma_1 = \sigma_2$.

Gemiddelde lengte met schoenen.

$$184.2 \pm 3.18 * 4.9/\sqrt{4} = 176.3 \leftrightarrow 192.1\text{cm}; \alpha = 5\% \text{ (tweezijdig)}. \quad (32.4)$$

Beter de s^2 -en voor zonder en met schoenen samenvoegen:

$$s^2 = \frac{v_1 s_1^2 + v_2 s_2^2}{v_1 + v_2} = \frac{17 * 5.9^2 + 3 * 4.9^2}{17 + 3} = 32.7 \text{ cm}^2,$$

$$s = 5.7 \text{ cm}; v = 20. \quad (32.5)$$

Hiermee voor de gemiddelde lengte zonder schoenen,

$$180.6 \pm 2.09 * 5.7/\sqrt{18} = 177.8 \leftrightarrow 183.4\text{cm}; \alpha = 5\%, \quad (32.6)$$

voor de gemiddelde lengte met schoenen

$$184.2 \pm 2.09 * 5.7/\sqrt{4} = 178.2 \leftrightarrow 190.2\text{cm}; \alpha = 5\%, \quad (32.7)$$

en voor het verschil met schoenen - zonder schoenen

$$184.2 - 180.6 \pm 2.09 * 5.7 \sqrt{\frac{1}{4} + \frac{1}{18}} = -3.0 \leftrightarrow + 10.2\text{cm}; \alpha = 5\%. \quad (32.8)$$

Voor de gewichten vinden we evenzo

$$s^2 = \frac{11 * 7.4^2 + 11 * 8.4^2}{22} = \frac{572 + 781}{22} = 61.5 \text{ kg}^2,$$

$$s = 7.9 \text{ kg}, v = 22. \quad (32.9)$$

en hiermee voor het verschil in gewicht met en zonder kleren

$$73.5 - 72.2 \pm 2.08 * 7.9 * \sqrt{\frac{1}{12} + \frac{1}{12}} = -5.2 \leftrightarrow + 7.8\text{kg}; \alpha = 5\%. \quad (32.10)$$

Een nauwkeuriger bepaling van dit verschil verkrijgen we uit $d_{12} = \bar{x}_1 - \bar{x}_2$ in Tabel 27, nl.

$$3.0 \pm 2.20 * 2.7/\sqrt{12} = +1.3 \leftrightarrow 4.7 \text{ kg}; \alpha = 5\% . \quad (33.1)$$

Op \bar{x}_1 en \bar{x}_2 uit Tabel 27 mag men een betrouwbaarheidsinterval in de vorm (32.8) of (32.10) niet toepassen, omdat de waarnemingsreeksen van x_1 en x_2 niet onafhankelijk zijn!

In de reeks voor d_{12} komt één uitschieter voor, -10kg. Laten we deze weg dan vinden we i.p.v. (33.1) een verschil

$$2.4 \pm 2.23 * 1.6/\sqrt{11} = +1.3 \leftrightarrow 3.5 \text{ kg}; \alpha = 5\% . \quad (33.2)$$

Mogelijk is dit beter, maar of we die uitschieter zo maar mogen weglaten is niet met zekerheid te zeggen. Eventueel moet men beide intervallen (33.1) en (33.2) vermelden.

Leiden beide grenzen van een betrouwbaarheidsinterval tot eenzelfde conclusie dan weten we genoeg; leiden zij tot verschillende conclusies dan zijn er meer waarnemingen gewenst.

C. Normale verdeling. Intervallen voor σ , en σ^2/σ_2^2 . Tabellen 3.2 en 4.1-4.4.

Lengten zonder schoenen: $s_1 = 5.9\text{cm}$, $v_1 = 17$.

$$\left. \begin{array}{l} \sigma_1 \text{ ligt tussen } 0.75 * 5.9 = 4.4 \text{ cm} \\ \text{en } 1.50 * 5.9 = 8.8 \text{ cm} \end{array} \right\} \alpha = 5\% . \quad (33.3)$$

Lengten met schoenen: $s_2 = 4.9 \text{ cm}$, $v_2 = 3$.

$$\left. \begin{array}{l} \sigma_2 \text{ ligt tussen } 0.57 * 4.9 = 2.8 \text{ cm} \\ \text{en } 3.37 * 4.9 = 16.5 \text{ cm} \end{array} \right\} \alpha = 5\% . \quad (33.4)$$

Onder aan Tabel 3.2 toevoegen:

Benadering voor $v > 30$: $\underline{s} = \sigma + \underline{u}\sigma/\sqrt{2v}$,

waaruit volgt

$$a_1 = \frac{\sqrt{2v}}{\sqrt{2v} + u_\alpha} , \quad a_2 = \frac{\sqrt{2v}}{\sqrt{2v} - u_\alpha} . \quad (33.5)$$

Wil men σ schatten met een nauwkeurigheid $\pm \delta$ en onbetrouwbaarheid α dan moet

$$\frac{\sigma + \delta}{\sigma - \delta} = \frac{a_2}{a_1} = \frac{\sqrt{2v + u_\alpha}}{\sqrt{2v - u_\alpha}},$$

waaruit volgt

$$v = \frac{1}{2} \left(\frac{u_\alpha \sigma}{\delta} \right)^2. \quad (34.1)$$

Verder

$$\frac{1}{F_{17}^3} \frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < F_{17}^3 \frac{s_1^2}{s_2^2},$$

of

$$\frac{1}{14.3} \frac{5.9^2}{4.9^2} < \frac{\sigma_1^2}{\sigma_2^2} < 4.01 \frac{5.9^2}{4.9^2}, \quad \text{Tabel 4.2,}$$

of

$$0.10 < \frac{\sigma_1^2}{\sigma_2^2} < 5.81,$$

$$0.32 < \frac{\sigma_1}{\sigma_2} < 2.41; \quad \alpha = 2 \times 0.025 = 5\% !! \quad (34.2)$$

De onbetrouwbaarheden α bij Tabel 4.1 4.4 zijn éézijdig. Dit bij de Tabellen vermelden.

Twee series van ieder n waarnemingen

						\bar{x}_i	
x_{1j}	=	60.0	59.4	58.1	59.6	58.0	59.02
x_{2j}	=	60.8	60.1	58.4	59.6	59.2	59.62
$\bar{x}_{.j}$	=	60.40	59.75	58.25	59.60	58.60	
d_j	=	0.8	0.7	0.3	0.0	1.2	

De interpretatie kan geschieden volgens verschillende modellen.

Model I:
$$\begin{aligned} x_{ij} &= \mu + e_{ij}, \\ E(e) &= 0, \quad \text{var } e = \sigma^2, \quad \text{cov}(e, e') = 0. \end{aligned} \quad (35.1)$$

De beide reeksen vormen samen één reeks onafhankelijke waarnemingen uit één populatie.

Bewerking: $\bar{x}_{..} = 59.32, \quad s = 0.91, \quad v = 9$
5% betrouwbaarheidsgrenzen voor μ .

$$\bar{x}_{..} \pm t_{9, \alpha} * \frac{s}{\sqrt{n}} = 59.32 \pm 2.26 * \frac{0.91}{\sqrt{10}} = 58.67 \leftrightarrow 59.97; \alpha = 5\%.$$

Model II:
$$\begin{aligned} x_{ij} &= \mu + a_i + e_{ij}, \\ E(e) &= 0, \quad \text{var } e = \sigma^2, \quad \text{cov}(e, e') = 0. \end{aligned} \quad (35.2)$$

De beide reeksen komen uit twee verschillende populaties met gemiddelden

$$\mu_1 = \mu + a_1, \quad \mu_2 = \mu + a_2.$$

Voorbeelden: De metingen x_{1j} en x_{2j} zijn uitgevoerd aan telkens n producten vervaardigd uit twee verschillende partijen grondstoffen of op twee verschillende machines enz. a_1, a_2 vertegenwoordigen de verschillen tussen de partijen grondstoffen of de machines.

Bewerking: $\bar{x}_1 = 59.02, \quad \bar{x}_2 = 59.62$ en door samenvoegen $s = 0.84, \quad v = 8$.

95% betrouwbaarheidsintervallen

voor μ_1 : $59.02 \pm 2.31 * \frac{0.85}{\sqrt{5}}$,

" μ_2 : $59.62 \pm 2.31 * \frac{0.84}{\sqrt{5}}$,

" $\mu_2 - \mu_1$: $59.62 - 59.02 \pm 2.31 * 0.84 * \sqrt{\frac{1}{5} + \frac{1}{5}}$.

Model III A:
$$\underline{x}_{ij} = \mu + b_j + e_{ij} \quad (36.1)$$

$$E(\underline{e}) = 0, \quad \text{var } \underline{e} = \sigma^2, \quad \text{cov}(\underline{e}, \underline{e}') = 0$$

De waarnemingen vormen duplo's. De schatting van σ^2 is

$$s^2 = \frac{\sum d_j^2}{2n} = \frac{2.66}{2 \cdot 5} = 0.266$$

$$s = 0.51, \quad v = 5.$$

Hiermee 95% betrouwbaarheidsintervallen:

$$\begin{aligned} \text{voor } \mu + b_1: & \quad 60.40 \pm 2.57 * \frac{0.51}{\sqrt{2}}, \\ \text{" } \mu + b_2: & \quad 59.75 \pm 2.57 * \frac{0.51}{\sqrt{2}}, \quad \text{enz.} \end{aligned}$$

Model III B:
$$\underline{x}_{ij} = \mu + \underline{b}_j + e_{ij} \quad (36.2)$$

$$E(\underline{e}) = 0, \quad \text{var } \underline{e} = \sigma^2, \quad \text{cov}(\underline{e}, \underline{e}') = 0$$

$$E(\underline{b}) = 0, \quad \text{var } \underline{b} = \sigma_b^2, \quad \text{cov}(\underline{b}, \underline{b}') = 0$$

Dit is een variant van III A. De \underline{b}_j 's worden nu opgevat als een serie van n onafhankelijke trekkingen uit een populatie met verwachting $E(\underline{b}) = 0$ en $\text{var } \underline{b} = \sigma_b^2$. Voorbeeld: de n paren waarnemingen zijn n chemische analyses in duplo uitgevoerd aan n monsters uit een partij van een chemische grondstof. De \underline{b}_j 's vertegenwoordigen dan verschillen tussen deze monsters ten gevolge van inhomogeniteit van de partij.

Geldt Model III B dan kunnen we voor μ een betrouwbaarheidsinterval construeren, door de n gemiddelden \underline{x}_j als n originele waarnemingen te beschouwen. Dit geeft

$$\bar{x}_.. = 59.32, \quad s = 0.88, \quad v = 4.$$

95% betrouwbaarheidsinterval voor μ :

$$59.32 \pm 2.78 * \frac{0.88}{\sqrt{5}} = 58.23 \leftrightarrow 60.41; \quad \alpha = 5\%.$$

Model IV:
$$\underline{x}_{ij} = \mu + a_i + b_j + e_{ij}, \quad (36.3)$$

$$E(\underline{e}) = 0, \quad \text{var } \underline{e} = \sigma^2, \quad \text{cov}(\underline{e}, \underline{e}') = 0.$$

Bij dit model nemen we systematische effecten a_i en b_j aan in beide richtingen. Voorbeeld: de twee series van n waarnemingen betreffen twee partijen grondstoffen. De waarnemingen werden paarsgewijze op n verschillende dagen uitgevoerd, omdat men slechts 2 waarnemingen per dag kan uitvoeren. Tussen dagen treden verschillen b_j op, bijv. ten gevolge van vochtigheidsfluctuaties.

We vinden nu een schatting van s^2 uit de verschillen $d_j = x_{2j} - x_{1j}$ volgens

$$s^2 = \frac{\sum (d_j - \bar{d})^2}{2(n-1)} = 0.45, \quad v = 4. \quad (36.a.1)$$

en hiermee een betrouwbaarheidsinterval voor het verschil $\mu_2 - \mu_1 = \mu_{d.i}$

$$\bar{x}_2 - \bar{x}_1 \pm t_{4, \alpha} * s * \sqrt{\frac{1}{n} + \frac{1}{n}} = \quad (36.a.2)$$

$$= 0.60 \pm 2.78 * 0.45 * \sqrt{\frac{1}{5} + \frac{1}{5}} = -0.18 \leftrightarrow +1.38; \alpha = 5\%$$

hier: $s^2 = \frac{\sum (d_j - \bar{d})^2}{n-1}$

minimale $\mu_1 - \mu_2$: $\pm \frac{s}{\sqrt{n}}$

Nel met $\pm 2.78 * \frac{0.45}{\sqrt{5}}$

Discrete verdelingen; $x = 0, 1, 2, \dots$

Ook deze kunnen op grond van de centrale limietstelling vaak door een normale verdeling worden benaderd. Dan gelden de volgende formules:

$$P(\underline{x} \leq x) = \int_{-\infty}^{x+\frac{1}{2}} \varphi(\xi; \mu, \sigma) d\xi = \Phi\left(\frac{x+\frac{1}{2}-\mu}{\sigma}\right), \quad (37.1)$$

$$P(\underline{x} = x) = \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} \varphi(\xi; \mu, \sigma) d\xi = \Phi\left(\frac{x+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{x-\frac{1}{2}-\mu}{\sigma}\right). \quad (37.2)$$

enz.

Hierin is $\Phi(u)$ de cumulatieve normale verdeling Comp. Tabel 1.1; μ en σ zijn de toepasselijke waarden van deze parameters.

Voorbeelden:

1) Voor één dobbelsteen geldt $\mu = 3.5$, $\sigma^2 = 2.92$.

Voor de som \underline{x} van de worpen met 3 dobbelstenen geldt dan

$$\mu = 10.50, \quad \sigma = \sqrt{3 \cdot 2.92} = 2.95.$$

Daarmee wordt benaderd

$$P(\underline{x} \leq 6) = \Phi\left(\frac{6.5-10.5}{2.95}\right) = \Phi(-1.36) = 0.0869 \quad (\text{exact } 0.0925),$$

$$P(\underline{x} < 6) = \Phi\left(\frac{5.5-10.5}{2.95}\right) = \Phi(-1.69) = 0.0455 \quad (\text{exact } 0.0462)$$

waaruit

$$P(\underline{x} = 6) = 0.0869 - 0.0455 = 0.0414 \quad (\text{exact } 0.0462).$$

2) Voor de binomiale verdeling met $n = 20$, $p = 0.40$ geldt

$$\mu = np = 8, \quad \sigma = \sqrt{np(1-p)} = 2.2$$

en dus benaderd

$$P(\underline{x} \leq 12) = \Phi\left(\frac{12.5-8.0}{2.2}\right) = \Phi(2.05) = 0.9798 \quad (\text{exact } 0.9790).$$

Overschrijdingskansen

Is x een waarneming van een variabele \underline{x} met een bekende verdeling $f(x)$ dan noemen we

de linker overschrijdingskans van $x =$

$$\alpha_l = P(\underline{x} \leq x) = \int_{-\infty}^x f(\xi) d\xi, \quad (37.3)$$

en de rechter overschrijdingskans van $x =$

$$\alpha_r = P(\underline{x} \geq x) = \int_x^{\infty} f(\xi) d\xi. \quad (37.4)$$

In principe kunnen we deze overschrijdingskansen op twee manieren definiëren, bijv. α_r als

$$\alpha_r = P(\underline{x} > x) \text{ en } \alpha_l = P(\underline{x} > x) \quad (38.1)$$

Is \underline{x} een continue variabele dan maakt dit geen verschil, omdat $P(\underline{x} = x) = 0$.

Is \underline{x} discreet dan maakt het wel verschil en hebben we de keuze uit twee definities:

$$\alpha_l = P(\underline{x} \leq x), \quad \alpha_r = P(\underline{x} \geq x), \quad (38.2)$$

en

$$\alpha_l = P(\underline{x} < x) + \frac{1}{2}P(\underline{x} = x), \quad \alpha_r = P(\underline{x} > x) + \frac{1}{2}P(\underline{x} = x). \quad (38.3)$$

Kan de discrete verdeling van \underline{x} door een normale verdeling worden benaderd dan geldt

$$\alpha_l = \Phi\left(\frac{x + \frac{1}{2} - \mu}{\sigma}\right), \quad \alpha_r = 1 - \Phi\left(\frac{x - \frac{1}{2} - \mu}{\sigma}\right), \quad (38.4)$$

daarentegen

$$\alpha_l = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad \alpha_r = 1 - \Phi\left(\frac{x - \mu}{\sigma}\right), \quad (38.5)$$

en dus ook

$$\alpha_r + \alpha_l > 1, \quad \text{maar } \alpha_l + \alpha_r = 1. \quad (38.6)$$

De definitie (38.2) is de meest gebruikelijke, maar heeft het grote nadeel dat men bij het berekenen van betrouwbaarheidsgrenzen steeds een correctie van $+\frac{1}{2}$ of $-\frac{1}{2}$ moet aanbrengen, hetgeen de formules onnodig compliceert. De definitie berust op een conventie. In feite is (38.3) een logischer definitie (zie (38.6)) en bovendien in de praktijk eenvoudiger. De in deze syllabus gegeven betrouwbaarheidsintervallen voor discrete verdelingen zijn allen op de definitie (38.3) gebaseerd. De definitie (38.2) wordt echter in vele leerboeken over wiskundige statistiek toegepast.

Een betrouwbaarheidsinterval voor μ wordt bepaald door die twee grenswaarden μ_1 en μ_2 waarvoor, bij een gegeven reeks waarnemingen, de rechter resp. de linker overschrijdingskans gelijk wordt aan $\frac{1}{2}\alpha$. Fig. 38 licht dit toe.

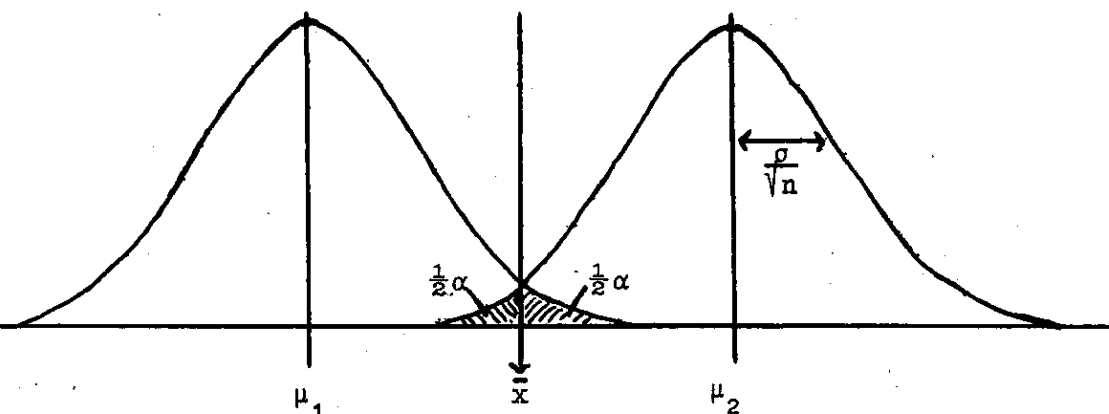


Fig. 38. Principe van betrouwbaarheidsgrenzen voor μ .

Ook bij een binomiale verdeling of een Poisson verdeling is ditzelfde principe algemeen toepasbaar. De situatie is eenvoudiger omdat we in deze gevallen slechts met één onbekende parameter, p of μ , te maken hebben waardoor de verdeling geheel wordt bepaald.

Voor

$$np > 5 \quad \text{en} \quad n(1-p) > 5 \quad (39.1)$$

kan een binomiale verdeling redelijk door een normale worden benaderd.

Dan is

$$\mu = np, \quad \sigma = \sqrt{np(1-p)}, \quad (39.2)$$

en met de definities (38.3) en (38.5) wordt het betrouwbaarheidsinterval voor p gegeven door

$$\frac{x - np_1}{\sqrt{np_1(1-p_1)}} = -u_\alpha, \quad \frac{x - np_2}{\sqrt{np_2(1-p_2)}} = +u_\alpha, \quad (39.3)$$

waarbij u_α tweezijdig moet worden gekozen.

De grenzen p_1 en p_2 zijn de wortels van de vergelijking

$$(x - np)^2 = u_\alpha^2 np(1-p),$$

te weten

$$\left. \begin{array}{l} p_1 \\ p_2 \end{array} \right\} = \frac{\hat{p} + \frac{u_\alpha^2}{2n} \mp u_\alpha \sqrt{\hat{p}(1-\hat{p})/n + \frac{u_\alpha^2}{4n^2}}}{1 + \frac{u_\alpha^2}{n}} \quad (39.4)$$

waarin $\hat{p} = x/n$ = de fractie \hat{p} in de steekproef waargenomen.

Formule (39.4) is te onhandig voor praktische gebruik. Een goede benadering is

$$\left. \begin{array}{l} p_1 \\ p_2 \end{array} \right\} = \hat{p} + \left(\frac{1}{2} - \hat{p}\right) \frac{u_\alpha^2}{n} \mp u_\alpha \sqrt{\hat{p}(1-\hat{p})/n}. \quad (39.5)$$

Deze formule geeft redelijke resultaten wanneer

$$x = n\hat{p} > 5, \quad \text{en} \quad n - x = n(1-\hat{p}) > 5. \quad (39.6)$$

Voor

$$x = n\hat{p} > 50 \quad \text{en} \quad n - x = n(1-\hat{p}) > 50 \quad (39.7)$$

kan men de 2e term nog zonder bezwaar weglaten en vereenvoudigen tot

$$\left. \begin{array}{l} p_1 \\ p_2 \end{array} \right\} = \hat{p} \mp u_\alpha \sqrt{\hat{p}(1-\hat{p})/n}. \quad (39.8)$$

Dit is de formule die verreweg het meest wordt toegepast. $\hat{p}(1-\hat{p})$ is een schatting van σ^2 , en (39.8) is een directe toepassing van de formule $\bar{x} \pm u_\alpha s/\sqrt{n}$ voor een populatiegemiddelde.

De formules voor een Poisson verdeling volgen uit dienvoor een binomiale verdeling uit de limietovergang $\mu = np$, $n \rightarrow \infty$.

(39.5) geeft

$$\left. \begin{array}{l} \mu_1 \\ \mu_2 \end{array} \right\} = x + \frac{1}{2} \frac{u_\alpha^2}{x} \mp u_\alpha \sqrt{x}, \quad (39.9)$$

bruikbaar voor $x > 5$ en (39.8) geeft

$$\left. \begin{array}{l} \mu_1 \\ \mu_2 \end{array} \right\} = x \pm u_\alpha \sqrt{x} \quad (40.1)$$

bruikbaar voor $x > 50$.

Voor $x = 0, 1, \dots, 20$ zijn betrouwbaarheidsgrenzen a_1, a_2 voor μ direct gegeven in de Compendium Tabel 5.3. Deze kunnen voor $\hat{p} < 10$ ook als benadering voor de binomiale verdeling worden gebruikt.

Dan zijn de grenzen voor p

$$\left. \begin{array}{l} p_1 \\ p_2 \end{array} \right\} = \frac{a_1}{n}, \frac{a_2}{n} \quad \text{Tabel } \overset{6}{\cancel{5}.3}. \quad (40.2)$$

Toepassingen van betrouwbaarheidsintervallen bij tellingen.

1. Is een steekproef van $n = 20$ producten uit een grote partij worden 4 exemplaren gevonden die niet aan de gestelde eisen voldoen. Gevraagd een 90% betrouwbaarheidsinterval voor $p =$ het percentage foute producten in de partij. Tabel geeft de resultaten volgens verschillende formules.

Tabel 41 90% betrouwbaarheidsintervallen voor p bij $n = 20$, $x = 4$.

$p_1\%$	$p_2\%$	
8.8	37.0	exact, Tabel 5.3
9.4	37.8	Formule (39.4)
8.4	37.8	" (39.5)
5.3	34.7	" (39.8)
8.3	42.3	" (40.2)

Formule (39.5) is een goede benadering en veel eenvoudiger dan (39.4); (40.2) is goed bij de benedengrens p_1 , maar niet bij p_2 .

(39.8) geeft een interval van de juiste breedte maar naar beneden verschoven.

2. In 1963 en 1964 werden 96515 en 98324 rekruten gekeurd en hiervan werden respectievelijk 22.64% en 22.30% afgekeurd. Is het verschil tussen deze percentages zuiver toevallig of niet?

De variantie voor beide percentages is ongeveer $0.16 \times 10^{-5} = 0.016(\%)^2$.

De variantie voor het verschil is dan $0.032(\%)^2$ en een 95% betrouwbaarheidsinterval:

$$22.64 - 22.30 \pm 2\sqrt{0.032} = -0.02 \leftrightarrow +0.70\%; \alpha = 5\%.$$

Het vermoeden dat men in 1963 iets strenger keurde dan in 1964 is dus gerechtvaardigd; doch het is niet helemaal zeker of dit vermoeden juist is.

3. Uit twee grote partijen producten, I en II, heeft men steekproeven van 200 en van 150 stuks genomen en hierin resp. 6 en 9 foute producten waargenomen. Bewijst dit een verschil in kwaliteit?

$$n_1 = 200, x_1 = 6, \hat{p}_1 = 0.030 = 3.0\%, \frac{\hat{p}_1(1-\hat{p}_1)}{200} = 1.5(\%)^2,$$

$$n_2 = 150, x_2 = 9, \hat{p}_2 = 0.060 = 6.0\%, \frac{\hat{p}_2(1-\hat{p}_2)}{150} = 4.0(\%)^2.$$

Het 95% betrouwbaarheidsinterval voor $p_2 - p_1$ is

$$6.0 - 3.0 \pm 2\sqrt{1.5 + 4.0} = -1.6 \leftrightarrow +7.6\%; \alpha = 5\%.$$

Men mag dus nog niet tot een verschil besluiten.

4. Komen beide partijen van eenzelfde productieproces dan geldt voor deze productie op grond van beide steekproeven samen:

$n = 350$, $x = 15$, en daaruit met de normale benadering voor p een interval:

$$\frac{1}{350}\left(15 + \frac{2^2}{2}\right) \pm 2\sqrt{15} = 2.6\% \leftrightarrow 7.1\%; \alpha = 5\%,$$

of met Tabel 6.3:

$$p_1 = \frac{8.75}{350} = 2.5\%; p_2 = \frac{24.1}{350} = 6.9\%; \alpha = 5\%.$$

5. Een radioactief preparaat geeft voor een Geigerteller 251 pulsen in 8 minuten. Voor het aantal pulsen per minuut, λ , geldt een interval.

$$\frac{251}{8} \pm 2\sqrt{\frac{251}{8}} = 27.4 \leftrightarrow 35.4; \alpha = 5\%.$$

Het toetsen van hypothesen.

Nauw verwand aan betrouwbaarheidsintervallen is het toetsen van hypothesen.

Een voorbeeld: Voor 18 THE-studenten vonden we voor de lengten zonder schoenen:

$$n = 18, \bar{x} = 180.6 \text{ cm.}, s = 5.9 \text{ cm.}$$

Voor ≈ 90.000 rekruten werd in 1964 een gemiddelde lengte gevonden van 176.0 cm.

Bewijst dit gegeven dat T.H.E.-studenten gemiddeld langer zijn dan de rekruten?

Gaan we uit van de hypothese dat voor de THE-studenten geldt:

$$H_0: \mu = 176.0 \text{ cm.},$$

dan is het gemiddelde \bar{x} in goede benadering een waarneming uit een normale verdeling met $\mu = 176.0 \text{ cm.}, \sigma_{\bar{x}} = s/\sqrt{n} = 5.9/\sqrt{18} = 1.4 \text{ cm.}$ Dit geeft

$$\mu = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{180.6 - 176.0}{1.4} = 3.3,$$

en hieruit $\alpha_r = 0.05\%$.

Het waargenomen gemiddelde heeft een rechter overschrijdingskans van slechts 0.05%. Dit is zó klein dat men de waarneming \bar{x} strijdig acht met de hypothese $\mu = 176.0 \text{ cm.}$, en men deze hypothese verwerpt.

Algemeen gaat men als volgt te werk.

- 1) Men maakt onderstellingen over de populatie waaruit een steekproef is waargenomen.
- 2) Men berekent uit de steekproef een karakteristieke grootte ($\bar{x}, \hat{\beta}, s$).
- 3) Men bepaald van de waargenomen grootte de overschrijdingskansen in de door de onderstellingen bepaalde steekproefpopulatie.
- 4) Is één van de overschrijdingskansen klein dan acht men de waarnemingen in strijd met de onderstellingen.

Meestal is men geïnteresseerd in een bepaalde populatieparameter en voert de toets uit met behulp van een steekproefkenmerk, dat daarmee overeenkomt, men toetst een waarde van μ met een gemiddelde \bar{x} , en een waarde van σ met een standaardafwijking s .

Men onderscheidt dienovereenkomstig bijkomstige onderstellingen en de nulhypothese die men wenst te toetsen. In het beschouwde voorbeeld gaan we uit van de onderstelling dat de 18 lengten een onderling onafhankelijke serie waarnemingen vormen uit een populatie met een standaardafwijking $\sigma = s = 5.9$ cm., en dat de centrale limietstelling van toepassing is. De nulhypothese is

$$H_0: \mu = 176.0 \text{ cm.}$$

Verder kiest men dikwijls een kritische drempelwaarde α voor de overschrijdingskans, en beschouwt een waargenomen \bar{x} in strijd met de nulhypothese, wanneer één van de overschrijdingskansen van \bar{x} kleiner is dan α .

Men zegt dan dat de waarneming significant afwijkt van de hypothese bij een signifikantieniveau α of bij een onbetrouwbaarheidsdrempel α .

Verder onderscheidt men:

$$\begin{array}{ll} \text{de tweezijdige hypothese: } \mu = \mu_0, & (43.1) \\ \text{en de éézijdige hypothesen: } \mu \leq \mu_0, & (43.2) \\ \text{en } \mu \geq \mu_0. & (43.3) \end{array}$$

De tweezijdige hypothese (43.1) kan worden verworpen zowel wanneer $\bar{x} > \mu_0$ als wanneer $\bar{x} < \mu_0$; de éézijdige hypothese (43.2) kan alleen worden verworpen wanneer $\bar{x} > \mu_0$, en (43.3) alleen wanneer $\bar{x} < \mu_0$.

Bij een significantie niveau α verwerpt men de tweezijdige hypothese (43.1) wanneer $\alpha_r < \frac{1}{2}\alpha$ of $\alpha_e < \frac{1}{2}\alpha$; men spreekt van een tweezijdige toets.

De éézijdige hypothese (43.2) wordt verworpen wanneer $\alpha_r < \alpha$, en de éézijdige hypothese (43.3) wanneer $\alpha_e < \alpha$; men spreekt dan van éézijdige toetsen.

De drempelwaarde α geeft een bovengrens voor de kans dat men een juiste hypothese ten onrechte verwerpt.

Het toetsen van hypothesen is nauw verwant aan betrouwbaarheidsintervallen. Indien de hypothese $\mu = \mu_0$ bij een drempel α niet wordt verworpen ligt μ_0 in het betrouwbaarheidsinterval met onbetrouwbaarheid α . Één- en tweezijdige betrouwbaarheidsintervallen omvatten al die waarden van μ_0 die bij corresponderende één- of tweezijdige toets niet worden verworpen. Toetsen komt dus neer op het vaststellen of de waarde μ_0 al of niet in een bepaald betrouwbaarheidsinterval ligt.

Waar toepasbaar verdienen betrouwbaarheidsintervallen de voorkeur omdat zij een volledig beeld geven van de interpretaties die een gegeven serie waarnemingen toelaten. Betrouwbaarheidsintervallen bestaan echter alleen, wanneer we slechts met één enkele onbekende parameter (μ, σ, p) te maken hebben. Daarnaast komen vele situaties voor waarbij meerdere onbekende parameters optreden, en dan kan men wel toetsen, maar meestal geen eenvoudig betrouwbaarheidsinterval construeren. Voorbeelden volgen later.

Hieronder volgt een samenvatting van een aantal veel voorkomende toetsen, corresponderend met de betrouwbaarheidsintervallen van Tabel 27.

Tabel

Enige gangbare toetsen.

	Hypothese	Toetsingsgrootheid	Onderstellingen
1	$\mu = \mu_0$	$u = \frac{(\bar{x} - \mu_0)}{s/\sqrt{n}}$	C.L. stelling; $s \approx \sigma$
2	"	$t_v = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	Normale verdeling
3	$\mu_1 = \mu_2$	$u = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$	C.L. stelling; $s \approx \sigma_1$ $s \approx \sigma_2$
4	"	$t_v = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{1/n_1 + 1/n_2}}$	Normale verdelingen $\sigma_1 = \sigma_2$
5	$\sigma_1^2 \leq \sigma_2^2$	$F_{v_1, v_2} = s_1^2/s_2^2$	Normale verdelingen
6	$p = p_0$	$u = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	Binomiale verdeling $np_0 > 5$
7	$p_1 = p_2$	$u = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$ $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$	Binomiale verdeling
8	$\mu = \mu_0$	$u = \frac{\bar{x}_\cdot - \mu_0}{\sqrt{\mu_0/n}}$	Poisson verdeling $\mu_0 > 5$
9	$\mu_1 = \mu_2$	$u = \frac{\bar{x}_{1\cdot} - \bar{x}_{2\cdot}}{\sqrt{\bar{x}_\cdot(1/n_1 + 1/n_2)}}$ $\bar{x}_\cdot = \frac{x_{1\cdot} + x_{2\cdot}}{n_1 + n_2}$	Poisson verdeling

Toelichting

De toetsingsgrootheid is de grootheid waarmee de overschrijdingskans kan worden afgelezen uit de betreffende tabellen; Tabel 1.1 voor u , Tabel 2.1 voor t_v , en Tabel 4 voor F .

Men kan de toetsen ook zó uitvoeren, dat men de waarde van u , t , of F die uit de waarnemingen volgen vergelijkt met de tabel waarden corresponderend met een gegeven drempel α . De hypothese $\mu = \mu_0$ wordt dan bijv. verworpen wanneer $|u| > u_\alpha$ (tweezijdig), de hypothese $\mu \leq \mu_0$ wanneer $u > u_\alpha$, en de hypothese $\mu \geq \mu_0$ wanneer $u < -u_\alpha$ (éénzijdig).

De toetsen 1, 2, 3, en 4 spreken verder voor zichzelf, en komen direct overeen met de betrouwbaarheidsintervallen van Tabel 27.

Wat de toetsen voor fracties 6 en 7 betreft is er een klein verschil. Door de hypothese $p = p_0$ ligt ook de var $\hat{p} = p_0(1-p_0)/n$ vast, en deze passen we toe in de noemer van de toets.

Luidt de hypothese $p_1 = p_2 = p$ dan is de beste schatting van deze fractie $\hat{p} = (x_1 + x_2)/(n_1 + n_2)$ en voor de variantie van het verschil

$$\text{var}(\hat{p}_1 - \hat{p}_2) \approx \hat{p}(1-\hat{p}) \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad (45.1)$$

en deze schatting is gebruikt in de noemer van toets 7. Bij een betrouwbaarheidsinterval gaan we er van uit dat $p_1 \neq p_2$ kan zijn, en dan gebruiken we als schatting

$$\text{var}(\hat{p}_1 - \hat{p}_2) = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}. \quad (45.2)$$

Het is een principieel verschil. In praktijkgevallen geven (45.1) en (45.2) nagenoeg dezelfde numerieke uitkomst.

Wat de variantie betreft is (5) verreweg de belangrijkste toets.

Er doet zich in de z.g. variantie-analyse veelvuldig de situatie voor dat we twee varianties s_1^2 en s_2^2 wensen te vergelijken, terwijl we weten dat $\sigma_1^2 \geq \sigma_2^2$, en waar men wil beslissen tussen de twee mogelijkheden, $\sigma_1^2 = \sigma_2^2$ en $\sigma_1^2 > \sigma_2^2$.

Men toetst dan de éénzijdige hypothese

$$H_0 : \sigma_1^2 \leq \sigma_2^2, \quad (45.3)$$

en wanneer deze hypothese wordt verworpen beschouwt men te hebben aangetoond dat $\sigma_1^2 > \sigma_2^2$. Is $s_1^2 < s_2^2$ dan is een toets overbodig; is $s_1^2 > s_2^2$ dan toetst men met behulp van

$$F_{v_1, v_2} = s_1^2 / s_2^2. \quad (45.4)$$

Waarden van F_{v_1, v_2}^1 met een voorgeschreven éénzijdige overschrijdingskans zijn in Tabel 4 gegeven.

Wil men de tweezijdige hypothese $\sigma_1^2 = \sigma_2^2$ toetsen dan neemt men als toetsingsgrootheid

$$F_{v_2, v_1}^1 = s_1^2 / s_2^2 \quad \text{wanneer} \quad s_1^2 > s_2^2, \quad (46.1)$$

en

$$F_{v_1, v_2}^1 = s_2^2 / s_1^2 \quad \text{wanneer} \quad s_2^2 > s_1^2.$$

De overschrijdingskansen uit Tabel 4 moeten dan echter worden verdubbeld, omdat we nu een tweezijdige toets uitvoeren.

Het onderscheid tussen één- en tweezijdig toetsen is een wiskundige verfijning die in de praktijk vaak verwarrend werkt. De gemiddelde lengte van 18 THE-studenten, 180.6 cm, wijkt significant af van de Nederlandse rekruten, 176.0 cm, met een éénzijdige overschrijdingskans van $\alpha = 0.5\%$ (t-toets). Bij een tweezijdige toets noemt men dit significant bij een drempel van $\alpha = 1\%$; men houdt daarbij rekening met het feit dat een even grote significantie zou zijn geconstateerd wanneer een gemiddelde van 171.4 cm voor de studenten was geworden. Voor praktisch ingestelden is het helemaal niet duidelijk dat het zin heeft met dit alternatief rekening te houden.

Een eenvoudige oplossing is, bij een gegeven waarnemingsresultaat de éénzijdige overschrijdingskans te vermelden, en aan de hand daarvan een oordeel te vellen. Wie tweezijdig wil toetsen kan de kans dan altijd zelf verdubbelen.

De termen "significant" en "het verwerpen van een hypothese" zijn verder weinig zeggend en kan men beter vermijden. Heeft men voldoende waarnemingen, dan wordt zelfs het kleinste verschil tussen gemiddelen significant. Het woord significant geeft dus geen enkele aanwijzing over de orde van grootte van de waargenomen verschillen.

Ook de overschrijdingskans heeft niet de scherpe kritische betekenis, die men er in de wiskundige statistiek aan toekent. Een overschrijdingskans van ongeveer 5% wil zeggen dat de waarnemingen en de nul-hypothese, en dat men dus eens na dient te gaan of hiervoor een verklaring is te vinden. Soms zal men bij een overschrijdingskans van 1% toch zijn hypothese handhaven omdat men uit vroeger ervaring in deze hypothese een groot vertrouwen heeft. In andere gevallen, -bijv. wanneer men het percentage successen met een bepaalde operatie in een bepaald ziekenhuis vergelijkt met het landelijk gemiddelde- zal men zelfs bij een overschrijdingskans van 10% of 20% toch tot een nader onderzoek naar mogelijke oorzaken van het waargenomen verschil overgaan, omdat het hier om mensenlevens gaat.

Algemeen dient men toetsen te beschouwen als een internationale beoordelingsnorm waarmee men bepaalde veelvuldig voorkomende vragen op uniforme wijze beantwoordt. De uniformiteit van oordelen is daarbij het waardevolle element. Of een onbetrouwbaarheidsdrempel α exact 5% of alleen van de orde van grootte van 5% is doet niet ter zake, zolang als de uniformiteit van ons handelen niet door een verschil van mening hierover in gevaar wordt gebracht.

Steekproefkeuringen

Toetsen van een hypothese wordt als routine toegepast bij keuringen van partijen producten. Stel men ontvangt een partij staaldraad waarvoor is voorgeschreven dat de gemiddelde treksterkte groter moet zijn dan 140 kg/mm^2 ; stel verder dat we op grond van vele vroegere metingen weten dat de standaardafwijking van een trekvastheidsbepaling $\sigma = 5.7 \text{ kg/mm}^2$ bedraagt, en dat men doorgaans goed materiaal geleverd krijgt. Voor alle zekerheid wordt niettemin uit iedere geleverde partij een steekproef onderzocht en de partij wordt afgekeurd wanneer hieruit blijkt dat $\mu < 140 \text{ kg/mm}^2$. Daartoe toetst men dan de hypothese $H_0: \mu \geq 140 \text{ kg/mm}^2$ met een drempelwaarde α en keurt een partij af wanneer deze hypothese wordt verworpen.

Is de centrale limietstelling van toepassing, hetgeen praktisch steeds het geval is, dan luidt het criterium:

$$\text{Afkeuren wanneer } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -u_\alpha$$

of

$$\bar{x} < \mu_0 - u_\alpha \sigma/\sqrt{n}. \quad (47.1)$$

$$\text{of nog eenvoudiger } \bar{x} < G, \quad (47.2)$$

waarbij G een waarde heeft die afhangt van n en α . Bijv. voor $\mu_0 = 140 \text{ kg/mm}^2$, $\sigma = 5.7 \text{ kg/mm}^2$, $\alpha = 5\%$ (éénzijdig) en $n = 10$ is

$$G = 140 - 1.645 \times 5.7/\sqrt{10} = 137.0 \text{ kg/mm}^2.$$

Is het ware partijgemiddelde gelijk aan μ dan is de goedkeurkans, P_g , gegeven door

$$P_g = P(\bar{x} > G) = 1 - \Phi\left(\frac{G - \mu}{\sigma/\sqrt{n}}\right). \quad (47.3)$$

Voor het beschouwde numerieke geval is P_g in Fig. 48 uitgezet als functie van μ .

Een dergelijke kromme heet het onderscheidingsvermogen van de toets, of de keuringskarakteristiek. Bij het ontwerpen van steekproefkeuringen spelen de keuringskarakteristieken een beslissende rol, omdat zij overzichtelijk in beeld brengen wat een voorgenomen keuring presteert.

De keuze $\mu_0 = 140 \text{ kg/mm}^2$ en $\alpha = 5\%$ heeft tot gevolg dat een partij met een gemiddelde treksterkte van 140 kg/mm^2 een kans van 5% heeft te worden afgekeurd.

Fig. 48 laat zien dat een steekproefkeuring geen scherpe scheiding maakt tussen goed en slecht; er is een geleidelijke overgang van partijen met $\mu < 134 \text{ kg/mm}^2$ die praktisch steeds worden afgekeurd, naar $\mu > 140 \text{ kg/mm}^2$ die praktisch steeds worden goedgekeurd.

De steilheid waarmee de karakteristiek verloopt hangt af van de standaardafwijking van \bar{x} : σ/\sqrt{n} . Voor $n = 40$ vinden we bijv. $G = 138.5 \text{ kg/mm}^2$ en de steilere kromme in Fig. 48.

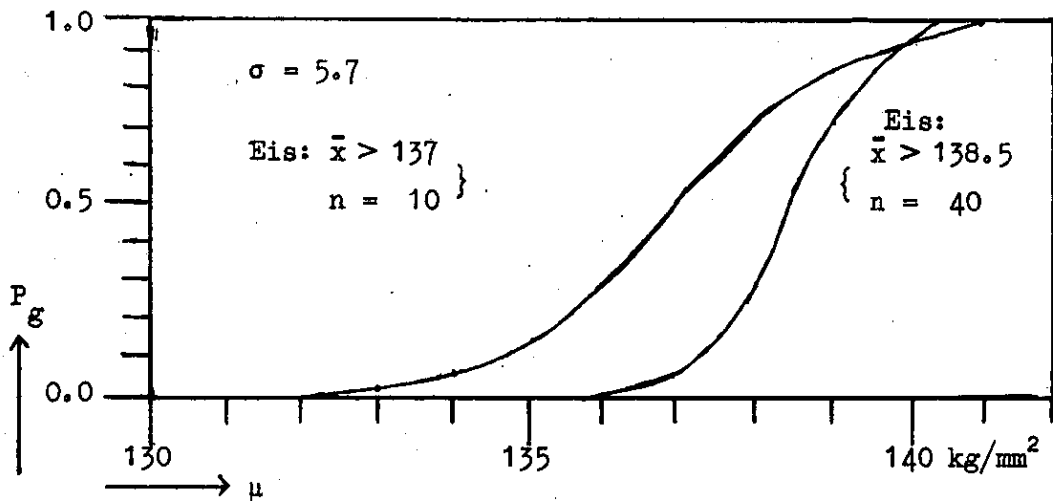


Fig. 48. Keuringskarakteristieken of het onderscheidingsvermogen van twee toetsen.

Soms stelt men twee grenzen vast waar \bar{x} tussen moet liggen; een partij wordt goedgekeurd wanneer $G_1 < \bar{x} < G_2$. Dan krijgen we twee krommen als die in Fig. 48, die samen de keuringskarakteristiek bepalen; zie Fig. 48.a, die overeenkomt met een tweezijdige toets van de hypothese $\mu_0 = 140 \text{ kg/mm}^2$, met $\alpha = 10\%$ (tweezijdig).

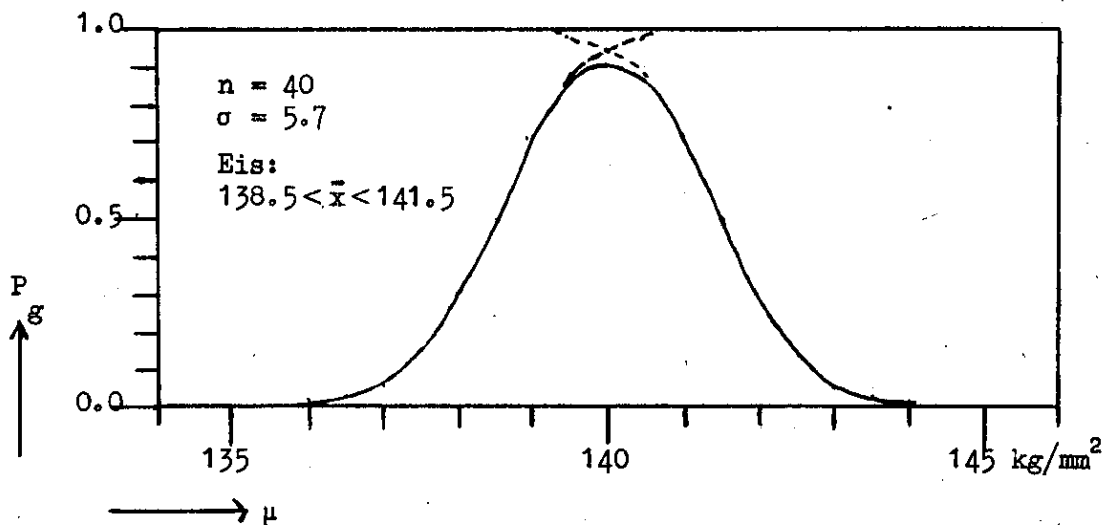


Fig. 48.a. Keuringskarakteristiek bij een tweezijdig criterium $138.5 < \bar{x} < 141.5 \text{ kg/mm}^2$.

Gaan we er niet van uit dat σ bekend is, en onderstellen we een normale verdeling dan zal de voorwaarde (47.1) waarbij wordt afgekeurd worden vervangen door

$$\bar{x} < \mu_0 - t_{v, \alpha} s / \sqrt{n}. \quad (48.1)$$

Nu moeten we echter met variaties van μ en σ beide rekening houden; de keuringskarakteristiek is dan een functie van μ en σ en kan niet meer door een eenvoudige kromme worden weergegeven.

Keuringen als hierboven beschouwd, gebaseerd op metingen, staan bekend als keuringen op variabelen. In de praktijk maakt men ook gebruik van een keuring op attributen. Daarbij worden de producten uitsluitend onderscheiden

in goede en defektieve. Defektieven zijn producten die niet aan de gestelde kwaliteitseisen voldoen. Het percentage defektieve producten in een partij heet het uitvalpercentage, of ook de kwaliteit van de partij.

De eenvoudigste attributenkeuring bestaat uit een enkelvoudig steekproef-schema. Men keurt een steekproef van n producten en keurt de partij goed indien hierin c of minder defektieven worden aangetroffen. Worden meer dan c defektieven gevonden, dan wordt de partij afgekeurd.

n heet de steekproefgrootte, c het goedkeurkriterium.

De keuringskarakteristiek geeft nu de goedkeurkans P_g als functie van het uitvalpercentage p . Daar p doorgaans kleiner is dan 10% kan de keuringskarakteristiek uit de Poissonverdeling worden bepaald. Tabel 49 geeft een voorbeeld.

Tabel 49.

Constructie van een keuringskarakteristiek van het enkelvoudig steekproef-schema $n = 80$, $c = 3$, met behulp van Compendium Tabel 6.2.

$\mu = np =$	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
$P_g =$	0.98	0.86	0.65	0.43	0.26	0.15	0.08	0.04
$p =$	1.25	2.50	3.75	5.00	6.25	7.50	8.75	10.00%

Dit geeft de punten in Fig. 49.

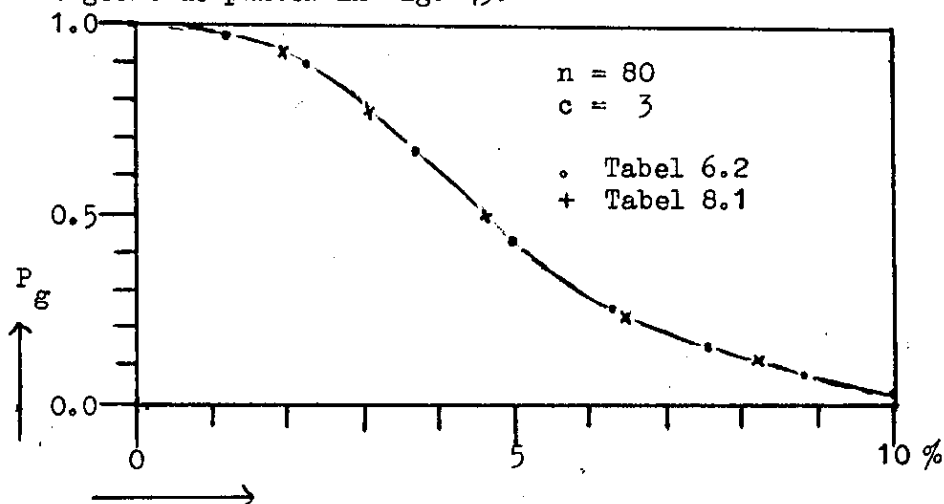


Fig. 49. Een keuringskarakteristiek voor een percentage defektieven p .

Deze keuringskarakteristieken kan men ook doelmatig construeren met behulp van Tabel 8.1, die waarden van $\mu = np$ geeft bij gegeven waarden van P_g en c .

Tabel 50.

Berekening van de keuringskarakteristiek voor het enkelvoudig steekproef-schema $n = 80$, $c = 3$, met behulp van Compendium Tabel 8.1.

$P_g =$	0.90	0.75	0.50	0.25	0.10
$np =$	1.745	2.535	3.675	5.109	6.681
$p =$	2.2	3.2	4.6	6.4	8.3 %

Dit geeft de kruisjes in Fig. 49.

Fig. 50 brengt in beeld hoe de keuringskarakteristiek veranderd bij variaties van n en c .

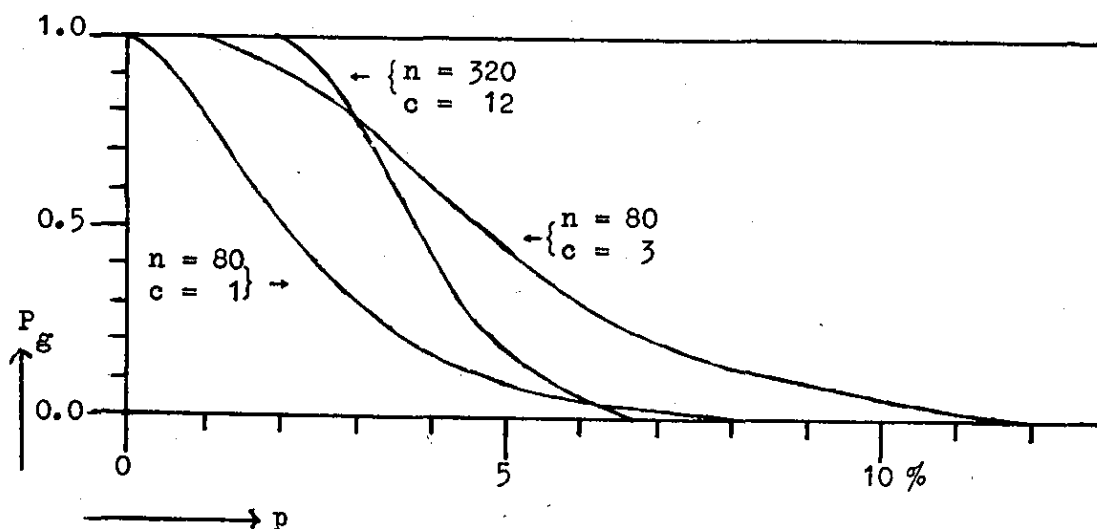


Fig. 50. Verloop van keuringskarakteristieken bij veranderingen van n en c .

Houdt men n vast en neemt c toe dan verschuift de karakteristiek naar rechts; houdt men de verhouding c/n konstant dan blijft de kromme ongeveer op dezelfde plaats doch wordt steiler naarmate n toeneemt.

Ook een attributenkeuring kan men zien als het toetsen van een hypothese. Bij $n = 80$, $c = 3$ is de benedengrens van een ééNZijdig 95% betrouwbaarheidsinterval $p_1 = 1.4\%$. Het keuringschema $n = 80$, $c = 3$ keurt die partijen af waarvan met een betrouwbaarheidsdrempel van 5% is aangetoond dat het uitvalpercentage hoger is dan 1.4%.

Naast enkele eenvoudige kent men dubbele, en meervoudige steekproefschema's. Een dubbel schema werkt als volgt.

Men keurt eerst een eerste steekproef van n_1 stuks, is x_1 het aantal defektieven hierin gevonden, dan wordt

de partij goedgekeurd wanneer $x_1 \leq c$,
 " " afgekeurd " $x_2 > c$.

Is $c_1 < x_1 \leq c_2$ dan wordt een tweede steekproef van n_2 stuks genomen en uiteindelijk

de partij goedgekeurd wanneer $x_1 + x_2 \leq c_3$,
 " " afgekeurd " $x_1 + x_2 > c_3$.

Het voordeel is dat men uitgesproken goede of uitgesproken slechte partijen reeds bij de eerste steekproef onderkend, en alleen in twijfelgevallen een tweede steekproef neemt. Men spaart daardoor waarnemingen uit. Een nadeel is dat het aantal te verrichten waarnemingen niet van te voren vast ligt en ten dele van toevallige omstandigheden afhangt.

De formule van de goedkeurkans luidt nu

$$P_g = P(\underline{x}_1 \leq c) + \sum_{i=c_2+1}^{c_3} P(\underline{x}_1 = i) P(\underline{x}_2 \leq c_3 - i), \quad (51.1)$$

en de gemiddelde steekproefgrootte is

$$E(\underline{n}) = n + n \sum_{i=c_2+1}^{c_3} P(\underline{x}_1 = i). \quad (51.2)$$

De kansen P kunnen in het algemeen uit de Poisson verdeling worden afgelezen.

Tabel 51.

Voorbeeld van de berekening van een keuringskarakteristiek en van $E(\bar{n})$ voor een dubbel steekproefschema.

Schema: $n_1 = n_2 = 50$, $c_1 = 0$, $c_2 = 2$, $c_3 = 3$.				
P%	$P(\underline{x}_1 \leq 0) + P(\underline{x}_1 = 1) \times P(\underline{x}_2 \leq 2) + P(\underline{x}_1 = 2) \times P(\underline{x}_2 \leq 1) =$			P_g
2	0.3679	+ 0.3679 × 0.9197	+ 0.1839 × 0.7358 =	0.841
4	0.1353	+ 0.2707 × 0.6767	+ 0.2707 × 0.4060 =	0.428
6	0.0498	+ 0.1494 × 0.4232	+ 0.2240 × 0.1991 =	0.158
8	0.0183	+ 0.0733 × 0.2381	+ 0.1465 × 0.0961 =	0.059
P%	n_1	+ $n_2 \times \{P(\underline{x}_1 = 1) + P(\underline{x}_2 = 2)\}$	=	$E(\underline{n})$
2	50	+ 50 × { 0.3679 + 0.1839 }	=	77.6
4	50	+ 50 × { 0.2707 + 0.2707 }	=	77.1
6	50	+ 50 × { 0.1494 + 0.2240 }	=	68.7
8	50	+ 50 × { 0.0733 + 0.1465 }	=	61.0

De keuringskarakteristiek heeft eenzelfde vorm als in de Figuren 49 en 50.

Fig. 52 geeft $E(n)$ weer als functie van p .

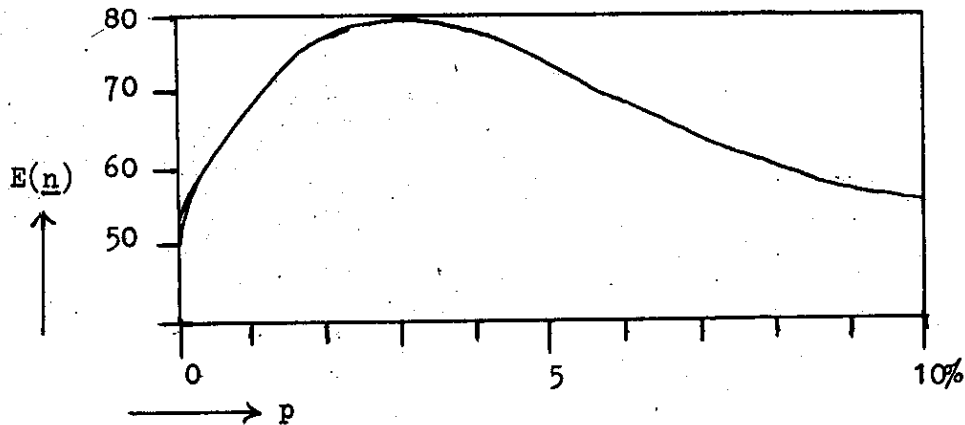


Fig.52 $E(n)$ als functie van p .

$E(n)$ gaat als functie van p door een maximum. Zeer goede en zeer slechte partijen vereisen slechts een eerste steekproef, en dus $E(n) = 50$. Bij $p = 3\%$ ligt het maximum; daar komen het meeste twijfelgevallen voor, die een tweede steekproef verlangen.

Door een juiste keuze van de steekproefgrootten, n_1, n_2 , en de criteria, c_1, c_2 , en c_3 , kan men dubbele steekproefschema's construeren die praktisch dezelfde keuringskarakteristieken bezitten als enkelvoudige schema's, en deze dus zonder meer kunnen vervangen. Globaal kan men door gebruik van een dubbel schema 20% à 25% op het aantal waarnemingen besparen.

Een meêrvoudig steekproefschema verloopt in meer dan twee stappen. Na iedere steekproef neemt men één van 3 beslissingen: goedkeuren, afkeuren, of een volgende steekproef nemen.

Een steeds verdere verfijning leidt tot de z.g. sequente steekproefmethode. Hierbij ligt de steekproefgrootte niet vast. Men keurt de producten één voor één en zet stap voor stap tegen elkaar uit, het totaal aantal gekeurde producten, en het aantal gevonden defektieven. Dit geeft een trapvormig omhooggaande serie punten als in Fig. 52.a. In dit diagram worden nu twee evenwijdige lijnen getrokken en de partij die wordt onderzocht wordt goedgekeurd indien het toevalspad dat door de keuring wordt beschreven de lijn G het eerst snijdt, en afgekeurd indien de lijn A het eerst wordt gesneden. Men kan bewijzen dat deze methode van keuren een gegeven keuringskarakteristiek gemiddeld met het kleinst mogelijke aantal waarnemingen kan realiseren. Voor de industrie is deze methode echter onpraktisch omdat de steekproefgrootte te variabel is. De methode vindt wel toepassing bij medisch onderzoek naar het effect van een geneesmiddel; daar ligt dit voor de hand omdat de patiënten zich één voor één aanmelden.

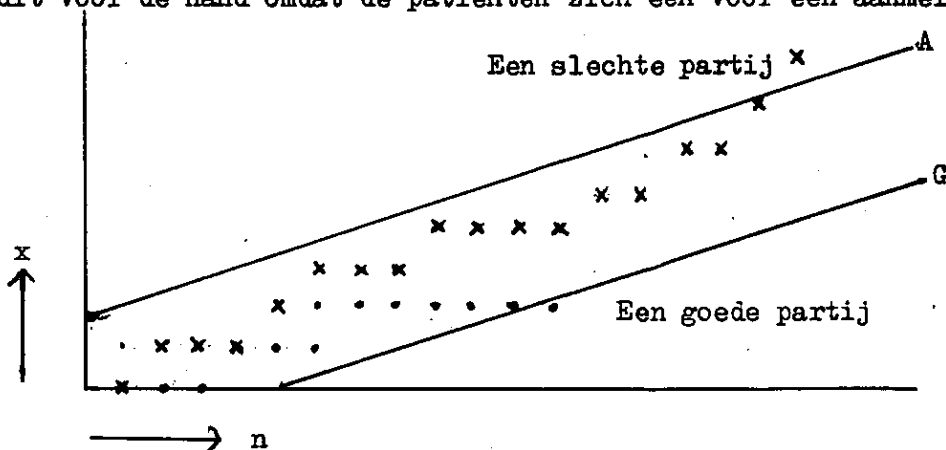


Fig.52.a. Principe van een sequente steekproefkeuring.
 n = aantal gekeurd, x = aantal defektieven.

Een keuring op variabelen kan steeds worden uitgevoerd als een keuring op attributen. In plaats van de eis te stellen dat een steekproefgemiddelde \bar{x} boven een gegeven grens G_0 moet liggen, als in Fig. 50, kan men bijv. eisen dat van de n waarnemingen er niet meer dan c beneden een andere grens G_1 mogen liggen. Men definieert dan producten waarvoor $x < G_1$ als defektieven.

Het is niet moeilijk voor deze methode van keuren een keuringskarakteristiek te berekenen. De fractie producten p waarvoor $x < G_1$ wordt bij een normale verdeling gegeven door

$$p = \Phi\left(\frac{G_1 - \mu}{\sigma}\right)$$

en met behulp van de binomiale verdeling of een van haar benaderingen kan men met deze waarde de goedkeurkans P_g berekenen; Tabel en Fig. 53. G_1 werd bij dit voorbeeld zó gekozen dat wederom, als in Fig. 50, $P_g = 0.95$ voor $\mu = 140 \text{ kg/mm}^2$.

Tabel 53

Voorbeeld van berekening van de karakteristiek van een variabelenkeuring uitgevoerd als attributen keuring.

Schema: $n = 40$, $c = 3$; ondersteld: $\sigma = 5.7$.					
P_g	np Tabel 8.1	$pf\%$	Tabel 1.1	$-u\sigma$	$G_1 - u\sigma$ $G_1 = 129.6$
0.95	1.366	3.42	- 1.82	10.4	140.0
0.90	1.745	4.36	- 1.71	9.7	139.3
0.75	2.535	6.34	- 1.53	8.7	138.3
0.50	3.672	9.18	- 1.33	7.6	137.2
0.25	5.109	12.77	- 1.14	6.5	136.1
0.10	6.681	16.70	- 0.97	5.5	135.1
0.05	7.754	19.38	- 0.86	4.9	134.5

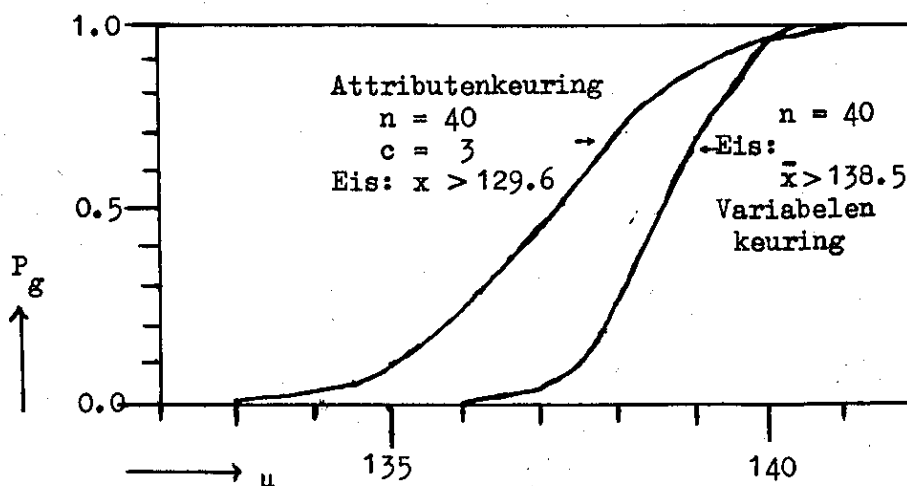


Fig. 53. Vergelijking van een keuring op variabelen met een attributenkeuring.

We zien uit deze figuur dat de attributenkeuring veel minder efficiënt is dan de keuring op variabelen. De steekproefgrootte is dezelfde, doch de keuringskarakteristiek loopt veel minder steil.

De attributenkeuring met $n = 40$, $c = 3$ is niet veel beter dan de variabelen keuring met $n = 10$ uit Fig. 50.

Niettemin wordt in de industrie in de meeste gevallen een attributenkeuring toegepast. Bij een variabelenkeuring is de keuringskarakteristiek alleen berekenbaar als de vorm van de verdeling van de gemeten grootte bekend is. Wanneer men echter niet weet onder welke omstandigheden een partij producten is vervaardigd is iedere onderstelling over de vorm van de verdeling riskant.

Voorts kan een attributenkeuring worden uitgevoerd met behulp van kalibers; zij vereist een eenvoudiger instrumentarium en minder geschoold personeel en dat weegt voor een groot deel op tegen het nadeel van een geringere efficiëntie.

We maken er nog op attent dat de karakteristiek voor attributenkeuring in Fig. 53 nadrukkelijk een normale verdeling onderstelt, terwijl de variabelenkeuring alleen op de geldigheid van de centrale limietstelling is gebaseerd. Men kan de karakteristiek voor attributen ook uitzetten tegen het percentage defektieven, d.w.z. het percentage produkten waarvoor $x < G$. Gebruikt men attributenkeuring dan is dit de aangewezen methode. Er is tussen μ en p alleen een duidelijk verband, wanneer de vorm van de verdeling bekend is.

Bij de praktische toepassing van een steekproefkeuring gaat men veelal uit van een gewenste karakteristiek, en construeert vervolgens een steekproefschema dat aan deze wens voldoet. De keuringskarakteristieken zijn steeds krommen van eenzelfde algemene vorm, die voldoende wordt vastgelegd door twee parameters. Hiervoor kiest men bij een attributenkeuring bijv.

1) het risicopunt van de leverancier = p_{95} =

= het uitvalspercentage met een goedkeurkans van 95%.

2) het risicopunt van de afnemer = p_{10} =

= het uitvalpercentage met een goedkeurkans van 10%.

De leverancier of producent weet dat hij weinig moeilijkheden met afgekeurde partijen zal krijgen zolang de kwaliteit van de geleverde partijen beter is dan p_{95} . De afnemer weet dat partijen waarvoor $p > p_{10}$ slechts zelden door de

keuring heen zullen glippen.

Schrijft men voor

$$p_{95} = 2.0\%, \quad p_{10} = 6.0\%, \quad \text{dus} \quad \frac{p_{10}}{p_{95}} = 3.0$$

dan volgt een bijpassend enkelvoudig steekproefschema uit Tabel 8.1. Uit de laatste kolom zien we dat $c = 7$ het eerste schema is waarvoor $\frac{p_{10}}{p_{95}} \leq 3.0$.

Kiezen we $c = 7$ dan volgt uit $p_{95} = 2\%$, $n = 199$ en uit $p_{10} = 6\%$, $n = 196$. In de praktijk zal men dit afronden tot $n = 200$ en is

$$n = 200, \quad c = 7$$

het gezochte enkelvoudige steekproefschema.

Dat men p_{95} en p_{10} kiest is puur conventie, doch algemeen gebruikelijk.

Een andere nuttige parameter is het

$$\text{controlepunt} = p_{50} =$$

= het uitvalpercentage met een goedkeurkans van 50%.

Dit houdt zowat het midden tussen wat men beschouwt als goede en slechte partijen. Tussen p_{50} , n , en c bestaat een eenvoudige relatie (zie Tabel 8.1)

$$np_{50} = c + 2/3. \quad (55.1)$$

Is p_{50} gegeven dan blijft nog een zekere vrijheid in de keuze van n en c over en dit is in de praktijk vaak nuttig. Een andere veel toegepaste parameter is de

$$\text{AQL} = \text{Acceptable Quality Level} = p_a =$$

het verlangde kwaliteitsniveau.

Hieronder verstaat men het gemiddeld uitvalpercentage dat men in geleverde partijen toelaatbaar acht.

p_a stelt een eis aan de productie, het is niet een eigenschap van een keuringskarakteristiek zoals p_{95} , p_{50} , en p_{10} . Is p_a gegeven, dan zal men steeds een steekproefkeuring kiezen die bij $p = p_a$ een hoge waarde voor de goedkeurkans geeft, doch deze waarde van P_g is niet strikt voorgeschreven, en dit is een praktisch voordeel.

Tenslotte noemen we de

$$\text{AOQL} = \text{Average Outgoing Quality Limit} = \bar{p}_m =$$

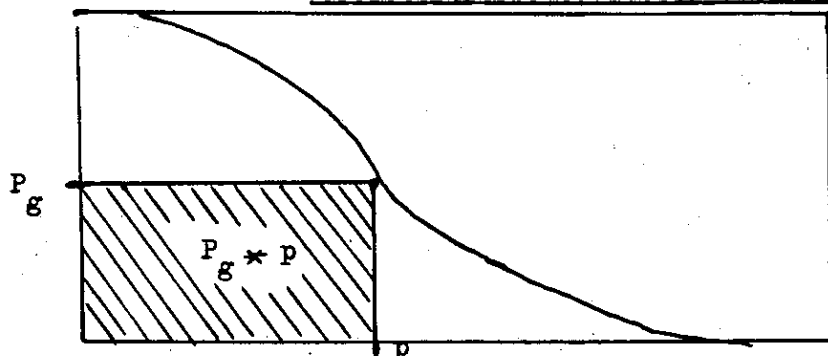
het maximaal gemiddeld-doorgelaten uitvalpercentage.

Hierbij wordt ondersteld dat afgekeurde partijen aan een 100% keuring worden onderworpen en dat daarbij alle defektieve producten worden geëlimineerd.

Nemen we verder aan dat alle gekeurde partijen eenzelfde uitvalpercentage, p , bevatten dan zal het gemiddeld uitvalpercentage in de doorgelaten partijen

$$E(p) = P_g * p + (1 - P_g) * 0 = P_g * p,$$

bedragen, afgezien van een klein aantal defektieven dat in de steekproeven van de goedgekeurde partijen wordt gevonden. Dit gemiddeld doorgelaten uitvalpercentage, gegeven door het gearceerde oppervlak in Fig. H, gaat als functie van p door een maximum: het maximaal gemiddeld-doorgelaten uitvalpercentage, \bar{p}_m .



Kromme moet gladder verlopen. Er zit een ongewenste knik in het midden.

Fig. 55. Illustratie bij het begrip maximaal gemiddeld-doorgelaten uitvalpercentage.

\bar{p}_m geeft een conservatieve grens, omdat men aanneemt dat alle partijen een zelfde uitvalpercentage p bevatten, en dat deze p de meest ongunstige waarde aanneemt.

Practische aspecten

Een keuringskarakteristiek ligt praktisch voldoende vast indien voor twee van de 4 parameters:

$$p_{95}, p_{50}, p_{10}, \bar{p}_m$$

een numerieke waarde wordt voorgeschreven.

Vervolgens kan een steekproefschema worden geconstrueerd, dat aan deze numerieke eisen voldoet.

In de praktijk werkt men echter bij voorkeur met standaard tabellen, waaruit het te gebruiken steekproefschema direct kan worden afgelezen als functie van a) de grootte van de partij, en b) één van de boven besproken parameters. Compendium Tabel 8.2 geeft een dergelijke tabel bij Philips in gebruik en gebaseerd op partijgrootte en p_{50} . Voor partijen van 1000 of minder producten worden enkelvoudige, voor grotere partijen dubbele steekproefschema's voorgeschreven.

Een volledig stel keuringskarakteristieken is toegevoegd, en kan geraadpleegd worden bij het bepalen van onze keuze. Een hulptabel op pag. 67 geeft tevens de bij ieder schema behorende waarden van p_{95} en p_{10} .

Een andere zeer veel gebruikte Tabel is de

Military Standard 105 D
U.S. Government Printing Office
Washington D.C. 10402, \$0.40.

Deze tabel geeft steekproefschema's als functie van partijgrootte en AQL. Zij bevat bovendien equivalente tabellen met enkelvoudige, dubbele, en meervoudige schema's. De Mil.Std. 105 D is ontworpen in onderlinge samenwerking door de ministeries van oorlog van Amerika, Canada, en Engeland, ten behoeve van de keuring van legergoederen. Deze standaard wordt echter ook uitgebreid toegepast in de industrie voor andere keuringen. Deze en andere tabellen gaan meestal begeleid door uitvoerige instructies voor hun gebruik. Men moet er zich echter voor wachten hieraan een te absolute waarde toe te kennen, want de praktische toepassing van steekproefkeuringen berust steeds op het vinden van een compromis tussen wensen en mogelijkheden. Daarbij spelen naast statistische ook vele technische overwegingen een rol. Men moet bijv. rekening houden

- a) met de mate waarin slechte partijen voorkomen;
- b) met de kosten van de keuring;
- c) met het aantal man-uren dat voor keuringswerk beschikbaar is;
- d) met de wijze van behandeling van afgekeurde partijen;
- e) met de aard en ernst van de fouten die men in de defekte producten aantreft; en
- f) met de omvang van aanwezige voorraden.

Een steekproefkeuring laat zich in de praktijk dan ook niet volgens strakke wetten voorschrijven; soepelheid moet steeds worden betracht.

Voor details verwijzen we naar leerboeken op dit gebied; bijv.

A.H. Schaafsma en F.G. Willemze,
Modern Kwaliteitsbeleid
Philips Technische Bibliotheek.

De χ^2 - verdeling en haar toepassingen

Is u_i ; $i = 1, \dots, v$

een reeks van v standaard normale variabelen dan is per definitie

$$\chi_v^2 = \sum_{i=1}^v u_i^2. \quad (57.1)$$

χ_v^2 = de som der kwadraten van v onafhankelijke standaard-normale variabelen.
De parameter v = het aantal vrijheidsgraden.

De verdeling van χ_v^2 is

$$f(\chi_v^2) d\chi_v^2 = A(\chi_v^2)^{\frac{1}{2}v-1} \exp \left\{ -\frac{\chi_v^2}{2} \right\} d\chi_v^2, \quad 0 \leq \chi_v^2 < \infty. \quad (57.2)$$

Compendium Tabel 3.1 geeft waarden van χ^2 met voorgeschreven overschrijdingskansen, bijv.

$$P(\chi_5^2 < 9.24) = 0.90,$$

$$P(\chi_5^2 > 1.61) = 1 - 0.10 = 0.90,$$

$$P(1.15 < \chi_5^2 < 11.1) = 0.90.$$

Tabel 3.1 speelt in de toegepaste statistiek een zeer belangrijke rol.

Uit de definitie volgt

$$E(\chi_v^2) = \sum_i E(u_i^2) = v, \quad (57.3)$$

$$\text{var}(\chi_v^2) = \sum_i \text{var}(u_i^2) = 2v,$$

$$\text{daar } \text{var } u^2 = E(u^4) - \{E(u^2)\}^2 = 3 - 1 = 2.$$

Voor grote waarden van v is de χ^2 -verdeling praktisch normaal. Een nog betere benadering is

$$\chi_v \cong \sqrt{v} + u \sqrt{\frac{1}{2}}. \quad (57.4)$$

Hieruit volgt bijv. voor $v = 30$

$$P \{16.8 < \chi_{30}^2 < 47.3\} = 0.95$$

terwijl tabel 3.1 als grenzen geeft 16.8 en 47.0.

Uit de definitie (57.1) volgt verder

$$\chi_{v_1}^2 + \chi_{v_2}^2 + \dots + \chi_{v_k}^2 = \chi_{v_1+v_2+\dots+v_k}^2 \quad (58.1)$$

De som van een aantal onafhankelijke χ^2 -variabelen is weer een χ^2 -variabele; met een aantal vrijheidsgraden gelijk aan de som van de vrijheidsgraden der componenten.

De omgekeerde stelling luidt: Bestaat tussen χ^2 -variabelen de betrekking (58.1) dan zijn $\chi_{v_1}^2, \chi_{v_2}^2, \dots, \chi_{v_k}^2$ onderling onafhankelijk.

Het bewijs valt buiten het bestek van dit college. De stelling speelt in de z.g. variantie analyse een belangrijke rol.

De verdeling van s^2 .

$$\text{Is } x_i \cong \mu + u_i \sigma, \quad i = 1, \dots, n$$

een aselechte steekproef uit een normale verdeling dan is

$$\Sigma \left\{ \frac{x_i - \mu}{\sigma} \right\}^2 \cong \chi_{n-1}^2 \quad (58.2)$$

Vervangen we μ door \bar{x} dan kan worden bewezen dat

$$\sum_i \frac{(x_i - \bar{x})^2}{\sigma^2} = \frac{KS}{\sigma^2} \cong \frac{v s^2}{\sigma^2} = \chi_v^2, \quad v = n-1. \quad (58.3)$$

Hiermee is de verdeling van $\frac{s^2}{\sigma^2}$ en dus ook van s/σ gegeven. Zo is bijv. voor $v = 8$. (Tabel 3.1)

$$P(2.18 < \frac{8s^2}{\sigma^2} < 17.5) = 0.95$$

en dus ook

$$P\left(\underline{s} \sqrt{\frac{8}{17.5}} < \sigma < \underline{s} \sqrt{\frac{8}{2.18}}\right) = 0.95,$$

of

$$P(0.68\underline{s} < \sigma < 1.92\underline{s}) = 0.95.$$

Dit zijn de betrouwbaarheids grenzen voor σ gegeven in Tabel 3.2; deze tabel is dus rechtstreeks uit Tabel 3.1 afgeleid, volgens de formules

$$a_1 = \sqrt{\frac{v}{\chi_v^2(\alpha_T)}}, \quad a_2 = \sqrt{\frac{v}{\chi_v^2(\alpha_L)}} \quad (58.4)$$

De benaderingsformules

$$\left. \begin{array}{l} a_1 \\ a_2 \end{array} \right\} = \frac{\sqrt{2v}}{\sqrt{2v + u_\alpha}} \quad (59.1)$$

volgen uit de benadering (57.4).

Uit de formule (58.1) volgt verder dat een s^2 volgens (30.1) verkregen door samenvoegen van een aantal varianties, wederom een χ^2 -verdeling bezit.

Het vergelijken van fracties of percentages

Veelvuldig komt het voor dat we een uit een aantal steekproeven fracties of percentages $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_v$ hebben gevonden en dat we één van de hypothesen

$$H_0: p_1 = p_2 = \dots = p_k = p_0^*, \quad (p_0^* \text{ gegeven}) \quad (59.2)$$

$$\text{of } H_0: p_1 = p_2 = \dots = p_k, \quad (59.3)$$

$$\text{of } H_0: p_i = p_i^*, \quad (p_i^* \text{ gegeven}) \quad (59.4)$$

willen toetsen. Dit geschiedt dan doorgaans met behulp van de χ^2 -toets.

De volgende voorbeelden lichten dit toe.

A. Een tollende gulden valt vaker met "munt" naar boven dan met "kruis". Tabel 59 geeft het resultaat van een experiment met 4 verschillende guldens.

Tabel 59. Tollen met 4 guldens; toetsen van de hypothese; $p_i = p_0 = \frac{1}{2}$.

Muntstuk	Aantal worpen n_i	Aantal "Munt" x_i	$E(\underline{x}_i) = n_i p_0$ $p_i = \frac{1}{2}$	$d_i =$ $x_i - n_i p_i$	u_i
A	30	20	15	+ 5	+ 1.82
B	40	19	20	- 1	- 0.32
C	40	28	20	+ 8	+ 2.53
D	50	32	25	+ 7	+ 1.98
Totaal	160	99	80	+ 19	

De waarnemingen \underline{x}_i zullen een binomiale verdeling bezitten en op grond van de normale benadering zullen de grootheden

$$u_i = \frac{x_i - n_i p_i}{\sqrt{n_i p_i (1 - p_i)}}$$

een onafhankelijke steekproef uit een standaard normale verdeling vormen. De waarden u_i corresponderend met de nulhypothese

$$H_0: p_i = \frac{1}{2} \quad (59.5)$$

zijn gegeven in de laatste kolom. Hiermee vinden we

$$\chi_4^2 = (1.82)^2 + (0.32)^2 + (2.53)^2 + (1.98)^2 = 13.75, \quad \alpha_T = 1\%.$$

Er is dus goede redenen aan de juistheid van de hypothese (59.5) te twijfelen.

We kunnen deze hypothese doelmatig in twee delen splitsen:

$$H'_0: p_i = p_0, \quad (60.1)$$

en

$$H''_0: p_0 = \frac{1}{2}. \quad (60.2)$$

H'_0 kunnen we apart toetsen door de berekeningen van Tabel 59 uit te voeren met de schatting

$$\hat{p}_0 = \frac{99}{160} = 0.62 \quad (60.3)$$

die uit de totalen in Tabel 59 volgt.
Dit geeft Tabel 60.

Tabel 60. Tollen met 4 munten.
Toetsen van de hypothese: $p_i = p_0$, $i = 1, \dots, 4$; p_0 onbekend.

Muntstuk	Aantal worpen n_i	Aantal "Munt" x_i	$E(x_i) = n_i \hat{p}_0$ $\hat{p}_0 = 0.62$	$d_i =$ $x_i - n_i \hat{p}_0$	u_i
A	30	20	18.6	+ 1.4	+ 0.53
B	40	19	24.8	- 5.8	- 1.89
C	40	28	24.8	+ 3.2	+ 1.04
D	50	32	31.0	+ 1.0	+ 0.30
Totaal	160	99	99.2	- 0.2	

Het verschil met tabel 59 is dat nu, op afrondingsfouten na, de som der verwachtingen gelijk is aan de som der waarnemingen; d.w.z.

$$\sum n_i \hat{p}_0 = \sum x_i \quad \text{of} \quad \sum d_i = 0.$$

Door deze aanpassing wordt χ^2 kleiner, maar dit kan correct in rekening worden gevraagd door het aantal vrijheidsgraden met 1 te verminderen. Dus nu vinden we

$$\chi_3^2 = (0.53)^2 + (1.89)^2 + (1.04)^2 + (0.30)^2 = 5.02; \quad \alpha_T \approx 20\%.$$

Er is dus geen redenen aan hypothese (60.1) te twijfelen.

De hypothese (60.2) kunnen we nu afzonderlijk toetsen op grond van de totalen. Dit geeft volgens

$$u_i = \frac{99 - 80}{\sqrt{160 \times \frac{1}{2} \times \frac{1}{2}}} = +3.01, \quad \alpha_T = 0.13\%$$

of

$$\chi_1^2 = (3.01)^2 = 9.1, \quad \alpha_T = 0.26\%$$

De hypothese $p_0 = \frac{1}{2}$ is dus moeilijk te handhaven.

Dit voorbeeld toont verder dat een u-toets gelijkwaardig is met een χ_1^2 -toets.

Alleen gaat bij χ_1^2 het teken verloren en de éézijdige overschrijdingskans bij χ_1^2 is gelijk aan de tweezijdige overschrijdingskans bij de u-toets.

B. Een enigszins andere situatie doet zich voor in Tabel 61.

Tabel 61

120 worpen met één dobbelsteen.

Toetsen van de hypothese $p_i = 1/6; i = 1, \dots, 6$.

Aantal ogen	Frekwentie x_i	$E(\underline{x}_i) = np_i$ $p_i = 1/6$	d_i	$\Delta \chi^2 = \frac{d_i^2}{np_i}$
"1"	15	20	-5	1.25
"2"	26	"	+6	1.80
"3"	21	"	+1	0.05
"4"	10	"	-10	5.00
"5"	27	"	+7	2.45
"6"	21	"	+1	0.05
Totaal	120	120	0	10.60 χ_5^2

Het verschil met het voorgaande geval is dat nu de som

$$\sum x_i = n = 120$$

van te voren vast ligt. De waarnemingen \underline{x}_i zijn niet onafhankelijk; zij bezitten samen een multinomiale verdeling. Dit brengt twee correcties met zich mee:

- 1) het aantal vrijheidsgraden moet met 1 worden verminderd.
- 2) de bijdrage van iedere rij in tabel 61 tot χ^2 moet worden berekend volgens

$$\Delta \chi^2 = \frac{(x_i - np_i)^2}{np_i}; \quad (61.1)$$

de factor $(1-p_i)$ in de noemer komt te vervallen. Het bewijs kan hier niet worden gegeven.

Tabel 61 geeft aldus

$$\chi^2 = 10.60, \quad \alpha_r \approx 3\%.$$

De juistheid van de hypothese $p_i = 1/6$ is dus aan twijfel onderhevig; er is een aanwijzing dat de dobbelsteen niet zuiver is.

Uit Tabel 61 zien we dat vooral de frekwentie van het ogetal "4" te laag is.

In hoeverre we inderdaad aanvaarden dat deze dobbelsteen vals is zal er van afhangen of we al of niet reden hebben valsheid te vermoeden.

C. Formules.

Het is nuttig de formules die voor de berekening van χ^2 werden toegepast duidelijk te onderscheiden.

$$\text{Tabel 59:} \quad \chi^2_k = \sum_{i=1}^k \frac{(x_i - n_i p_0)^2}{n_i p_0 (1-p_0)}, \quad p_0 \text{ gegeven} \quad (62.1)$$

$$\text{Tabel 60:} \quad \chi^2_{k-1} = \sum_{i=1}^k \frac{(x_i - n_i \hat{p}_0)^2}{n_i \hat{p}_0 (1-\hat{p}_0)}, \quad \hat{p}_0 \text{ geschat} \quad (62.2)$$

$$\text{Tabel 61:} \quad \chi^2_{k-1} = \sum_{i=1}^k \frac{(x_i - n p_i)^2}{n p_i}, \quad p_i \text{ gegeven, } \sum p_i = 1. \quad (62.3)$$

De laatste formule is de meest gangbare die in alle leerboeken voorkomt. Ook de Tabellen 59 en 60 kunnen met deze formule behandeld worden mits we naast de waarneming x_i voor "Munt" ook de waarneming $(n_i - x_i)$ voor "Kruis" in rekening brengen. Tabel 63 licht dit nader toe voor het geval van Tabel 60.

Tabel 63 Berekening van χ^2 voor Tabel 60 met behulp van formule.

Munt- stuk	Aantal worpen n_i	"Munt"		"Kruis"		$\Delta\chi^2$	
		Waarn. x_{1i}	Verwachting $n_i\hat{p}_1$	Waarn. x_{2i}	Verwachting $n_i\hat{p}_2$	"Munt"	"Kruis"
A	30	20	18.6	10	11.4	0.11	0.17
B	40	19	24.8	21	15.2	1.35	2.21
C	40	28	24.8	12	15.2	0.41	0.67
D	50	32	31.0	18	19.0	0.03	0.05
						$\chi^2 = 5.00$	

χ^2 is nu opgebouwd uit 8 bijdragen.

De waarnemingen x_{1i} en x_{2i} moeten echter aan 4 voorwaarden voldoen:

$$x_{1i} + x_{2i} = n_i, \quad i = 1, \dots, 4.$$

Daarmee gaan 4 vrijheidsgraden verloren.

De 5de vrijheidsgraad verliezen we omdat $\hat{p}_1 = 0.62$ en $\hat{p}_2 = 1 - \hat{p}_1 = 0.38$ uit de totalen werden geschat.

De Tabel 63 is een speciaal voorbeeld van een algemener geval, de z.g. $m \times n$ verwandschapstabel, waarvan Tabel 64 een voorbeeld geeft. Deze Tabel bevat de aantallen E, N, W en T-studenten die bij een wiskunde tentamen in 1961 resp. niet deelnamen, een voldoende, en een onvoldoende (cijfer behaalden).

De vraag is nu, is het gedragspatroon wat de E, N, W en T studenten betreft eender of zijn er significante verschillen.

Om dit na te gaan schatten we uit de rand totalen de kansen $\hat{p}_{i.}$ en $\hat{p}_{.j}$. Is het gedragspatroon onafhankelijk van de studierichting dan is de verwachting in de cel i, j ,

$$E(x_{ij}) = n\hat{p}_{ij} = n\hat{p}_{i.} \cdot \hat{p}_{.j}. \quad (63.1)$$

Met deze formule zijn de tussen haakjes geplaatste verwachtingen berekend, en de bijdrage tot χ^2 gegeven door.

$$\Delta\chi^2 = \frac{(x_{ij} - n\hat{p}_{i.}\hat{p}_{.j})^2}{n\hat{p}_{i.}\hat{p}_{.j}}. \quad (63.2)$$

Tabel 64.

Gedragpatroon van E, N, W, en T-studenten bij een wiskunde tentamen.

Studie- richting		Niet deel- genomen	Voldoende	Onvoldoende	Totaal	\hat{p}_i
E	x_{1j}	51	81	51	183	0.327
	$E(\underline{x}_{1j})$	(51.0)	(80.1)	(52.0)	(183.1)	
	$\Delta\chi^2$	0.00	0.01	0.02		
N	x_{2j}	31	45	17	93	0.166
	$E(\underline{x}_{2j})$	(25.9)	(40.7)	(26.4)	(93.0)	
	$\Delta\chi^2$	1.00	0.45	3.35		
T	x_{3j}	25	61	39	125	0.223
	$E(\underline{x}_{3j})$	(34.9)	(54.8)	(35.5)	(125.2)	
	$\Delta\chi^2$	2.80	0.70	0.35		
W	x_{4j}	49	58	52	159	0.284
	$E(\underline{x}_{4j})$	(44.4)	(69.6)	(45.1)	(159.1)	
	$\Delta\chi^2$	0.48	1.93	1.38		
Totaal	Totaal	156	245	159	560	
\hat{p}_j	\hat{p}_j	0.279	0.438	0.284		

De verwachtingen zijn nu zo aangepast dat de marginale sommen $x_{i.}$ en $x_{.j}$ gelijk zijn aan hun verwachtingen $E(\underline{x}_{i.})$ en $E(\underline{x}_{.j})$, en dat ook de som $x_{..}$ gelijk is aan $E(\underline{x}_{..})$. Bevat de tabel m rijen en n kolommen, dan zijn de verwachtingen voor $m+n+1$ rand totalen aan de waarnemingen aangepast. Daar echter steeds geldt

$$\sum x_{i.} = \sum x_{.j} = x_{..}$$

kunnen van deze $m+n+1$ waarden er slechts $m+n-1$ worden aangepast, de resterende 2 zijn dan vanzelf goed. De aanpassing omvat derhalve $m+n-1$ vrijheidsgraden en er blijven voor

$$mn - m - n + 1 = (m-1)(n-1) \quad (64.1)$$

vrijheidsgraden over. In het geval van Tabel 64, $(4-1)(3-1) = 6$.

Tellen we de bijdragen aan χ^2 op dan resulteert

$$\chi^2_6 = 12.47 ; \quad \alpha_T = 5\%$$

Dit wijst erop dat enige twijfel aan de hypothese van onafhankelijkheid gerechtvaardigd is, maar de afwijkingen zijn niet overtuigend.

Wil men een eventuele afhankelijkheid verklaren, dan dient de aandacht te worden gericht op die cellen die de hoogste bijdragen leveren tot χ^2 . Het valt dan op dat bij de N-studenten relatief weinig onvoldoendes zijn gemaakt, en dat relatief weinig T-studenten niet aan het tentamen hebben deelgenomen.

E. De χ^2 -toets vindt ook toepassing wanneer we willen nagaan in hoeverre een waargenomen frekwentie verdeling met een bepaalde theoretische verdeling klopt.

Tabel 65. Toetsen van een normale verdeling met de χ^2 -toets. Gewichten in mgr. van pag. 3.

Klassen	Frekwentie x_i	u_i bij de klasse- grenzen	Φ_i	$p_i =$ $\Delta\Phi_i$	np_i
1080 - 89	1	- 2.19	0.0143	0.0143	1.1
1090 - 99	4 } 5	- 1.61	0.0537	0.0395	3.0 } 4.1
1100 - 09	7	- 1.03	0.1515	0.0978	7.3
1110 - 19	7	- 0.45	0.3264	0.1749	13.7
1120 - 29	11	0.12	0.5478	0.2214	16.6
1130 - 39	24	0.70	0.7580	0.2102	15.8
1140 - 49	15	1.27	0.8980	0.1400	10.5
1150 - 59	5	1.85	0.9678	0.0690	5.2
1160 - 69	1 } 6			0.0322	2.4 } 7.6
mgr.					
$\bar{x} = 1127.5$ mgr.		$s = 17.3$ mgr.			
$\chi^2_4 = \frac{(5-4.1)^2}{4.1} + \frac{(7-7.3)^2}{7.3} + \dots + \frac{(6-7.6)^2}{7.6} = 11.5$					$\alpha_T = 2.5\%$

Pag. 3 geeft een tabel met de gewichten van 75 hardpapieren schijfjes. De eerste twee kolommen van Tabel 65 geven een frekwentietabel op deze gegevens gebaseerd. Hieruit volgt

$$\bar{x} = 1127.5 \text{ mgr.}, \quad s = 17.3 \text{ mgr.}$$

Met deze waarden berekenen we de waarden u_i corresponderen met de bovengrenzen van de klassen; bijv. voor de 2e klasse:

$$u_i = \frac{1099.5 - 1127.5}{17.3} = -1.61$$

enz. Met deze waarden van u_i zoeken we vervolgens de bijbehorende waarden van de kumulative normale verdeling Φ_i in Tabel 1.1, en daaruit voor iedere klasse het verschil

$$\Delta\Phi_i = \Phi_i - \Phi_{i-1}$$

met $\Phi_0 = 0$, en $\Phi_n = 1$. Deze verschillen geven de kansen dat de waarnemingen uit de aangepaste normale verdeling in de betreffende klassen zullen vallen. Door vermenigvuldiging met het totaal aantal waarnemingen n verkrijgen we de verwachtingen np_i in de laatste kolom.

De χ^2 -toets is gebaseerd op de onderstellingen dat de binomiale verdeling van de waarnemingen x_i door normale verdelingen mogen worden benaderd. Bij de verwachtingen van 1.1 en 2.4 in de eerste en laatste klasse is dit niet toelaatbaar. Daarom worden de eerste twee en de laatste twee klassen paarsgewijze bijeengevoegd tot twee klassen met verwachtingen van 4.1 en 7.6 resp. Het totaal aantal klassen verminderd hierdoor van 9 tot 7.

Uit de waarnemingen x_i en de verwachtingen np_i volgt nu

$$\chi_4^2 = 11.5 ; \quad \alpha_T = 2.5\%$$

Deze χ^2 bezit slechts 4 vrijheidsgraden omdat de verwachtingswaarden in 3 opzichten aan de waarnemingen zijn aangepast, en wel voor n , \bar{x} , en s . Iedere aanpassing geeft een vermindering van χ^2 die door vermindering met één vrijheidsgraad moet worden gecompenseerd. Dit geldt algemeen.

De overschrijdingskans is vrij gering en de conclusie gerechtvaardigd dat de waargenomen frekwentie verdeling niet geheel bevredigend door een normale verdeling kan worden beschreven. Vooral in de klasse 1130 - 1139 mgr. is de waargenomen frekwentie abnormaal hoog.

Grafische methoden

Tabel 68 geeft het aantal 1ste jaars studenten in Delft in de jaren 1945 - 1960 en de percentages W; B; E; en T-studenten onder hen.

We kunnen nu bv. de hypothese

$$H_0 : p_{W_i} = p_W = \text{konstant}, \quad (67.1)$$

dus de hypothese dat het percentage W-studenten in deze jaren niet varieerde, grafisch toetsen als in de Figuren 68 en 69 uitgevoerd. In Fig. 1 zijn om het totale gemiddelde

$$\bar{p}_W = \sum n_i p_{W_i} / \sum n_i \quad (67.2)$$

95% intervallen geconstrueerd. Gemiddeld zal dan 1 op de 20 waargenomen percentages buiten deze grenzen vallen. In werkelijkheid vallen 4 van de 16 waarnemingen er buiten en dit is een duidelijke indicatie dat de hypothese (67.1) niet kan worden gehandhaafd. De kans dat 4 of meer punten buiten deze grenzen zullen vallen is ongeveer 0.5%.

In Fig. 69 is voor iedere waarneming afzonderlijk een 95% betrouwbaarheidsinterval getekend. De vraag is nu of de horizontale lijn $p = 20.9\%$ al deze of bijna al deze, intervallen snijdt. De lijn gaat niet door 4 van deze intervallen, corresponderend met dezelfde 4 punten die in Fig. 1 buiten de 95% grenzen vallen. Ook dit is een duidelijke aanwijzing dat de hypothese (67.1) niet deugt.

Men kan deze hypothese ook toetsen met χ^2 , dit geeft

$$\chi_{15}^2 = 40.0, \alpha_T < 0.1\%. \quad (67.3)$$

De χ^2 toets is een scherper gedefinieerd criterium, doch de grafische toets met behulp van Fig. 68 of Fig. 69 is voor de niet-statisticus begrijpelijker, en geeft bovendien een inzicht in de aard van de afwijkingen. Het percentage W-studenten gaat in de jaren '53 - '54 duidelijk door een maximum en neemt daarna vrij duidelijk af.

Uit één enkele waarde van χ^2 is dit niet te zien.

De praktische vraag is dan bv. of men voor een dergelijk verloop een verklaring kan vinden, of er maatregelen bestaan waarmee men de toeloop van W-studenten kan beïnvloeden, en, zo ja, hoe en wanneer men daarvan gebruik dient te maken.

Tabel 08. Aantal eerstejaars studenten in Delft van 1954 - 1960 en % W, B, E en N studenten onder hen.

Jaar	Aantal	P_W %	P_B %	P_E %	P_N %
1945	2355	20.8	6.6	15.4	4.2
46	1357	20.0	6.0	17.3	5.2
47	1092	21.8	4.2	17.6	5.1
48	980	21.9	6.6	21.6	5.9
49	908	21.3	6.5	21.3	6.5
50	870	22.6	8.2	16.6	7.2
51	660	22.9	6.7	18.6	8.2
52	642	24.6	8.1	13.9	7.8
53	665	23.0	7.1	14.0	8.0
54	773	22.5	7.2	14.4	9.7
55	947	19.8	8.9	18.6	8.7
56	1123	24.1	7.3	17.5	9.3
57	1050	19.2	9.5	14.1	11.8
58	1091	18.9	9.2	14.3	14.3
59	1051	17.1	9.2	15.1	15.9
60	1196	18.0	10.3	14.4	13.9
Totaal	16760	20.9%			

Het totale percentage werd bepaald met de formule $P_W = \frac{\sum n_i P_{W_i}}{\sum n_i}$; het is dus een gewogen gemiddelde

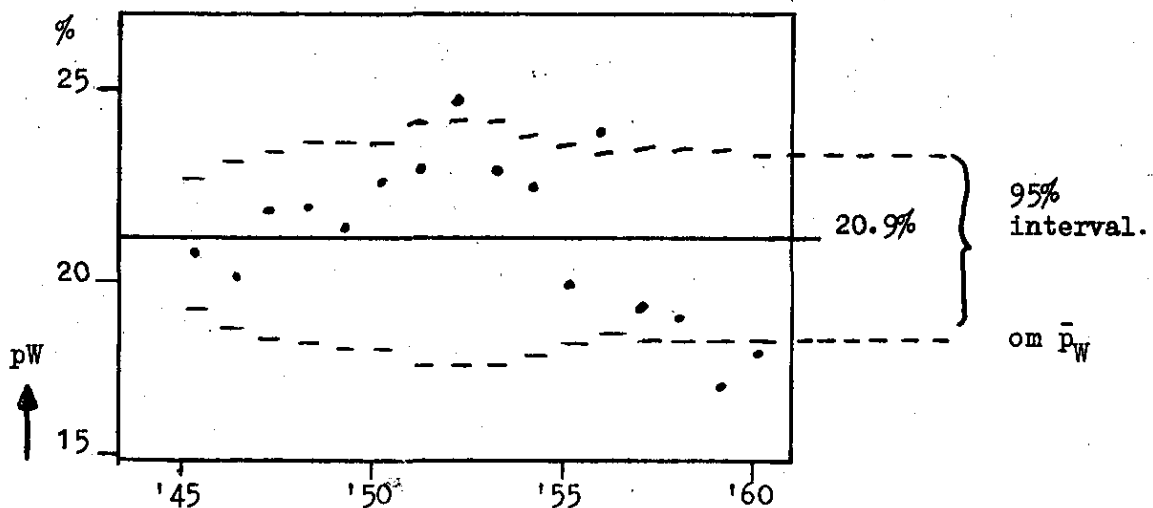


Fig. 68. Grafische analyse van de gegevens voor P_W uit Tabel 68. Toetsing van de hypothese $p_W = \text{constant} = 20.9\%$

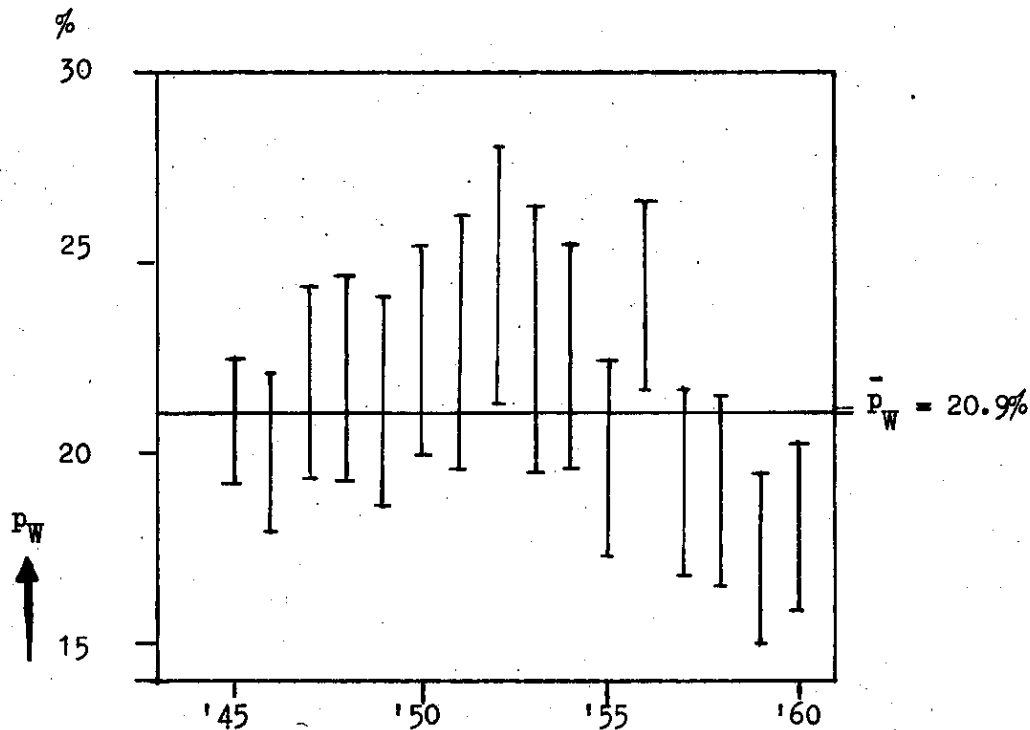


Fig. 69 Grafische analyse van de gegevens voor p_W van Tabel 68 2e methode; iedere waarneming p_W met een eigen betrouwbaarheidsinterval.

Kruisjeskaarten

Een soortgelijk probleem doet zich vaak voor bij de controle van een doorlopende massaproductie. Men neemt regelmatig, bv. om het uur of om het halve uur, een steekproef uit de productie en bepaald hierin het aantal defektieven, d.w.z. het aantal producten dat niet aan de gestelde kwaliteitseisen voldoet. Aan de hand van deze gegevens wenst men dan de hypothese

$$p = \text{konstant,} \quad (69.1)$$

te toetsen, waarbij p = het uitvalpercentage dat wordt geproduceerd. Zo gauw er aanwijzingen zijn dat de hypothese (69.) niet langer op gaat, en dat p is opgelopen bv. door slijtage van het gereedschap, zal men dan het proces stoppen en opnieuw bijstellen.

Het gaat hier steeds om kleine percentages en de theorie is daarom steeds gebaseerd op de Poisson verdeling. Als illustratie-materiaal

geeft Tabel 70 series van 30 waarnemingen uit Poisson verdelingen met $\mu = 0.4, 0.6, 0.8,$ en 1.0 respectievelijk.

Stel nu we nemen om het halve uur een steekproef van 20 stuks en het productieproces loopt met een uitvalpercentage van 2%. Dan geeft de serie A in Tabel 70 een beeld van de waarnemingen die men zal verkrijgen. In 30 steekproeven neemt men in totaal 16 defectieven waar, en hieruit volgt een geschat fouten-

Tabel 70.

Series van 30 waarnemingen uit 4 Poisson verdelingen.

$\mu = 0.4$ A	0 1 1 0 1 0 0 1 0 2 1 1 0 0 0 0 0 0 2 1 1 0 1 0 0 1 1 0 1 0 $x. = 16$
$\mu = 0.6$ B	3 1 1 0 1 1 1 1 1 0 0 1 0 1 0 1 0 2 1 2 0 2 1 1 1 0 0 0 0 2 $x. = 25$
$\mu = 0.8$ C	0 2 1 1 0 1 2 0 0 1 1 1 0 1 0 0 0 2 2 0 0 0 0 0 0 0 0 0 1 0 $x. = 16$
$\mu = 1.0$ D	0 2 2 0 0 1 2 0 0 1 1 0 1 0 2 1 2 3 0 2 1 1 1 1 0 1 1 0 2 0 $x. = 28$

percentage van

$$\frac{16}{600} = 0.027 = 2.7\%$$

en $\mu = 20 \times 0.027 = 0.54.$

Heeft het proces bevredigend gelopen dan zal men concluderen dat men met een dergelijk uitvalpercentage moet rekenen. Teneinde het proces regelmatig onder controle te houden zet men nu de waargenomen aantallen defectieven naast elkaar grafisch uit als in Fig. 71 is uitgevoerd. Dit heet algemeen een kruisjeskaart.

Hierin trekt men nu twee horizontale lijnen; een Waarschuwinglijn, en een Actielijn. In het geval van Fig. 71 ligt de Waarschuwinglijn tussen 1 en 2, de Actielijn tussen 2 en 3 defectieven. Komt een waarneming boven de Actielijn, dan is dit een duidelijke aanwijzing dat er iets

mis is; men stopt de productie en stelt bij. Komt een punt tussen de Waarschuwings- en de Actielijn terecht, dan is dit een waarschuwing dat men op moet letten. Men kan dan bv. voorschrijven dat direct nog een steekproef wordt genomen en de productie wordt gestopt en bijgesteld indien in deze 2de steekproef één of meer defectieven worden gevonden.

Men kan dan verder weer een werkkarakteristiek berekenen die de kans op "actie" weergeeft als functie van het uitvalpercentage.

Fig.71 geeft de karakteristiek behorend bij het zojuist beschreven voorschrift.

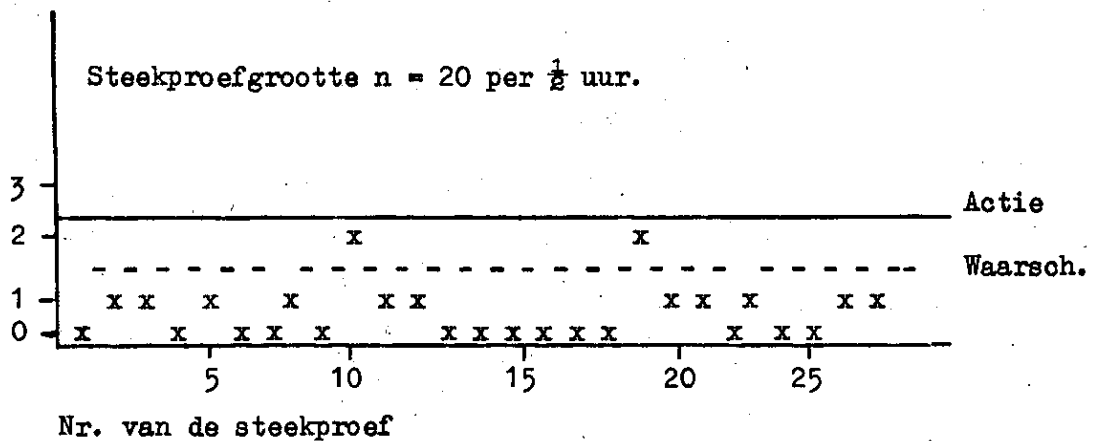


Fig.71. Principe van een "kruisjeskaart".

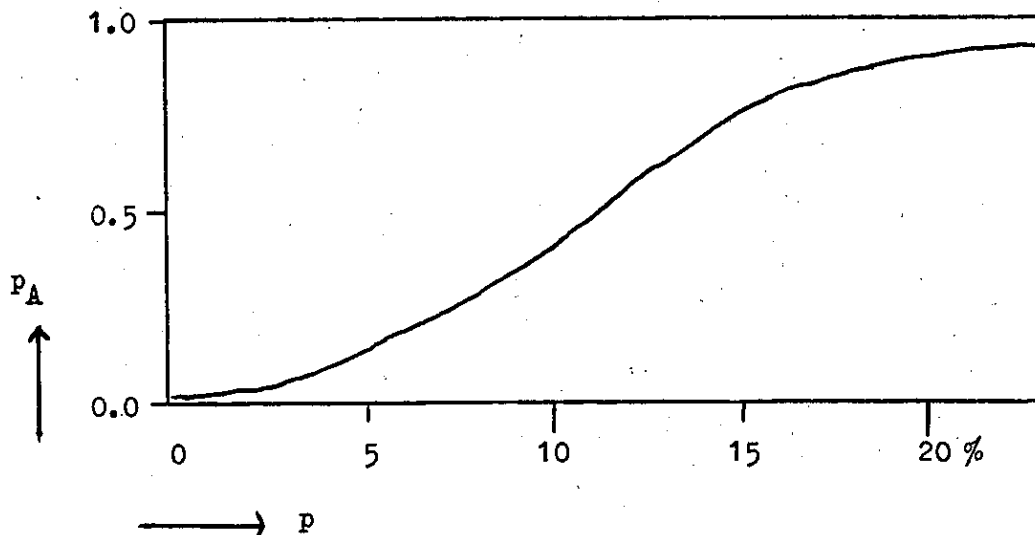


Fig.71a Werkkarakteristiek voor een kruisjeskaart.

$$n_1 = 20, n_2 = 20. \text{ "Actie"}$$

$$\text{wanneer } x_1 \geq 3, \text{ of } x_1 = 2, x_2 \geq 1.$$

Deze figuur laat duidelijk zien dat een dergelijk systeem niet snel reageert. Bij een $p = 10\%$ is de kans op "Actie" 0.40. Dit is echter niet zo erg omdat men een half uur later weer een steekproef neemt. Is de kans op "Actie" p_A dan zal men gemiddeld na

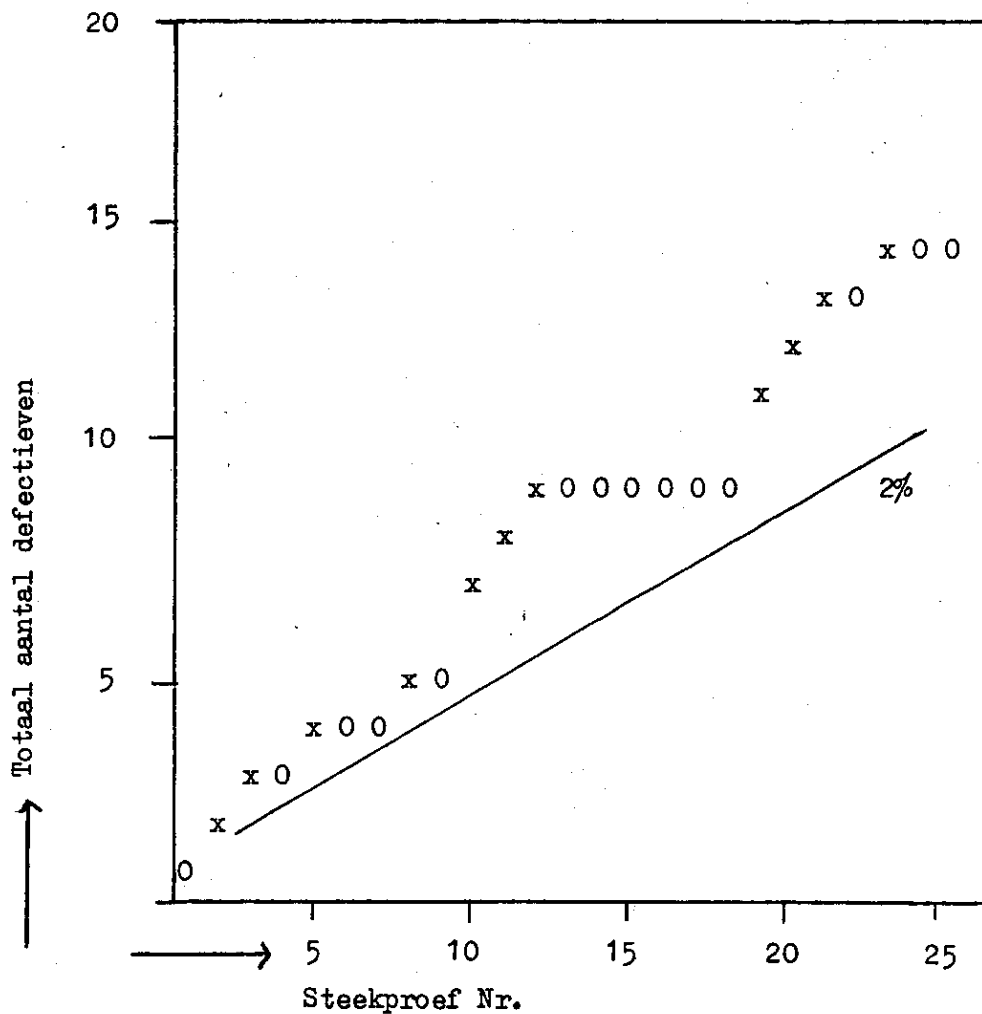


Fig. 72. Serie A van Tabel 70 cumulatief uitgezet.

$1/p_A$ steekproeven tot "Actie" overgaan. Een $p = 10\%$ komt dus gemiddeld na $2\frac{1}{2}$ steekproef, dus na 1 uur aan het licht.

Een kruisjeskaart reageert alleen snel op grove storingen. Wil men een nauwkeuriger controle dan zijn grotere steekproeven noodzakelijk. Daartoe kan men bv. de steekproeven die in de loop van de dag genomen zijn aan het eind van de dag bijeen voegen tot een dagresultaat, en

5-dag resultaten kunnen weer tot een week totaal worden samengevoegd. Dergelijke totaalcijfers zijn dan vooral voor de bedrijfsleiding van belang.

Verder wordt een kruisjeskaart doorgaans ook voorzien van systematische aantekeningen over de aard van de geconstateerde fouten, zodat men ook een inzicht krijgt in welke fouten het meest voorkomen en in welk opzicht dus naar verbetering kan worden gestreefd.

Een andere wijze om de gegevens van Tabel 70 in beeld te brengen is door ze z.g. cumulatief uit te zetten.

In Fig. 72 is dit uitgevoerd, wederom voor serie A. De helling waarmee de cumulatieve waarnemingen omhoog lopen geeft een maat voor het uitvalpercentage. Uit de Figuur leest men direct of dat in 25 steekproeven 14 defectieven zijn waargenomen.

Bij een steekproefgrootte van 20 stuks correspondeert de getrokken lijn met 2% uitval, en men kan een dergelijke lijn bv. in de Figuur aangeven als een aansporing om te trachten beneden deze lijn te blijven.

De \bar{x} - en R-kaarten

Kruisjeskaarten zijn van toepassing wanneer we alleen een attributenkeuring uitvoeren. Bij een variabelenkeuring, dus metingen, bereikt men het zelfde doel met z.g. \bar{x} - en R-kaarten. Als voorbeeld nemen we een partij Uranium oxyd tabletten, geleverd in 120 dozen van ieder 75 stuks, dus 9000 in totaal aan het Reactor Centrum in Petten.

Van 5 stuks uit iedere doos werd het gewicht bepaald. De metingen voor de eerste 3 dozen geven de uitkomsten:

Doos 1	8191	8292	8261	8277	8220	mgr.,
Doos 2	8015	7976	7990	7960	7978	mgr.,
Doos 3	8051	8081	8024	8043	7939	mgr..

We vinden hieruit $\bar{R} = 99$ mgr., $a \leq 20$ mgr.

De waarnemingen kunnen dus zonder bezwaar op 10 mgr. worden afgerond. Ze zijn daarna gecoduceerd volgens

$$x = \left(\frac{x-7800}{10} \right) \times 0.01 \text{ gr.}$$

De aldus gecodeerde waarnemingen voor de eerste 15 dozen zijn weer-gegeven in Tabel 74. Voor iedere doos werd gemiddelde, en de spreidingsbreedte berekend, en daaruit het totaal gemiddelde $\bar{\bar{x}}$, en \bar{R} .

We gaan nu uit van de veronderstelling dat we met 15 steekproeven uit normale verdelingen te doen hebben. \bar{x} - en R-kaarten dienen dan in wezen voor een grafische toets van de hypothesen

$$H_0: \mu_i = \mu., \quad (74.1)$$

$$\text{en } H_0: \sigma_i = \sigma, \quad (74.2)$$

dus dat alle dozen eenzelfde gemiddelde, of eenzelfde standaardafwijking hebben.

Uit $\bar{R} = 9.1$ volgt $s = 3.9$ en daar deze waarde uit vele waarnemingen is afgeleid mogen we in eerste benadering ook $\sigma = 3.9$ aannemen. Op grond van de verdeling van \bar{R} , die dan kan worden berekend kunnen we grenzen aangeven waarbinnen \bar{R} met voorgeschreven kans $(1-\alpha)$ zal liggen. Voor α zijn de waarden 5% en 0.1% gebruikelijk. De grenzen kunnen dan direct uit \bar{R} worden berekend met de coëfficiënten in Compendium Tabel 8.3 gegeven.

Tabel 74. De gecodeerde gewichten voor de eerste 15 dozen.

Doos							R
1	39	49	46	48	42	44.8	9
2	22	18	19	16	18	18.6	6
3	25	28	22	24	14	23.6	14
4	23	7	26	25	27	21.6	20
5	16	15	24	17	17	17.8	9
6	22	20	23	19	20	20.8	4
7	25	26	25	26	19	24.2	7
8	24	19	21	28	22	22.8	9
9	21	24	22	24	24	23.0	3
10	18	21	25	26	26	23.2	8
11	20	19	24	23	18	20.8	6
12	17	23	17	17	18	18.4	6
13	17	20	10	23	10	16.0	13
14	14	17	9	13	14	13.4	8
15	13	8	6	-1	8	6.8	14
						$\bar{\bar{x}} = 21.05$	$\bar{R} = 9.1$

$$x = (x-7800)/10 \text{ in } 0.01 \text{ gr.}$$

Deze z.g. controlegrenzen voor R zijn

$$\begin{aligned} \text{voor } \alpha = 5\% : \quad R_1 &= 0.37 * 9.1 = 3.4 \\ R_2 &= 1.81 * 9.1 = 16.5 \end{aligned} \quad (75.1)$$

$$\begin{aligned} \text{voor } \alpha = 0.1\% : \quad R_1 &= 0.16 * 9.1 = 1.5 \\ R_2 &= 2.36 * 9.1 = 21.5 \end{aligned} \quad (75.2)$$

De 5% controlegrenzen zijn in Fig. 75 getekend tezamen met de 15 waarden van R.

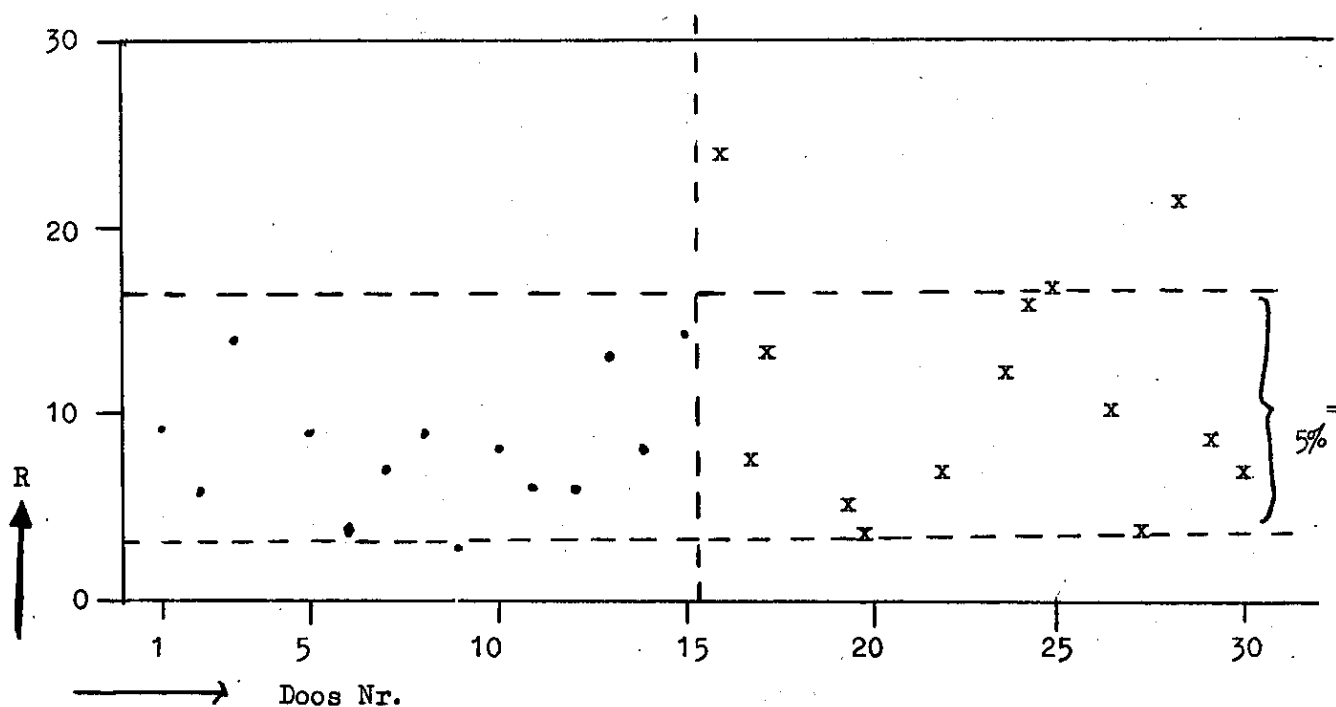


Fig. 75. Een R-kaart gebaseerd op de gegevens van Tabel 74. De kruisjes geven de R-waarden voor nog 15 dozen.

De waarde $R = 20$ voor doos nr. 4 ligt vrij ver buiten de 5% grenzen en doet vermoeden dat de waarneming $x = 7$ in deze doos een uitschieter is. Over het geheel liggen de R-waarden redelijk binnen de grenzen en wat betreft schijnt de partij vrij goed onder controle.

We kunnen de 5% grenzen nu doortrekken en ook voor de verdere dozen de R-waarden uitzetten. De kruisjes in Fig. 75 geven de waarden voor de dozen 16 - 30. Ook hier zijn bij de dozen nr. 16 en 28 vrij sterke

afwijkingen zichtbaar. Deze bleken evenals bij doos nr. 4 aan een enkele abnormaal lage uitkomst te wijten. Dit maakt aannemelijk dat we hier met een specifieke storing te doen hebben, bv. doordat soms een schilfer van de brossen tabletten is afgesprongen.

Naast de R-kaarten, die dienen om te toetsen in hoeverre de standaardafwijking van een proces beheerst of onder controle is, d.w.z. in hoeverre de hypothese (74.2) kan worden gehandhaafd, kennen we \bar{x} -kaarten die dienen als controle op de hypothese (74.1).

De grenzen voor \bar{x} vinden we eveneens met coëfficiënten gegeven in Tabel 8.3. Voor de gegevens van Tabel 74 zijn deze grenzen:

$$\alpha = 5\%: \quad \bar{\bar{x}} \pm A_{0.025} * \bar{R} = 21.0 \pm 0.38 \quad 9.1 = 17.6 \quad 24.5$$

$$\alpha = 0.1\%: \quad \bar{\bar{x}} \pm A_{0.001} * \bar{R} = 21.0 \pm 0.59 \quad 9.1 = 15.7 \quad 26.4$$

In Fig. 77a zijn de 15 gemiddelden met de 5% grenzen getekend. Het eerste gemiddelde \bar{x}_1 , valt ver buiten de grenzen en de laatste 3 gemiddelden eveneens en in toenemende mate.

Laten we \bar{x}_1 en \bar{x}_{15} buiten beschouwing dan daalt het totaal gemiddelde tot $\bar{\bar{x}} = 20.3$ en we krijgen nu Fig. 77.b.

Wel is waar liggen de laatste 2 gemiddelden nu nog buiten de grenzen, doch de afwijkingen zijn niet zo ernstig als in Fig. 77.a.

Een verklaring voor de afwijkingen zou kunnen zijn dat men bij doos 1 de machine waarop de tabletten op de juiste maat werden geslepen nog net goed had ingesteld, en dat bij doos no. 12 of 13 een verloop in het procédé begint op te treden en bijstellen dus gewenst is.

Fig. 77b geeft dan een redelijker beeld van wat het proces onder goede controle kan presteren.

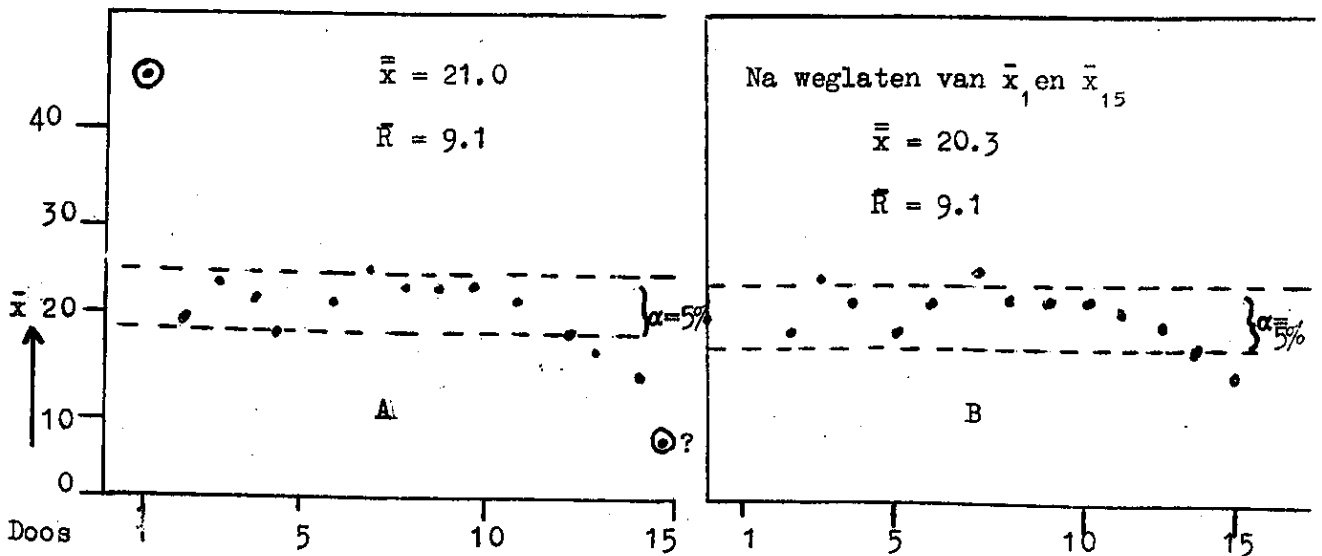


Fig. 77.a. Controle kaarten voor \bar{x} van Tabel 74:
A met alle 15 waarden; B met weglating van $\bar{x}_1 = 44.8$ en $\bar{x}_{15} = 6.8$

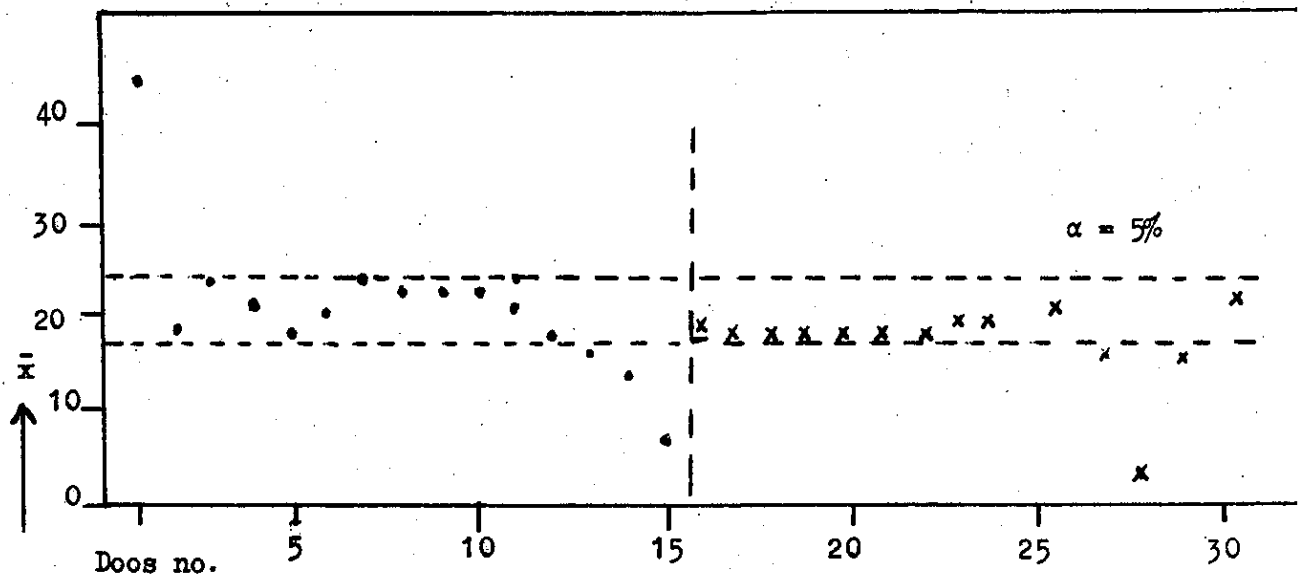


Fig. 78. Controlekaart voor \bar{x} voor Tabel 74 en de gemiddelden voor nog 15 dozen. De grenzen zijn bepaald op grond van Tabel 74 alleen.

Of men daarbij \bar{x}_{14} wel of niet mee wil rekenen is een vraag die men op grond van technische overwegingen dient te beslissen. Dat hangt er van af of men het onderste uit de kan wil halen of niet. Indien men te hoge eisen stelt leidt dat in de praktijk tot moeilijkheden.

In Fig. 78 zijn de grenzen uit Fig. 77.b. doorgetrokken en de gemiddelden van nog 15 dozen erbij uitgezet. Vanaf doos 16 is het proces weer goed beheerst, maar bij doos 28 treedt opnieuw een duidelijke storing op.

Inplaats van \bar{x} kan men ook een controle kaart construeren op grond van de mediaan M of op grond van de individuele waarnemingen x_{ij} . In dit laatste geval zet men de 5 waarnemingen in iedere rij van Tabel 74 samen op een verticale lijn uit. Voor M of voor de individuele waarnemingen kan men dan weer grenzen construeren met een overschrijdingskans van 5% of 0.1%. De benodigde formules en coëfficiënten zijn eveneens in Tabel 8.3 opgenomen. Voor de Mediaan geldt in goede benadering dat $\sigma_M = 1.25 \frac{\sigma}{\sqrt{5}}$.

De mediaan is dus minder nauwkeurig dan het gemiddelde doch heeft het voordeel dat zij geen rekenwerk vereist.

Kruisjes kaarten en \bar{x} -, M -, R -kaarten worden veel toegepast voor de regelmatige controle van een lopend productieproces. De kaarten worden daarbij zichtbaar bij de machine opgehangen. Daarbij wordt dan ook op de kaart aangetekend welke wijzigingen in het proces zijn aangebracht: wanneer een nieuwe partij grondstof in gebruik is genomen, wanneer de machine is bijgeregeld, enz. Uit het verloop dat op kaart grafisch in beeld wordt gebracht kan men dan zien of en in hoeverre deze wijzigingen van invloed zijn geweest.

Variantie analyse

In Tabel 74 vonden we $\bar{R} = 9.1$ en daaruit

$$s_x = s_b = 0.43 * 9.1 = 3.9 \quad (79.1)$$

Dit is een schatting van de standaardafwijking binnen dozen. Laten we in de gemiddelden in Tabel \bar{x} en \bar{x}_i buiten beschouwing en dan vinden we uit 3 groepen van 4 gemiddeld ($\bar{x}_2^4, \bar{x}_5^4, \bar{x}_6^4, \bar{x}_9^4, \bar{x}_{10}^4, \bar{x}_{13}^4$)

$R_{\bar{x}} = 5.8, 3.4, 7.2$ waaruit $R_{\bar{x}} = 5.5$ en

$$s_{\bar{x}} = 0.49 * R_{\bar{x}} = 2.7. \quad (79.2)$$

Op grond van de spreiding binnen dozen verwachten we

$$s_{\bar{x}}^i = \frac{s_b}{\sqrt{5}} = 1.7. \quad (79.3)$$

Dus de standaardafwijking (79.2) uit de gemiddelden zelf bepaald is hoger dan we uit de binnen dozen variaties kunnen verklaren. Het vermoeden is daarom gerechtvaardigd dat systematische verschillen optreden tussen de doosgemiddelden. We komen zo tot het model:

$$x_{ij} = \mu + a_i + e_{ij}, \quad (79.4)$$

waarin a_i de variaties van doos tot doos weergeven, en e_{ij} de variaties binnen de dozen.

We kunnen daarbij zonder bezwaar

$$E(a_i) = 0, E(e_{ij}) = 0 \quad (79.5)$$

stellen en definiëren verder

$$E(a_i^2) = \sigma_d^2 = \text{de tussendozenvariatie}, \quad (79.6)$$

$$E(e_{ij}^2) = \sigma_b^2 = \text{de binnendozenvariatie},$$

terwijl we tevens onderstellen dat a_i en de e_{ij} 's alle onderling onafhankelijk zijn.

Een elementaire beschouwing leert dan dat

$$\sigma_{\bar{x}}^2 = \sigma_d^2 + \frac{\sigma_b^2}{n} \quad (79.7)$$

wanneer \bar{x}_i het gemiddelde is van n elementen per doos. De waarde (79.3), $s_{\bar{x}}^i = 1.7$, is een schatting van σ_b / \sqrt{n} en de waarde (79.2), $s_{\bar{x}} = 2.7$ een schatting van $\sigma_{\bar{x}}$ als gegeven door (79.7). Stellen we de schattingen gelijk aan de verwachtingen dan vinden we

$$\hat{\sigma}_b = s_b = 2.7,$$

$$\hat{\sigma}_d = \sqrt{2.7^2 - 1.7^2} = 2.1$$

We hebben daarmee de totale variantie die in Tabel 74 optreedt in twee variantie componenten, σ_b^2 en σ_d^2 , gesplitst.

Bij vele industriële toepassingen is een dergelijke splitsing uiterst belangrijk. In plaats van een component binnen- en tussen-dozen onderscheidt men dan bv. componenten binnen-dagen en tussen-dagen, binnen-machines en tussen-machines, enz. De variantie binnen-machines geeft een maat voor de nauwkeurigheid waarmee één machine produceert, de variantie tussen machines geeft een inzicht in de mate waarin verschillende machines op het zelfde gemiddelde niveau zijn afgesteld.

Evenzo geeft een variantie tussen-dagen een inzicht in de mate waarop het niveau van één machine van dag tot dag aan varianties onderhevig is.

De waarden (79.8) zijn slechts schattingen en dientengevolge aan fouten onderhevig.

Bij industriële toepassingen kan men doorgaans over zeer vele waarnemingen beschikken en de variantie componenten daaruit met voldoende nauwkeurigheid bepalen.

Meestal heeft men er dan geen behoefte aan die nauwkeurigheid nog numeriek vast te leggen.

Een bepaling van variantie componenten is technisch van groot belang. Is bv. de component tussen-machines groter dan die binnen-machines, dan zal men door de afstelling van de machines beter te verzorgen, een homogener product kunnen maken. In het omgekeerde geval kan men op deze wijze weinig bereiken en dient men, zo verbetering van de productie gewenst is, de machines zelf te verbeteren.

Meestal kan men bij een productieproces verscheidene varianten componenten onderscheiden en een bepaling van deze is dan van belang voor een juiste beoordeling van het procédé, voor het vast stellen van realistische toleranties, en voor het beantwoorden van de vraag of verbeteringen mogelijk zijn.

Een statistisch meer geraffineerde methode met hetzelfde doel staat bekend als de variantie analyse; wij lichten dit toe aan de hand van de gegevens van doos 2 tot en met 14 uit Tabel 74; doos 1 en 15 laten we weer wegens de afwijkende gemiddelden buiten beschouwing. In Tabel 81 zijn de gecoördeneerde waarnemingen nog eens weergegeven. De berekening loopt nu als volgt.

We berekenen eerst de sommen $x_{.j}$, voor iedere rij en de totale som $x_{..}$, en dan de grootheden A, B, en C als onder Tabel 81 aangegeven. Een tafelrekenmachine is daarbij een belangrijk hulpmiddel. De variantie analyse ziet er dan als volgt uit.

Tabel 81. Berekeningen voor een variantie analyse van de gegevens voor doos no. 2 tot 14 uit Tabel 74:

→ j = 1, ..., k

↓
i

1, ..., r

	x_{ij}					x_i
	22	18	19	16	18	93
	25	28	22	24	14	113
	23	7	26	25	27	108
	16	15	24	17	17	89
	22	20	23	19	20	104
	25	26	25	26	19	121
	24	19	21	28	22	114
	21	24	22	24	24	115
	18	21	25	26	26	116
	20	19	24	23	18	104
	17	23	17	17	18	92
	17	20	10	23	10	80
	14	17	9	13	14	67
	$x_{..} = 1316$					
Berekening:						
$A = \sum_{ij} x_{ij}^2 = 28094$						
$B = \sum x_i^2 / k = 27261$						
$C = x_{..}^2 / kr = 26644$						

Tabel 82.
Variantie analyse van Tabel 81.

I In formules: r rijen, k kolommen

Variatiebron	KS = Kwadraten- som	ν = Aantal vrij- heidsgr.	GK = Gemidd- Kwar- draat	E(GK)
Totaal	$x_{i.}^2 / kr$	1	KS/ ν	$\sigma_b^2 + k\sigma_d^2 + kr\mu^2$
Tussen rijen	$\sum x_{ij}^2 / k - x_{i.}^2 / kr$	r - 1	KS/ ν	$\sigma_b^2 + k\sigma_d^2$
Binnen rijen	$\sum x_{ij}^2 - \sum x_{i.}^2 / k$	r(k-1)	KS/ ν	σ_b^2

II Numeriek:

Variatiebron	KS	ν	GK	E(GK)
Totaal	C = 26644	1	26644	$\sigma_b^2 + 5\sigma_d^2 + 65\mu^2$
(of dozen) Tussen rijen	B-C = 617	12	51.4	$\sigma_b^2 + 5\sigma_d^2$
(of dozen) Binnen rijen	A-B = 833	52	16.1	σ_b^2

Het principe dat hieraan ten grondslag ligt is het volgende. Overeenkomstig het

$$\text{Model: } \bar{x}_{ij} = \mu + \bar{a}_i + \bar{e}_{ij} \quad (82.1)$$

worden de waarnemingen gesplitst in 3 componenten:

$$x_{ij} = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.}), \quad (82.2)$$

met een overeenkomstige splitsing van de aantallen vrijheidsgraden behorend bij deze componenten

$$rk = 1 + (r-1) + r(k-1). \quad (82.3)$$

De componenten in (82.3) zijn orthogonaal; d.w.z. hun productsommen zijn steeds nul, bv.

$$\sum_{ij} \bar{x}_{..} (x_{ij} - \bar{x}_{i.}) = 0, \text{ enz.} \quad (83.1)$$

Dientengevolge geldt:

$$\sum_{ij} x_{ij}^2 = \sum_{ij} \bar{x}_{..}^2 + \sum_{ij} (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_{ij} (x_{ij} - \bar{x}_{i.})^2 \quad (83.2)$$

en dit is de splitsing in 3 kwadraatsommen die bij de variantie analyse wordt uitgevoerd. De Gemiddelde Kwadraten zijn dan telkens de Kwadraten-sommen gedeeld door het bijbehorend aantal vrijheidsgraden.

Is het Model (82.1) van kracht dan zullen bij herhaling van het gehele experiment de a_i 's en de e_{ij} 's andere waarden aannemen en we dus ook onder G.K.'en vinden. Onder het Model (82.1) zijn de G.K.'en stochastische variabelen en de waarden $E(\underline{GK})$ in de laatste kolom geeft de verwachtingswaarden.

Een schatting van de parameters σ_b en σ_d krijgen we door deze verwachtingswaarden gelijk te stellen aan de numerieke waarden die het experiment heeft opgeleverd. Dit geeft:

$$\begin{aligned} \hat{\sigma}_b^2 &= 16.1 & \hat{\sigma}_b &= 4.0 \\ \text{en } \hat{\sigma}_d^2 &= \frac{51.4 - 16.1}{5} = 7.1, & \hat{\sigma}_d &= 2.7 \end{aligned} \quad (83.3)$$

De waarde van $\hat{\sigma}_b$ stemt vrijwel overeen met s_b volgens (79.8), $\hat{\sigma}_d$ is wat groter dan in (79.8) gevonden. Dit is niet zo vreemd, want bij de berekening van s_d werd het gemiddelde voor doos no. 14, dat juist nogal sterk afwijkt, niet meegerekend.

De F-toets

Met de F-toets toetst men de hypothese:

$$H.: \quad \sigma_d^2 = 0 \quad (83.4)$$

die in feite gelijkwaardig is met $H_0: \mu_i = \text{constant}$.

De F-toets is gebaseerd op de onderstelling dat de variaties e_{ij} binnen-dozen een normale verdeling bezitten. De theorie leert dat, wanneer de hypothese (83.4) juist is de gemiddelde kwadraten binnen rijen en tussen rijen onafhankelijke schattingen zijn van de variantie σ_b^2 . Hun verhouding bezit dan een F-verdeling. De hypothese (83.4) kan met behulp van compendium Tabel 4 worden getoets.

We vinden uit Tabel 82:

$$F_{52}^{12} = \frac{51.4}{16.1} = 3.2; \alpha_f < 0.5\%$$

We verwerpen de hypothese en beschouwen daarom als aangetoond dat $\sigma_d^2 > 0$. De toets is kennelijk een éenzijdige omdat $\sigma_d^2 < 0$ onmogelijk is.

De bovenste regel, die voor het Totaal in de Tabel 82 laat men meestal weg. Men zou met het G.K. in principe de hypothese $\mu = 0$ kunnen toetsen doch meestal weet men dat μ van nul verschilt. We hebben deze term alleen volledigheidshalve vermeld.

Deze tabellen geven het eenvoudigste voorbeeld van een variantie analyse. Deze methode van analyse vindt uitgebreide toepassingen vooral in meer gecompliceerde situaties waar ingewikkelder modellen dan (82.1) van toepassing zijn. In het algemeen geeft de variantie analyse dan één G.K. voor iedere term in het model, en kan men door toetsen nagaan welke termen een significant effect vertegenwoordigen.

De F-toets heeft echter ook zijn beperkingen. Heeft men voldoende waarnemingen binnen rijen dan wordt tenslotte ieder verschil tussen rijen significant. Een significante F-waarde geeft geen inzicht in de technische betekenis van de tussen rijen of dozen variantie; de variantie component σ_d^2 geeft dit wel.

Maar ook σ_d^2 geeft op zich geen inzicht waarde tussen dozen variantie door wordt veroorzaakt. Wil men dit onderzoeken dan is een controlekaart als van Fig. 77 doelmatig. In het algemeen dient men daarom de effecten die men op grond van een variantie analyse constateert ook grafisch in beeld te brengen. Dit verhoogt het inzicht en is voor ieder begrijpelijk. De variantie analyse is alleen voor de "insiders" van waarde.

Regressie analyse

Nauw verwant aan de variantie analyse is de regressie analyse. Hierbij beschikken we over n paren waarnemingen =

$$x_i, y_i, i = 1, \dots, n$$

en we vragen ons af in hoeverre y kan worden geschreven als lineaire functie van x . Het model luidt:

$$y_i = \alpha + \beta x_i + e_i, \quad (84.1)$$

$$E(e) = 0, E(e_i^2) = \sigma^2.$$

Uit de gegeven waarnemingen krijgen we een schatting van de parameters α en β volgens de methode der kleinste kwadraten. D.w.z. we kiezen $\hat{\alpha}$ en $\hat{\beta}$ zo dat

$$\sum_i \{y_i - \hat{\alpha} - \hat{\beta}x_i\}^2 \rightarrow \text{minimum} \quad (84.2)$$

Differentieren naar $\hat{\alpha}$ geeft

$$\bar{y} - \hat{\alpha} - \hat{\beta} \bar{x} = 0, \quad (85.1)$$

m.a.w. de lijn door $\hat{\alpha}$ en $\hat{\beta}$ gegeven gaat door het zwaartepunt \bar{x} , \bar{y} van de puntenwolk.

Met (85.1) vereenvoudigt (84.2) tot

$$\Sigma \{y_i - \bar{y} - \hat{\beta} (x_i - \bar{x})\}^2 \rightarrow \text{minimum} \quad (85.2)$$

en door differentieren naar $\hat{\beta}$ vinden we

$$\hat{\beta} = \frac{\Sigma (x_i - \bar{x}) (y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} \quad (85.3)$$

Stellen we

$$y_i^* = \hat{\alpha} + \hat{\beta} x_i = \bar{y} + \hat{\beta} (x_i - \bar{x}) \quad (85.4)$$

dan wordt de zg. rest kwadratensom.

$$\Sigma (y_i - y_i^*)^2 = \Sigma (y_i - \bar{y})^2 - \frac{\{\Sigma (x_i - \bar{x}) (y_i - \bar{y})\}^2}{\Sigma (x_i - \bar{x})^2} \quad (85.5)$$

Deze rest bezit nu (n-2) vrijheidsgraden omdat twee constanten $\hat{\alpha}$ en $\hat{\beta}$ bij de waarnemingen zijn aangepast.

$$y_i = \bar{y} + \hat{\beta} (x_i - \bar{x}) + (y_i - y_i^*), \quad (85.6)$$

dan zijn de waarnemingen y_i weer gesplitst in 3 orthogonale componenten, met een splitsing van de vrijheidsgraden in

$$n = 1 + 1 + (n-2) \quad (85.7)$$

Als numeriek voorbeeld kiezen we de gewichten uit Tabel 27A. We noemen bij de controle gevonden gewichten x_1 en de eigen opgave y . Voor de berekening is het doelmatig nog een grootheid x_0 toe te voegen die de constante waarde 1 bezit. De waarnemingen en de gang van de berekeningen zijn in Tabel 86 weergegeven.

Tabel 86

Berekeningen voor een regressie analyse.

x_0	x_1	y		x_0	x_1	y
1	74	69		1	73	70
1	84	82		1	80	78
1	59	57		1	71	69
1	72	72		1	80	70
1	76	75		1	81	80
1	70	65		1	83	80

A Matrix der ruwe produktsommen.

	x_0	x_1	y
x_0	12	903	867
x_1		68493	65760
y			63213

B Matrix der gecorrigeerde produktsommen.

	x_1	y
x_1	543	519
y		573

C. Rest kwadratensom voor y .

	y
y	77

D. Regressie konstanten en vergelijking.

$$\bar{y} = 867/12 = 72.25, \quad \bar{x} = 903/12 = 75.25$$

$$\beta = 519/543 = 0.956.$$

$$\hat{\sigma}^2 = 77/10 = 7.7, \quad v = 12 - 2 = 10$$

$$y^* = 72.75 + 0.956 (x - 75.25) = 0.31 + 0.956 x.$$

Onderstellen we dat bij een herhaling van het experiment de waarden van x_i hetzelfde blijven doch de fouten e_i toevallige veranderingen ondergaan. Dan zullen \bar{y} en $\hat{\beta}$ stochastische grootheden zijn. Het kan dan worden bewezen dat

$$E(\bar{y}) = \alpha + \beta \bar{x}, \quad \sigma_{\bar{y}}^2 = \frac{\sigma^2}{n},$$

$$E(\hat{\beta}) = \beta, \quad \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \quad (87.1)$$

en $\text{cov}(\bar{y}, \hat{\beta}) = 0$.

Uit de gegevens van Tabel 86 vinden we bijvoorbeeld

$$\sigma_{\bar{y}}^2 = \frac{7.7}{12} = 0.64, \quad \sigma_{\bar{y}} = 0.80$$

en

$$\sigma_{\hat{\beta}}^2 = \frac{7.7}{543} = 0.014, \quad \sigma_{\hat{\beta}} = 0.12.$$

Men zou verwachten dat $\beta = 1.00$, namelijk dat het opgegeven gewicht en het waargenomen gewicht gelijk opgaan.

Onderstellen we dat de fouten e normaal verdeeld zijn dan kunnen we de hypothese $\beta = 1.00$ toetsen met de t-toets.

$$t = \frac{\hat{\beta} - \beta}{\sigma_{\hat{\beta}}} = \frac{0.956 - 1.000}{0.12} = -3.7, \quad v = 12, \quad \alpha_2 = 0.2\%.$$

De hypothese $\beta = 1.00$ is dus aan ernstige twijfel onderhevig.

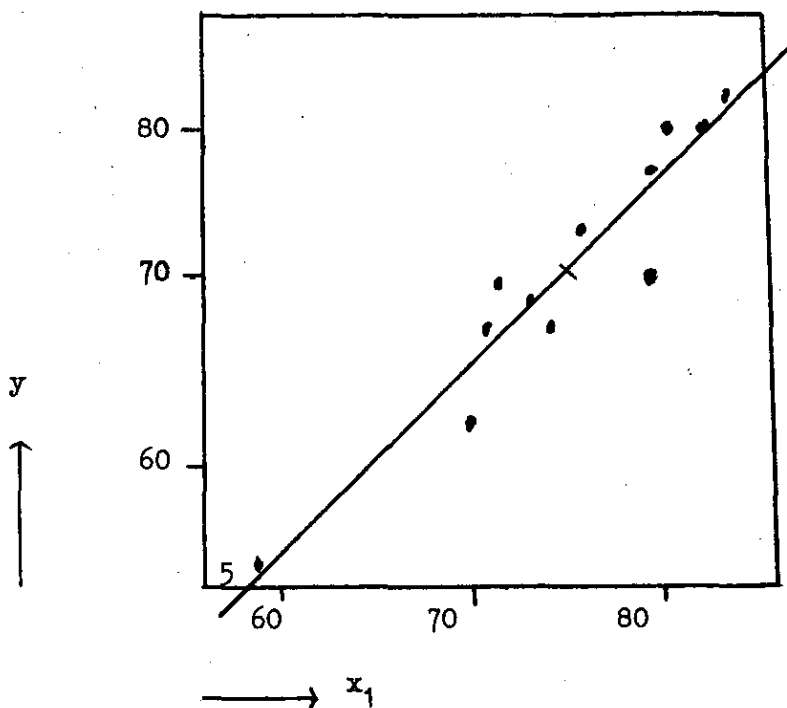


Fig. 87 . de puntenwolk voor Tabel en de regressielijn.

In fig. 87 zijn de waarnemingen van x , en y tegen elkaar uitgezet. Ook hier geldt dat naast berekening een grafische voorstelling voor een juist inzicht noodzakelijk is. We zien bijvoorbeeld, dat één punt, $x_1 = 80$, $y = 70$, stuk afwijkt. Dientengevolge liggen slechts 3 van de 12 punten beneden en 9 punten boven de regressielijn. Het vermoeden ligt daarom voor de hand dat we hier met een uitschieter te doen hebben en het heeft zin de regressie analyse ook nog eens uit te voeren met weglating van dit punt, en dan de hypothese $\beta = 1.00$ opnieuw te toetsen.

Ten slotte zij erop gewezen dat het model (84) nadrukkelijk onderstelt dat alleen de waarneming y met fouten e behebt is en dat de waarneming x zonder fouten is, of althans dat de fouten in x te verwaarlozen zijn t.o.v. die in y .

Is dit niet het geval dan zou het model luiden

$$y_i = \alpha + \beta(x_i + e_{1i}) + e_{2i}, \quad (88.1)$$

maar de bespreking van dit model valt buiten het kader van dit college en stuit statistisch op bepaalde moeilijkheden.

Niettemin is in vele gevallen waarin het model (84) niet exact opgaat de regressie analyse als hierboven gegeven toch van nut. De methode der kleinste kwadraten heeft een praktische waarde geheel afgezien van de statistische interpretatie.

Foutendiscussies en de wet van de voortplanting van fouten.

Wanneer men in de chemie een oplossing titreert, of in de natuurkunde het gewicht of volumen van een lichaam bepaalt, heeft de gemeten grootheid in principe een

$$\text{ware waarde} = \mu_0 : \quad (89.1)$$

het ware gehalte, gewicht, volume.

Voeren we herhaald metingen x_i uit, dan stemmen deze onderling niet exact overeen. We vatten een serie metingen x_i daarom steeds op als een aselechte steekproef uit een populatie van mogelijke metingen, \underline{x} , waarvoor we als model nemen

$$\underline{x} = \mu + \underline{e}, \quad E(\underline{e}) = 0, \quad E(\underline{e}^2) = \sigma^2. \quad (89.2)$$

μ is het populatiegemiddelde, \underline{e} de toevallige waarnemingsfout.

In vele gevallen is $\mu \neq \mu_0$; dan noemen we

$$\delta = \mu_0 - \mu = \text{de systematische fout} \quad (89.3)$$

in de metingen. In het engels noemt men dit vaak de "accuracy". Het is de correctie die moet worden aangebracht om uit het waarnemingsgemiddelde μ de ware waarde μ_0 te vinden: $\mu_0 = \mu + \delta$.

Verder wordt het resultaat van een natuurkundig, chemisch, of ander experiment vaak gevonden in de vorm

$$y = f(\underline{x}_1, \underline{x}_2, \underline{x}_3),$$

waarbij $\underline{x}_1, \underline{x}_2, \underline{x}_3$ gemeten grootheden zijn en y de grootheid die we willen weten. Een soortgelijk gewicht vinden we als het quotiënt van een gemeten gewicht en een gemeten volume: $y = \underline{x}_1/\underline{x}_2$; een specifieke weerstand als het product weerstand x doorsnede gedeeld door een lengte; $y = \underline{x}_1 \underline{x}_2 / \underline{x}_3$; enz. De vraag rijst dan hoe groot zijn de systematische en toevallige fouten in y .

We schrijven daartoe

$$\underline{x}_1 = \mu_{10} - \delta_1 + \underline{e}_1, \text{ enz.}$$

en ontwikkelen de functie f in het punt $\mu_{10}, \mu_{20}, \mu_{30}$ in een Taylor reeks. In vele gevallen zijn δ en \underline{e} klein zodat kwadratische termen in deze grootheden kunnen worden verwaarloosd. Dan geldt

$$y = f_0 + f'_1 \delta_1 + f'_2 \delta_2 + f'_3 \delta_3 + f''_1 \underline{e}_1 + f''_2 \underline{e}_2 + f''_3 \underline{e}_3, \quad (89.4)$$

waarin $f_0 = f(\mu_{10}, \mu_{20}, \mu_{30})$, $f'_1 = f'_{x_1}(\mu_{10}, \mu_{20}, \mu_{30})$ enz.

Uit () volgt:

$$E(\underline{y}) = f_0 + f'_1 \delta_1 + f'_2 \delta_2 + f'_3 \delta_3$$

Meestal zullen we aannemen dat de "ware" waarde van y gegeven wordt door f_0 .

De systematische fout in y is dan

$$\delta y = f'_1 \delta_1 + f'_2 \delta_2 + f'_3 \delta_3 \quad (90.1)$$

Verder volgt uit ()

$$\begin{aligned} \text{var } \underline{y} = & f_1'^2 \text{var } \underline{x}_1 + f_2'^2 \text{var } \underline{x}_2 + f_3'^2 \text{var } \underline{x}_3 + 2f_1'f_2' \text{cov}(\underline{x}_1, \underline{x}_2) + \\ & 2f_1'f_3' \text{cov}(\underline{x}_1, \underline{x}_3) + 2f_2'f_3' \text{cov}(\underline{x}_2, \underline{x}_3). \end{aligned}$$

In vele gevallen vallen de termen met de covarianties weg omdat \underline{x}_1 , \underline{x}_2 , en \underline{x}_3 onafhankelijk van elkaar worden bepaald. Formule () staat bekend als de wet van de voortplanting van fouten.

Bovenstaande formules gelden alleen wanneer we ons in de Taylor reeks () tot de lineaire termen mogen beperken. Zo nodig kan men ook verder gaan en tevens de kwadratische termen meerekenen, doch de formules worden dan gecompliceerd. We vermelden slechts één speciaal geval; zijn de waarnemingen \underline{x}_1 , \underline{x}_2 en \underline{x}_3 vrij van systematische fouten dan geldt

$$\begin{aligned} E(\underline{y}) = & f_0 + \frac{1}{2!} \{ f_{11}'' \text{var } \underline{x}_1 + f_{22}'' \text{var } \underline{x}_2 + f_{33}'' \text{var } \underline{x}_3 \\ & + 2f_{12}'' \text{cov}(\underline{x}_1, \underline{x}_2) + 2f_{13}'' \text{cov}(\underline{x}_1, \underline{x}_3) + 2f_{23}'' \text{cov}(\underline{x}_2, \underline{x}_3) \} \end{aligned} \quad (90.2)$$

De verwachtingswaarde van \underline{y} is dan niet gelijk aan f_0 doch bezit een systematische fout die van de varianties en covarianties afhangt.

Vaak beschrijft men de nauwkeurigheid van waarnemingen niet door de standaardafwijking σ maar door de

$$\text{variantiecoëfficiënt} = V = \frac{\sigma}{\mu} \text{ of } = 100 \frac{\sigma}{\mu} \% \quad (90.3)$$

In vele gevallen is men meer in deze relatieve fout geïnteresseerd dan in de standaardafwijking zelf. Voor een Poisson verdeling bijv. is

$$\sigma = \sqrt{\mu} \quad \text{en} \quad V = \frac{1}{\sqrt{\mu}}$$

De standaardafwijking σ neemt evenredig met $\sqrt{\mu}$ toe, doch de variatie coëfficiënt V neemt omgekeerd evenredig met $\sqrt{\mu}$ af.

De variatie coëfficiënt is vooral van waarde wanneer we met een betrekking

van de algemene vorm

$$y = A \begin{matrix} a_1 & a_2 & a_3 \\ \underline{x}_1 & \underline{x}_2 & \underline{x}_3 \end{matrix} \quad (91.1)$$

te maken hebben. Zijn \underline{x}_1 , \underline{x}_2 en \underline{x}_3 onafhankelijk dan leidt de wet van de voortplanting van fouten nu tot de relatie

$$V^2(y) = a_1^2 V^2(\underline{x}_1) + a_2^2 V^2(\underline{x}_2) + a_3^2 V^2(\underline{x}_3), \quad (91.2)$$

dus nu een additieve betrekking tussen de kwadraten van de variatie coëfficiënten.

Voorbeelden

A. Heeft \underline{x} een Poisson verdeling, dan is in eerste benadering

$$E(\sqrt{\underline{x}}) = \sqrt{\mu}, \text{ en var } \sqrt{\underline{x}} = \frac{1}{4}.$$

De variantie van $\sqrt{\underline{x}}$ is dus een konstante onafhankelijk van μ .
In tweede benadering is

$$E(\sqrt{\underline{x}}) = \sqrt{\mu} \left(1 - \frac{1}{8\mu}\right).$$

B. Voor \underline{x}_v^2 geldt $E(\underline{x}_v^2) = v$, var $\underline{x}_v^2 = 2v$

Hieruit volgt in eerste benadering.

$$\text{var } \underline{x}_v = \frac{1}{2},$$

dus een konstante onafhankelijk van v .

C. Stel we hebben twee radioactieve bronnen met intensiteiten λ_1 en λ_2 en we wensen het verschil $\lambda_1 - \lambda_2$ te bepalen. Voor bron I tellen we x_1 deeltjes in een tijd T_1 en voor bron II, x_2 in tijd T_2 . Dan is de schatting voor $\lambda_1 - \lambda_2$

$$\hat{\lambda}_1 - \hat{\lambda}_2 = \frac{x_1}{T_1} - \frac{x_2}{T_2},$$

en var $(\hat{\lambda}_1 - \hat{\lambda}_2) = \frac{\lambda_1}{T_1} + \frac{\lambda_2}{T_2}.$ (a)

Is de totale beschikbare waarnemingstijd

$$T = T_1 + T_2 \quad (b)$$

gegeven, dan is het het voordeligst T_1 en T_2 zó te kiezen dat (a) minimaal wordt onder de voorwaarde (b).

Met de methode van Lagrange moeten we

$$\frac{\lambda_1}{T_1} + \frac{\lambda_2}{T_2} + k(T_1 + T_2 - T) \text{ minimaal}$$

maken als functie van T_1 , T_2 , en k .

Dit geeft

$$T_1 = \sqrt{\frac{\lambda_1}{k}}, \quad T_2 = \sqrt{\frac{\lambda_2}{k}}$$

of

$$\frac{T_1}{T_2} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \quad (c)$$

D. Is het ons niet om het verschil maar om de verhouding λ_1/λ_2 te doen, dan vinden we

$$\sqrt{V^2(\hat{\lambda}_1 / \hat{\lambda}_2)} = \sqrt{V^2(\hat{\lambda}_1)} + \sqrt{V^2(\hat{\lambda}_2)} = \frac{1}{\lambda_1 T_1} + \frac{1}{\lambda_2 T_2}, \quad (d)$$

en wanneer we (d) minimaal maken onder de voorwaarden (b) resultaat.

$$\frac{T_1}{T_2} = \frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}} \quad (e)$$

De gunstigste verdeling van de waarnemingstijd hangt dus af van de vraag waarvoor we ons interresseren, en is voor het verschil $\lambda_1 - \lambda_2$ een andere dan voor de verhouding λ_1/λ_2 .

Natuurlijk kan men van de formules (d) en (e) alleen gebruik maken wanneer men λ_1 en λ_2 ongeveer kent, doch een voorlopige schatting is voldoende, want zo nauwkeurig komt het er niet op aan.

E. Voor de lenzen proef geldt,

$$\frac{1}{f} = \frac{1}{b} + \frac{1}{v} \text{ of } f = \frac{bv}{b+v}$$

waaruit volgt

$$df = \frac{b^2 dv + v^2 db}{(b+v)^2} \quad (f)$$

en verder

$$\sigma_f^2 = \frac{b^4 \sigma_v^2 + 2b^2 v^2 \text{cov}(b, v) + v^4 \sigma_b^2}{(b+v)^4} \quad (g)$$

Indien men wil uitmaken wat de optimale keuze van b en v is, dient men er nu echter rekening mee te houden dat σ_b^2 en σ_v^2 van b en v afhankelijk kunnen zijn. Men moet dus eerst de instelfouten van b en v experimenteel voor verschillende waarden van b en v bepalen.

Ook kan in dit geval, al naar de methode van werken een correlatie tussen b en v bestaan, die dan natuurlijk in rekening moet worden gebracht.

Tot slot zij er ook op gewezen dat de scheiding tussen systematische en toevallige fouten geen absolute scheiding is. Voert men een chemische analyse steeds uit met één en hetzelfde pipet, dan zal een fout in het volume van deze pipet optreden als een systematische fout; gebruikt men steeds andere pipetten dan treden volume fouten op als toevallige fouten.

Statistische methoden geven geen informatie over één en-kele systematische fout. Door vergelijk van resultaten met verschillende pipetten bereikt kan de statistische analyse ons wel leren of verschillen tussen pipetten onderling bestaan. Zij leert ons echter niet of de verzameling beschikbare pipetten als geheel beschouwd een systematische fout bezit.