TECHNISCHE HOGESCHOOL EINDHOVEN
Afdeling Algemene Wetenschappen
Onderafdeling der Wiskunde

INLEIDING STATISTIEK

....

Najaarssemester 1979

Bibl. Mag

Technische Hogeschool Eindhoven

Onderafdeling der Wiskunde

Inleiding Statistiek

Dictaatnr. 2.270 Prijs f 3,00

TECHNISCHE HOGESCHOOL EINDHOVEN

Onderafdeling der Wiskunde

INLEIDING STATISTIEK

najaarssemester 1979

Inhoudsbeschrijving

INLEIDING STATISTIEK

....

Najaarssemester 1979

	Onderwerp	**blz**
	Formuleblad	2-3
I	FREQUENTIEVERDELING VAN EEN STEEKPROEF	4
§1	Tabellarisch en grafische weergeven van steekproeven. Frequentie	4
§2	Groeperen	6
§3	Verwerking van tabellen	7
	STEEKPROEFGEMIDDELDE EN STEEKPROEFVARIANTIE	9
§1	Definities	9
§2	Berekening van steekproefgemiddelde en -variantie	9
§3	Berekening van $\overline{x}$ en s^2 uit de frequentiefunctie	10
§4	Standaardiseren	11
	GRAFIEKEN	13
	CORRELATIE EN REGRESSIE	17
§1	Inleidende opmerkingen	17
§2	Regressievergelijking (methode van de kleinste kwadraten)	19
	LIJST VAN DE BELANGRIJKSTE TERMEN IN DE KANSREKENING EN STATISTIEK (A.J. BOSCH)	22
	Opgaven 1-12	35-39
	Antwoorden opgaven 1-12	40
	Opgaven 13-21	41-42
	Examens/Tentamens 1977-1980	43-54
	Antwoorden Examens/Tentamens	55-56

JdG, 25 Juli 2005

Inhoudsopgave

Formuleblad

I Verdelingsfunctie van een steekproef.

§1 Tabellarisch en grafisch weergeven van steekproeven. Frequentie.

§2 Groeperen.

§3 Verwerking van tabellen.

II Steekproefgemiddelde en steekproefvariantie.

§1 Definities.

§2 Berekening van gemiddelde en variantie.

§3 Berekening van $\bar{x}$ en s^2 uit de frequentiefunctie.

§4 Standaardiseren.

III Grafieken.

IV Correlatie en regressie.

§1 Inleidende opmerkingen.

§2 Regressievergelijking (Methode van de kleinste kwadraten).

Lijst van belangrijkste engelse termen in de kansrekening en statistiek (A.J. Bosch).

Opgaven.

Antwoorden.

Tentamens.

Antwoorden.

Formules te gebruiken bij het tentamen Inleiding Statistiek.

Algemene somregel: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Voorwaardelijke kans: $P(B|A) = \frac{P(A \cap B)}{P(A)}$.

A en B stochastisch onafhankelijk: $P(A \cap B) = P(A).P(B)$.

Binomiale verdeling: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$; $E[X] = n.p$; $\sigma[X] = \sqrt{np(1-p)}$.

Poisson verdeling: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$; $E[x] = \lambda$; $\sigma[x] = \sqrt{\lambda}$.

Normale verdeling: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$; $E[X] = \mu$; $\sigma[X] = \sigma$.

Uniforme verdeling: $f(x) = \frac{1}{b-a}$ op $[a,b]$; $E[x] = \frac{a+b}{2}$; $\sigma^2[x] = \frac{1}{12}(b-a)^2$.

Verdelingsfunctie: $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$.

$$P(a < X \le b) = F(b) - F(a).$$

Bepaling kritiek gebied: $P(X \in Z|H_0) = \alpha$.

Verwachtingswaarde: $\mu = E[X] = \sum_{u \in U} X(u).P(u) = \sum_x x.P(X = x)$.

$$E[aX + b] = a.E[X] + b.$$

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E[X_i].$$

X en Y onafhankelijk: $E[X.Y] = E[X].E[Y]$.

Variantie: $Var[X] = E[(X-\mu)^2] = E[X^2] - \mu^2$.

Standaardafwijking: $\sigma[X] = \sqrt{Var[X]}$.

$$Var[aX + b] = a^2 Var[X].$$

$X_1 \ldots X_n$ onafhankelijk: $Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i]$.

$$Z = \frac{X_1 + X_2 + \ldots + X_n}{n} = \bar{X}: \; E[Z] = E[\bar{X}] = E[X]; \; \sigma[Z] = \sigma[\bar{X}] = \frac{\sigma[X]}{\sqrt{n}}.$$

Wet van de grote aantallen: $P(|X - E[X]| \geq d) \leq \frac{Var[X]}{d^2}$.

$$P(|\bar{X} - E[X]| \geq d) \leq \frac{Var[X]}{d^2 n} .$$

Is Z een N(0,1) verdeelde stochast: $P(Z \leq z) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} = \Phi(z)$.

$$P(a < Z < b) = \Phi(b) - \Phi(a).$$

Gestandaardiseerde stochast: $X^* = \frac{X - \mu}{\sigma}$.

Centrale limietstelling: $\lim_{n\to\infty} P(X^* \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$.

Rekenkundig gemiddelde $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
of steekproefgemiddelde
Steekproefvariantie $s^2 := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1}[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2]$.

Coderen: $x_i = c_1 x_i^* + c_2$: $\bar{x} = c_1 \bar{x}^* + c_2$
$s^2 = c_1^2 (s^*)^2$.

$\bar{x} = \sum_{i=1}^{m} x_i \tilde{f}(x_i)$ ($n\tilde{f}(x_i)$ = frequentie van waarneming x_i)

$s^2 = \frac{1}{n-1}\{\sum_{i=1}^{m} x_i^2 n\tilde{f}(x_i) - \frac{1}{n}[\sum_{i=1}^{m} x_i n\tilde{f}(x_i)]^2\}$.

Correlatiecoëfficiënt: $r(\underline{x},\underline{y}) = s_{xy}/s_x s_y$.

met $s_{xy} := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)]$.

Regressielijn: $y - \bar{y} = b(x - \bar{x})$ met $b = s_{xy}/s_x^2$.

I Frequentieverdeling van een steekproef.

In dit hoofdstuk gaan we steeds uit van een representatieve steekproef. Voor het hoe en waarom van zo'n steekproef wordt verwezen naar b.v.

Nijdam: Statistiek en Kansrekening voor het V.W.O. §1-1.

Dictaat Wiskunde 31/49 §5-1.

Dr. P.C. Sander: Statistiek. Een aanval van gezond verstand (Openbare les, 19 mei 1978).

§1. Het tabellarisch en grafisch weergeven van steekproeven. Frequentie

a). Als voorbeeld nemen we de gegevens uit tabel 1.1, waar we te maken hebben met een steekproef uit een zekere populatie. Deze steekproef bestaat uit 50 getallen, de steekproefwaarden. De steekproefgrootte n is dus 50.

Om dit getallenmateriaal te verwerken, gebruiken we tabel 1.2:

In kolom 1 staan de steekproefwaarden van de kleinste tot en met de grootste waarde.

In kolom 2 tellen we door middel van turfstreepjes het aantal keren dat ieder van de waarden voorkomt, in kolom 3 geven we dat aan door middel van een getal. We noemen het aantal keren dat een waarde x in de steekproef optreedt, de frequentie van die waarde x in de steekproef.

In kolom 4 komt de relatieve frequentie, dit is de frequentie gedeeld door de steekproefgrootte.

Als we voor een zekere x alle frequenties optellen, corresponderend met steekproefwaarden die kleiner of gelijk zijn aan x, krijgen we de cumulatieve frequentie, corresponderend met die x. (kolom 5).

Delen we deze getallen door de steekproefgrootten, dan krijgen we de cumulatieve relatieve frequentie (kolom 6).

. De relatieve frequentie is tenminste 0 en ten hoogste 1.

. Beschouw een gegeven steekproef ter grootte n, bestaande uit m numerieke verschillende waarden

$$x_1, x_2, \ldots, x_m \qquad (m \leq n).$$

met corresponderende relatieve frequenties

$$\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_m$$

We voeren dan in de frequentie functie van de steekproef

$$\tilde{f}(x) = \begin{cases} \tilde{f}_j & \text{als } x = x_j \qquad (j = 1,2,\ldots,m) \\ 0 & \text{anders.} \end{cases}$$

Deze functie laat zien hoe de waarden van de steekproef zijn verdeeld. We zeggen daarom dat deze functie de frequentie verdeling van de steekproef bepaalt.

. De som van alle relatieve frequenties in een steekproef is gelijk 1

$$\sum_{j=1}^{m} \tilde{f}(x_j) = \tilde{f}(x_1) + \tilde{f}(x_2) + \ldots + \tilde{f}(x_m) = 1.$$

. Vervolgens voeren we in de functie

$\tilde{F}(x)$ = som van de relatieve freqyenties van al de steekproefwaarden die kleiner zijn dan of gelijk aan x.

ofwel

$$\tilde{F}(x) = \sum_{t \leq x} \tilde{f}(t)$$

Deze functie wordt genoemd de cumulatieve verdelingsdichtheid van de steekproef of de verdelingsfunctie van de steekproef $\tilde{F}(x)$ is een stapfunctie, die sprongen ter grootte $\tilde{f}(x)$ heeft bij die x met $\tilde{f}(x) \neq 0$.

b. Er zijn verscheidene manieren om steekproeven grafisch weer te geven. In fig. 1.1 geven we een afbeelding van de turfstaat uit tabel 1.2 (kolom 2). In fig. 1.2 geven we door middel van een staafdiagram, uit tabel 1.2 de kolom 4 weer. Op dezelfde manier is het ook mogelijk om de (absolute) frequenties weer te geven. In fig. 1.3 hebben we te maken met een histogram. In plaats van verticale lijnsegmenten, hebben we nu rechthoekjes, waarbij de oppervlakte van ieder van deze rechthoekjes overeenkomt met de corresponderende (relatieve) frequentie. (let wel: totale oppervlak = 1).
Het is duidelijk dat een staafdiagram de voorkeur verdient in experimenten waarbij we tellen, en dat een histogram vooral bij experimenten waarbij gemeten wordt in aanmerking komt.
In fig. 1.4 hebben we te maken met een lijndiagram, die we kunnen verkrijgen door de middens van de bovenzijden van de rechthoekjes uit het histogram met elkaar te verbinden.

§2. Groeperen

. Als een steekproef uit teveel numeriek verschillende steekproefwaarden bestaat is het beter om uit de de frequentiefunctie en de daarmee samenhangende grafieken onnodige details te verwijderen.
Dit kan gedaan worden door de steekproef te groeperen, dat wil zeggen door de waarnemingen in klassen in te delen.
Is een steekproef gegeven, dan kiezen we een interval I dat alle steekproef waarden bevat. We verdelen I in deelintervallen die klasseintervallen genoemd worden. De middens van deze intervallen zijn de klassemiddens.
De steekproefwaarden in zo'n interval vormen een klasse. Het aantal waarden in een klasse is de overeenkomstige klasse-frequentie. Delen door de steekproefgrootte geeft de relatieve klassefrequentie.
Deze frequentie, beschouwd als een functie op de klasse middens, wordt de frequentiefunctie van de groepeerde steekproef genoemd, en kan genoteerd worden als $\tilde{f}(x)$. De cumulatieve frequentiefunctie van de gegroepeerde steekproef kan genoteerd worden als $\tilde{F}(x)$ en is gedefinieerd door

$$\tilde{F}(x) = \sum_{t \leq x} \tilde{f}(t).$$

. In veel toepassingen zal het mogelijk zijn om aan de volgende regels te voldoen

(1) Alle klasseintervallen zullen dezelfde lengte hebben.
(2) De klasseintervallen zullen zo gekozen worden dat de klassemiddens overeenkomen met eenvoudige getallen.
(3) Als een steekproefwaarde samenvalt met een gemeenschappelijk eindpunt van twee klasseintervallen, kunnen we dit in rekening brengen door ½ op te tellen bij de klassefrequentie van ieder van die twee intervallen (waar mogelijk zullen we dit samenvallen proberen te vermijden).

. Hoe minder klassen we kiezen, hoe eenvoudig de gegroepeerde steekproef zal zijn, maar hoe meer informatie verloren zal gaan. In de meeste van de practische toepassingen zullen we het aantal klassen kiezen tussen 10 en 20. (Een ander vuistregel is: Kies het aantal klassen ongeveer $\sqrt{n}$, waarbij n = aantal waarnemingen, maar in ieder geval >5 en <16).

§3 Verwerking van tabellen

a) Voorbeelden

1) Verbruik tabaksartikelen

Jaar	in miljoenen guldens				in miljoenen(aantal)		in tonnen
	sigaren	sigaretten	tabak	totaal	sigaren	sigaretten	tabak
1	2	3	4	5	6	7	8
1955	194	468	120	782	1056	11734	8315
1956	205	527	120	852	1093	13149	8208
1957	219	598	134	951	1163	13457	8166
1958	216	618	152	986	1170	13043	8956
1959	226	638	163	1027	1260	13452	9429

2) Luchtvlootcapaciteit van de KLM.

jaar	vliegtuigen in gebruik (jaargemiddelde)	gemiddeld per vliegtuig				totaal		
		aantal vlieguren	snelheid in km. per uur	laadvermogen in tonnen	aantal zitplaatsen	plaatskm	tonkm produktie	vlieguren
						mln		x 1000
1958	.	.	369	6,8	.	.	475	188
1963	60,2	2270	496	12,1	101	6472	855	142
1965	39,6	2700	567	15,1	116	6884	997	117
1966	38,7	2910	585	15,5	118	7704	1119	123
1967	39,5	2920	605	15,8	121	8490	1239	129

Bron: jaarverslag KLM.

b) Eisen waaraan een tabel moet voldoen.

a) Het opschrift (of onderschrift) van een tabel moet beknopt en volledig de inhoud beschrijven. Ook de bron moet worden vermeld.

b) De kop van een kolom of kolomgroep moet een beknopte omschrijving bevatten van de inhoud van de kolom of kolomgroep met vermelding der eenheden. Ditzelfde geldt uiteraard ook voor de regels van een tabel.

c) De tabel mag niet te gedetailleerd zijn.

d) Gebruik bij wat grotere tabellen kolom- en regelgroepen.

e) Als kolommen en regels genummerd zijn wordt verwijzing naar de tabel gemakkelijker.

f) Rond getallen van meer dan 4 cijfers af.

g) Bij percentages moeten steeds de aantallen worden vermeld.

h) Zorg dat te vergelijken getallen in aan elkaar grenzende velden staan.

i) De volgende notatie wordt toegepast:

.   : gegeven ontbreekt.

-      : gegeven is exact nul.

0   : gegeven is te klein om in de gegeven eenheid te worden uitgedrukt

*   : gegeven is voorlopig.

blanco: gegeven kan logischerwijze niet voorkomen.

x   : gegeven is geheim.

1977-1978 betekent: 1977 tot en met 1978. (ook wel ÷)

1977/1978 betekent: het gemiddelde van de jaren 1977 en 1978,

1977/'78 betekent: oogstjaar, boekjaar, schooljaar enz. beginnend in 1977 en eindigend in 1978.

Opgaven: 7a,8a,8b,9a,9b.

Tentamens: 14-12-77: 1a,4.
22-4-78: 4.
19-12-78: 3.

II Steekproefgemiddelde en steekproefvariantie

Gegeven is steeds een steekproef ter grootte n met als steekproefwaarden $x_1, x_2, \ldots, x_n$.

§1 Definities

a) (rekenkundig) gemiddelde $\bar{x} = \dfrac{x_1+\ldots+x_n}{n} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$ (II-1-1).

b) steekproefvariantie $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$ (II-1-2).

c) standaarddeviatie = s

d) range = verschil tussen grootste en kleinste waarneming.

e) Q_p = getal zodat tenminste p% van de steekproefwaarden kleiner of gelijk zijn aan Q_p

Q_{50}= mediaan

Q_{25}= eerste kwartiel

Q_{75}= derde kwartiel

f) $Q_{75}-Q_{25}$ = interkwartielafstand van de steekproef

g) modus = steekproef waarde die het meeste voorkomt.

Opmerking: De steekproefwaarden kunnen verdeeld worden in
in 4 groepen door middel van 3 kwartielen,
in 10 groepen door middel van 9 decielen,
in 100 groepen door middel van 99 percentielen.

§2 Berekening van steekproefgemiddelde en -variantie.

a) $(x_i - \bar{x})^2 = x_i^2 - 2x_i\bar{x} + (\bar{x})^2$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}\{x_i^2 - 2x_i\bar{x} + (\bar{x})^2\} = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n(\bar{x})^2 =$$

$$= \sum_{i=1}^{n} x_i^2 - 2.\frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i) + n(\frac{1}{n}\sum_{i=1}^{n} x_i)^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2$$

$$s^2 = \frac{1}{n-1}[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2]. \qquad \text{(II-2-1)}$$

b) Verplaatsen van de oorsprong:

$x_i = x_i^* + c$ of $x_i^* = x_i - c$.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n}(x_i^* + c) = \frac{1}{n}\sum_{i=1}^{n} x_i^* + \frac{1}{n}nc = \overline{x^*} + c. \qquad \text{(II-2-2)}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^* + c - \overline{x^*} - c)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^* - \overline{x^*}) = (s^*)^2$$

(II-2-3)

c) Coderen: $x_i = c_1 x_i^* + c_2$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n}(c_1 x_i^* + c_2) = c_1\frac{1}{n}\sum_{i=1}^{n} x_i^* + \frac{1}{n}nc_2 = c_1\overline{x^*} + c_2. \qquad \text{(II-2-4)}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(c_1 x_i^* + c_2 - c_1\overline{x^*} - c_2)^2 =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} c_1^2(x_i^* - \overline{x^*})^2 = c_1^2(s^*)^2. \qquad \text{(II-2-5)}$$

§3 Berekening van $\bar{x}$ en s^2 uit de frequentiefunctie

Stel gegeven een steekproef ter grootte n: $y_1,\ldots,y_n$ met m numeriek verschillende waarden: $x_1,\ldots,x_m$.

a) $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{m} x_i\, n\tilde{f}(x_i) = \sum_{i=1}^{m} x_i\, \tilde{f}(x_i). \qquad \text{(II-3-1)}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{m}(x_i - \bar{x})^2 n\tilde{f}(x_i) =$$

$$= \frac{n}{n-1}\sum_{i=1}^{m}(x_i - \bar{x})^2\tilde{f}(x_i) \qquad \text{(II-3-2)}$$

Maken we tevens gebruik van (II-2-1), dan wordt

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{m}(x_i - \bar{x})^2 n\tilde{f}(x_i) =$$

$$= \frac{1}{n-1}\Big[\sum_{i=1}^{m} n\tilde{f}(x_i)x_i^2 - \frac{1}{n}\Big(\sum_{i=1}^{m} n\tilde{f}(x_i)x_i\Big)^2\Big] \qquad \text{(II-3-3)}$$

b) Coderen: $x_i = c_1 x_i^* + c_2$

(i) Omdat $\tilde{f}(x_i^*) = \tilde{f}(x_i)$ wordt $\overline{x^*} = \sum_{i=1}^{m} x_i^* \tilde{f}(x_i)$. (II-3-4)

Dan is $\bar{x} = c_1 \overline{x^*} + c_2$.

(ii) $s^2 = c_1^2 (s^*)^2$ waarin

$$(s^*)^2 = \frac{1}{n-1} \sum_{i=1}^{m} n\tilde{f}(x_i)(x_i^* - \overline{x^*})^2$$

of, met gebruikmaking van (II-2-1):

$$(s^*)^2 = \frac{1}{n-1} [\sum_{i=1}^{m} n\tilde{f}(x_i)(x_i^*)^2 - \frac{1}{n} (\sum_{i=1}^{m} n\tilde{f}(x_i) x_i^*)^2]$$

c) In het geval van een gegroepeerde steekproef, definiëren we het gemiddelde als

$$\bar{x} = \sum_{i=1}^{k} x_i \tilde{f}(x_i) = \frac{1}{n} \sum_{i=1}^{k} x_i \; n\tilde{f}(x_i) \qquad \text{(II-3-5)}$$

en de steekproefvariantie als

$$s^2 = \frac{n}{n-1} \sum_{i=1}^{k} (x_i - \bar{x})^2 \tilde{f}(x_i) = \frac{1}{n-1} \sum_{i=1}^{k} (x_i - \bar{x})^2 n\tilde{f}(x_i). \qquad \text{(II-3-6)}$$

Hierin is k = aantal klasse intervallen

x_i = klassemidden van het i-de interval

$n\tilde{f}(x_i)$ = overeenkomstige klassefrequentie.

evenals in §3b ii) is (II-3-6) te herschrijven als

$$s^2 = \frac{1}{n-1} [\sum_{i=1}^{k} x_i^2 \; nf(x_i) - \frac{1}{n}(\sum_{i=1}^{k} x_i \; n\tilde{f}(x_i))^2] \qquad \text{(II-3-7)}$$

§4 Standaardiseren

Heeft een steekproef $x_1, \ldots, x_n$ een gemiddelde $\bar{x}$ en een standaard deviatie s, dan heeft de steekproef van getransformeerde waarden $y_1, \ldots, y_n$ met

$$y_i = \frac{x_i - \bar{x}}{s}$$

een gemiddelde 0 en een standaarddeviatie 1.

Bewijs:

(i) $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\bar{x}}{s} = \frac{1}{ns}\sum_{i=1}^{n}(x_i-\bar{x}) =$

$= \frac{1}{ns}[(x_1+x_2+\ldots+x_n)-n\bar{x}] = \frac{1}{ns}[(x_1+\ldots+x_n)-n.\frac{1}{n}(x_1+\ldots+x_n)] = 0$

(ii) $s^2(y) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{x_i-\bar{x}}{s})^2 =$

$\frac{1}{s^2}\cdot\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{s^2}\cdot s^2 = 1.$

Dit resultaat is ook gemakkelijk te verkrijgen door gebruik te maken van §2c. We schrijven dan

$x_i = s\, y_i + \bar{x}$ dus met $c_1 = s$ en $c_2 = \bar{x}$

Toepassen van (II-2-4):

$\bar{x} = s\,\bar{y} + \bar{x}$ waaruit $\bar{y} = 0$.

Toepassen van (II-2-5):

$s^2 = s^2\,(s(y))^2$ waaruit $(s(y))^2 = 1$.

Opgaven: 7b,8c,9c,9d.

Tentamens: 14-12-77: 3.
22-4-78: 3.
19-12-78: 1a,3.
31-3-79: 5.

III Grafieken

Een grafiek wordt vaak toegepast in die gevallen waarbij het verloop van een gegeven of, wanneer men meer gegevens in één grafiek uitbeeldt, het onderlinge verband tussen de gegevens een belangrijker rol speelt dan de absolute hoogten van de uitgebeelde gegevens.

Het voordeel van een grafiek ten opzichte van een tabel is dat men in één oogopslag de daarin voorkomende gegevens kan overzien. Een bezwaar van de grafiek ten opzichte van de tabel is dat men uit een grafiek geen exacte gegevens kan afleiden.

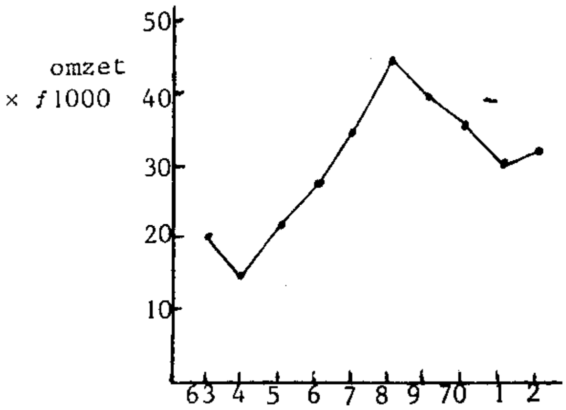
De eisen die aan een grafiek gesteld worden zijn

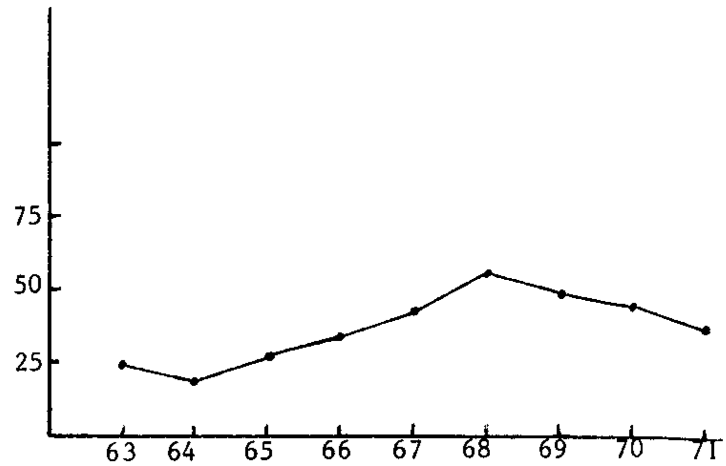
1) De grafiek moet op zichzelf leesbaar zijn. Er moet een opschrift zijn en bij de assen moet vermeld staan welke grootheden daar uitgezet zijn en welke eenheden gebruikt worden.
2) Komen er meerdere lijnen in de grafiek voor, dan dienen deze van elkaar onderscheiden te worden, hetzij door kleur, hetzij door tekening (getrokken lijn, stippellijn etc.).
3) Bij de grafiek moet vermeld worden wat deze lijnen voorstellen.
4) De grafiek moet het papier goed vullen (keuze schaalverdeling).
5) De assen mogen niet te veel cijfers bevatten.

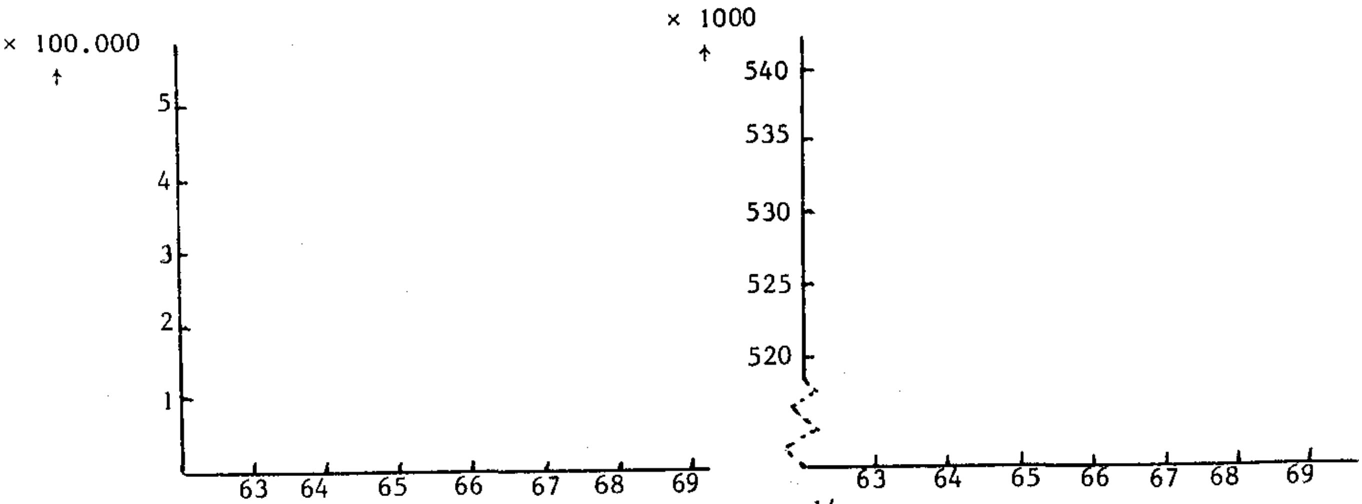
Voorbeelden van diverse grafische voorstellingen.

a) Lijndiagram. (opgave 13)

Bij een lijndiagram worden de opeenvolgende punten door rechte lijnen verbonden.



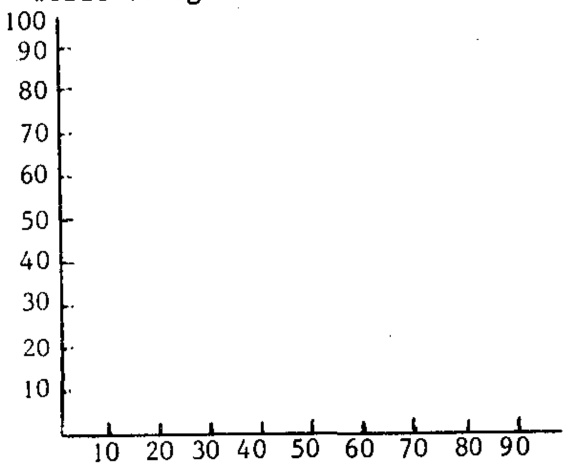




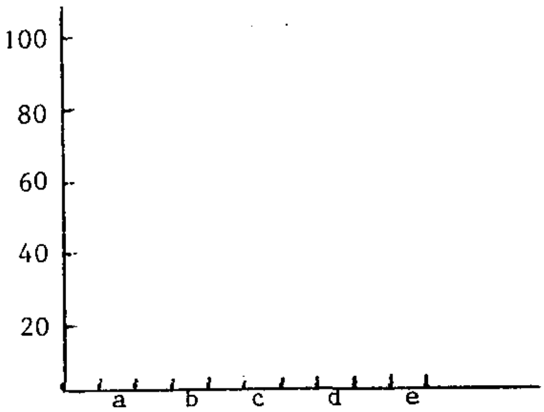
opgave 14

b) <u>Curve</u> (opgave 15)

Een curve is een diagram waarin het verloop door een vloeiende lijn wordt voorgesteld.



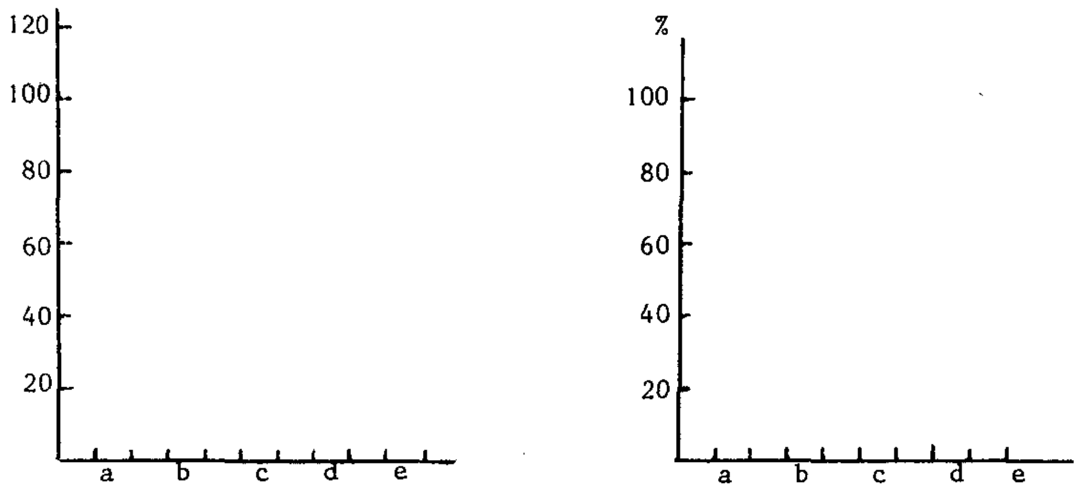
c) <u>Staafdiagrammen</u> (opgave 16)



Opm.: Bij een staafdiagram mag geen scheurlijn gebruikt worden.

Een staafdiagram waarin één gegeven wordt uitgebeeld, noemt men een enkelvoudig staafdiagram.

Een staafdiagram waarin meer gegevens per staaf worden uitgebeeld, noemt men een samengesteld staafdiagram.

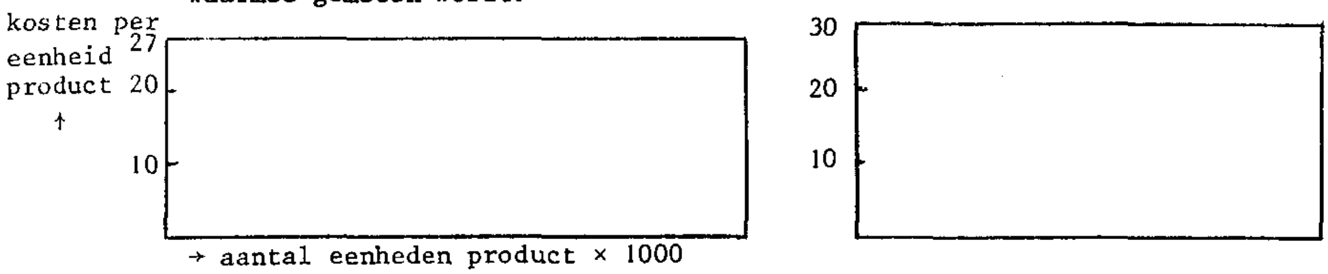


opgave 17

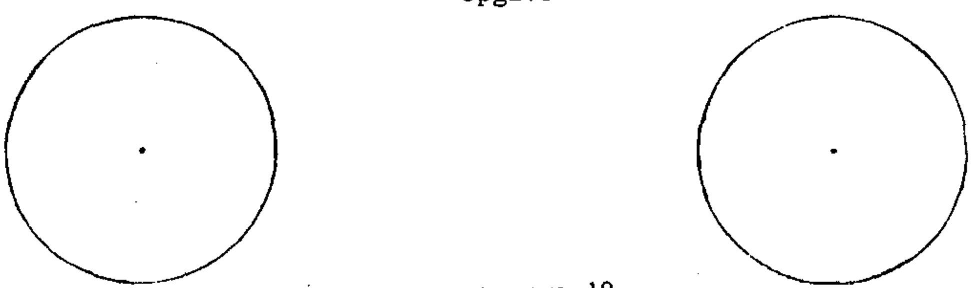
Opm.: Betreft het diagram een uitbeelding van opzichzelf staande objecten, dan moeten de kolommen los van elkaar staan. Betreft het echter gegevens *waarbij de onderscheiding berust op* een verdeling b.v. in jaren, dan is het aan te bevelen de staven tegen elkaar te plaatsen (zie histogrammen in §I-1).

d) <u>Vlakdiagrammen</u>

In tegenstelling tot de staafdiagrammen, waarbij alleen de lengte een rol speelt, is bij vlakdiagrammen zowel de lengte als de breedte van de staaf van belang. In een vlakdiagram is de oppervlakte de maat waarmee gemeten wordt.



opgave 18

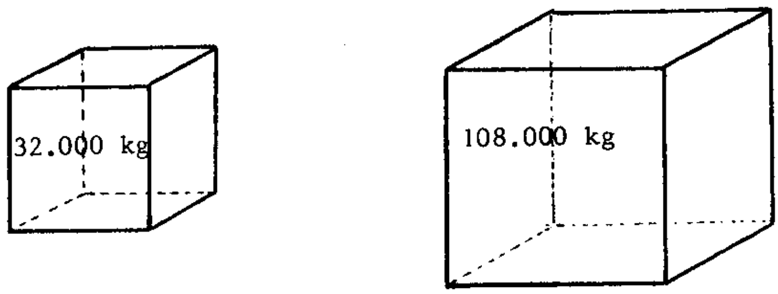


opgave 19

Vlakdiagrammen zijn moeilijk leesbaar en worden daarom weinig gebruikt. Het enige vlakdiagram dat wel regelmatig gebruikt wordt, is het cirkel diagram, dat een verdeling geeft van een bepaalde grootheid.

e) Blokdiagram

Hierbij wordt de inhoud als maatstaf genomen. Het bezwaar van de moeilijke leesbaarheid spreekt hier nog sterker, omdat vergelijken van twee gegevens, anders dan heel ruw, niet goed mogelijk is.
Om dit te ondervangen kan de absolute waarde van de gegevens vermeld worden



f) Cartogrammen

Cartogrammen zijn landkaarten, waarop men de gegevens die men wil vermelden, heeft aangebracht.

g) Beeldstatistieken

In deze soort grafieken beeldt men een verschijnsel door een typerende figuur.

h) Stroomdiagram

Dit soort diagram wordt samengesteld in die gevallen dat een bepaald gegeven uit verschillende bronnen stamt en verder gedistribueerd wordt.

Opgaven: 4,5,6,11,13 t/m 20.
Tentamens: 19-12-78: 6.
31-12-79: 6.

IV Correlatie en regressie

§ 1 Inleidende opmerkingen

a) Het gaat om het verband tussen twee verschijnselen, waarbij het ene verschijnsel invloed uitoefent op het andere. Is er sprake van dat de twee verschijnselen op te vatten zijn als twee toevalsvariabelen, dan spreekt men van correlatie. Gaat het om het verband van twee grootheden waarvan de een te kiezen is, dan spreekt men over regressie. Constateert men bij stijging van de ene variabele ook een stijging van de andere, dan noemt men dit positieve correlatie. Is de werking tegengesteld, dus dat bij stijging van de ene variabele, daling in de andere optreedt, dan noemt men dit negatieve correlatie

b) Een grafische voorstelling van correlatie kan gegeven worden in een spreidingsdiagram.

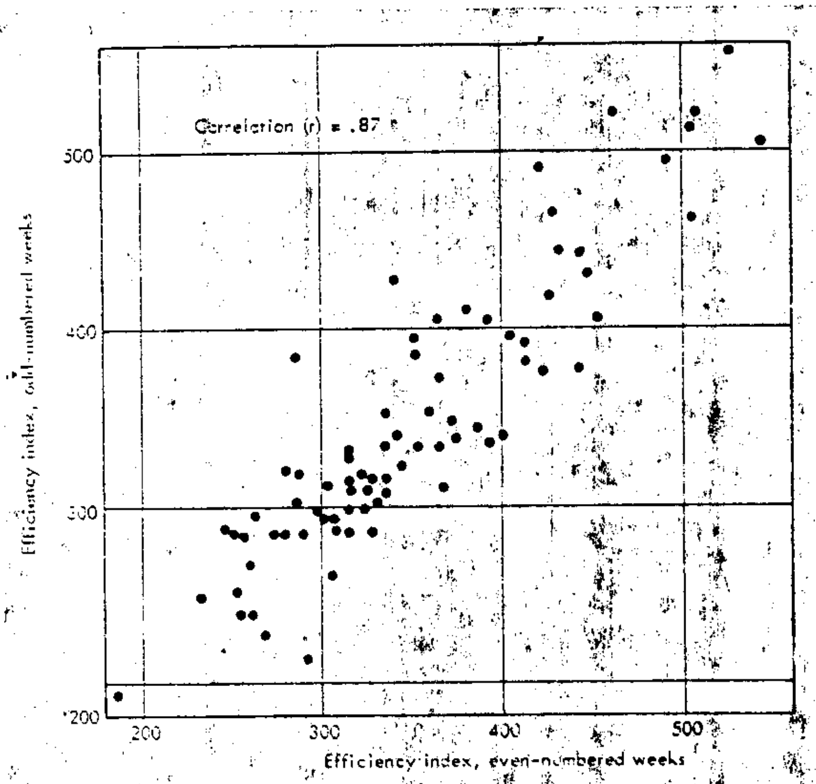
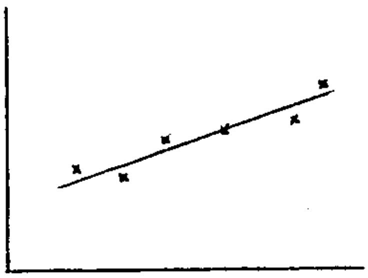


Fig. 2.14. Scattergram showing the relationship between efficiency indexes of a group of 79 unit wirers for two 5-week periods (five even-numbered weeks and five odd-numbered weeks). (Courtesy Tetronix, Inc., and Dr. Guyot Frazier.)

In sommige gevallen zien we dat punten van zo'n puntenwolk nagenoeg op een rechte lijn liggen. Deze lijn wordt de regressielijn genoemd.



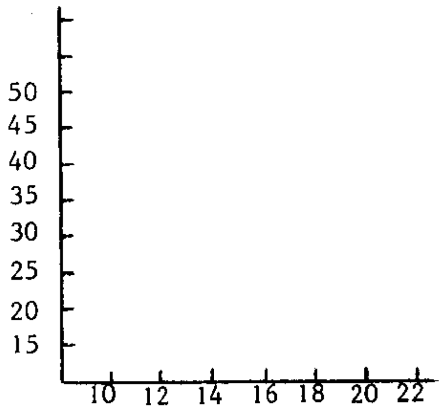
c) Wil men het verband van twee verschijnselen onderzoeken waarvoor de waarden in 10 achter volgende perioden resp. zijn.

9	10	10	9	11	13	16	21	19	11

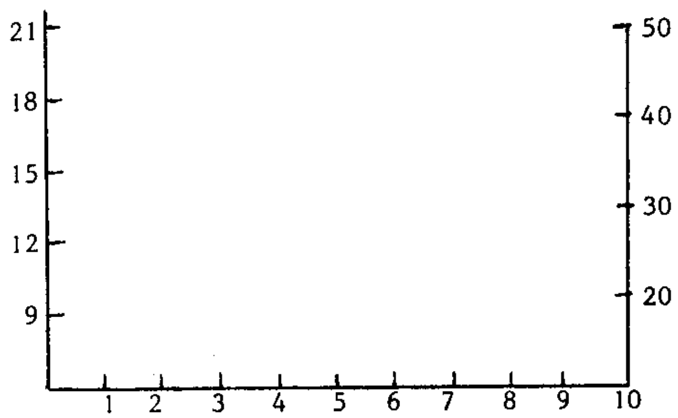
en

22	17	22	27	32	47	42	27	22	32

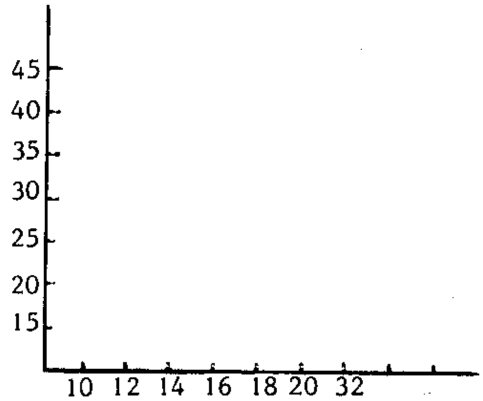
dan levert het spreidingsdiagram het volgende beeld



Zet men deze gegevens in een tijddiagram uit, dan blijkt dat de twee lijnen eenzelfde verloop te hebben. Echter is het wel zo dat de ene reeks 2 perioden achter loopt op de ander. Men spreekt hier over een time-lag of fase verschil.



Door nu de eerste waarneming van de tweede reeks te vergelijken met de derde waarneming van de eerste reeks, krijgen we het volgende spreidingsdiagram.



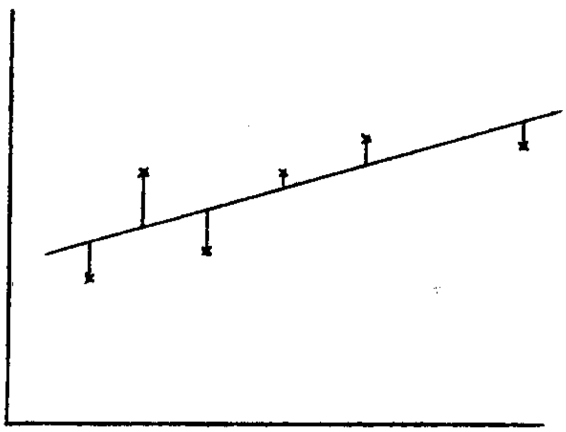
opgave 21

§2 Regressievergelijking (Methode van de kleinste kwadraten)

a) We gaan uit van een toevalsexperiment waarbij we te maken hebben met een gewone variabele x en toevalsvariabele Y. Veronderstel dat bij n gekozen waarden $x_1,\dots,x_n$ de waarden $y_1,\dots,y_n$ als waarden van Y waargenomen worden. We hebben dan de waarde paren.

$$(x_1y_1),\dots,(x_n,y_n)$$

die we op de gebruikelijke manier in het xy vlak afbeelden.



Om de vergelijking van de rechte lijn te vinden die zo goed mogelijk bij de gevonden punten aansluit, gebruiken we de methode van de kleinste kwadraten (Gauss):
De rechte lijn moet zo getrokken worden dat de kwadraten van de afstanden van de punten tot de rechte lijn (waarbij de afstand in verticale richting gemeten is), minimaal is.
Stel de vergelijking van de gevraagde rechte lijn

$$y = ax + b \tag{IV-2-1}$$

De verticale afstand van een punt (x_i, y_i) tot deze rechte lijn is

$$|y_i - b - ax_i|. \qquad \text{(IV-2-2)}$$

De som van de kwadraten van de verticale afstanden van de n punten is

$$q = \sum_{i=1}^{n} (y_i - b - ax_i)^2. \qquad \text{(IV-2-3)}$$

q is een functie van a en b. Opdat q een minimum heeft, is het nodig dat

$$\frac{\partial q}{\partial a} = 0 \text{ en } \frac{\partial q}{\partial b} = 0 \qquad \text{(IV-2-4)}$$

$$\frac{\partial q}{\partial a} = -2 \sum_{i=1}^{n} (y_i - b - ax_i) \qquad \text{(IV-2-5a)}$$

$$\frac{\partial q}{\partial b} = -2 \sum_{i=1}^{n} x_i(y_i - b - ax_i) \qquad \text{(IV-2-5b)}$$

De voorwaarden worden dus

$$\sum_{i=1}^{n} y_i - nb - a \sum_{i=1}^{n} x_i = 0 \qquad \text{(IV-2-6a)}$$

$$\sum_{i=1}^{n} x_i y_i - b \sum_{i=1}^{n} x_i - a \sum_{i=1}^{n} x_i^2 = 0 \qquad \text{(IV-2-6b)}$$

Gebruiken we $\bar{x} = \frac{1}{n}(x_1+\ldots+x_n)$ en $\bar{y} = \frac{1}{n}(y_1+\ldots+y_n)$, dan krijgen we de volgende twee vergelijkingen in a en b:

$$a\,\bar{x} + b = \bar{y} \qquad \text{(IV-2-7a)}$$

$$a \sum_{i=1}^{n} x_i^2 + n\bar{x}b = \sum_{i=1}^{n} x_i y_i \qquad \text{(IV-2-7b)}$$

Veronderstellen we dat niet alle x waarden gelijk zijn, dan is dit stelsel lineaire vergelijkingen in a en b oplosbaar:

$$b = \bar{y} - a\bar{x} \qquad \text{(IV-2-8)}$$

Dit ingevuld in (IV-2-7b):

$$a \sum_{i=1}^{n} x_i^2 + n\bar{x}\,(\bar{y} - a\bar{x}) = \sum_{i=1}^{n} x_i y_i$$

of $a(\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$.

waaruit $$a = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \qquad \text{(IV-2-9)}$$

b) Uit II.§2 weten we dat

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} [\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2]$$

ofwel $$(n - 1)s_x^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2 \qquad \text{(IV-2-10)}$$

Evenzo definiëren we nu

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \qquad \text{(IV-2-11)}$$

waaruit volgt (zie opgave 1)

$$(n - 1)\ s_{xy} = [\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}] \qquad \text{(IV-2-12)}$$

Toepassen van (IV-2-10) en (IV-2-12) in (IV-2-9) geeft

$a = \frac{s_{xy}}{s_x^2}$. (a wordt genoemd de regressie coëfficiënt).

c) Substitueren we (IV-2-8) in (IV-2-1) dan staat er

$$y = ax + \bar{y} - a\bar{x}$$

ofwel $y - \bar{y} = a(x - \bar{x})$.

d) De (steekproef) correlatie coëfficiënt wordt gedefinieerd door

$$r(\underline{x},\underline{y}) = \frac{s_{xy}}{s_x s_y}$$

Opgaven: 1,2,3,10,12,21.

Tentamens: 14-12-77: 5. 19-12-78: 4.
22-4-78: 2d,5. 31-3-79: 1b.

A.J. Bosch.
Lijst van de belangrijkste engelse termen in de kansrekening en statistiek.
(Nieuw tijdschrift voor wiskunde, jaargang 64, aflevering 2, november 1976, blz. 98-101).

accept	- aanvaarden (beter: niet verwerpen)
acceptance region	- acceptatiegebied
addition rule	- optelregel
admissible hypothesis	- toegelaten hypothese
alternative hypothesis	- alternatieve hypothese
analysis of variance	- variatieanalyse
a-posteriori probability	- aposteriori-kans
application	- toepassing
approximation	- benadering
a-priori probability	- apriori-kans
arithmetic mean	- rekenkundig gemiddelde
asymptotically unbiased	- asymptotisch zuiver
average	- gemiddelde
bell-shaped	- klokvormig
Bernoulli distribution	- bernoulliverdeling, alternatief (met kansen p en 1-p)
bias	- onzuiverheid
bias(s)ed	- onzuiver
bimodal	- tweetoppig
binomial distribution	- binomiale verdeling
central limit theorem	- centrale limietstelling
central moment	- centraal moment
characteristic function	- karakteristieke functie
chi-square(d)	- chi-kwadraat
coding	- codering
coefficient of correlation	- correlatiecoëfficiënt
coefficient of variation	- variatiecoëfficiënt
coin	- munt
combination	- combinatie
combinatorics	- combinatoriek
complementation rule	- complementregel
composite hypothesis	- samengestelde hypothese
conditional	- voorwaardelijk

confidence interval	- betrouwbaarheidsinterval
confidence limits	- betrouwbaarheidsgrenzen
consistent	- asymptotisch raak
consumer's risk	- fout van de tweede soort
contingency table	- n × m-tabel, afhankelijkheidstabel
correction for continuity	- discontinuïteitscorrectie
correlation	- correlatie
covariance	- covariantie
critical region	- kritiek gebied
critical value	- kritieke waarde
cumulative frequency	- cumulatieve frequentie
curvilinear	- kromlijnig
data	- gegevens, waarnemingen
decision function	- beslissingsfunctie
defective	- defect
degree of freedom	- vrijheidsgraad
density (function)	- dichtheid(-sfunctie)
dependence	- afhankelijkheid
dependent	- afhankelijk
derivative	- afgeleide
descriptive statistics	- beschrijvende statistiek
dichotomy	- dichotomie
die (mv: dice)	- dobbelsteen
discontinuous distribution	- discontinue verdeling
discrete distribution	- discrete verdeling
disjoint	- disjunct
dispersion	- spreiding (algemeen)
distribution	- verdeling
distribution-free	- verdelingsvrij
distribution function	- verdelingsfunctie, cumulatieve verdeling
dot diagram	- puntenwolk
double-tailed	- tweezijdig
drawing with replacement	- trekking met teruglegging
efficiency	- doeltreffendheid
efficient (estimator)	- nauwkeurigst, meest doeltreffend
enumerable	- aftelbaar

equally likely	- even waarschijnlijk
error of measurements	- meetfout
error of the first kind	- fout van de eerste soort
estimate	- schatting
estimator	- schatter
event	- gebeurtenis
expectation	- (mathematische) verwachting
expected value	- verwachtingswaarde
experimental design	- proefopzet
exponential distribution	- exponentiële verdeling
extreme value	- uiterste waarde
factorial	- faculteit
factorial moment	- factoriëel moment
failure	- mislukking
fair	- zuiver
F-distribution	- F-verdeling
figures	- getallen, waarnemingen, gegevens
fit	- aanpassen, passen in
fourfold contingency table	- 2 × 2-tabel
frequency	- frequentie
frequency distribution	- frequentieverdeling
frequency function	- verdelingsdichtheid
Gaussian distribution	- verdeling van Gauss
generating function	- genererende-, voortbrengende functie
geometric distribution	- geometrische verdeling
geometric mean	- meetkundig gemiddelde
goodness of fit	- aanpassing (mate van-)
harmonic mean	- harmonisch gemiddelde
head and tail	- kruis en munt
histogram	- histogram
hypergeometric distribution	- hypergeometrische verdeling
hypothesis	- hypothese

impossible event	- onmogelijke gebeurtenis
independent	- onafhankelijk
inequality	- ongelijkheid
interaction	- interactie
intercept	- intercept
intersection	- doorsnede
interval estimate	- intervalschatting
joint density	- simultane dichtheid
joint distribution	- simultane verdeling
law of large numbers	- wet van de grote aantallen
least squares	- kleinste kwadraten
level of sifnificance	- onbetrouwbaarheidsdrempel
likelihood	- aannemelijkheid
likelihood ratio	- aannemelijkheidsquotiënt
linear regression	- lineaire regressie
lot	- partij (producten)
lower confidence limit	- ondergrens (van een betrouwbaarheidsinterval)
marginal distribution	- marginale verdeling
mass distribution	- massaverdeling
maximum likelihood method	- methode van de meeste aannemelijkheid
maximum likelihood estimator	- meest aannemelijke schatter
mean (value)	- gemiddelde
measurement	- waarneming, meting
median	- mediaan
mode	- modus
moment	- moment
moment generating function	- momentvoortbrengende functie
most powerful test	- meest onderscheidende toets
multionimial distribution	- multinomiale verdeling
multiplication rule	- productregel
mutually exclusive events	- elkaar uitsluitende gebeurtenissen

nomogram	- nomogram
nonparametric	- parametervrij
normal distribution	- normale verdeling
normal probability paper	- normaal waarschijnlijkheidspapier
null hypothesis	- nulhypothese
observation	- waarneming
one-sided test	- éénzijdige toets
operating characteristic curve	- keuringskarakteristiek
outcome	- uitkomst, elementaire gebeurtenis
outlier	- uitschieter, uitbijter
parameter space	- parameterruimte
peakedness	- scherptoppigheid
percentile	- percentiel
permutation	- permutatie
point estimate	- puntschatting
Poisson distribution	- poissonverdeling
pool (of variances)	- samenvoegen (o.a. van varianties)
population	- populatie
power (function)	- onderscheidingsvermogen
probability	- kans, waarschijnlijkheid
probability density	- kansdichtheid
probability field	- kansveld
probability function	- kansfunctie
probability of excedance	- overschrijdingskans
probability space	- kansruimte
probability theory	- kansrekening, waarschijnlijkheidsrekening
producer's risk	- fout van de eerste soort
propagation of error	- foutenvoortplanting
proportion	- fractie
P-value	- P-waarde
quality control	- kwaliteitsbeheersing
quantile	- kwantiel
random	- toevallig, aselect, stochastisch
randomization	- verloting, randomisering
random numbers	- aselecte getallen

random sample	- aselecte steekproef
random variable, -variate	- stochast, stochastische grootheid, toevalsvariabele
random walk	- stochastische wandeling
range	- variatie-, spreidingsbreedte, range
rankcorrelation	- rangcorrelatie
rectangular distribution	- rechthoekige, homogene verdeling
regression analysis	- regressie-analyse
reject	- verwerpen
rejection region	- kritiek gebied
relative frequency	- frequentiequotiënt, relatieve frequentie
replication	- herhaling
residual	- residu, rest-
run	- run, serie
sample	- steekproef, monster
sample mean	- steekproefgemiddelde
sample size	- steekproefgrootte, -omvang
sample space	- steekproefruimte, uitkomstenruimte
sample variance	- steekproefvariantie
sampling plan	- steekproefschema
sampling with replacement	- trekken met teruglegging
scatter diagram	- puntenwolk, spreidingsdiagram
sequential analysis	- sequente analyse
significant	- significant
sign-test	- tekentoets
simple hypothesis	- enkelvoudige hypothese
single tail	- éénzijdig
skewness	- scheefheid
slope	- helling
squared error consistent	- asymptotisch nauwkeurig
standard deviation	- standaardafwijking, -deviatie
standardized variable	- gestandariseerde variabele
statistic	- statistische grootheid, steekproeffunctie
statistical inference	- verklarende statistiek
statistics	- statistiek
stepfunction	- trapfunctie
stochastically independent	- stochastisch onafhankelijk
stochastic variable	- stochast, stochastische variabele, toevalsvariabele

Student's t-distribution	- t-verdeling van Student
subevent	- deelgebeurtenis
success	- succes
sufficient	- voldoende, uitputtend
sum of squares	- kwadratensom
symmetric distribution	- symmetrische verdeling
systematic error	- systematische fout
tail	- staart (muntzijde)
tailprobability	- overschrijdingskans
t-distribution	- t-verdeling
test (of sifnificance)	- toets
test statistic	- toetsingsgrootheid
ties, tied ranks	- gelijken, gelijk rangnummers
tolerance limits	- tolerantiegrenzen
toss (coin)	- (op)werpen (munt)
treatment	- behandeling
trial	- experiment
truncated	- afgeknot
two-sided test	- tweezijdige toets
type I-error	- fout van de eerste soort
unbias(s)ed	- zuiver
uncorrelated	- ongecorreleerd
uniform distribution	- uniforme-, homogene verdeling
unimodal	- ééntoppig
union	- vereniging
universe	- universum
upper confidence limit	- bovengrens (van een betrouwbaarheidsinterval)
variable	- variabele
variance	- variantie
variate	- toevalsvariabele
variation	- variatie
Venn-diagram	- venndiagram
weighted mean	- gewogen gemiddelde

Tabel 1.1: Resultaten van een steekproef, ter grootte 50

201	206	207	203	211	202	208	204	213	198
209	208	199	210	204	218	203	206	200	206
206	201	210	202	208	195	207	209	205	214
211	200	212	213	199	205	204	197	207	203
204	207	197	205	215	206	202	210	201	211

Tabel 1.2.

1	2 3 Frequentie	4 Relatieve frequentie	5 Cumulatieve frequentie	6 Cumulatieve relatieve frequentie
195				
196				
197				
198				
199				
200				
201				
202				
203				
204				
205				
206				
207				
208				
209				
210				
211				
212				
213				
214				
215				
216				
217				
218				

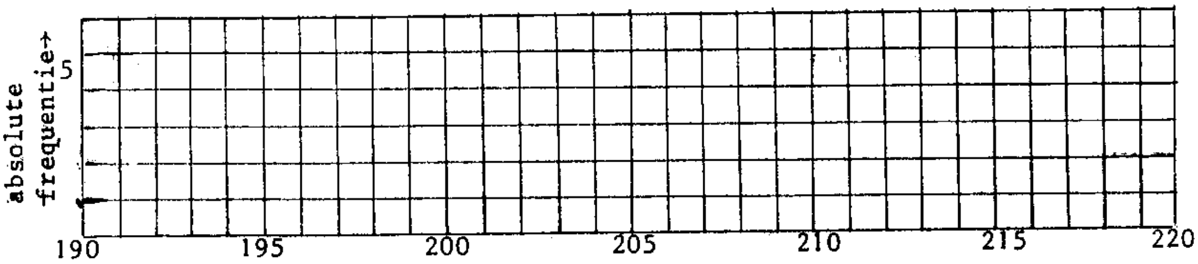


Fig. 1.1.

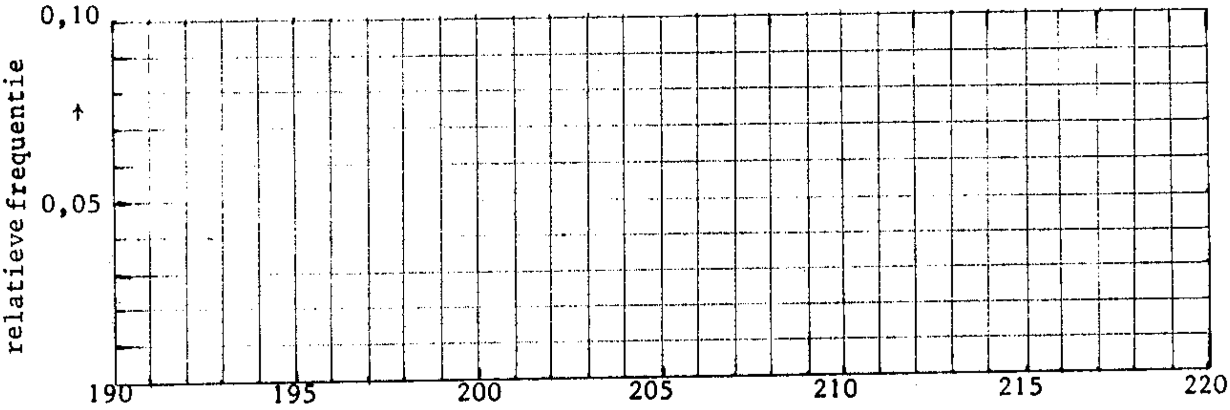


Fig. 1.2. Staafdiagram

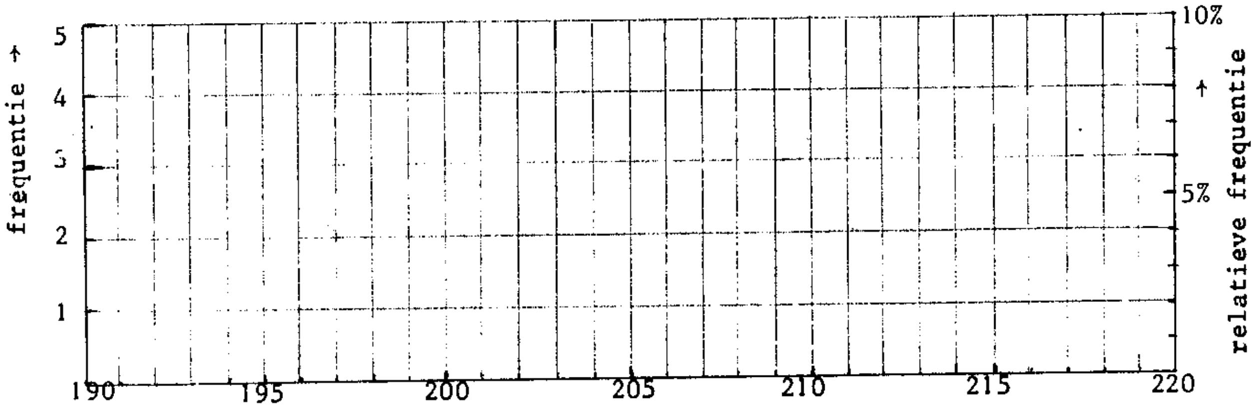


Fig. 1.3. Histogram

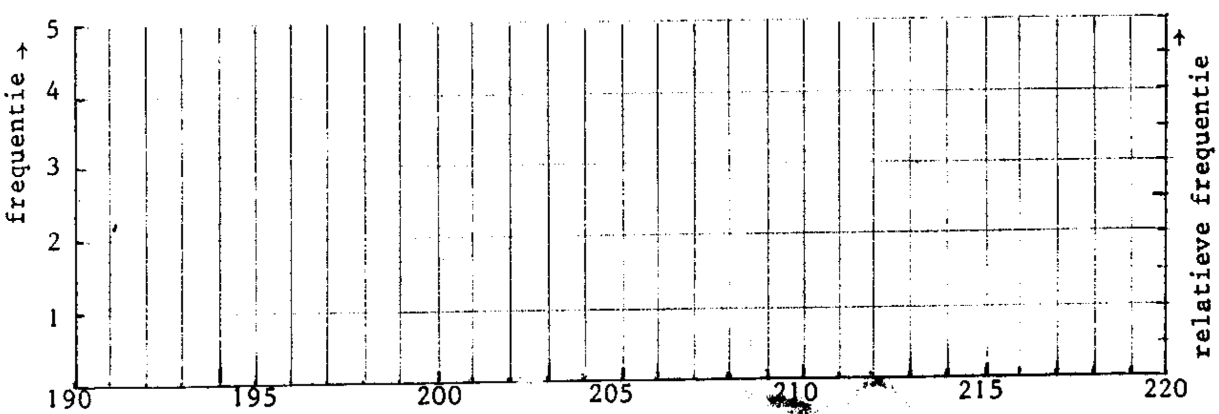


Fig 1.4. Polygoon

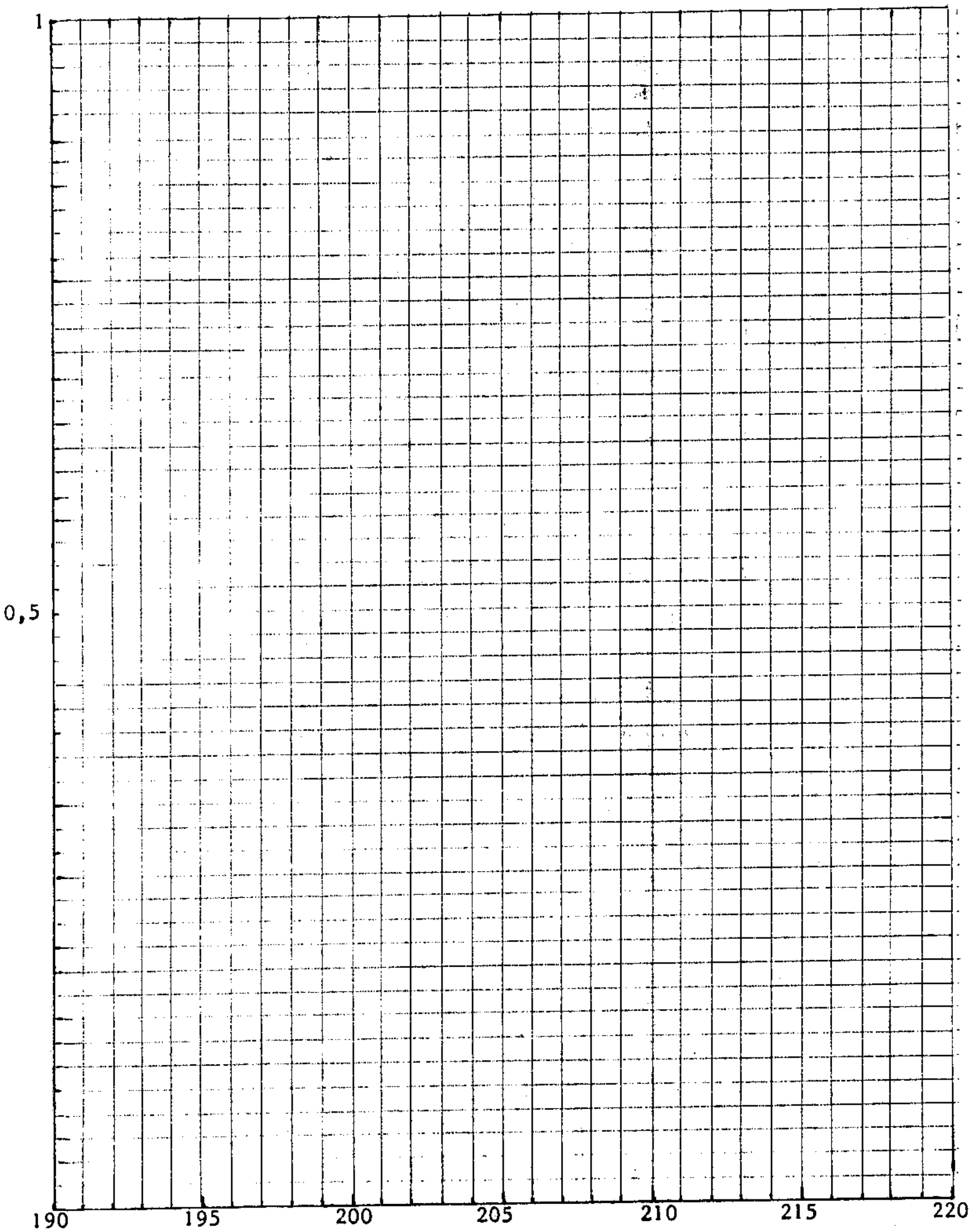


Fig. 1.5. Verdelingsfunctie

Tabel 2.1: Resultaten van een steekproef, ter grootte 125

423	435	430	458	416	441	426	439	444	427
438	440	450	436	447	437	448	434	411	449
460	412	438	419	445	420	438	429	430	432
426	443	432	443	424	435	438	452	421	442
448	453	435	446	434	427	441	443	429	437
420	420	438	422	436	440	408	435	448	454
452	424	444	427	450	448	419	440	453	417
428	437	448	430	435	425	461	433	444	436
432	442	411	440	455	425	427	442	423	446
441	433	445	438	420	439	434	414	435	430
417	448	440	429	460	437	445	450	428	440
443	423	435	453	431	455	426	464	443	465
461	445	458	414	460					

Tabel 2.2:

klassegrenzen	klassenmidden	klassefrequentie	$\tilde{f}(x)$	$\tilde{F}(x)$
407.5-412.5	410			
412.5-417.5	415			
417.5-422.5	420			
422.5-427.5	425			
427.5-432.5	430			
432.5-437.5	435			
437.5-442.5	440			
442.5-447.5	445			
447.5-452.5	450			
452.5-457.5	455			
457.5-462.5	460			
462.5-467.5	465			

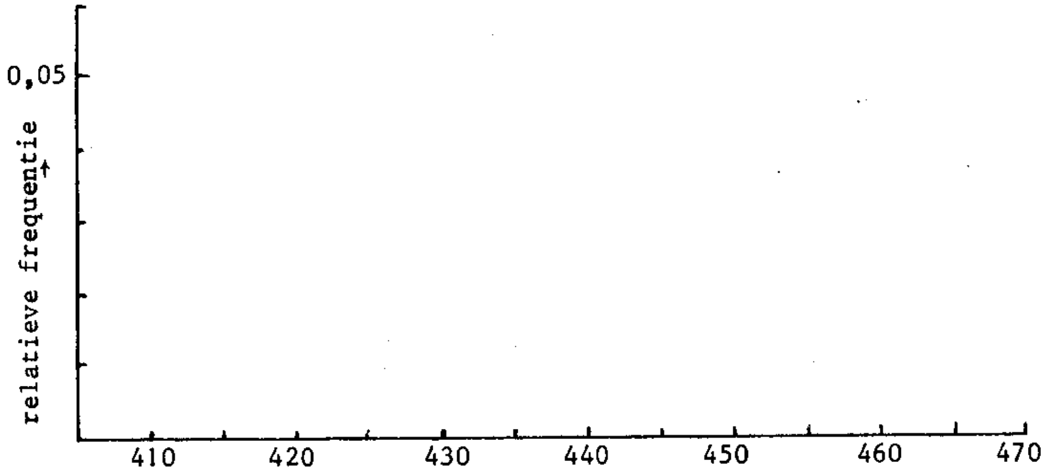


Fig. 2.1. Niet gegroepeerd. (staafdiagram)

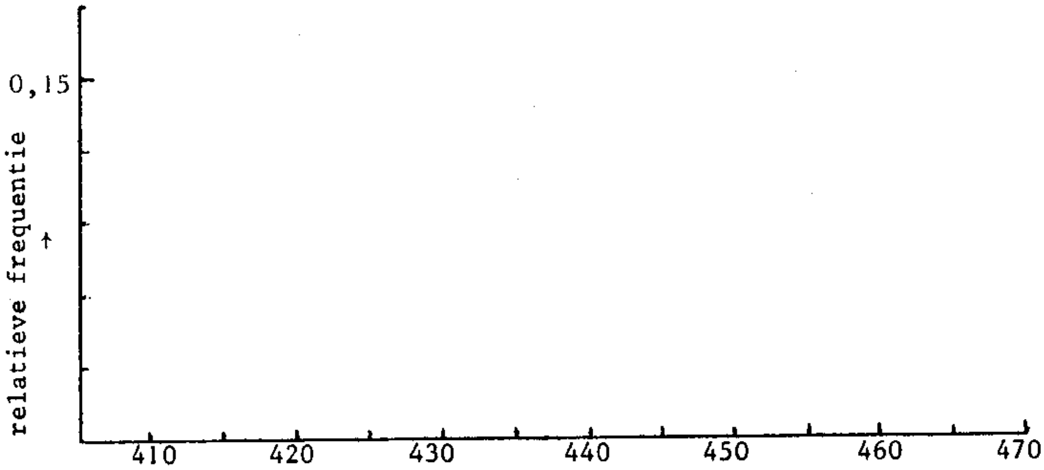


Fig. 2.2. Gegroepeerd. (histogram)

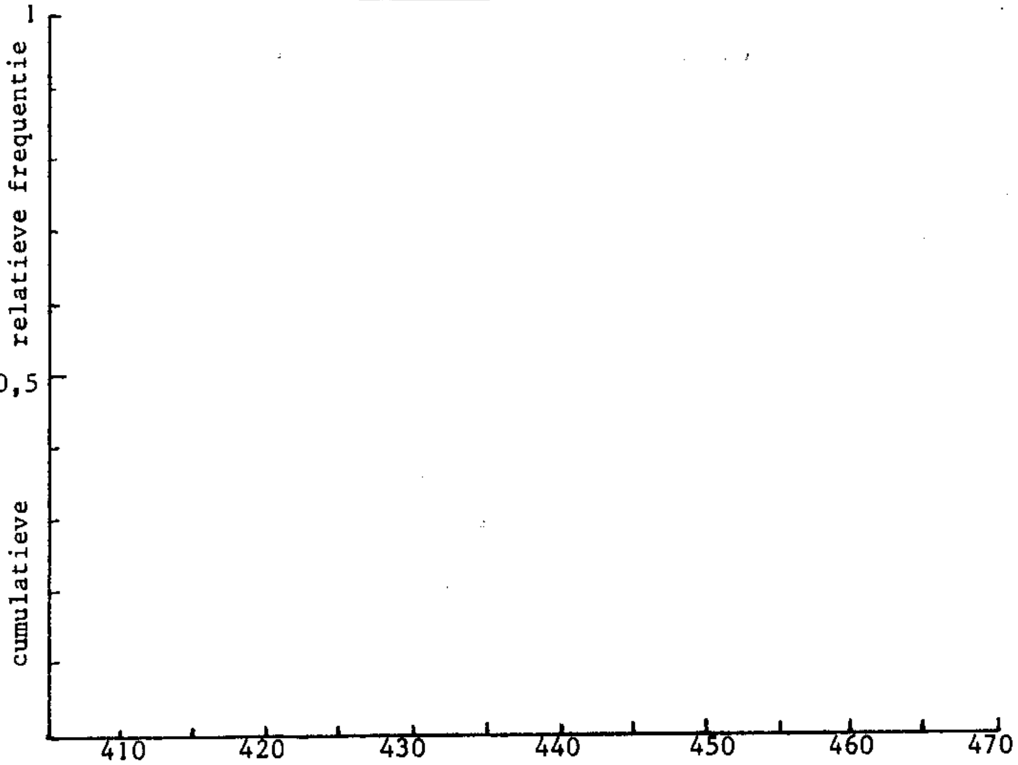


Fig. 2.3. Gegroepeerd.

Tabel 3.1.

1	2	3	4	5	6
x_j	x_j^*	$n\tilde{f}(x_j)$	$nx_j^*\tilde{f}(x_j)$	$(x_j^*)^2$	$n\tilde{f}(x_j)(x_j^*)^2$
0.195					
0.196					
0.197					
0.198					
0.199					
0.200					
0.201					
0.202					
0.203					
0.204					
0.205					
0.206					
0.207					
0.208					
0.209					
0.210					
0.211					
0.212					
0.213					
0.214					
0.215					
0.216					
0.217					
0.218					

1. $$s_{xy} := \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y}) .$$

Bewijs:

$$s_{xy} = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j y_j - \frac{1}{n} (\sum_{i=1}^{n} x_i)(\sum_{j=1}^{n} y_i)\right].$$

2. Bepaal met behulp van de kleinste kwadraten de regressielijn uit de volgende steekproeven.

a)

x	72	66	67	69	74	61	66	62	70	63
y	178	141	158	165	180	133	159	140	160	136 .

b)

x	20	30	40	50
y	50	95	150	210 .

3. Bepaal de correlatiecoëfficiënt

$$r = \frac{s_{xy}}{s_x s_y}$$

en teken het spreidingsdiagram van de volgende steekproeven:

a)

x	0,14	0,17	0,18	0,21	0,23	0,17	0,18
y	0,16	0,18	0,18	0,12	0,20	0,20	0,20 .

b)

x	80	75	45	75	70	60	45	80	80
y	80	75	75	95	100	80	45	70	90 .

4. Bruto investeringen in vaste activa 1959 in miljoenen guldens:

	Nijverheid	Overige bedr.	Overheid	Totaal
1e kw.	576	1040	259	1875
2e kw.	638	1228	380	2246
3e kw.	657	1181	407	2245
4e kw.	736	1360	436	2532

Breng deze gegevens in een samengesteld staafdiagram.

5. Produktie boter en margarine in 1000 ton.

	Boter	Margarine
1951	84	181
1952	74	187
1953	83	202
1957	76	232
1958	92	226
1959	80	239

Breng bovenstaande gegevens in een lijndiagram.

6. A. Bereken de vijf getallen, die in de laatste kolom geplaatst moeten worden.

Per 1 jan.	Bevolking (× 10.000)	Nominaal bedrag van gevestigde schulden v/h Rijk × (ƒ 1.000.000)	Schuld per hoofd van de bevolking (in gehele guldens)
1937	856	2870	
1938	864	3253	
1939	873	3212	
1940	883	3140	
1941	892	3587	

B. Stel de gegevens van het staatje, met inbegrip van de sub A verkregen uitkomsten, grafisch voor met behulp van een vlakdiagram.

7. Gegeven is dat bij een 40-tal ondernemingen in dezelfde bedrijfstak de expeditiekosten in % van het totaal van de kosten als volgt zijn:

9,2	9,5	11,5	10,8	10,8	7,3	9,0	12,5
8,1	9,2	9,9	12,0	9,2	9,4	10,0	11,4
10,1	7,9	9,4	10,5	12,5	12,4	11,7	10,3
9,3	10,0	8,5	11,0	11,7	8,8	8,1	9,7
11,5	7,2	10,0	9,8	8,7	9,1	10,9	8,7

a. Stel uit deze gegevens een frequentietafel samen.

b. Bepaal $\bar{x}$, mediaan, modus,

c. Bepaal s, range, interkwartiel-afstand.

8. De netto-winstpercentages (in percenten van de omzet) van 55 ondernemingen in een bepaalde branche van de grossierderij waren in 1943 als volgt:

6,3	0,1	1,1	3,6	-0,3	6,4	8,3	1,4	2,2	8,2
0,3	-1,8	-1,9	1,2	1,2	4,2	0,2	3,9	5,5	1,5
3,0	-1,5	2,8	7,7	2,6	3,5	2,5	5,3	1,7	4,9
9,4	5,2	5,9	5,4	3,5	4,6	2,1	1,4	5,8	1,3
-1,5	5,2	3,4	0,3	1,0	1,9	6,8	-1,0	1,6	3,8
4,2	7,0	9,0	7,7	2,9					

(N.B. Een ongunstig resultaat is aangegeven door een min-teken vóór het getal.)

a. Stel uit deze gegevens een frequentietafel samen. Motiveer de keuze van de klassebreedte.

b. Bereken de relatieve frequenties (in 1 decimaal).

c. Bepaal steekproefgemiddelde $\bar{x}$ en steekproefvariantie s^2.

9. Van 100 eieren wordt het gewicht bepaald. De uitkomsten hiervan zijn in grammen.

48	52	50	45	57	50	47	55	49	53
46	54	56	50	48	49	50	44	51	43
49	43	52	53	51	41	56	47	53	48
49	46	54	49	54	52	52	48	47	49
51	48	46	50	52	59	51	42	52	45
40	51	58	42	50	48	49	50	55	47
50	49	53	46	53	54	50	57	51	48
55	50	51	52	49	44	55	52	47	44
45	57	48	58	46	56	53	45	48	53
53	55	49	44	56	46	48	47	50	51

a. Maak van deze gegevens een frequentieverdeling.
b. Breng deze in grafiek (polygoon).
c. Bepaal $\bar{x}$, mediaan, modus.
d. Bepaal s, range, interkwartielafstand.

10. Een industriële onderneming wenst te onderzoeken, of er correlatie bestaat tussen de omzet van haar produkt per hoofd van de bevolking en de dichtheid van de bevolking (= aantal inwoners per km^2). Zij heeft daartoe voor een twaalftal gemeenten, die willekeurig gekozen zijn, de volgende gegevens verzameld uit de eigen omzetstatistiek en de bevolkingsstatistiek.

Gemeente	Omzet per hoofd v/d bevolking in glds.	Dichtheid van de bevolking × 100
1	4,3	6,8
2	4,7	5,8
3	5,7	8,3
4	0,4	0,7
5	1,1	2,2
6	3,3	3,6
7	3,3	4,4
8	3,4	5,7
9	6,2	8,8
10	4,7	6,6
11	1,1	1,3
12	1,8	3,2

Gevraagd:

1. Teken een spreidingsdiagram.
2. Is er naar Uw mening sprake van correlatie tussen beide bovenbedoelde verschijnselen? Motiveer Uw antwoord.
3. Geef, indien U dit verantwoord acht, door een lijn in het diagram de correlatie weer.
4. Welke betekenis moet aan de door U getrokken lijn worden gehecht?

11. Van drie groepen werknemers in een fabriek wordt de efficiëntie-index bepaald. (efficiëntieindex = $\frac{\text{standaardproductietijd}}{\text{feitelijke productietijd}} \times 100$).

eff.index (klassemidden)	groep 1 n = 40	groep 2 n = 35	groep 3 n = 138
40	1	3	3
50	2	3	15
60	13	15	43
70	11	6	33
80	7	2	20
90	4	3	18
100	1	2	4
110	1	0	2
120	0	1	0

Verwerk deze gegevens in een drietal histogrammen (horizontale as: efficiëntie index, verticale as: aantal werknemers).

12. In een bedrijfsvergelijkende statistiek komen in een tabel de volgende cijfers voor:

Omzetklasse × *f* 1000,--	Aantal ondernemingen	Gem. omzet per ondern. × *f* 1000,-	Gem. kosten per ondern. × *f* 1000,-
200- < 500	5	363	39
500- < 750	5	688	55
750- < 1000	5	860	63
1000- < 1500	6	1201	80
1500- < 2000	5	1737	105
2000- < 3000	7	2345	134
3000- < 5000	4	4449	244

1. Maak een spreidingsdiagram waaruit duidelijk het verband tussen omzet en kosten blijkt en geef dit verband door middel van een lijn aan.
2. In deze branche wordt 7,5% brutowinst gemaakt. Teken in dezelfde grafiek de lijn van de brutowinst.
3. Bij welke omzet wordt noch winst noch verlies gemaakt?
4. Hoe groot zal de vermoedelijke winst zijn bij een omzet van 6 miljoen gulden?

Antwoorden opgaven

2a) $y = 3{,}72x - 94{,}43$ of $y = 4x - 94$
b) $y = 5{,}35x - 61{,}00$ of $y = 5x - 61$

3a) - 0.03
0,55

6a) 335,377,368,356,402

7b) $\bar{x} = 9{,}9$; mediaan = 9,8; modi: 9,2 en 10,0.
c) $s = 1{,}4$; range = 5,3; interkwartielafstand = 1,9.

8c) $\bar{x} = 3{,}5$ $s^2 = 7{,}1$

9c) $\bar{x} = 50{,}0$ $Q_{50} = 50$; modus = 50
d) $s = 4{,}1$; range = 19 ; $Q_{75} - Q_{25} = 6$

12) 3) 10^6 gulden
4) 8.10^4 gulden

13. Geef in een lijndiagram de volgende resultaten weer:

jaar	1963	64	65	66	67	68	69	70	71	72
omzet (× ƒ 1000,-)	20	15	21	28	35	45	40	38	30	31

14. Geef in een lijndiagram de volgende resultaten weer:

jaar	1963	64	65	66	67	68	69
omzet (× ƒ 1000,-)	520	515	525	530	535	532	529

15. Geef in een curve de volgende resultaten weer:

x	0	10	20	30	40	50	60	70	80	90
y	0	4	12	24	41	63	81	90	94	100

16. Geef in een staafdiagram de volgende resultaten weer:

periode	a	b	c	d	e
uitkomst	90	100	80	60	90

17. Geef in een samengesteld staafdiagram de resultaten t.a.v. de producten I, II en III in de perioden a,b,c,d,e, zowel in absolute als in relatieve maat.

	I	II	III	totaal
a	30	40	20	90
b	40	45	25	110
c	30	40	25	95
d	45	40	30	115
e	40	35	25	100

18. Geef in twee vergelijkbare vlakdiagrammen de kosten per eenheid product weer voor:

	lonen	grondstoffen	overige kosten
80.000 producten	800.000	960.000	400.000
60.000 producten	720.000	720.000	360.000

19. Geef in twee vergelijkbare cirkeldiagrammen het volgende verloop:

	periode 1	periode 2
grondstoffen	37	40
lonen	44	40
overige kosten	19	20

20. Geef in een stroomdiagram de volgende gegevens weer van het exploitatieoverzicht sociale verzekeringen 1958 van de sociale verzekeringsfondsen.

Ontvangsten:

premies van bedrijven	1090	miljoen	gulden
premie overheid	30	"	"
bijdrage overheid	160	"	"
interest	90	"	"
premies gezinnen	1590	"	"

Uitgaven:

Uitkeringen	2580	"	"
administratiekosten	180	"	"
Overschot	200	"	"

21. Geef de volgende puntenwolk weer:

(10,22), (9,17), (11,22), (13,27), (16,32), (21,47), (19,42), (11,27).

TECHNISCHE HOGESCHOOL EINDHOVEN

Examen/tentamen Inleiding Statistiek op woensdag 14 december 1977, 9.00-12.00 uur.

De uitwerkingen van de opgaven dienen duidelijk geformuleerd en overzichtelijk opgeschreven te worden.

-.-

1. a) De resultaten van een steekproef zijn $x_1, x_2, \ldots, x_n$.
   De steekproefvariantie van deze steekproef is s^2.
   Leid af:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right].$$

   b) Is X een stochast en zijn a en b constanten, dan geldt

$$E[aX + b] = a.E[X] + b .$$

   Bewijs dit.

   c) In een doos zitten 4 gele en 3 blauwe knikkers. Men pakt er aselect 2 uit.
   X = het aantal blauwe knikkers.
   Bereken E[X].

   d) De stochast X is uniform verdeeld op het interval [0,6].
   i) Wat is de kansdichtheid?
   ii) Wat is de verwachtingswaarde?
   iii) Laat zien dat de variantie 3 is.

2. a) X is binomiaal verdeeld. Bepaal $P(X \geq 15 \mid n = 50; p = 0{,}65)$.
   b) Op een kantoor komen gemiddeld 4 telefoongesprekken per uur binnen.
   Bereken de kans dat er in twee uur minder dan 6 gesprekken binnenkomen.
   c) Z is normaal verdeeld met $\mu = 30$ en $\sigma = 5$. Bereken $P(10 \leq Z \leq 20)$.

3. Een steekproef ter bepaling van het koolstofpercentage in kolen, gaf de volgende resultaten:

87	86	85	87	86	87	86	81	77	85
86	84	83	83	82	84	83	79	82	73

   a) Bepaal het steekproefgemiddelde $\bar{x}$ en de steekproef variantie s^2.
   b) Wat zijn de mediaan en de modus van deze steekproef?
   c) Wat zijn de standaarddeviatie, de range en de interkwartielafstand van deze steekproef?

Examen/tentamen Inleiding Statistiek op woensdag 14 december 1977.

-.-

4. Beschouw de volgende steekproef

74	100	90	99	97	89	108	94	87	79
101	90	105	83	91	96	81	98	81	98
105	110	91	99	101	94	106	98	93	82
90	86	96	88	97	103	85	106	92	115
97	101	102	96	100	76	96	81	101	93

a) Maak een frequentietabel van deze steekproef.

b) Groepeer de steekproef en kies klasseintervallen van gelijke lengte, met klassemiddens 75,80,85,90,...

c) Teken het histogram van de gegroepeerde steekproef.

5. a) Bepaal met behulp van de methode der kleinste kwadraten de vergelijking van de regressielijn, uit de volgende gegevens

x	15	20	30	40	50	60	70
y	6,5	5,6	5,4	6,0	4,6	1,4	0,1

(y = zuurstofgehalte in milligrams/liter in een meer
x = afstand in meters onder het wateroppervlak).

b) Teken het spreidingsdiagram en de regressielijn.

c) Bepaal de correlatiecoëfficiënt:

$$r = \frac{s_{xy}}{s_x s_y} .$$

6. De resultaten van de afdeling BDK voor de proeftentamens wis 10 en 20 geven het volgende beeld:

cursusjaar	69/70	70/71	71/72	72/73	73/74	74/75	75/76	76/77
percentage voldoendes wis 10	85	43	76	61	41	53	31	46
percentage voldoendes wis 20	72	83	55	63	52	72	56	46

Als aan de eisen van vergelijkbaarheid is voldaan, is hieruit dan op te maken dat één van beide proeftentamens lastiger is dan de andere?
Gebruik de tekentoets. Neem $\alpha = 0,1$.

TECHNISCHE HOGESCHOOL EINDHOVEN

Examen/tentamen Inleiding Statistiek op zaterdag 22 april 1978, 9.00-12.00 uur.

De uitwerkingen van de opgaven dienen duidelijk geformuleerd en overzichtelijk opgeschreven te worden.

-.-

1. a) X is binomiaal verdeeld. Bepaal $P(10 \le X < 30 \mid n = 50; p = 0{,}55)$.
   b) Y is normaal verdeeld met $\mu = 100$, $\sigma = 2$. Bepaal $P(Y > 106)$.
   c) In een boek met 200 bladzijden komen in totaal 70 drukfouten voor. Bereken op hoeveel bladzijden resp. 0,1,2,3,4 drukfouten te verwachten zijn.

2. a) Is X een stochast en zijn a en b constanten, dan geldt

$$\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X] .$$

   b) Bewijs

$$P(|X - E[X]| \ge d) \le \frac{\mathrm{Var}[X]}{d^2} .$$

   c) Van een continue kansverdeling is de kansdichtheid

$$f(x) = \begin{cases} 0 & \text{voor } x < 0 \\ \lambda e^{-\lambda x} & \text{voor } x \ge 0 \quad (\lambda > 0) \end{cases} .$$

   Bepaal de verdelingsfunctie $F(x)$.

   d) Bewijs dat de regressielijn, met behulp van de methode der kleinste kwadraten te bepalen uit n waardenparen $(x_1,y_1),\ldots,(x_n,y_n)$, voldoet aan de vergelijking

$$y - \bar{y} = b(x - \bar{x})$$

   waarin $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ en $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

3. Een steekproef ter bepaling van een cyclusduur (in minuten), gaf de volgende resultaten

1.18	1.09	1.14	1.13	1.13	1.13	1.09	1.11	1.20	1.12
1.13	1.11	1.12	1.07	1.14	1.09	1.10	1.11	1.11	1.10

   a) Bepaal het steekproefgemiddelde $\bar{x}$ en de steekproefvariantie s^2.
   b) Wat zijn de mediaan en de modus van deze steekproef?
   c) Wat zijn de standaarddeviatie, de range en de interkwartiel afstand van deze steekproef?

Examen/tentamen Inleiding Statistiek op zaterdag 22 april 1978.

-.-

4. Beschouw de volgende steekproef

14,35	39,17	30,65	43,08	48,10	33,10
36,25	49,45	40,02	63,50	34,20	74,45
49,90	17,83	54,95	20,15	64,35	21,30
29,01	12,20	15,23	76,02	49,00	48,50
67,50	38,50	54,15	42,40	29,98	36,80
84,80	54,34	37,95	73,04	65,17	81,56
40,57	30,02	22,80	19,80	37,45	44,78
60,00	50,03	47,80	53,75	10,47	53,18
35,70	27,65	9,15	42,25	65,88	31,85
56,18	61,75	50,55	71,50	18,64	55,60
28,20	41,15	24,50	32,50	43,69	44,90
60,95	68,90	69,50	51,70	77,50	58,70
23,05	44,01	51,25	66,69	52,35	45,70
47,25	62,29	34,45	25,20	44,25	59,07
68,10	48,60	58,34	58,00	35,00	52,60
33,74	26,17	41,40	70,25	89,25	88,90
4,80	78,20	79,80	1,65	14,85	51,5
44,70	52,70	26,03	68,25	34,90	75,00
84,95					

Neem klassen met breedte 10 en klassemiddens 5,15,25,... .

a) Maak een frequentietabel van de gegroepeerde steekproef.

b) Teken het histogram van de gegroepeerde steekproef.

5. a) Bepaal met behulp van de methode der kleinste kwadraten de vergelijking van de regressielijn, uit de volgende gegevens

x	10	15	30	40	50	55	80	100
y	50	46	43	43	36	39	37	33

b) Teken het spreidingsdiagram en de regressielijn.

c) Bepaal de correlatiecoëfficiënt

$$r = \frac{s_{xy}}{s_x s_y} .$$

Examen/tentamen Inleiding Statistiek op zaterdag 22 april 1978.

-.-

6. **Y is een binomiaal verdeelde toetsingsgrootheid.**

Men wil de nulhypothese H_0: $p = 0{,}2$ toetsen tegen de alternatieve hypothese H_1: $p > 0{,}2$, met een onbetrouwbaarheidsdrempel 0,1 en een onderscheidingsvermogen van minstens 0,75 bij $p = 0{,}5$.

De voorwaarden zijn dus $P(Y \geq g \mid n;\ p = 0{,}2) \leq 0{,}1$ en $P(Y \geq g \mid n;\ p = 0{,}5) \geq 0{,}75$.

Bereken voor n = 6 t/m 12 de waarden van g die voldoen aan de eerste voorwaarde en controleer of die waarden voldoen aan de tweede voorwaarde.

Wat is Uw conclusie?

-.-

Voor de vraagstukken kunnen de volgende aantallen punten worden behaald:

Vraagstuk 1	:	2,1,2
2	:	2,3,2,3
3	:	2,1,2
4	:	3,3
5	:	3,3,2
6	:	6

Het cijfer wordt bepaald door het totaal der behaalde punten door 4 te delen.

Technische Hogeschool Eindhoven

Examen/tentamen Inleiding Statistiek op dinsdag 19 december 1978, 9.00-12.00 uur. De uitwerkingen van de opgaven dienen duidelijk geformuleerd en overzichtelijk opgeschreven te worden.

-.

1. a) De resultaten van een steekproef zijn $x_1, x_2, \ldots, x_n$.
   Deze resultaten worden gecodeerd door de betrekking

   $$x_i = x_i^* + c \qquad ; i = 1,\ldots,n.$$

   Bewijs: i) $\bar{x} = \bar{x}^* + c$.

   ii) $s^2 = s^{*2}$.

   b) Een binomiale stochast X heeft parameters n en p.
   Leid af: i) $E[X] = n.p.$

   ii) $Var[X] = np(1 - p)$.

   c) Een stochast X heeft een Poissonverdeling

   $$P(x = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad ; k = 0,1,\ldots$$

   Bewijs: $E[X] = \lambda$.

2. a) X is binomiaal verdeeld. Bepaal $P(20 \leq X \leq 60 \mid n = 100;\ p = 0,7)$.
   b) Y is normaal verdeeld. Bepaal $P(Y > 100 \mid \mu = 50;\ \sigma = 10)$.
   c) Z is uniform verdeeld op het interval $[-2,3]$. Bepaal $P(Z > 4)$.

3. Teken het histogram van de volgende (gegroepeerde) steekproef:

klasse interval	aantal
0.28-0.31	10
0.32-0.35	100
0.36-0.39	157
0.40-0.43	125
0.44-0.47	81
0.48-0.51	18
0.52-0.55	8
0.56-0.59	0
0.60-0.63	1
	500

Bereken tevens het steekproefgemiddelde en geef een spreidingsmaat. Wat valt aan te merken op de klasse-intervallen?

Examen/tentamen Inleiding Statistiek op dinsdag 19 december 1978, 9.00 - 12.00 uur.

-.

4. Bepaal met behulp van de methode der kleinste kwadraten de vergelijking van de regressielijn, uit de volgende gegevens:

x	0	1	1	2
y	0	0	-2	2

5. Piet beweert dat 75% van de leerlingen van zijn school aanhanger is van club A en 25% aanhanger van club B.
Jan trekt de mening van Piet ernstig in twijfel.
Zij nemen een steekproef van 50 leerlingen. Piet krijgt gelijk als van deze 50 het aantal supporters van club A ligt in het interval [32,42].
a) Hoe groot is de kans dat Piet gelijk krijgt als hij ook heeft?
b) Hoe groot is de kans dat Piet gelijk krijgt, terwijl de fractie aanhangers van club A maar 50% is?
c) Wat versta je onder het 95%-betrouwbaarheidsinterval?
d) Bepaal het 95%-betrouwbaarheidsinterval voor de fractie supporters van club A op school bij een steekproefresultaat van 38 aanhangers, van club A.

6. Door het CBS werden de volgende gegevens gepubliceerd betreffende de inleggingen bij de spaarbanken in Nederland voor 1947.

(in miljoenen guldens)

maand	inleggingen		terugbetalingen	
	RPS	particuliere spaarbanken	RPS	particuliere spaarbanken
januari	20,5	31,2	26,2	24,2
februari	21,8	22,8	20,3	18,3
maart	18,9	26,1	24,6	22,7
april	23,8	22,4	29,3	25,2
mei	19,8	25,5	28,9	25,6
juni	20,9	24,7	26,2	23,0

a) Bepaal, hoeveel het saldo-tegoed van de spaarders bij de Rijkspostspaarbank en hoeveel dit saldo bij de particuliere spaarbanken tengevolge van inleggingen en terugbetalingen in de opvolgende maanden januari tot en met juni 1947 meer of minder was dan op 1 januari 1947.
b) Ontwerp een grafiek, waarin het verloop van de onder a gevonden verschillen tot uitdrukking komt.

TECHNISCHE HOGESCHOOL EINDHOVEN

Examen/tentamen Inleiding Statistiek op zaterdag 31 maart 1979, 9.00-12.00 uur.

De uitwerkingen van de opgaven dienen duidelijk geformuleerd en overzichtelijk opgeschreven te worden.

-.-

1. a) De stochast X is normaal verdeeld met kansdichtheid

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

i) Wat weet U van $\int_{-\infty}^{+\infty} f(x)dx$?

ii) Leid af $E[X] = \mu$.

b) In de definitie van de correlatiecoëfficiënt, $r(\underline{x},\underline{y}) = \frac{s_{xy}}{s_x s_y}$, wordt de teller gedefinieerd door

$$s_{xy} := \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y}) .$$

Bewijs:

$$s_{xy} = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j y_j - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right)\right] .$$

2. Experiment: het eenmaal werpen van 2 zuivere dobbelstenen.
Noem Y de som van beide worpen.

a) Wat is de uitkomstenverzameling en wat zijn de kansen op deze uitkomsten?

b) Bepaal de verwachtingswaarde van Y.

c) Bepaal de standaarddeviatie van Y.

3. a) X is normaal verdeeld. Bereken $P(X \le 1{,}5 \mid \mu = 2, \sigma = 3)$.

b) Y is binomiaal verdeeld. Bereken $P(10 \le Y < 20 \mid n = 50, p = 0{,}55)$.

c) Z is Poisson verdeeld. Bereken $P(Z > 10 \mid \lambda = 5)$.

4. Iemand wil de zuiverheid van een dobbelsteen toetsen. In een eerste reeks van 40 worpen, komt 4 maal een zes boven; in een tweede reeks van 400 worpen, komt 40 maal een zes boven.
Toets in beide gevallen de hypothese van zuiverheid. Kies $\alpha = 0{,}05$.
Geef zowel een berekening als een schatting m.b.v. nomogrammen.

Examen/tentamen Inleiding Statistiek op zaterdag 31 maart 1979.

-.-

5. De resultaten van het tentamen Inleiding Statistiek van 19-12-1978 waren als volgt:

cijfer:	1	2	3	4	5	6	7	8	9	10
aantal:	0	3	3	10	16	32	30	9	3	0

a) Bereken het gemiddelde en de steekproefvariantie.

b) Bereken de mediaan en de interkwartielafstand.

c) Geef de resultaten weer in een histogram.

d) Waarom is hier een histogram beter op zijn plaats dan een staafdiagram?

6. Uit een historisch overzicht van Robeco worden de volgende gegevens ontleend:

	kapitaal	vermogen	dividend
1933	ƒ 1.103.000,--	ƒ 1.103.023,--	-
1939	13.670.000,--	14.876.182,--	2,50
1950	32.100.000,--	60.915.362,--	4,25
1960	150.200.000,--	681.374.866,--	10,71
1970	665.500.000,--	2.788.914.225,--	15,89
1977	1.265.000.000,--	4.134.383.406,--	17,54

Verwerk deze gegevens in één grafische voorstelling.

-.-

Voor de vraagstukken kunnen de volgende aantallen punten worden behaald:

Vraagstuk 1 : 1,4,4
2 : 4,3,3
3 : 2,2,2
4 : 7
5 : 3,2,2,2
6 : 9

Het cijfer wordt bepaald door het totaal der behaalde punten door 5 te delen.

TECHNISCHE HOGESCHOOL EINDHOVEN

Examen/tentamen Inleiding Statistiek op zaterdag 29 maart 1980, 9.00-12.00 uur.

De uitwerkingen van de opgaven dienen duidelijk geformuleerd en overzichtelijk opgeschreven te worden.

-.-

1. Voor het tentamen Inleiding Statistiek op 19-12-78 slaagden 32 van de 106 deelnemers niet. Op 31-3-79 waren er 28 deelnemers. In de hieronder volgende lijst van puntenparen stelt het eerste getal het aantal behaalde punten voor, gescoord op 19-12-78, het tweede getal dat op 31-3-79 behaald werd, achtereenvolgens voor alle kandidaten die aan beide zittingen deelnamen.
   Teken hiermee de puntenwolk (op de horizontale as het resultaat van 19-12-78)
   (16,37), (12,33), (15,35), (16,28), (16,41), (10,24), (13,27), (16,38), (20,33), (19,35), (19,34), (18,31), (19,33), (21,41), (19,40), (18,44), (18,39), (7,37), (19,41), (16,35), (20,18), (15,23), (20,32).
   Geef een schema waarmee de correlatiecoëfficiënt berekend kan worden (berekening dus niet noodzakelijk).

2. Een stochast X heeft een Bernoulliverdeling:

$$P(X = 1) = p, \; P(X = 0) = q, \quad (p+q = 1) \; .$$

   a) Bewijs $E[X] = p$.

   b) Bewijs $\sigma^2[X] = pq$.

3. a) X is binomiaal verdeeld. Bepaal $P(X \leq 8 \mid n = 20, p = 0{,}65)$.

   b) Y is normaal verdeeld. Bepaal $P(42 < Y < 63 \mid \mu = 50; \sigma = 7)$.

   c) Z is Poisson verdeeld. Bepaal $P(Z = 9 \mid \lambda = 9)$.

4. Om het 100 maal werpen met een dobbelsteen te simuleren, maken we gebruik van tabel 1: toevalsgetallen.

   a) Lees deze tabel kolomsgewijs en tel het aantal malen dat resp. 1, 2, 3, 4, 5 en 6 voorkomt. Ga door tot het totale aantal 100 is.

   b) Bereken de relatieve frequenties, steekproefgemiddelde en steekproefvariantie.

   c) Zet de resultaten in een staafdiagram uit.

   d) Als de stochast X het geworpen ogental voorstelt, bereken dan ook $E[X]$ en $Var[X]$.

Z.O.Z.

Examen/tentamen Inleiding Statistiek, zaterdag 29 maart 1980

-.-

5. Een helderziende beweert de uitkomst van een worp met een zuivere munt in acht van de tien gevallen goed te kunnen voorspellen. Een sceptisch journalist wil het bestaan van deze B(oven) N(atuurlijke) G(ave) onderzoeken, en besluit tot het volgende onderzoek:
Met een zuivere munt wordt twintig maal geworpen. Het aantal goede voorspellingen wordt genoteerd. Bij 15 of meer goede voorspellingen wordt het bestaan van B.N.G. geaccepteerd; bij 14 of minder wordt aangenomen dat de helderziende naar de uitkomst raadt.

   a) Bereken de kans dat de helderziende ten onrechte een B.N.G. wordt toegekend.

   b) Bereken de kans dat de helderziende zijn B.N.G. wordt onthouden.

6. a) Stel uit de gegevens van bijgaande enquête een frequentietabel samen. Groepeer daarbij de getallen in de volgende 11 klassen:
   minder dan 150,150-<250,250-<350,...,950-<1050,1050 en meer.

   b) Bereken de relatieve frequenties in 3 decimalen.

   c) Bereken het steekproefgemiddelde, door gebruik te maken van bovenstaande groepering. Geef daarbij de overwegingen hoe de eerste en de elfde klasse in de berekening zijn meegenomen.

-.-

Voor de vraagstukken kunnen de volgende aantallen punten worden behaald:

Vraagstuk 1 : 6
          2 : 3,3
          3 : 2,2,2
          4 : 1,2,1,2
          5 : 3,3
          6 : 3,3,4.

Het cijfer wordt bepaald door het totaal der behaalde punten door 4 te delen.

Enquête van 200 Nederlandse werknemers van 16 jaar en ouder met volledige werkweek naar verdiend bruto weekloon, gehouden in oktober 1977.

378,12	334,73	368,85	250,23
560,46	422,95	444,72	394,95
528,23	1319,42	556,80	1489,36
233,19	377,93	492,74	449,04
428,23	686,16	483,60	499,51
682,41	467,80	276,90	217,53
334,73	681,65	415,97	431,55
480,34	535,36	572,70	741,12
571,31	446,21	384,36	301,74
522,95	636,14	518,54	744,39
859,45	113,73	605,39	357,31
352,79	435,22	548,66	469,22
979,01	473,78	421,32	1651,96
415,21	375,36	652,67	421,44
422,01	535,22	472,56	268,37
546,21	896,01	247,61	598,99
576,15	465,07	593,03	493,15
673,59	654,97	467,71	746,53
287,00	486,25	539,61	416,09
470,02	606,45	888,11	519,33
465,02	337,58	417,43	649,33
515,21	524,85	956,12	385,11
808,84	439,61	469,56	287.93
550,02	582,83	642,54	593,15
324,76	544,72	327,85	497,62
481,02	750.19	585,33	1028,07
521,32	356,65	714,19	473,90
756,44	909.80	420,51	976,82
351,07	515,97	387.65	207,54
435,36	451,61	472,48	377,99
614,08	569,88	517,43	420,40
1098,39	224,34	760.32	864.27
463,11	636,21	520,51	502.05
288,01	419,33	402,05	549,04
586,64	700,04	702,51	550.06
1715,38	500,70	524,14	324,04
479,18	672,09	590,75	1188,13
741,17	467,73	239,38	531,55
418,54	585,06	380,77	461,24
1170,60	705,36	664,04	641,01
464,22	497,72	631.72	498,19
175,38	581,81	400,70	521,44
260,58	343,00	451,59	654,70
448,66	391,93	1287,78	276,90
635,09	881,73	520,40	584,53
486,15	516,09	828,05	504,37
522,01	644,03	318,45	628,64
570,39	570,57	660,79	424.14
354,73	424,85	607,54	763,25
632,40	266,44	384,82	593,10

Antwoorden tentamens

14 december 1977:

1c) $E[x] = \frac{6}{7}$

1d:i) $f(x) = \frac{1}{6}$ op $[0,6]$, $f(x) = 0$ elders

ii) $E[x] = 3$

2a) $P(x \geq 15 \mid n = 50;\ p = 0{,}65) = 1{,}0000$

b) 0,1912

c) $P(10 \leq Z \leq 5) = 0{,}0228$

3a) $\bar{x} = 83{,}3$; $s^2 = 13{,}16$

b) $Q_{50} = 84$; modus = 86

c) $s = 3{,}62$; range = 14 ; $Q_{75} - Q_{25} = 4$

5a) $y = -0{,}11x + 8{,}63$

c) $r = 0{,}90$

6) H_0: beide tentamens even lastig, niet verwerpen.

22 april 1978:

1a) 0,7138

b) 0,0013

c) 141,49,9,1,0.

3a) $\bar{x} = 1{,}12$; $s^2 = 10{,}53.10^{-4}$

b) $Q_{50} = 1{,}11$; modus: 1,11 en 1,13

c) $s = 0{,}03$; range = 0,13; $Q_{75} - Q_{25} = 0{,}03$

5a) $y = -0{,}17x + 48{,}87$

c) $r = -0{,}92$

6) Voor $n = 6$ is $g = 3$; voor $n = 7,8,9$ is $g = 4$; voor $n = 10,11,12$ is $g = 5$.
Voor $n = 6,7,8,9,10,11$ wordt met bijbehorende g niet aan de tweede voorwaarde voldaan, voor $n = 12$ met $g = 5$ wel.
Conclusie: om op de voorgeschreven manier te toetsen, moet de steekproef grootte minstens 12 zijn.

19 december 1978

2a) 0,0125

b) 0

c) 0

3) $\bar{x} = 0,396$;

$s = 0,050$

4) $y = x - 1$

5a) 0,9260

b) 0,0325

d) $0,61 < p < 0,87$

6a) RPS: meer +5,7; -4,2; -9,9; -15,4; -24,5; -29,8

Post: meer +7,0; +11,5; +14,9; +12,1; +12,0; +13,7 (in miljoen guldens).

31 maart 1979

1a) i) 1

2a) $\mathcal{U} = \{2,3,4,5,6,7,8,9,10,11,12\}$.

36P(Y = y):	1	2	3	4	5	6	5	4	3	2	1

b) 7

c) $\frac{1}{6}\sqrt{210}$.

3a) 0,4325

b) 0,0116

c) 0,0137

4) $H_0: p = \frac{1}{6}$; $H_1: p \neq \frac{1}{6}$.

$n = 40$: Met Poissonverdeling. H_0 accepteren; $0,035 < p < 0,24$.

$n = 400$: Met normale verdeling. H_0 verwerpen; $0,07 < p < 0,14$.

5a) $\bar{x} = 6.00$; $s^2 = 2,133$.

b) $Q_{50} = 6$; $Q_{75} - Q_{25} = 2$.

---

29 maart 1980

3a) 0,0196 ; b) 0,8415 ; c) 0,1317.

4d) 350 ; $\frac{3500}{12}$.

5a) 0,0207 ; b) 0,1958.

6a) 1,7,18,42,55,35,18,6,6,4,8 ; c) 550.