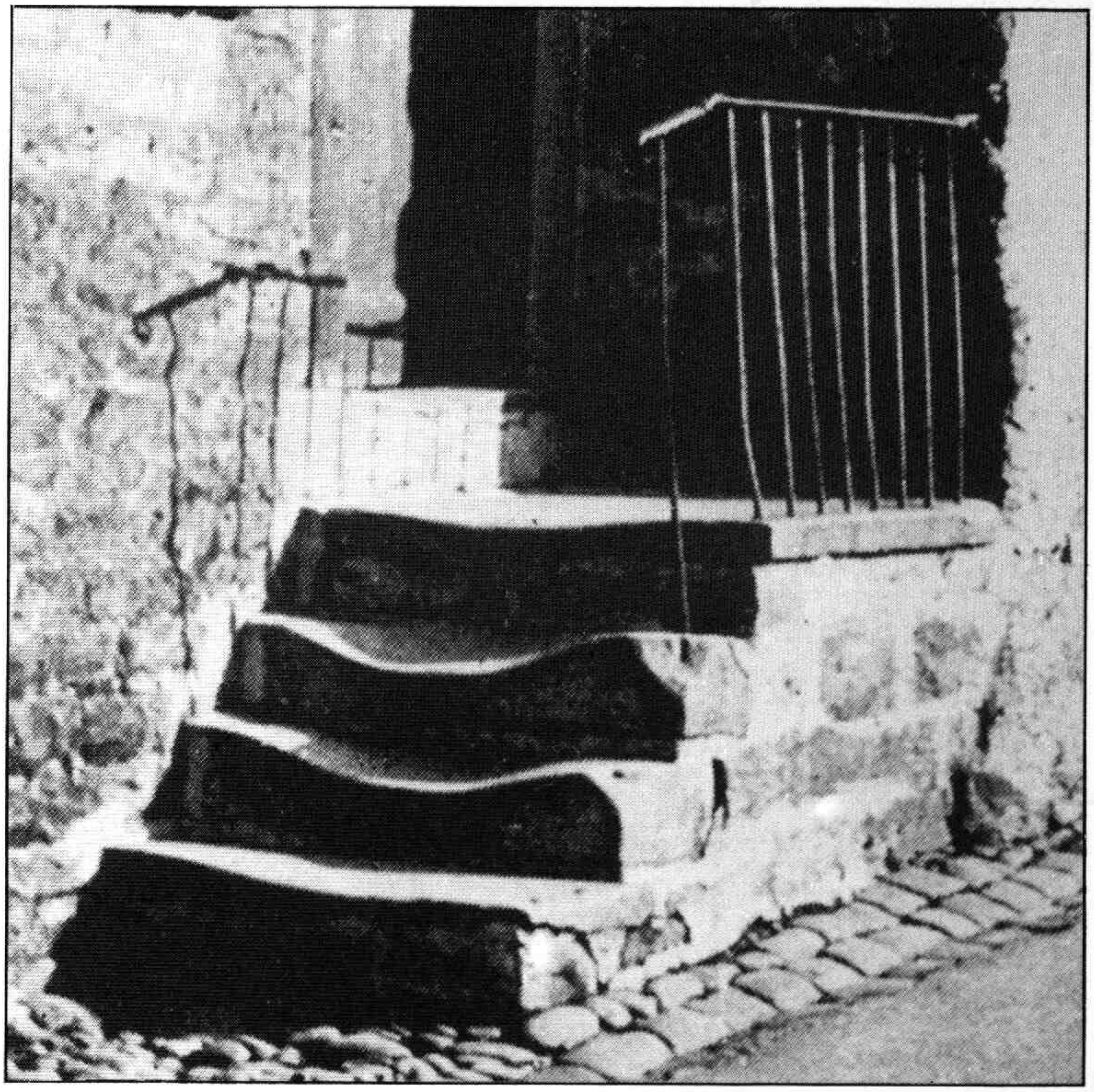
De normale verdeling

https://hdl.handle.net/1874/10246



DE NORMALE VERDELING

DE NORMALE VERDELING

Tiberdreef 4 - 3561 GG Utrecht

DE NORMALE VERDELING

Een produkt ten behoeve van de experimenten in het kader van de Herverkaveling Eindexamenprogramma's Wiskunde I en II V.W.O.

Samenstelling: Martin Kindt
Jan de Lange Jzn

Vormgeving: Ellen Hanepen

Utrecht, december 1983.

INHOUDSOPGAVE

1 VERDELINGEN VAN LICHAAMSAFMETINGEN

In 1947 werd in opdracht van N.V. Magazijn 'De Bijenkorf' een statistisch onderzoek verricht naar de lichaamsafmetingen van de Nederlandse vrouwen.
Dit onderzoek had ten doel de fabricage van damesconfectiekleding op een hoger peil te brengen: beter passende kleding en minder geldverspilling aan vermaakkosten. Ten behoeve van dit onderzoek werden bij een aantal willekeurig gekozen vrouwelijke klanten van De Bijenkorf vijftien lichamelijke kenmerken gemeten, o.a. lichaamslengte en gewicht, taille, bovenwijdte, heupbreedte, kniehoogte. De proefpersonen werkten mee op vrijwillige basis en werden door De Bijenkorf na afloop van de metingen onthaald op koffie en gebak.

Gedurende twee weken werden zo de gegevens van 5001 vrouwen boven de 18 jaar verzameld, hoofdzakelijk afkomstig uit de drie grote steden.
Met betrekking tot de lichaamslengte werd de volgende frequentieverdeling vastgesteld:

Lengte (cm)	*Frequentie*	*Lengte* (cm)	*Frequentie*
136	-	161	321
137	-	162	313
138	-	163	290
139	1	164	294
140	1	165	291
141	4	166	261
142	3	167	222
143	2	168	184
144	8	169	157
145	4	170	167
146	17	171	109
147	18	172	86
148	32	173	65
149	51	174	62
150	54	175	29
151	71	176	49
152	78	177	28
153	115	178	17
154	149	179	5
155	170	180	10
156	208	181	6
157	208	182	3
158	231	183	1
159	301	184	2
160	302	185	-
		186	1
		Totaal	5001

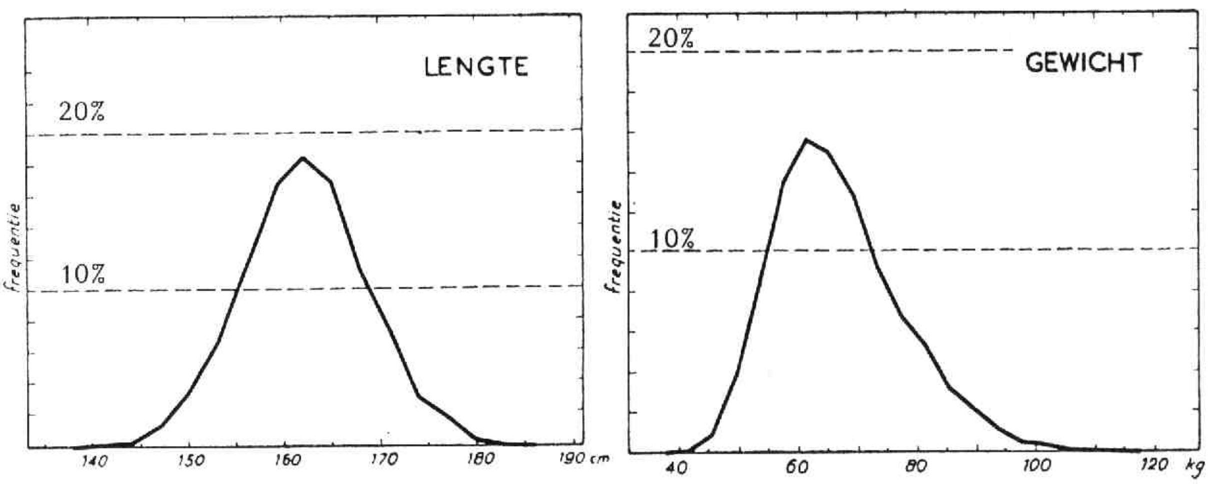
» 1. a. Wat is de mediaan van deze verdeling?

b. Waaraan kun je zien dat de gemiddelde lengte ongeveer gelijk is aan de mediaan?

» 2. Er is vastgesteld dat ongeveer 2½% van die 5001 vrouwen een lengte heeft die meer dan twee keer de standaarddeviatie (2 × S.D.) groter is dan de gemiddelde lengte (de 'uitschieters' naar boven). Hoe kun je uit dit gegeven de S.D. schatten?

≫ 3. Controleer of het aantal 'uitschieters naar beneden' (d.w.z. vrouwen met een lengte die meer dan 2 × S.D. onder het gemiddelde ligt) ook $2\frac{1}{2}$% van het totaal bedraagt.

Van de 5001 vrouwen werd ook het lichaamsgewicht gemeten.
Hieronder zie je de *relatieve-frequentiepolygonen* van lengte en gewicht van de betrokken personen. De horizontale streepjeslijnen geven het 10%- resp. 20%-niveau aan.



≫ 4. Wat is een opmerkelijk verschil tussen de beide grafieken?

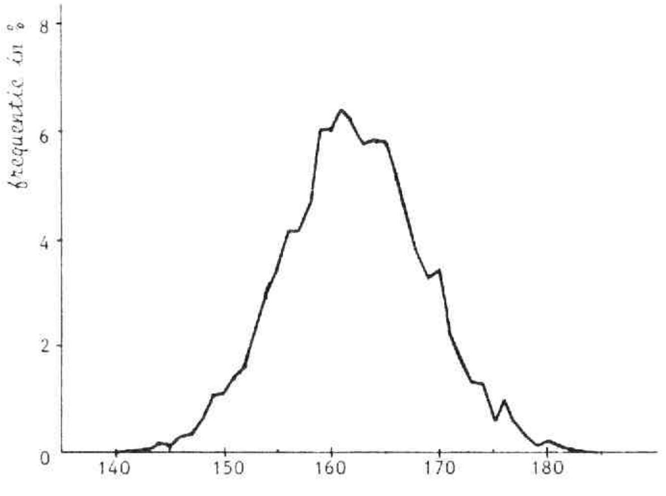
≫ 5. a. Bij de frequentiecurve van de lengte is de klassebreedte niet, zoals in de tabel, 1 cm. Hoe kun je dat zien aan die grafiek?

b. Uit de hoogte van de top kun je afleiden hoeveel cm de klassebreedte is. Hoe?

≫ 6. In de tweede grafiek kun je aflezen dat de modus van de gewichtsverdeling in de steekproef rond de 62 kg ligt. (Eigenlijk moet je spreken van 'modale klasse').
Hoeveel % van de vrouwen waren naar schatting zwaarder dan het modale gewicht van 62 kg?

≫ 7. Voor het gemiddelde gewicht werd gevonden 66,80 kg en de S.D. bedroeg 10,91 kg.
Schat uit de grafiek hoeveel % van de vrouwen meer dan 1 S.D. zwaarder of lichter was dan het gemiddelde gewicht.

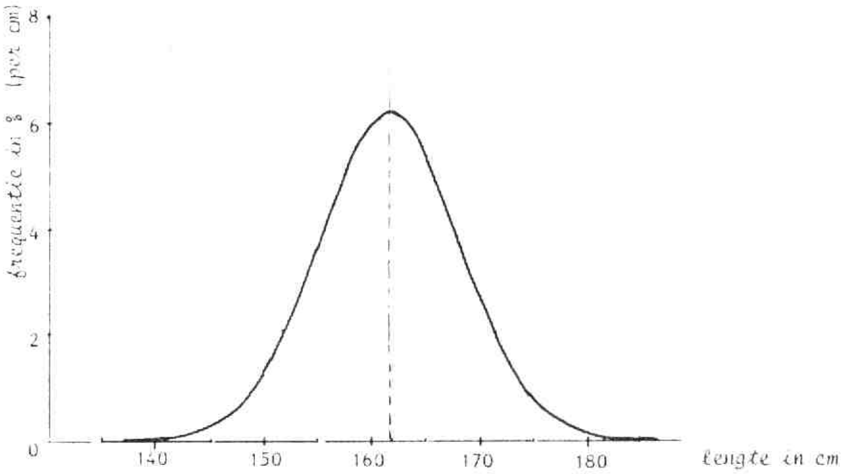
De vorm van een frequentiepolygoon bij een gevonden verdeling hangt af van de gekozen klassebreedte! Bij de frequentiepolygonen op blz. 3 is een klassebreedte van resp. 3 cm en 4 kg genomen.
Als je de grafiek van de lichaamslengte op basis van een klassebreedte van 1 cm tekent, wordt het verloop wat grilliger. Ook wordt de symmetrie wat minder duidelijk.



Neem je daarentegen een klassebreedte van 5 cm en zorg je ervoor dat de gemiddelde lengte (162 cm) het midden van een klasse is, dan krijg je een vrijwel volmaakt symmetrische grafiek.

≫ 8. Maak die grafiek. Vermeld langs de verticale as de frequentiepercentages.

Bij een goed op elkaar afstemmen van steekproefomvang en klassebreedte kun je een 'mooie' grafiek verwachten. Had men in 1947 de steekproef flink uitgebreid (tot bijv. 50.000 vrouwen), dan zou de frequentiecurve bij een klassebreedte van 1 cm er vermoedelijk 'glad' en symmetrisch hebben uitgezien. In het algemeen kun je verwachten dat bij vergroting van de steekproef en verkleining van de klassebreedte, de grafiek van de lichaamslengte zal gaan lijken op een 'klokvormige' kromme. (Zie pag. 5).



De klokvormige kromme geeft een wiskundig model van de verdeling van de Nederlandse vrouwen (anno 1947) naar lichaamslengte. Dit model noemt men de *normale verdeling*.
De naam 'normale verdeling' is enigszins misleidend, omdat zij suggereert dat elke verdeling die hiervan afwijkt abnormaal zou zijn, hetgeen zeker niet het geval is. De verdeling van het gewicht van de 5001 vrouwen (figuur op blz. 3) ziet er vanwege de asymmetrie al aanzienlijk minder 'normaal' uit.

≫ 9. Bij het Bijenkorf-onderzoek werd ook de voetlengte gemeten.
Resultaat:

voetlengte in cm	*frequentie*	*voetlengte in cm*	*frequentie*
20	4	25	848
20,5	6	25,5	532
21	11	26	303
21,5	33	26,5	179
22	68	27	76
22,5	213	27,5	20
23	419	28	13
23,5	582	28,5	7
24	718	29	1
24,5	968		5001

Onderzoek of deze verdeling lijkt op de normale verdeling.

≫ 10. Dezelfde vraag voor de verdeling van de 5001 vrouwen naar leeftijd.

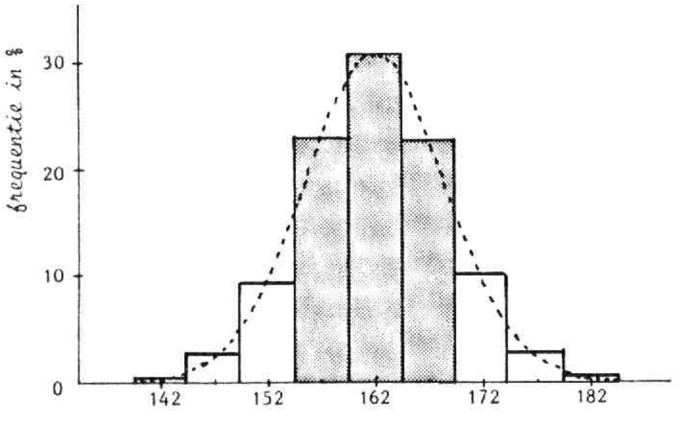
leeftijd	*frequentie*
18 - 27	1186
28 - 37	947
38 - 47	1183
48 - 57	1048
58 - 67	515
> 67	122
	5001

≫ 11. De Nederlandse vrouw van 1983 is gemiddeld aanzienlijk langer dan die van 1947.
Neem aan dat het verschil 8 cm is.

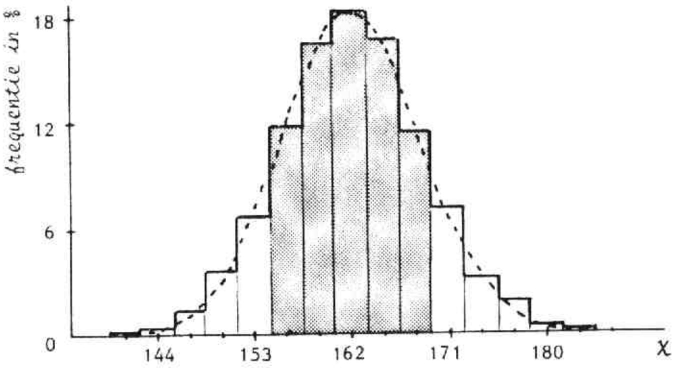
a. Schets in één figuur de (normale) verdeling van de Nederlandse vrouwen naar lichaamslengte in de jaren 1947 en 1983. Neem aan dat de spreiding in 1983 even groot is als in 1947.

b. Hoe zal de normale curve veranderen bij een meer gespreide verdeling van de lichaamslengte (bijv. met een twee keer zo grote S.D.)?

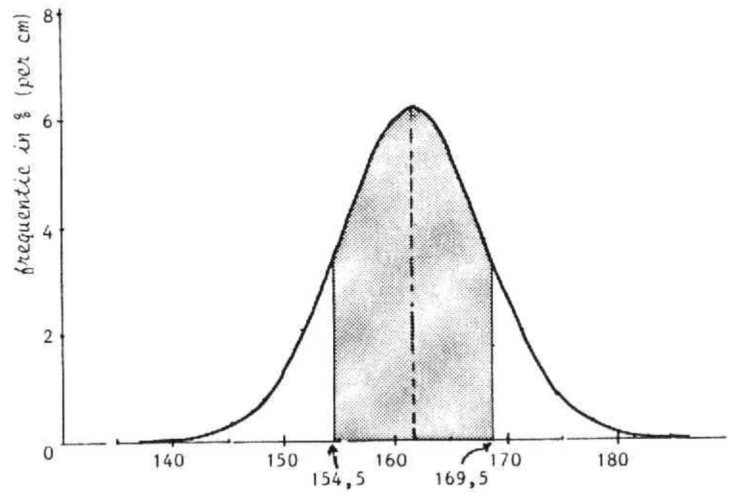
In het voorbeeld van de lichaamslengte werd de normale kromme gevonden door 'glad strijken' van de frequentiecurve. We hadden ook uit kunnen gaan van het *histogram* van de relatieve frequenties.



Bij een grotere steekproef en een kleinere klassebreedte worden de treden van de trap kleiner en benadert de kromme de contouren van het histogram.



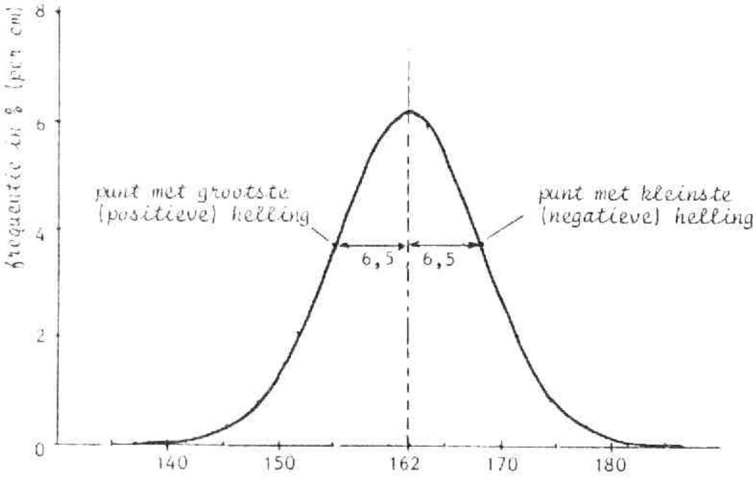
De oppervlakte van het gebied onder de kromme tussen $x = 154,5$ en $x = 169,5$ geeft de relatieve frequentie van het aantal vrouwen met lengte tussen 154,5 en 169,6 cm.



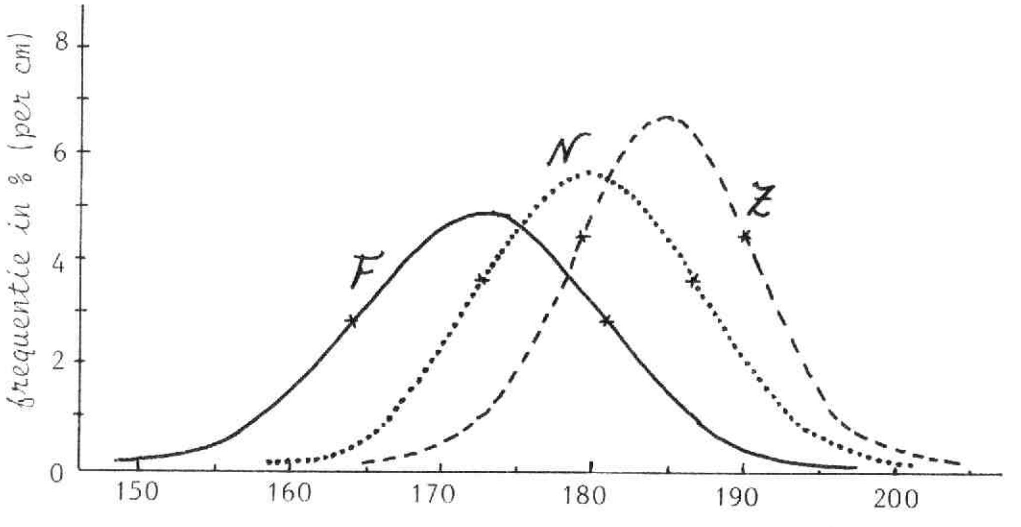
≫ 12. a. Uit de Bijenkorf-steekproef blijft dat ongeveer 75% van de vrouwen een lengte heeft tussen 154,5 en 169,5 cm.
Hoe kun je zien dat die 75% aardig klopt met de laatste grafiek?

b. Welk frequentiepercentage komt overeen met de oppervlakte van het *gehele* gebied onder de kromme?

De standaarddeviatie (in het geval van de 5001 vrouwen bedroeg die 6,5 cm) kun je in de grafiek terugvinden als de afstand van de symmetrie-as tot de punten op de kromme waar die kromme het steilst is. (De zgn. *buigpunten* van de grafiek).



≫ 13. Hieronder zie je de (fictieve) verdeling van Nederlandse, Franse en Zweedse 18 jarige mannen naar lichaamslengte. De buigpunten zijn aangegeven met een kruisje.



a. Lees uit de figuur af hoe groot de gemiddelde lengte van de resp. Nederlandse, Franse en Zweedse man is.

b. Lees ook de S.D. voor elk van de drie verdelingen af.

c. Verklaar: hoe smaller de 'klok', hoe hoger de 'top'.

Biometrische kenmerken van mensen (en ook van dieren en planten) beantwoorden dikwijls aan de normale verdelingen. De Belgische wiskundige Quételet (1796 - 1874) was de eerste die hierover gepubliceerd heeft (L'anthropométrie ou mesure de différentes facultés de l'homme).
Een voorwaarde voor het normaal zijn van de verdeling is wel dat de groep waarvan je een bepaald kenmerk meet, redelijk 'homogeen' is.

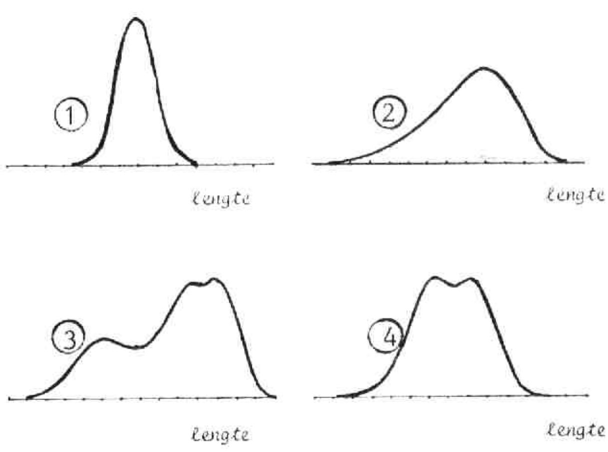
≫ 14. Als je bijv. de lengte van alle brugklassers en eindexamenkandidaten van je school meet en de resultaten in één grafiek verwerkt, heb je te maken met een niet-homogene groep.
Hoe zal de grafiek van die lengte-verdeling er globaal uitzien?

≫ 15. Bij het Bijenkorf-onderzoek bleken lichaamslengte, ruglengte, mouwlengte, handomvang, lengte middelvinger, kniehoogte, voetlengte en - breedte redelijk normaal verdeeld te zijn.
Van de diktematen (gewicht, taille, heupomvang e.d.) gold het niet of in veel mindere mate. Wat voor verklaring kun je hiervoor bedenken?

≫ 16. De (globale) grafieken van de lengteverdeling van:

a. alle Nederlanders;
b. alle gehuwde Nederlanders;
c. alle standaard-matrassen (voor volwassenen);
d. alle leden van gezinnen waarvan de ouders jonger zijn dan 30 jaar.

Welke hoort bij welke? Verklaar.

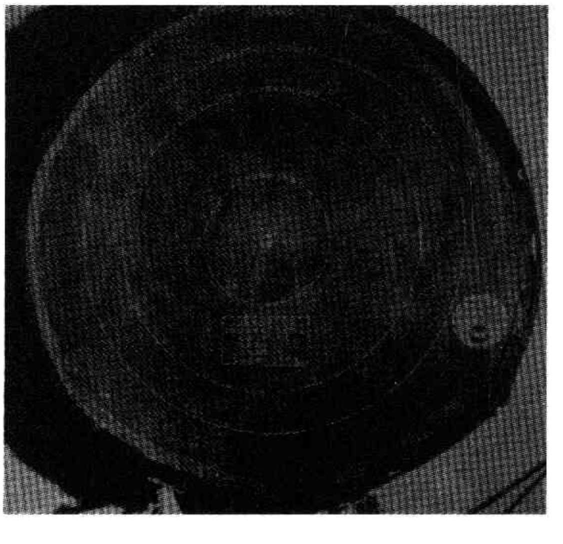


CONTINU EN DISCREET

Lichaamslengte, gewicht e.d. zijn voorbeelden van (stochastische) variabelen.
In de statistiek kennen we zowel *continue* variabelen als *discrete* variabelen.

Voorbeelden van continue variabelen:	*Voorbeelden van discrete variabelen:*
- lichaamsgewicht; - bloeddruk; - voetlengte.	- aantal bloembladen van een roos; - aantal korrels in een aar; - schoenmaat.
Bij continue variabelen kan het verschil tussen twee waarden 'willekeurig' klein zijn.	Discrete variabelen veranderen sprongsgewijs.
Het 'bereik' van een continue variabele is een *interval*.	Het 'bereik' van een discrete variabele bestaat uit 'geïsoleerde' punten.
Voetlengte (volwassen mens): 10 cm 40 cm	Schoenmaat: 37½ 38 38½ 39 39½ 40 40½

≫ 17. In onderstaande tekst is sprake van de waarden van een aantal variabele kenmerken van een boom.
Welke van die variabelen zou je continu, welke discreet willen noemen?



Deze Douglasspar werd in 1381 geplant en in 1922 geveld: hij telde precies 540 jaarringen, was 76 m hoog, 2 m dik, woog 28.000 kilo en leverde 40 m^3 hout op.

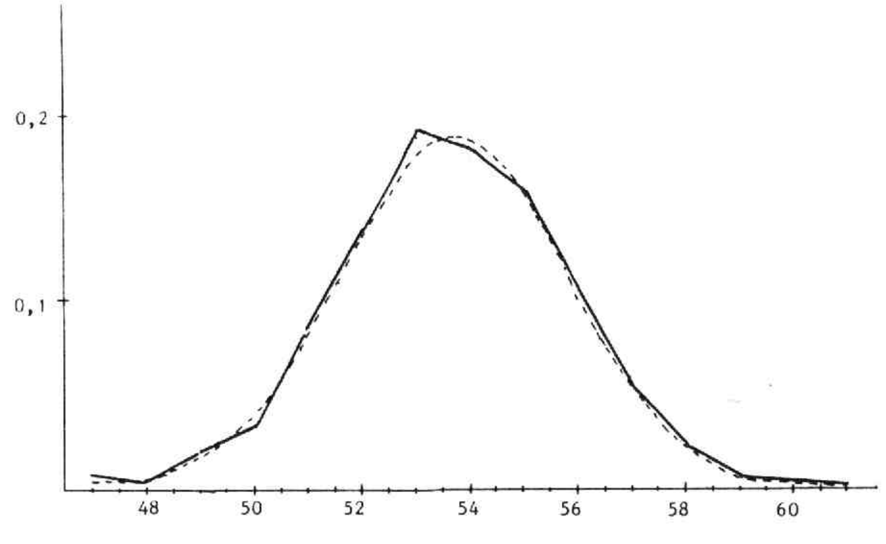
De normale verdeling geeft een continu model dat vaak ook bij discrete variabelen wordt gebruikt.

Voorbeeld:

Het aantal vinnerven bij bot.
Van 703 vissen werd het aantal vinnerven geteld.
Het aantal varieerde van 47 tot 61. De verdelingsgrafiek benadert de normale kromme heel dicht.


bot

aantal vinnerven	*frequentie*
47	5
48	2
49	13
50	23
51	58
52	96
53	134
54	127
55	111
56	74
57	37
58	16
59	4
60	2
61	1

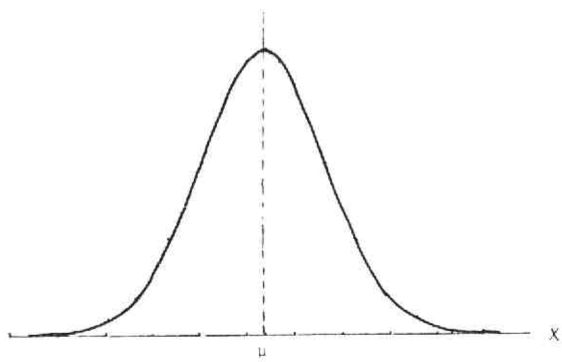


≫ 18. Schat uit de grafiek het gemiddelde en de S.D. van het aantal vinnerven bij bot.

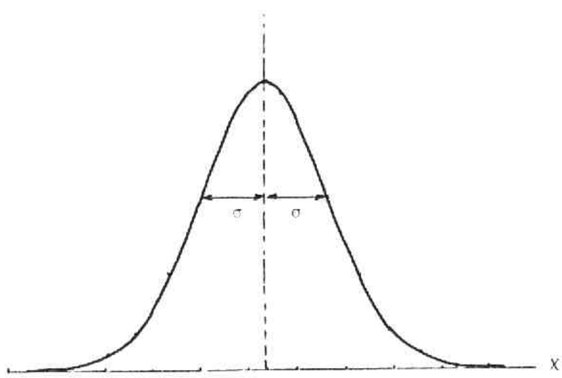
SAMENVATTING

Voor de normale verdeling geldt:

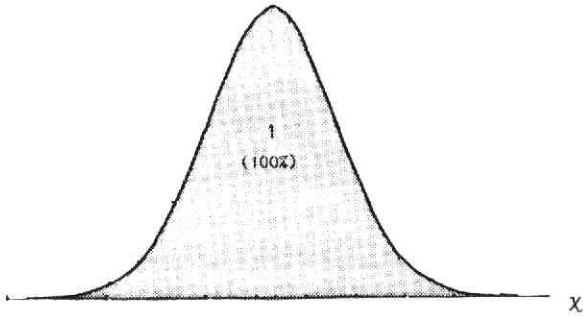
- de verdelingscurve is symmetrisch en klokvormig;
- de as van symmetrie gaat door het gemiddelde (μ);
- de afstand van symmetrie-as tot de punten waar de helling van de kromme het steilst is (de zgn. buigpunten) is gelijk aan de standaarddeviatie (σ);
- de totale oppervlakte van het gebied onder de kromme is gelijk aan 1 (of 100%).



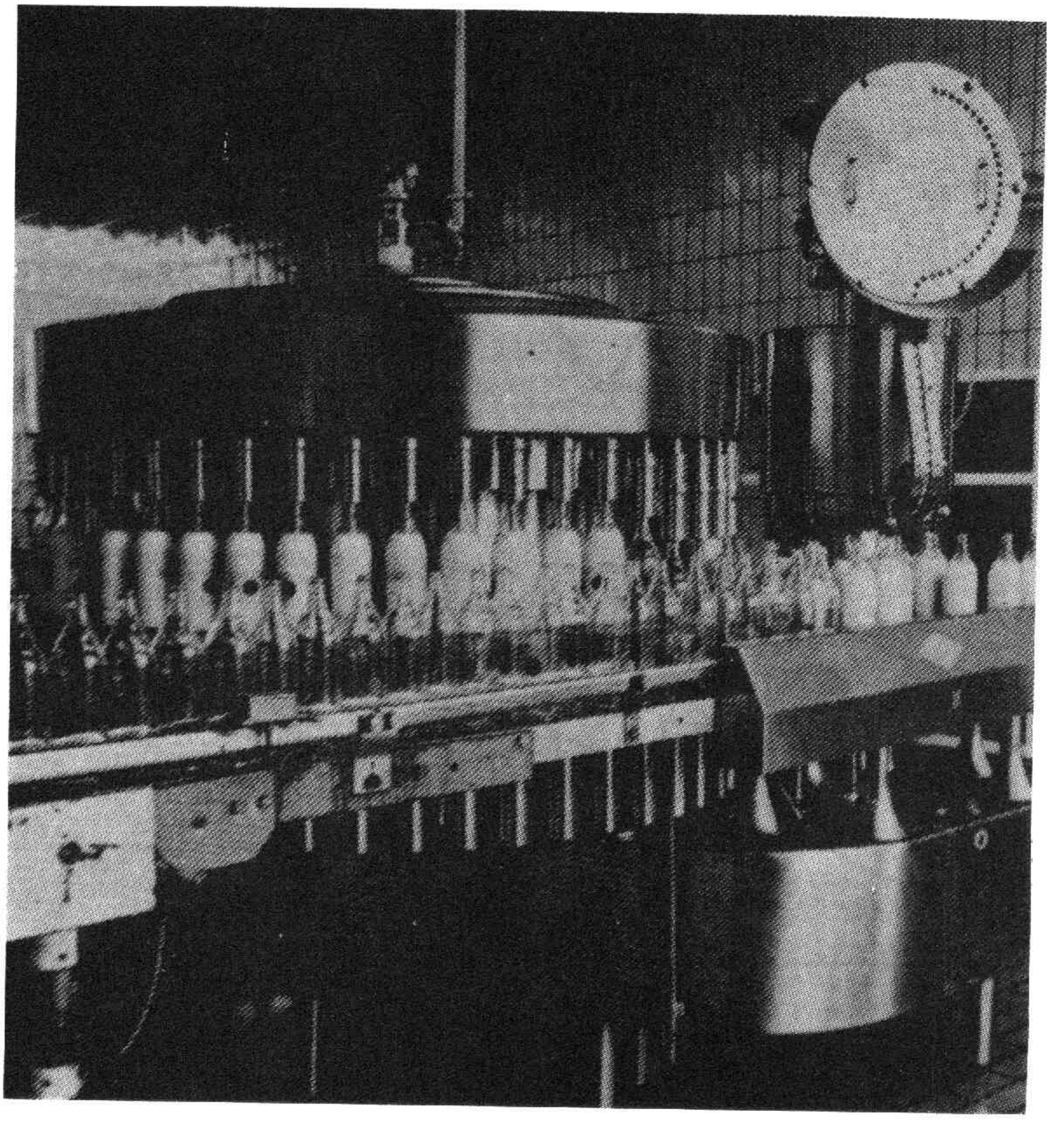
gemiddelde = μ



S.D. = σ



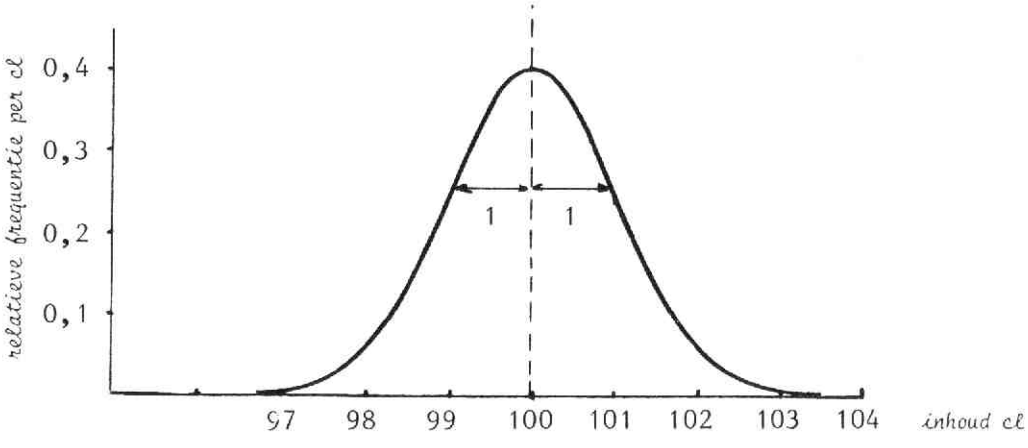
totale oppervlakte
= *100%*



Volautomatische installatie voor het vullen van melkflessen.

2

DE STANDAARD-NORMALE VERDELING



Een fles melk zal vrijwel nooit precies één liter melk bevatten. Als de vulmachine 'eerlijk' is ingesteld, kun je hoogstens verwachten dat de gemiddelde inhoud van de flessen (ongeveer) 1 liter is.
Bij een controle, waarbij de werkelijke inhoud in centiliters nauwkeurig werd gemeten, werd inderdaad een gemiddelde van 100 cl gevonden. De gemeten inhouden bleken normaal verdeeld te zijn met een standaarddeviatie van 1 cl. (Zie grafiek).

≫ 19. De controleur was alleen geïnteresseerd naar de afwijkingen van de voorgeschreven inhoud. Daarom noteerde hij voor elke fles uit de steekproef als meetresultaat het verschil met 100 cl (positief/negatief als de fles meer/minder bevatte).

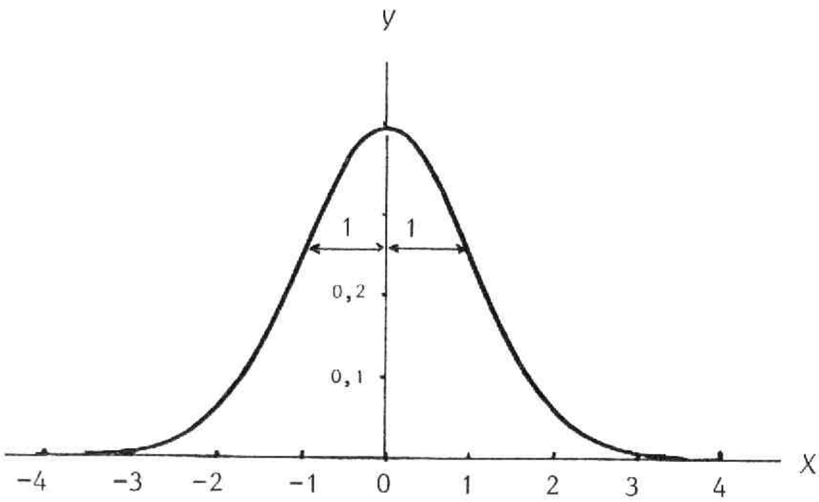
a. Wat was het gemiddelde van zijn meetresultaten?

b. En hoe groot was de standaarddeviatie?

De normale verdeling in opgave ≫ 19 is een hele bijzondere.
Het bijzondere zit hem hierin dat het *gemiddelde* van alle meetresultaten gelijk is aan *nul* en de *standaarddeviatie* gelijk is aan *één*.

Kortweg: $\mu = 0$ en $\sigma = 1$.

Men spreekt in dit geval van de *standaard-normale verdeling*.



De bijbehorende standaard-normale kromme beantwoordt aan een 'keurige' formule:

$$y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$$

De 'constanten' π en e zijn bekende getallen in de wiskunde.
In 5 decimalen afgerond: $\pi = 3{,}14159$
$e = 2{,}71828$

≫ 20. De standaard-normale kromme is symmetrisch t.o.v. de lijn $x = 0$. Hoe kun je dat 'zien' aan de formule $y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$?

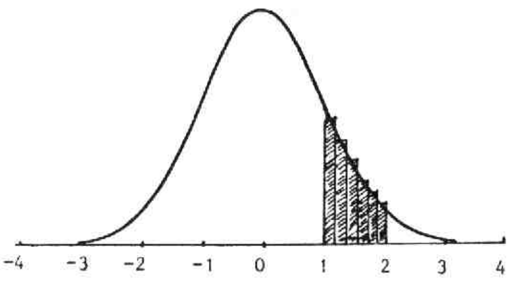
≫ 21. Bereken uit de formule een aantal punten van de standaard-normale kromme (neem $x = 0, 1, 2, 3, 4$) en controleer of je resultaat klopt met het plaatje.

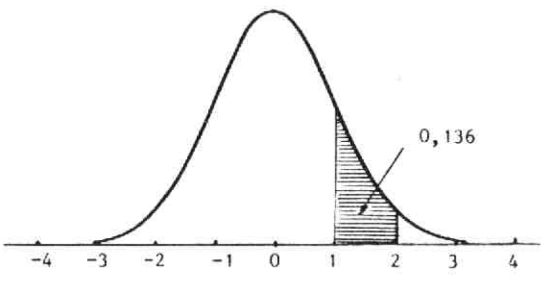
Opmerking:

Bij toenemende positieve waarde van x wordt de bijbehorende y-waarde kleiner en zelfs 'willekeurig klein'.
Voor $x = 5$ bijv. is y bij benadering $1\frac{1}{2}$ miljoenste.

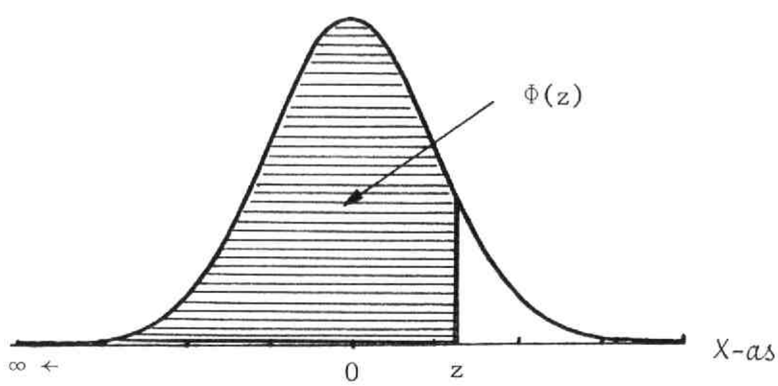
De x-as is een *asymptoot* van de standaard-normale kromme.
De afstand van de kromme tot de x-as is rechts van $x = 3,5$ en links van $x = -3,5$ zó miniem, dat het lijkt of de grafiek daar samenvalt met de x-as.

In hoofdstuk 1 heb je gezien dat een relatieve frequentie bij een normale verdeling overeenkomt met de oppervlakte van een gebied 'tussen de kromme en de x-as'.
Met behulp van de formule $y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$ kan de oppervlakte van zo'n gebied in bijv. 3 decimalen nauwkeurig worden berekend. Het principe bij zulke berekeningen is dat het gebied in smalle strookjes wordt verdeeld; elk van die strookjes kan worden benaderd door een rechthoekje. De som van de oppervlakten van die rechthoekjes is bij benadering de oppervlakte van het gebied.
Zo geldt: de oppervlakte van het gebied tussen $x = 1$ en $x = 2$ is gelijk aan 0,136 (of 13,6%).





Omdat het toepassen van deze methode nogal bewerkelijk is, is er een tabel gemaakt van oppervlakten onder de standaard-normale kromme. Die tabel geeft de oppervlakte van het gebied links van de lijn 'x = z' (z > 0).
Die oppervlakte wordt meestal genoteerd als $\Phi(z)$. (Φ is de Griekse hoofdletter 'phie').



z	opp. = $\Phi(z)$	z	opp. = $\Phi(z)$	z	opp. = $\Phi(z)$
0,00	0,500	1,00	0,841	2,00	0,977
0,05	0,520	1,05	0,853	2,05	0,978
0,10	0,540	1,10	0,864	2,10	0,982
0,15	0,560	1,15	0,875	2,15	0,984
0,20	0,579	1,20	0,885	2,20	0,986
0,25	0,599	1,25	0,894	2,25	0,988
0,30	0,618	1,30	0,903	2,30	0,989
0,35	0,637	1,35	0,912	2,35	0,991
0,40	0,655	1,40	0,919	2,40	0,992
0,45	0,674	1,45	0,926	2,45	0,993
0,50	0,692	1,50	0,933	2,50	0,994
0,55	0,709	1,55	0,939	2,55	0,995
0,60	0,726	1,60	0,945	2,60	0,995
0,65	0,746	1,65	0,950	2,65	0,996
0,70	0,758	1,70	0,955	2,70	0,996
0,75	0,773	1,75	0,960	2,75	0,997
0,80	0,788	1,80	0,964	2,80	0,997
0,85	0,802	1,85	0,968	2,85	0,998
0,90	0,816	1,90	0,971	2,90	0,998
0,95	0,829	1,95	0,974	2,95	0,998

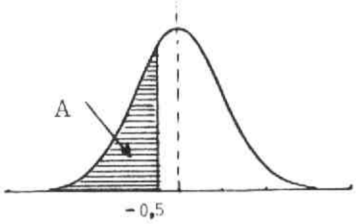
≫ 22. Hoe kun je uit de tabel vinden dat de oppervlakte 'tussen 1 en 2' gelijk is aan 0,136?

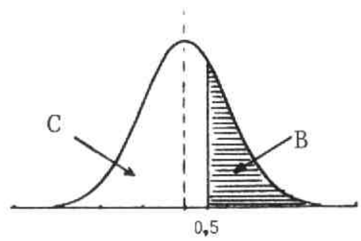
De tabel kan blijkens opgave ≫ 22 ook worden gebruikt om de oppervlakte te vinden van andere gebieden dan 'tussen $-\infty$ en z' (voor $z > 0$).
Daarbij zul je dikwijls gebruik maken van twee eigenschappen van de standaard-normale kromme:

- de kromme is symmetrisch t.o.v. de lijn $x = 0$;
- de totale oppervlakte van het gebied onder de kromme is 1.

Voorbeeld:

Gevraagd: oppervlakte A.





opp. A = opp. B (spiegeling in de lijn $x = 0$);

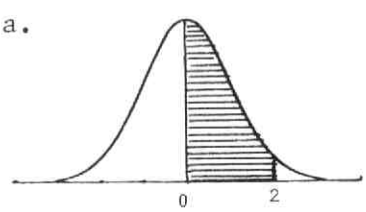
opp. B = 1 - opp. C (overgang op het 'complement');

opp. C = $\Phi(0,5)$ = 0,692 (tabel);

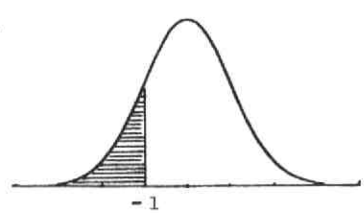
dus: opp. A = 1 - 0,692 = 0,308.

≫ 23. Vind de volgende oppervlakte uit de tabel van blz. 18).

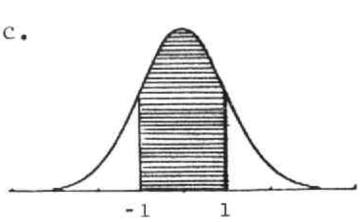
a.



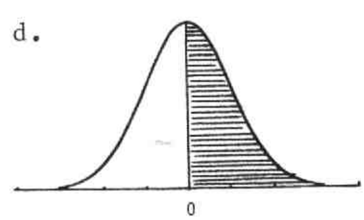
b.



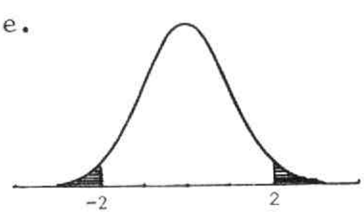
c.



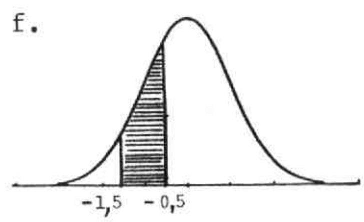
d.



e.



f.



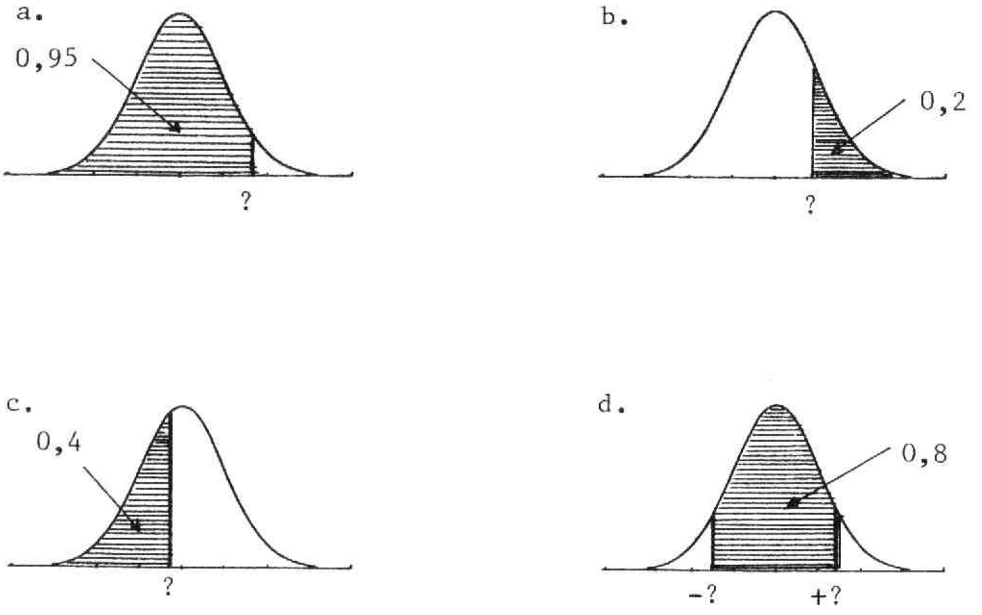
≫24. Even terug naar blz. 15.

Beantwoord de volgende vragen m.b.v. de Φ - tabel.

a. Hoeveel % van de melkflessen bevatte meer dan 99 cl en minder dan 101 cl?

b. En van hoeveel % lag de inhoud tussen 98 en 102 cl?

≫25. Terugzoeken in de tabel: vind de grenspunten van de volgende gebieden.



≫26. De controleur van de melkflessen beschouwde de 90% van de flessen, waarvan de inhoud het minst van het gemiddelde afweek, als 'normaal' en de overige 10% als 'uitschieters' (hetzij naar boven, hetzij naar beneden).

Bij welke gemeten inhouden kun je van een uitschieter spreken?

≫27. De melkfabrikant wil dat niet meer dan 5% van de door hem geleverde flessen minder dan 99 cl melk bevatten.

Op welk gemiddelde moet de vulmachine worden ingesteld?

Achterin dit boek is een meer uitgebreide tabel van Φ-waarden (in 4 decimalen) afgedrukt.

In de tabel vind je bijvoorbeeld:

$\Phi(0,34) \approx 0,6331$

z	0	1	2	3	4	5	6	7	8	9
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224

Terugzoeken:

$\Phi(z) \approx 0,719 \Rightarrow z \approx 0,58.$

z	0	1	2	3	4	5	6	7	8	9
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224

≫ 28. Zoek op in de uitgebreide tabel (blz. 71):

a. $\Phi(1,96)$ b. $\Phi(2)$ c. $\Phi(-1,96)$ d. $\Phi(-0,04)$

≫ 29. Voor welke z geldt:

a. $\Phi(z) \approx 0,96$?

b. $\Phi(z) \approx 0,50$?

c. $\Phi(z) \approx 0,25$?

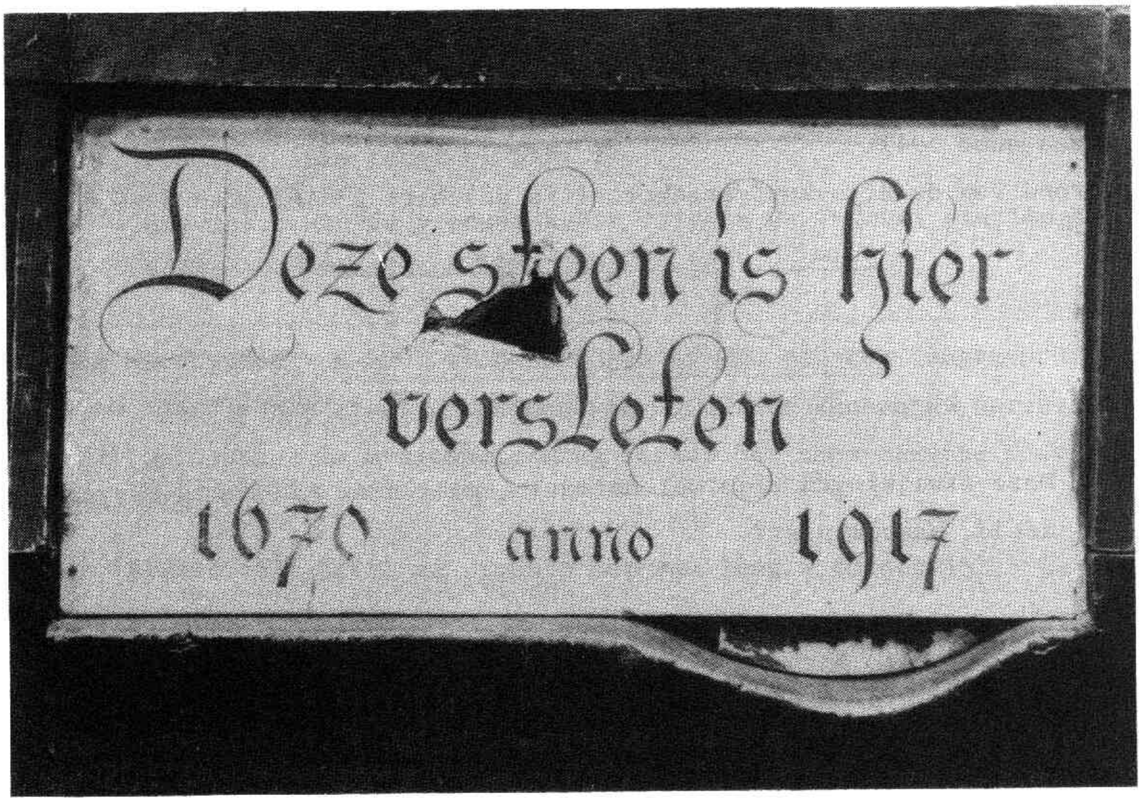
d. $\Phi(z) \approx 0,04$?

≫ 30. In een fabriek wordt frisdrank in halve-liter pakken verpakt. De werkelijke inhoud van de pakken is normaal verdeeld met een S.D. van 1 cl.
De fabrikant wil dat ten hoogste $2\frac{1}{2}\%$ van de pakken meer dan 52 cl bevatten.
Op welk gemiddelde moet de vulmachine worden ingesteld?

Café Hoppe op het Spui in Amsterdam bestaat al eeuwen. De drempelsteen die tussen 1670 en 1917 moest worden betreden om toegang tot de taveerne te krijgen is uitgesleten volgens een vrijwel normale kromme, zoals de detailfoto op blz. 23 laat zien.

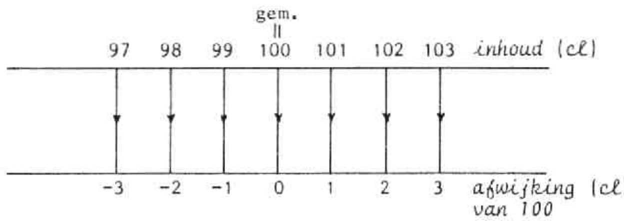
3

STANDAARDISEREN



Een normale verdeling met 'gemiddelde = 0' en 'S.D. = 1' kom je in de praktijk niet dagelijks tegen.
Toch is de standaard-normale verdeling van groot belang in de statistiek.
De clou zit hem in het feit dat elke normale verdeling via een *schaaltransformatie* teruggebracht kan worden tot de standaard-normale verdeling.
Met behulp van *één tabel* (de Φ-tabel) kunnen allerlei problemen worden opgelost die betrekking hebben op een normale verdeling met S.D. $\neq$ 1!

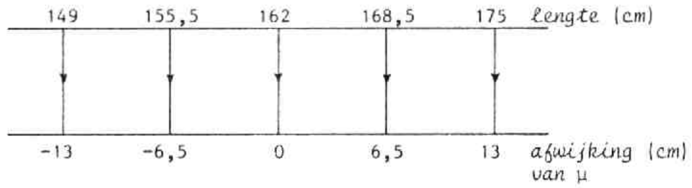
In het voorbeeld van de literflessen melk (zie blz. 15) hebben we via een voor de hand liggende ingreep het gemiddelde op nul gebracht. Die ingreep was: neem de *afwijking van het gemiddelde* als nieuwe variabele. Het 'nomogram' laat dat duidelijk zien:



Anders gezegd: je kiest een nieuwe schaalverdeling met de 'oorsprong' in 100.

Iets dergelijks kun je doen bij de lengte-verdeling van de Nederlandse vrouwen anno 1947.

Op grond van de Bijenkorf-steekproef is geschat: $\mu = 162$ cm en $\sigma = 6{,}5$ cm.

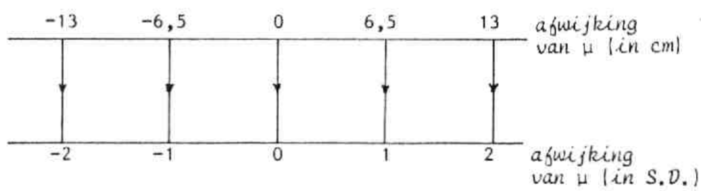


≫ 31. Deze afwijkingen zijn wel normaal-, maar niet *standaard*-normaal verdeeld. Waarom niet?

De grootte van de standaarddeviatie σ wordt hier uitgedrukt in cm.

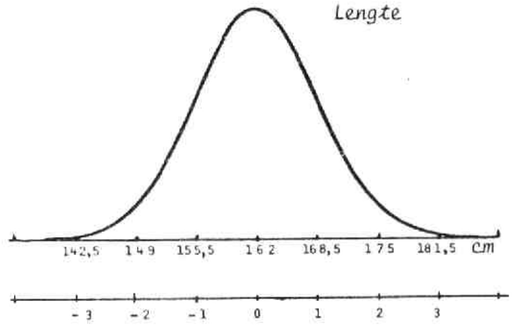
Door een andere keuze van de lengte-eenheid kun je een andere σ vinden, bijv. $\sigma = 65$ (in mm) of $\sigma = 0{,}065$ (in m).

Een listige keuze van de lengte-eenheid is nu: 6,5 cm.



Hiermee bereiken we dat de standaarddeviatie van de lengte-afwijkingen (gemeten in de nieuwe eenheid) gelijk is aan 1.

Door de keuze van de nieuwe eenheid is de lengte-verdeling van de Nederlandse vrouwen *gestandaardiseerd*.



≫ 32. a. Hoe kun je in de Φ-tabel vinden dat 68,26% van de vrouwen een lengte heeft tussen 155,5 en 168,5 cm?

b. Komt dat percentage overeen met dat in de steekproef van 5001 vrouwen? (Zie tabel op blz. 2).

c. Hoeveel % van de vrouwen heeft volgens de Φ-tabel een lengte tussen 149 en 175 cm?

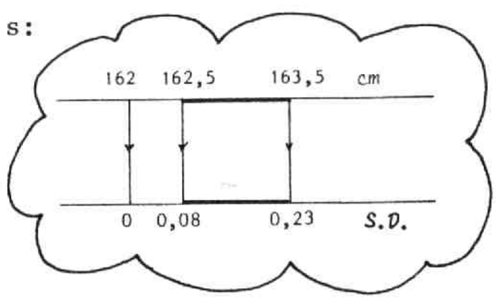
In de steekproef van 5001 vrouwen waren er 290 met een lengte van 163 cm; dat is een fractie van ongeveer 5,8%. We kunnen nu narekenen of dit percentage in overeenstemming is met het theoretische (d.w.z. normale) model:

- de 'klasse 163' bevat de vrouwen met een lengte tussen 162,5 en 163,5 cm;
- de klassegrenzen wijken resp. 0,5 en 1,5 cm af van het gemiddelde (= 162 cm);
- die afwijkingen druk je uit in S.D.'s:

$6{,}5 \text{ cm} = 1 \text{ S.D.}$

$0{,}5 \text{ cm} = \frac{0{,}5}{6{,}5} \text{ S.D.} \approx 0{,}08 \text{ S.D.}$

$1{,}5 \text{ cm} = \frac{1{,}5}{6{,}5} \text{ S.D.} \approx 0{,}23 \text{ S.D.}$



- de oppervlakte (= relatieve frequentie) onder de normale kromme, vind je nu uit de Φ-tabel:

$\Phi(0{,}23) - \Phi(0{,}08) = 0{,}5910 - 0{,}5319 = 0{,}059$ (ofwel 5,9%).

≫ 33. Hoeveel vrouwen (van de 5001) zou je volgens de Φ-tabel verwachten in de klasse van 155 cm? (De klassebreedte is weer 1 cm). Klopt dit met het resultaat in de Bijenkorf-steekproef?

≫ 34. In een zeker land is de gemiddelde lengte van dienstplichtige mannen gelijk aan 180 cm met een standaarddeviatie van 10 cm. Bij een militaire keuring worden er 10.000 mannen gekeurd.

a. Hoeveel van die mannen kun je verwachten met een lengte tussen 1.75 en 1.85 m? (teken een normale kromme met twee schalen).

b. Een dienstplichtige wordt afgekeurd op lengte als hij langer dan 2.00 m en kleiner dan 1.60 m is. Hoeveel afkeuringen op lengte kun je verwachten?

c. Door de minister van defensie wordt beslist dat in het vervolg 5% van de dienstplichtigen zal worden afgekeurd omdat ze te lang zijn en 5% omdat ze te klein zijn. Waar komen de grenzen van afkeuring te liggen?

≫ 35. De jaarproduktie aan eieren van een pluimveehouder is normaal verdeeld (naar gewicht) met een gemiddelde van 56,3 gram en S.D. van 7,6 gram.

a. Hoeveel % van de eieren is zwaarder dan 60 gram?

b. De eieren worden in acht gewichtsklassen verdeeld:

klasse	*gewicht (gr)*
0	> 68½
1	64 - 68½
2	59½ - 64
3	54½ - 59½
4	49½ - 49½
5	44½ - 44½
6	39½ 39½
7	< 39½

Teken de normale kromme bij de gewichtsverdeling en geef in je tekening de acht gewichtsklassen aan als gebieden onder de kromme. Schrijf in elk gebied het percentage eieren dat tot die klasse behoort.

≫ 36. Een postorderbedrijf verkoopt o.a. bouwpakketten die een normaal verdeeld brutogewicht hebben (d.w.z. inclusief verpakking) met een gemiddelde van 4,2 kg en een S.D. van 0,8 kg.
Men verwacht in het volgende jaar 20.000 van die pakketten te verkopen.
De geldende PTT-tarieven van de verzending van pakjes:

gewicht (gr)	prijs
250 - 500	4,25
500 - 1000	5,25
1000 - 3000	6,50
3000 - 5000	8,00
5000 - 7000	10,--
7000 - 10000	12,--

Welk totaalbedrag aan portokosten kan het postorderbedrijf voor de verzending van de bouwpakketten verwachten?

≫ 37. Een droog korrelig poeder bevat deeltjes die zuiver bolvormig zijn en waarvan de diameter normaal verdeeld is met $\mu = 170$ mikron en $\sigma = 11,6$ mikron.
Men wil dit poeder verdelen in drie soorten, namelijk: grof, middel en fijn en wel zó dat deze drie klassen gelijke aantallen korrels bevatten.

a. Hoe groot moeten de diameters van de gaten van de benodigde zeven zijn waarmee je de gewenste verdeling kunt krijgen?

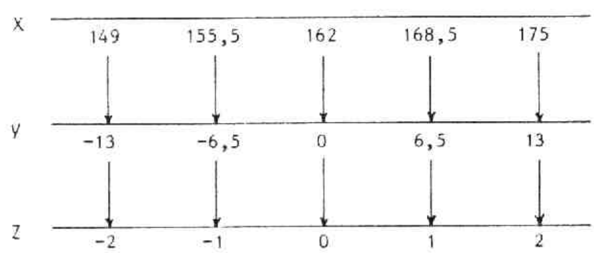
b. Zal de diameter van de fijne korrels ook normaal verdeeld zijn?

≫ 38. De straatlantaarns in een zekere stad worden door de gemeente van nieuwe gloeilampen voorzien: in totaal 2000 stuks.
De levensduur van deze lampen is normaal verdeeld met een gemiddelde levensduur van 2500 uur en een standaarddeviatie van 500 uur. Om economische redenen is de gemeente van plan *alle* lampen door nieuwe te vervangen op het moment dat 20% van de geplaatste lampen stuk is (en dus niet tussentijds lampen te vervangen).
Na hoeveel tijd zal dit het geval zijn?

≫ 39. In een fabriek worden pakken suiker gevuld waarvan het nettogewicht 1 kg behoort te zijn.
Uit steekproeven is gebleken dat dit nettogewicht normaal is verdeeld met (inderdaad) een gemiddelde van 1 kg en een standaardafwijking van 15 gram.
De fabrikant wil dat 95% van de te produceren pakken suiker tenminste 1 kg suiker bevatten. Op welk gemiddeld nettogewicht (afgerond in g.) zal de vulmachine moeten worden ingesteld?

TRANSFORMATIEFORMULES

De voorafgaande vraagstukken heb je steeds opgelost door overgang van 'gewone eenheden' (cm, kg, ...) op nieuwe-eenheden (S.D.).
Wiskundig gezegd: je hebt steeds een 'schaaltransformatie' toegepast.



X = lichaamslengte (in cm).
Y = afwijking van de gemiddelde lichaamslengte (in cm).
Z = afwijking van het gemiddelde in S.D.'s.

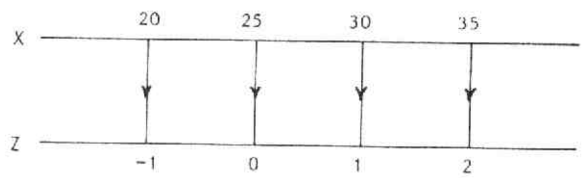
Er geldt: $Y = X - 162$

$$Z = \frac{Y}{6,5}$$

Dus: $$Z = \frac{X - 162}{6,5} \qquad \text{.........} \; (*)$$

≫ 40. In (*) is Z uitgedrukt in X.
Omgekeerd kun je ook X uitdrukken in Z. Hoe?

≫ 41. Bekijk de schaaltransformatie:

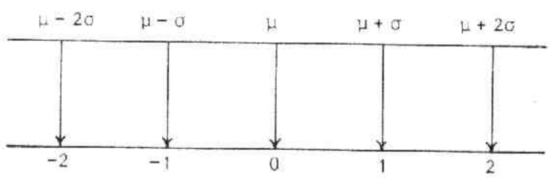


a. Druk Z uit in X.

b. Druk X uit in Z.

ALGEMEEN:

Laat X een variabele zijn met gemiddelde μ en standaarddeviatie σ. De overgang op de 'standaardvariabele' Z met gemiddelde 0 en standaarddeviatie 1:



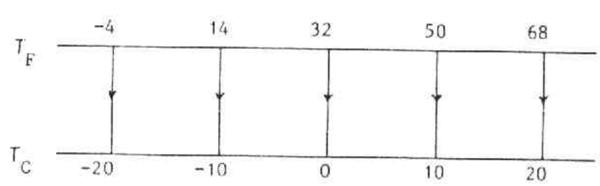
wordt gekenschetst door de formule:

$$Z = \frac{X - \mu}{\sigma}$$

ofwel door:

$$X = \sigma Z + \mu$$

Schaaltransformaties zoals bij het standaardiseren kom je ook in andere situaties tegen. Bekend is het voorbeeld van de temperatuurschalen Celsius-Fahrenheit. De laatste eenheid wordt nog steeds vaak gebruikt in Angelsaksische landen.



≫ 42. Bekijk het plaatje onderaan blz. 29.
Je ziet daarin o.a. dat 32° F overeenkomt met 0° C.

a. Druk T_C uit in T_F.

b. Druk T_F uit in T_C.

≫ 43. In 1983 beleefde men in Engeland een ongekend mooie zomer. In de maand juli werd op zekere plaats een gemiddelde maximum temperatuur van liefst 76° F gemeten met een standaarddeviatie van 6,3° F.

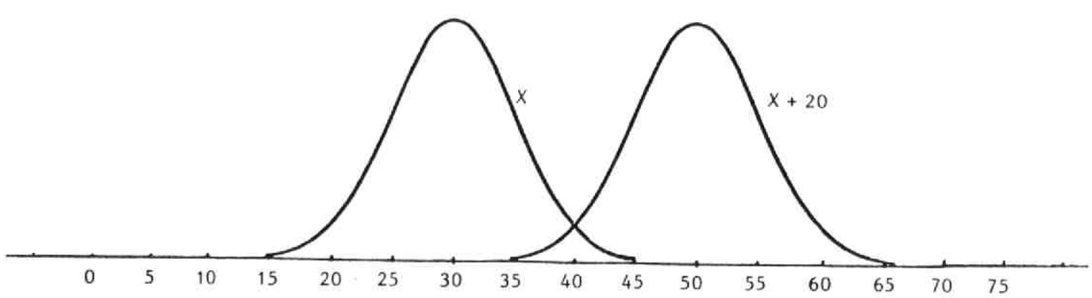
a. Hoeveel °C bedroeg de gemiddelde maximum temperatuur in juli 1983 in die plaats?

b. En hoeveel °C bedroeg de S.D.? (Pas op!).

≫ 44. Een variabele X is normaal verdeeld met $\mu = 30$ en $\sigma = 5$.
De variabelen X_1, ..., X_5 zijn op de volgende wijze van X afgeleid.

$X_1 = X + 20$; $X_3 = \frac{1}{2}X$; $X_2 = X - 15$; $X_4 = 2X - 60$; $X_5 = 0{,}2(X - 30)$.

De variabelen X_1, ..., X_5 zijn elk ook weer normaal verdeeld.

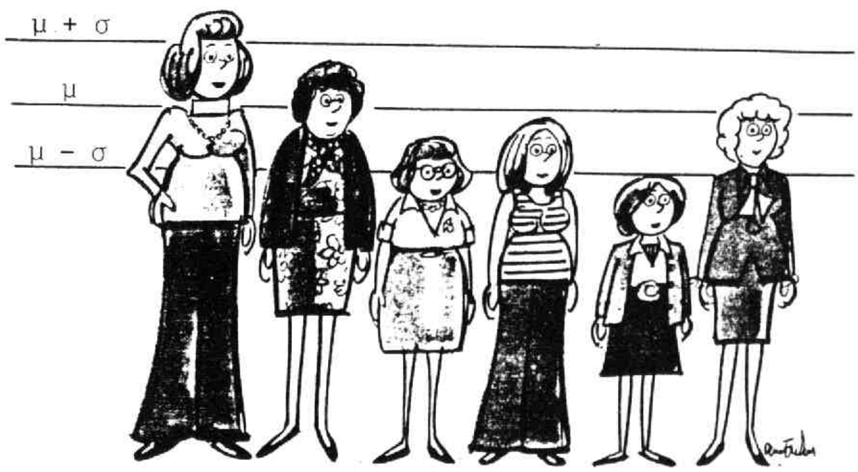


a. De grafieken van X en X_1 (= $X + 20$) zijn in één figuur getekend.
Wat is de waarde van μ en σ voor de variabele X_1?

b. Teken ook de grafieken van X_2, X_3, X_4, X_5 samen met X in één figuur. (N.B. De totale oppervlakte onder de kromme moet steeds hetzelfde blijven!).
Geef bij elke kromme de waarde van μ en σ.

4

STATISTISCHE VUISTREGELS/ NORMAAL WAARSCHIJNLIJKHEIDSPAPIER

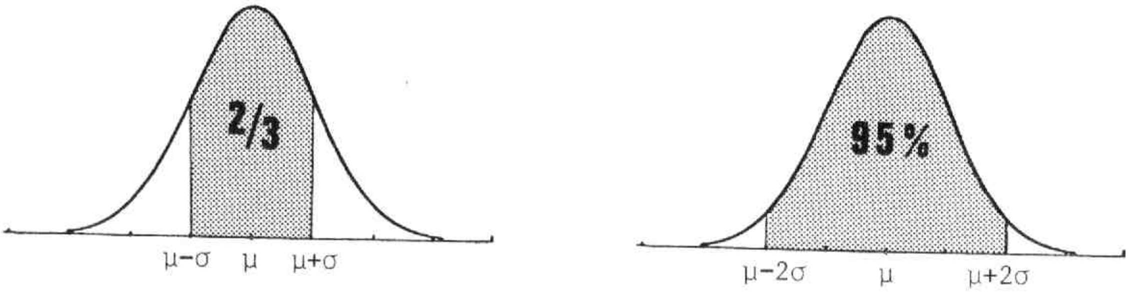


Twee van de drie volwassen Nederlandse vrouwen zal een lichaamslengte hebben die ten hoogste één keer de standaarddeviatie afwijkt van de gemiddelde lengte.

≫45. Controleer of deze slogan opgaat voor de 'Bijenkorf-vrouwen'. ($\mu = 162$; $\sigma = 6{,}5$).

≫46. Een uitspraak als hierboven kun je voor elke normaal-verdeelde populatie doen!
Hoe kun je deze vuistregel 'bewijzen' met behulp van de Φ-tabel?

Twee bekende statistische vuistregels in beeld en woord gebracht:



Bij een representatieve steekproef uit een normaal-verdeelde populatie kun je verwachten dat:

- *ongeveer 2/3 van de waarnemingsuitkomsten niet meer dan 1 keer de S.D. afwijkt van het gemiddelde;*
- *ongeveer 95% van de waarnemingsuitkomsten niet meer dan 2 keer de S.D. afwijkt van het gemiddelde.*

Deze vuistregels zijn een gevolg van:
de oppervlakte van het gebied tussen −1 en 1 onder de standaard-normale kromme is 0,6826 en de oppervlakte van het gebied tussen −2 en 2 is 0,9544.
Je ziet dat '$\frac{2}{3}$' en '95%' tamelijk ruwe benaderingen zijn.

≫ 47. 'Dit blik verf is goed voor 9 tot 12 m²' staat er op het etiket. Neem aan dat de oppervlakte (in m²) die je met de inhoud van één zo'n blik kunt verven een normaal-verdeelde variabele is. De fabrikant bedoelt met zijn 'garantie' dat ongeveer 95% van zijn blikken aan de gestelde eisen voldoen.

a. Met welk gemiddelde oppervlakte en met welke S.D. heeft de fabrikant (vermoedelijk) gerekend?

b. Hoe groot schat je de kans dat je met één blik verf niet meer dan 9,75 m² kunt verven?

≫ 48. Bekijk nog eens de frequentie-tabel van het aantal vinnerven bij bot:

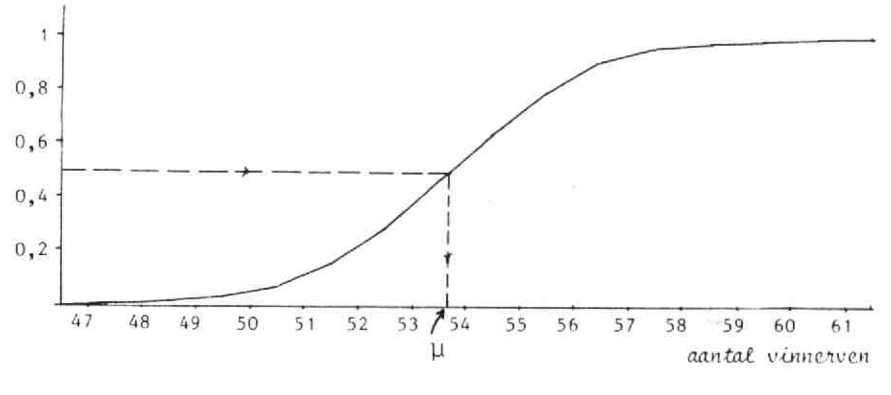
aantal vinnerven	*frequentie*
47	5
48	2
49	13
50	23
51	58
52	96
53	134
54	127
55	111
56	74
57	37
58	16
59	4
60	2
61	1

a. Maak een *cumulatieve-frequentie-tabel* bij deze gegevens.

b. Hoe kun je uit die tabel snel (bij benadering) het gemiddelde vinden?

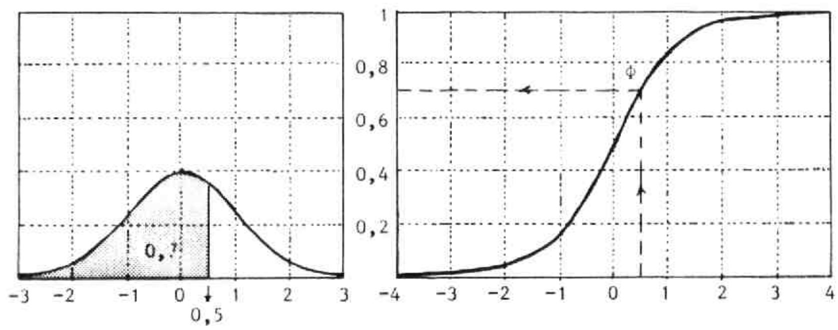
c. En hoe de S.D.? (Gebruik één van de vuistregels).

Bij opgave ≫ 48 had je ook met frequentiepercentages of relatieve frequenties kunnen werken.
Hieronder zie je de cumulatieve relatieve-frequentiecurve op grond van de 703 gegevens:



≫ 49. In de figuur op blz. 33 zie je hoe je de plaats van het gemiddelde (μ) op de horizontale as met behulp van de grafiek kunt bepalen. Gebruik de vuistregels om de plaatsen van $\mu \pm \sigma$, $\mu \pm 2\sigma$ op de horizontale as te vinden.

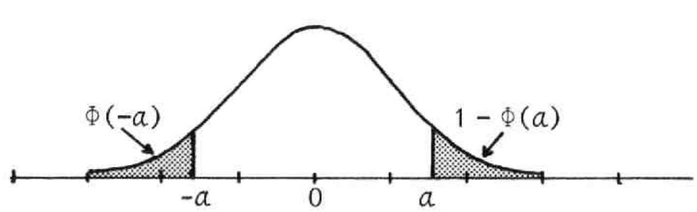
De grafiek op blz. 33 benadert de 'cumulatieve normale curve'. Hieronder rechts zie je de cumulatieve grafiek bij de standaardnormale curve (in feite de grafiek van de functie Φ):



≫ 50. Het aflezen uit de cumulatieve grafiek gaat gemakkelijker dan uit de gewone klok-kromme. (Zie bijv.: $\Phi(0,5) \approx 0,7$).
Hoe kun je uit de cumulatieve grafiek de beide vuistregels voor ... male verdeling aflezen?

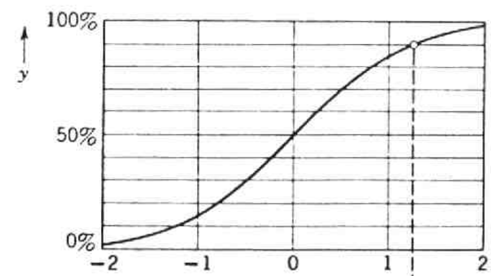
≫ 51. a. Die grafiek van Φ (de 'S-kromme') heeft twee horizontale asymptoten, namelijk: $y = 0$ en $y = 1$.
Hoe kun je dit verklaren?

b. Laat a een positief getal zijn.
In de 'klok-grafiek' kun je zien: $\Phi(-a) = 1 - \Phi(a)$.
Hoe kun je die eigenschap uit de 'S-kromme' aflezen?

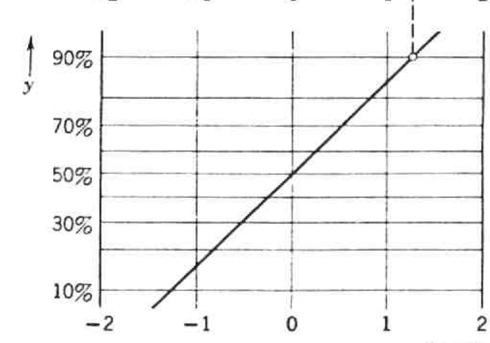


Het werken met de cumulatieve normaalkromme (de S-kromme) heeft zekere voordelen, zoals het direct kunnen aflezen van het gemiddelde, en, met gebruikmaking van de vuistregel(s) van de standaarddeviatie.
Je moet daarbij natuurlijk wel weten of je inderdaad met een normale verdeling te doen hebt.
Er is papier ontworpen met een speciale schaalverdeling, waarop je dit laatste gemakkelijk kunt aflezen. Dit zgn. *normaal waarschijnlijkspapier* heeft langs één zijde (de horizontale as) een gewone ('lineaire') schaalverdeling en langs de andere zijde (verticale as) een schaalverdeling die zodanig is, dat een cumulatieve normale-verdelingskromme er uitziet als een rechte lijn.

Cumulatieve standaard-normale kromme (grafiek van Φ):



Grafiek van Φ op normaal-waarschijnlijkheidspapier:



≫ 52. De schaalverdeling op het normaal waarschijnlijkheidspapier is zo opgerekt, dat de S-kromme een rechte lijn wordt.

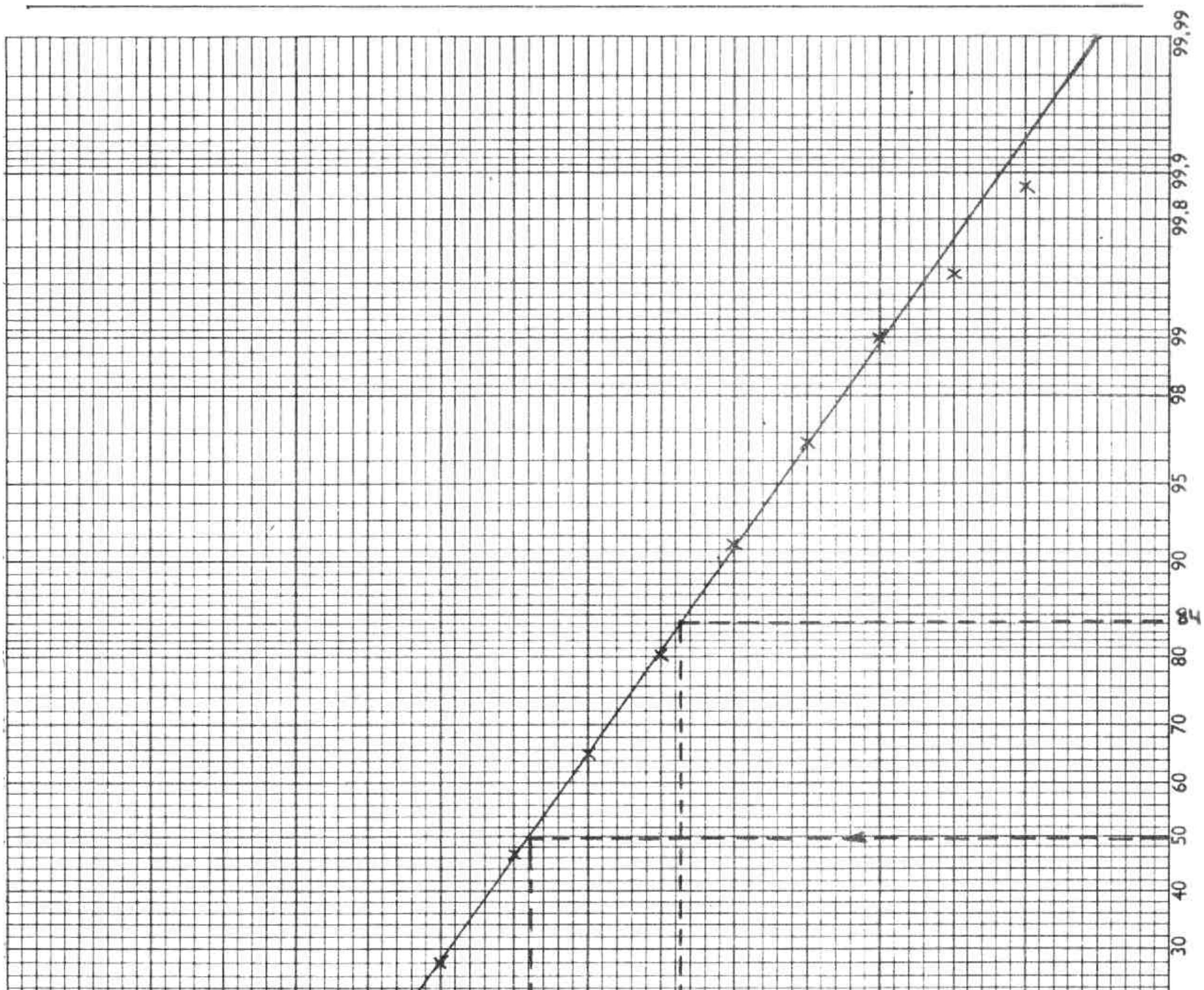
a. Hoe kun je verklaren dat de 'mazen' in het rooster naar boven toe en naar beneden verticaal steeds wijder worden?

b. De onderste grafiek kun je niet zo ver uitbreiden dat de 100%-lijn (en de 0%-lijn) te zien is. Verklaar.

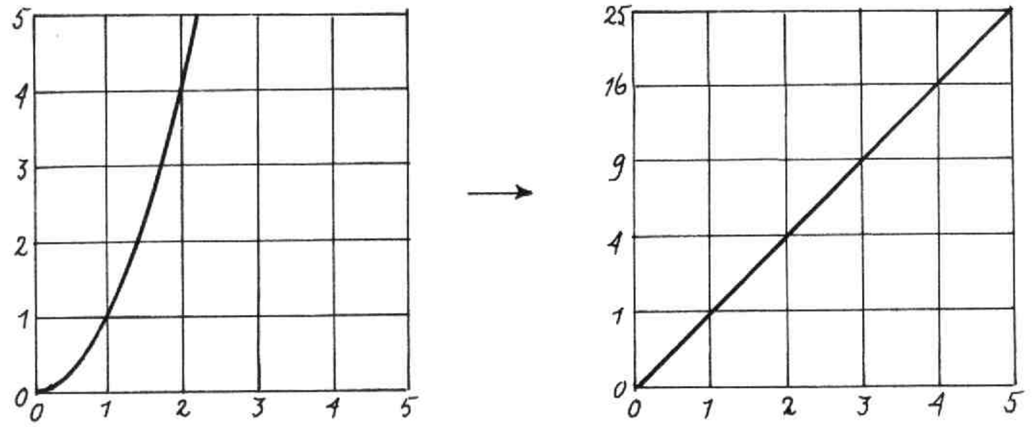
Het normaal waarschijnlijkheidspapier gebruik je als volgt:

- langs de horizontale as is de schaalverdeling lineair en kun je de getallen bijplaatsen afhankelijk van de gegevens waarover je beschikt;
- langs de verticale as zijn cumulatieve frequentiepercentages volgens de normale verdeling geplaatst; het papier is dan ook uitsluitend geschikt voor cumulatieve grafieken;
- als de gegevens van een steekproef een normale verdeling representeren zullen de bijbehorende punten in de grafiek (nagenoeg) op een rechte lijn moeten liggen;
- is dat laatste het geval, dan kun je via de 50%-lijn het gemiddelde van de verdeling schatten;
- de standaarddeviatie bij een normaal verdeelde steekproef vind je bijv. door de 84%-lijn (of 16%-lijn) te tekenen; je kunt dan de waarnemingsuitkomst aflezen die 1 keer de S.D. van het gemiddelde afligt.

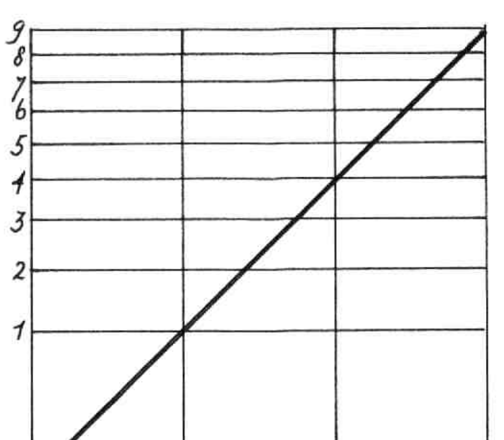
Op blz. 37 is dit toegepast op de verdeling van het aantal vinnerven.
De meetpunten zijn uitgezet bij 47,5; 48,5; enz. zoals te doen gebruikelijk bij een cumulatieve frequentie-curve.
De 'meetpunten' blijken inderdaad nagenoeg op een rechte lijn te liggen; de verdeling van het aantal vinnerven kan goed benaderd worden door de normale verdeling.
De afwijkingen zitten vooral in het begin en de staart. Zie ook de grafiek in hoofdstuk 1 (blz. 11).
De afgelezen S.D. is 55,65 - 53,6 = 2,05.
Ter controle hebben we het gemiddelde en de S.D. ook berekend in 1 dec. nauwkeurig. Resultaten: 53,7 en 2,1!


99,99
99,9
99,8
99
98
95
90
84
80
70
60
50
40
30
20
10
5
2
1
0,5
0,2
0,1
0,05
0,01
47.5
48.5
49.5
50.5
51.5
52.5
53.5
54.5
55.5
56.5
57.5
58.5
59.5
60.5
61.5
σ
aantal vinnerven

Via een geschikte schaaltransformatie langs de assen, kun je in feite elke grafiek omvormen tot een rechte lijn.
Als voorbeeld nemen we $y = x^2$ (voor $x \geqq 0$).



De verticale schaalverdeling in het rechter plaatje is niet-lineair, d.w.z. bij gelijke afstanden horen ongelijke getalverschillen:
afstand 0 tot 1 = afstand 1 tot 4 = afstand 4 tot 9 = ...
Meestal worden bij zo'n afwijkende schaal de getalverschillen gelijk genomen, waardoor de verdeelstrepen op ongelijke afstanden van elkaar komen te liggen.
Bovenstaande grafiek komt er dan, uitvergroot, zó uit te zien:



≫ Hoe kun je de plaats van het getal 2 op de verticale as precies bepalen?

≫ 53. De Belg A. Quetelet verzamelde veel gegevens om zijn ontdekkingen over het normaal verdeeld zijn van biometrische gegevens te staven. Zo mat hij de borstomvang van 1516 soldaten ('the Army of Potomac') in inches.

Resultaat:

Borstomvang in inches	*Frequentie*
28	2
29	4
30	17
31	55
32	102
33	180
34	242
35	310
36	251
37	181
38	103
39	42
40	19
41	6
42	2

Verwerk deze gegevens op normaal waarschijnlijkheidspapier en lees af hoeveel inches de gemiddelde borstomvang en hoeveel inches de S.D. is.

≫ 54. Van 300 gloeilampen van een zeker type werd de levensduur (in uren) bepaald:

Levensduur (uren)	*Frequentie*
950 - 1050	4
1050 - 1150	9
1150 - 1250	19
1250 - 1350	36
1350 - 1450	51
1450 - 1550	58
1550 - 1650	53
1650 - 1750	37
1750 - 1850	20
1850 - 1950	9
1950 - 2050	3
2050 - 2150	1

Ga na of de levensduur van deze gloeilampen normaal verdeeld is en geef (m.b.v. normaal waarschijnlijkheidspapier) een schatting van gemiddelde en S.D.

≫ 55. In de Volkskrant van dinsdag 1 november 1983 schreef Machteld Roede onder de titel 'Blijven de Nederlanders maar groeien?' onder andere:

De Nederlanders zijn langer geworden. Kinderen van nu zijn langer dan leeftijdgenootjes van vroeger. De lichaamslengte bij het bereiken van de volwassen leeftijd verschilt aanzienlijk van die van jonge mensen een eeuw geleden. In 1865 was de helft van de keurlingen kleiner dan 1,65 m. Ruim een kwart werd afgekeurd vanwege een lengte minder dan 1,57 m. Extreem was de lengte van 1,80 m, bereikt door minder dan twee percent. Groei bij mannen kwam toen omstreeks het 25ste jaar tot stilstand; na de keuring kon nog een doorgroei van enkele centimeters optreden. Thans zijn mannen in hun twintigste jaar of eerder uitgegroeid.

Sinds 1975 is meer dan de helft van de keurlingen groter dan 1,80 m. Bij een landelijk groei-onderzoek van 1980 bleek drie percent van 14,0 jaar oude jongens, tien percent van 14,5-jarigen en meer dan de helft van 17,5-jarigen de grens van 1,80 m reeds te zijn gepasseerd. Ook bij vrouwen zijn verschuivingen in het groeipatroon opgetreden. In 1980 bleek drie percent van de 18/20-jarige vrouwen 1,80 m of langer te zijn. Extreem is nu te noemen de lengte van 1,95 m, bereikt door drie percent van de twintigjarige mannen. In 1980 was een op elke duizend twintigjarige mannen 2,02 m en een op elke duizend twintigjarige vrouwen 1,87 m.

Verschuivingen in de lichaamslengte van de bevolking over verloop van tijd zijn waargenomen in alle westerse landen, terwijl nu ook in ontwikkelingslanden dit verschijnsel op gaat treden. De verklaring ligt in veranderingen in het milieu rond het opgroeiend kind. Lengte-groei is immers slechts ten dele erfelijk bepaald. Gedurende de gehele periode is het proces milieugevoelig. Tot nu toe zijn alleen maar remmende invloeden aangetoond. Een kind dat steeds weer ernstig ziek is, regelmatig hoge koorts heeft en/of over langere tijd te weinig en te eenzijdig eet, blijft steeds meer achter bij het erfelijk uitgestippelde groeipatroon. Bij een algehele verbetering van de omgeving neemt daarentegen de gezondheidstoestand van de bevolking toe, ziekte en sterfte-percentages bij kinderen nemen af, en de gemiddelde lengte neemt toe.

Midden vorige eeuw waren de omstandigheden in Nederland evenals in de rest van Europa erbarmelijk. Epidemieën en ernstige infectieziekten kwamen veel voor. De hygiëne was zeer slecht: nauwelijks of geen watervoorziening, geen riool, de kinderen speelden op afvalhopen. De behuizing was slecht en het voedselpakket schaars.

Het is navrant te noemen dat de eerste gerichte aandacht voor de slechte gezondheidstoestand en inzicht in de invloed van het milieu voortkwam uit de zorg voor een goed, gezond leger. Villermé toonde in 1829 aan dat in Frankrijk de lengte van de keurlingen uit de armste departementen geringer was en het afkeuringspercentage hoger. In Nederland start omstreeks 1850 een soortgelijk onderzoek. Zeeman wees er in 1861 op hoe de jaarlijkse schommelingen in het afkeuringspercentage van keurlingen een relatie vertoonde met de schommelingen in de prijs van het graan.

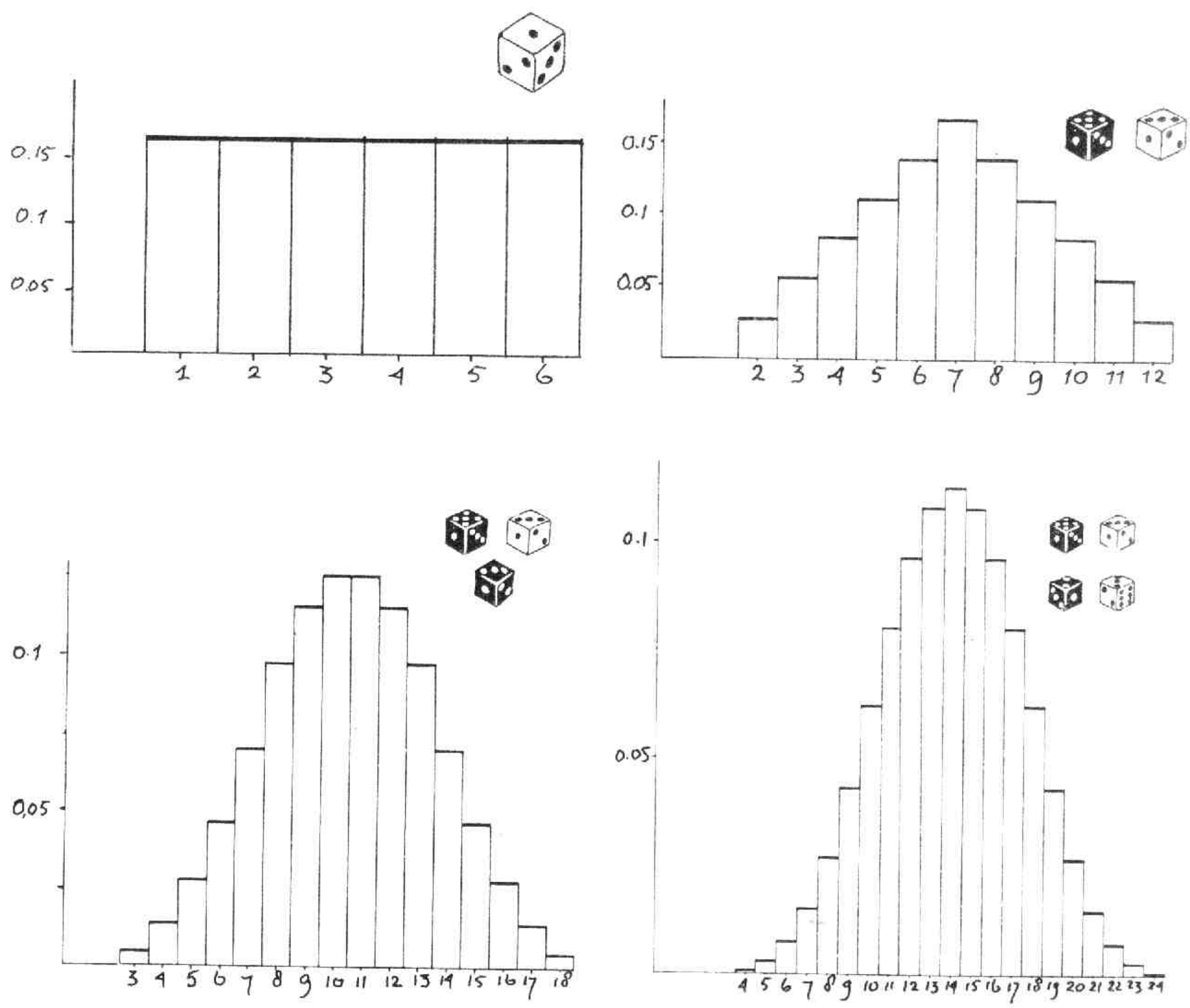
a. Uit de gegevens over de ter militaire keuring opgeroepen mannen in 1865 kun je afleiden dat de lengte-verdeling van de 20-jarige mannen in dat jaar *niet* normaal verdeeld was. Hoe?

b. Welke verklaring kun je voor die afwijking bedenken?

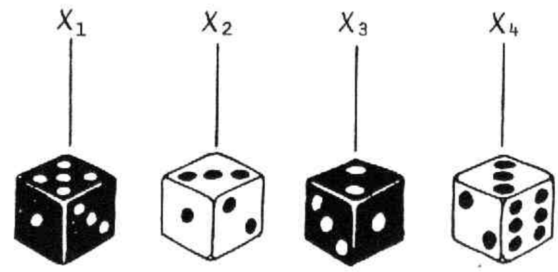
c. Geef een schatting van de gemiddelde lengte en de S.D. (beide in cm) van de Nederlandse twintigjarige mannen en vrouwen in 1980.

5

SPREIDING BIJ KANSVERDELINGEN



Kanshistogrammen voor het werpen met resp. 1, 2, 3, 4 dobbelstenen, waarbij gelet wordt op de som van de aantallen ogen.



≫ 56. Er wordt geworpen met hoogstens vier (van elkaar te onderscheiden) dobbelstenen. De ogenaantallen die boven komen zijn resp. X_1, X_2, X_3, X_4.

a. Bereken: $P(X_1 = 6)$, $P(X_1 + X_2 = 12)$, $P(X_1 + X_2 + X_3 = 18)$, $P(X_1 + X_2 + X_3 + X_4 = 24)$.

b. Bereken ook: $P(X_1 + X_2 = 6)$, $P(X_1 + X_2 + X_3 = 6)$, $P(X_1 + X_2 + X_3 + X_4 = 6)$.

Op blz. 41 zie je de kanshistogrammen van X_1, $X_1 + X_2$ $(= S)$, $X_1 + X_2 + X_3$ $(=T)$ en $X_1 + X_2 + X_3 + X_4$ $(=U)$.
Net als bij steekproefverdelingen zijn 'gemiddelde' en 'spreiding' bij zo'n kansverdeling belangrijke grootheden.

Voor wat betreft 'gemiddelde' is het een oud verhaal. Als je een groot aantal keren met één dobbelsteen werpt, kun je verwachten dat je in $\frac{1}{6}$ van het aantal beurten 1, in $\frac{1}{6}$ van het aantal beurten 2, enz. werpt.

Het 'verwachte gemiddelde' is dus:

$$1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6} = 3{,}5$$

Bij dit experiment heb je natuurlijk ook een zekere verwachting met betrekking tot de standaarddeviatie van de uitkomsten.
Immers: in $\frac{1}{6}$ van het aantal beurten kun je een 'zes', dus een afwijking van $2\frac{1}{2}$ van het gemiddelde verwachten, in $\frac{1}{6}$ van het aantal beurten een afwijking van $1\frac{1}{2}$, enz.

In tabelvorm:

uitkomst	*afwijking van gemiddelde* ($=3\frac{1}{2}$)	*gekwadrateerde afwijking*	*kans*
1	$-2\frac{1}{2}$	$6\frac{1}{4}$	$\frac{1}{6}$
2	$-1\frac{1}{2}$	$2\frac{1}{4}$	$\frac{1}{6}$
3	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{6}$
4	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{6}$
5	$1\frac{1}{2}$	$2\frac{1}{4}$	$\frac{1}{6}$
6	$2\frac{1}{2}$	$6\frac{1}{4}$	$\frac{1}{6}$

≫ 57. Reken na dat de verwachte *gemiddelde gekwadrateerde afwijking* gelijk is aan $2\frac{11}{22}$.
Hoe groot is de verwachte *standaarddeviatie* (in 3 dec.)?

AFSPRAKEN:

- Het 'verwachte gemiddelde' van een stochast X noemt met de *verwachtingswaarde* van X.
  Notatie: $E(X)$ of μ_X.
- De 'verwachte gemiddelde gekwadrateerde afwijking' van X, dus het verwachte gemiddelde van $(X-\mu_X)^2$, noemt men de *variantie* van X.
  Notatie: $\mathrm{Var}(X)$.
- De 'verwachte standaarddeviatie' van X, dus de vierkantswortel uit de variantie, noemt men kortweg de *standaarddeviatie* van X.
  Notatie: $SD(X)$ of σ_X.

Opmerking:

De 'variantie' wordt niet als spreidingsmaat gebruikt, maar als tussenstap bij het berekenen van de S.D.

Voor X_1 (= aantal ogen bij werpen met één dobbelsteen) geldt dus:

$$E(X_1) = \frac{7}{2} = 3{,}5$$
$$Var(X_1) = \frac{35}{12} \approx 2{,}917$$
$$SD(X_1) = \sqrt{\frac{35}{12}} \approx 1{,}708$$

≫ 58. Werpen met twee dobbelstenen.

Bekijk nu de stochast $X_1 + X_2$ $(=S)$.

a. Verklaar: $E(S) = 7$.

b. Maak een tabel als bovenaan blz. 43 en bereken $Var(S)$.

c. Bereken ook: $SD(S)$.

De verwachtingswaarde van $X_1 + X_2$ kun je op drie manieren vinden:

- Recht toe, recht aan berekenen:

  $E(X_1 + X_2) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \ldots + 12 \cdot \frac{1}{36} = 7.$

- Toepassing van de 'somregel' van verwachtingswaarden:
  het verwachte aantal ogen is voor de eerste, zowel als de tweede steen $3\frac{1}{2}$, de verwachte gemiddelde som is dan $3\frac{1}{2} + 3\frac{1}{2} = 7$.
  Of in formule: $E(X_1 + X_2) = E(X_1) + E(X_2)$.

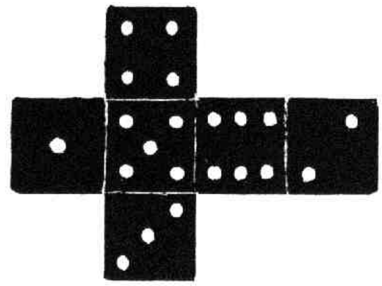
- Gebruikmaking van de symmetrie van de kansverdeling:
  7 is het middelste getal tussen 2 en 12 (zie het histogram op blz. 41).

De vraag is nu of je ook de variantie handig had kunnen vinden.
Als je opgave ≫ 58b goed hebt berekend, kun je inzien dat geldt:

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2). \quad (5\tfrac{5}{6} = 2\tfrac{11}{12} + 2\tfrac{11}{12}).$$

De verklaring hiervoor is veel minder eenvoudig dan bij de 'somregel' voor het gemiddelde.
Sterker: de regel gaat niet altijd op!
De volgende opgave levert een tegenvoorbeeld.

≫59. Noem het aantal ogen dat *boven* komt bij *één* dobbelsteen: X en het aantal ogen dat *onder* komt: Y.



a. Wat weet je (zeker) van $X+Y$?

b. Verklaar: $\mathrm{Var}(X+Y)=0$.

c. Hoe groot is $\mathrm{Var}(X)+\mathrm{Var}(Y)$?

In het voorbeeld van opgave ≫59 zijn de stochasten X en Y duidelijk niet onafhankelijk.

Dat wil zeggen: de kans op een uitkomst van Y wordt beïnvloed door de uitkomst van X.

$P(Y=1)=\frac{1}{6}$, maar bijvoorbeeld: $P(Y=1\,|\,X=5)=0$.

Voor afhankelijke stochasten hoeft de somregel voor varianties dus niet te gelden; voor onafhankelijke stochasten echter wel![*)]

Als A en B onafhankelijke stochasten zijn, dan $\mathrm{Var}(A+B)=\mathrm{Var}(A)+\mathrm{Var}(B)$. De somregel geldt ook voor drie of meer onafhankelijke stochasten!

≫60. In opgave ≫58 heb je $SD(X_1+X_2)$ uitgerekend.
Geldt: $SD(X_1+X_2)=SD(X_1)+SD(X_2)$?

≫61. Werpen met drie dobbelstenen. $T=X_1+X_2+X_3$.
Bereken (zo handig mogelijk): $E(T)$, $\mathrm{Var}(T)$, $SD(T)$.

---

*) Twee stochasten A en B noem je onafhankelijk als de kans op een uitkomst van A niet beïnvloed wordt door de uitkomst van B.
In formule: $P(A=k\,|\,B=m)=P(A=k)$ voor alle uitkomsten k en m van resp. A en B.

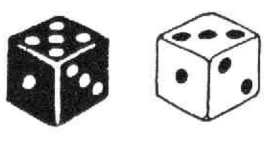
≫ 62. Vier dobbelstenen. $U = X_1 + X_2 + X_3 + X_4$.
Bereken E(U), Var(U), SD(U).

≫ 63. Vergelijk de kanshistogrammen van X_1, $X_1 + X_2$, $X_1 + X_2 + X_3$ en $X_1 + X_2 + X_3 + X_4$. (Zie blz. 41).
De standaarddeviatie is een maat voor de spreiding van de uitkomsten. Dat deze spreiding groter wordt bij toename van het aantal dobbelstenen is niet geheel onverwacht. Het aantal mogelijke uitkomsten neemt immers toe, als je met meer dobbelstenen werpt.

Stochast	*Bereik*	*'Spreidingsbreedte'*
X_1	$\{1,2,\ldots,6\}$	5
$X_1 + X_2$	$\{2,3,\ldots,12\}$	10
$X_1 + X_2 + X_3$	$\{3,4,\ldots,18\}$	15
$X_1 + X_2 + X_3 + X_4$	$\{4,5,\ldots,24\}$	20

De spreidingsbreedten verhouden zich als 1, 2, 3 en 4.
Hoe verhouden de respectievelijke S.D.'s zich?

≫ 64. Het maakt een groot verschil of je te maken hebt met de stochasten X_1, $X_1 + X_2$, $X_1 + X_2 + X_3$, enz. of met X_1, $2X_1$, $3X_1$, enz.!



$X_1 + X_2 = 8$

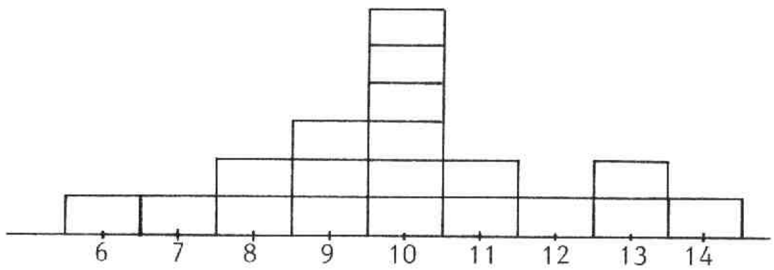
$2X_1 = 10$

a. Teken het kanshistogram voor $2X_1$.

b. Ga na dat $2X_1$, $3X_1$ en $4X_1$ dezelfde verwachtingswaarden hebben als $X_1 + X_2$, $X_1 + X_2 + X_3$, $X_1 + X_2 + X_3 + X_4$.

c. Hoe verhouden de S.D.'s van X_1, $2X_1$, $3X_1$ en $4X_1$ zich?

≫ 65. X is een stochast waarvan bekend is: $E(X) = 10$, $Var(X) = 4$ en $SD(X) = 2$.



Bereken $E(Y)$, $Var(Y)$ en $SD(Y)$ in het geval:

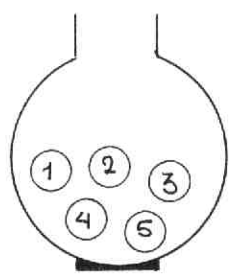
a. $Y = X + 2$

b. $Y = 3X$

c. $Y = -X$

d. $Y = X - 3$

≫ 66. In een vaas zitten balletjes met nummer 1 tot en met 5. Iemand trekt zonder teruglegging twee balletjes uit de vaas.
Het nummer op het eerste balletje noemen we X_1, dat op het tweede X_2.



a. Hoeveel verschillende trekkingen zijn er mogelijk? Schrijf alle mogelijkheden op.

b. Bereken $Var(X_1)$, $Var(X_2)$ en $Var(X_1 + X_2)$.

c. Waarom geldt hier niet: $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$?

d. K is het kleinste nummer van de twee getrokken balletje. Bereken $E(K)$ en $SD(K)$.

≫ 67. Nog een keer twee dobbelstenen. We letten nu op het verschil van de aantallen ogen van de eerste en de tweede steen: $V = X_1 - X_2$.

a. Geef de complete kansverdeling van V.

b. Bereken $E(V)$, $Var(V)$ en $SD(V)$.

c. Welk verband bestaat er tussen $Var(X_1 - X_2)$, $Var(X_1)$ en $Var(X_2)$?

d. Kun je dat verklaren zonder dat je die varianties stuk voor stuk hebt uitgerekend? (Aanwijzing: $X_1 - X_2 = X_1 + (-X_2)$).

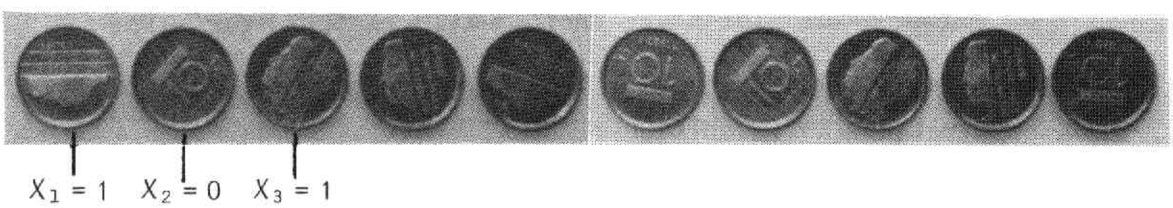
≫ 68. Asterix werpt tien Romeinse geldstukken omhoog en telt het aantal 'koppen' dat boven komt ($=X$).
Stel je voor dat hij dit experiment een groot aantal keren uitvoert. Wat is het gemiddelde aantal 'koppen' dat hij kan verwachten, m.a.w. hoe groot is $E(X)$?

De vraag is nu: welke spreiding kan Asterix verwachten in de aantallen 'kop', m.a.w. hoe groot is $SD(X)$?
Er staan verschillende wegen open om die standaarddeviatie te bepalen:

- Je voert het experiment een groot aantal keren uit, bijv. door middel van een computersimulatie en berekent de S.D. (Zie blz. 49).
- Je gebruikt de binomiale tabel voor $n=10$ en $p=0{,}5$ voor de kansen $P(X=0)$, $P(X=1)$, enz. te vinden.
  Daarmee bereken je dan de S.D.
- In plaats van de tabel, gebruik je de kansformule: $P(X=k)=\binom{10}{k}\cdot\frac{1}{2}^{10}$

≫ 69. Welk van die drie methoden levert het meest exacte antwoord op?
Welke zal het minst nauwkeurig zijn?

Alle drie de methoden zijn nogal bewerkelijk.
Gelukkig is er een veel kortere.
Net als bij de dobbelstenen kun je handig gebruik maken van de 'somregel' bij varianties. Daartoe voer je de stochasten X_1, X_2, ..., X_{10} in.



Door alle 'enen' en 'nullen' in een reeks-van-tien op te tellen, vind je X. De stochasten X_1, X_2, ..., X_{10} hebben allemaal dezelfde verwachtingswaarde (nl. $\frac{1}{2}\cdot 0+\frac{1}{2}\cdot 1=0{,}5$) en dezelfde variantie (nl. $\frac{1}{2}\cdot(-0{,}5)^2+\frac{1}{2}\cdot(0{,}5)^2=0{,}25$).

≫ 70. Bereken nu: $Var(X)$ en $SD(X)$. Klopt dat met het resultaat van de computersimulatie?

COMPUTEROUTPUT

Reeks	Waarde van X
1011000001	4
1000111011	6
0110100101	5
1010000010	3
0111001001	5
1110000110	5
1010001001	4
0101010100	4
0001100000	2
0000001010	2
0100010100	3
1110110101	7
1000011010	4
1111100010	6
0000001011	3
1010111001	6
1000111100	5
1001010010	4
1110110011	7
0111111101	8
1011010111	7
0100011100	4
1101000010	4
1111100001	6
1111101111	9
1001110000	4
0110111010	6
0100000111	4
1110010110	6
0010100111	5
0110110000	4
1100100101	5
1100100101	5
1101011100	6
1101001010	5
0101000000	2
0000101111	5
1110011001	6
1010110111	7
0111101011	7
0011101100	5
0100010111	5
1010011000	4
0101011100	5
0111011111	8
1010100001	4
1001111011	7
1100000001	3
0011010010	4
0111111101	8

gemiddelde: 5.03
st.deviatie: 1.56496

Reeks	Waarde van X
1001011101	6
0100000010	2
1110000101	5
0001100001	3
0101111111	8
0010011011	5
1110010110	6
0111010000	4
0000011100	3
1111001010	6
1000001010	3
0111100000	4
0011100000	3
0111101111	8
1000011111	6
0111000100	4
1000101011	5
0001111010	5
0000101101	4
1101110010	6
1101010111	7
1110100111	7
0011100111	6
1100110100	5
1001010100	4
0111110111	8
0001101110	5
0110100110	5
0011100100	4
0011110010	5
0111010001	5
1001111100	6
0010000001	2
0101110001	5
1101000101	5
1100101110	6
0000010010	2
0000011110	4
1100101110	6
0000001101	3
1111000101	6
0100101100	4
0111001110	6
1011101100	6
1100000110	4
1110110010	6
0001101011	5
0001000111	4
1101100001	5
1011011111	8

≫ 71. Bereken de S.D. van het aantal keren 'kop' (= U) bij het werpen met 100 muntstukken.

≫ 72. a. Druk de standaarddeviatie van X (opgave ≫ 70) en U (opgave ≫ 71) uit in % van het aantal munten dat je werpt.

b. Doe hetzelfde ook voor het werpen met 1000 muntstukken.

c. Is het 'logisch' dat de procentuele spreiding afneemt bij toename van het aantal muntstukken?

≫ 73. Bij een vierkeuzewerk worden 40 vragen op de gok beantwoord. X is het totaal aantal goede antwoorden (de 'score').

a. X_1 is de serie bij de eerste vraag.
Laat door berekening zien dat: $E(X_1) = 0{,}25$ en $Var(X_1) = 0{,}1875$.

(Maak een tabel als op blz. 43).

b. Bereken E(X) en SD(X).

We letten op een binomiaal experiment met parameters n en p.
Dat wil zeggen:

- een uitkomstenreeks bestaat uit n onafhankelijke uitkomsten 'succes' of 'mislukking';
- de kans op 'succes' is p en de kans op 'mislukking' 1 - p (= q).

We noemen het aantal successen: X.
Voor het 'gemiddelde' en de 'spreiding' van de stochast X gelden de volgende formules:

$$E(X) = np$$
$$Var(X) = npq$$
$$SD(X) = \sqrt{npq}$$

≫ 74. Controleer of deze formules kloppen met je uitkomsten bij opgave ≫ 71 en ≫ 73.

AFLEIDING VAN DE n-p-q-FORMULES

We voeren de hulpstochasten X_1, X_2, ..., X_n in met:

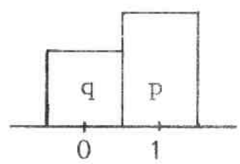
k^e uitkomst — *succes* $\longrightarrow X_k = 1$ — (k = 1, 2, ..., n)

k^e uitkomst — *mislukking* $\longrightarrow X_k = 0$

$X = X_1 + X_2 + \ldots + X_n$ = totale aantal successen.

Kijk nu eerst naar X_1

$E(X_1) = q \cdot 0 + p \cdot 1 = p$



We berekenen nu de gemiddelde gekwadrateerde afwijking van X_1:

uitkomst	*afwijking van gem.*	*gekwadr. afwijking*	*kans*
0	$-p$	p^2	p
1	$1 - p = q$	q^2	q

Dus: $\mathrm{Var}(X_1) = q \cdot p^2 + p \cdot q^2 = pq(p+q) = pq \cdot 1 = pq$

Hetzelfde geldt natuurlijk voor X_2, ... X_n.

Uit de somregel voor verwachtingswaarde en variantie volgt nu:

$E(X) = p + p + \ldots + p = np$ en $\mathrm{Var}\ (X) = pq + pq + \ldots + pq = npq$.

SAMENVATTING

Bij het berekenen van de (verwachte) *standaarddeviatie* van de stochast X is het vaak handig om als tussenstap de *variantie* te berekenen.

Kan X de waarden $x_1, x_2, \ldots, x_n$ aannemen, resp. met kansen $p_1, p_2, \ldots,$ p_n en geldt: $E(X) = \mu$, dan:

$$\mathrm{Var}(X) = p_1 \cdot (x_1 - \mu)^2 + p_2 \cdot (x_2 - \mu)^2 + \ldots + p_n \cdot (x_n - \mu)^2$$

ofwel:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

De standaarddeviatie van X wordt dan gevonden uit:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

Zijn X en Y *onafhankelijke* stochasten, dan kan behalve de verwachtingswaarde van $X + Y$ ook de variantie van $X + Y$ met behulp van de zgn. somregel worden gevonden:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Tussen de standaarddeviaties van X, Y en $X + Y$ bestaat een wat ingewikkelder verband:

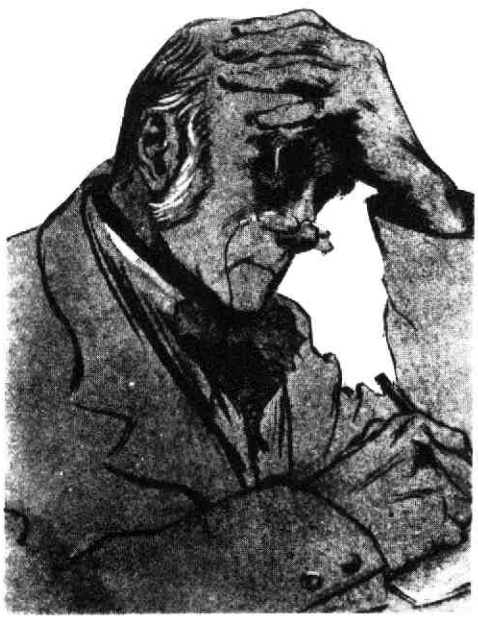
$$\mathrm{SD}(X + Y) = \sqrt{\mathrm{SD}^2(X) + \mathrm{SD}^2(Y)}$$

Deze regels gelden i.h.a. niet in het geval dat X en Y afhankelijk zijn!

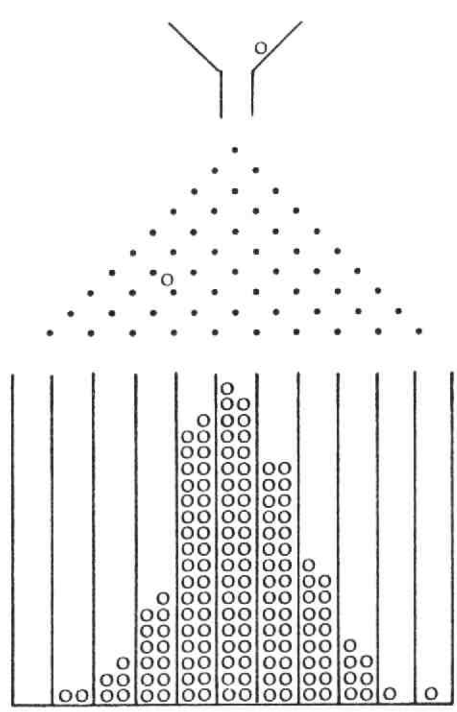
Voor een binomiale stochast X met parameters n(= aantal 'beurten'), p(= kans op succes) en q(= kans op mislukking) geldt:

$$E(X) = np \quad \text{en} \quad \mathrm{SD}(X) = \sqrt{npq}$$

6 DE NORMALE BENADERING VAN KANSEN



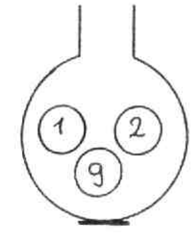
Sir Francis Galton (England, 1822–1911)
Biometrika, 1903.



De kogeltjes op het bord van Galton verdelen zich, na een zig-zag-weg tussen de pinnen, volgens een bijkans normale kromme. Hoe meer pinnenrijen er worden aangebracht hoe 'normaler' die verdeling gaat lijken ...

≫ 75. Het bord van Galton demonstreert een binomiaal kansexperiment. Wat zijn de parameters (p en n) in de situatie op het plaatje?

In een vaas zitten drie balletjes met daarop de nummers 1, 2 en 9. Er wordt, *met teruglegging*, een aantal keren (zeg n) een balletje uit de vaas getrokken. Het nummer van het eerste balletje noemen we X_1, van het tweede balletje X_2, enz.

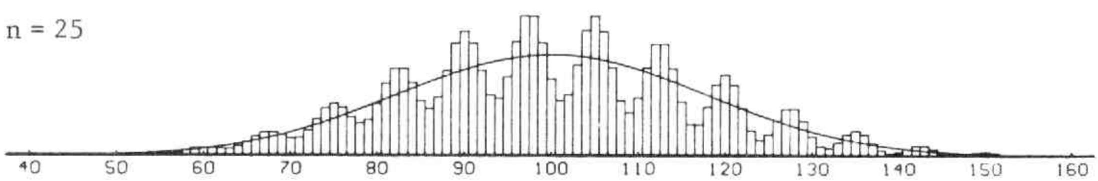


De getrokken nummers worden opgeteld. De som van die nummers is een stochast S_n.

Dus: $S_n = X_1 + X_2 + \ldots + X_n$.

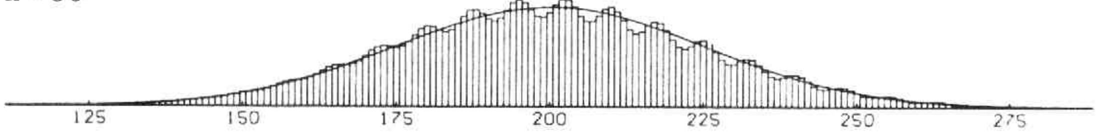
Hieronder zie je de kanshistogrammen van S_{25}, S_{50} en S_{100}.

n = 25



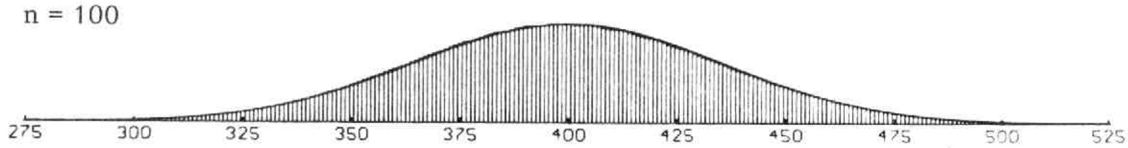
waarde van de som (S_{25})

n = 50



waarde van de som (S_{50})

n = 100



waarde van de som (S_{100})

Je ziet dat bij een toenemend aantal balletjes de verdeling van de som S_n steeds beter gaat lijken op de normale verdeling!

≫ 76. a. Geef de kansverdeling van S_2 (= som van de nummers bij trekken van twee balletjes) en teken het bijbehorende kanshistogram.

b. Dezelfde opdracht voor S_3.

≫ 77. Bekijk de figuren op blz. 54.

a. Behalve de normale 'trend' vertonen de histogrammen voor n = 25 en n = 50 een bijna periodiek beeld.
Hoe groot schat je de 'periode'?
Kun je dat verklaren uit de samenstelling van de vaas?

b. De histogrammen van S_2 en S_3 bestaan uit geïsoleerde 'bergjes'.
Die van S_{25} lijkt een aaneengesloten 'bergketen' te zijn. Al zie je dat niet in de figuur, toch kun je weten dat het histogram van S_{25} ook geïsoleerde bergjes bevat!
Waarom?

c. Enig idee waarom die losse bergjes niet getekend zijn?

≫ 78. a. Bereken de verwachtingswaarde en de standaarddeviatie in het geval je één nummertje uit de vaas {1,2,9} trekt. (We noemen dit wel: de verwachtingswaarde en de SD van die vaas).

b. Bereken hieruit de verwachtingswaarde en de standaarddeviatie van S_{25}, S_{50} en S_{100}.

c. Controleer je antwoorden op vraag b. in de drie histogrammen.

≫ 79. We vervangen het nummer 9 in de vaas {1,2,9} door het nummer 3.

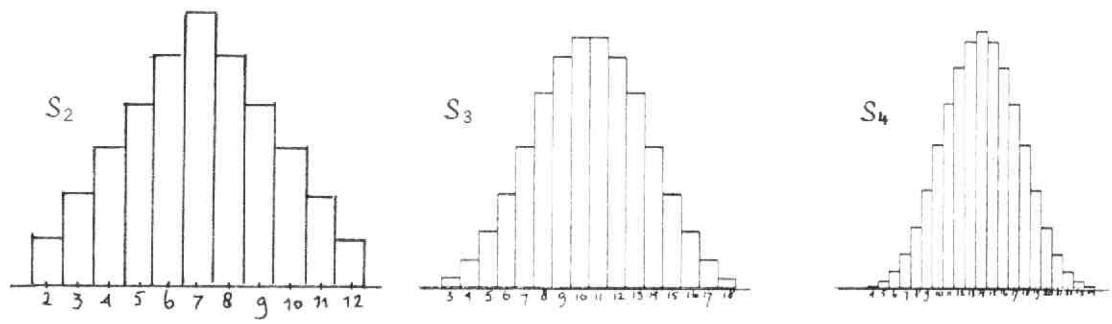
a. Welke invloed zal dat hebben op het histogram van S_{25} denk je?

b. Schets de normale kromme die de verdeling van S_{25} in dat geval benadert. (Denk aan de getallen bij de horizontale as!).

Kansexperimenten laten zich meestal gemakkelijk 'vertalen' in het zgn. *vaasmodel*.

Neem bijv. het werpen met een aantal dobbelstenen. Je trekt een aantal keren (met teruglegging) een balletje uit een vaas, gevuld met de nummers 1 tot en met 6.

S_n is opnieuw de som van de getrokken nummers bij het trekken van n balletjes.



Merk op:

De histogrammen van S_2, S_3, S_4, ... vertonen in het geval er getrokken wordt uit de vaas $\{1,2,3,4,5,6\}$ heel wat eerder gelijkenis met de normale verdeling dan bij de vaas $\{1,2,9\}$.
Het histogram voor $n = 4$ ziet er al aardig klokvormig uit, terwijl in het geval van de vaas $\{1,2,9\}$ je in het histogram voor $n = 50$ nog duidelijke afwijkingen van de normale kromme kunt constateren!

≫ 80. a. Laat zien dat geldt: $P(S_4 \leqq 7) = \frac{35}{1296}$.

b. Verwachtingswaarde en S.D. van de vaas $\{1,2,3,4,5,6\}$ zijn resp. 3,5 en $\sqrt{4 \times 2\frac{11}{12}} = 3{,}42$.
Op grond hiervan had je de kans $P(S_4 \leqq 7)$ kunnen schatten op $2\frac{1}{2}\%$. Leg dat uit.

≫ 81. Nu het bord van Galton, ofwel de binomiale verdeling met $n = 10$ en $p = \frac{1}{2}$. Hieronder is de cumulatieve tabel van deze kansverdeling afgedrukt.

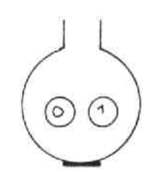
$n = 10,\ p = \frac{1}{2}$	
k	$P(X \leq k)$
0	0,0010
1	0,0107
2	0,0547
3	0,1719
4	0,3770
5	0,6230
6	0,8281
7	0,9453
8	0,9893
9	0,9990
10	1

a. Maak een cumulatieve grafiek van deze kansverdeling op normaal-waarschijnlijkheidspapier.
(N.B. Schrijf bij de 'meetpunten' op de horizontale as niet 0, 1, 2, 3, enz. maar 0,5; 1,5; 2,5; enz.

b. Lees uit die grafiek de verwachtingswaarde en de standaarddeviatie van de verdeling af.

c. Controleer je antwoorden bij b met behulp van de n-p-q-formules.

Ook de binomiale verdeling met $n = 10$ en $p = \frac{1}{2}$ kun je vertalen in het vaasmodel: 'trek tien keer (met teruglegging) een balletje uit de vaas $\{0,1\}$ en tel de nummers op'.



In opgave ≫ 81 is gebleken dat ook in dit geval de verdeling van de 'som' goed lijkt op de normale verdeling.

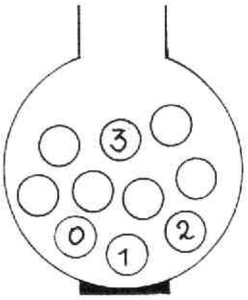
≫ 82. Ook een binomiaal experiment met bijv. $n = 50$ en $p = 0,2$ kan worden vertaald in het trekken van nummertjes uit een vaas.

a. Hoe zou je die vaas samenstellen?

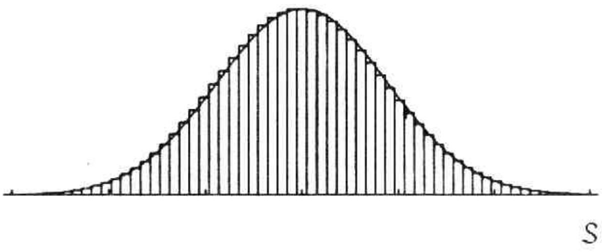
b. Ga met behulp van de binomiale tabel na of de kans dat het aantal successen minder dan 2 S.D. afwijkt van de verwachtingswaarde ongeveer 95% is.

DE CENTRALE LIMIETSTELLING

Er geldt:
Bij het aselect trekken (met teruglegging) van een aantal nummertjes uit een vaas, zal de *som van de nummertjes* een verdeling hebben die lijkt op de normale verdeling.
Die gelijkenis wordt sterker naarmate het aantal nummers dat getrokken wordt groter is!
Deze regel staat in de statistiek bekend als de *centrale limietstelling*.



$S = X_1 + X_2 + \ldots + X_n$



Nog wat algemener geformuleerd luidt de stelling:

Wanneer een aantal onafhankelijke stochasten bij elkaar wordt opgeteld, dan gaat de verdeling van de som beter lijken op een normale verdeling, naarmate het aantal termen groter wordt.

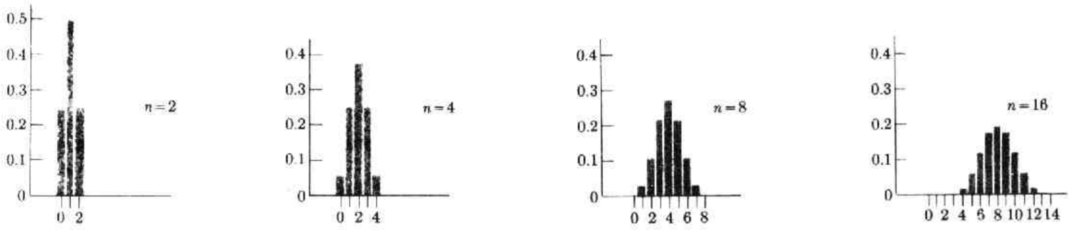
Opmerkingen:

1. Behalve het aantal balletjes heeft ook de samenstelling van de vaas invloed op de mate van gelijkenis met de normale verdeling.
   Bij een 'onregelmatige samenstelling' (zoals {1,2,9}) zal het aantal te trekken balletjes groot moeten zijn om de normale verdeling redelijk te benaderen.

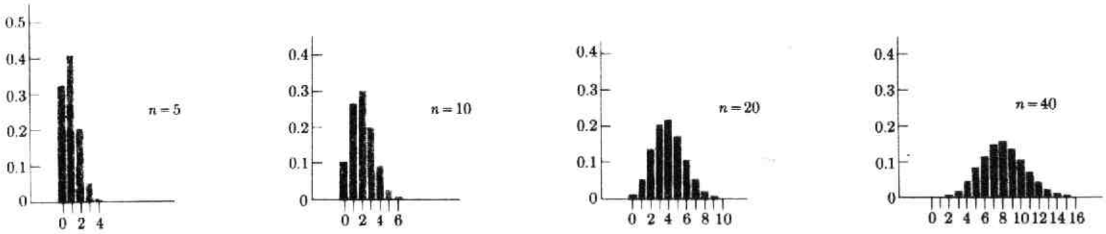
2. Met de centrale limietstelling kan ook worden verklaard waarom de normale verdeling zo vaak in de natuur voorkomt.
Neem bijv. de lichaamslengte. Die wordt beïnvloed door een groot aantal 'kansfactoren' tengevolge van erfelijkheid en milieu.
De lichaamslengte kan worden opgevat als de som van een aantal min of meer onafhankelijke stochasten en is dan volgens de centrale limietstelling normaal verdeeld.

Als gevolg van de centrale limietstelling zijn binomiale kansen te berekenen (eigenlijk: te benaderen) via de normale tabel.
Voor $p = \frac{1}{2}$ is die benadering al gauw nauwkeurig.
Voor $p \neq \frac{1}{2}$, bijv. $p = 0,2$, is de binomiale verdeling voor kleine waarden van n nog 'scheef' en moet n tamelijk groot zijn, wil de normale benadering enige nauwkeurigheid hebben.
(De vaas met twee 'enen' en acht 'nullen' is niet erg regelmatig samengesteld!).
Vergelijk de histogrammen:

p = 0,5



p = 0,2



Tot 'op zekere hoogte' kun je binomiale kansen berekenen met de normale verdeling.

Dat is een groot voordeel, omdat je nu eenmaal niet over binomiale tabellen voor alle waarden van n en p kunt beschikken, terwijl je voor alle normale verdelingen met één standaard-tabel (de Φ-tabel) kunt volstaan!

Als voorbeeld nemen we een geval dat in onze binomiale tabel te vinden is: $n = 100$, $p = 0,2$.

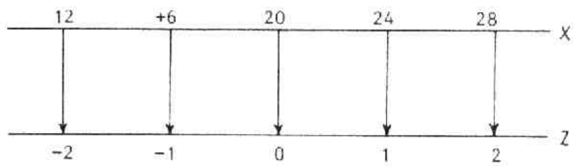
Volgens de binomiale tabel geldt: $P(X \leqq 25) = 0,9125$.

Om deze kans met de Φ-tabel te kunnen berekenen, moet er eerst worden gestandaardiseerd! Daartoe berekenen we μ ($= E(X)$ en σ ($= SD(X)$).

$\mu = n \cdot p = 20$

$\sigma = \sqrt{npq} = \sqrt{16} = 4.$

Stel nu $Z = \dfrac{X - \mu}{\sigma} = \dfrac{X - 20}{4}$



Z is bij benadering standaard normaal verdeeld.

$$
\begin{aligned}
P(X \leqq 25) &= P(X \leqq 25,5) \\
&= P(Z \leqq \tfrac{5,5}{4}) \\
&= \Phi(1,375) \\
&\approx 0,91545
\end{aligned}
$$

overgang van discreet naar continu.

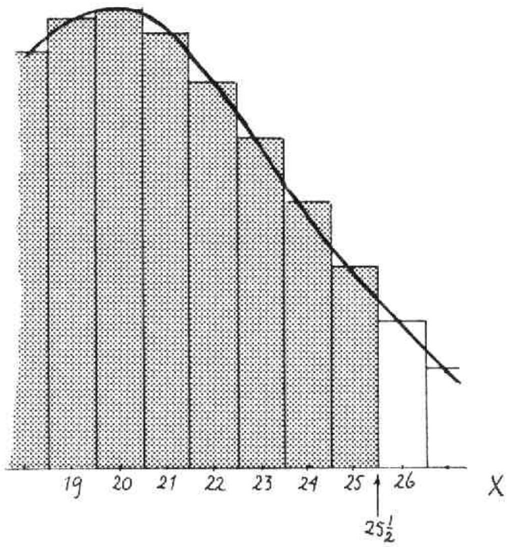
uit de tabel via interpolatie.

Je ziet: een afwijking van niet meer dan 0,003 met 0,9125!

Opmerking:

Bij vervanging van 25 door 25,5 spreekt men wel van een *continuïteitscorrectie.*

Aan de hand van onderstaand plaatje kun je gemakkelijk verklaren waarom die continuïteitscorrectie een nauwkeuriger antwoord oplevert.



P(X ≦ 25) is gelijk aan de oppervlakte van het histogram tot en met de 'balk' bij 25 en dit komt overeen met de oppervlakte onder de normale kromme tot 25,5.

Voor erg grote waarden van n is de invloed van de continuïteitscorrectie gering. In zulke gevallen laat men het toepassen van de continuiteits-correctie soms achterwege. In het algemeen is het veiliger om de correctie wel toe te passen.

≫ 83. Ga na welk resultaat je met de Φ-tabel had gekregen zonder toepassing van de continuïteitscorrectie.

≫ 84. Bereken m.b.v. de standaard-normale tabel:

a. P(X ≦ 20) in het geval: p = 0,4; n = 50.

b. P(X ≧ 320) in het geval: p = 0,75; n = 400.

c. P(X = 12) in het geval: p = 0,5; n = 25.

≫ 85. Volgens de politie is bij één van de vier fietsen het achterlicht niet in orde. Bij een steekproef worden 60 fietsen gecontroleerd. Hoe groot is de kans dat van tenminste 20 fietsen het achterlicht niet goed functioneert?

≫ 86. Het trekken van tien balletjes (met teruglegging) uit een vaas met één wit en één zwart balletje werd met de computer 10000 keer gesimuleerd.

Dit leverde de volgende frequentie-tabel op:

```
simulatie balletjes trekken

gemiddelde:    5.0041
st.deviatie:   1.575272

frekwentie tabel bij het 10000 keer trekken van 10balletjes uit een vaas met 1
 witte balletjes en
1 zwarte balletjes


              0      6
              1     78
              2    453
              3   1231
              4   1997
              5   2391
              6   2124
              7   1192
              8    425
              9     97
             10      6
```

Welke frequentie had je volgens de normale tabel bij '5 witte balletjes' kunnen verwachten? En volgens de binomiale tabel?

≫ eerste 3000 geboorten in 1962 in Graz (Oostenrijk) waren er 1578 jongens.

Hoe groot is de kans op tenminste 1578 jongens bij 3000 geboorten, als je aanneemt dat er gemiddeld evenveel jongens als meisjes worden geboren?

≫ 88. In 1976 werden er in Nederland ca 177.000 kinderen geboren. Volgens statistieken is de kans op een jongensgeboorte in ons land 51,5%.

a. Welk aantal jongensgeboorten had men in dat jaar kunnen verwachten?

b. In werkelijkheid waren er ongeveer 91.000 jongens- en 86.000 meisjesgeboorten. Hoe bijzonder is dit, m.a.w. hoe groot had je bij voorbaat de kans op *niet meer dan* 91.000 jongensgeboorten kunnen schatten?

≫ 89. Een bedrijf koopt zekeringen in partijen van 8000 stuks. Bij aankomst van zo'n partij wordt een aselecte steekproef van 50 stuks aan keuring onderworpen.
Als alle 50 exemplaren van de steekproef deugdelijk zijn, wordt de partij geaccepteerd, anders niet.
Aan het bedrijf wordt een partij toegestuurd, waarvan 5% niet aan de eisen voldoen.

a. Hoewel er hier zonder teruglegging wordt getrokken, maak je slechts een kleine fout als je doet alsof er met teruglegging wordt gekozen. Waarom?

b. Bereken de kans dat zo'n partij van 8000 stuks wordt geaccepteerd.

≫ 90. Iemand werpt 200 dobbelstenen.

a. Hoe groot is de kans dat hij een totaal van ten hoogste 675 ogen gooit?

b. Bij 50 van de 200 stenen is het resultaat een zes.
Vind je dit resultaat erg toevallig?
(Aanwijzing: bereken de kans op tenminste 50 zessen).

≫ 91. Een basketballspeler zegt dat hij bij 80% van zijn afstandschoten doel treft. Zijn trainer beweert dat het schotpercentage van de bewuste speler 60% is. Ze spreken af dat, als de speler kans ziet met tenminste 70% van zijn schoten tijdens de eerstvolgende tien wedstrijden te scoren, hij van zijn trainer gelijk krijgt. Scoort hij minder dan wordt de trainer in het gelijk gesteld.
Hij blijkt in die tien wedstrijden 90 keer van afstand te hebben geschoten.

a. Hoe groot is de kans dat de speler gelijk krijgt, terwijl de trainer het bij het rechte eind had?

b. Hoe groot is de kans dat de trainer ten onrechte gelijk krijgt?

≫ 92. Experiment: n keer werpen met een geldstuk.
Uitkomst waar op gelet wordt: het aantal keren 'kop' ($= X$).

a. Ga na: $E(X) = \frac{1}{2}n$ en $SD(X) = \frac{1}{2}\sqrt{n}$.

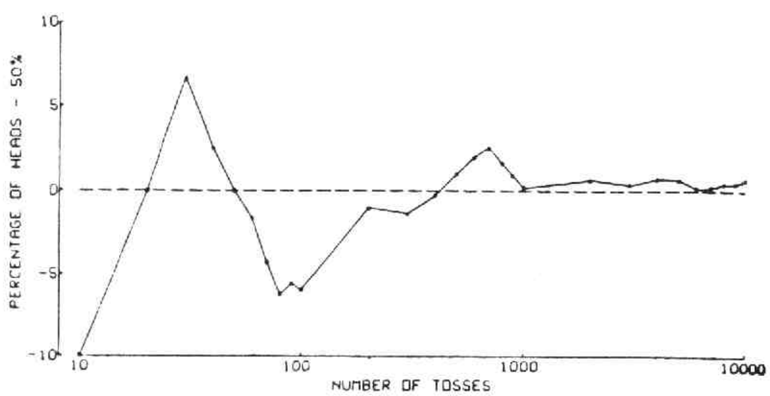
b. Met behulp van een bekende vuistregel kun je schatten dat de kans op een resultaat tussen 40 en 60 keer 'kop' bij 100 worpen ongeveer 95% zal zijn.
Leg dat uit.

c. Hoe groot schat je bij honderd worpen de kans op een resultaat tussen 45 en 55 keer 'kop'?

≫ 93. Experiment als in opgave ≫ 92.
Iemand beweert: de kans dat het aantal keer 'kop' tussen 49% en 51% van het totale aantal ($= n$) worpen ligt, is gelijk aan 0,95.
Hoe groot is n? (Gebruik de vuistregel).

John Kerrick bracht zijn tijd van gevangenschap in de Tweede Wereldoorlog door met het uitvoeren van een kansexperiment: hij toste tienduizend keer met één geldstuk en noteerde regelmatig het aantal keren 'kop'.

Resultaat:



Number of tosses	Number of heads
10	4
20	10
30	17
40	21
50	25
60	29
70	32
80	35
90	40
100	44
200	98
300	146
400	199
500	255
600	312
700	368
800	413
900	458
1000	502
2000	1013
3000	1510
4000	2029
5000	2533
6000	3009
7000	3516
8000	4034
9000	4538
10000	5067

De grafiek van het experiment van John Kerrick illustreert prachtig de 'wet van de grote getallen': als je een kansexperiment erg vaak uitvoert, zal het gemiddelde aantal 'successen' in de buurt komen van de theoretische kans.

De sterke invloed van het aantal kan worden verklaard uit het feit dat de S.D. van het aantal successen bij toenemende n niet evenredig is met n, maar met $\sqrt{n}$.

Bij 10000 worpen betekent dit het volgende.

Stel X = het aantal keer 'kop' en $Y = \frac{X}{10000}$ (= de relatieve frequentie van het aantal keer 'kop').

Er geldt: $E(X) = 10000 \cdot \frac{1}{2} = 5000$

$SD(X) = 100 \cdot \frac{1}{2} = 50$

En dus: $E(Y) = \frac{5000}{10000} = 0{,}5$ (50%)

$SD(Y) = \frac{50}{10000} = 0{,}005$ (0,5%)

Volgens de statistische vuistregel is er een kans van ongeveer 95% dat het resultaat niet meer dan 0,01 (= 2 SD) van 0,5 afwijkt.

Kerricks resultaat voldeed hier ruimschoots aan.

≫ 94. Hoe groot is de 95%-marge bij 5000 keer uitvoeren van het experiment.

Vergelijk Kerrick's resultaat met je antwoord.

In bovenstaand voorbeeld heb je gezien dat bij toenemende n de spreiding van het gemiddelde aantal successen afneemt.

Iets preciezer:

de SD van het gemiddelde aantal successen is *omgekeerd evenredig* met de wortel uit het aantal beurten.

In het voorbeeld:

De SD van het gemidd. aantal successen bij 1 beurt = 0,5.

De SD van het gemidd. aantal successen bij 10000 beurten = $\frac{0{,}5}{\sqrt{10000}} = 0{,}005$.

De SD van het gemidd. aantal successen bij 5000 beurten = $\frac{0{,}5}{\sqrt{5000}} = 0{,}007$.

SAMENVATTING

Centrale limietstelling:

De kansverdeling van de som van een aantal onafhankelijke stochasten kan worden benaderd met de normale verdeling.

Binomiale verdeling benaderd door standaard-normale verdeling:

Binomiale kansen kunnen dus met de Φ-tabel worden benaderd.
Daarbij moet je de stochast X met parameters n en p eerst standaardiseren:

$$Z = \frac{X - np}{\sqrt{npq}} \qquad (q = 1 - p)$$

Bij gebruikmaking van de continuïteitscorrectie geldt dan bijvoorbeeld:

$$P(X \leqq k) = \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{npq}}\right)$$

$$P(X \geqq m) = 1 - \Phi\left(\frac{m - \frac{1}{2} - np}{\sqrt{npq}}\right)$$

$\sqrt{n}$ - wet:

Bij toename van het aantal beurten (= n) bij een binomiaal experiment neemt de SD van het aantal successen evenredig toe met $\sqrt{n}$.
De SD van het gemiddelde aantal successen per beurt is dan omgekeerd evenredig met $\sqrt{n}$.

7 GEMENGDE OPGAVEN

≫ 95. Van de 20-jarige inwoners van een zeker land behoort 48% tot het vrouwelijk geslacht.
De gemiddelde lengte van de 20-jarige vrouwen is 170 cm met een S.D. van 6 cm.
Voor de 20-jarige mannen zijn die gegevens resp. 180 cm en 8 cm.

a. Hoeveel % van alle 20-jarige inwoners van dat land zal langer zijn dan 1,76 cm?

b. Laat het getal L (in cm) een lichaamslengte voorstellen.
De kans dat een willekeurig gekozen 20-jarige man *kleiner* is dan L is even groot als de kans dat een willekeurig gekozen vrouw (van 20 jaar) *groter* is dan L.
Bereken L.

≫ 96. Dezelfde gegevens als in opgave ≫ 95.
Hoe groot is de kans dat een willekeurig gekozen 20-jarige man langer is dan een willekeurig gekozen vrouw van 20 jaar?

Aanwijzing:
De lengte van een 20-jarige man is een stochast X, die van een dito vrouw Y.
Het lengteverschil $X - Y$ ($= V$) is ook een normaal verdeelde stochast.

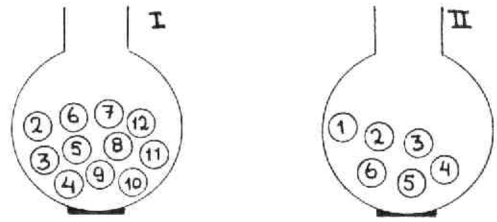
a. Hoe groot is $E(V)$?

b. Verklaar: $SD(V) = 10$ cm.

c. Bereken $P(V > 0)$.

d. Wat is dus de gevraagde kans?

≫ 97. Gegeven zijn de vazen:



X is het nummer van een willekeurig gekozen balletje uit vaas I; Y is de som van twee willekeurig gekozen balletjes uit vaas II. Hebben X en Y dezelfde kansverdeling? Verklaar je antwoord.

≫ 98. Volgens een NIPO-enquête bedroeg het aantal aanhangers van een politieke partij in een zekere stad 20%.
Een NOS-medewerker vroeg 81 aselect gekozen stemgerechtigden in die stad naar hun politieke voorkeur.
In zijn steekproef bleken niet meer dan 12 mensen voorkeur voor genoemde partij te hebben.
De NOS-medewerker trok de uitslag van de NIPO-enquête in twijfel. Volgens hem zou bij aanname van de juistheid van die 20% de kans op niet meer dan 12 aanhangers in zijn steekproef kleiner zijn dan 5%. Had hij gelijk?

≫ 99. Van het Station Utrecht C.S. vertrekken van twee kanten busdiensten (resp. stadsvervoer en streekvervoer) naar het buiten de stad gelegen universiteitscentrum 'de Uithof'. Tussen 09.00 en 09.15 uur komen er 92 mensen per trein aan die van één van beide busdiensten gebruik willen maken om een congres in de Uithof te bezoeken. Bij beide maatschappijen is voor dat doel één bus met 52 zitplaatsen gereserveerd. Om 09.00 uur vertrekken beide bussen. Veronderstel dat die 92 mensen elk met kans $\frac{1}{2}$ voor één van beide busdiensten kiezen.
Hoe groot is de kans dat er in een van de twee bussen mensen moeten staan?

≫ 100. Op de veiling van Delft werden ook deze zomer weer veel komkommers aangevoerd. Eigenlijk teveel en op een gegeven moment was het overschot zelfs 25% van de aanvoer. Het bestuur van de veiling besloot toen, na overleg met de kwekers, de wat kleinere komkommers niet te laten veilen.
Ga er bij deze opgave vanuit dat de lengte van de aangevoerde komkommers normaal is verdeeld met een gemiddelde lengte van 50 cm en een standaarddeviatie van 5 cm.

a. Welk percentage komkommers zal een lengte van meer dan 57 cm hebben?

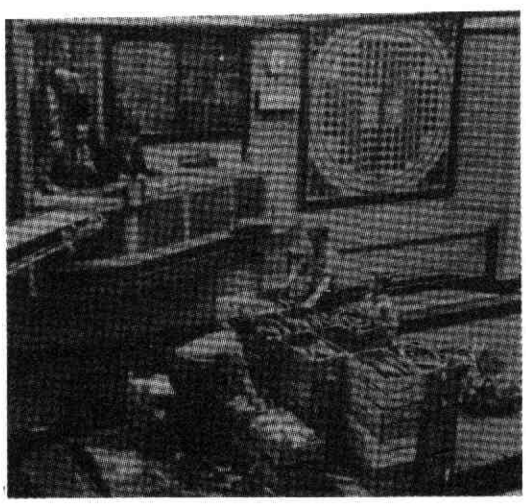
b. De 25% kleinste komkommers zullen niet worden geveild. Vanaf welke lengte worden de komkommers geveild?

c. Hoe groot is de kans dat van een aselecte partij van 100 komkommers er tenminste 80 zullen worden geveild?

d. In een kist met 25 komkommers zijn er 7 'onder de maat'.
Iemand pakt aselect 10 komkommers uit de kist.
Hoe groot is de kans dat daar 3 te kleine komkommers bij zijn?

e. Iemand haalt één komkommer uit een kist met geveilde komkommers. In de kans dat die komkommers groter dan 50 cm is kleiner dan $\frac{1}{2}$, gelijk aan $\frac{1}{2}$, of groter dan $\frac{1}{2}$? Licht je antwoord toe. Hoe groot is die kans precies?

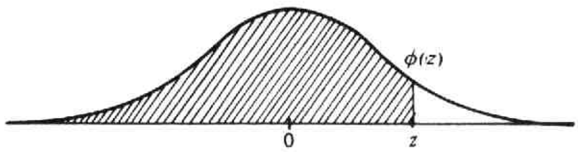
f. Schets in de figuur op het grafiekenblad de lengteverdeling van de geveilde komkommers. (Gebruik de aangegeven schaalverdeling).



Groenteveiling in 1961 (Delft).

DE VERDELINGSFUNCTIE Φ

van de standaardnormale stochast Z, vermenigvuldigd met 10 000.



z	0	1	2	3	4	5	6	7	8	9
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9891	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3,0	9987	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,1	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,2	9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3	9995	9995	9995	9996	9996	9996	9996	9996	9996	9997
3,4	9997	9997	9997	9997	9997	9997	9997	9997	9997	9998
3,5	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,6	9998	9998	9999	9999	9999	9999	9999	9999	9999	9999