Willem van Ravenstein

Lesbrief de normale verdeling

Inhoudsopgave

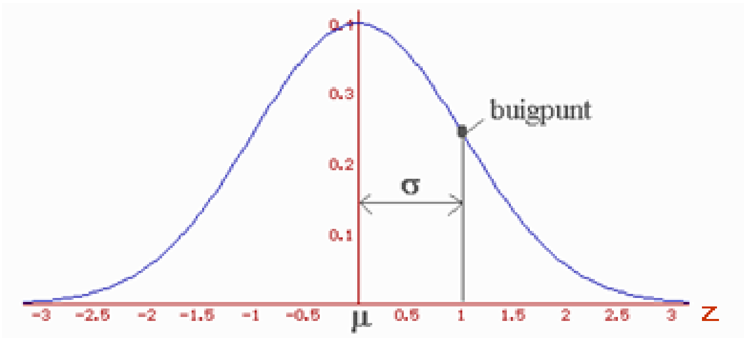
Hoofdstuk 1 – de normale verdeling

De **normale verdeling** is een **continue** kansverdeling. Kansverdelingen waarbij een continue variabele een rol speelt komen veel voor. Als je bijvoorbeeld kijkt naar het gewicht van een pak koffie van een bepaald merk, of naar de gemiddelde opbrengst van een hectare grond of naar de gemiddelde lengte van een groot aantal personen dan heb je steeds te maken met een continue kansverdeling. Bij veel discrete kansverdelingen wordt vaak gedaan alsof ze continu zijn, vooral bij grote aantallen.

Een paar eigenschappen van een normale verdeling:

- klokvormig
- de oppervlakte onder de kromme komt overeen met 100% van de gegevens
- symmetrisch t.o.v. het gemiddelde.
- gemiddelde, mediaan en modus vallen samen
- de verdeling wordt bepaald door de verwachtingswaarde en de standaarddeviatie
- vuistregels:
  - 68% van de gegevens wijkt op z'n hoogst één keer de standaarddeviatie af van de verwachtingswaarde
  - 95% van wijkt op z'n hoogst twee keer de standaarddeviatie af van de verwachtingswaarde

Voor een normale dichtheidskromme is het mogelijk de standaarddeviatie op het oog te schatten. De afstand van het buigpunt tot het centrum (gemiddelde en mediaan) is namelijk de standaarddeviatie.

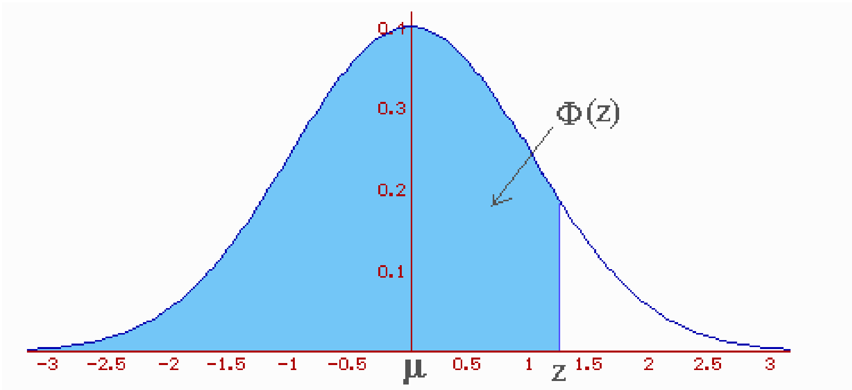


Hierboven zie je de standaardnormale verdeling. Hierbij is de verwachtingswaarde μ gelijk aan 0 en de standaarddeviatie σ gelijk aan 1. Met de standaardnormale verdeling en een tabel kun je allerlei uitspraken doen over alle normale verdelingen.

De standaard normale verdeling

Om bij een willekeurige normale verdeling uitspraken te kunnen doen gebruik je een tabel van de standaardnormale verdeling. Zo'n standaard normale verdeling is een normale verdeling met een gemiddelde van 0 en een standaarddeviatie van 1. In zo'n tabel kun je voor een willekeurige z-waarde (de horizontale as in bovenstaande grafiek) de bijbehorende waarde voor ϕ vinden.

ϕ (z) is de oppervlakte onder de kromme die hoort bij de waarde van z.



Hierboven is z = 1,25 en ϕ (1,25) = 0,8944. Dat betekent dat 89,4% van de gegevens een z-waarde heeft lager dan 1,25.

In de bijlage kan je een tabel vinden van de standaard normale verdeling. Uiteraard kan je dit ook met de grafische rekenmachine doen. Daarover straks meer.

Van normaal naar standaardnormaal

Om van een willekeurige normale verdeling een standaardnormale verdeling te maken kun je een waarde (x) omrekenen naar een standaardnormale waarde (z).

Hierbij gebruik je de formule:

$$z = \frac{x - \mu}{\sigma}$$

Hierbij is μ de verwachtingswaarde en is σ de standaarddeviatie.

Voorbeeld 1

Van een grote groep leerlingen berekenen we het gemiddelde rapportcijfer voor wiskunde. De rapportcijfers hebben een gemiddelde van 5,6 en een standaarddeviatie van 1,5. We nemen aan dat de rapportcijfers normaal verdeeld zijn.

- Hoeveel procent van de leerlingen heeft lager dan een 3?

Uitwerking

- Je weet μ = 5,6, σ = 1,5 en x=3. Bereken eerst z met de formule:
  $$z = \frac{3 - 5{,}6}{1{,}5} \approx -1{,}73$$
  In de tabel zoek je bijbehorende oppervlakte onder de grafiek op: ϕ(-1,73) = 0,0418
  Zie de volgende bladzijde voor een toelichting op de tabel

Conclusie: 4,2% van de leerlingen heeft lager dan een 3.

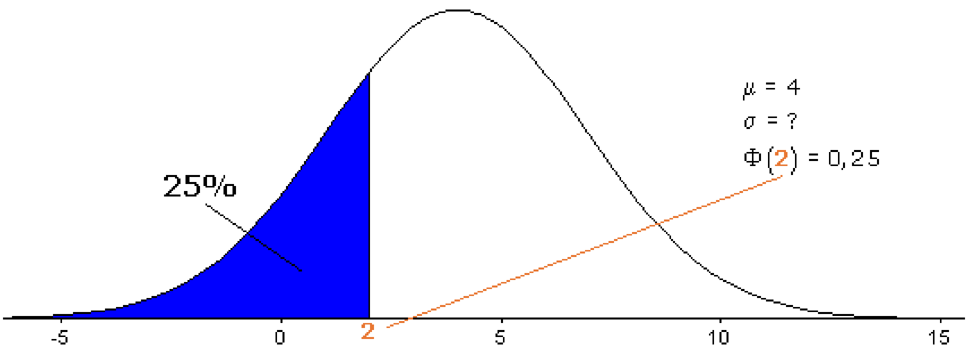
Tabel bij voorbeeld 1

z	0	1	2	3	4	5	6	7	8	9
0,0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0,1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0,2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0,3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0,4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0,5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0,6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0,7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0,8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1921	.1894	.1867
0,9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1,0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1,1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1,2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1,3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1,4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1,5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1,6	.0548	.0537	.0527	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1,7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1,8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1,9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2,0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183

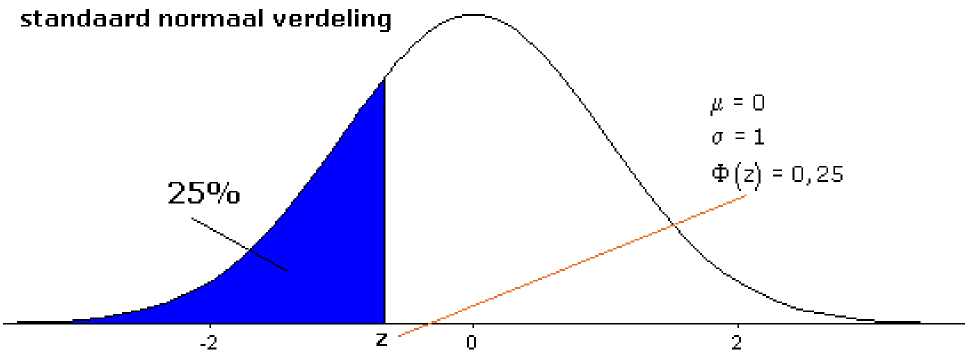
In de tabel worden alleen de waarden voor de oppervlakte onder de grafiek gegeven voor z groter dan nul. Vanwege de symmetrie is dat natuurlijk geen probleem. De oppervlakte onder de grafiek wordt gegeven als de kans dat een waarde **groter** is dan z. Dat is dan steeds even omrekenen.

Voorbeeld 2

Neem aan dat X normaal verdeeld is met een gemiddelde van 4 en een onbekende standaarddeviatie. Wel weet ik dat 25% van de gevallen kleiner is dan 2. Hoe bereken je dan de standaarddeviatie?



Vertaald naar de standaard normale verdeling:



Van de standaard normale verdeling weet ik het gemiddelde en de standaarddeviatie en ik weet welke waarde van **z** er hoort bij 25%.

Standaardnormale verdeling P(z > z)

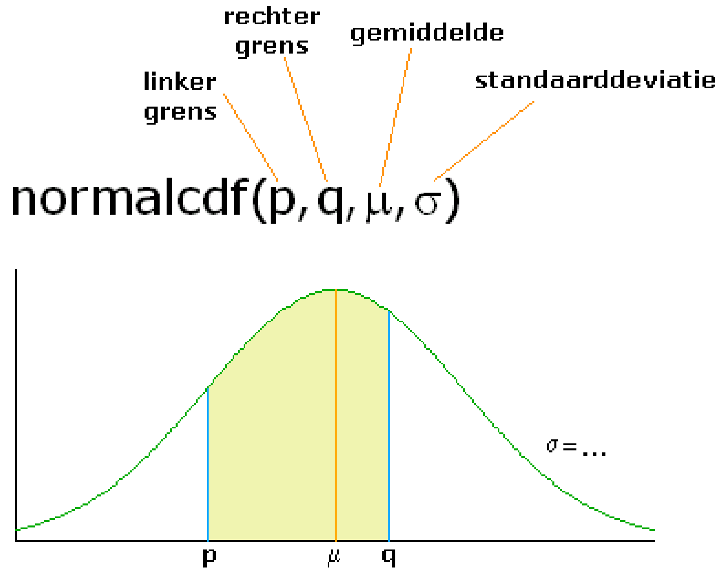
z	0	1	2	3	4	5	6	7	8	9
0,0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0,1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0,2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0,3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0,4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0,5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0,6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0,7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0,8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1921	.1894	.1867
0,9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1,0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Dus z ≈ -0,675

Gebruik de formule en vul in wat je weet:

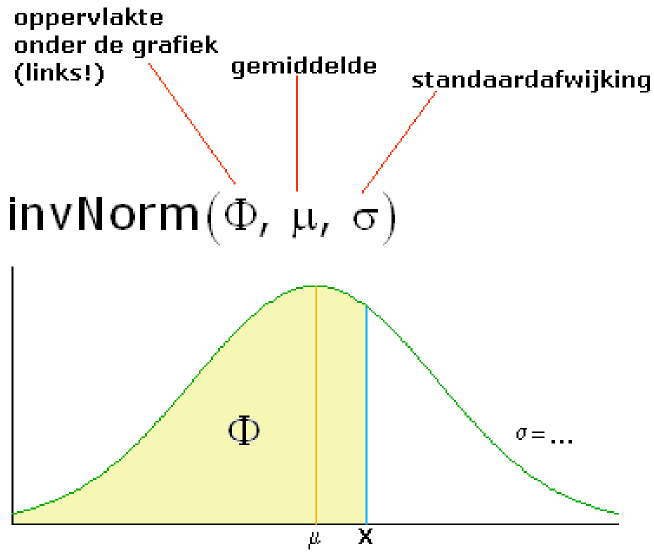
$$z = \frac{x - \mu}{\sigma} \rightarrow -0{,}675 = \frac{2 - 4}{\sigma} \rightarrow \sigma \approx 2{,}96$$

Met de grafische rekenmachine



In het geval je geen linker grens hebt neem dan voor de linkergrens 9E-99 of als je geen rechtergrens hebt neem dan 9E99.

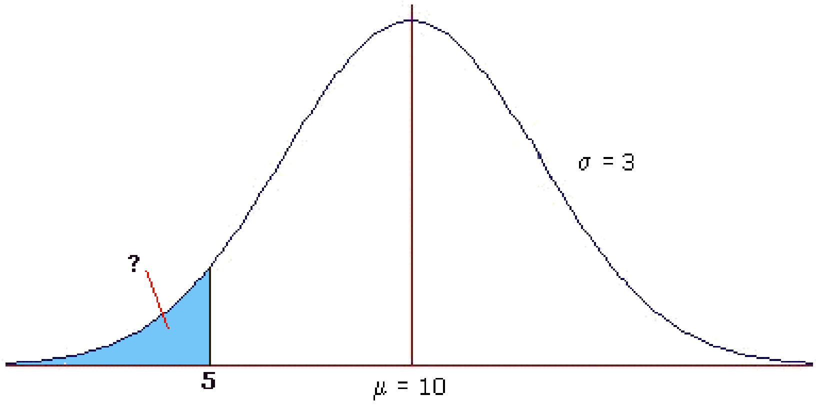
In het menu onder [DISTR] kan je allerlei kansverdelingen vinden. Met normalcdf() kan je bij gegeven grenzen, gemiddelde en standaarddeviatie de oppervlakte berekenen onder de grafiek tussen de linker- en rechtergrens van de normale verdeling.



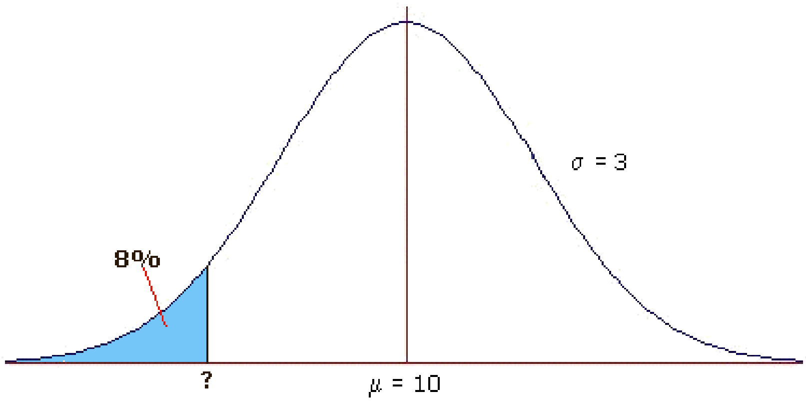
Daarnaast gebruik je ook de opdracht **invNorm()**.

Bij gegeven oppervlakte onder de grafiek, het gemiddelde en de standaarddeviatie kan je de waarde voor X berekenen.

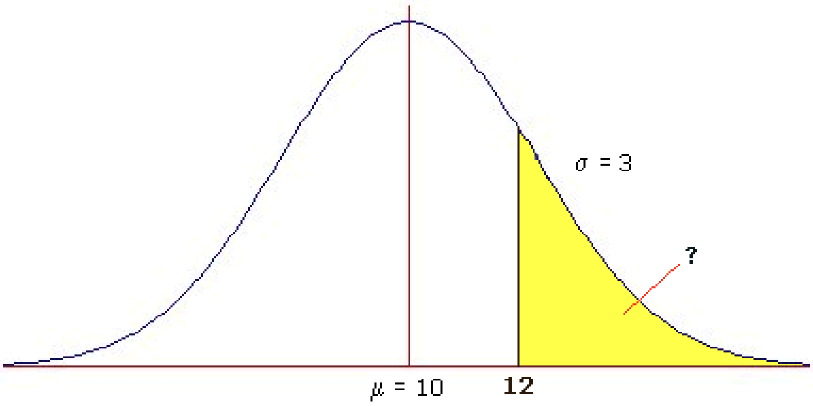
Voorbeelden



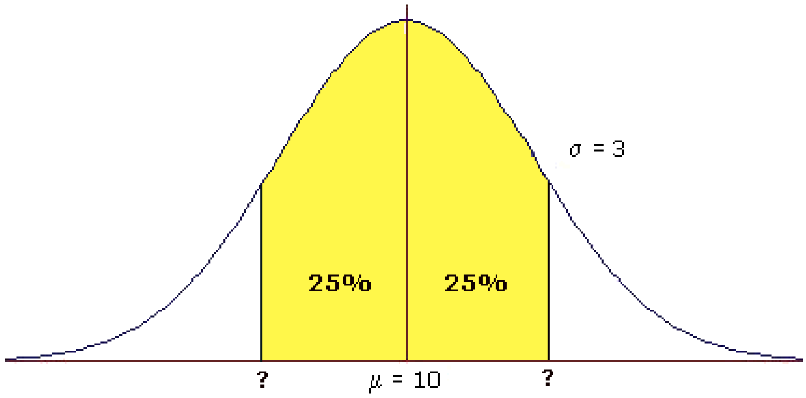
normalcdf(-9E99,5,10,3) ⟶ 0,048



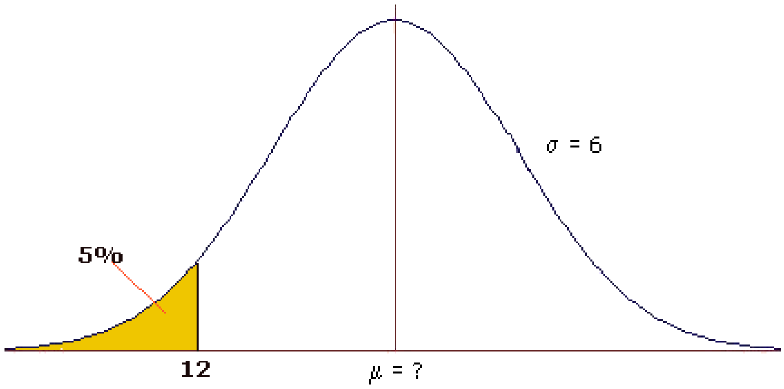
invNorm(.08,10,3) ⟶ 5,8



normalcdf(12,9E99,10,3) ⟶ 0,252

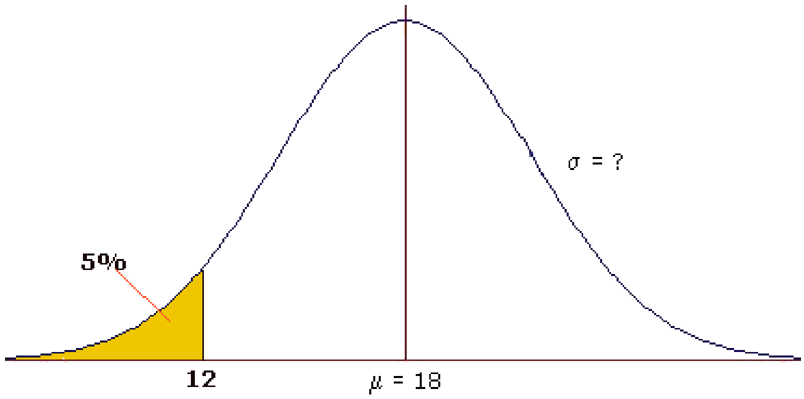


- **invNorm(.25,10,3) → 8,0**
- **invNorm(.75,10,3)→ 12,0**



$$\Phi(z)=0{,}05 \xrightarrow{\substack{\text{invNorm(.05,0,1)}\\ \downarrow}} z=-1{,}64$$

$$-1{,}64=\frac{12-\mu}{6}\Rightarrow \mu=21{,}9$$



$$\Phi(z)=0{,}05 \xrightarrow{\substack{\text{invNorm(.05,0,1)}\\ \downarrow}} z=-1{,}64$$

$$-1{,}64=\frac{12-18}{\sigma}\Rightarrow \sigma=3{,}65$$

Oefeningen

Opgave 1
Op het eindexamen is de gemiddelde uitslag 73% met een standaardafwijking van 9%. De top 10% van de geslaagden krijgen een eervolle vermelding op hun diploma. We gaan er van uit dat de uitslag normaal verdeeld is. Welke is de minimale uitslag om deze vermelding te krijgen.

Opgave 2
In een klas zitten 80 studenten. De gemiddelde uitslag is 65 met een standaardafwijking van 10. Student A behaalt 58 en student B behaalt 75. We gaan er van uit dat de uitslag normaal verdeeld is. Hoeveel studenten komen er in de rangschikking tussen A en B?

Opgave 3
Een boomkweker koopt een grote partij jonge sparrenboompjes. Uit onderzoek is bekend dat de lengte van jonge sparrenboompjes bij benadering normaal verdeeld is met een gemiddelde van 25 cm en dat 5% van de boompjes korter is dan 20 cm. De partij jonge sparrenboompjes is te beschouwen als een aselecte steekproef.

1. Hoeveel procent van de boompjes is naar verwachting langer dan 30 cm? Licht je antwoord toe.
2. Bereken de standaardafwijking van de lengteverdeling van jonge sparrenboompjes. Geef je antwoord in twee decimalen nauwkeurig.

Na een aantal jaren wordt een groot aantal van deze sparrenboompjes voor de kerstverkoop gerooid. Je kunt er nu van uitgaan dat de lengte van deze partij bomen bij benadering normaal verdeeld is met een gemiddelde van 145 cm en een standaardafwijking van 15 cm.

3. Bereken de kans dat een aselect gekozen boom uit deze partij een lengte heeft die ligt tussen de 140 en de 170 cm. Rond je antwoord af op twee decimalen.

De bomen worden ingedeeld in twee prijsklassen, namelijk: kleine bomen van € 10,- per stuk en grote bomen van € 15,- per stuk. De kweker wil dat de te verwachten opbrengst per 100 bomen € 1300,- is.

4. Bereken bij welke lengte de grens tussen de beide prijsklassen dan moet liggen. Rond je antwoord af op hele centimeters.

Uit het HAVO-examen wiskunde B1 van 2003

Opgave 4
De lampen van straatverlichting in een bepaalde gemeente hebben een levensduur die normaal verdeeld is met een gemiddelde van 1500 uur en een standaardafwijking van 100 uur. De gemeente vervangt de lampen op het moment dat 2,5 % defect is.

- Na hoeveel uur gaat de gemeente tot vervanging over?

Uitwerkingen van de oefeningen

Opdracht 1

Gegeven: m=73% en s=9%. Met de GR: invNorm(.9,73,9) geeft x=84,5. Dus, zeg maar, bij 85% of meer krijgen de kandidaten een eervolle vermelding.

Opdracht 2

Gegeven: μ = 65 en σ = 10

Gevraagd: P(58 < X < 75)

Met de GR: Normalcdf(58,75,65,10) geeft P(58X75)=0,599

59,9% van de studenten tussen 58 en 75, dat zijn er ongeveer 48.

Opdracht 3

1. De grafiek van de normale verdeling is symmetrisch. Dus 5% van de boompjes is langer dan 30 cm.
2. P(X<20)=0,05 met m=25. Bereken s:

   $\Phi(z) = 0{,}05$

   $z \approx -1{,}645$

   $z = \dfrac{X-\mu}{\sigma} \rightarrow -1{,}645 = \dfrac{20-25}{\sigma}$

   Oplossen geeft $\sigma \approx 3{,}04$
3. P(140 < X < 170) met μ = 145 en σ = 15 geeft 0,5828.
4. Kies voor het aantal kleine boompjes n.

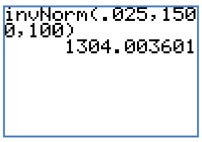
   Er geldt: n·10 + (100-n)·15 = 1300 $\Rightarrow$ n = 40

   De oppervlakte onder de grafiek bij de grens klein/groot is 0,4.

   P(X < ?)=0,4 met μ = 145 en σ = 15 geeft 141,2.

Opdracht 4

Met je GR:



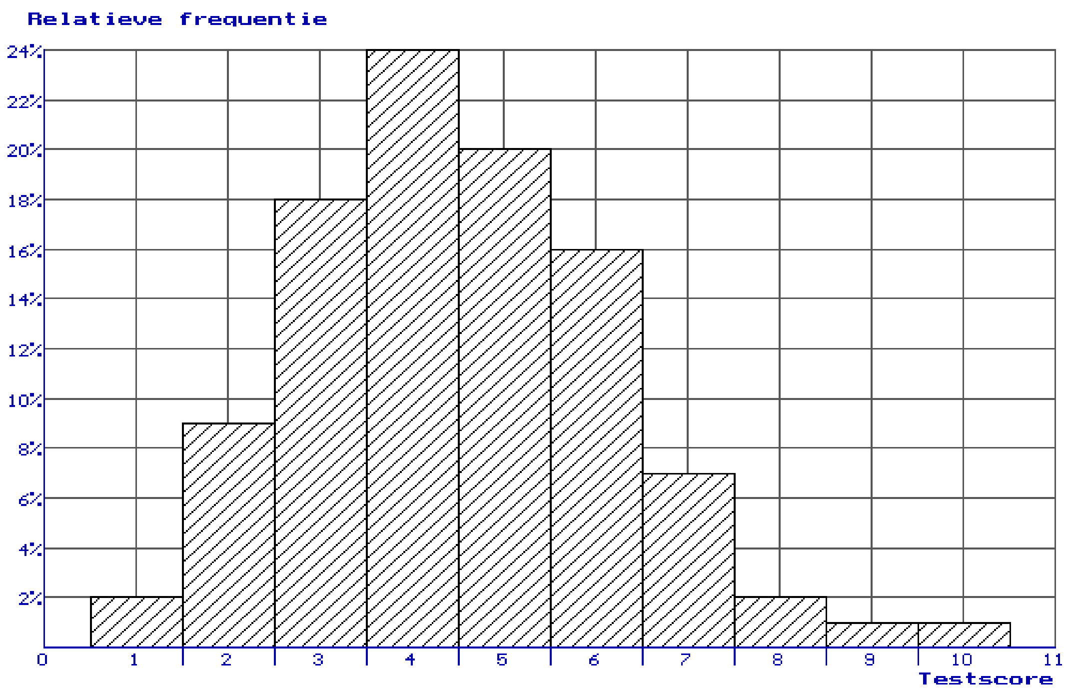
Na ongeveer 1304 uur.

Hoofdstuk 2 – meer over de normale verdeling

Om vast te stellen of een gemeten stochast normaal verdeeld is er speciaal papier ontworpen met een speciale verticale schaalverdeling (de horizontale verdeling is lineair). Langs de verticale as is de verdeling zo, dat de cumulatieve grafiek van de normale verdeling een rechte lijn wordt. Dit wordt bereikt door de procentlijnen ter weerszijden van de 50%-lijn steeds op groter afstand van elkaar te tekenen. Dergelijk papier wordt **normaal waarschijnlijkheidspapier** genoemd.

Normaal waarschijnlijkheidspapier

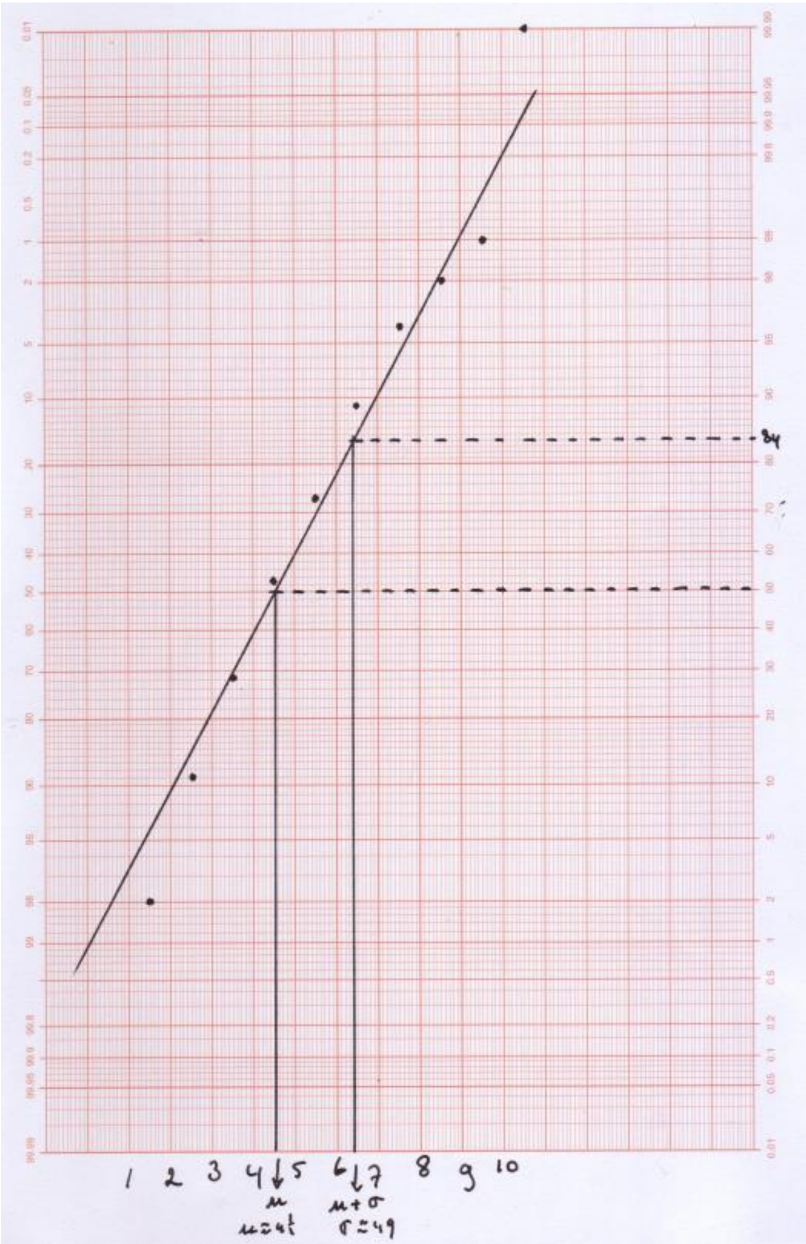
Hieronder zie je een frequentieverdeling met relatieve frequenties van de uitslag op een test van 100 proefpersonen.



Je kunt nu controleren of de verdeling op een normale verdeling lijkt en zo ja dan kun je het gemiddelde en de standaarddeviatie schatten.

- Zet de gegevens in een frequentietabel en bepaal de cumulatieve relatieve frequenties.
  Zie rechts.
- Teken de punten op normaal waarschijnlijkheidspapier.
- Gebruik daarbij de rechter klassengrens!
- Teken door de punten een 'passende' rechte lijn.
- Lees het gemiddelde en de standaarddeviatie af uit de grafiek.
  Zie volgende bladzijde.

Testscore	Rel.Freq.	Cum.Freq.
1	2	2
2	9	11
3	18	29
4	24	53
5	20	73
6	16	89
7	7	96
8	2	98
9	1	99
10	1	100



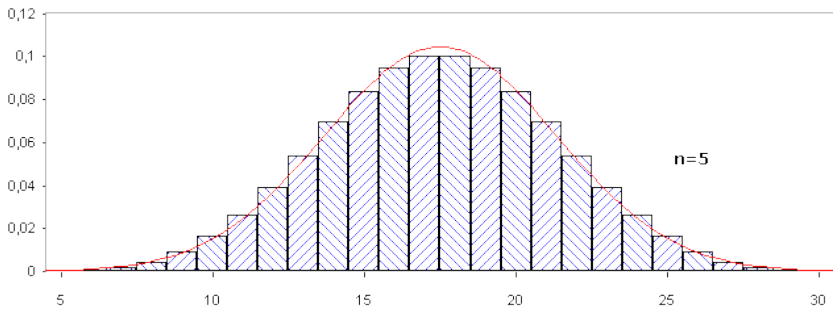
μ ≈ 4,5 en σ ≈ 1,9

Dat is geen gekke benadering. Berekend kom je uit op μ ≈ 4,5 en σ ≈ 1,7

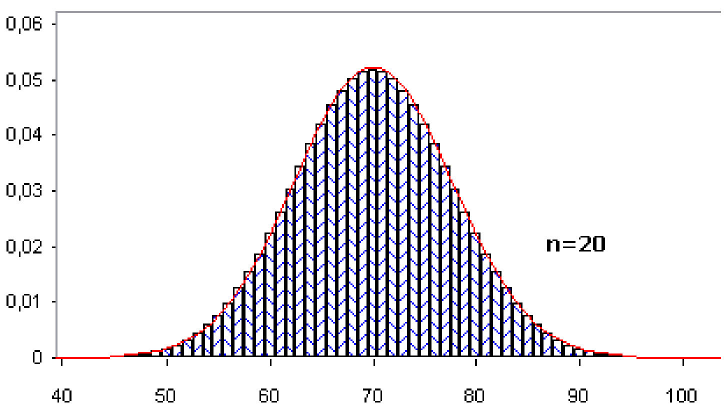
Centrale limietstelling

De som en het gemiddelde van een (voldoende) groot aantal onafhankelijke stochasten is (nagenoeg) normaal verdeeld. Daarbij moeten alle stochasten dezelfde verdeling hebben.

Als je met 5 dobbelstenen gooit en je kijkt naar de som van de ogen dan krijg je deze verdeling:



De rode kromme is de normale verdeling. Zoals je ziet lijkt het al erg veel op elkaar.
Doe je hetzelfde met 20 dobbelstenen, dan lijkt het nog beter...



Benaderen binomiale verdeling

Voor (voldoende) grote waarden van n kan de binomiale stochast Bin(n,k) met verwachting μ = n·p en standaarddeviatie σ = $\sqrt{n \cdot p \cdot (1-p)}$ worden benaderd door de normale stochast met verwachting μ en standaarddeviatie σ.

Voorbeeld

Bereken de kans op minder dan 25 keer munt in 70 worpen met een zuiver geldstuk.

Uitwerking

We hebben te maken met een binomiaal experiment met n = 70 en p = 0,5.

1e manier

X:aantal keren munt
X~Bin(70,0.5)
Met de GR: **Binomcdf**(70,.5,24) $\rightarrow$ 0,0058
P(X $\leq$ 24)=0,0058

2e manier

μ = n·p = 70 · 0,5 = 35

$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{70 \cdot 0{,}5 \cdot (1-0{,}5)} \approx 4{,}18$

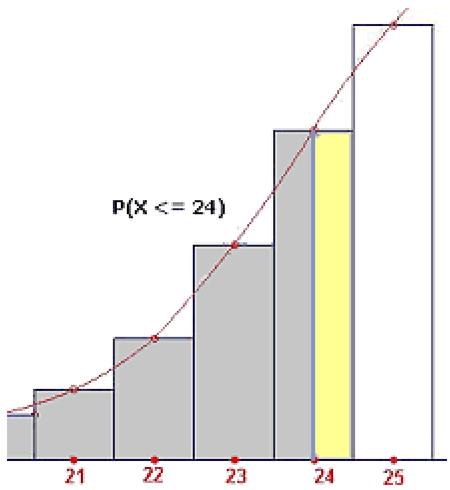
Benaderen met de normale verdeling:

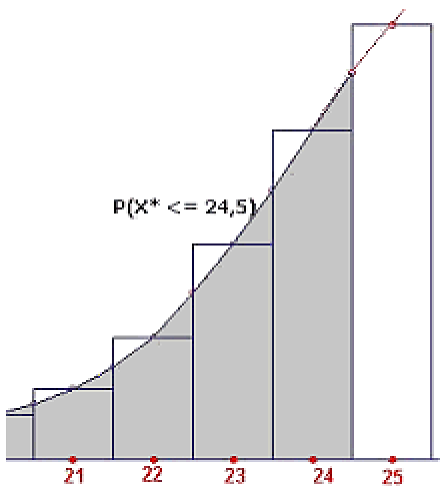
X~Norm(35,4.18)

P(X < 24,5) ≈ 0,006

Continuïteitscorrectie

Je zou je moeten afvragen waarom we nu **24,5** nemen en niet **24**. Aan onderstaande plaatjes kan je dat zien:





Als je P(X<24) dan mis je net een stukje. Dat zou een grotere fout opleveren dan nodig en met P(X<24,5) ondervang je dat.

Noot: bij normaal verdeelde stochasten met uitkomsten die, door afronding, discreet zijn pas je dezelfde correctie toe.

Voorwaarde

Een voorwaarde voor het benaderen van een binomiale door een normale verdeling is:

n·p>5 **en** n·(1-p)>5

"Als ze beide groter dan 5 moeten zijn dan is of n heel groot of p en 1-p verschillen niet te sterk en liggen dus dicht bij 0,5"

Voorbeeld

Van peren is bekend dat 1% van de peren wegens een vervelende schimmel niet te eten zijn. Helaas is dit aan de buitenkant niet waar te nemen. Van een partij van 100 peren wordt onderzocht hoeveel peren aangetast zijn.

a. Bereken met de binomiale verdeling de kans dat er meer dan 2 peren zijn aangetast.
b. Mag je (volgens de regels) deze kans benaderen met de normale verdeling?
c. Benader deze kans met de normale verdeling.
   Maakt het veel verschil?

Uitwerking

a. $X \sim \text{Bin}(100; 0{,}01)$ dus $P(X > 2) = 1 - P(X \leq 2) \approx 1 - 0{,}9206 = 0{,}0794$
b. $0{,}01 \cdot 100 = 1$ en dat is niet groter dan 5, dus nee eigenlijk niet.
c. $X \sim \text{Norm}(1; 0{,}995)$ dus $P(X > 2{,}5) \approx 0{,}0658$. Dat scheelt wel, maar niet heel veel…

Oefeningen

Opgave 1

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

Levensduur in 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 verdeel is en geef m.b.v. normaal waarschijnlijkheidspapier) een schatting van het gemiddelde en de standaarddeviatie.

Opgave 2

In een fabriek worden flessen automatisch gevuld. De inhoud van de gevulde flessen is **normaal verdeeld** rond de inhoud waarop de machine wordt ingesteld. De standaardafwijking bedraagt 5cc. Men wenst dat 90% van de flessen een inhoud hebben van minstens 500cc.

- Op welke maat moet de machine ingesteld worden?

Opgave 3

Van abrikozen is bekend dat 2% van de abrikozen wegens een vervelende schimmel niet te eten zijn. Helaas is dit aan de buitenkant niet waar te nemen. Van een partij van 100 abrikozen wordt onderzocht hoeveel abrikozen aangetast zijn.

a. Bereken met de binomiale verdeling de kans dat er meer dan 3 abrikozen zijn aangetast.
b. Benader deze kans met de normale verdeling.

Uitwerkingen van de oefeningen

Opgave 1

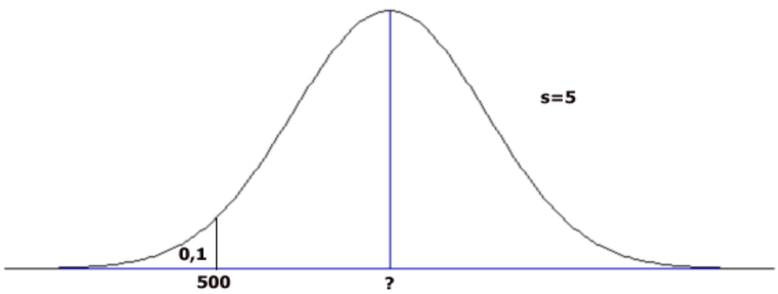
Eerst maar 's de relatieve cumulatieve frequenties uitrekenen.

Levensduur in uren	Klassenmidden	Frequentie	Somfrequentie	Relatief
950-1050	1000	4	4	1%
1050-1150	1100	9	13	4%
1150-1250	1200	19	32	11%
1250-1350	1300	36	68	23%
1350-1450	1400	51	119	40%
1450-1550	1500	58	177	59%
1550-1650	1600	53	230	77%
1650-1750	1700	37	267	89%
1750-1850	1800	20	287	96%
1850-1950	1900	9	296	99%
1950-2050	2000	3	299	100%
2050-2150	2100	1	300	100%

..en dan tekenen op normaal waarschijnlijkheidspapier.

Zie volgende bladzijde.

Opgave 2



$\Phi(z) = 0{,}10$

GR : invNorm(.10,0,1) $\rightarrow -1{,}28$

$$-1{,}28 = \frac{500 - m}{5} \Rightarrow m \approx 506{,}4$$

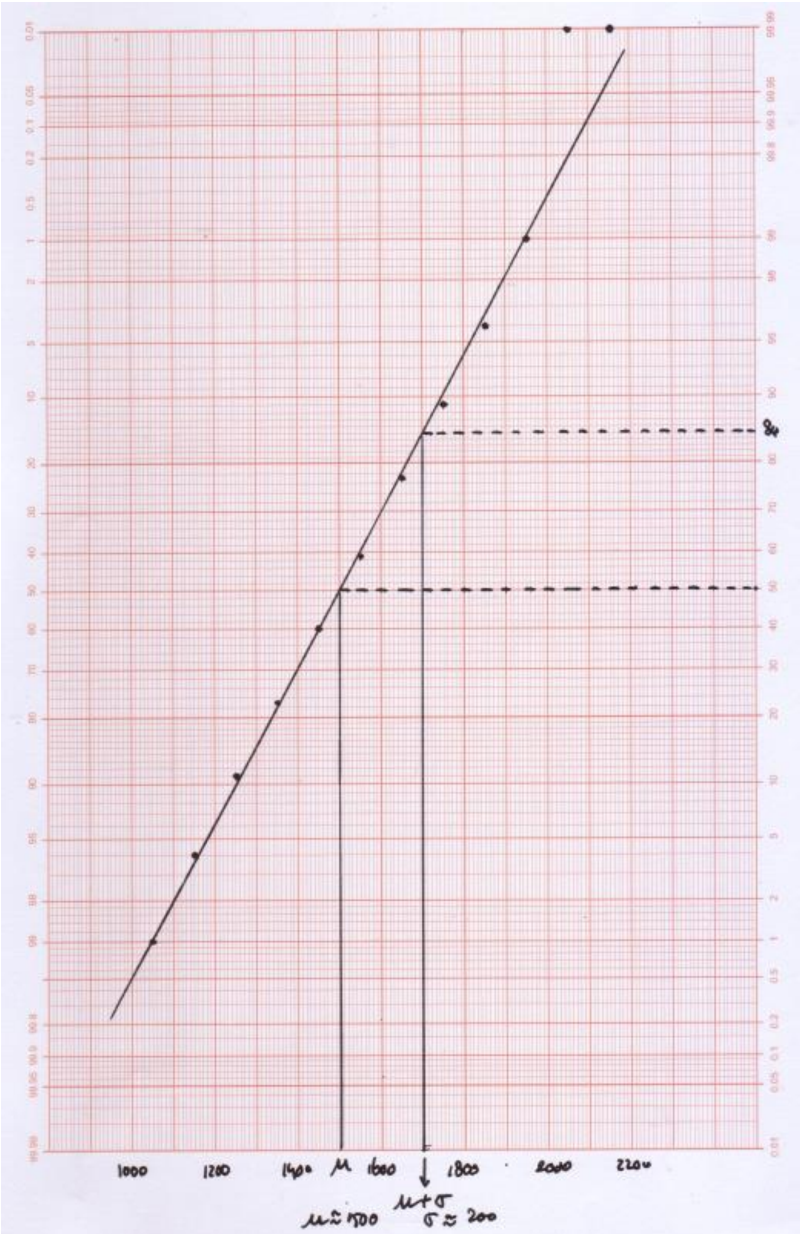
De machine moet worden ingesteld op 506,4

Opgave 3

a. X~Bin(100,.02)
P(X>3)=1-P(X≤3)=0,141
GR: 1-Binomcdf(100,.02,3)

b. X*~Norm(…,…)
$E(X^*) = 2$
$\sigma(X^*) = \sqrt{100 \cdot 0{,}02 \cdot 0{,}98} \approx 1{,}4$
$P(X^* > 3) \approx 0{,}142$
GR: Normalcdf(3.5, ∞,2,1.4)

Uitwerking opgave 1



$\mu \approx 1500$ en $\sigma \approx 200$

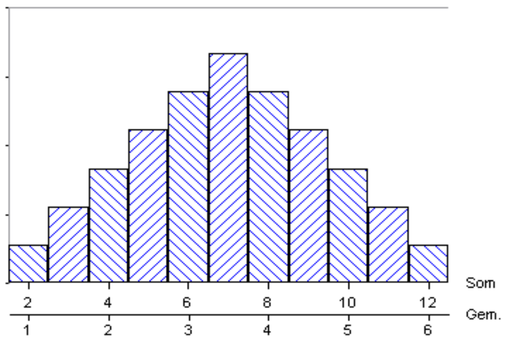
Hoofdstuk 3 – de √n-wet

Je gooit me een dobbelsteen. X noemen we het aantal ogen van de worp. Je kunt de verwachtingswaarde en de standaarddeviatie van X uitrekenen met een tabel:

X	P(X)	X·P(X)	X – E(X)	(X – E(X))²	(X – E(X))²·P(X)
1	$\frac{1}{6}$	$\frac{1}{6}$	$-2\frac{1}{2}$	$6\frac{1}{4}$	$1\frac{1}{24}$
2	$\frac{1}{6}$	$\frac{2}{6}$	$-1\frac{1}{2}$	$2\frac{1}{4}$	$\frac{3}{8}$
3	$\frac{1}{6}$	$\frac{3}{6}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{24}$
4	$\frac{1}{6}$	$\frac{4}{6}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{24}$
5	$\frac{1}{6}$	$\frac{5}{6}$	$1\frac{1}{2}$	$2\frac{1}{4}$	$\frac{3}{8}$
6	$\frac{1}{6}$	$\frac{6}{6}$	$2\frac{1}{2}$	$6\frac{1}{4}$	$1\frac{1}{24}$
	E(X) =	$3\frac{1}{2}$		Var(X) =	2,92
	μ(X) =	**3,5**		**σ(X) =**	**1,71**

We zien dat $\mu = 3{,}5$ en $\sigma \approx 1{,}71$

Als je nu met twee keer met die dobbelsteen gooit en je noemt het aantal ogen van de eerste worp X en het aantal ogen van de tweede worp Y. Wat is dan de verwachtingswaarde van X+Y en wat is dan de standaarddeviatie van X+Y?



Je zou natuurlijk een nieuwe tabel kunnen maken, maar **handig** is dat niet.

In het algemeen geldt voor onafhankelijke stochasten X en Y:

E(X + Y) = E(X) + E(Y)
Var(X + Y) = Var(X) +Var(Y)

Gevolg:

- E(X+Y)=7
- Var(X+Y)=5,84 $\Rightarrow$ σ(X+Y)=2,42

De som van n onafhankelijke stochasten

$X_1, X_2, X_3, \ldots X_n$ zijn n onafhankelijke stochasten,
elk met dezelfde kansverdeling met $E(X_i)$ en $\sigma(X_i)$
Wat is dan $E(X_1 + X_2 + X_3 + \ldots + X_n)$?
$E(X_1 + X_2 + X_3 + \ldots + X_n) = n \cdot E(X_i)$
En wat is dan $\sigma(X_1 + X_2 + X_3 + \ldots + X_n)$?
$\sigma(X_1 + X_2 + X_3 + \ldots + X_n) = \sqrt{n} \cdot \sigma(X_i)$

Voorbeeld
Je gooit 3 keer met een dobbelsteen en telt de worpen bij elkaar op.

- Bereken de verwachtingswaarde en de standaarddeviatie.

Uitwerking
$S = X_1 + X_2 + X_3$ met $E(X_i) = 3{,}5$ en $\sigma(X_i) = 1{,}71$

$E(S) = 3 \cdot E(X_i) = 3 \cdot 3{,}5 = 10{,}5$

$\sigma(S) = \sqrt{3} \cdot \sigma(X_i) = \sqrt{3} \cdot 1{,}71 \approx 2{,}96$

Het gemiddelde van n onafhankelijk stochasten
$X_1, X_2, X_3, \ldots X_n$ zijn n onafhankelijke stochasten,
elk met dezelfde kansverdeling met $E(X_i)$ en $\sigma(X_i)$

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}$$

Wat is dan $E(\bar{X})$?

$E(\bar{X}) = E(X_i)$

En wat is dan $\sigma(\bar{X})$?

$$\sigma(\bar{X}) = \frac{\sigma(X_i)}{\sqrt{n}}$$

Voorbeeld
Je gooit 3 keer met een dobbelsteen en berekent het gemiddelde dan de worpen.

- Bereken de verwachtingswaarde en de standaarddeviatie.

Uitwerking
$\bar{X} = \dfrac{X_1 + X_2 + X_3}{3}$ met $E(X_i) = 3{,}5$ en $\sigma(X_i) = 1{,}71$

$E(\bar{X}) = E(X_i) = 3{,}5$ en $\sigma(\bar{X}) = \dfrac{\sigma(X_i)}{\sqrt{3}} = \dfrac{1{,}71}{\sqrt{3}} \approx 0{,}99$

Oefeningen

Opgave 1
In een brood gaat gemiddeld 2 gram broom. Er wordt een steekproef van 100 broden gedaan en vervolgens wordt het steekproefgemiddelde hiervan vastgesteld. De standaarddeviatie van de populatie is 0,1 gram.

- Bereken de standaarddeviatie van de steekproef.

Opgave 2
In de bussen van Noordwest hangt een bordje '42 zitplaatsen en 39 staanplaatsen, max. laadvermogen 6.129 kg'. Het gewicht van passagiers is normaal verdeeld met een gemiddelde van 71 kg en een standaardafwijking van 21 kg.

- Bereken de kans dat twee passagiers samen meer wegen dan 150 kg.

Opgave 3

De hoeveelheid drank uit een drankautomaat is normaal verdeeld met een verwachtingswaarde van 150 ml en standaardafwijking van 8 ml. Dit wordt gevuld in een beker van 165 ml met een standaardafwijking van 6 ml.

a. Bereken de kans dat in de beker nog minstens 5 ml ruimte over blijft.

De fabrikant wil dat de instelling het vulgemiddelde verlaagd wordt, zodat de kans op een overlopende beker hoogstens 0,1% is.

b. Hoe hoog mag het vulgemiddelde (in tienden van millimeters nauwkeurig) hoogstens zijn, zodat aan de eis van de fabrikant is voldaan?

Uitwerkingen van de oefeningen

Opgave 1

$$\sigma(\overline{X}) = \frac{0,1}{\sqrt{100}} = 0,01$$

Opgave 2

x_1 en x_2 zijn twee onafhankelijke stochasten, elk met dezelfde kansverdeling met $\mu = 71$ en $\sigma = 21$.
$E(x_1+x_2) = 2 \cdot 71 = 142$ en $\sigma(x_1+x_2) = \sqrt{2} \cdot 21 \approx 29,7$

Gegeven:

- X: gewicht van 2 passagiers
- X~Norm(142,29.7)
- P(X > 150) = 0,394

Opgave 3

We definiëren een nieuwe stochast V = inhoud beker – hoeveelheid drank.
$\mu(V) = 165 - 150 = 15$ ml en $\sigma(V) = \sqrt{6^2 + 8^2} = 10$

a. Gevraagd: $P(V > 5) \rightarrow 0,841$
b. Kies $\mu(V)=p$ dan $\sigma=10$ en $P(V < 0) < 0,001$
   Met de GR: Y1=normalcdf(-9E99,0,X,10) $\rightarrow$ table $\rightarrow$X minimaal 31
   Dus neem als vulgewicht 181 ml.

Hoofdstuk 4 – experiment

We doen het volgende experiment. We hebben 22.000 gezonde proefpersonen. We verdelen de proefpersonen in twee groepen van 11.000. Eén groep krijgt vijf jaar lang een pil met werkzame stof (aspirine) en de andere groep krijgt een pil zonder werkzame stof (controlegroep).

Controlegroep:	189 hartaanvallen
Testgroep:	165 hartaanvallen

- Heeft aspirine een gunstig effect op de kans op een hartaanval?

De vraag is nu: is 165 wel 'echt' minder. Kan het ook toeval zijn? Met andere woorden neem aan dat de aspirine geen invloed heeft op de kans op een hartaanval, wat is dan de kans dat je dan toch 165 vindt?

In het algemeen hanteren we daarbij een grenswaarde van 5% (of soms ook 1%). Als de kans dat je 165 vindt onder de veronderstelling dat het middel niet werkt kleiner is dan 5% dan besluit je dat het geen toeval kan zijn en dat de aspirine wel degelijk een positief effect heeft op de kans op een hartaanval.

Uitwerking

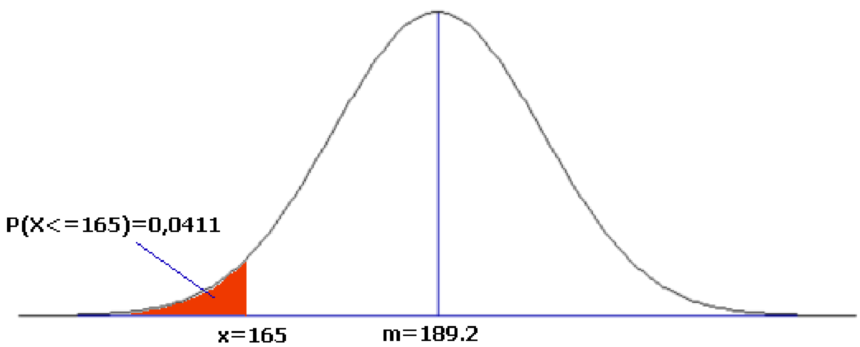
X: aantal mensen met een hartaanval bij 11.000 personen in de testgroep.
X ~ binomiaal verdeeld met n=11.000 en p=0,0172.

We gaan er vanuit dat de kans in de controlegroep een goede schatting is van de kans op een hartaanval.

We kunnen deze kans benaderen met de normale verdeling met continuïteitscorrectie.
Dit geeft $\mu = 189{,}2$[1] en $\sigma \approx 13{,}6$.

$P(X < 165) \approx 0{,}041$

De kans dat je 165 vindt (of lager) is kleiner dan 5%. We concluderen dat aspirine een gunstig effect op de kans op een hartaanval heeft.



[1] Dat gemiddelde is eigenlijk 189. Door afronding van σ=0,0172 geeft dit 189,2.

Opgave

Na jarenlange ervaring is bekend dat ongeveer 40% van de mensen bereid is vragen te beantwoorden bij een mondelinge enquête in de binnenstad.

Bij een enquête van 100 personen in Den Haag blijkt dat slechts 30% van de aangesproken mensen bereid is een aantal vragen te beantwoorden.

a. Als toch 40% van de mensen bereid is vragen te beantwoorden, bereken dan op 3 decimalen de kans dat je in een steekproef van 100 toch 30% of minder vindt.
b. Is er reden om te twijfelen aan de 40%-bereidheid in Den Haag?

Uitwerking

X: aantal mensen dat bereid is vragen te beantwoorden
X~ binomiaal verdeeld met p=0,4 en n=100

a. Gevraagd: $P(X \leq 30) = 0{,}0248$
b. De kans is kleiner van 0,05. Als je 0,0248 vindt zou dat toch wel heel toevallig zijn. Er is statistisch gezien reden om aan te nemen dat die 40%-bereidheid niet klopt.

Einde

Zie ook de lesbrief **Hypothese toetsen**.

februari 2010

Willem

Tabel normale verdeling

Standaardnormale verdeling P($\underline{z} > z$)

z	0	1	2	3	4	5	6	7	8	9
0,0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0,1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0,2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0,3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0,4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0,5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0,6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0,7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0,8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1921	.1894	.1867
0,9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1,0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1,1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1,2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1,3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1,4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1,5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1,6	.0548	.0537	.0527	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1,7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1,8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1,9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2,0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2,1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2,2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2,3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2,4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2,5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2,6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2,7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2,8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2,9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3,0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Inhoudsopgave compleet