Chemometrie, ISBN: 9789491764332
Vierde herziene druk
www.syntaxmedia.nl

Uitwerkingen

In de uitgewerkte voorbeelden worden vanwege de leesbaarheid afgeronde tussenresultaten gepresenteerd. De eindresultaten zijn echter altijd berekend zonder tussentijds afronden.

Hoofdstuk 2

Opmerking.
Als gewerkt wordt met tabellen 1, 2 en 3 uit bijlage 1 van het boek dan worden berekeningen uitgevoerd met afgeronde *z*- of *t*-waarden. Daardoor kunnen kleine verschillen optreden met de exacte waarden die berekend kunnen worden met Excelfuncties zoals blijkt uit de uitwerkingen van de opgaven in dit hoofdstuk.

Zie voor de gebruikte Excelfuncties de helpfunctie in Excel of het boek J.W.A. Klaessens, *Statistiek in het laboratorium met Excel*, Syntax Media, 3e druk, 2016.

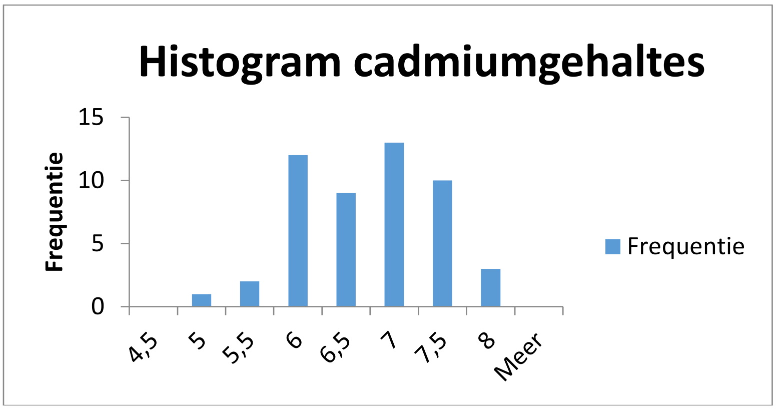
Antwoord 2.1

1	2	3	4	5	6	7	8	9	10
5,16	5,80	7,35	6,07	4,93	6,57	7,26	6,09	5,57	6,80
6,48	6,68	6,78	5,87	7,91	5,78	6,73	7,34	6,29	5,62
6,26	6,23	6,94	6,50	7,49	6,59	6,55	5,86	6,58	5,46
6,14	5,92	7,29	7,85	6,57	6,63	7,07	6,55	6,54	7,04
5,97	7,05	7,07	5,54	5,89	5,84	7,08	5,99	6,21	7,74

Voor de cadmiumgehaltes in voorgaande tabel kan in Excel een frequentietabel en een histogram worden gemaakt in de Analyse Toolpak met de optie ‘Histogram’, waarbij de optie ‘Grafiek maken’ moet worden aangevinkt. Hiervoor moeten eerst alle data in één rij of kolom worden gezet. Bovendien moeten de klassen worden gedefinieerd, zie volgende tabel. Dit wordt in Excel het ‘Verzamelbereik’ genoemd.

Frequentietabel

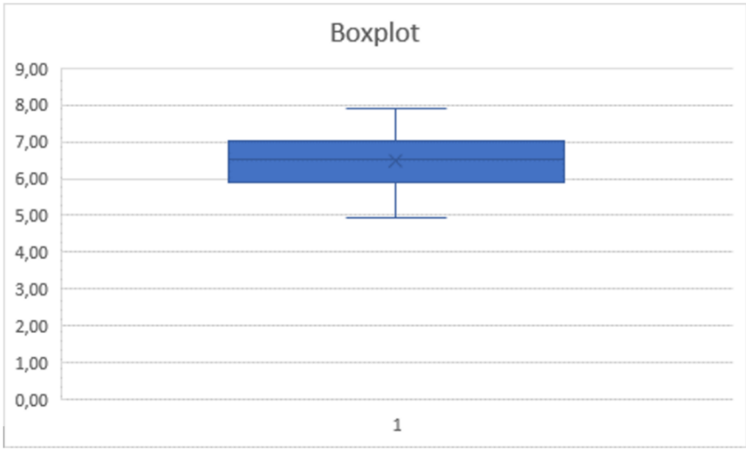
Verzamelbereik	Frequentie
4,5	0
5,0	1
5,5	2
6,0	12
6,5	9
7,0	13
7,5	10
8,0	3
Meer	0



De statistische gegevens kunnen in de Analyse Toolpak van Excel worden samengevat met de optie ‘Beschrijvende statistiek’, zie volgende tabel.

Gemiddelde	6,4704
Standaardfout	0,09717
Mediaan	6,545
Modus	6,57
Standaarddeviatie	0,68712
Steekproefvariantie	0,47213
Kurtosis	-0,38873
Scheefheid	0,07390
Bereik	2,98
Minimum	4,93
Maximum	7,91
Som	323,52
Aantal	50

Vanaf Excel 2016 is een sjabloon beschikbaar voor het maken van een boxplot.



Het gemiddelde van de cadmiumgehalten is $\bar{x}$ = 6,4704 µg/L, de standaarddeviatie s = 0,68712 en het aantal metingen is n = 50. Het aantal vrijheidsgraden is $f = (n - 1) = 50 - 1 = 49$. Voor een 95%-betrouwbaarheidsinterval geldt, met een tweezijdige $t_{(0,05;49)}$ = 2,010, met vergelijking (2.18):

$$\mathrm{BI}(\bar{x}) = \bar{x} \pm \frac{t_{(1-\alpha;n-1)}s}{\sqrt{n}}$$

$$\mathrm{BI}(\bar{x}) = 6{,}4704 \pm \frac{2{,}010 \cdot 0{,}68712}{\sqrt{50}} = 6{,}470 \pm 0{,}195$$

<u>Berekening van de kans dat het cadmiumgehalte groter is dan 8,0 µg/L.</u>

Eerst moet de grenswaarde van 8,0 µg/L met vergelijking (2.9) worden geschaald naar een *z*-waarde:

$$z = \frac{x - \mu}{\sigma} = \frac{8{,}0 - 6{,}4704}{0{,}68712} = 2{,}2261$$

Voor een afgeronde *z*-waarde van $z = 2{,}23$ kan de kans worden berekend met behulp van tabel 2 van bijlage 1. De kans $P(z < 2{,}23) = 0{,}9871$, zie tabel 2 van bijlage 1.
Daaruit volgt dat de kans $P(z > 2{,}23) = 1 – 0{,}9871 = 0{,}0129$.

De kans $P(x > 8{,}0)$ kan exact worden berekend in één stap, zonder eerst een *z*-waarde te berekenen, met een Excelfunctie volgens:
1 - NORM.VERD.N(*grenswaarde*;*gemiddelde*;*standaarddeviatie*;1)
1 - NORM.VERD.N(8,0;6,4704;0,68712;1) = 0,0130

De hierbij ingevulde waarden voor het gemiddelde en de standaarddeviatie zijn weer afgeronde tussenresultaten. Door te werken met celverwijzingen kan echter gerekend worden met de exacte waarden hiervan.

Berekening van de kans dat het cadmiumgehalte ligt tussen 5,0 en 8,0 µg/L.

Eerst moeten de grenswaarden met vergelijking (2.9) worden geschaald naar *z*-waarden:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{5{,}0 - 6{,}4704}{0{,}68712} = -2{,}1400$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{8{,}0 - 6{,}4704}{0{,}68712} = 2{,}2261$$

Met behulp van tabel 1 en 2 van bijlage 1 wordt met afgeronde *z*-waarden gevonden:
$P(5{,}0 < x < 8{,}0) = P(x_2 < 8{,}0) – P(x_1 < 5{,}0) = P(z_2 < +2{,}23) – P(z_1 < -2{,}14) = 0{,}9871 – 0{,}0162 = 0{,}9709$.

De kans $P(5{,}0 < x < 8{,}0) = P(x_2 < 8{,}0) – P(x_1 < 5{,}0)$ kan in één stap, zonder eerst een *z*-waarde te berekenen, exact worden berekend met een Excelfunctie volgens:
NORM.VERD.N(8,0;6,4704;0,68712;1) - NORM.VERD.N(5,0;6,4704;0,68712;1) = 0,9870 – 0,0162 = 0,9708.

Antwoord 2.2

1	36,5
2	45,2
3	40,5
4	42,0
5	37,7
6	39,4
7	41,6
8	38,8
9	39,0
10	43,3

gem	40,40
sdev	2,6558
sdev_gem	0,8398
n	10
1-alpha	0,01
t	3,2498
halve BI	2,7293
ondergrens	37,67
bovengrens	43,13

Het 99% betrouwbaarheidsinterval van het gemiddelde van een steekproef kan worden berekend met vergelijking (2.18), met $t_{(0,01;9)} = 3,2498$:

$$\mathrm{BI}(\bar{x}) = \bar{x} \pm \frac{t_{(1-\alpha;n-1)}s}{\sqrt{n}} = 40,40 \pm \frac{3,2498 \cdot 2,6558}{\sqrt{10}} = 40,40 \pm 2,73 \text{ mg/L.}$$

$\mathrm{BI}(\bar{x})$ = [37,7; 43,1] mg/L.

Antwoord 2.3

Volumepipetten van 10 ml van een zekere kwaliteitsklasse.
De tolerantie is 0,02 mL.
De standaarddeviatie van de populatie pipetten: $\sigma = 0,02/\sqrt{3} = 0,01155$.
De normale verdeling is $N(\mu = 10,00; \sigma = 0,02/\sqrt{3})$.

gemiddelde	10,00
sdev	0,01155
grenswaarden	
x1	9,99
x2	10,01
z1	-0,8660
z2	0,8660
Kansen	
P(X<x1)	0,1932
P(X<x2)	0,8068
p(x1<X<x2)	0,6135
95%-interval: bovenste grenswaarde	10,0226
95%-interval: onderste grenswaarde	9,9774
99%-interval: bovenste grenswaarde	10,0297
99%-interval: onderste grenswaarde	9,9703

Berekening van het percentage van de pipetten die meer dan 10,01 mL leveren.

De grenswaarden moeten eerst met vergelijking (2.9) worden geschaald naar *z*-waarden. Daarna kunnen voor de op twee decimalen afgeronde *z*-waarden de bijbehorende kansen worden opgezocht in tabel 1 en 2 van bijlage 1.

Met *z*-tabel in bijlage 1:

$$P(x > 10{,}01) = P\left(z > \frac{10{,}01 - 10{,}00}{0{,}02/\sqrt{3}}\right) = P(z > 0{,}8660)$$

$$P(x > 10{,}01) \approx P(z > 0{,}87) = 1 - 0{,}8078 = 0{,}1922$$

19,22 % van de 10,00 mL-pipetten levert meer dan 10,01 mL.

$P(x > 10{,}01)$ kan met Excel in één stap, zonder eerst een *z*-waarde te berekenen, exact worden berekend:
1 - NORM.VERD.N(10,01;10,00;0,01155;1) = 0,1932
Hieruit volgt dat 19,32 % van de 10,00 mL-pipetten meer dan 10,01 mL levert.

Berekening van het percentage van de pipetten die een volume tussen 9,99 en 10,01 mL leveren.

Eerst moeten de grenswaarden met vergelijking (2.9) worden geschaald naar *z*-waarden:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{9{,}99 - 10{,}00}{0{,}02/\sqrt{3}} = -0{,}8660$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{10{,}01 - 10{,}00}{0{,}02/\sqrt{3}} = 0{,}8660$$

Met behulp van tabel 1 en 2 van bijlage 1 wordt met afgeronde *z*-waarden gevonden:
$P(9{,}99 < x < 10{,}01) = P(x_2 < 10{,}01) – P(x_1 < 9{,}99) = P(z_2 < +0{,}87) – P(z_1 < -0{,}87) = 0{,}8078 – 0{,}1922 = 0{,}6156$.

61,56 % van de 10,00 mL-pipetten levert een volume tussen 9,99 en 10,01 mL.

De kans $P(9{,}99 < x < 10{,}01) = P(x_2 < 10{,}01) – P(x_1 < 9{,}99)$ kan in één stap, zonder eerst een *z*-waarde te berekenen, exact worden berekend met een Excelfunctie volgens:
NORM.VERD.N(10,01;10,00;0,01155;1) - NORM.VERD.N(9,99;10,00;0,01155;1) = 0,8068 – 0,1932 = 0,6135.

Berekening van de grenzen waartussen 95% van de geleverde volumes zich bevinden.

2,5% van de pipetten levert een volume af dat boven de bovenste grenswaarde ligt en 2,5% van de pipetten levert een volume af dat onder de onderste grenswaarde ligt.

Berekening van de bovengrens x_{BG}:

$$P\left(z > \frac{x_{BG} - 10{,}00}{0{,}02/\sqrt{3}}\right) = 0{,}025$$

De overeenkomstige *z*-waarde is $z = 1{,}96$ (zie tabel 2 van bijlage 1).
Dan geldt dus voor de bovengrens x_{BG}:

$$1{,}96 = \frac{x_{BG} - 10{,}00}{0{,}02/\sqrt{3}} \rightarrow x_{BG} - 10{,}00 = 1{,}96 \cdot 0{,}02\sqrt{3}$$

$$x_{BG} = 10{,}00 + 1{,}96 \cdot 0{,}02\sqrt{3} = 10{,}00 + 0{,}023 = 10{,}023$$

Vanwege de symmetrie van de verdeling ligt de ondergrens op
x_{OG} =10,00 - 0,023 = 9,977.
95% van de pipetten levert een volume af dat ligt tussen 9,977 en 10,023 mL.

De grenswaarde die overeenkomt met een zekere kans voor een normale verdeling met bekend gemiddelde en standaarddeviatie kan worden berekend met de volgende Excelfunctie:
=NORM.INV.N(*kans*;*gemiddelde*;*standaarddeviatie*)

De bovengrens van het 95% interval is:
=NORM.INV.N(0,975;10,00;0,01155) = 10,0226
De ondergrens van het 95% interval is:
=NORM.INV.N(0,025;10,00;0,01155) = 0,9774
Deze grenswaarden zijn gelijk aan de hiervoor berekende waarden.

Berekening van de grenzen waartussen 99% van de geleverde volumes zich bevinden.

0,5% van de pipetten levert een volume af dat boven de bovenste grenswaarde ligt en 0,5% van de pipetten levert een volume af dat onder de onderste grenswaarde ligt.

Berekening van de bovengrens x_{BG}:

$$P\left(z > \frac{x_{BG} - 10,00}{0,02/\sqrt{3}}\right) = 0,005$$

De overeenkomstige z-waarde is z = 2,58 (zie tabel 2 van bijlage 1).
Dan geldt dus voor de bovengrens x_{BG}:
$$2,58 = \frac{x_{BG} - 10,00}{0,02/\sqrt{3}} \rightarrow x_{BG} - 10,00 = 2,58 \cdot 0,02\sqrt{3}$$

$$x_{BG} = 10,00 + 2,58 \cdot 0,02\sqrt{3} = 10,00 + 0,030 = 10,030$$

Vanwege de symmetrie van de verdeling ligt de ondergrens op
x_{OG} =10,00 - 0,030 = 9,970.
99% van de pipetten levert een volume af dat ligt tussen 9,970 en 10,030 mL.

De bovengrens van het 99% interval kan met een Excelfunctie exact worden berekend:
=NORM.INV.N(0,995;10,00;0,01155) = 10,030
De ondergrens van het 99% interval is:
=NORM.INV.N(0,005;10,00;0,01155) = 0,970
Deze grenswaarden zijn gelijk aan de hiervoor berekende waarden.

Antwoord 2.4

De grenswaarde in 1998 voor het 90 percentielniveau van zink is 24 μg Zn^{2+}/L.
De standaarddeviatie is σ = 15 μg Zn^{2+}/L.

$$P\left(z < \frac{x - \mu}{\sigma}\right) = \left(z < \frac{24 - \mu}{15}\right) = 0,90$$

De overeenkomstige z-waarde is z = 1,28 (zie tabel 2 van bijlage 1).

Dan geldt:
$1{,}28 = \frac{24-\mu}{15} \rightarrow 24 - \mu = 1{,}28 \cdot 15$

$\mu = 24 - 1{,}28 \cdot 15 = 24 - 19{,}20 = 4{,}80$

De kans dat de streefwaarde van 12 μg Zn^{2+}/L wordt overschreden is:

$P(x > 12) = P\left(z > \frac{x-\mu}{\sigma}\right) = P\left(z > \frac{12-4{,}80}{15}\right) = P(z > 0{,}48) = 1 - 0{,}6844 = 0{,}3156.$

Er is dus 31,56% kans dat de streefwaarde van 12 μg Zn^{2+}/L wordt overschreden.