Samenvatting statistiek havo4 boek 1 H4
Centrummaten:
Modus (modaal) = wat het vaakst voorkomt, zowel kwalitatief als kwantitatief
Mediaan = het middelste getal, in een rij getallen die op volgorde staat
Gemiddelde= $\overline{X}$ = m = $\dfrac{allegetallenoptellen}{aanta\lg etallen}$
Spreidingsmaten:
Spreidingsbreedte = Xmax – Xmin
Kwartielafstand = Q3 – Q1 (boxplot, elk gedeelte = 25%)
Standaardafwijking = σ = s (deze vindt je op de GR, of in de tekst)
GR: voer in x = L1 = wat er opgemeten wordt, ook klassenmiddens
Y = L2 = frequentie (= aantal keer dat iets voorkomt)
Stat calc 1: 1-VarStats L1, L2 Enter, en je krijgt bijna alle bovenstaande gegevens
Daarbij kan je aan n = aantal zien, of je goed hebt ingevoerd.
Boxplot 2nd stat plot enter On enter type 5 enter xlist = L1 en freq = L2,
denk aan x window instellen, graph

Teken de relatieve cumulatieve frequentie polygoon
Relatief = % cumulatief = bij elkaar optellen polygoon= lijn met hoekjes
X = wat opgemeten wordt, Y = frequentie in %, opgeteld
Let op: als je klasseverdelingen hebt, neem je bij de frequentiepolygoon het midden van de klasse, maar bij de cumlatieve frequentie polygoon de rechtergrens van de klasse (dus het hoogste getal).

Normale verdeling havo 4 boek 2 H8
(GR 2nd Vars normalcdf en invnorm)
a. Bij de **standaardnormale verdeling** heb je te maken met
De oppervlakte onder het klokje, en de z-as.
Bij het gemiddelde geldt: z = 0
1 σ naar rechts, vanuit het midden : z = 1 , opp = 34%
1 σ naar links, vanuit het midden : z = -1 , opp = 34%
2 σ naar rechts, vanuit het midden : z = 2 , opp = 13,5
2 σ naar links, vanuit het midden : z = -2 , opp = 13,5%
Als z geen 1, -1 , 2, of -2 is gebruik je de GR
1. Oppervlakte of % uitrekenen
Opp = $\Phi(z)$ = normalcdf (linkergrens , rechtergrens, μ , σ)
Als z de rechtergrens is kies je linkergrens -10^{99}
Als z de linkergrens is kies rechtergrens $+10^{99}$
Opp $\rightarrow$ % keer 100 ; % $\rightarrow$ opp gedeeld door 100
2. z uitrekenen
z = invnorm(opp) , heb je nodig als je μ of σ moet uitrekenen
b. normale verdeling (vooral bij tekstsommen)
Stappenplan : Schrijf al je gegevens op, (er zijn er 5):
μ = σ = z = **linkerOpp** (of %) =
score X = (feitelijk de linker of rechtergrens)

Zet een vraag teken bij wat je uit moet rekenen
GR: 1. als je μ, σ en X weet, en Opp (of % of kans) uit moet rekenen :
Normalcdf(linkergrens, rechtergrens, μ, σ) = opp
2. als je μ, σ, en Opp (of % of kans) weet, en score X uit moet rekenen :
invnorm(opp, μ, σ) = score X
Zonder GR: als je μ of σ uit moet rekenen: Meestal eerst: z = invnorm(linkeropp) ,
en daarna: $z = \frac{X-\mu}{\sigma}$, en vergelijking oplossen

Boek 1 hoofdstuk 4 Havo 4 Statistiek.

4.1frequentietabellen

1a. Kwantitatief = wat je kan tellen, bv gewicht, afstand, zakgeld
Kwalitatief = niet telbaar, kenmerkend, bv bloedgroep, soort vervoer.
b. kwantitatief: lengte,
kwalitatief: welk muziekinstrument, haarkleur
c. frequentie jongens = aantal jongens = 12

2a.

Bloedgroep	O	A	B	AB
frequentie	12	10	2	4

3a.

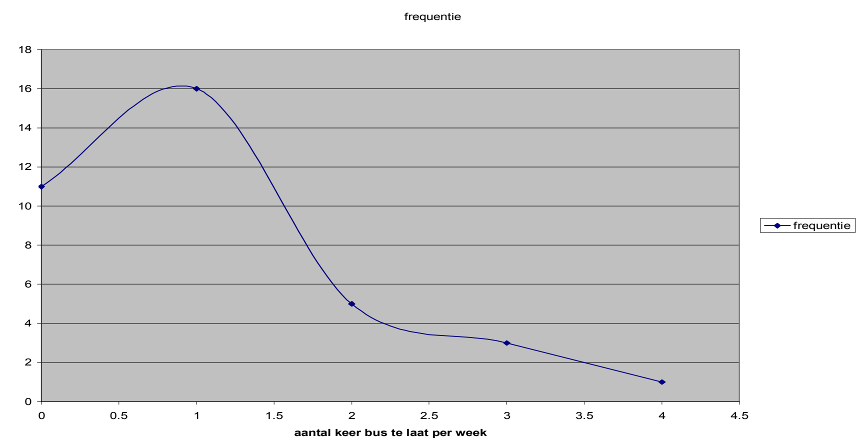
Aantal gezinsleden	1	2	3	4	5	6	7
frequentie	0	3	7	9	5	3	1

d. Minder dan 4 dus 2 of 3, dat zijn er 10 van de 28 Het percentage= $\frac{10}{28} * 100 = 35.7\%$

Minstens 4: Het percentage= $\frac{18}{28} * 100 = 64.3\%$

4b. Er komt 18 keer voor: 2 slakeen per m ^{2}, dat zijn dus 18 m ^{2}
Zo kom je op 41m ^{2}.
c. $2*18 + 3*16 + 4*5 + 5*7 + 6*3+ 7*2 = 171$

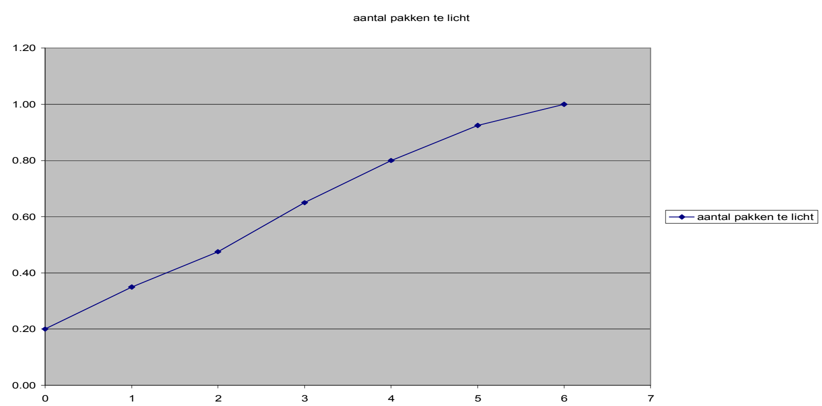
5a. Het komt 11 weken voor dat de bus 0 keer per week te laat is, dat zijn 11 weken, net zo: 11 + 16 + 5 + 3 + 1= 36 weken
c. minder dan twee keer, dat is 0 of 1 keer $\frac{27}{36} * 100 = 75\ \%$
d. $1*16 + 2*5 + 3*3 + 4*1 = 39$ keer te laat
totaal 10 keer per week met de bus, 36 weken, dat is 360 keer $\frac{39}{360} * 100 = 10,8\ \%$
b. zie vlg blz.



6a. 8 + 6 + 5 + 7 + 6 + 5 + 3 = 40 controlebeurten, dus 40 dagen
b. Er zijn 40 $*$ 50 pakken onderzocht, dat is 2000 pakken
Te weinig gewicht hadden 1 $*$ 6 + 2 $*$ 5 + 3 $*$ 7 + 4 $*$ 6 + 5 $*$ 5+ 6 $*$ 3 = 104 pakken

$$\frac{104}{2000} * 100 = 5.2\%$$

c. Teken de relatieve cumulatieve frequentie polygoon
Relatief = %   cumulatief = bij elkaar optellen   polygoon= lijn met hoekjes
X = wat opgemeten wordt, Y = frequentie in %, opgeteld

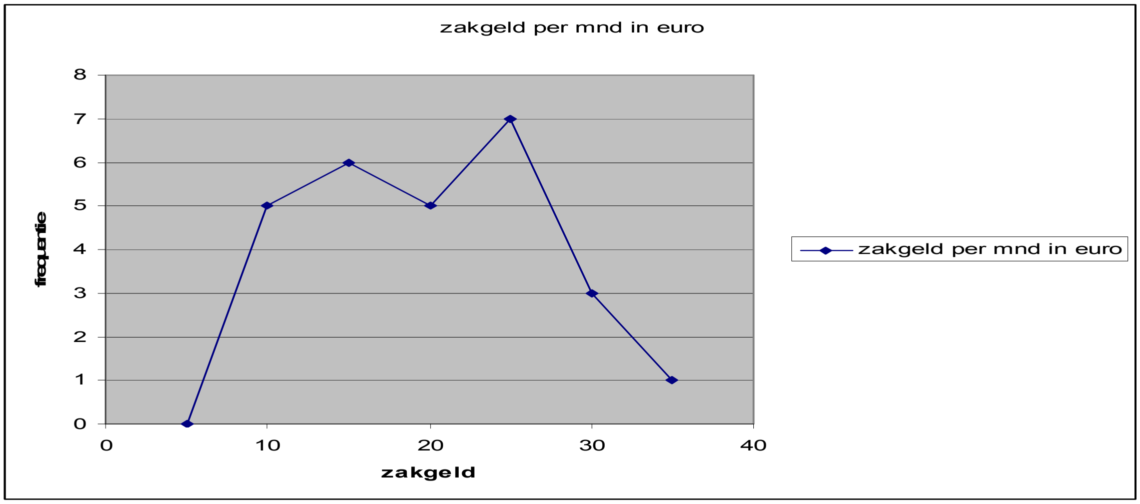


4.2 Frequentieverdelingen.

7. Er zijn te veel verschillende getallen, dat geeft geen inzicht

8a en c. zie ook blz. 121

Zakgeld per mnd. in euro's	5 -< 10	10 - < 15	15 -< 20	20-< 25	25-<30	30-<35
frequentie	5	6	5	7	3	1



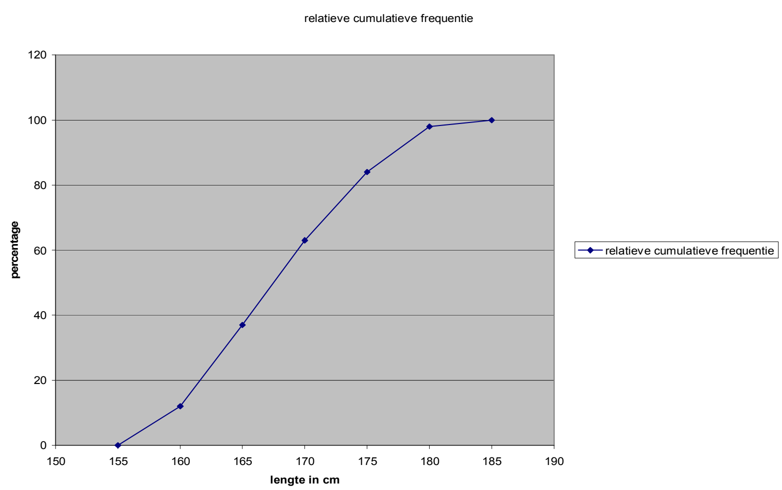
9a. twee keer, op de tweede regel **b.** 6 euro **c.** 20 euro
d. De klassen zijn tientallen, dus 0-<10, 10-<20 enz.
10a.16 echtparen **b.** 1 man was 33, en niet een vrouw was 33
c. 5 mannen en 6 vrouwen, totaal 11 personen
d. 6 mannen en 8 vrouwen = 14 mensen

11

meisjes		**jongens**
123346667	**0**	125788
00234	**1**	1599
14	**2**	23
eenheden	**tientallen**	**eenheden**

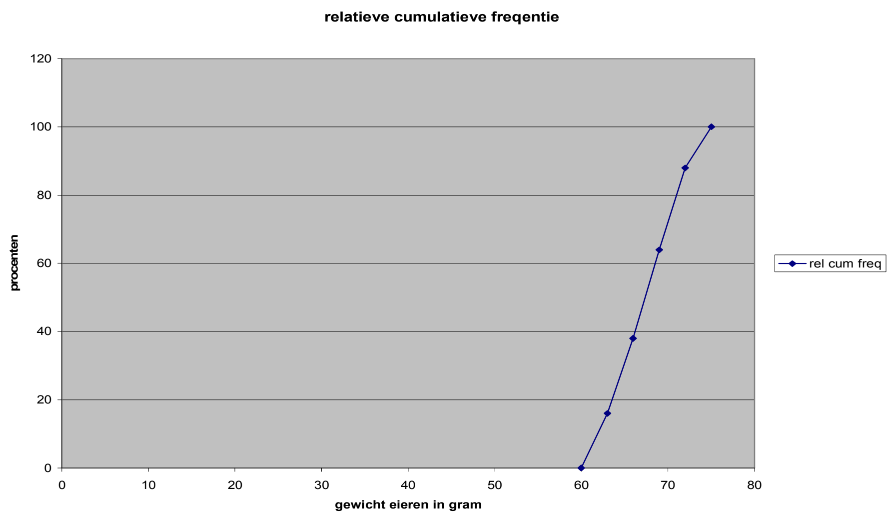
12a en b.

klasse	cum freq	rel cum freq
155	0	0
160	538	12
165	1673	37
170	2891	63
175	3832	84
180	4489	98
185	4572	100



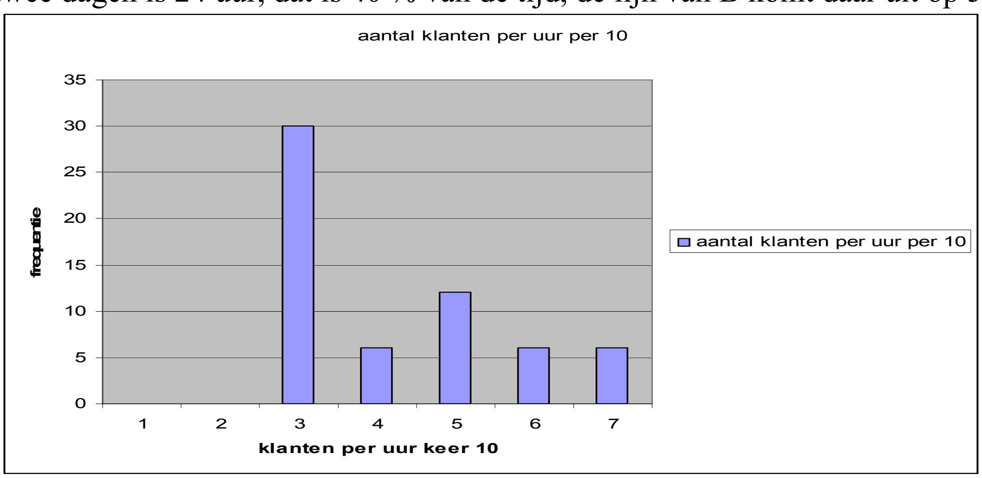
13.

klasse	cum freq	rel cum freq
60	0	0
63	8	16
66	19	38
69	32	64
72	44	88
75	50	100



14. Zodat alle vorige klassen en de laatste klasse bij elkaar geteld worden.
Let op: als je klasseverdelingen hebt, neem je bij de frequentiepolygoon het midden van de klasse, maar bij de cumlatieve frequentie polygoon de rechtergrens van de klasse (dus het hoogste getal).

15a. Het onderzoek duurde $5 * 12 = 60$ uur, daarvan tankten bij A 50% minder dan 30 klanten per uur, dus dat is 30 uur.
b. Minstens 40 klanten is 40 of meer per uur, dat is 80% van de tijd = 48 uur
c. twee dagen is 24 uur, dat is 40 % van de tijd, de lijn van B komt daar uit op 55 klanten.



d.

e. Bij pomp B is het juist drukker, want daar zijn er van 20 tot 60 klanten per uur, in 50% van de tijd, bij pomp A zijn dat er maar 30

16a. perceel 1, minder dan 50 kg per boom : 40% van 200 bomen = 80 bomen
perceel 2, minstens 60 kg per boom : 75% van 160 bomen = 120 bomen
perceel 1, tussen 50 en 70 kg per boom : 20% van 200 bomen = 40 bomen
b. de meeste kg per boom krijg je op perceel 2.
Perceel 1: 12200/200 = 61 kg per boom   Perceel 1 levert 240 kg meer op.

Perceel 2: 11760?160 = 73,5 kg per boom
Als de twee percelen even groot zijn staan de bomen ook wijder uit elkaar.
Welk perceel het meest oplevert moet je echt uitrekenen

perceel 1 :

Aantal bomen	Kg per boom	Totaal gewicht
80	45	3600
10	55	550
30	65	1950
70	75	5250
10	85	850
200		12200

perceel 2:

Aantal bomen	Kg per boom	Totaal gewicht
16	45	720
24	55	1320
8	65	520
32	75	2400
80	85	6800
160		11760

4.4 Centrum en spreidingsmaten.

24a. gemiddelde = $\frac{somvanallegetallen}{aanta\lg etallen} = \frac{200000+40000+250000}{12}$ = 40833 euro
b. rare manier om met gemiddelde om te gaan, vakbond?

25a. Tel alle frequenties bij elkaar . 20 dagen = 4 schoolweken.
b. 3∗0 +3∗1+ 7∗2 + 0∗3 + 3∗4 + 2∗5 + 2∗6 = 51
51/20 = 2,55 = gemiddelde
c. mediaan is het middelste getal van 20 getallen, dat zijn de getallen 2 en 2,
0 0 0 1 1 1 2 2 2 **2 2** 2 2 4 4 4 5 5 6 6 dus de mediaan is 2
Modus is wat het vaakst voorkomt dus 2

26a. 2∗0 +4∗1+ 5∗2 + 3∗3 + 3∗4 + 2∗5 + 1∗32 = 77 77/20 = 3,85 = gemiddelde
b. 0 0 1 1 1 1 2 2 2 **2 2** 3 3 3 4 4 4 5 5 32
mediaan en modus zijn ook allebei 2, net als bij de andere klas.
c. Gemiddelde verandert door uitschieter 32, andere twee maten niet.

27a. modus = 5, komt het vaakst voor (6 keer), mediaan = 6 (middelste getal)

cijfer	3	4	5	6	7	8	9	10
frequentie	2	4	6	5	4	4	3	2

3∗2 +4∗4+ 5∗6 + 6∗5 + 7∗4 + 8∗4 + 9∗3 + 2∗10 = 189 189/ 30 = 6.3 = gemiddelde
b. gemiddelde = 6.3, mediaan = 6, modus = 5
c. gemiddelde = 6.5 34∗6.5 = 221 punten totaal 189 + 27 = 216
het vierde cijfer is een 5, want 221 – 216 = 5

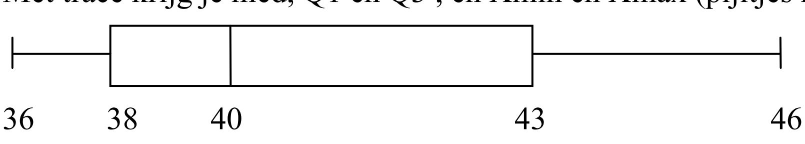
28a. kwalitatief (kenmerkend, zonder getal)
b en c. alleen de modus

29a. mediaan **b.** modus, want kwalitatief **c.** gemiddelde
d. modus (iemand heeft een fout gemaakt)

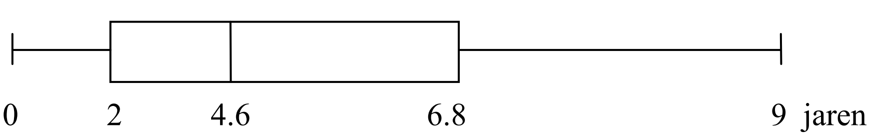
30a. Hoogst mogelijke gemiddelde,
18∗34 +3∗44+ 40∗54 + 9∗64= 3480 3480/70 = 49,7
Laagst mogelijke gemiddelde,
18∗25 +3∗35+ 40∗45 + 9∗55= 2850 2850/70 = 40,7
b. mediaan in klasse 45-< 55
c. 38 ligt in een klasse die maar drie keer voorkomt Stel dat er 3 leerk. zijn van 36.
In de klasse 45-< 55 zitten 40 lk. Dus zelfs bij een evenwichtige verdeling komt elke leeftijd 45, 46, 47, enz toch al 4 keer voor.

31a. 1800∗85 + 2200∗75 + 2600∗63 + 3000∗58 + 3400∗19 = 720400
720400/300 = 2501 branduren (300 = alle freq. opgeteld)
b. mediaan = middelste getal dus de klasse 2000-<2400
c. modale klasse komt het vaakst voor, dus 1600-<2000
d. Ik vermoed kleiner, kijkend naar de opbouw van de tabel.

GEBRUIK GR ! BELANGRIJK, dit moet je goed kunnen.
32. GR, Stat Edit aantal =L1= {36……46}, freq. =L2 = {6, 12, 18 enz zie tabel}
Stat calc 1 var stats enter enter (voer in met 2nd 1 en 2nd 2) L1,L2 enter
(vergeet de , niet)
Bovenste getal $\overline{X}$ = gemiddelde= 40.4
vijfde getal = σ x =standaardafwijking (komt later voor, ook als SD) = 2,693
zesde getal = n = aantal waarnemingen (dus alle frequenties opgeteld) = 105
pijltje naar beneden
minX = kleinste waarneming = 36
Q1 = 38 Mediaan = 40 Q3 = 43
maxX = grootste waarneming = 46
Boxplot maken van deze gegevens:
GR (2nd Y)Stat Plot 1 enter On enter, type 5 (pijltjestoets naar rechts 5 keer)
Xlist = L1 en Freq = L2 Zoom:9 Stat
Stel je window in x = [35,50] (zie tabel) , y hoeft niet
Met tracé krijg je med, Q1 en Q3 , en Xmin en Xmax (pijltjes nar links en rechts)



33a. Q1 = 25% van alle waarnemingen = 2 jaar Mediaan = 50% = 4.6 jaar
Q3 = 75% van alle waarnemingen = 6.8 jaar



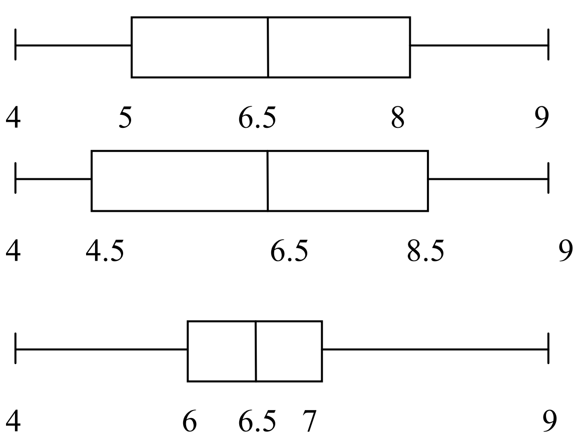
34a. 25% van de 50 staten = 12 of 13 staten
b. 25% van de staten heeft van 4 tot 6,3 miljoen inwoners,
dus ongeveer (3 a 4%) + 25% = 28.5% van de staten heeft meer dan 6 miljoen inwoners
c. 25 + 13 = 38 % van de staten heeft minder dan 3 miljoen inwoners
d. van Q1 tot Q3 is van 1.8 tot 6.3 miljoen inwoners, gemiddeld 4 miljoen per staat
het gaat dus om 25 staten ∗ 4 miljoen inwoners = 100 miljoen inwoners
e. Als je rekent totaal 290 miljoen inwoners, 100 miljoen = middelste 50%
eerste 25% heeft gemiddeld 1.2 miljoen inwoners ∗ 13 staten = 15 a 16 miljoen inwoners
Voor de grootste staten zijn 290 – 100 – 16 = 174 miljoen inwoners (gemiddeld 14 miljoen per staat)

35a. mediaan is steeds 3 km **b.** geeft geen goede indruk
c. de meeste leerlingen (50%) woont op een afstand tussen 2 en 4 km van school B
de meeste leerlingen (50%) woont op een afstand tussen 1 en 5 km van school A
d. spreiding het kleinst bij school C, het grootst bij school A

36a. Bij alledrie 70 – 30 = 40
b. A ; 44 -35 = 9 B: 43 -37 = 6 C: 55 – 32 = 23
c. grootste spreiding bij ll. C
d. spreidingsbreedte wordt 50, kwartielafstand verandert niet
e. spreidingsbreedte is gevoelig voor uitschieters.

37a. Alanya : spreidingsbreedte=12 – 4 = 8, kwartielafstand = 11 – 7 = 4
kleinste spreiding is Alanya
Mallorca: spreidingsbreedte=12 – 0 = 12, kwartielafstand = 10 – 6 = 4
Amsterdam; spreidingsbreedte=12 – 4 = 8, kwartielafstand = 8 – 2 = 6

38.



Aklas

Bklas meeste spreiding

Cklas minste spreiding

39. gebruik je GR stat edit L1 en L2 (freq), stat calc 1 var stats L1, L2
$\overline{X}$ = gemiddelde= 4.02
vijfde getal = σ x =standaardafwijking (komt later voor, ook als SD) = 1,517

40 zelfde manier, maar zonder freq, en daardoor alleen L1
$\overline{X}$ = gemiddelde= 7.224en σ x =standaardafwijking =0.253

41. L1 = klassemiddens = 67,5 ; 72,5 ; 77,5 ; 82,5 L2 = freq.
$\overline{X}$ = gemiddelde= 75,9
σ x =standaardafwijking = 2,9 (de meeste banden zitten tussen 73 en 78,8 km)
n = aantal waarnemingen = 500

42a. σ x =8 cm **b.** σ x =1,8

43a en c. gebruik je GR stat edit L1 en L2 (freq), stat calc 1 var stats L1, L2
$\overline{X}$ = gemiddelde = 5.094 en σ x =standaardafwijking = 0.124
Q1 = 5 mediaan = 5.1 en Q3 = 5.2, kwartielafstand =Q3 – Q1 = 0.2
b. Je ziet al dat de meeste waarnemingen rondom 5.1 liggen, dus is de σ x < 0.1

4.5 steekproeven

44a. suggestie door het woord “zeker ook”
b. suggestie door de woorden “vind U niet”
c. wat wordt bedoeld met “ veel”
46a. Allemaal mensen die boodschappen doen.
b. Mensen die allemaal ‘s morgens naar hun werk gaan.
c. Uit elke provincie 1 is te weinig.
d. Mensen die van natuur houden.

47a. 5% vlekken blijft, dat is nogal veel.
b. op 1 klas kan je geen conclusie baseren
c. de leeftijdsopbouw binnen de wijk is niet vermeld.
d. vergelijkingen met een onbekend product geven geen info.
e. het ligt eraan wanneer die neerslag valt
f. misschien is er een kortingsactie onder artsen geweest
g. dit is met alle opgepompte banden het geval

48 Uit deze tabel lijkt de ziekte vaker voor te komen bij mannen dan bij vrouwen,
Uit de tweede tabel blijkt de ziekte echter met grote waarschijnlijkheid aan roken gekoppeld te zijn.

50 toevalsgetallen maken met de GR:
MATH_$\leftarrow$_PRB_5:randInt(kleinste toevalsgetal, grootste toevalsgetal, aantal toevalsgetal) Daarbij betekent randInt: random = toeval en Int = integer = geheel getal
In dit geval heb je: MATH_$\leftarrow$_PRB_5:randInt(1, 480, 8)
55 = E55 445 = I45

51a. twee lettercodes met herhaling, dus $26 * 26 = 676$ codes
b. codes met letter A vooraan 1 – 26, B vooraan 27 – 52 , dus nr 65 is code CL
$26 * 16 = 416$ Q is de 17^{de} letter dus nr 430 is code QN

52a. de groepen verschillen sterk in grootte
b. 100 van de 5000 = 1 van de 50
dus 1 directielid, 90 winkelmedewerkers en 9 magazijnmedewerkers.

Boek 2 hoofdstuk 8 De normale verdeling.

8.1 Vuistregels bij de normale verdeling

1a. 155-< 160 ; 160-< 165 t/m 185-< 190 (dus per 5 cm, vanaf 1,55 t/m 1,90)
b. 1000 personen (getallen boven staven optellen)
c. L1 = klassemiddens = 157,5 ; 162,5 ; enz t/m 187,5
L2 = freq. = getallen boven staven
σ **x** (namen standaard deviatie en standaard afwijking is hetzelfde)
$\overline{X}$ = gemiddelde = μ (spreek uit mu) = 172,3 cm
σ x =standaardafwijking = 5,7 cm (de meeste mannen zitten tussen 166,6 en 178 cm
n = aantal waarnemingen = 1000
d. σ **x** (namen standaard deviatie en standaard afwijking is hetzelfde)
680/1000 = 68% **e.** 95 %
2a. klassenbreedtes van 1 cm **b.** 375 (aflezen grafiek)
c. nee, alleen al de groep van 170 tot 175 bestaat uit $5*370$ personen

4 zie vuistregel blz 188, die heb je bij de komende sommen nodig
a. $\mu - 2*\sigma$ x = 50 $\mu - \sigma$ x = 60 $\mu = 70$ $\mu + \sigma$ x = 80 $\mu + 2*\sigma$ x = 90
b. 68 % **c.** 95% **d.** 2,5% **e.** 47,5%

6. 16% = 2,5% + 13,5% , dus 76 gram = $\mu - \sigma$ x invullen μ = 80 gr. $\Rightarrow \sigma$ x = 4 gr.

7a. Het gemiddelde μ zie je bij 50%, **b.** $\mu + \sigma$ x = 84 % (want 50% + 34%)
c. μ = 68 uur en $\mu + \sigma$ x = 75 uur dus **d.** σ x = 7 uur

9a. De top ligt bij allebei op hetzelfde getal 167cm.
b. standaardafwijking is groter bij de groene lijn, die wijkt verder uit
c. De oppervlakte onder beide grafieken is even groot, dus platter, dan ook breder
d. De lijn heeft dezelfde vorm en hoogte en breedte als lijn A, maar is naar links verschoven. (top boven het gemiddelde van lijn C)

10a. lijn A, de reactie tijd van jonge mensen is korter dan van oudere mensen.
b. lijn C, de 60 jarigen want σ x is bij deze lijn het grootst

11a. μ = 65 σ x = 1 **b.** μ = 66,5 σ x = 1
c. μ = 67,5 σ x = 1,25 **d.** μ = 70 σ x = 1