

Basisprincipes

Deel 1: databundels en instellingsprofielen: interpretatieoefening

Thema 1: leerlingenkenmerken

Thema 2: leerlingenstromen IN

Thema 3: leerprestaties

1. We gaan in groepen aan de slag met databundels en instellingsprofielen.

Beschrijf welke gegevens je over deze school kan afleiden.
Wat valt je op?
Vind je dezelfde gegevens terug in beide bronnen? Zijn ze even rijk of complementair?
Kan je vergelijken met anderen? Welke referentiegroepen?
Wat vind je goed, wat vind je niet goed?
Welke vragen zou je nog stellen?

Lees de gegevens met volgende vragen in het achterhoofd:

--> Welke gegevens vind je het meest waardevol?

--> Welke wijze van voorstellen is het makkelijkst om een overzicht te hebben/te lezen.

2. Plenum

Stel je bevindingen voor aan de rest van de groep.
Hoe kader je dit binnen schoolontwikkeling?
Wat kunnen deze gegevens voor de school betekenen?
Hoe kunnen de gegevens je helpen?

3. Hoe kunnen we nu zelf aan de slag gaan met de gegevens uit de databundels en instellingsprofielen?

--> grafische voorstellingen maken met de gegevens uit de bundels.

Deel 2: basisprincipes beschrijvende statistiek

Beschrijvende statistiek helpt je om

--> algemene trends of tendensen in je data samen te vatten.

--> de spreiding van de scores te begrijpen.

--> een inzicht te krijgen in hoe een score zich tot de andere scores verhoudt.

Beschrijvende statistiek biedt je mogelijkheden om een verhaal te construeren op basis van je geselecteerde gegevens.

We maken hiervoor gebruik van frequenties (absoluut en relatief), centrale tendensmaten (modus, mediaan, gemiddelde, kwartielen) en spreidingsmaten (standaardafwijking).

Hieronder beschrijven we de verschillende maten en tonen we hoe je ze in Excel¹ kan bepalen of berekenen. We geven ook enkele voorbeelden van grafische voorstellingen.

Frequentie & proportie

De frequentie is het aantal keer een bepaalde waarde of score voorkomt. Dit kan men uitdrukken in aantallen (absolute frequentie) of percentages (relatieve frequentie of proportie).

Bv: Hoeveel jongens/meisjes zitten er in een bepaald leerjaar?

Hoeveel jongens/meisjes zitten er in elke klas?

Hoeveel leerlingen vinden het nooit/meestal niet/meestal wel/altijd leuk in de klas?

Hoeveel leerlingen scoren 16/20 voor de toets wiskunde?

De samenvatting van het aantal keer de verschillende waarden voorkomen noemen we de frequentieverdeling van een variabele.



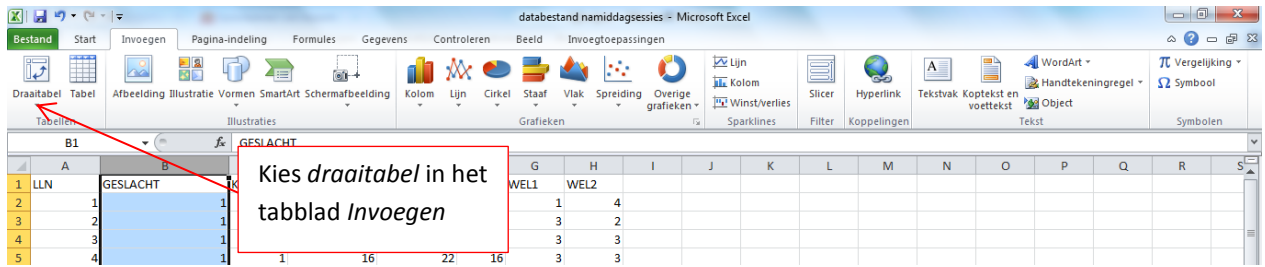
Hoe kunnen we een frequentieverdeling, een overzicht van de scores op een variabele, weergeven?

Draaitabel

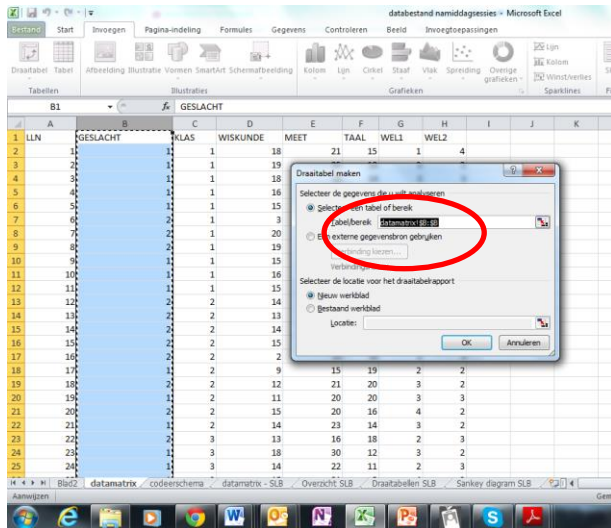
Een draaitabel is een eenvoudige tabel waarin je de frequentie kan aflezen. Je wil bv de verdeling tussen het aantal jongens/meisjes in je databestand weergeven.

1. Je selecteert de kolom waarin de variabele *geslacht* staat.
2. Kies in het tabblad *invoegen* voor *draaitabel*.

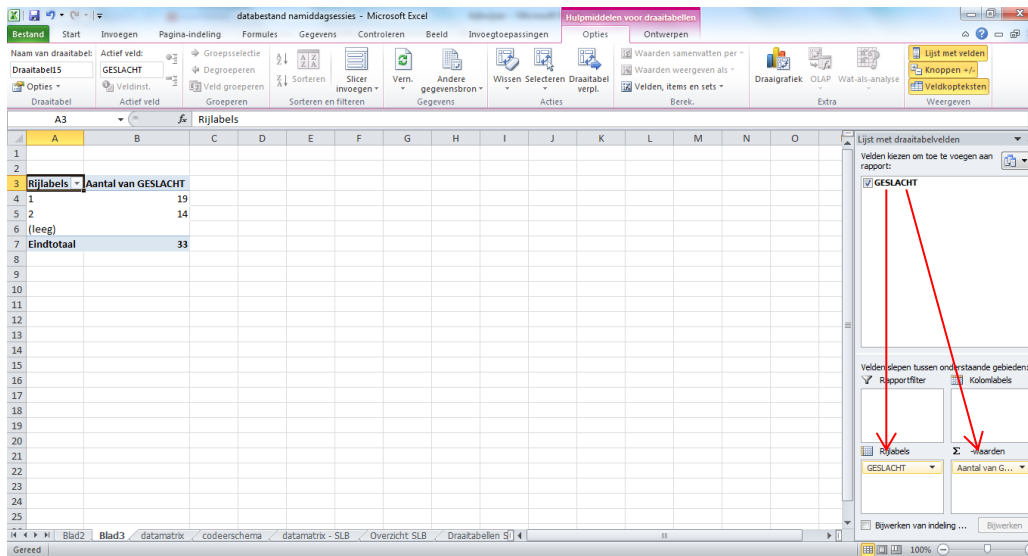
¹ We werken met Excel 2010



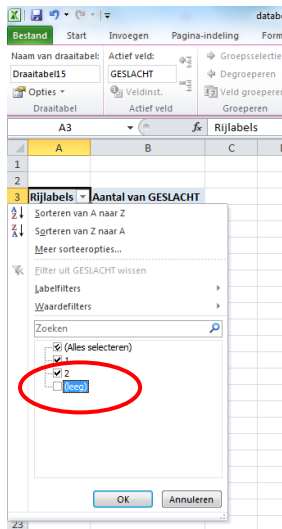
3. In het dialoogvenster is het celbereik waarin de waarden voor je tabel staan, ingevuld. Kies OK.



4. In het volgende venster sleep je geslacht naar het vak rijlabels en het vak waarden.

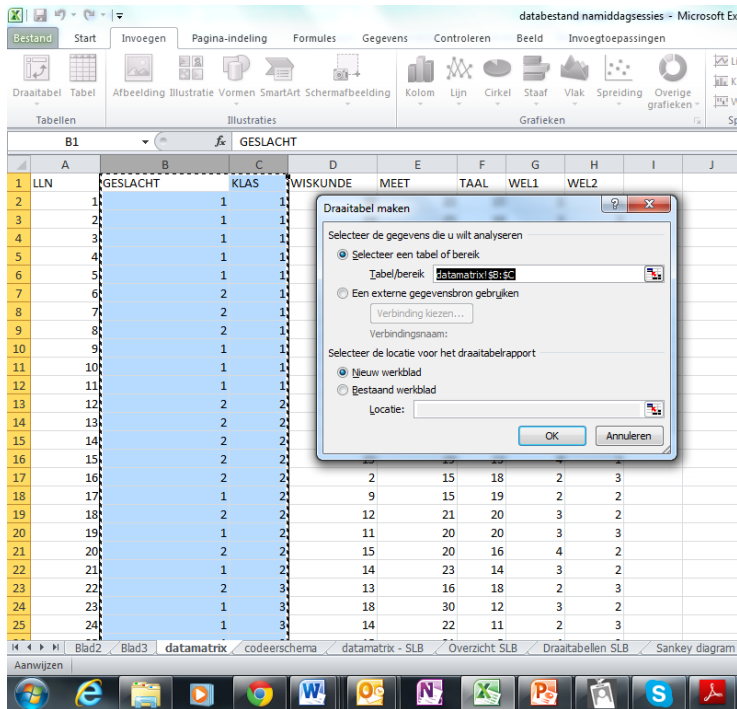


5. Als resultaat krijg je een tabel waarin je kan aflezen hoeveel jongens/meisjes er in je steekproef/populatie zitten. Je kan de categorie *leeg* wegdoen door op het pijltje naast *rijlabels* te klikken en af te vinken.

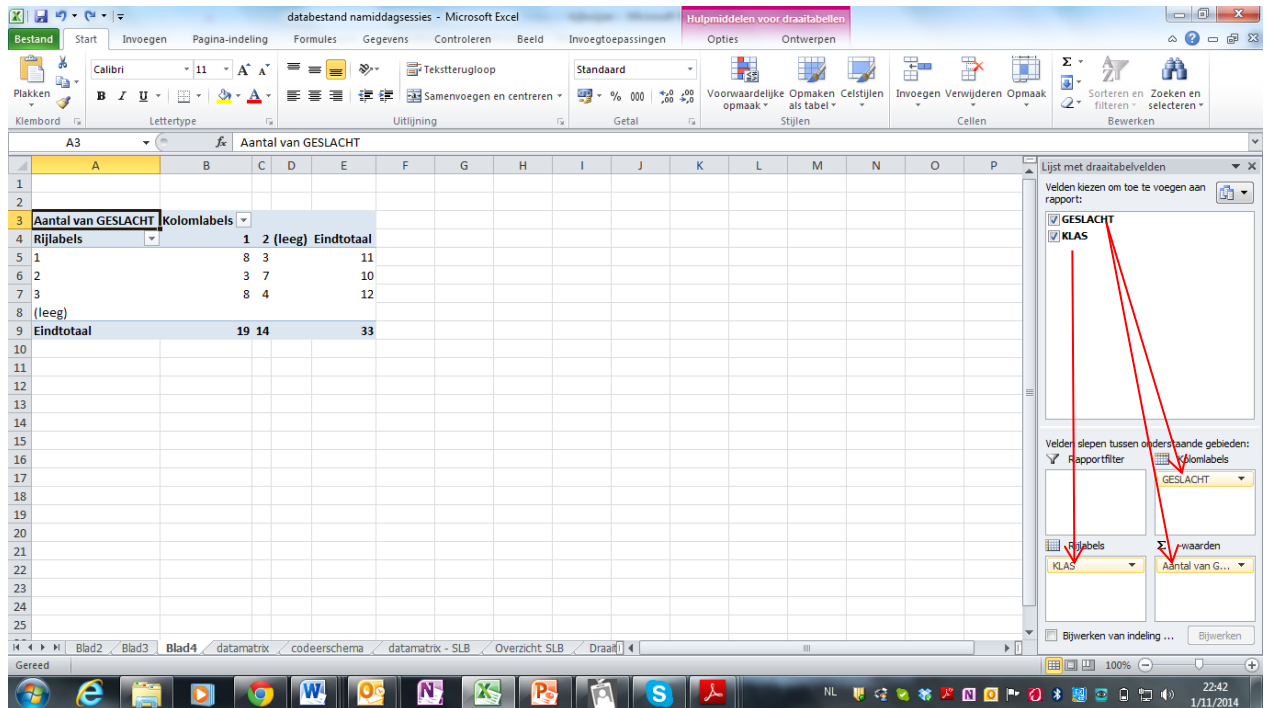


Je kan ook een draaitabel maken met verschillende variabelen. Je wil misschien weten hoeveel jongens/meisjes er in elke klas zitten. Hiervoor ga je op dezelfde manier tewerk.

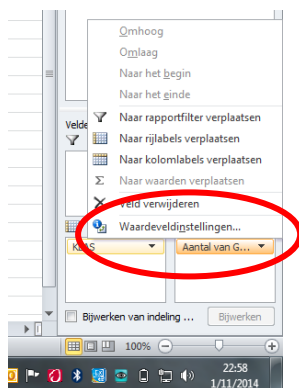
1. Je selecteert voor deze draaitabel 2 kolommen: *geslacht* en *klas*.



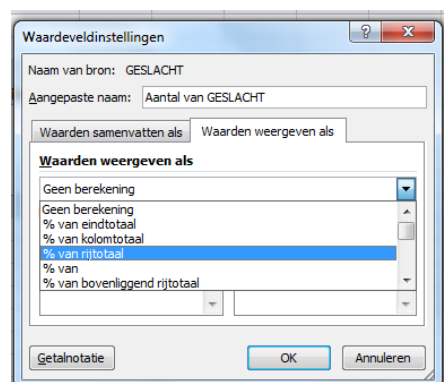
2. Je sleept *klas* naar het veld *rijlabels* en *geslacht* naar zowel *kolomlabels* als naar *waarden*. Dan krijg je volgende tabel:



3. Je kan de categorieën (*leeg*) opnieuw wegvinken via het pijltje naast *rijlabels* en *kolomlabels*.
4. Kies je ervoor om percentages weer te geven in de tabel klik dan op het pijltje naast *aantal van geslacht* in het vak *waarden* en kies voor *waardeveldinstellingen*.



5. Kies het tabblad *waarden weergeven als* en selecteer *% van rijtotaal* als je de proportie binnen de klas wil weergeven.



6. Volgende tabel is het resultaat

	A	B	C	D	E
1					
2					
3	Aantal van GESLACHT		Kolomlabels		
4	Rijlabels	1	2	Eindtotaal	
5	1	72,73%	27,27%	100,00%	
6	2	30,00%	70,00%	100,00%	
7	3	66,67%	33,33%	100,00%	
8	Eindtotaal	57,58%	42,42%	100,00%	
9					
10					

Je kan de percentages afronden door de cellen te selecteren en in het menu *getal* op het tabblad *start* te kiezen voor *minder decimalen*.

Tenslotte kan je de tabel duidelijker maken door de rijlabels en de kolomlabels aan te passen.

klas	geslacht		Eindtotaal
	meisje	jongen	
klas A	73%	27%	100%
klas B	30%	70%	100%
klas C	67%	33%	100%
Eindtotaal	57,58%	42,42%	100,00%

Zo kan je via draaitabellen in Excel verschillende overzichten creëren.



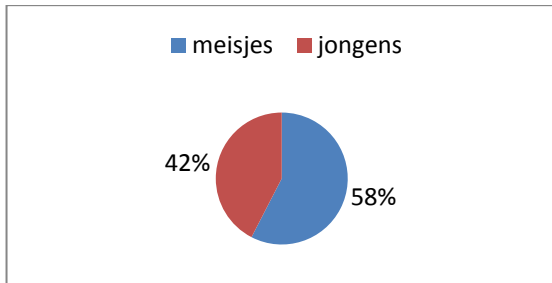
Hoe kunnen we een frequentieverdeling grafisch weergeven?

Taartdiagram

Met een cirkeldiagram of taartdiagram kan je een categorische of een ordinale variabele grafisch voorstellen.

We nemen opnieuw de vraag naar de frequentieverdeling jongens/meisjes als voorbeeld.

1. Maak een draaitabel zoals hierboven uitgelegd.
2. Je zorgt ervoor dat de categorie *leeg* is afgevinkt.
3. Je selecteert de tabel en kiest in het tabblad invoegen voor cirkel in het grafiekenmenu.
4. Je kiest voor de eerste 2D-cirkel.
5. Je kan de grafiek verder opmaken via het tabblad *hulpmiddelen voor draaigrafieken*.

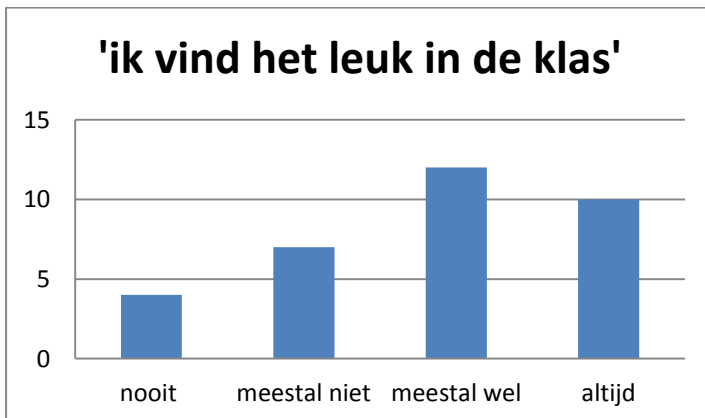


Kolomdiagram

Een andere mogelijkheid om de frequentie of proportie van een categorische of een ordinale variabele weer te geven is een kolomdiagram. Tussen de kolommen is er ruimte om duidelijk te maken dat het om verschillende categorieën gaat.

Als voorbeeld nemen we de vraag 'Hoeveel leerlingen vinden het nooit/meestal niet/meestal wel/altijd leuk in de klas?'.

Je gaat op dezelfde manier tewerk als bij een taartdiagram. Alleen selecteer je nu *kolomdiagram* in het menu *grafieken*. Je kan de grafiek opnieuw opmaken via het tabblad *hulpmiddelen voor draaigrafieken*.



Centrale tendensmaten

Centrale tendensmaten zijn cijfers die een samenvatting geven van de verschillende waarnemingen. De keuze van het kengetal hangt af van het niveau van de variabele.

meetniveau	modus	mediaan	gemiddelde
nominaal	X		
ordinaal	X	X	
kwantitatief (interval/ratio)	X	X	X

Modus

De **modus** is de categorie of score die het meeste voorkomt in een reeks waarnemingen.

Voor sommige variabelen kunnen er meerdere modi zijn. Dit wil zeggen dat er twee of meer categorieën het meest voorkomen.

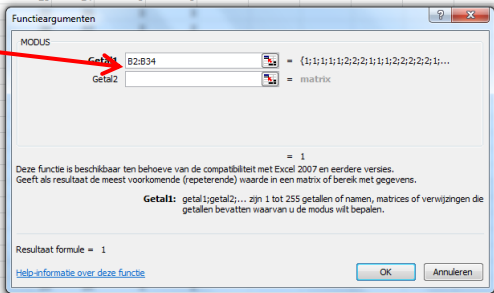


Hoe bepaal je de modus van een aantal waarnemingen in Excel?

1. Selecteer de cel waar je het resultaat wil.
2. Klik op *Fx* (functie)
3. Selecteer *MODUS* in het dialoogvenster

4. Selecteer het gegevensbereik waarvoor je de modus wil berekenen --> **OK**

1	LLN	GESLACHT	KLAS	WISKUNDE	MEET	TAAI	WEL1	WEL2
2	1	1	1	18	21	15	1	4
3	2	1	1	19	25	18	3	2
4	3	1	1	28	14	3	3	
5	4	1	1					
6	5	1	1					
7	6	2	1					
8	7	2	1					
9	8	2	1	19				
10	9	1	1	15				
11	10	1	1	16				
12	11	1	1	15				
13	12	2	2	14				
14	13	2	2	13				
15	14	2	2	14				
16	15	2	2	15				
17	16	2	2	2				
18	17	1	2	9				
19	18	2	2	12				
20	19	1	2	11				
21	20	2	2	15				
22	21	1	2	14				
23	22	2	3	13	16	18	2	3



5. Het resultaat verschijnt in de cel die je bij de start geselecteerd hebt.

32	31	1	3
33	32	2	3
34	33	1	3
35			
36			
37		1	
38			
39			
40			
41			
42			
43			

Mediaan

De **mediaan** is de middelste waarde als je de scores rangschikt van laag naar hoog. De mediaan verdeelt de waarnemingen in twee gelijke groepen. 50% van de scores liggen boven de mediaan en 50% van de scores liggen beneden de mediaan.

De mediaan kan een niet waargenomen score zijn.

Stel dat je 10 waarnemingen hebt voor de variabele welbevinden in de klas. Je rangschikt de waarnemingen in oplopende volgorde: 1-1-1-2-**2-3**-3-3-4-4. De mediaan is dus de waarde tussen de 5^{de} en de 6^{de} waarneming, tussen het getal 2 en 3.

De mediaan kan ook een fictieve waarde hebben.

Volgende 10 waarnemingen geven het aantal leerlingen per klas weer dat met de fiets naar school komt: 8-8-10-11-**11-12**-12-14-15-18. De mediaan bedraagt 11,5 wat een fictieve waarde is voor deze variabele.



Hoe bepaal je de mediaan in Excel?

Voor mediaan ga je op dezelfde wijze tewerk als voor de modus. In het dialoogvenster bij stap 3 selecteer je nu *mediaan*.

Gemiddelde

Het **gemiddelde** van een reeks getallen is de som van die getallen gedeeld door het aantal waarnemingen/respondenten. Het is een kengetal voor variabelen op intervalniveau.

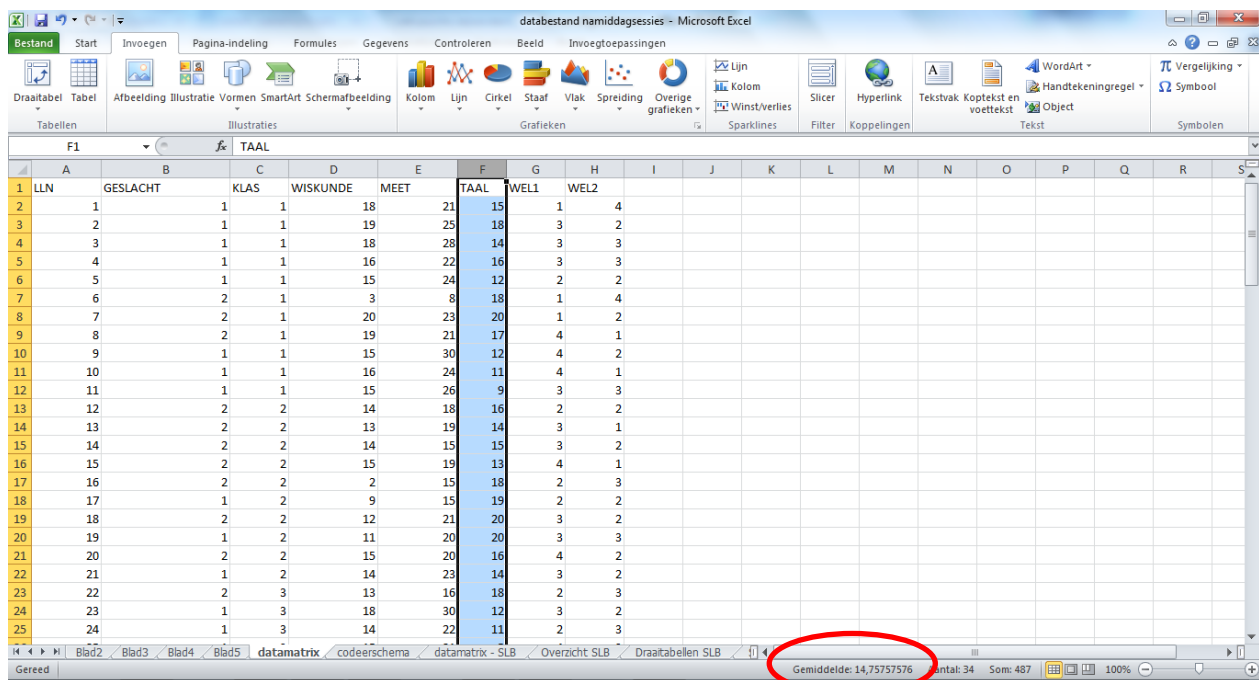
Het gemiddelde is de meest gebruikte statistiek om geobserveerde waarden van alle respondenten (leerlingen) samen te vatten. Het gemiddelde is gevoelig voor extreme waarden (uitbijters). Die kunnen het gemiddelde beïnvloeden.

Ook het gemiddelde kan een niet waargenomen waarde hebben, zoals bij de mediaan.



Hoe bereken je het gemiddelde in Excel?

Als je de kolom met de variabele waarvoor je het gemiddelde wil weten selecteert, krijg je rechts onderaan het scherm het gemiddelde.



Spreidingsmaten

Het gemiddelde kan een misleidend beeld geven van de variabele. Je hebt dan meer informatie nodig om de gegevens te kunnen beschrijven.

Stel, je wil de resultaten van een toets voor twee parallelklassen vergelijken. Het gemiddelde voor beide klassen is 7. Maar als je verder kijkt naar de cijfers op leerlingniveau geeft het een heel ander beeld:

	Klas A	Klas B
gemiddelde score op toets	7	7
individuele cijfers leerlingen	7-7-7-6-8-7-7-6-8-7	10-10-5-4-9-7-10-7-7-1

De cijfers tonen aan dat de klassen toch heel verschillend zijn. In klas B is er veel meer **spreiding** in de scores. **Spreidingsmaten** geven een indicatie van de spreiding binnen de waargenomen scores. Ze kunnen ons helpen om de spreiding van de cijfers te interpreteren. Op basis van de spreidingsmaten kunnen we twee groepen binnen een variabele vergelijken bv de score op een toets wiskunde voor jongens en meisjes.

Kwartielen

Om een eerste idee te krijgen van de spreiding van een variabele van ten minste ordinaal niveau kunnen we gebruik maken van kwartielen.

Kwartielen verdelen de geordende waarnemingen in 4 gelijke groepen. Elke deel omvat dus 25% van de waarnemingen. Ze worden aangeduid met kwartiel 1 (K1), kwartiel 2 (K2) en kwartiel 3 (K3). K2 komt overeen met de mediaan. De afstand tussen K1 en K3 is het interkwartiel en omvat de middelste 50% van de waarnemingen.

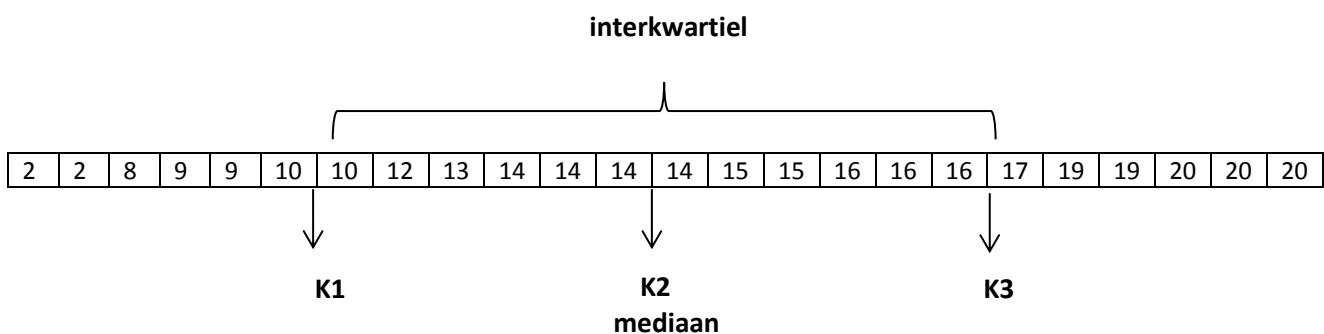
In onderstaand voorbeeld is

$$K1 = 10$$

$$K2 = 14$$

$$K3 = 16,5$$

$$\text{Het interkwartiel} = 16,5 - 10 = 6,5$$

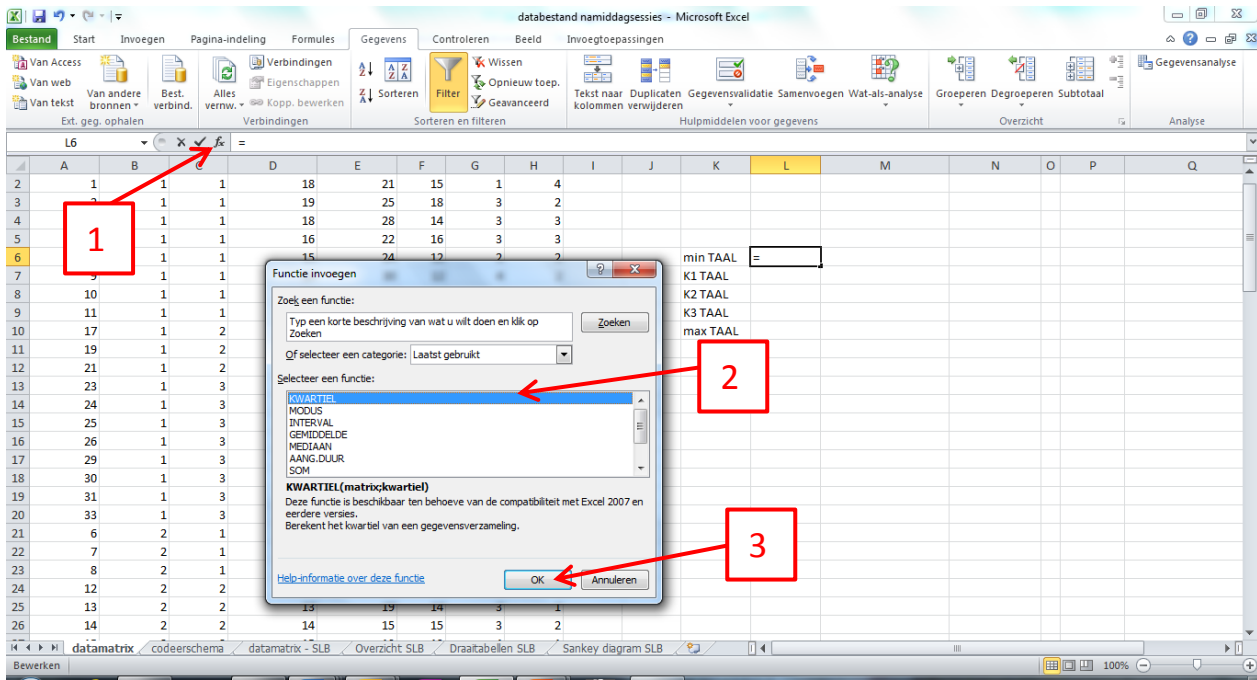


Hoe bepaal ik de kwartielen van een variabele in Excel?

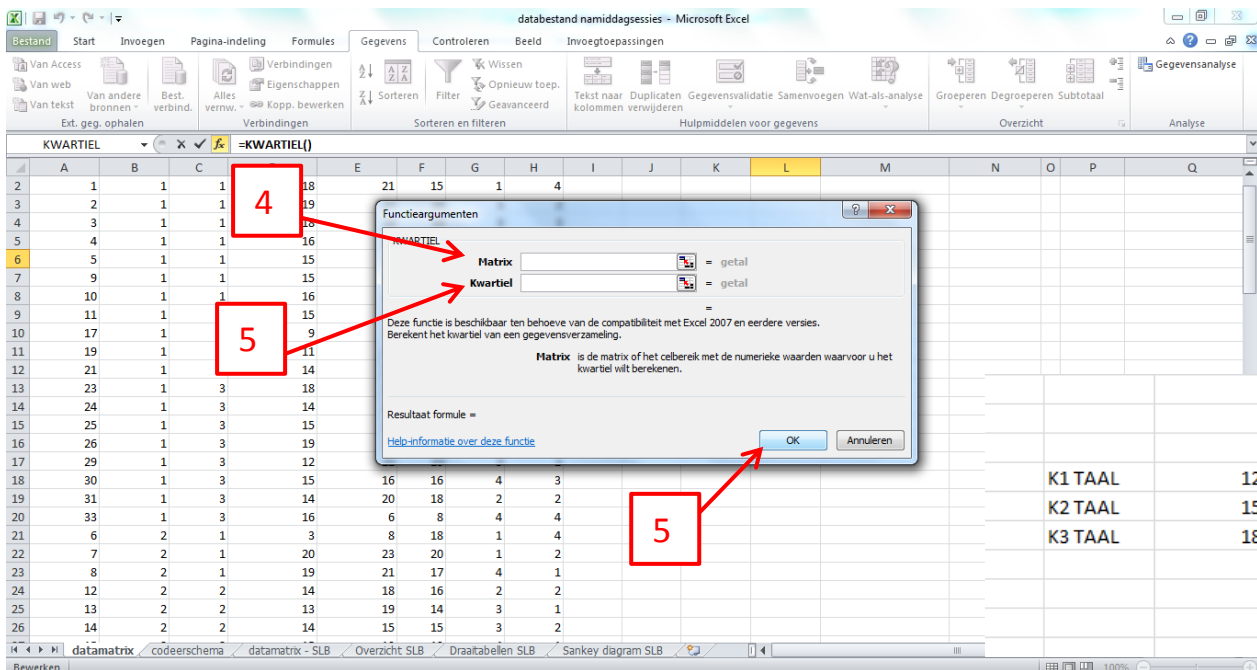
In dit voorbeeld berekenen we de kwartielscores voor de variabele TAAL in het voorbeeldbestand.

Dit kan met de formulefunctie *kwartiel*. Je moet dit telkens voor elk kwartiel apart doen.

1. Klik *Fx* (functie)
2. Selecteer 'kwartiel'
3. OK



4. Selecteer de toetsscores voor TAAL
5. Duid aan welk kwartiel je wil berekenen: 1, 2 of 3
6. OK



Standaarddeviatie

De **standaarddeviatie (SD)** of **standaardafwijking** is een maat waarmee je zicht krijgt op de spreiding van de scores rondom het gemiddelde. Een hoge standaardafwijking wijst op veel verschillen tussen de scores. Als de standaardafwijking op een toets hoog is betekent dit dat de leerlingen zeer gevarieerd gescoord hebben ten opzichte van elkaar. Om te bepalen of de standaarddeviatie hoog is moet je steeds de vergelijking maken met het gemiddelde.

Als je bij een vergelijking van 2 groepen naast het gemiddelde ook de SD vermeldt, wordt ook de verdeling binnen de groepen duidelijk.



Hoe kan je de standaardafwijking berekenen in Excel?

Om dit te doen moet je eerst extra statistische functies installeren. Hoe je dit doet zie je in volgend filmpje:

https://www.youtube.com/watch?v=TIcucKmsf_Y

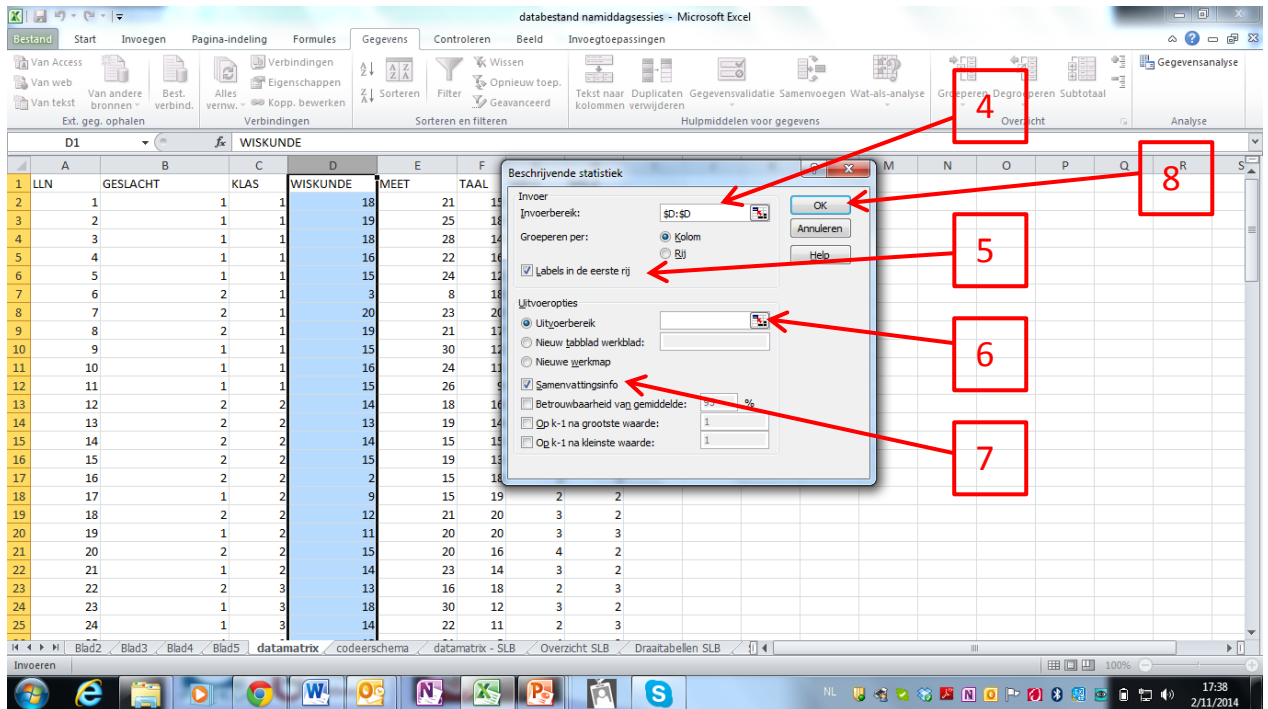
In het tabblad *gegevens* is nu een extra tabblad toegevoegd: *gegevensanalyse*.

Stel dat je de standaardafwijking van de score voor WISKUNDE in het oefenbestand wil weten:

1. Kies in het tabblad *gegevens* voor *gegevensanalyse*.
2. Selecteer in het dialoogvenster *beschrijvende statistiek*
3. OK

The screenshot shows the Microsoft Excel interface with the 'Gegevensanalyse' (Data Analysis) dialog box open. The dialog box is titled 'Gegevensanalyse' and has a list of functions. The 'Beschrijvende statistiek' (Descriptive Statistics) option is selected and highlighted in blue. Red boxes and arrows point to the 'OK' button (labeled 3), the 'Beschrijvende statistiek' option (labeled 2), and the 'Gegevensanalyse' button in the ribbon (labeled 1). The spreadsheet in the background shows a table with columns labeled LLN, GESLACHT, KLAS, WISKUNDE, MEET, TAAL, WEL1, and WEL2. The 'WISKUNDE' column is highlighted in blue.

4. In het dialoogvenster *beschrijvende statistiek* geef je eerst het invoerbereik aan.
5. Als je de volledige kolom selecteert vink je ook *labels in eerste rij* aan omdat de naam van de variabele in de eerste rij staat.
6. Bij uitvoerbereik geef je aan waar op het werkblad je de uitdraai wil.
7. Vink *samenvattingsinfo* aan.
8. Tenslotte klik je op *OK*.



9. Dit is de output:

<i>WISKUNDE</i>	
Gemiddelde	14,39394
Standaardfout	0,692136
Mediaan	15
Modus	15
Standaarddeviatie	3,976017
Steekproefvariantie	15,80871
Kurtosis	3,421554
Scheefheid	-1,56174
Bereik	18
Minimum	2
Maximum	20
Som	475
Aantal	33

Je merkt dat er naast de standaardafwijking ook andere beschrijvende statistieken die we besproken hebben gerapporteerd worden.