**Statistiek met grote datasets - HAVO Wiskunde A

# Auteurs

Marc Bouman; e-mail: mbouman@vszutphen.nl

Jossy Carmelia; e-mail: j.carmelia@osg-erasmus.nl

Marian Gall; e-mail: m.gall@rsg-sneek.nl

Bert Kraai; e-mail: bkraai@vszutphen.nl

Jan Mostert; e-mail: j.mostert@wolfsbos.nl

Eva Vermeulen; e-mail: e.vermeulen@annamaria.nl

Yde van Wieren; e-mail: y.vanwieren@singelland.nl

# Inleiding

In de voorlopig vastgestelde examenprogramma's Wiskunde A (Havo en VWO) en Wiskunde C (VWO) is het domein Statistiek uitgebreid, vanuit de wens om leerlingen beter voor te bereiden op vervolgstudies op HBO en Universitair niveau. De belangrijkste veranderingen zijn:

1. het domein Statistiek wordt ook bij het centraal examen getoetst;
2. de nadruk verschuift van reproductie van kennis en opgaven, naar inzicht en productieve vaardigheden;
3. leerlingen dienen statistische vaardigheden toe te kunnen passen bij grote datasets met behulp van ICT en deze praktische vaardigheden dienen getoetst te worden bij het schoolexamen.

*Speciaal voor de NWD is dit document ingekort om beter aan te sluiten bij de doelgroep en het doel van de workshop. Het volledige document kan gevonden worden op de site van de leergang wiskunde*

<http://www.leergangwiskunde.nl/lesmateriaal.html>

# 4.Spreidingsplot

We gaan een spreidingsplot maken in het werkblad spreidingsplot bij de minimum temperatuur en het aantal mm neerslag. Zo zouden we bijvoorbeeld kunnen zien dat er meer neerslag valt als de minimum temperatuur laag is.

\* Selecteer de cellen C3 t/m C33 en E3 t/m E33.

\* Klik op invoegen en bij grafieken op spreiding.

\* Bij het menu indeling kun je verticale rasterlijnen aanzetten zodat je een beter beeld krijgt.

\* As titels maken: Labels, astitels. Zet horizontaal “minimum temperatuur” en verticaal

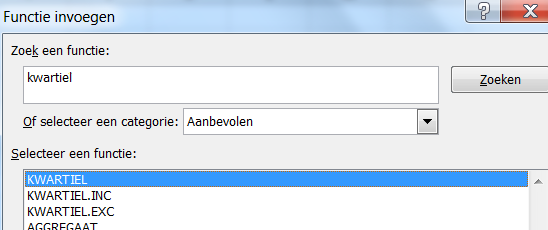
“neerslag”.

\* Grafiektitel: Geef de grafiek de titel: “neerslag bij minimum temperatuur”.

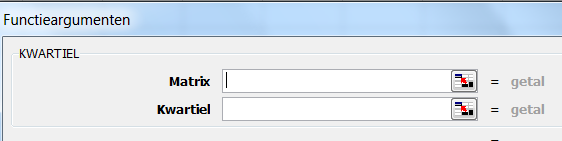
# 5. Boxplot

Voor 2012 en 2013 ga je een boxplot maken van de maximum temperatuur.

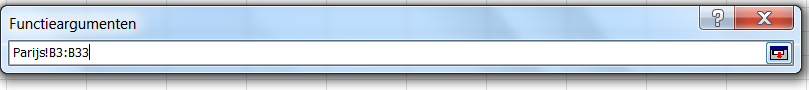
Daarvoor moet je eerst de kwartielen berekenen met de formule in het werkblad boxplot.

\* Maak in cel B2 met fx de functie: = KWARTIEL

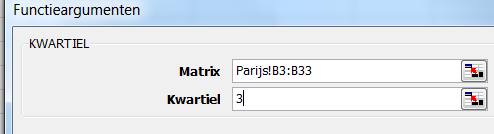
\* Druk op OK



\* Op het blokje bij matrix klikken, naar het werkblad Parijs en B3:B33 selecteren



\* Enter en bij kwartiel 3 invullen



\* Druk op OK

\* Op dezelfde manier zet je de mediaan=kwartiel 2 in de tabel en kwartiel 1.

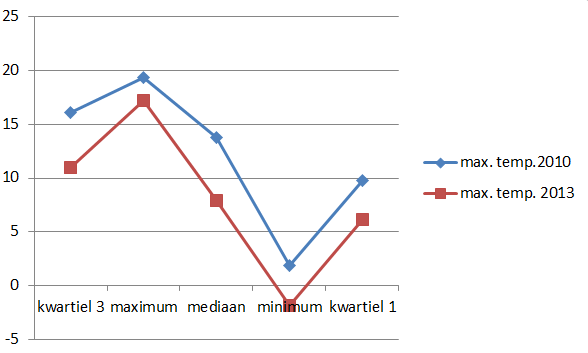
\* Max en min van de maximum temperatuur kun je kopiëren(die had je al).

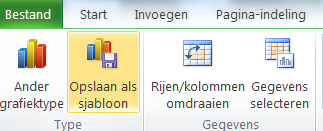
\* Maak de tabel af met de kwartielen,mediaan,max/min van 2013

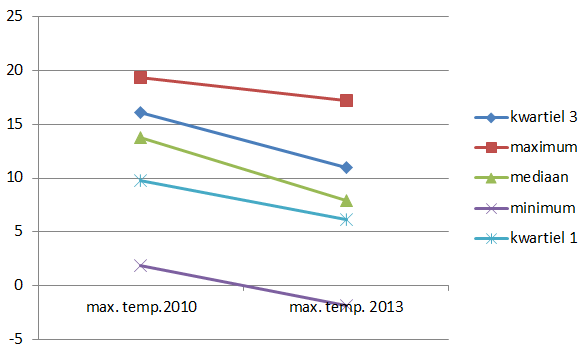
\* Deze gegevens gaan we in een boxplot zetten:

\* Maak van de tabel een lijngrafiek met gegevensmarkeringen, dus met puntjes

De volgende grafiek is het resultaat:



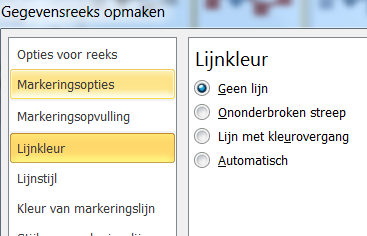
**\*** Bij het ontwerpen kies je dan voor “Rijen en kolommen omdraaien”.

\* Je krijgt dan

\* Klik hier op een van de gegevenspunten en klik op de rechtermuisknop. Kies hier voor

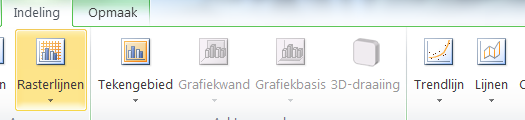
Gegevensreeks opmaken. Kies voor lijnkleur en vervolgens “Geen lijn”. Doe dat voor alle

vijf lijnen afzonderlijk.

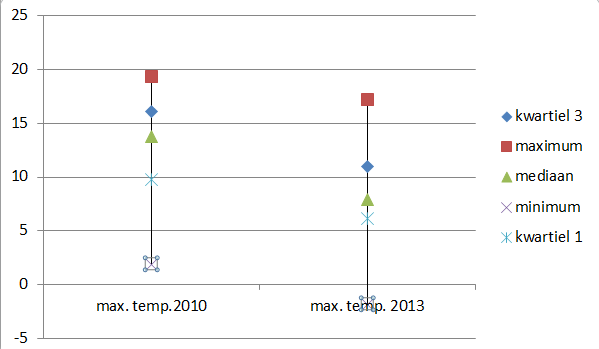


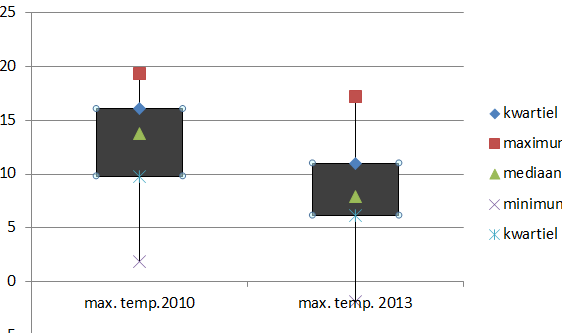
\* Als alle tussenlijnen zijn verwijderd, markeer je weer een gegevenspunt en kies in het menu

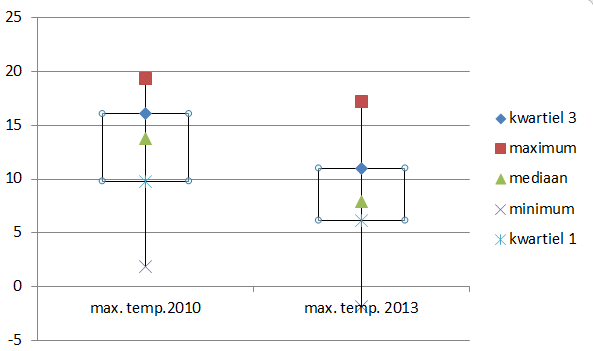
‘Indeling’ voor ‘Lijnen’ en vervolgens voor ‘Hoog/laag lijnen’.



\* Excel trekt nu twee verticale lijnen door de gegevens

Door nu te kiezen voor Omhoog/omlaag balken voegt Excel boxen toe





\* Wanneer je op de box gaat staan en op de rechtermuisknop

klikt, krijg je een menu waarmee je de opmaak van de box

kunt aanpassen. Kies hiervoor ‘Dalingsbalk opmaken’.

Je kunt kiezen voor geen vulling

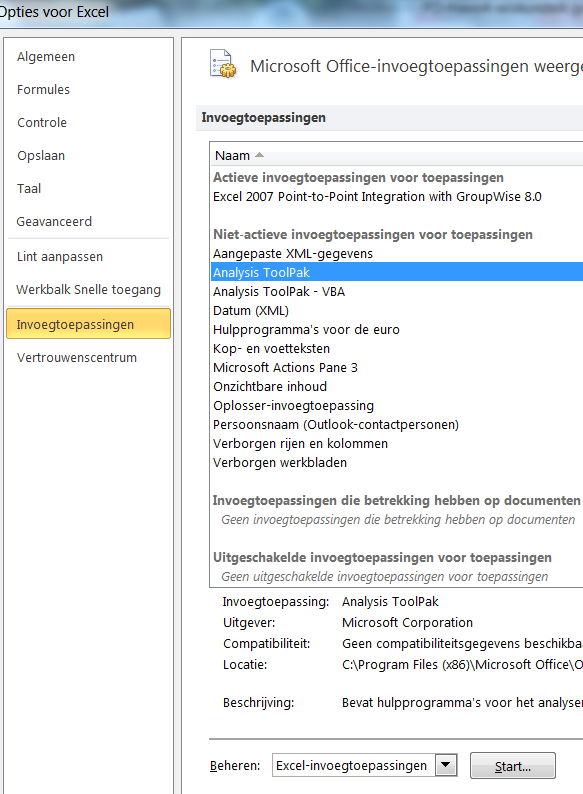
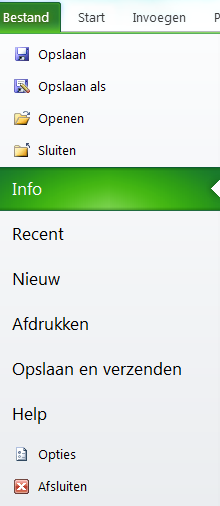
\* Sla deze boxplot op

# Gegevensanalyse

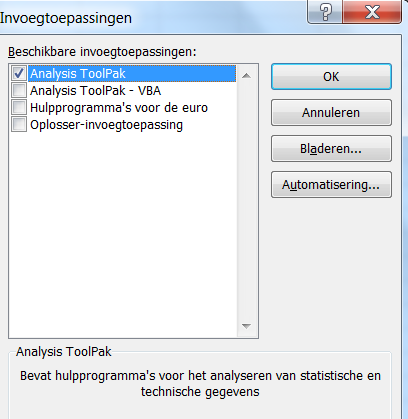
Je moet eerst Toolpak installeren en dat doe je als volgt:

\* Ga naar bestand,

Dan naar opties vervolgens invoegtoepassingen,selecteer Analysis Toolpak zie hieronder

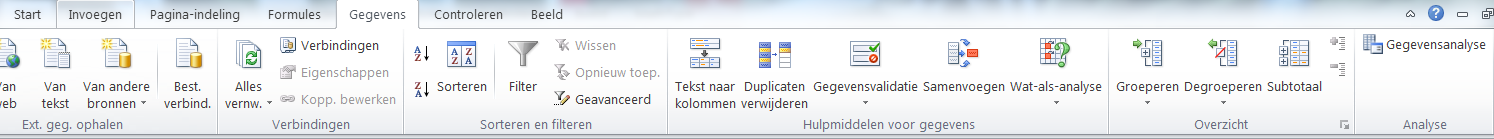


\* Druk op Start bij Beheren: Excel-invoegtoepassingen zie hierboven



\* Bij invoegtoepassingen aanvinken Analysis

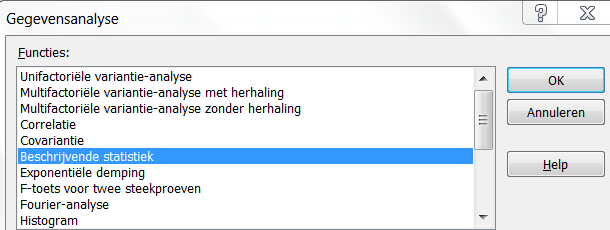
Toolpak, dan Ok



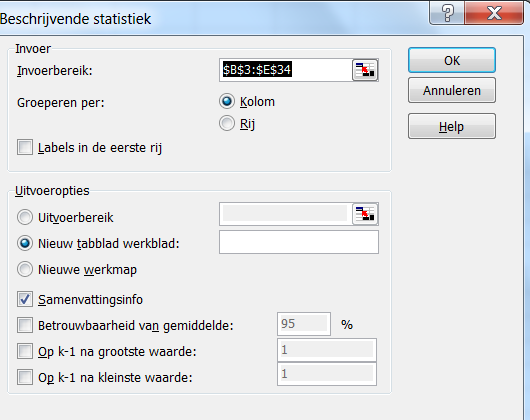
In de bovenste balk staat Gegevens, klik deze aan en klik op Gegevensanalyse.

\* **Selecteer eerst B3 t/m E33 uit het werkblad Parijs**

\* in de Gegevensanalyse Beschrijvende statistiek selecteren, dan ok,



\* Vink "Samenvattingsinfo" aan.



\*

Als je op ok drukt krijg je een overzicht van allerlei statistische gegevens. Voor ons zijn van belang het gemiddelde, de mediaan, de modus, de standaarddeviatie, enz. Voor dit overzicht wordt een nieuw tabblad aangemaakt. Noem deze "statistische gegevens”.

Sla het bestand op !!

# Opdrachten

Middelen die nodig zijn bij deze opdracht:

* Je krijgt een groot Excelbestand met de leerlinggegevens van een school. Met deze gegevens ga je aan de slag.

**Opdracht 1**

Maak een spreidingsplot waarbij op de horizontale as de cijfers voor het vak Ned (Nederlands). staan en op de verticale as de cijfers voor het vak Eng (Engels). Doe dit voor klas 11.

Selecteer de betreffende kolommen en kopiëer deze naar het tabblad “opdracht 1”.

Benoem de assen en geef de grafiek een titel.

Zet in de grafiek een lineaire trendlijn.  
Vragen:

Waarom zou je zo’n grafische voorstelling willen maken?  
Welke conclusie kun je trekken uit de grafische voorstelling die je hebt gemaakt?

**Opdracht 2**

We gaan cirkeldiagrammen maken die de keuze voor de profielen laten zien.

Maak een cirkeldiagram voor:

* havo jongens
* havo meisjes
* vwo jongens
* vwo meisjes

Je krijgt dus 4 cirkeldiagrammen.

Zorg dat de verschillende profielen telkens dezelfde kleuren krijgen

*Door eerst te selecteren op profiel, daarna op schooltype en tenslotte op geslacht, krijg je alle 16 soorten (categorieën) netjes bij elkaar om te verwerken.*

*Bij deze selectievolgorde heb je alle* ***havo jongens*** *bij elkaar staan enz*

Ook andere volgordes van selectie leveren de verschillende groepen op, maar vaak niet zo netjes geordend. Ga dit na.

De laatste ordening levert de hoofdordening M/V op, binnen de hoofdordening zie vervolgens de eerdere ordeningen terug komen, eerst havo/vwo en daarbinnen weer de profielen. De havo-jongens staan dus mooi bij elkaar.

Als je de groepn bij elkaar hebt staan kun je tellen en een tabel maken waaruit je het cirkeldiagram laat ontstaan. Zoals je al eerder hebt gezien.

Let er wel op dat bij het op volgorde zetten telkens de gehele rij (horizontaal) wordt meegenomen.

**Vragen**:  
Waarom zou je zo’n cirkeldiagram willen maken?  
Welke conclusie kun je trekken uit cirkeldiagrammen die je hebt gemaakt?

**Opdracht 3**

Je hebt nu na opdracht 2 het data bestand zo geordend dat je voor elk van de 16 categorieën (bijvoorbeeld Havo- CM- M ) kunt berekenen wat de gegevens voor de boxplots zijn van de cijfers voor het vak Nederlands.   
Maak de boxplots voor de 8 Havo categorieën.  
Zet deze boxplots bij elkaar in één grafiek.  
**Vragen**:

Waarom zou je zo’n grafische voorstelling willen maken?  
 Welke conclusie kun je trekken uit de grafische voorstelling die je hebt gemaakt?

**Opdracht 4**

Maak nu boxplots van de cijfers van de drie kernvakken: ne, eng, wi voor de leerjaren 1,2 en 3   
Zet deze negen boxplots in één of meerdere figuren, maak **zelf** keuzes.

**Vragen**:

Waarom zou je zo’n grafische voorstelling willen maken?  
 Welke conclusie kun je trekken uit de grafische voorstelling die je hebt gemaakt?

Zet telkens bij de figuren welke conclusie je kunt trekken uit jou gekozen figuur.

**Opdracht 5a**

In een boxplot verdeel je de reeks getallen in vier groepen met elk 25% van de getallen.   
Je hebt in het boekje “Statistiek met Excel” ook geleerd hoe je van een rij getallen de *standaarddeviatie* kunt berekenen.

Veel frequentieverdelingen hebben een nette symmetrische klokvorm, zeker als het veel gegevens betreft. Voorbeelden daarvan zijn de verdeling van de lengtes van een grote groep mensen, de gewichten van volle pakken suiker, de inhoud van een grote groep colaflessen van dezelfde soort en maat, enzovoorts...

Bij dit soort verdelingen kan de standaard deviatie kan gebruikt worden voor een indeling van de groep getallen in zes onderdelen:  
het gemiddelde geven we aan met ‘μ’ en de standaarddeviatie met ‘σ’.

Hier zie je een histogram van de lengtes van een groep soldaten op een bepaalde kazerne.

http://www.mathunited.nl/getresource?blob-key=AMIfv96I40xq2pS6fyTMIC-ZxeMtLk3UgvKc65fzw98TkXs3IcQ7BnjPnRmu496dcpgBJVRyTv8YEzuiG_tIw5s5A1sqihbKD7WW2ZSQMMIj6ur-GMxipdtNdSqjBBOof0XxbVMC8y1AORyY3iwpFeUH1bJNPi_w0PpJDxdZtKApZ55wZTOuZSY


De vuistregel voor de normale verdeling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grenzen | 🡨, μ-2σ | μ-2σ 🡨μ-σ | μ -σ 🡨μ | μ🡪 μ+σ | μ+σ🡪 μ+2σ | μ+2σ, 🡪 |
| % | 2.5% | 13.5% | 34% | 34% | 13.5% | 2,5% |
| soldaten | 🡨,168 | 168-175 | 175-182 | 182-189 | 189-196 | 196,🡪 |

Voor meer uitleg over de normale verdeling zie bv.:

<http://www.math4all.nl/website/view.php?page=overzichten/havo-45-a&repo=m4a&item=introduction>

Voor de boxplot gelden de percentages 25% telkens precies omdat de boxplot zo is gemaakt.

De onderverdeling m.b.v. de μ en de σ geldt ongeveer, maar steeds beter bij grotere bestanden.  
**Vraag/opdracht** :

Bereken het gemiddelde en standaarddeviatie van de cijfers voor het vak Engels in het tweede leerjaar. En ga na of deze onderverdeling ook een beetje klopt voor de cijfers voor het vak Engels in het tweede leerjaar.

**Opdracht 5b**

Geef een schatting voor grenzen van de middelste 68% van de cijfers voor het vak *Engels* in het tweede leerjaar. Gebruik hierbij de systematiek uit opgave 5a.  
Geef een schatting van de grenzen voor de laagste 16% van de cijfers op het van *Engels* in het tweede leerjaar.