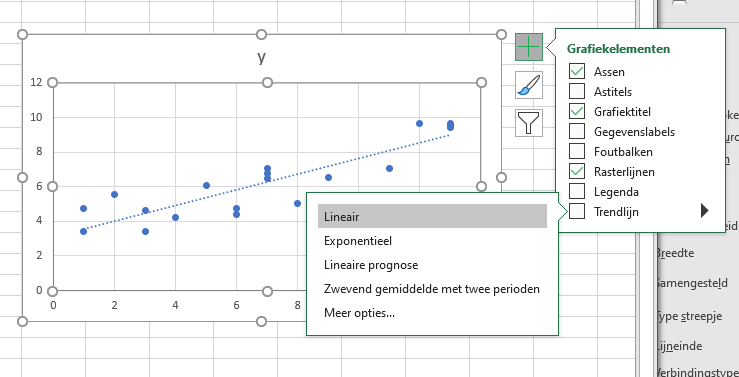
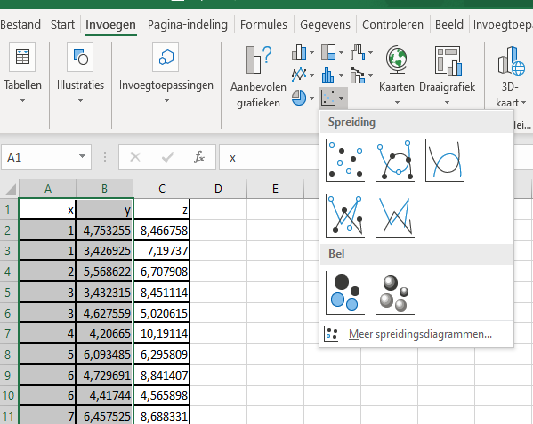
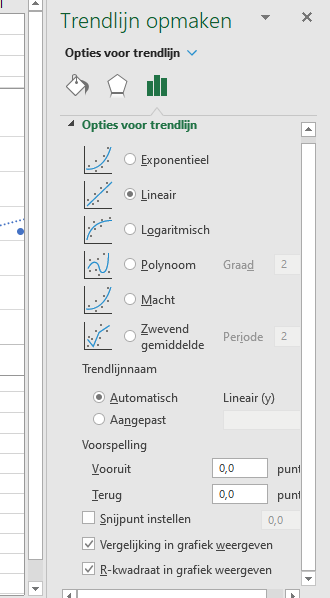
# Appendix: Trendlijnen in Excel

In deze appendix leggen we uit hoe je met de standaard instellingen van Excel trendlijn kan maken in een spreidingsdiagram. Het opstellen van een trendlijn noemt men regressie. Je kunt een best passende rechte lijn laten opstellen, maar ook andere lijnen zijn mogelijk. Hieronder is een voorbeeld gegeven. We spelen dit voorbeeld na.



|  |  |  |
| --- | --- | --- |
| x | y | z |
| 1 | 4,753255 | 8,466758 |
| 1 | 3,426925 | 7,19737 |
| 2 | 5,568622 | 6,707908 |
| 3 | 3,432315 | 8,451114 |
| 3 | 4,627559 | 5,020615 |
| 4 | 4,20665 | 10,19114 |
| 5 | 6,093485 | 6,295809 |
| 6 | 4,729691 | 8,841407 |
| 6 | 4,41744 | 4,565898 |
| 7 | 6,457525 | 8,688331 |
| 7 | 7,048074 | 3,161889 |
| 7 | 6,775916 | 7,86127 |
| 8 | 5,003786 | 7,820007 |
| 9 | 6,525547 | 5,522803 |
| 11 | 7,027179 | 5,368564 |
| 12 | 9,662825 | 5,553985 |
| 13 | 9,629671 | 3,810277 |
| 13 | 9,513795 | 5,118965 |
| 13 | 9,398465 | 5,561814 |

* Start Excel en kopieer de gegevens x, y en z in de tabel naar een leeg werkblad.
* Selecteer de kolommen x en y inclusief de titels
* Kies uit het hoofdmenu **Invoegen** en kies het **spreidingsdiagram** linksboven
* In het plaatje dat je krijgt klik je op de ***plus box*** naast de figuur en kies met het zwarte pijltje naast de optie trendlijn voor **meer opties**.
* Kies voor **vergelijking in grafiek weergeven** en **R-kwadraat in grafiek weergeven**.
* Sleep de box met de regressie gegevens naar een plek waar je die goed kunt lezen.
* Je hebt nu de lijn y = 0,4536x + 3,0991 gekregen en   
  R² = 0,7861
* R² is een maat die aangeeft hoe goed de trendlijn past bij de data. Als alle punten precies op één lijn liggen dan is R² gelijk aan 1, er is dan sprake van volledige correlatie ofwel ieder x bepaalt precies één y. Een R² gelijk aan 0 geeft aan dat voor iedere x er een willekeurige y waarde kan optreden. In dit geval is de a in y=a x + b gelijk aan 0. In de tweede grafiek hieronder zie je een voorbeeld van een trendlijn die nauwelijks een voorspellende waarde heeft, R² is bijna 0.

Bovenstaande lijnen zijn dus gemaakt met behulp van de voeg trendlijn toe. Echter de formule en bijbehorende R² geven nog niet zo goed weer of de lijn er ook werkelijk is. Beter zou zijn als we een betrouwbaarheidsinterval voor de helling zouden hebben. Als helling = 0 in dit interval ligt, dan is er geen sprake van een aantoonbaar lineair verband. Beter is dan om de LIJNSCH(y-waarden;x-waarden; door; nul; met statsistieken)

functie te gebruiken ('lijn schatting'). Deze biedt de mogelijkheid om te bepalen of de gevonden lijn niet op toeval berust.  
Kijk op <https://support.office.com/nl-nl/article/lijnsch-functie-84d7d0d9-6e50-4101-977a-fa7abf772b6d> voor de uitleg. Hieronder is LIJNSCH ingezet om de intervallen voor a te bepalen. Voor de lijnen y = a x + b en z = a x + b

# LIJNSCH(y-waarden;x-waarden; door; nul; met statistieken)

Hieronder staan de resultaten van het gebruik van LIJNSCH(y-waarden;x-waarden; door; nul; met statsistieken) om te bepalen of de lijn al dan niet op toeval berust bij een onbetrouwbaarheid (significantie niveau) α = 0,05 . In het bestand ExampleRegressionData.xlsm kun je simulaties maken met verschillende foutmarges om een gevoel te krijgen wat lineaire regressie te bieden heeft.

Algemeen:

y = range y waarden  
x = range x waarden  
geschatte a = INDEX(LIJNSCH(y;x;1;1);1;1)  
geschatte σ a = INDEX(LIJNSCH(y;x;1;1);2;1)  
geschatte b = INDEX(LIJNSCH(y;x;1;1);1;2)  
geschatte σ b = INDEX(LIJNSCH(y;x;1;1);2;2)  
R²= INDEX(LIJNSCH(y;x;1;1);3;1)  
geschatte σ y = INDEX(LIJNSCH(y;x;1;1);3;2)  
F toets waarde fw = INDEX(LIJNSCH(y;x;1;1);4;1)  
vrijheidsgraden vg = INDEX(LIJNSCH(y;x;1;1);4;2)  
P( F > F toets waarde) =1-F.VERD(fw;AANTAL(y)-vg-1;vg;1)  
Tinv(α;vrijheidsgraden) = TINV(α;vg)  
Geschaald naar Std fout (gf) = TINV(α;vg) \* geschatte σ a   
bovengrens a = geschatte a + gf  
ondergrens a = geschatte a - gf  
Betrouwbaarheidsinterval: ondergrens < a < bovengrens

Als P( F > F toets waarde) < α dan is er een lineair verband anders niet.

Als a=0 in het betrouwbaarheidsinterval ligt dan is er geen lineair verband.

|  |  |
| --- | --- |
| y |  |
| geschatte a | geschatte σ a |
| 0,453555546 | 0,057383464 |
| geschatte b | geschatte σ b |
| 3,09910255 | 0,453555546 |
| R² | geschatte σ y |
| 0,786088971 | 0,990242199 |
| F | vrijheidsgraden |
| 62,47229297 | 17 |
| P( F > 62,47) | 4,30076E-07 |
| P( F > 62,47) < α |  |
| Tinv(α;vrijheidsgraden) | 2,109815578 |
| Geschaald naar Std fout | 0,121068526 |
| bovengrens a | 0,574624072 |
| ondergrens a | 0,332487021 |
| Betrouwbaarheids interval: 0,332 < a < 0,575 | |

|  |  |
| --- | --- |
| z |  |
| geschatte a | geschatte σ a |
| -0,234572191 | 0,098933095 |
| geschatte b | geschatte σ b |
| 8,15446742 | 0,786568043 |
| R² | geschatte σ z |
| 0,248510118 | 1,707246635 |
| F | vrijheidsgraden |
| 5,621728389 | 17 |
| P( F > 5,62) | 0,029821389 |
| P( F > 62,47) < α |  |
| Tinv(α;vrijheidsgraden) | 2,109815578 |
| Geschaald naar Std fout | 0,208730584 |
| bovengrens a | -0,025841606 |
| ondergrens a | -0,443302775 |
| Betrouwbaarheids interval: -0,443 < a < -0,026 | |