

Korte uitleg van twee veelvoorkomende statistische toetsen

Veel wetenschappelijke hypothesen kunnen statistisch worden getoetst. Aan de hand van een statistische toets beslis je of een hypothese waar is. Dit geldt ook voor de twee hypothesen in de volgende filmpjes: [filmpje 1](#), [filmpje 2](#) (vervolg op filmpje 1).

Hypothese 1 luidde:

“Voetbalsupporters plegen meer geweld naarmate ze sterker verbonden zijn met een groep waar een delinquente subcultuur heerst.”

Dit is een hypothese naar *samenhang*. Er is sprake van samenhang tussen twee eigenschappen van mensen als bij toename van de ene eigenschap, de ander over het algemeen toeneemt (positieve samenhang) of afneemt (negatieve samenhang). Hieronder zullen we allereerst kort laten zien hoe je een hypothese met betrekking tot samenhang kunt toetsen in Excel.

Hypothese 2 luidde:

“Door het stadionverbod verplaatst het geweld zich van het stadion naar daarbuiten.”

Dit is een hypothese naar het verschil in *gemiddelde* tussen twee groepen op één bepaalde eigenschap. Er is een verschil in gemiddelde als de ene groep over het algemeen hoger scoort op de eigenschap dan de andere groep. Ook het toetsen van deze hypothese wordt hieronder kort toegelicht met Excel.

Voor het uitleggen van de toetsen van de hypothesen gaan we uit van andere hypothesen dan in het filmpje. Deze voorbeelden zijn gemakkelijker, maar de hypothesen zijn wetenschappelijk niet interessant; we weten van de hypothesen namelijk al dat ze waar zijn. Dat is niet erg, want het gaat ons hier alleen maar om de uitleg van hoe je de hypothesen statistisch kunt toetsen. We hopen dat je aan de hand van de voorbeelden zelf in staat bent om de twee hypothesen statistisch te toetsen. Als je meer wilt weten over de achtergrond van statistisch toetsen in het algemeen of deze twee toetsen in het bijzonder, dan verwijzen we je naar een boek over statistiek of je wiskundeleraar.

Uitleg toets van hypothese naar samenhang

De hypothese naar samenhang die we statistisch gaan toetsen is: “Gewicht van mensen neemt toe naarmate hun lengte toeneemt.” Stel dat we de gegevens met betrekking tot hun lengte en gewicht hebben van 12 personen. De eerste stap is dat we deze gegevens invoeren in Excel.

1. Open het bestand “toets samenhang.xls” in Excel. In de vierde rij en kolommen B, C, en D, zie je dat de eerste persoon een lengte heeft van 180 cm en een gewicht van 82 kg. Deze gegevens dien je zelf in te voeren, voor alle personen. Zo zie je dat rijen 4 tot en met 15 de gegevens bevatten van alle 12 personen.
2. De volgende stap is het berekenen van de samenhang tussen de lengte en het gewicht van de personen. De meest gebruikte maat voor samenhang is de *correlatie*. De correlatie bereken je als volgt. Je gaat op een lege cel staan. Dan ga je naar de formuleregel bovenaan, (waarvoor staat f_x), en tikt daarin: “=CORREL(C4:C15;D4:D15)”. Als je dan op Enter drukt berekent Excel voor jou de correlatie. Je krijgt dan de waarde 0,869 (afgerond), zoals je die ziet in de geel gearceerde cel F6. Als de waarde van de correlatie positief is, zoals hier, dan is de samenhang positief. Als je op cel F6 gaat staan zie je precies de hierboven gegeven commandoregel “=CORREL(C4:C15;D4:D15)”.

Om kort op de commandoregel in te gaan, “=” betekent “Bereken in deze cel”, “CORREL” is het commando voor de correlatie, “C4:C15” verwijzen naar de scores van de personen op de ene eigenschap in rijen 4 tot en met 15 van kolom C, en “D4:D15” verwijzen naar de scores van de personen in rijen 4 tot en met 15 van kolommen D op de andere eigenschap. Let op: in ons voorbeeld gaan we uit van een groep van 12 personen. Als je een groep hebt van 20 personen moet je de commandoregel zelf aanpassen en veranderen in “=CORREL(C4:C23;D4:D23)”. Let ook op het gebruik van de dubbele punt “:” en de semicolon “;” in de commandoregel.

3. De volgende stap is het uitvoeren van de toets van de hypothese. Deze hypothese wordt uitgevoerd met een zogenaamde *t*-toets. Eerst bereken je dan de toetswaarde met de formule die je vindt als je met je cursor gaat staan op de groen gearceerde cel G6. In deze formule gebruiken we de correlatie (hier in F6) en het totaal aantal personen (hier in B15). “SQRT” staat voor de wortel. Als je op Enter drukt levert dit de waarde 5,56 op (afgerond).
4. Als laatste bepalen we dan nog de kans op deze toetswaarde. Dat wordt gedaan in de paars gearceerde cel H6. Als je op H6 gaat staan zie je dat je Excel een commando “T.DIST.2T” laat uitvoeren, waar je zowel de toetswaarde (G6) meegeeft als ook het aantal personen verminderd met 2 (B15-2). Na drukken op Enter wordt in dit voorbeeld een kans verkregen van 0,00024 (afgerond).

De interpretatie van deze kans komt zeer nauw. De kans betekent hier:

“Als in de populatie van mensen de correlatie tussen hun lengte en gewicht gelijk is aan 0 (geen samenhang), dan is de kans op een correlatie gelijk aan 0,869 of extremer in een steekproef van 12 personen uit deze populatie, gelijk aan 0,00024.”

De gewoonte is om als deze kans kleiner is dan 0,05 de hypothese over samenhang aan te accepteren. Omdat de kans in ons geval kleiner is dan 0,05, accepteren we dus de hypothese, en concluderen we dat gewicht van mensen neemt toe naarmate hun lengte toeneemt. Als de kans kleiner zou zijn geweest dan 0,05, dan accepteren we de hypothese niet, en zouden we concluderen dat we niet kunnen uitsluiten dat er geen correlatie is tussen lengte en het gewicht van mensen.

Uitleg toets van hypothese naar gemiddelde

De hypothese naar gemiddelde die we statistisch gaan toetsen is: “de gemiddelde lengte van mannen is ongelijk aan de gemiddelde lengte van vrouwen”. Stel dat we de gegevens hebben van 20 personen met betrekking tot hun lengte en geslacht.

1. De eerste stap is dat we deze gegevens invoeren in Excel. Open het bestand “toets gemiddelden.xls” in Excel. In de vierde rij en kolommen B, C, en D, zie je dat de eerste persoon een lengte heeft van 180 cm (kolom C) en man is (kolom D). De gegevens van de 20 personen dien je zelf in te voeren. Zo zie je dat rijen 4 tot en met 23 de gegevens bevatten van alle 20 personen. Tien personen zijn man, tien personen zijn vrouw.

Verder zie je in de vierde rij, in kolommen E en F, de gemiddelden van beide groepen (mannen en vrouwen) staan. Ga op de velden E4 en F4 staan als je wilt weten hoe deze door Excel worden berekend.

2. Na het invoeren van de gegevens kun je de toets van de hypothese uitvoeren. Deze hypothese wordt uitgevoerd met een zogenaamde *t*-toets.¹ Met één commando geeft Excel je meteen de kans die je nodig hebt je uiteindelijke conclusie te trekken. Als je op de paars gearceerde cel G6 gaat staan zie je dit commandoregel; “=T.TEST(C4:C13;C14:C23;2;2)”.

Om kort op de commandoregel in te gaan, “=” betekent “Bereken in deze cel”, “T.TEST” is het commando voor de *t*-toets van de hypothese over de vergelijking van gemiddelden van twee groepen, “C4:C13” verwijst naar de scores van de personen in de ene groep (mannen) op lengte, “C14:C23” verwijst naar de scores van de personen in de andere groep (vrouwen) op lengte. Let op: in ons voorbeeld gaan we uit van twee groepen van tien personen. Als je twee groepen hebt van 20 personen moet je de commandoregel zelf aanpassen en veranderen “=T.TEST(C4:C23;C24:C43;2;2)”. Let op het gebruik van de dubbele punt “:” en de semicolon “;” in de commandoregel.

3. Als je op cel G6 staat en op Enter drukt geeft dit de kans 0,0029 (afgerond). De interpretatie van deze kans komt zeer nauw. De kans betekent hier:

“Als er geen verschil is in de gemiddelde lengte van mannen en vrouwen, dan is de kans op een verschil in lengte zoals hier in de steekproeven van 10 mannen en 10 vrouwen, gelijk aan 0,0029.”

De gewoonte is om als deze kans kleiner is dan 0,05 de hypothese over gemiddeld verschil te accepteren. Omdat de kans in ons geval kleiner is dan 0,05, accepteren we dus de hypothese, en concluderen we dat de gemiddelde lengte van mannen groter is dan van vrouwen. Als de kans

¹ Het betreft een andere *t*-toets dan de *t*-toets die je gebruikt bij het toetsen van een hypothese over een correlatie.



kleiner zou zijn geweest dan 0,05, dan accepteren we de hypothese niet, en zouden we concluderen dat we niet kunnen uitsluiten dat er geen verschil is in gemiddelde lengte tussen mannen en vrouwen.