

Concept “Significant verschil”

Om significantie toe te lichten, vertrekken we van een populatie, genoteerd als P , een sociodemografische variabele (kortweg “socio-demo”) die bestaat uit een aantal klassen die we als $K_{i|1 \leq i \leq t}$ noteren en een vooraf gekozen KPI. Met een significantieanalyse wordt nagegaan in hoeverre er voor elke klasse uit de socio-demo een significant gemiddeld verschil optreedt in KPI scores tussen deze klasse en alle overige klassen samen.

Het begrip “significant verschil” dient hierbij in een statistisch kader te worden opgevat. Als generiek voorbeeld beschouwen we als populatie de werknemers uit een bepaald bedrijf en als socio-demo “beroep” met klassen arbeiders, bedienden en kaderleden. In een significantieanalyse voor een welbepaalde KPI wordt dus nagegaan in hoeverre de KPI-score voor arbeiders gemiddeld verschilt van die van bedienden en kaderleden, de KPI-score voor bedienden gemiddeld verschilt van de score van arbeiders en kaderleden en in hoeverre tenslotte de score van de kaderleden gemiddeld afwijkt van die van arbeiders en bedienden.

Bij elke stap in de analyse dient dus bepaald te worden in hoeverre de KPI-scores voor twee gescheiden subgroepen uit de populatie (waarbij de eerste wordt gevormd door de vooraf gekozen KPI klasse en de tweede door de overige klassen samen te nemen) gemiddeld verschillen. Deze groepen worden genoteerd als G_1 en G_2 . Aangezien in de statistiek altijd gewerkt wordt met exacte getallen, gebruiken we populatiegemiddelden μ_{G_1} en μ_{G_2} of populatiemedianen η_{G_1} en η_{G_2} om de gemiddelde score per groep te bepalen. Een populatiegemiddelde μ_{G_1} of μ_{G_2} is hierbij de gemiddelde KPI-score voor alle elementen uit de populatie die tot de groep G_1 of G_2 behoren. Een populatie mediaan η_{G_1} of η_{G_2} is de waarde die kleiner is dan de helft van de elementen uit een groep G_1 of G_2 . Bij elke stap van de analyse dient dus voor elke KPI klasse bepaald te worden in hoeverre μ_{G_1} en μ_{G_2} of η_{G_1} en η_{G_2} van elkaar verschillen. (Het feit of het verschil tussen μ_{G_1} en μ_{G_2} of η_{G_1} en η_{G_2} in rekening wordt gebracht zal afhangen van de grootte van de groepen G_1 en G_2 .)

In de statistiek noemt men waarden berekend uit een populatie ook wel populatieparameters. In de praktijk zijn zulke populatieparameters meestal niet exact gekend. Een populatie is immers in vele gevallen omschreven door middel van conceptuele terminologie bijvoorbeeld “de Belgische beroepsbevolking” of bevat teveel elementen zodat deze niet allen in rekening gebracht kunnen worden bijvoorbeeld “alle Belgische inwoners”. Om schattingen voor populatieparameters te kunnen maken wordt daarom in de statistiek vertrokken van een steekproef. Dit is een verzameling van een aantal door toeval geselecteerde objecten uit de populatie die representatief is voor deze populatie. Meestal noteert men een steekproef die representatief is voor een populatie P met het symbool S_p .

In het voorgaande voorbeeld zou de steekproef een aantal ad random geselecteerde werknemers uit het bedrijf bevatten. Op basis van deze steekproef kunnen dan schattingen gedaan worden voor de waarden van de populatieparameters. Zo’n schatting noemt men ook wel een statistiek. Van essentieel belang is echter wel dat elke statistiek een geassocieerde kansverdeling bezit. Dit is een wiskundige functie die met elk getal de kans associeert dat de statistiek deze waarde aanneemt.

Voor de praktische berekeningen in de *ZebraZone Toolset* wordt dus uitgegaan van een steekproef S_p die representatief is voor P . Een schatting van μ_{G_1} zal dan berekend worden als de som van alle KPI-scores van de elementen uit de steekproef die tot groep G_1 behoren gedeeld door het totale aantal elementen uit de steekproef die tot G_1 behoren. Men noemt dit dan ook wel het steekproefgemiddelde voor de KPI voor subgroep G_1 uit S_p en noteert dit als m_{G_1} . Analogie stelt m_{G_2} het steekproefgemiddelde voor groep G_2 voor.

Om een schatting voor η_{G_1} te bepalen, worden de elementen uit S_p die tot G_1 behoren eerst geordend van klein naar groot. Indien n_1 oneven is, wordt de schatting gegeven door het $(n_1 + 1)/2$ -ste element uit deze geordende rij. Indien n_1 even is, gebruiken we het gemiddelde van het $n_1/2$ -ste en $(n_1 + 2)/2$ -ste element uit deze rij als schatting voor de populatiemediaan. Dit getal noemt men ook wel de steekproefmediaan voor de elementen uit S_p die tot G_1 behoren en men noteert dit als med_{G_1} . Analogie stelt med_{G_2} de schatting voor van η_{G_2} . In ons generiek voorbeeld stelt dus $m_{Arbeiders}$ het steekproefgemiddelde voor van de KPI scores voor alle arbeiders en $med_{Arbeiders}$ de mediaan voor de KPI scores van alle arbeiders in de steekproef.

In het algemeen worden statistieken naast hun functie als schatting tevens gebruikt om populatieparameters te vergelijken. De significantieberekening in de *ZebraZone Toolset* volgt deze filosofie volledig. Indien μ_{G_1} en μ_{G_2} exact aan elkaar gelijk zijn, zullen niet noodzakelijk m_{G_1} en m_{G_2} exact dezelfde waarde aannemen. Steekproefgemiddelden zijn immers slechts benaderingen van populatiegemiddelden. Natuurlijk verwacht je dat indien de populatiegemiddelden gelijk zijn, dan ook het verschil tussen m_{G_1} en m_{G_2} relatief klein zal zijn. Dezelfde logica is volledig van toepassing op de populatiemedianen η_{G_1} en η_{G_2} met schattingen med_{G_1} en med_{G_2} . Om de grens te bepalen vanaf wanneer je uit het gemeten verschil tussen m_{G_1} en m_{G_2} of med_{G_1} en med_{G_2} kan besluiten dat de populatieparameters μ_{G_1} en μ_{G_2} of η_{G_1} en η_{G_2} tevens gelijk zijn, kan gebruik worden gemaakt van een aantal statistische procedures. Zo een procedure noemt men ook wel een statistische test.

Aangezien hierbij wordt uitgegaan van schattingen van populatieparameters, is zo'n heuristiek niet voor 100 procent sluitend. Er bestaat namelijk altijd een kans dat hoewel μ_{G_1} en μ_{G_2} gelijk zijn dit niet weerspiegeld wordt in de gemeten steekproef en je dus foutief besluit dat μ_{G_1} en μ_{G_2} verschillend zijn. Een dergelijke fout wordt kortweg een fout van type I genoemd. Daarnaast zou het natuurlijk ook kunnen dat een verschil tussen μ_{G_1} en μ_{G_2} niet in de test weerspiegeld wordt. Je besluit dat de populatiegemiddelden gelijk zijn, impliceert dan wat men aanduidt als een fout van type II. De kans op het maken van een fout van type I bij een test wordt significantieniveau genoemd (genoteerd als α). De kans op het maken van een fout van type II noteert men kortweg als β . Het getal $(1-\beta)$ stelt dus de kans voor dat je het verschil tussen μ_{G_1} en μ_{G_2} of η_{G_1} en η_{G_2} ook in de test terugvindt. In de statistische theorie wordt dit aangeduid als de power van de test.

In praktisch alle gevallen kan je het significantieniveau van de test op voorhand bepalen maar niet de power. Power en significantieniveau zijn echter niet onafhankelijk van elkaar. Er geldt namelijk algemeen dat hoe hoger het significantie niveau (lagere α -waarde) is, hoe lager de power van de test zal zijn. Dus een test met significantieniveau 0.05 zal over het algemeen minder power hebben dan een test met een α -waarde van bijvoorbeeld 0.03.

Er dient dus een keuze voor α te worden gemaakt zodanig dat de power groot genoeg is. In de praktijk wordt de waarde 0.05 veelal voorop gesteld als significantieniveau. De analyses in de *ZebraZone Toolset* zijn conform met deze richtlijn.

Naast een keuze in significantieniveau is er bij de significantieanalyse tevens de keuze van het soort test die een cruciale rol zal spelen. We kiezen voor de parametrische Student t-Test of de niet-parametrische Wilcoxon's Rank-Sum Test. De t-Test steunt hierbij op een aantal voorwaarden op de kansverdeling van de steekproefgemiddelden terwijl de Wilcoxon's Rank-Sum test zulke eisen niet stelt. De keuze welke test bij welke situatie past, wordt hoofdzakelijk bepaald aan de hand van deze voorwaarden en de powerwaarde van de test.

□ ***Eerste geval: beide groepen G_1 en G_2 bevatten meer dan 30 elementen.***

In dit geval is de "Centrale Limietstelling"¹ van toepassing die impliceert dat de kansverdelingen van zowel m_{G_1} als m_{G_2} de Gauss klok curve² volgen. Dit is juist een cruciale voorwaarde om een Student t-Test te kunnen uitvoeren. Bovendien zal de Student t-Test in dit geval een hogere power waarde impliceren als de Wilcoxon's Rank-Sum Test. Vandaar dat er dan ook geopteerd wordt voor de t-test met een significantieniveau van 0.05 of 5%. Globaal gezien wordt met deze test nagegaan in hoeverre de populatiegemiddelden voor de groepen G_1 en G_2 van elkaar verschillen. Aan de hand van m_{G_1} en m_{G_2} en gegevens uit de steekproef berekenen we een getal t en een interval I. Indien t tot dit interval behoort, besluiten we dat de populatiegemiddelden μ_{G_1} en μ_{G_2} gelijk zijn. In het andere geval besluiten we het tegenovergestelde.

□ ***Tweede geval: een van beide groepen G_1 en/of G_2 bevatten minder dan 30 elementen.***

De "Centrale Limietstelling" is in dit geval niet meer van toepassing en de Student t-Test is hierdoor niet meer betrouwbaar. We opteren in dit geval dan ook om de niet parametrische alternatieve Wilcoxon's Rank-Sum Test te gebruiken die niet afhangt van voorwaarden op de kansverdeling van de steekproefgemiddelden. Met de Wilcoxon's Rank-Sum test wordt nagegaan in hoeverre de populatiemedianen η_{G_1} en η_{G_2} verschillen. Aan de hand van de steekproefgegevens wordt een getal W en een interval I berekend. Als W tot het interval I behoort, besluiten we dat de medianen η_{G_1} en η_{G_2} gelijk zijn en in het andere geval zullen η_{G_1} en η_{G_2} dus verschillend zijn. Het significantieniveau wordt hierbij ingesteld op 0.05 of 5%.

¹ Een van de fundamentele stellingen uit de statistiek. Ruwweg zegt de stelling dat voor een steekproef S van voldoende grootte de kansverdeling van het steekproefgemiddelde juist de Gauss klok curve is.

² De Gauss klok curve is de belangrijkste kansverdeling uit de statistiek die voorkomt in tal van natuurlijke processen. Ze is in feite de wiskundige vertaling van het fenomeen van middelmaat. Zo zijn er bijvoorbeeld heel veel mensen met een gemiddeld gewicht, een minder aantal met een laag gewicht of hoog en een paar uitzonderingen met een extreem laag of hoog gewicht.